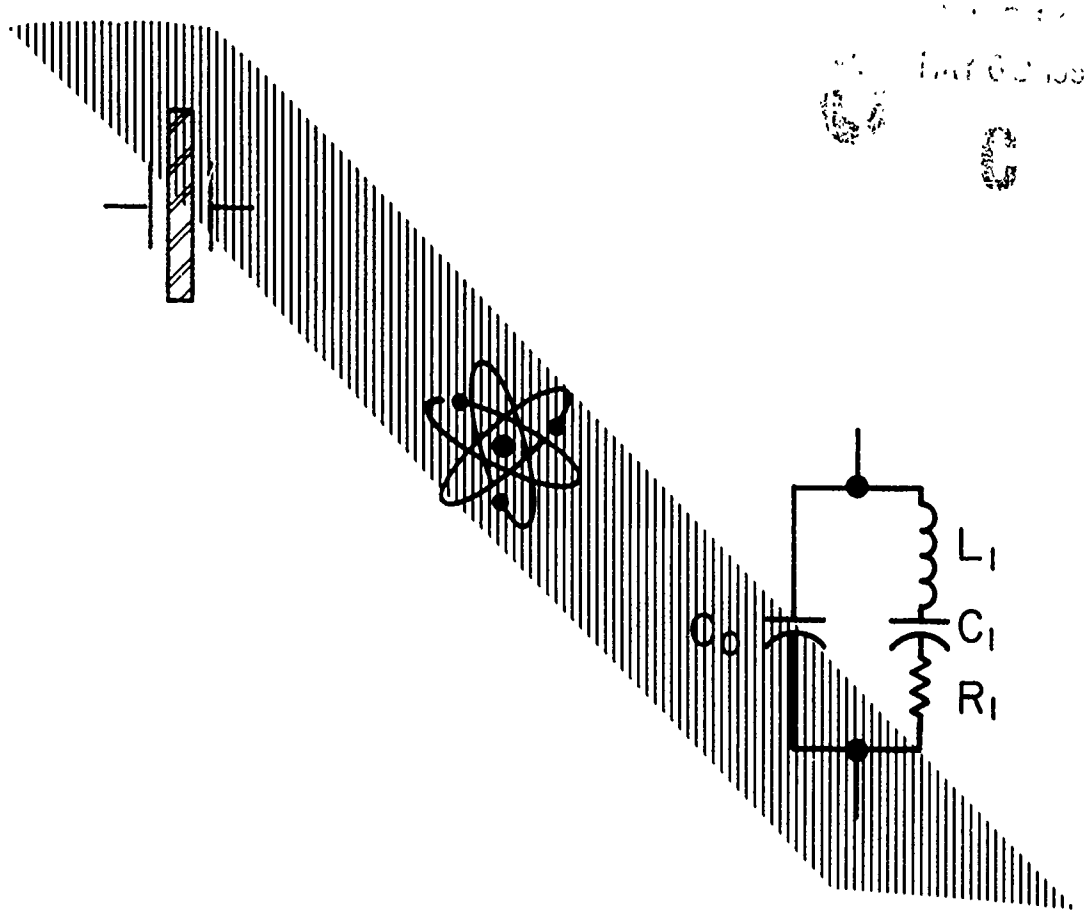
(2)

Proceedings of the **43**rd Annual

Symposium on Frequency Control

1989



31 May - 2 June 1989

89CH2690-6

58-60781

NOTE TO DTIC for CITATION

Suggested Descriptors:

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

(Note: Headings in Table of Contents are good to use for Abstract information.)

Proceedings of the 43rd Annual

Symposium on Frequency Control

1989

Cosponsored by the

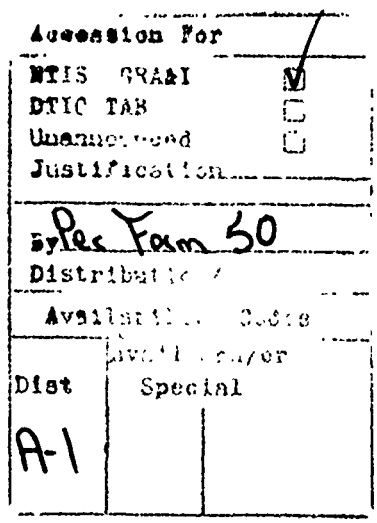
U.S. Army
Electronics Technology and
Devices Laboratory

and

The Institute of Electrical and
Electronics Engineers, Inc.
Ultrasonics, Ferroelectrics and
Frequency Control Society

IEEE Catalog No. 89CH2690-6

Library of Congress No. 58-60781



31 May - 2 June 1989

Denver Marriott Hotel
City Center
Denver, Colorado

SYMPOSIUM CHAIRMEN

General Chairman ..... David W. Allan
Technical Program Chairman ..... Thrygve R. Meeker
Finance Chairman ..... Thomas E. Parker
Local Arrangements Chairman..... Samuel R. Stein
Editorial Chairman ..... John R. Vig
Publicity Chairman ..... Raymond L. Filler

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Michael Driscoll ..... Westinghouse Systems Center
Larry Halliburton ..... Oklahoma State University
Donald Malocha ..... University of Central Florida
Lauren Rueger ..... Applied Physics Laboratory

TECHNICAL PROGRAM COMMITTEE

David Allan, National Institute of Standards and Technology
Arthur Ballato, U.S. Army LABCOM
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, National Research Council of Canada
Janet Brown, Fischer Controls International, Inc.
Mary Chiu, Johns Hopkins University, Applied Physics Laboratory
Leonard Cutler, Hewlett-Packard Company
Michael Driscoll, Westinghouse Systems Center
Errol EerNisse, Quartztronics, Inc.
Raymond Filler, U.S. Army LABCOM
Marvin Frerking, Rockwell International
Jean-Jacques Gagnepain, CNRS-LPMO, France
Edward Garber, TRW
Michael Garvey, Frequency & Time Systems, Inc.
Helmut Hellwig, National Institute of Standards and Technology
James Pon-Wei Hou, Quartz Frequency Technology, Ltd., Japan
Charles Jensik, Piezo Crystal Company
Gary Johnson, Sawyer Research Products, Inc.
Robert Kinsman, Motorola, Inc.
William Klepczynski, U.S. Naval Observatory
Jack Kusters, Ball Corporation
Theodore Lukaszek, U.S. Army LABCOM
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Joel Martin, Oklahoma State University
Thrygve Meeker, Consultant
Yuzo Nakazawa, Toyo Communication Equipment Company, Japan
Thomas Parker, Raytheon Research Division
Robert Peach, GEC Research Ltd., England
Victor Reinhardt, Hughes Aircraft Company
Gerald Roberts, General Electric Company
Vincent Rosati, U.S. Army LABCOM
Lauren Rueger, Johns Hopkins University
Stanley Schodowski, U.S. Army LABCOM
Tadashi Shiosaki, Kyoto University, Japan
Robert Smythe, Piezo Technology, Inc.
Samuel Stein, Ball Corporation
Richard Sydnor, Jet Propulsion Laboratory
John Vig, U.S. Army LABCOM
Rolf Weglein, Hughes Aircraft Company
Joseph White, Naval Research Laboratory
Gernot Winkler, U.S. Naval Observatory
Nicholas Yannoni, U.S. Air Force/RADC

TECHNICAL SESSION CHAIRMEN

GRAVITY
S. Stein, Ball Communication Systems Division

PROCESSING
J. Vig, U.S. Army LABCOM

MICROWAVE OSCILLATORS
M. Driscoll, Westinghouse Systems Center

MICROWAVE RESONATORS & OSCILLATORS
R. Weglein, Hughes Aircraft Company

RESONATOR EFFECTS AND PHASE NOISE
R. Smythe, Piezo Technology, Inc.

ACCELERATION SENSITIVITY
A. Ballato, U.S. Army LABCOM

CHARACTERIZATION OF FREQUENCY SOURCES
L. Cutler, Hewlett Packard Co.

ATOMIC CLOCKS
G. Winkler, U.S. Naval Observatory

QUARTZ MATERIAL
G. Johnson, Sawyer Research Products

MATERIALS AND CONSTANTS
J. Gualtieri, U.S. Army LABCOM

TIME & FREQUENCY MEASUREMENT
J. White, Naval Research Laboratory

CRYSTAL OSCILLATORS
C. Jensik, Piezo Crystal Company

MEASUREMENTS & STATISTICAL ANALYSIS
R. Garvey, Frequency & Time Systems, Inc.

CRYSTAL OSCILLATORS
M. Frerking, Rockwell International

CRYSTAL OSCILLATORS
W. Smith, AT&T Bell Laboratories (retired)

TIME & FREQUENCY TRANSFER
V. Reinhardt, Hughes Aircraft

PIEZOELECTRIC RESONATORS & FILTERS
F. EerNisse, Quartztronics, Inc.

SAW DEVICES AND PHASE NOISE
E. Garber, TRW

ENVIRONMENTAL EFFECTS & TESTING
H. Hellwig, National Institute of Standards and Technology

TESTING AND MEASUREMENT
L. Maleki, Jet Propulsion Laboratory

TABLE OF CONTENTS

ATOMIC FREQUENCY STANDARDS

TIME AND FREQUENCY TRANSFER, CLOCKS, GRAVITATIONAL WAVE EXPERIMENTS

ENVIRONMENTAL EFFECTS, MEASUREMENT, TESTING, STATISTICS

ACCELERATION SENSITIVITY

MATERIAL GROWTH AND PROPERTIES

PIEZOELECTRIC RESONATORS, FILTERS, SAW DEVICES

RESONATOR PROCESSING

1989 AWARD WINNERS

The Cady Award

The Cady Award was presented to Darrell E. Newell of the Northern Illinois University, "for contributions to the development of the Temperature Compensated Crystal Oscillator and for stimulation and education of students in the field of frequency control." The award was presented by Marvin Frerking, Rockwell International.

The Rabi Award

The Rabi Award was presented to Leonard S. Cutler of the Hewlett Packard Co., "for consistent technical and managerial contributions to the development of atomic cesium, rubidium, and mercury ion frequency standards." The award was presented by Charles Adams, Hewlett Packard Co.

The Sawyer Award

The Sawyer Award was not presented due to insufficient nominations.

Leonard S. Cutler, Rabi Award winner; Darrell E. Newell, Cady Award winner; after the award presentation.

43rd Annual Symposium on Frequency Control - 1989

RESONATOR SELF-TEMPERATURE-SENSING USING A DUAL-HARMONIC-MODE CRYSTAL OSCILLATOR

Stanley S. Schodowski

U.S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

Summary

A thermometry method employing resonator self-temperature-sensing is described. Self-temperature-sensing, which provides for the resonator to determine its own temperature as opposed to using external temperature sensors, permits the realization of high-accuracy Microcomputer-Compensated Crystal Oscillators (MCXO). Previously, the only means available for resonator self-temperature-sensing was the b-mode of an SC-cut crystal when used in a dual b- and c-mode oscillator. However, the b-mode has not proved to be reliable for application in wide temperature range MCXO.

The new method uses a harmonically related pair of c-modes of an SC-cut that is simultaneously excited in a dual-mode oscillator, and then combined to produce a thermometric frequency having a monotonic and nearly linear dependence on temperature. It avoids the activity dip problems of the b-mode and overcomes limitations of using external sensors that are subject to thermal lag and temperature gradients. This method provides a means of resonator temperature measurement with high accuracy, limited only by the stability difference between the c-mode frequencies. The concept for this dual c-mode thermometry method is presented and its implementation in MCXO discussed.

Introduction

Intrinsic to all high-performance quartz frequency- and- time sources is a precision temperature-sensing device, i.e., a thermometer, which detects crystal temperature to effect either temperature compensation or temperature control. The conventional method for sensing crystal temperature in Temperature Compensated Crystal Oscillators (TCXO), for example, makes use of a thermistor, or thermistors, placed in close proximity to the resonator. This method suffers from inaccuracies due to thermal lag stemming from a difference in crystal and thermistor effective thermal time constants, thermal gradients, and thermistor aging. To overcome these limitations, Kusters [1], at the 1978 Frequency Control Symposium, proposed using the b-mode of an SC-cut for temperature sensing in a dual b- and c-mode oscillator. However, excessive activity dips associated with the b-mode, subsequently reported by Vig [2] and Baltzer [3], have just about eliminated this method from current consideration in wide temperature range TCXO.

In this paper a resonator self-temperature-sensing method [4] using a pair of harmonically related c-modes in a dual c-mode oscillator is presented. By employing an SC-cut, the c-modes are almost guaranteed to be free of activity dips over the full operating temperature range. The stress and thermal-transient-compensated characteristics of the SC-cut can then be exploited to their fullest extent.

The dual c-mode method overcomes the limitations of existing temperature sensing methods, and makes possible a new generation of higher accuracy crystal oscillators. It is particularly suited for application in MCXO to provide 10 to 100 times improvement in frequency-temperature stability when compared to conventional TCXO [5-7].

The Harmonic Effect

Rotation of the apparent angle for an overtone resonance frequency-temperature (f-T) characteristic relative to that for its fundamental, as shown in Figure 1, has been known for over 35 years. Ballato [8], in 1975, explained the apparent angle change, i.e., harmonic effect, to be due to a factor G, or Onoe function, which multiplies the temperature coefficient of coupling. The effect gets progressively smaller with increasing harmonic number. It manifests

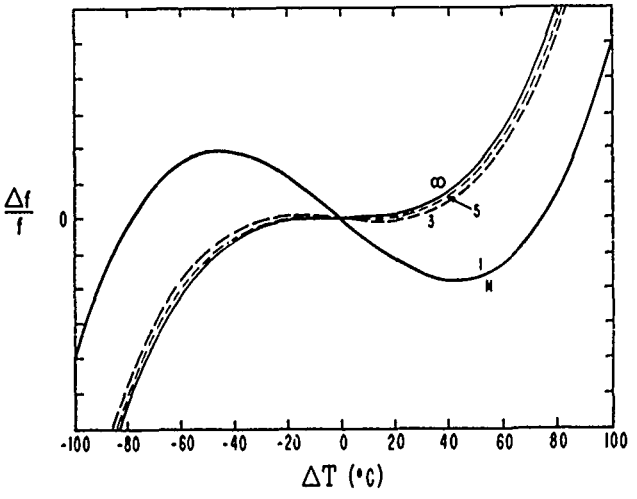


Figure 1. Frequency-temperature-harmonic characteristics of temperature stable resonators.

itself as a change in the first-order, i.e., linear, temperature coefficient of the cubic function:

$$\frac{\Delta f_M}{f_M} = a_M\Delta T + b_M\Delta T^2 + c_M\Delta T^3 \qquad (1)$$

where the normalized frequency $\Delta f_M/f_M$ and the difference temperature ΔT are referenced at the inflection temperature and a_M, b_M, and c_M are the first-, second-, and third-order temperature coefficients of frequency at harmonic number M. This "harmonic effect" provides the basis for the self-temperature-sensing method.

The Thermometric Beat Frequency

By simultaneously exciting a pair of harmonically related c-modes of a temperature-stable resonator in a dual-mode oscillator and combining their signals (Figure 2), a beat frequency (f_β) is obtained which can be used to compensate either one of the generated c-mode signals. Several means for combining the signals are possible.

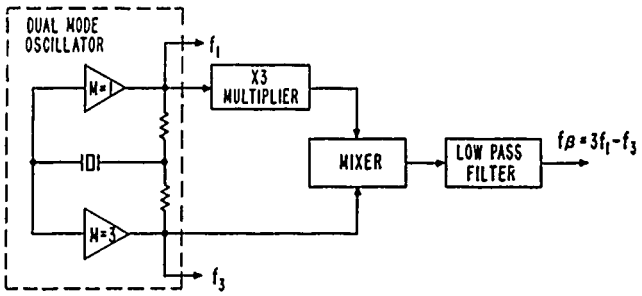


Figure 2. Illustration of the dual c-mode thermometry method.

In this illustration, the lower harmonic c-mode (M = 1), i.e., the fundamental frequency f_1, is multiplied by three and then mixed with the higher harmonic c-mode (M = 3), i.e., the third overtone frequency f_3 to obtain a difference product, or beat frequency f_β, which is

$$f_\beta = 3f_1 - f_3. \qquad (2)$$

The beat frequency may be described in normalized form as:

$$\frac{\Delta f_\beta(\upsilon)}{f_\beta} = \frac{3}{3-n}\frac{\Delta f_1(\upsilon)}{f_1} - \frac{n}{3-n}\frac{\Delta f_3(\upsilon)}{f_3} \qquad (3)$$

where n is the noninteger ratio of the frequencies for the harmonic pair at the reference temperature, i.e., $n = f_3/f_1$, and where υ is any variable, such as supply voltage, load, temperature, etc. A noninteger ratio for n is conveniently available, since the frequencies of harmonic modes of actual resonators are rarely exact multiples of each other. Values of n ranging from 2.94 to 2.98 were found to be typical for a variety of SC-cut designs inspected [9].

Using the f-T dependence of the two modes as given by (1), Eq, (3) can be rewritten solely as a function of temperature

$$\frac{\Delta f_\beta(T)}{f_\beta} = \frac{3a_1-na_3}{3-n}\Delta T + \frac{3b_1-nb_3}{3-n}\Delta T^2 + \frac{3c_1-nc_3}{3-n}\Delta T^3, \qquad (4)$$

where a_1, b_1, c_1 and a_3, b_3, c_3 are the temperature coefficients of frequency for the fundamental and third overtone, respectively. Equation (4) then yields the temperature coefficients for the beat frequency. The translation process results in a small difference between the multiplier and n, which has a profound effect on enhancing the linear coefficient for the beat frequency.

Table I
Typical SC-cut harmonic and beat frequency temperature coefficients.

MODE	FREQ (MHz)	a (10^{-6}/°C)	b (10^{-8}/°C²)	c (10^{-11}/°C³)
f_1	5.000	0.13	-1.69	5.45
f_3	14.843	1.15	-1.60	5.58
f_β	0.157	-96.33	-9.94	-7.24

Table 1 provides the computed temperature coefficients of an SC-cut crystal (n = 2.97) for the fundamental, third overtone and beat frequency. Equation (4) was used to determine the coefficients for the beat frequency using the coefficients for the fundamental and third overtone that were computed from measured frequency-temperature data. It is noted that the linear coefficient for the beat frequency is significantly larger than the first-order coefficients of the two modes, and that it also dominates its associated higher order coefficients. In fact, it is so dominant that a nearly linear beat-frequency versus temperature characteristic results. Evaluation of (4), using the beat frequency temperature coefficients listed in Table 1, produced no zero derivative in the temperature range 150K to 450K. This indicates monoticity of the beat frequency and suitability for practical thermometric application over a wide operating temperature range.

The resonator's fundamental and third overtone frequency are shown plotted against temperature in Figure 3. Since the temperature reference is taken at 0°C, the linear rotation between the two curves is not immediately obvious. However, the angle change is evidenced by shift of the turning point from +5°C to approximately +45°C.

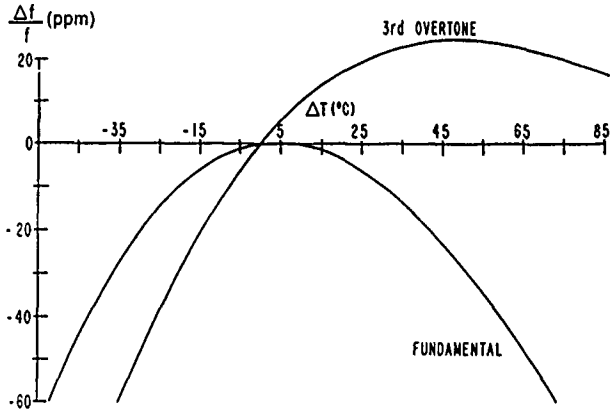


Figure 3. Fundamental and third overtone characteristics of a typical SC-cut resonator.

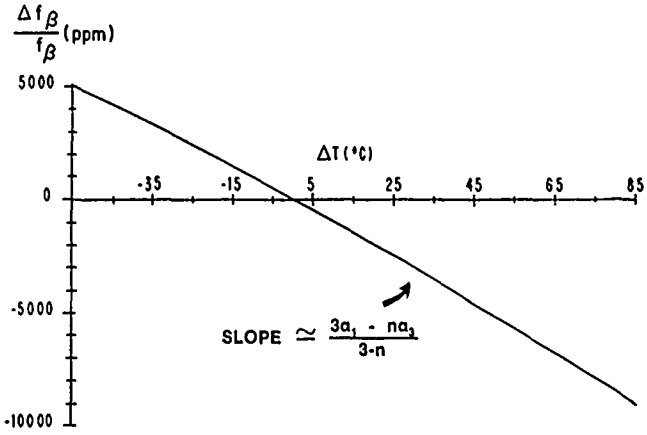


Figure 4. Temperature dependency of the beat frequency.

Figure 4 shows the temperature dependency of the beat frequency for the same SC crystal unit. The f-T curve is clearly monotonic. Its linearity is about 1.5 percent, comparable to that of a clean b-mode over the same temperature range. The high degree of linearity permits approximating the slope by its first-order temperature coefficient which, from Table 1, is -96.3 ppm/°C. Since n may be practically realized in the range of 2.94 to 2.98, the corresponding temperature sensitivities for resonators having similar harmonic first-order temperature coefficients would range from approximately -60 ppm/°C to -150 ppm/°C. By inspection of the first-order term in (4), it becomes evident that even greater temperature sensitivities may be realized simply by further reduction in the denominator, i.e., (3-n). This can be accomplished by using a noninteger multiplier to replace the times three multiplier. Or, additional circuit processing, e.g., using double heterodyne conversion, can be used advantageously to tailor the sensitivity to any desired value.

Temperature Sensing Error

Common to all electrical temperature sensing methods, the stability of the sensing device and associated circuitry are the primary causes of temperature sensing error. This error, which manifests itself as a change in calibration, determines the temperature measurement accuracy and limits the degree of temperature compensation. For the dual c-mode method, changes in either or both mode output frequencies at constant temperature can result in temperature sensing error.

The temperature sensing error ϵ_T due to a parameter variation, e.g., supply voltage change, and expressed in Kelvin, can be determined by dividing (3) by the derivative (slope) of (4). This results in the close approximation

$$\epsilon_T \simeq \frac{\Delta\delta}{\Delta a} \simeq \Delta\delta \times 10^9 \text{ mK} \qquad (5)$$

where $\Delta\delta$ is the differential, or difference stability, i.e., $\Delta\delta = \Delta f_1/f_1 - \Delta f_3/f_3$, and $\Delta a = a_1 - a_3$, the difference between the first-order temperature coefficients of the two modes; Δa is typically 1×10^{-6} K^{-1} for practical SC-cut designs.

This relationship tells us that the difference stability between the two modes essentially determines temperature sensing error, and that for a 1×10^{-9} difference stability the temperature sensing error will be approximately 1 mK. We now have the stability criteria that is required to be considered in the design of a dual-mode oscillator. In the case of a 2×10^{-8} (-55°C to +85°C) MCXO, the temperature sensing error must be no greater than 7 mK, as determined by the worst case f-T slope of either mode for an optimum design SC-cut resonator [9]. This corresponds to an allowable difference stability of about 7×10^{-9} to be maintained under all conditions.

The Dual-Harmonic-Mode Crystal Oscillator (DHMXO)

Figure 5 shows one example of a dual-mode oscillator capable of generating the two c-mode frequencies. This DHMXO can be described as two separate oscillators sharing a common crystal operating at series resonance. A single section resembles that of the Butler or Bridged Tee oscillator and falls into the family of crystal oscillators characterized by the crystal current being essentially equal the transistor emitter current. In this configuration the crystal is grounded to facilitate design. In each gain loop, emitter degeneration provides negative feedback that is sufficient to reduce net loop gain below unity for all frequencies except the desired c-mode frequency at which the crystal branch impedance is a minimum. This condition is assured by the series tuned networks C_s, L_s and C_s', L_s', which also serve to decouple the adjacent mode signal. Using this circuit, the coupled frequencies can be attenuated more than 50 dB below the primary output signal level. Positive feedback at each c-mode is provided by the phase shifting PI networks consisting of C_1, C_2 and L_3 and C_1', C_2' and L_3'.

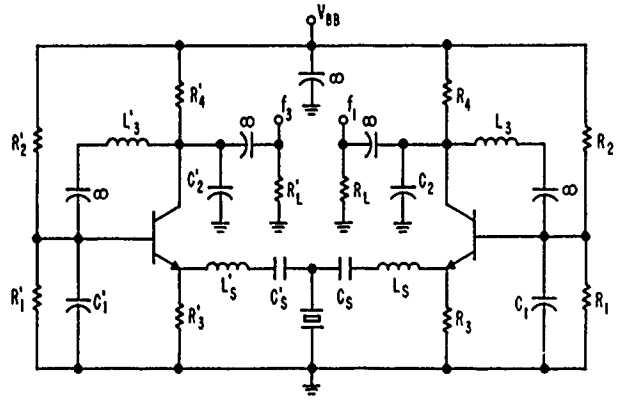


Figure 5. A double gain loop DHMXO (emitter degenerative type).

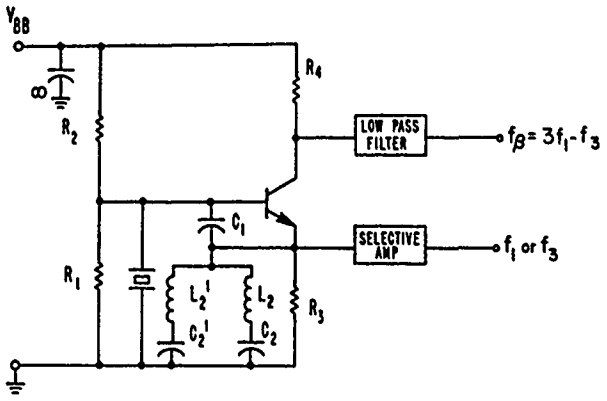


Figure 6. A single gain loop DHMXO (colpitts type).

Measured frequency-voltage stabilities for a ±1% supply voltage change provided a difference stability of $\pm 5 \times 10^{-9}$. This performance indicates that a maximum 1 mK temperature sensing error due to supply voltage variation is achievable by regulating the supply voltage to a tolerance of ±0.2%. Difference stabilities due to load variation, component and crystal aging and other parameter changes require similar consideration.

Simultaneous excitation of the c-modes was also obtained in the single-gain-loop Colpitts type DHMXO shown in Figure 6. Suppression of undesired SC-cut b-mode frequencies is accomplished by the network C_2, L_2, C_2', L_2', which is designed to appear capacitive, and provides the correct phase shift for oscillation, at only the two c-mode frequencies. Inherent nonlinearities in the transistor provide for multiplication and mixing such that the beat frequency is available at the collector output, following filtering of the higher frequency components. Either the fundamental or third overtone frequencies can be extracted at the emitter using a frequency selective amplifier. The single gain loop DHMXO basically offers lower parts count and lower input power than the double gain loop circuit. However, it does trade off the flexibility for separate control over crystal-mode-current-adjustment and stability-coefficient optimization. This limited capability can prevent the single gain loop DHMXO from achieving the same high level of performance as the double gain loop design. However, its use may be warranted in applications permitting somewhat less temperature sensing accuracy.

Transient Temperature Performance

Tests were conducted on a brassboard of the dual-mode oscillator shown in Figure 5 to demonstrate performance of the self-temperature-sensing method under thermal transient conditions. Figures 7 and 8 show the DHMXO fundamental frequency plotted against its case temperature, measured by an external temperature sensor, for a complete temperature cycle at air temperature slew rates of 1/4°C/minute and 2°C/minute, respectively. In one case, the crystal temperature lags the external sensor by about 2°C (Figure 7) and in the other, by about 10°C (Figure 8). The curves

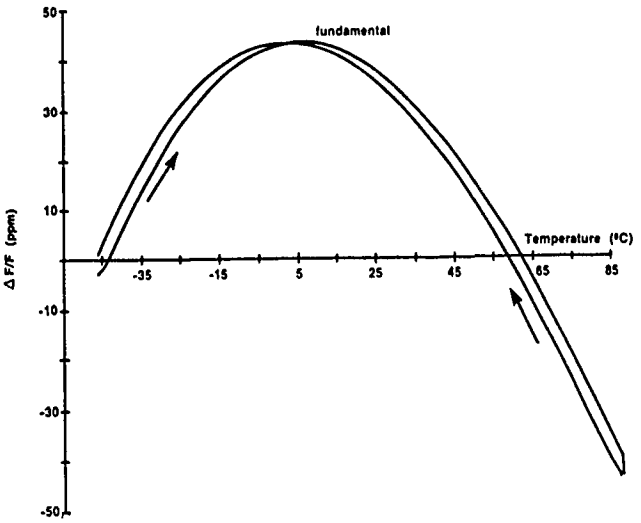


Figure 7. DHMXO f-T characteristic using an external sensor (slew rate - 0.25°C/min).

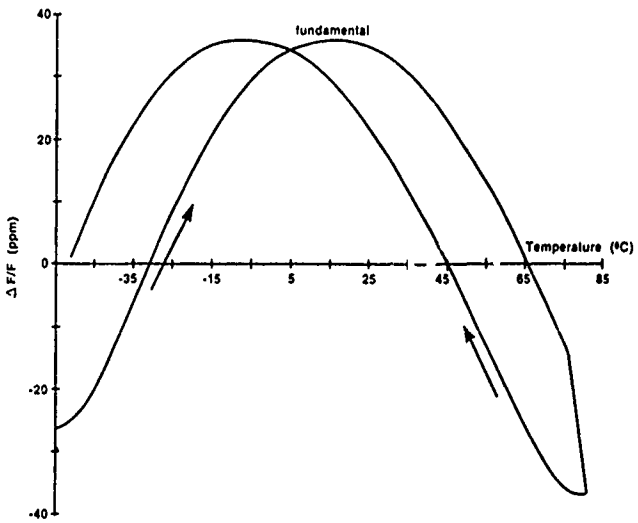


Figure 8. DHMXO f-T characteristic using an external sensor (slew rate - 2°C/min).

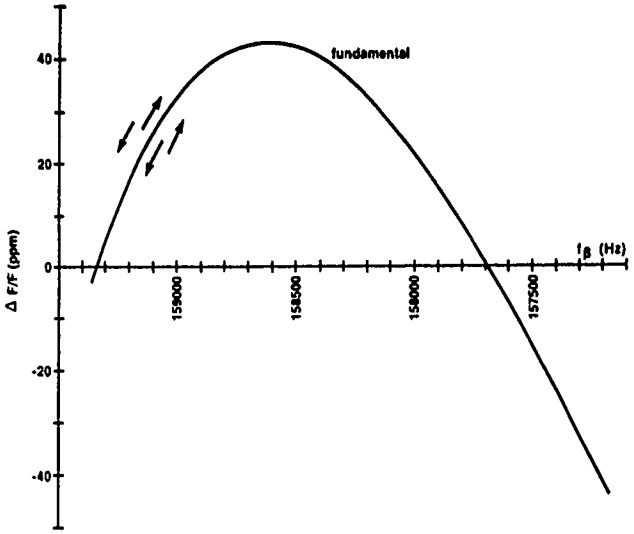


Figure 9. DHMXO f-T characteristic using self-temperature-sensing (slew rate - 0.25°C/min and 2°C/min).

illustrate, in somewhat exaggerated form, the thermal lag which might occur between the crystal and its compensating sensor due to their unequal effective thermal time constants. Figure 9 is a plot of fundamental frequency versus the thermometric beat frequency for the same two temperature runs. All four frequency curves superimposed over a broad temperature range, verifying the non-existence of thermal lag between the crystal and its thermometry. However, there are residuals in the data among these curves that primarily reflects resonator hysteresis [9,10].

Implementation

Implementation of the self-temperature-sensing method is shown by the MCXO simplified block diagram of Figure 10. The MCXO is defined as a microcomputer-based system that corrects for the temperature sensitivity of the quartz crystal unit's frequency by using external compensation means [11]. External compensation requires "no pulling" of the crystal frequency. Instead, it employs techniques such as pulse deletion [5], phase-locked-loop summing [6], digital phase shifting [12], and single-sideband mixing [13].

In this illustration, the thermometric beat frequency is obtained by dividing the dual-mode oscillator's third-overtone output frequency by three and then mixing the divided signal with the fundamental output frequency. Although f_β is now one-third of the value it had in the initial example, it retains the same temperature coefficient as previously. Also, f_β is further divided by M to obtain a convenient value for gating a reciprocal counter, which uses the third-overtone frequency as a time base. The reciprocal counter produces a number N1 that represents the temperature of the crystal. The divisor M serves to scale the N1 counter range consistent with the required temperature-sensing resolution and the number of counter

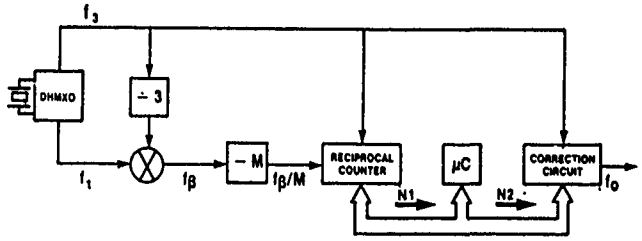


Figure 10. Typical implementation in an MCXO.

bits. This number, or temperature word, is read by the microcomputer and is used to compute a second number N2 that programs the correction circuitry. Correction algorithms may require either evaluation of a higher order polynomial or a look-up table using coefficients or table entries, respectively, that were previously stored in memory during the calibration process.

Summary of Characteristics

The dual c-mode thermometer circumvents the need for a separate temperature sensor in MCXO applications. Since the crystal determines its own temperature, thermal lag errors resulting from unequal effective thermal time constants of the resonator and sensor are eliminated. Thermal-gradient-related errors due to the physical location and self-heating of an external sensor are dispensed with.

The c-modes of the SC-cut are stress and thermal transient compensated. Therefore, thermometry errors due to either temperature gradients across the crystal or to induced stresses under dynamic temperature conditions are minimized.

A high level of stability results since the c-modes have Q's greater than 1 million and are relatively insensitive to circuit component changes. Also, a high order of uniformity and reliability is assured since the c-modes are virtually free of activity dips.

The method has a capability to limit temperature-sensing error to 1 mK over extended periods of time. In fact, μK sensing appears possible if the difference stability between the two modes can be sufficiently minimized to the extent where the limiting mechanism is oscillator noise.

Conclusions

A thermometry method of using stable c-modes in a dual-mode oscillator has been described. Its potential is now being realized in the development of high-performance MCXO having stabilities 10 to 100 times greater than analog TCXO. The method overcomes the limitations of existing temperature-sensing methods, is easy to implement, is microprocessor compatible, and makes possible a new generation of inexpensive, higher-accuracy, temperature-compensated crystal oscillators.

Acknowledgements

The author thanks Vincent Rosati for his helpful suggestions and Dennis Bowman for his valuable assistance in measurement of DHMXO performance. A special thanks is extended to John Vig and Raymond Filler for the many stimulating discussions on the dual c-mode method.

References

[1] J. A. Kusters, M. C. Fischer, and J. G. Leach, "Dual Mode Operation of Temperature and Stress Compensated Crystals," Proceedings of the 32nd Annual Symposium on Frequency Control, 1978, pp. 389-397.

[2] J. R. Vig, R. L. Filler, and J. A. Kosinski, "SC-cut Resonators for Temperature Compensated Oscillators," Proceedings of the 36th Annual Symposium on Frequency Control, 1982, pp. 181-186.

[3] O. J. Baltzer and C. S. Stone, "External Compensated Crystal Oscillator (ECXO) Study," Final Technical Report, RADC-TR-84-166 (July 1984).

[4] S. Schodowski, "Dual-Mode Quartz Thermometric Sensing Device," US Patent Application, SN 487,560.

[5] M. Bloch, M. Meirs, and J. Ho, "The Microcomputer Compensated Crystal Oscillator (MCXO)," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[6] A. Benjaminson and S. C. Stallings, "A Microcomputer-Compensated Crystal Oscillator Using a Dual-Mode Resonator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[7] M. Bloch, M. Meirs, J. Ho, J. R. Vig, and S. Schodowski, "Low Power Timekeeping," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[8] A. Ballato and T. Lukaszek, "Higher Order Temperature Coefficients of Frequency of Mass-Loaded Piezoelectric Crystal Plates," Proceedings of the 29th Annual Symposium on Frequency Control, 1975, pp. 10-25.

[9] R. L. Filler and J. R. Vig, "Resonators for the Microcomputer Compensated Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[10] R. L. Filler, J. A. Messina and V. J. Rosati, "Frequency-Temperature and Aging Performance of Microcomputer-Compensated Crystal Oscillators," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[11] Military Specification: "Oscillators, Crystal, General Specification for," MIL-0-55310B, 1988.

[12] O. J. Baltzer and D. J. Rybicki, "Spectrally Pure MCXO with Doppler Correction Input," Final Technical Report, SLCET-TR-84-0430-F (October 1988).

[13] B. Schodowski, "An External Compensated Crystal Oscillator Study," Proceedings of the 8th Annual Quartz Devices Conference, 1986, pp. 169-178.

43rd Annual Symposium on Frequency Control - 1989

RESONATORS FOR THE MICROCOMPUTER COMPENSATED CRYSTAL OSCILLATOR

Raymond L. Filler and John R. Vig
U.S. Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

Summary

A novel oscillator, based on dual c-mode self-temperature sensing, is described by Schodowski elsewhere in these proceedings. Whether or not the promise of Schodowski's novel method could be realized depended on whether or not a resonator could be developed which is well-behaved on two c-modes. Therefore, the goal of the resonator development program discussed in this paper was to develop dual c-mode resonators which have: a. frequency vs. temperature (f vs. T) characteristics that are free of significant anomalies, b. the minimum possible f vs. T slopes for the temperature ranges of interest, c. high Q, and d. small hysteresis.

Well-behaved dual c-mode SC-cut resonator designs have been successfully developed. Plano-convex resonators of the proper contour are shown to be well-behaved on both the third overtone, at 10 MHz, and on the fundamental mode over, at least, a -55°C to +85°C temperature range. Such resonators are shown to be suitable for microcomputer compensated crystal oscillators of a few parts in 10^8 stability.

Introduction

The best overall accuracy available in a wide temperature range (e.g., -55°C to +85°C) temperature compensated crystal oscillator (TCXO), has remained at about 1 ppm for the past 30 years. In 1978, Kusters et al.[1] described a dual-mode oscillator which seemed extremely promising as a means of overcoming the limitations of conventional, analog TCXO. Today, eleven years later, there are no such dual-mode oscillators on the market. The best TCXO's overall accuracy (including hysteresis, trim effect, and one year of aging) is still about 1 ppm.

That the promise of dual-mode oscillators has not been realized is due to the fact that the b-mode, which had been suggested by Kusters et al. as the temperature sensing mode, is not well behaved. When measured over a wide temperature range, the b-mode has been found to exhibit a high incidence of frequency versus temperature (f vs. T) and resistance versus temperature (R vs. T) anomalies. Moreover, the b-mode of SC-cut resonators is neither stress nor thermal-transient compensated.

A different, even more promising dual-mode oscillator, based on dual c-mode self-temperature sensing, is described by Schodowski[2] elsewhere in these proceedings. Whether or not the promise of Schodowski's novel method could be realized depended on whether or not a resonator could be developed which is well-behaved on two c-modes. Therefore, the goal of the resonator development program discussed in this paper was to develop dual c-mode resonators which have:

a. frequency vs. temperature (f vs. T) characteristics that are free of significant anomalies,

b. the minimum possible f vs. T slopes for the temperature ranges of interest,

c. high Q, and

d. small hysteresis.

The Resonators Studied

The characteristics of a variety of 5 MHz fundamental mode, 5 MHz 3rd overtone, and 10 MHz 3rd overtone SC-cut resonators, which had been fabricated for previous experiments,[3-5] were examined. The resonator plates were 14 mm diameter, plano-convex, and were mounted in a four-point mounting structure, as was described previously.[4] The electrodes were pure gold and 5 mm in diameter.

Anomalies of the b-mode

Figures 1 and 2 illustrate activity dips which are frequently encountered in the b-mode of SC-cut resonators. During the f vs. T measurement shown in Fig. 1a, the frequency was "lost" at the temperature indicated by the arrow. A careful measurement of the R vs. T of the same resonator indicates that, at the temperature where the frequency was "lost," the resistance had increased from 146 Ω to more than 6,000 Ω, as is shown in Fig. 1b. Examination of the residuals of the f vs.T data (after subtracting the least-squares-fit to the data from the data) shows that, in the vicinity of the activity dip, there had been a significant departure from the b-mode's nearly linear f vs.T characteristic, as is shown in Fig. 1c. Figure 2 shows the R vs. T for the b-mode of another 10 MHz 3rd overtone SC-cut resonator.

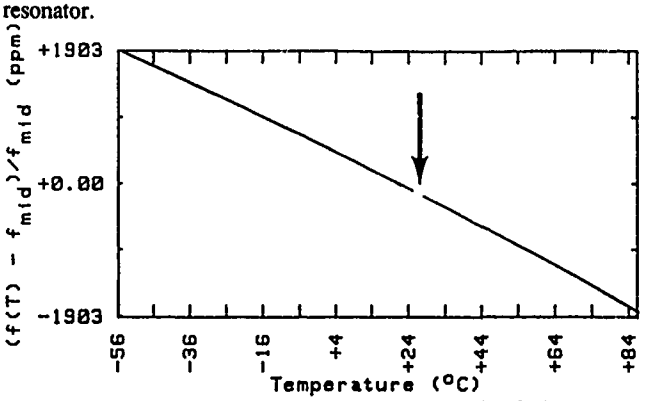


Figure 1a - Frequency vs. temperature characteristic of a b-mode.

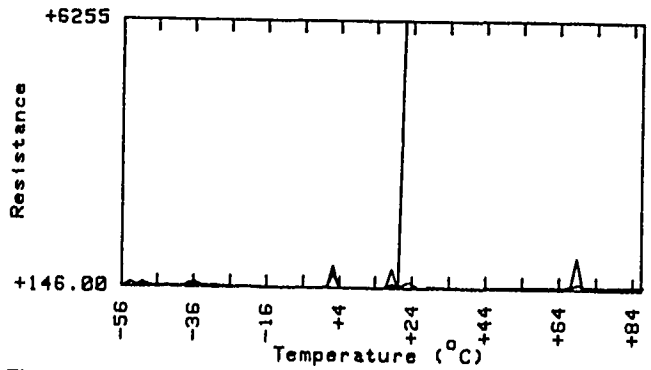


Figure 1b - Resistance vs. temperature characteristic of a b-mode.

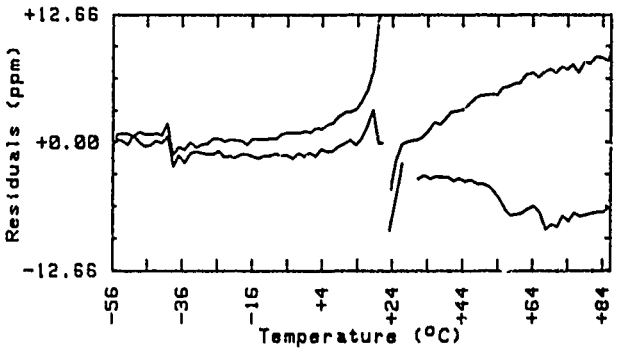


Figure 1c - Frequency vs. temperature residuals of a b-mode.

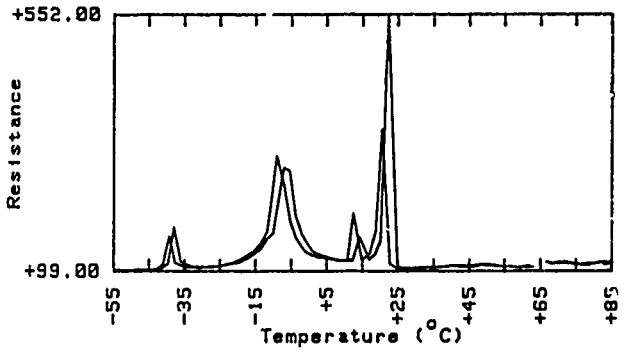


Figure 2 - Resistance vs. temperature characteristic of a b-mode.

MCXO Resonator Designs

Several 5 MHz fundamental mode and 10 MHz 3rd overtone designs were found to be "well-behaved" over the -55°C to +85°C range, on both the fundamental and the third overtones. No 5 MHz 3rd overtone design was found to be well behaved at the fundamental mode frequency (1.7 MHz). It appears that, for the fundamental mode of the 5 MHz 3rd overtone to be well-behaved, the plate diameter would need to be significantly larger than 14 mm.

The investigations of resonator properties concentrated on the 10 MHz 3rd overtone designs, primarily because: 1. everything else being equal, 10 MHz 3rd overtone resonators are generally of higher stability than 5 MHz fundamental mode resonators, and 2. the lower the frequency, the lower the power required by the digital (CMOS) circuitry in the microcomputer compensated crystal oscillator (MCXO) that was to utilize the resonators. The major variable studied was the plate contour. The range of contours extended from 0.5 diopters to 7.0 diopters (and included 0.5, 1.0, 2.0, 2.25, 2.5, 3.0, 3.87, 5.0, 5.25, 5.5, 6.0, and 7.0 diopters).

Table I shows a summary of the "good" MCXO resonator designs and the corresponding salient characteristics. Acceptable performance was found over a wide range of contours. In the range of 2.5 to 5.5 diopters, the achievable Q's were greater than 1.2 million for both the fundamental and the 3rd overtone, and the f vs. T characteristics were free of significant anomalies.

MCXO Resonator Properties

Fundamental and Third Overtone Q's

In the range of 1 to 7 diopters, the Q of the third overtone mode showed no systematic variation with contour; Q's greater than 1.0 million can be achieved in this range. The Q of the fundamental mode, however, was found to decrease at contours below 2.5 diopters. At 1.0 diopter, for example, the Q's were typically less than 300,000. Above 2.0 diopter, the Q's were typically in the 1.2 to 2.8 million range, with no obvious dependence on contour.

Curvature (diopters)	f_β (kHz)	df_β/dT (Hz/°C)	f_3/f_1	C_{11} (fF)	C_{13} (fF)
2.5	157	-13	2.954	2.84	0.156
4.0	200	-15	2.943	2.60	0.125
5.5	225	-18	2.932	2.26	0.106

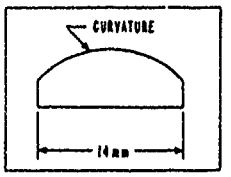


Table I - MCXO resonator characteristics.

Frequency Ratio and Beat Frequency

The third overtone to fundamental mode frequency ratio decreases (nearly linearly) from 2.974 at 0.5 diopter to 2.927 at 7.0 diopter, i.e., the ratio is less than 3.0 at all contours. The thermometric beat frequency f_β, where $f_\beta = 3f_1 - f_3$, increases correspondingly with increasing contour. As an interesting aside, for plano-plano resonators, the ratio can be varied from less than 3.0 to greater than 3.0 by adjusting the electrode dimensions. As the electrode becomes larger, so does the ratio.[6]

The Motional Capacitances

As can be expected, the motional capacitance of the fundamental mode, C_{11}, and of the third overtone, C_{15}, both decrease with increasing contour. More important for the stability of the MCXO, however, is the difference between C_{11} and C_{13}. An important consequence of this difference is the difference in sensitivities of the two modes to load reactance changes (which needs to be considered in the design of dual-mode oscillators).

The C_{11} to C_{13} ratio increases slightly, from 18 to 21, as the contour increases from 2.5 diopters to 5.5 diopters; i.e., everything else being equal, the fundamental mode is about 20 times more sensitive than the third overtone to reactance changes in the oscillator circuitry.

Frequency vs. Temperature Characteristics

The f vs. T characteristics of SC-cut resonators for TCXO applications were investigated previously.[7] The f vs. T characteristics appear to be parabola-like because the inflection temperature of SC-cut resonators is above the upper limit of the "normal" TCXO temperature ranges. (See Fig. 3.)

Figure 4 shows the typical relationship between the third overtone and the fundamental mode f vs. T of the same resonator. The apparent θ-angle difference, $\Delta\theta$, is about 31 minutes for 3.0 diopter plano-convex resonators. Since the fundamental mode's apparent angle-of-cut varies with contour more than the third overtone's does[3], $\Delta\theta$ varies with contour from about $\Delta\theta = 26$ minutes at 1 diopter to $\Delta\theta = 38$ minutes at 5 diopters.

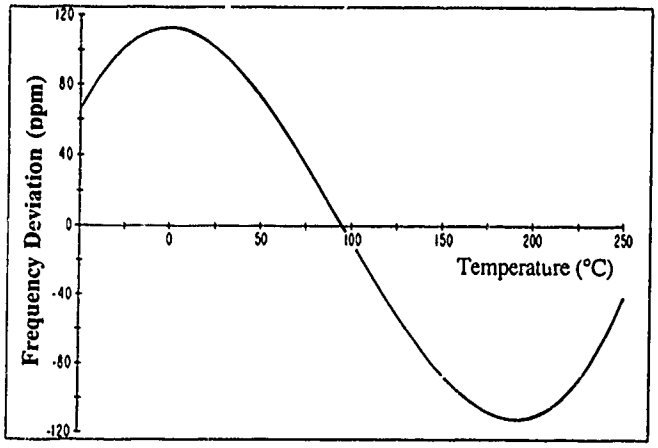


Figure 3 - SC-cut f vs. T, -50°C to +250°C.

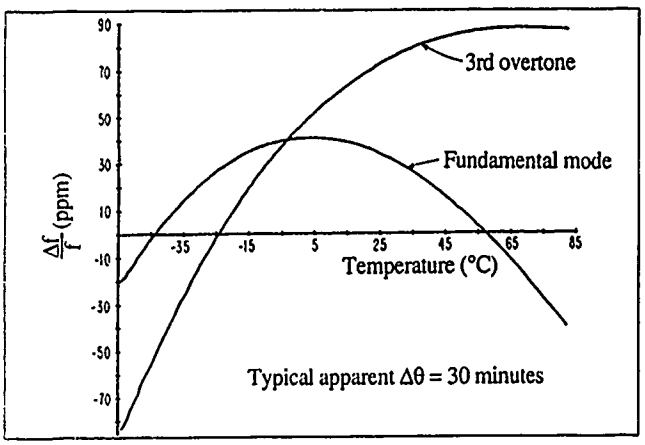


Figure 4 - Fundamental mode and 3rd overtone of an SC-cut resonator.

The beat frequency, f_β, and the change of the beat frequency with temperature, df_β/dT, are also functions of contour, as is shown in Table I. A typical f_β vs. T characteristic is shown in Fig. 5. The relationship is nearly linear. For example, for a -55°C to +85°C temperature range, the maximum deviation from the best fit straight line is about 2 percent of the total frequency change. When the data is fit to a quadratic expression, the ratio of the linear coefficient (80.19) to the quadratic coefficient (0.0978) is 820.

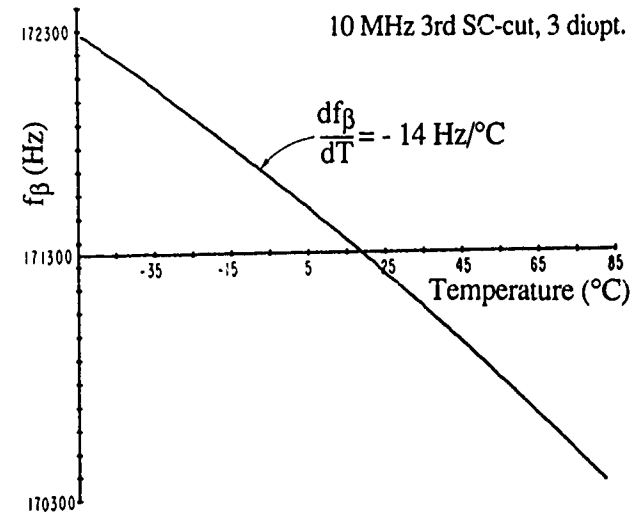


Figure 5 - Beat frequency vs. temperature.

In MCXO applications, for reasons discussed in the next section, the optimum resonator f vs. T characteristics are the ones where the maximum f vs. T slopes of the two modes are minimized. This condition is satisfied when the absolute value of the slope of the fundamental mode at the upper temperature extreme equals the slope of the third overtone at the lower temperature extreme. See for example, the optimum f vs. T for the -55°C to +85°C temperature range shown in Fig. 6. The absolute values of the maximum slopes are 2.7 ppm per °C, which occur when the turnover temperature of the 3rd overtone is at about +20°C. The exact condition for optimum f vs. T is that the maximum f vs. f_β slopes be minimized, however, since f_β vs. T is nearly linear, the difference between the two conditions is small.

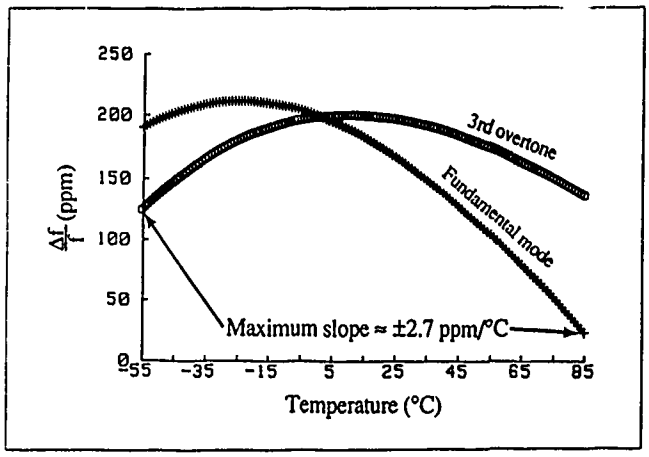


Figure 6 - Optimum resonator f vs. T for -55°C to +85°C MCXO.

Frequency vs. Temperature vs. Angles of Cut Characteristics

The f vs. T vs. angles-of-cut characteristics of SC-cut TCXO and MCXO resonators were investigated previously.[7] It was shown that the necessary angle-of-cut tolerances can be readily achieved with conventional cutting technology. For example, for the optimum resonator for a -55°C to + 85°C MCXO, for which the lower turnover temperature is at +20°C, a cutting error of 5 minutes results in an 8°C shift in the lower turnover temperature. For such an MCXO resonator, the rate of change in f vs. T slope with lower turnover temperature is 0.043 ppm/°C^2 at -55°C. Therefore, a -5 minute error increases the lower turnover temperature to +28°C, which increases the slope at -55°C from the optimum 2.7 ppm/°C to 3.0 ppm/°C. A +5 minute error causes a smaller maximum slope change (at +85°C), as the rate of change in f vs. T slope is -0.029 ppm/°C^2 at +85°C. Therefore, in many MCXO applications, the angle-of-cut tolarence requirement may be loose enough for x-ray orientation and angle-correction to be unnecessary.

Consequences of Using f_β

That the beat frequency, $f_\beta = 3f_1 - f_3$, is nearly linear with temperature is a consequence of the fact that:

$$f_1(T) = f_1(T_0) + A_1\Delta T + B_1\Delta T^2 + C_1\Delta T^3 + \ldots ,$$

and $$f_3(T) = f_3(T_0) + A_3\Delta T + B_3\Delta T^2 + C_3\Delta T^3 + \ldots$$

where the A's, B's, and C's are constants, T_0 is a reference temperature, and $\Delta T = T - T_0$.

Therefore,

$$f_\beta(T) = f_\beta(T_0) + (3A_1 - A_3)\Delta T + (3B_1 - B_3)\Delta T^2 + (3C_1 - C_3)\Delta T^3 ...$$

The first order term is much greater than the higher order terms for at least a 25°C ± 100°C temperature range. Therefore,

$$f_\beta(T) \approx f_\beta(T_0) + (3A_1 - A_3)\Delta T,$$

and, therefore, $\Delta f_\beta(T) \approx k\Delta T$ for all the customary oscillator temperature ranges, where $k = 3A_1 - A_3$.

Some consequences of measuring frequency vs. f_β instead of frequency vs. T are:

1. Since the resonator acts as its own thermometer, problems due to thermal lag between the resonator and an external temperature sensor are eliminated. To illustrate, Fig. 7 shows the frequency of a c-mode, from -55°C to +85°C and back to -55°C, plotted as a function of temperature, where the temperature was measured by means of a (HP quartz) thermometer in the temperature chamber. Figure 8 shows the data for the same temperature run; however, this time, the frequency is plotted as a function of f_β.

2. When frequency is measured as a function of temperature, frequency and temperature are independent variables. When there is an abnormal excursion in the frequency (e.g., an activity dip), the temperature is unaffected. However, when frequency is measured as a function of f_β, since f_β is a function of both c-mode frequencies, an anomaly in either c-mode frequency affects f_β, i.e., anomalies distort the f_β scale. Two important consequences of this, derived in the appendix, are:

a. After the data is fit to an approximating function, and that function is subtracted from the actual data, the fundamental mode residuals are equal to the third overtone residuals; i.e., from the point of view of MCXO stability, there is no advantage to using one mode as opposed to the other - either mode can provide the same stability as the other. Figure 9 shows an example of the frequency residuals vs. f_β for the fundamental mode and the third overtone of a 10 MHz 3rd overtone resonator. No differences are visible in the residuals.

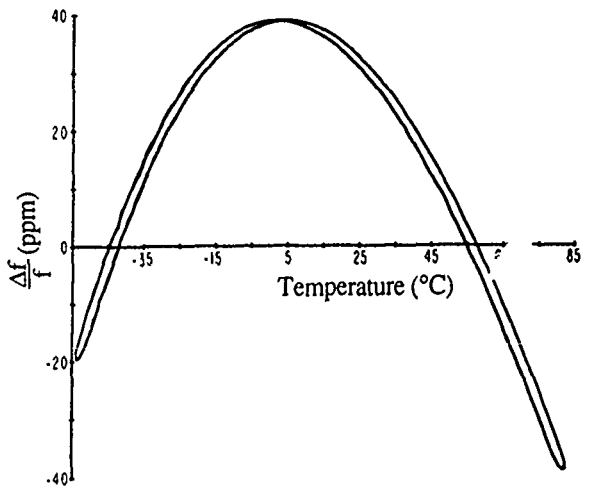


Figure 7 - Frequency vs. temperature shows apparent hysteresis due to thermal lag.

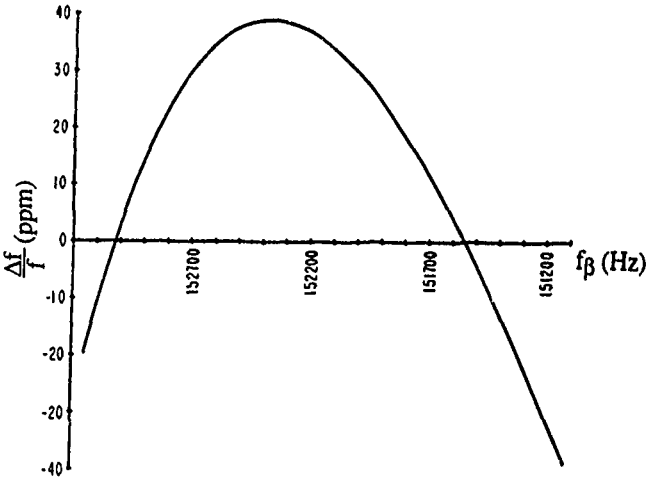


Figure 8 - Frequency vs. beat frequency shows no thermal lag.

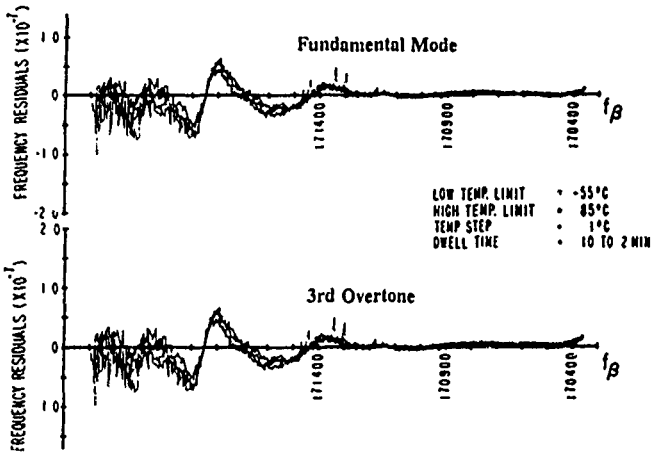


Figure 9 - Frequency residuals vs. f_β for the fundamental mode and third overtone of the same resonator.

b. In the residuals, deviation of one mode from the approximating function is scaled by the slope of the other mode at the same f_β (i.e., a deviation in the fundamental mode at an f_β is scaled by the slope of the third overtone at the same f_β, and a deviation in the third overtone at an f_β is scaled by three times the slope of the fundamental mode at the same f_β.) For example, if an excursion from the approximating function in one mode takes place at the turnover temperature of the other mode, there will be no evidence of this anomaly in the residuals. If, on the other hand, the anomaly takes place where the slope of the other mode's f vs. T is maximum, the anomaly will be amplified by the slope of the other mode. To illustrate, as is mentioned above, the worst case slopes for resonators optimized for the -55°C to +85°C temperature range are ±2.7 ppm per °C, or ±27 Hz per °C at 10 MHz. The slope of f_β vs. T for a (3 diopter plano-convex) 10 MHz 3rd overtone SC-cut resonator is -14 Hz per °C. Therefore, the worst case amplification factor, at -55°C and at +85°C, is 27/14 = 1.9. The normalized amplification factor at the fundamental mode is the same.

3. The slopes of the f vs. f_β curves are related by

$$3df_1/df_\beta - df_3/df_\beta = 1;$$

therefore, the amplification factor discussed in the previous paragraph is always greater than one.

4. When curve-fitting the f vs. f_β data, to prevent the systematic deviations from the approximating function from exceeding 1 x 10^{-8}, a simple sixth-order polynomial function is not sufficient, in general. A better curve-fitting procedure must be used, such as a higher-order polynomial fit, a segmented polynomial fit, etc. For example, in Fig. 9, the simple 6th-order polynomial fit that was used was clearly inadequate. Figure 10 shows the residuals vs. f_β for another resonator after fitting the data to a 6th-order polynomial function. Figure 11 shows the results for the same resonator after the data was fit to a three-segment 6th order polynomial. A significant improvement can be seen in the vicinity of f_β = +3,000 ppm, where there had been a systematic deviation from the approximating function. The "noise" in the vicinity of f_β = -5,000 ppm was due to measurement-system limitations (see next section), together with the fundamental mode having a steeper than optimum slope near +85 C.

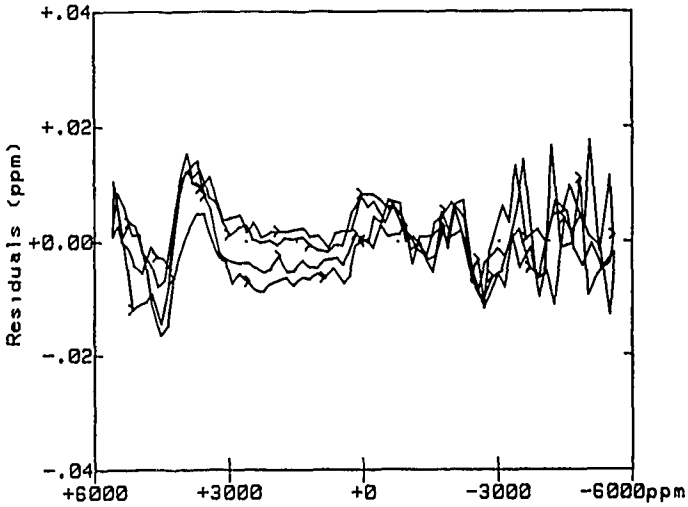


Figure 10 - Residuals vs. f_β and hysteresis after a fit to a 6th order polynomial.

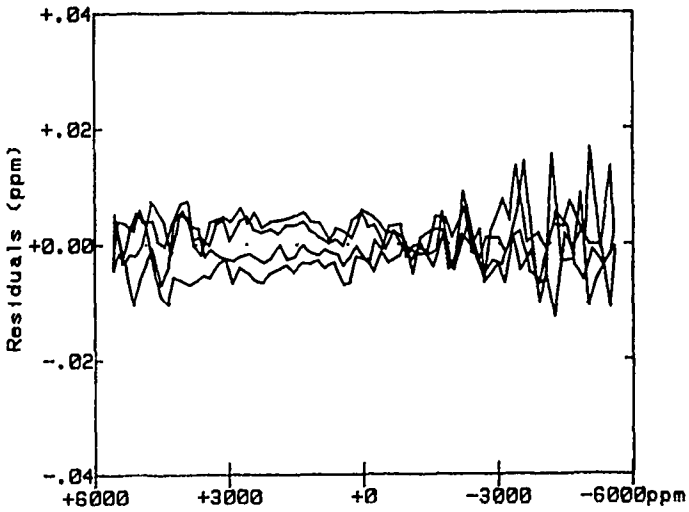


Figure 11 - Residuals vs. f_β, and hysteresis, after a fit to a three segment 6th order polynomial.

Ideally, the curve fitting should make the systematic contributions to the residuals much smaller than the random contributions; i.e., the f vs. T stability of the MCXO should be limited primarily by the hysteresis. Although a "good" curve-fitting algorithm will be able to fit f vs. f_β curves that contain activity dips, one must be careful to ensure that, in so doing, the activity dips do not change the monotonicity of the f_β vs. T characteristic. The f_β vs. T relationship must be monotonic for proper operation of the MCXO.

Hysteresis

MCXO resonator hysteresis was measured with the Π-network measurement system illustrated in Fig. 12. The synthesizer's resolution was 1 mHz; the vector voltmeter's was 0.1 degree. Data was collected during two temperature cycles: -55°C to + 85°C to -55°C to +85°C to -55°C. The temperature steps were 1°C, and the dwell time at each temperature ranged from 10 min at -55°C to 2 min at 20°C and above. At each temperature, five frequencies were recorded sequentially: f_3, f_1, f_3, f_1, and f_3. The first f_1 and the average of the first and second f_3 measurements were recorded as one data point set, and the second f_1 together with the average of the second and third f_3 measurements were recorded as another data-point set.

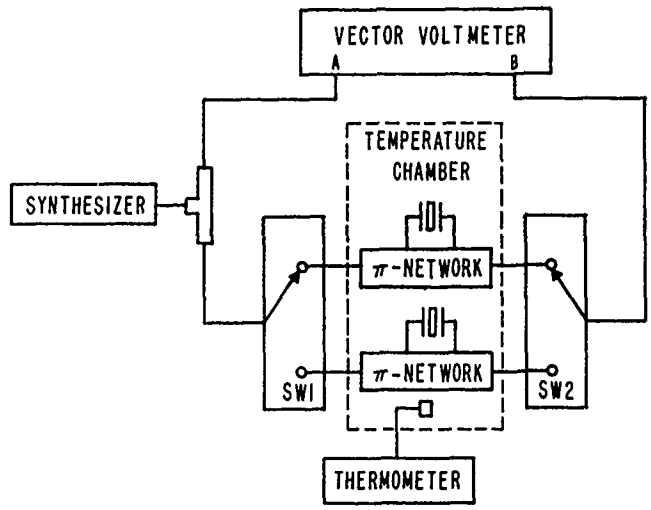


Figure 12 - Frequency vs. temperature teasurement system.

After a least-squares fit to a polynomial function, the residuals were plotted as a function of f_β. Typical results are shown in Figures 9 to 11. The measured hysteresis ranged from parts in 10^9 for the best units, to about 2 X 10^{-8} for the typical "good" unit, to several parts in 10^{-8} for the "bad" units. Further hysteresis results, for dual-mode oscillators and complete MCXO's, are presented elsewhere in these Procedings.[8]

Hysteresis is not well understood.[9, 10] The major mechanisms responsible for hysteresis are likely to be similar to those responsible for aging: mass transfer to and from the resonator surfaces due to contamination, strain relief due to stress relief in the resonator's mounting structure, changes in the oscillator circuitry, and, possibly, changes in the quartz (as is implied in references 9 and 10).

Comments on Differential Stability

Since f_β is a function of both f_1 and f_3, if the two modes, e.g., age differently, f_β, and, consequently, the compensation of the MCXO, also change with time. Hence, the factors that can affect the two modes differently have been of concern.[2]

Some inherent differences exist between the two modes which can affect the relative stabilities. One, already discussed, is the difference in C_1. Oscillator circuitry changes, which can affect both the resonator load reactance and the drive level, will affect the two modes differently. These differences can be minimized through proper oscillator circuit design.

One of the major aging mechanisms is mass transfer due to surface contamination. To first order, since $\Delta f/f = \Delta m/m$, where m is the mass loading due to the contamination, mass transfer will have the same effect on the two modes. However, since the fundamental and third overtone mode shapes are different, contamination in the region where the mode shapes differ will affect the fundamental mode more than the third overtone. The difference is likely to be negligible in 10^{-8} f-T stability MCXO's that use precision resonators.

Another significant aging mechanism is strain relief due to stress relief in the resonator's mounting structure. Again, due to the difference in mode shapes, the effect of strain relief will be somewhat higher on the fundamental mode frequency than on the third overtone frequency. This difference is also likely to be small in properly mounted precision resonators.

Evidence that differential stability is not a significant problem in 10^{-8} f-T stability MCXO's is presented elsewhere in these proceedings.[8]

Summary and Conclusions

SC-cut resonator designs for the MCXO have been developed. Plano-convex resonators of the proper contour are well-behaved on both the third overtone, at 10 MHz, and on the fundamental mode over, at least, a -55°C to + 85°C temperature range. Since the goal of the MCXO resonator development program was to develop resonator designs for high stability, low-power MCXO's, and since these goals were met, higher frequency MCXO resonators were not investigated. It is likely, however, that at higher frequencies, plano-plano resonators exist that are also well-behaved on both the fundamental and third overtone modes.

Table II shows a comparison between resonators for the MCXO and for precision analog TCXO. Not only is the angle-of-cut tolerance looser for MCXO resonators, but so are the blank frequency and plating tolerances. For example, in one MCXO implementation,[11] pulse deletion is used in order to generate an accurate time corrected pulse train. The resonator target frequency during plating is chosen so as to ensure that the resonator frequency is above the nominal clock frequency at all temperatures. Since the resonator's frequency excursion is greater than 100 ppm over the typical MCXO temperature ranges, and since no frequency pulling is used in the MCXO, there is no need to specify a tight plating tolerance. In fact, the accuracy achievable with "rough" plating may be sufficient; i.e., no frequency adjustment should be necessary if the rough plating is reasonably well controlled.

Parameter	MCXO	TCXO
Cut, overtone	SC-cut, 3rd	AT-cut, fund.
Angle-of-cut tolerance	Loose	Tight
Blank f and plating tolerance	Loose	Tight
Activity dip incidence	Low	Significant
Hysteresis (-55°C to +85°C)	10^{-9} to 10^{-8}	10^{-7} to 10^{-6}
Aging per year	10^{-8} to 10^{-7}	10^{-7} to 10^{-6}

Table II - MCXO-TCXO resonator comparison.

The hysteresis and aging are lower for MCXO resonators because 3rd overtone SC-cut resonators are inherently more stable than the (usuallly) fundamental mode AT-cut resonators used in wide-temperature-range TCXO. Another possible reason is that the interface between the rough plating and the fine plating can be eliminated in MCXO resonators.

Hysteresis is the major limitation on the f-T stability that is achievable with the MCXO. Although the majority of resonators exhibited hysteresis at the low 10^{-8} level, some were in the 10^{-9} range. This shows that MCXO's with an f-T stability in the 10^{-9}, or better, range is a reasonable goal for the future. For that goal to be met, further research is needed on gaining a better understanding of, and on methods of controlling, the mechanisms responsible for hysteresis.

Acknowledgement

The authors thank John A. Kosinski for his contributions during the early phases of this research.

APPENDIX

RESIDUALS of DUAL-MODE APPROXIMATING FUNCTIONS

By definition,

$$f_\beta(T) = Kf_1(T) - f_3(T) \qquad (1)$$

(where K=3 for the current MCXO implementation; however, K can be other than 3).

We approximate f_1 and f_3 with polynomials in f_β:

$$f_1(T) \sim F_1(f_\beta) = \Sigma\, C_{1j}(f_\beta)^j \qquad (2)$$

$$f_3(T) \sim F_3(f_\beta) = \Sigma\, C_{3j}(f_\beta)^j \qquad (3)$$

We substitute Eqs. (2) and (3) into (1) to obtain:

$$f_\beta \sim F_\beta = KF_1(f_\beta) - F_3(f_\beta), \qquad (4)$$

which can be rewritten as

$$f_\beta = \Sigma\, (KC_{1j} - C_{3j})(f_\beta)^j \qquad (5)$$

In Eq. (5), we equate terms of like order in f_β; therefore, only the coefficient of the term with j=1 is nonzero.

$$KC_{1j} - C_{3j} = 0 \;;\; j = 0, 2, 3, \ldots \qquad (6)$$
$$KC_{11} - C_{31} = 1$$

We define the residuals at each data point, as the difference between the actual data point and the approximating function at the same f_β, i.e.,:

$$r_{1i} = f_{1i} - \Sigma\, C_{1j}(f_{\beta i})^j \quad (7)$$

$$r_{3i} = f_{3i} - \Sigma\, C_{3j}(f_{\beta i})^j \quad (8)$$

where

$$f_{\beta i} = Kf_{1i} - f_{3i} \quad (9)$$

Using Eq. (6) in Eq. (8) we have,

$$r_{3i} = f_{3i} + f_{\beta i} - \Sigma\, KC_{1j}(f_{\beta i})^j. \quad (10)$$

Substituting for f_{3i} from Eq. (9) in Eq. (10) we obtain

$$r_{3i} = Kf_{1i} - f_{\beta i} + f_{\beta i} - \Sigma\, KC_{1j}(f_{\beta i})^j, \quad (11)$$

which, using Eq. (7), can be rewritten as

$$r_{3i} = Kr_{1i}. \quad (12)$$

We normalize to get

$$r_{3i}/f_3 = Kr_{1i}/f_3. \quad (13)$$

Since we choose K such that

$$Kf_1 \sim f_3, \quad (14)$$

we arrive at the fact that the normalized residuals are equal

$$r_{1i}/f_1 \sim r_{3i}/f_3. \quad (15)$$

Noise or Hysteresis

Assume a set of data (f_1, f_3) is taken. A pair of functions is determined to approximate that data

$$F_{1i} = F_1(f_{\beta i}) + r_{1i} \quad (16)$$

and

$$F_{3i} = F_3(f_{\beta i}) + r_{3i} \quad (17)$$

where the r's are the residuals determined above. For simplicity we will assume that the residuals are zero. The residuals can simply be added to the final error result.

If, on a subsequent determination of f_1 and f_3, a deviation such as noise or hysteresis occurs, we seek the effect of the deviation on the residual error.

The disturbed or "noisy" measurements are denoted by primes. Let us assume that

$$f'_1 = f_1 + \Delta f_1 \quad (18)$$

and

$$f'_3 = f_3 + \Delta f_3. \quad (19)$$

In general, the two error terms are unknown.

The measured value of f_β is

$$f'_\beta = Kf'_1 - f'_3 = f_\beta + \Delta f_\beta. \quad (20)$$

where

$$\Delta f_\beta = K\Delta f_1 - \Delta f_3. \quad (21)$$

The error we want to determine is the difference between the measured value of f_1, i.e., f'_1, and the value of the function F_1 evaluated at f'_β,

$$e_1 = f'_1 - F_1(f'_\beta) \quad (22)$$

and the corresponding error for f_3

$$e_3 = f'_3 - F_3(f'_\beta). \quad (23)$$

If we assume small deviations, we can express $F(f'_\beta)$ as the first two terms of a Taylor series, i.e.,

$$F(f'_\beta) = F(f_\beta + \Delta f_\beta) \sim F(f_\beta) + \Delta f\beta(dF/df_\beta). \quad (24)$$

Therefore,

$$e_1 = f'_1 - F_1(f_\beta) - \Delta f_\beta(dF_1/df_\beta) \quad (25)$$

$$e_3 = f'_3 - F_3(f_\beta) - \Delta f_\beta(dF_1/df_\beta). \quad (26)$$

Using Eqs. (16), (17), (18), and (19) in (25) and (26) we have, after subtracting

$$e_1 = \Delta f_1 - \Delta f_\beta(dF_1/df_\beta) \quad (27)$$

$$e_3 = \Delta f_3 - \Delta f_\beta(dF_3/df_\beta). \quad (28)$$

Using Eq. (21) in (27) and (28) we have

$$e_1 = \Delta f_1 - (K\Delta f_1 - \Delta f_3)(dF_1/df_\beta) \quad (29)$$

$$e_3 = \Delta f_3 - (K\Delta f_1 - \Delta f_3)(dF_3/df_\beta), \quad (30)$$

which can be rewritten as

$$e_1 = \Delta f_1 - K\Delta f_1(dF_1/df_\beta) + \Delta f_3(dF_1/df_\beta) \quad (31)$$

$$e_3 = \Delta f_3 - K\Delta f_1(dF_3/df_\beta) + \Delta f_3(dF_3/df_\beta). \quad (32)$$

We take the derivative of Eq. (4) with respect to f_β to obtain

$$K(dF_1/df_\beta) - (dF_3/df_\beta) = 1. \quad (33)$$

Using Eq. (33) in (31) and (32) we have

$$e_1 = \Delta f_1 - \Delta f_1(1 + dF_3/df_\beta) + \Delta f_3(dF_1/df_\beta) \quad (34)$$

$$e_3 = \Delta f_3 - K\Delta f_1(dF_3/df_\beta) + \Delta f_3(dF_1/df_\beta - 1) \quad (35)$$

which can be reduced to

$$e_1 = \Delta f_3(dF_1/df_\beta) - \Delta f_1(dF_3/df_\beta) \quad (36)$$

$$e_3 = K(\Delta f_3(dF_1/df_\beta) - \Delta f_1(dF_3/df_\beta)). \quad (37)$$

From Eqs. (36) and (37) we see that

$$e_3 = K\, e_1, \quad (38)$$

which can be normalized, in analogy with the residuals from the previous section, to give

$$e_1/f_1 = e_3/f_3 = E \quad (39)$$

where

$$E = (\Delta f_3/f_3)\, K\, (dF_1/df_\beta) - (\Delta f_1/f_1)(dF_3/df_\beta) \quad (40)$$

i.e., the normalized errors are the same. It can be seen from Eq. (40) that the error in one mode is multiplied by the "slope" (in f_β space) of the other mode. Since f_β is a function of both modes, the errors are not independent. It can also be seen from Eqs. (29) and (30) that if $K\Delta f_1 = \Delta f_3$, both e_1 and e_3 are independent of the slopes.

References

1. J. A. Kusters, M. C. Fisher, and J. G. Leach, "Dual Mode Operation of Temperature and Stress Compensated Crystals," Proc. 32nd Ann. Symp. on Frequency Control, pp. 389-397, 1978.

2. S. S. Schodowski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," elsewhere in these Proceedings.

3. J. R. Vig, W. Washington, and R. L. Filler, "Adjusting the Frequency vs. Temperature Characteristics of SC-cut Resonators by Contouring," Proc. 35th Ann. Symp. on Frequency Control, pp. 104-109, 1981.

4. R. L. Filler, J. A. Kosinski and J. R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," Proc. 37th Ann. Symp. on Frequency Control, pp. 265-271, 1983.

5. R. L. Filler, J. A. Kosinski and J. R. Vig, "The Effect of Blank Geometry on the Acceleration Sensitivity of AT and SC-cut Quartz Resonators," Proc. 36th Ann. Symp. on Frequency Control, pp. 215-219, 1982

6. R. C. Smythe, Piezo Technology, Inc., private communications, Sept. 1982.

7. J. R. Vig, R. L. Filler and J. A. Kosinski, "SC-cut Resonators for Temperature Compensated Crystal Oscillators," Proc. 36th Ann. Symp. on Frequency Control, pp. 181-186, 1982.

8. R. L. Filler, J. A. Messina and V. J. Rosati, "Frequency-Temperature and Aging Performance of Microcomputer-Compensated Crystal Oscillators," elsewhere in these Proceedings.

9. D. L. Hammond, C. A. Adams, and A. Benjaminson, "Hysteresis Effects in Quartz Resonators," Proc. 22nd Ann. Symp. on Frequency Control, pp. 55-66, 1968.

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 Applications," Proc. 39th Ann. Symp. on Frequency Control, pp. 223-229, 1985.

11. M. Bloch, M. Meirs, and J. Ho, "The Microcomputer Compensated Crystal Oscillator (MCXO)", elsewhere in these Proceedings.

43rd Annual Symposium on Frequency Control - 1989

THE MICROCOMPUTER COMPENSATED CRYSTAL OSCILLATOR (MCXO)

MARTIN BLOCH, MARVIN MEIRS, AND JOHN HO
Frequency Electronics, Inc.
55 Charles Lindbergh Blvd.
Mitchel Field, NY 11553

ABSTRACT

The MCXO uses a new technique to achieve temperature compensation without use of ovens or conventional temperature-compensating components. The crystal oscillator in the MCXO, which is free to vary with temperature, operates on two modes simultaneously -- the fundamental (f_{1c}) and the third overtone (f_{3c}). The fundamental is multiplied by 3 ($3f_{1c}$) and subtracted from f_{3c} to produce a beat frequency (f_B). This beat frequency is used as a temperature indicator to correct the frequency of the fundamental or the third overtone, either of which may be used to generate the output (f_o) of the MCXO.

Five MCXOs were delivered to the U.S. Army by FEI. Their testing showed that the units provide the following performance:

Frequency vs. temperature	5×10^{-8} over the temperature range of -55 to +85°C
Aging	1×10^{-10} per day
Input power	41 mW

The output of the MCXO is a pulse train from which pulses have been deleted under microcomputer control. The beat frequency is counted and the microcomputer, which stores the relationship between the oscillator output (f_{1c} or f_{3c}) and f_B, determines the number of pulses that must be deleted from the pulse train per unit of time.

Several advantages accrue because this method of temperature compensation does not resort to frequency pulling. In many applications, the MCXO will replace conventional temperature-compensated crystal oscillators, providing an order-of-magnitude improvement in frequency vs. temperature performance. In applications requiring low power consumption, the MCXO will replace OCXOs.

This paper presents the details of how the MCXO operates and the details of the performance of the delivered systems.

INTRODUCTION

Because of limitations related to thermal hysteresis in AT-cut crystal units, lack of precision in analog compensating networks and the trim effect, efforts to improve the frequency accuracy of conventional temperature-compensated crystal oscillators (TCXOs) over a wide temperature range have proved unsuccessful.

The impetus for the development of the MCXO was the need for a low-power, high-accuracy timekeeping and frequency control in application such as tactical spread spectrum systems.

This requirement is reflected in the initial specifications for the MCXO: Frequency stability of less than 5×10^{-8} over the temperature range of -55 to +85°C; hysteresis of 1×10^{-8} over that temperature range; aging of 1×10^{-10} per day and 3×10^{-8} per year; and power consumption of less than 50 mW.

The development of the MCXO is continuing, but has already succeeded in meeting or exceeding most specifications. Five units have already been tested by the U.S. Army.

OPERATION OF THE MCXO

The MCXO uses two modes of an SC-cut crystal, which are simultaneously energized. This technique is fully explained in references 1 and 2. The MCXO uses the relationship between one of the "C" mode frequencies (f_{1c} or f_{3c}) and the beat frequency (f_B), as shown in Figure 1, to obtain a digital correction.

A simplified block diagram of the MCXO is shown in Figure 2. The two outputs of the dual-mode oscillator are the crystal frequency (fundamental f_{1c} or third overtone f_{3c}) and f_B, a beat frequency derived by mixing $3f_{1c}$ and f_{3c}. The beat frequency is used as a thermometer to indicate the actual temperature of the resonator One of the crystal frequencies will be used as the basis for the output pulse train and will be corrected before outputting from the MCXO.

The f_B signal is counted, with f_{1c} or f_{3c} as the interval timing reference. The count is applied to the microprocessor, which determines, based on this count and on the known frequency/temperature characteristics of the crystal, the number of pulses that must be deleted in each unit of time. The crystal oscillator frequency (f_{1c} or f_{3c}) is selected so that, at all temperatures, it is always greater than the required output frequency. This ensures that pulses can always be deleted from the pulse train to provide the required number of pulses in each interval. Typical values are: f_{1c} = 3.38+ MHz, f_{3c} = 10 MHz, f_B = 150 kHz, and f_o = 3.38 MHz. If the calculated correction includes a fraction of a pulse, the fractional value is stored in memory, and an additional pulse will be deleted when the memory, incremented in subsequent cycles, has reached or exceeded a full pulse.

Figure 3 is a more detailed Digital Control Block Diagram. The gate period and reading of the f_{3c} counter data are controlled by a microcomputer. The microcomputer additionally extends the gate time (multiple gate times) generating a period of approximately 1 sec. Upon completion of the period, the frequency count is input to the microcomputer. The microcomputer solves a polynomial equation and, during the next period, generates the correction required to the output frequency.

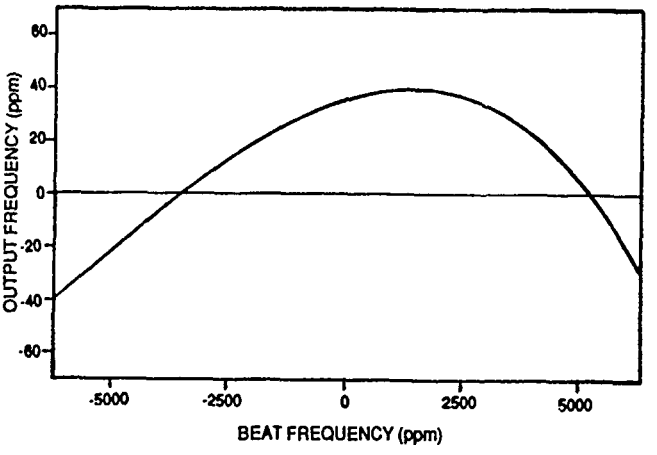


Figure 1.' Oscillator Output Frequency vs. Beat Frequency

CH2690-6/89/0000-016 $1.00 C 1989 IEEE

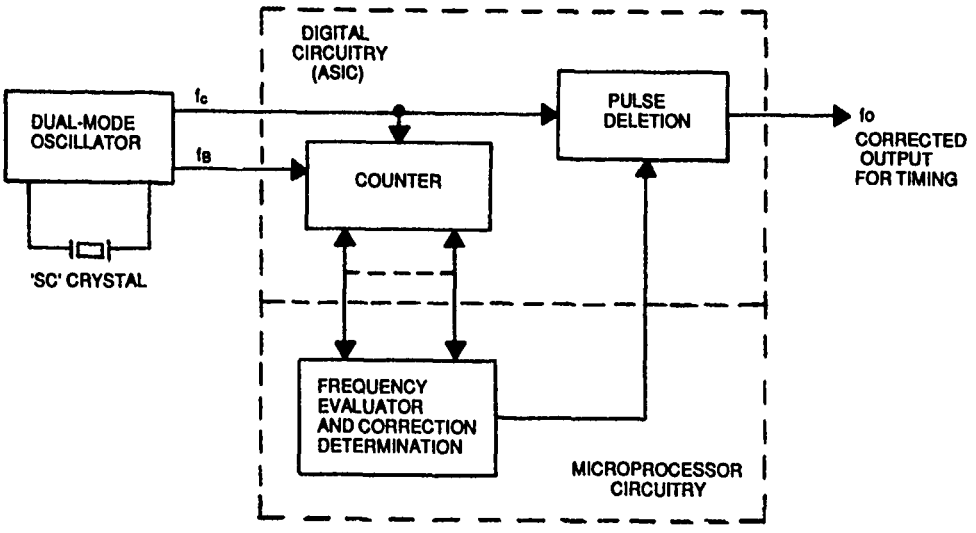


Figure 2. MCXO Block Diagram

SYSTEM EQUATIONS

The system equations for the resolution and accuracy under static and dynamic conditions are show below for an f_{3c} output frequency:

I. Beat Frequency:

$F_B = 3xF_{1C} - F_{3C}$

Where F_B = Beat Frequency

F_{IC} = Fundamental Frequency, C Mode

F_{3C} = 3rd Overtone Frequency, C Mode

II. Count Resolution Required:

F_{3C} Temp Coef = $100x10^{-6}$ from -55°C to +85°C

For Stability of $1x10^{-8}$

Enhancement Required = $\frac{100x10^{-6}}{1x10^{-8}} = 10,000$

F_B Temp Coef = $10,000 \times 10^{-6}$ from -55°C to +85°C

Using F_B Temp Coef

Measurement Resolution Required
= $\frac{10,000x10^{-6}}{10,000} = 1x10^{-6}$
(Required Enhancement)

Count F_{3C} (10+ MHz) for 1 sec period

System Measurement Resolution = $1x10^{-7}$

Yields Accuracy of $0.1x10^{-8}$

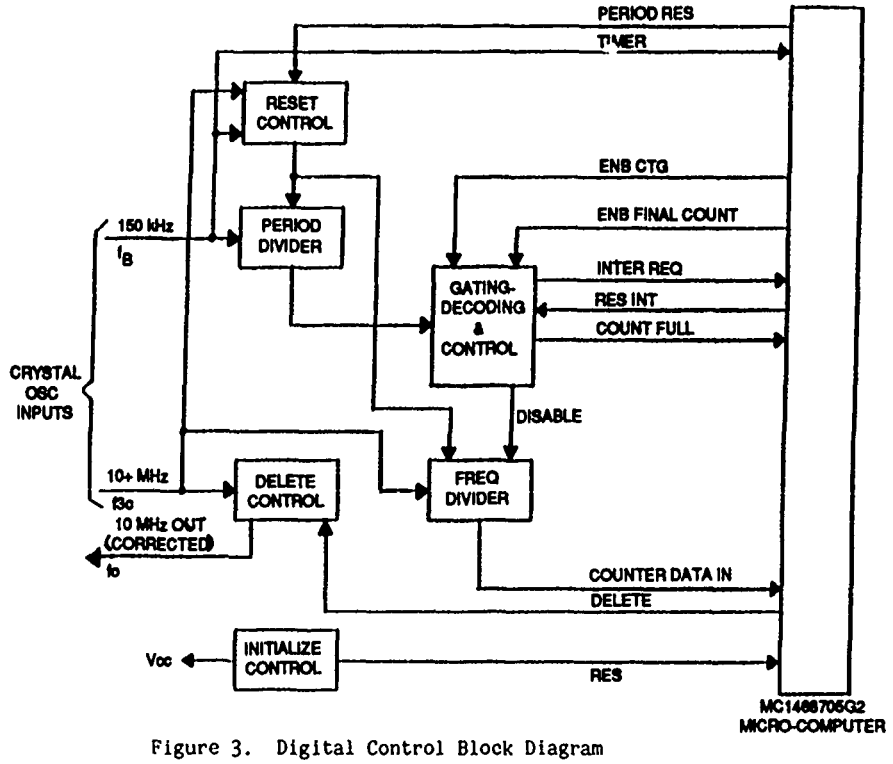


Figure 3. Digital Control Block Diagram

III. Maximum Temperature Slope Requirement:

F_{3C} Slope Max = 2×10^{-6}/°C

F_{3C} Slope Avg = $\frac{(100) \times (10^{-6})}{140} = 0.71 \times 10^{-6}$/°C

or

$\frac{F_{3C}\ Max}{F_{3C}\ Avg} = \frac{2 \times 10^{-6}/°C}{.7 \times 10^{-6}/°C} = 2.8$

F_{3C} Max = 2.8 F_{3C} Avg

..For F_{3C} Max Slope, Resolution Required

Yields Accuracy of 0.28×10^{-8}

IV. Maximum Temperature Slew:

Max Rate = 2°C/Min = 3.33×10^{-2}°C/Sec

Max Slope = 2×10^{-6}/°C

Max Slew = $(3.3 \times 10^{-2}$°C/Sec) x $(2 \times 10^{-6}$/°C)
= 6.6×10^{-8}/Sec

Max Change = $(6.6 \times 10^{-8}$/Sec) x (10^7 counts)
= .66 Counts/Sec

SYSTEM BUILDING BLOCKS

1. SC-cut quartz crystal.

   An SC-cut lateral-field quartz crystal is used in this application. The advantage of using the lateral field is that the unwanted "B" modes, which are 8% higher in frequency than the "C" modes, are suppressed. This allows the use of wide-band, low Q electronic circuits with less pulling effect upon the final output frequency. The quartz crystals used for the five units tested by the U.S. Army were in C holders. Future units are planned using ceramic flatpacks, which will have the advantage of providing even better aging than has been achieved to date.

2. Dual-mode hybrid oscillator.

   A custom hybrid has been manufactured in a 1 x 1 inch assembly. Internal to this hybrid are the dual-mode oscillator, a multiplier that multiplies the fundamental by 3, a mixer that provides an output at the beat frequency, and amplifiers for the beat frequency output and either the fundamental or third overtone.

3. Microcomputer.

   A CMOS microcomputer is used to perform the computations to determine the pulse deletions required. The frequency vs. beat frequency curve is stored as a 5th order curve-matching algorithm in the microcomputer.

4. Gate array.

   A custom gate array with approximately 2550 gates was developed to perform all of the additional digital functions not performed in the microcomputer. This gate array performs the high frequency counting and pulse deletion and provides the interface to the computer.

PERFORMANCE

The power consumption of the MCXO has been reduced since the early periods of development until it is now well below the objective of 50 mW. The power consumed in each of the major assemblies for an f_{1c} output is:

Dual-mode oscillator	8 mW
Microcomputer	18 mW
Gate array	15 mW
Total	41 mW

The frequency versus temperature, hysteresis, and aging characteristics of the MCXO were determined as part of the U.S. Army's testing of the five delivered units. Their data are presented in another paper (ref. 3). Their evaluation has verified the feasibility of achieving frequency vs. temperature stability of ± 5 x 10^{-8}, aging of 1 x 10^{-10}/day, and hysteresis of no more than 3 x 10^{-8}.

PHYSICAL DESIGN

The existing MCXO units have been supplied in a 3.3 x 2.1 x 1.75 inch package (5.2 cubic inches). This package is seen in Figures 4 and 5. A newer package is being developed to reduce the dimensions to 1.45 x 1.3 x 1.2 inches (2.2 cubic inches).

ADDITIONAL FEATURES

A calibration input from a more accurate frequency standard can be applied to the MCXO to automatically recalibrate the unit and provide improved accuracy. The

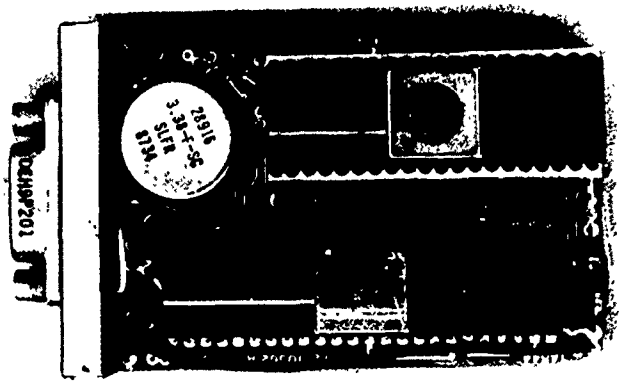


Figure 4. Top View, Cover Removed

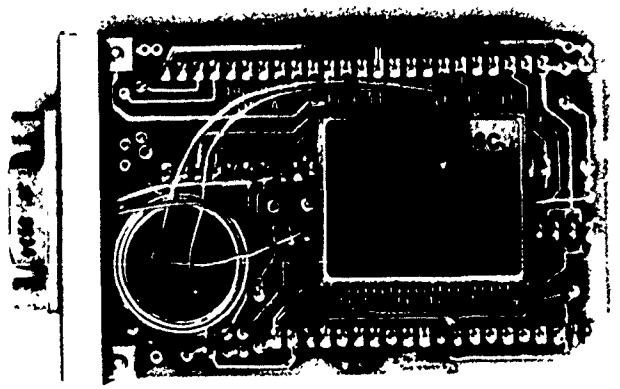


Figure 5. Bottom View, Cover Removed

calibration input is counted and compared with the output pulse count to generate an additional correction that gives the MCXO the accuracy of the frequency input.

The basic MCXO already described provides a precise pulse train output that is ideally suited for timing reference applications. As seen in the block diagram in Figure 6, a sine wave output at almost any frequency can be provided.

A spectrally pure sine wave output is generated by locking a VCXO to the MCXO. A submultiple of the VCXO is compared to a submultiple of the MCXO to develop an error frequency. This error is converted to a voltage that controls the VCXO.

A time of day (TOD) display can be incorporated in the MCXO. The output pulse train is divided down to 1 pps and this is further divided in the TOD circuit to provide 1 ppm and 1 pphr, or digital counts of seconds, minutes, and hours as required for the display.

SUMMARY

The new approach to temperature compensation in the MCXO compares the different temperature responses of the fundamental and third overtone to provide a sensitive indication of the correction needed in the output frequency. This approach eliminates the need for a temperature-controlling oven, which is needed in other approaches to maintain this accuracy. In many applications, an oven could not be used because it would increase power consumption as well as size.

This approach, which does not correct the frequency at the oscillator, also eliminates the need for external temperature-compensating components whose temperature characteristics would be used to correct those of the crystal over a limited temperature range. This MCXO approach, unlike temperature compensated crystal oscillators, does not result in frequency pulling of the oscillator. Instead, in the MCXO, the dual-mode oscillator signals are free to vary with temperature. However, the temperature characteristics of the crystal in use have been determined and stored in the microcomputer. Frequency compensation is achieved under microcomputer control by deleting pulses from an output pulse train derived from the fundamental or third overtone signal.

ACKNOWLEDGMENT

The authors thank the U.S. Army Electronic Technology and Devices Laboratory for their technical guidance and support provided under Contract DAAL01-85-C-0403.

REFERENCES

[1] S. Schodowski "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[2] R.L. Filler and J.R. Vig "Resonators for the Microcomputer Compensated Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

[3] R.L. Filler, J.A. Messina, and V.J. Rosati "Frequency-Temperature & Aging Performance of Microcomputer-Compensated Crystal Oscillators," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

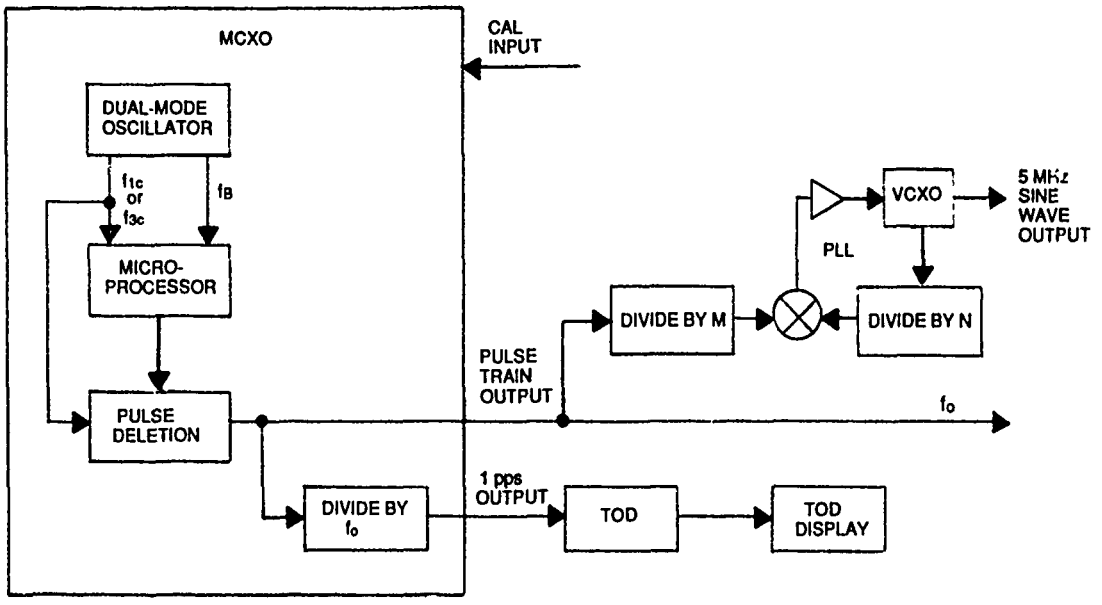


Figure 6. MCXO with Additional Features

43rd Annual Symposium on Frequency Control - 1989

A MICROCOMPUTER-COMPENSATED CRYSTAL OSCILLATOR USING A DUAL-MODE RESONATOR

A. BENJAMINSON General Technical Services, Inc.
105 Clermont Avenue, Alexandria, VA 22304

S. C. STALLINGS, Practical Micro Design, Inc.
105 Clermont Avenue, Alexandria, VA 22304

Abstract

Temperature compensation to date has been performed by using analog correction operating through a varicap to control a crystal oscillator. The system described uses microcomputer-controlled compensation wherein the correction is implemented by adding a computed offset frequency to the output of a stable oscillator. Temperature sensing is done by the use of a dual-mode oscillator system employing two C-modes of an SC-cut crystal, which are combined to produce a temperature-sensitive difference frequency. System accuracies of ±20 ppb have been achieved over a temperature range of –55 to +85 degrees Celsius.

Introduction

This paper describes a 10 MHz MCXO (microcomputer-compensated crystal oscillator) using an SC-cut resonator in a dual-mode oscillator without varicap control. The resonator vibrates at two C-mode frequencies simultaneously. The difference in frequency between the two modes is utilized as a temperature indicator. An initial calibration run from –55 to +85 degrees Celsius provides frequency versus temperature data. The calibration results are stored in the form of coefficients for a least-squares-fitted seventh-order polynomial which models the data. The microcomputer then uses this model to perform frequency correction calculations. Since the temperature information is obtained directly from the resonator itself rather than from an external sensor, temperature offset and lag effects are minimized. Temperature velocity effects are also minimized by the stress compensation characteristics of the SC-cut resonator. Overall system error is less than ±20 ppb over the temperature range of –55 to +85 degrees Celsius.

Four modes of operation are available in the MCXO: Frequency Mode, Clock Mode, Aging Correction Mode, and Calibration Mode. The Frequency Mode produces both a corrected 10 MHz output and a 1 pps clock output. The Clock Mode operates with reduced power and produces only the 1 pps clock output. The Aging Correction and Calibration modes are special modes used for setting the correction values into the unit.

The Frequency Mode Compensation System

The MCXO system is driven by two oscillators operating on a single SC-cut resonator.[1,2] One oscillator excites the third overtone C-mode, while the second excites the fundamental C-mode. Since both modes have essentially identical temperature characteristics except for their first-order terms, the difference in frequency between the overtone mode divided by three and the fundamental mode is a nearly linear function of temperature. Figure 1 shows a plot of seventh-order polynomials which have been least-squares-fitted to the two frequencies (with the fundamental multiplied by three) versus N_1, a number derived from the period of the difference frequency. The 3*ff curve has been normalized so that its intercept of $N_1 = 0$ is at the same point as the f_o intercept of $N_1 = 0$. A normalized plot of N_1 versus temperature is shown in Figure 2 to illustrate the excellent performance of this temperature sensing technique. The crystal is plated so that maximum frequency of the third overtone (f_o) is about 200 Hz below 10 MHz. This produces a fundamental frequency (f_f) at about 3.384 MHz. As suggested above, the frequency difference is a nearly linear function of temperature and serves as the thermometer for the compensation system.

As indicated in Figure 1, the third overtone frequency varies more than 700 Hz over the temperature range of –55 to +85 degrees Celcius. The MCXO system is designed to reduce this variation by a factor of 1 part in 3500 to a value of less than 0.2 Hz. The block diagram of the system is shown in Figure 3.

In Figure 3, the oscillators are represented by the two boxes at the far left of the block diagram. The digital output of the overtone mode (OXO) is divided by three and mixed with the digital output of the fundamental mode (FXO) using an exclusive-or gate. A sampling latch is used to remove the sum frequency, leaving only the difference frequency. This frequency is further divided by 2^{16} to produce f_t which is applied to a period-measuring counter. The clock for the counter is derived by dividing the output of the third overtone oscillator by two. These divisions provide range scaling as well as integration and smoothing of the temperature indication number N_1, derived from the period of the difference frequency.

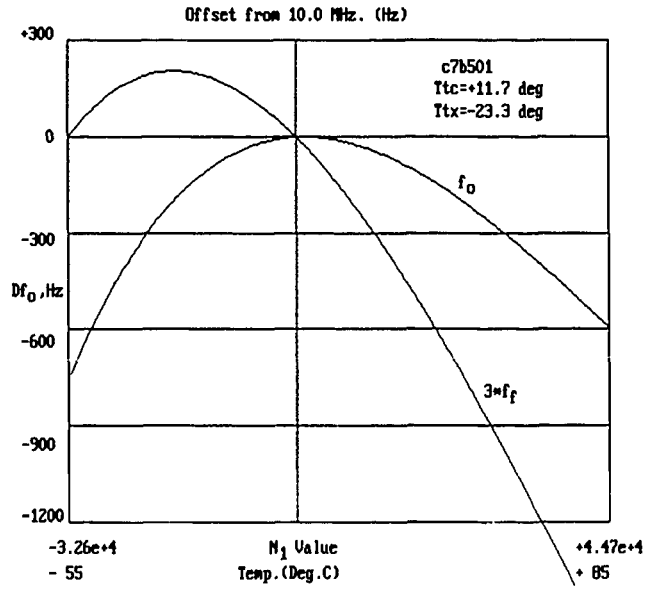


Figure 1. Dual-Mode Polynomials

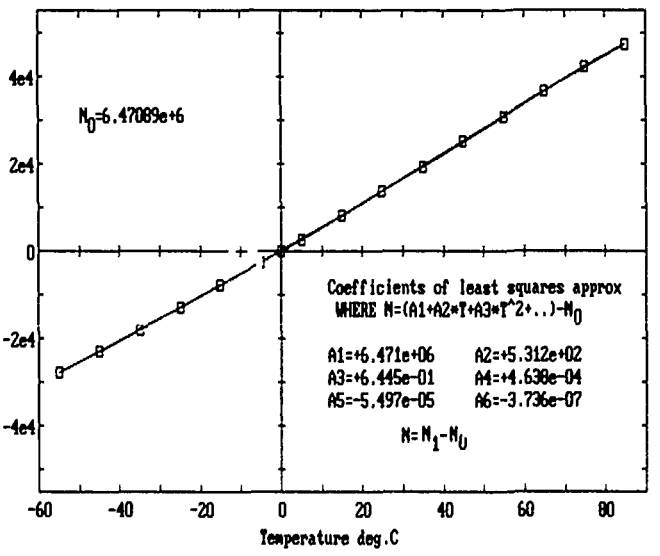


Figure 2. N versus Temperature

CH2690-6/89/0000-020 $1.00 C 1989 IEEE

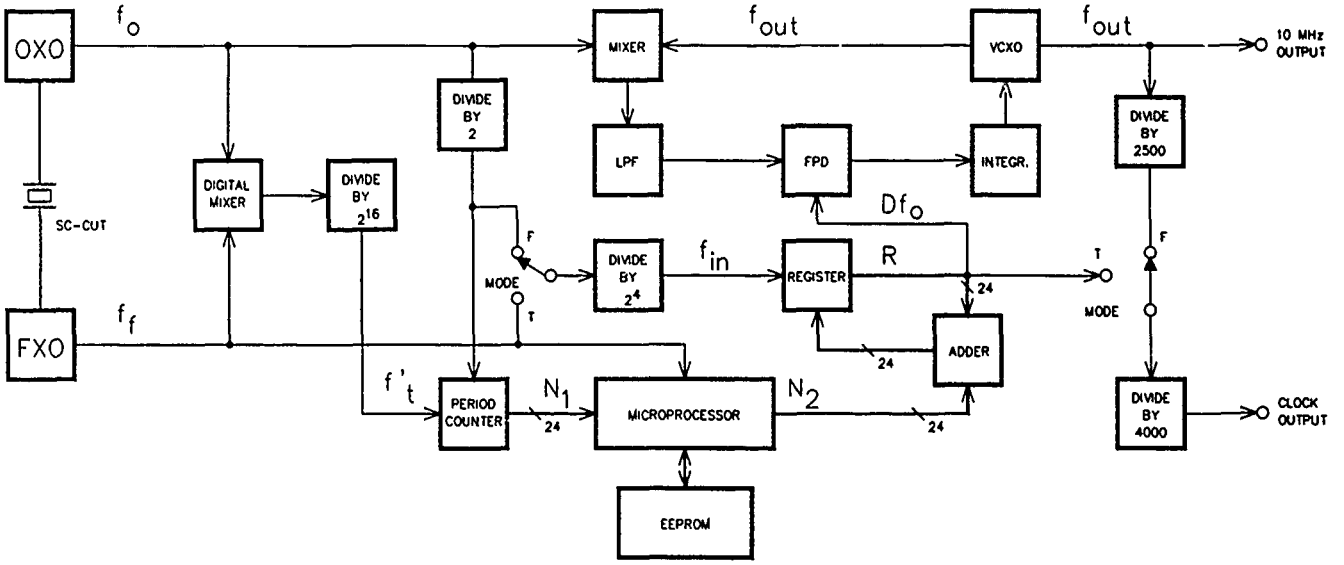


Figure 3. MCXO Block Diagram

N_1 can be expressed as the ratio of the two frequencies applied to the period counter, where:

$$N_1 = f_o / (2 f'_t) \quad \text{where} \quad f'_t = (f_f / 2 - f_o / 6) / 2^{16}$$

N_1 is used as the independent variable, representing temperature, for the seventh-order polynomial which is used to predict the actual frequency of the third overtone. The polynomial coefficients are derived from a least-squares curve-fitting routine using data taken during a temperature calibration run.

The microcomputer solves the polynomial during each correction cycle and subtracts the resulting frequency from the target 10 MHz frequency to get the difference Df_o. The microcomputer then computes a second number N_2 which is used to control a direct digital frequency synthesizer (DDS) operating from the third overtone divided by 2^5. The DDS produces a correction frequency which equals Df_o to within the DDS's resolution of 0.02 Hz. The DDS resolution establishes the resolution of the system.

In the Frequency Mode, a phase-locked loop (PLL) is employed to establish the final output frequency at 10 MHz. A VCXO (voltage-controlled crystal oscillator) using a 10 MHz AT-cut resonator is phase-locked to the sum of f_o, the actual third overtone frequency, and Df_o, the computed frequency error. The loop gain of the PLL is more than adequate to slave the VCXO to the sum frequency under all operating conditions. The phase-noise spectrum beyond the PLL bandwidth (nominally 10 Hz) is dominated by the spectrum of the VCXO, but within that band the stable third overtone oscillator controls the phase noise and short-term stability.

A "sleep period" between correction cycles is used to conserve power during frequency mode operation. During the sleep period the microcomputer shuts down portions of the digital logic and then enters a reduced-power mode itself. At the end of the sleep period the microcomputer powers up, re-enables the digital logic previously shut down, and proceeds with the next correction cycle.

During each cycle the microcomputer software compares the newest calculated value of Df_o with the last value calculated. If the difference exceeds 0.2 Hz, the sleep period is reduced to zero. If the difference is less than 0.2 Hz, the sleep time is set to one second and doubles every measurement cycle (as long as the difference is still less than 0.2 Hz), until a limit of 128 seconds is reached. The results of each measurement are utilized for correction even if the difference is below the 0.2 Hz threshold. This allows the system to follow a very slow drift even while operating with maximum sleep periods.

While in the Frequency Mode, the clock output is derived by dividing down the VCXO output. The division is done in two stages, first by 2500 and then by 4000, ultimately yielding a one pulse per second digital output. The division is done in two stages so that the division by 4000 can be shared with the Low-Power Clock Mode.

The Low-Power Clock Compensation System

In the Low-Power Clock Mode, the PLL and VCXO are always shut down and the synthesizer is driven from the fundamental-mode oscillator followed by a division by 2^4. Measurement of N_1 and computation of N_2 are done in a fashion similar to the one used in Frequency Mode. The difference is that a second set of polynomial coefficients describing the frequency versus temperature characteristic of the fundamental oscillator is used to calculate the actual frequency of the fundamental. N_2 is then computed so that the DDS produces a precise 4 KHz at its output despite the fact that the fundamental frequency is varying by as much as 600 Hz. The DDS output is then divided by 4000 using the shared portion of the clock divider chain. Switching between divider chains when changing modes is done in a nearly synchronous manner so as to reduce phase jitter in the one pulse per second output.

The Low-power Clock Mode also employs a sleep period. During the Low-Power Clock Mode's sleep period, the microcomputer shuts down the third overtone mode (OXO) oscillator and portions of the digital logic and then enters a reduced power mode itself. At the end of the sleep period the microcomputer powers up and re-enables all the digital logic and the OXO. It then waits a short period of time to allow the OXO to stabilize before measuring N_1 for the next correction cycle.

The microcomputer software compares the calculated value of N_2 with the value calculated during the previous measurement cycle. If the absolute difference is greater than 1 ppm, the sleep period is reduced to zero. If the difference is less than 1 ppm, the sleep time is set to one second and doubles every measurement cycle (as long as the difference is still less than 1 ppm), until a limit of 128 seconds is reached. Results of each measurement are utilized for correction even if the difference is below the 1 ppm threshold. This allows the system to follow a very slow drift even while operating with maximum sleep periods.

The Aging Correction System

The effects of crystal aging are represented by changes in the constant term of the polynomial which predicts the crystal's actual frequency. To

make adjustments in the constant term the system is placed in Aging Correction Mode. An external 10 MHz reference standard is fed into the unit. The unit operates as if in the Frequency Mode but also monitors a digital phase comparator which looks at the slippage between f_{out} and the external reference. If the slippage indicates an error in f_{out} of greater than 10 ppb the constant term is adjusted and the process is repeated. When the error is less than 10 ppb, the new value of the constant term is saved and the Aging Correction Cycle is complete. This method is used because it evaluates the actual output of the unit, is fully automatic, and is self-contained (with the exception of the external reference standard).

The Calibration Mode and Procedures

Calibration is performed on complete units, with the crystal and its supporting electronics henceforth treated as a matched pair, so that the corrective measures applied will take into account all sources of error. Performance data is collected on each unit as it is exposed to temperatures ranging from –55 to +85 degrees Celsius in an environmental test chamber. These temperature tests, undertaken for each system, accumulate the values of N_1, the variable used to represent temperature, and Df_o, the deviation of f_o from an external 10 MHz reference standard. A least-squares curve fit of a seventh-order polynomial is performed on the data. The resulting constant term and coefficients for Frequency Mode are stored in the unit. The data is then manipulated to extract information for the fundamental mode alone, another curve fit is done, and the constant term and coefficients for Clock Mode are stored in the unit.

The measurement data is taken by digital logic within the actual unit. N_1 data is taken from the same logic used in Frequency Mode. Df_o is taken from an additional logic system very similar to the N_1 logic. The processor in the unit collects both numbers and transmits them via a serial link to an external IBM-PC. The IBM-PC which records the data also controls the environmental chamber. When the temperature run is completed, the IBM-PC performs the curve fitting and sends the results back to the MCXO unit via the serial link. The MCXO unit stores the results for future use and the calibration procedure is complete.

Design of the Dual-Mode Oscillators

The sustaining amplifiers used in a dual-mode oscillator system must be designed to provide regeneration for their respective modes only, while rejecting all other modes and overtones. Since the impedance of the resonator at the fundamental mode is also much lower than it is at the third overtone, the oscillator circuits were selected to provide maximum stability on each mode.

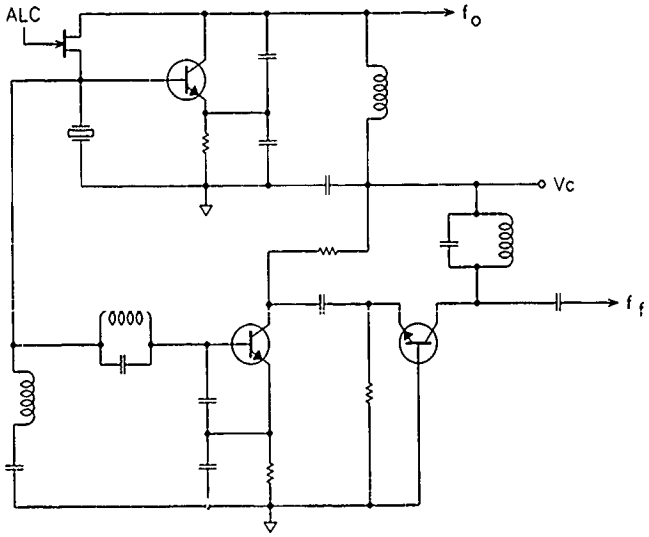


Figure 4. Dual-Mode Oscillators

A bridge oscillator circuit was selected for the third overtone oscillator and is shown in Figure 4. This design is based on the original Meacham bridge oscillator in which the resonator is treated as a series-tuned circuit in a resonance bridge that is precisely balanced in phase and partially balanced in amplitude. The sustaining amplifier magnifies the residual amplitude signal to provide drive for the bridge, as well as the load. A narrow-band tuned circuit produces the necessary phase reversal at the frequency of operation only, rejecting the B-mode and all other modes and overtones.

The configuration selected for the fundamental mode is an impedance-inverting oscillator. This is basically a Colpitts oscillator preceded by a lumped circuit quarter-wave line. The line inverts the low series-resonant impedance of the crystal to present a high impedance to the sustaining amplifier. Additional reactive elements are added to filter out the signal from the other oscillator and to prohibit oscillation in other modes.

Series-mode oscillators were selected for use in the dual-mode system because the crystal impedances and voltages at the sustained frequencies are at a minimum. This helps to reduce crosstalk between the circuits. The resonant impedance of the third overtone is typically 75 ohms, while that of the fundamental is less than 10 ohms. The two oscillator circuits were selected on the basis of these values, since the bridge oscillator is better suited for high impedance resonators than the Colpitts design is, and vice versa.

While the Colpitts circuit is satisfactory as the low-frequency oscillator, it does not produce as much output power as the bridge oscillator. Additional power gain and load isolation were achieved using the Isolated-Colpitts configuration shown in Figure 4. This design channels the collector current through a grounded-base amplifier so that the oscillator signal flows directly to the tuned load to achieve the necessary 3 volts peak-to-peak, with the excellent isolation afforded by a grounded-base circuit.[3,4]

The low crystal impedance of the fundamental mode can easily be degraded by the AC resistance of the series inductor. Therefore, only stable high Q coils are used to maximize the loaded Q and stability of the oscillators.

ALC (automatic level control) was also added to the bridge oscillator to increase the loaded Q of that oscillator. A characteristic of bridge oscillators is that the loaded Q increases as the bridge approaches balance. The sustaining amplifier gain, however, must also be increased to offset the reduced output from the bridge. When self-limiting action is used, the bridge must be unbalanced in favor of high loop gain to ensure start-up. Subsequent limiting action then reduces the amplifier gain and the loaded Q. In the MCXO, power restrictions limit the available gain, so that ALC is used to maximize the gain and thus the loaded Q. ALC also provides significant collateral benefits: it controls crystal current, and also reduces the starting time of the oscillator. The use of ALC may also mitigate the effects of radiation on crystal resistance.

Since the output level of bridge oscillators is high, only one cascode amplifier stage was needed to implement the ALC and provide the necessary isolation. This amplifier stage provides buffering between the bridge oscillator and the first CMOS gate, and also provides high-level drive for the ALC rectifier. The ALC circuit is fairly simple. It uses a JFET in place of the bridge feedback resistor, plus two each of Schottky diodes, resistors, and capacitors. The amplifier output is set at 3-4 volts peak-to-peak to ensure adequate drive for the CMOS logic.

The third overtone oscillator must be shut down during the Clock Mode, without disturbing the fundamental oscillator (which runs continuously). This is done by opening the base-emitter circuits of the overtone oscillator and its amplifier. The open drain terminals of a CMOS inverter are connected to the emitter circuits of the oscillator and amplifier. When the gate output is low, the drain-source path provides a low resistance to ground of about 30 ohms. When the gate output is high, the drain-source path is open. The emitters are then floating, and no base-emitter diode action occurs to degrade the FXO performance. Testing has shown that FXO frequency changes no more than 1 ppb when the OXO is turned off.

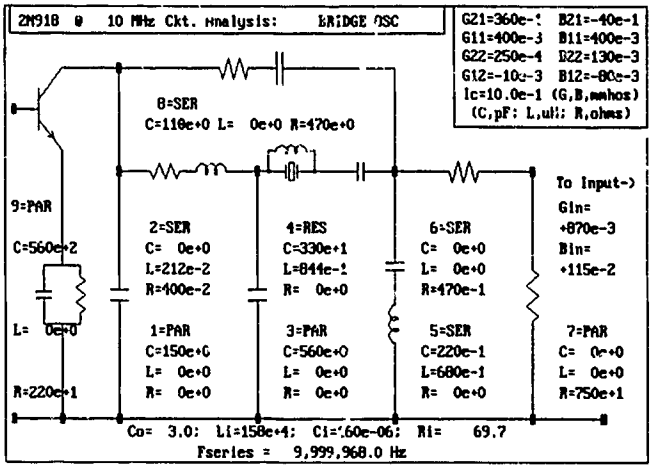


Figure 5. Analysis Circuit for Third Overtone Oscillator

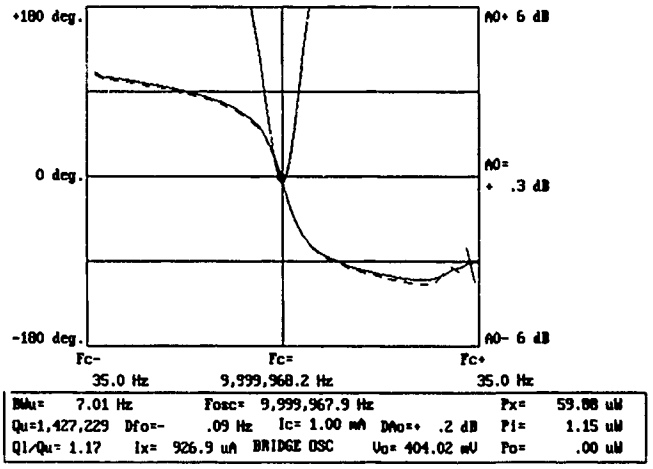


Figure 7. Open-loop Response of Bridge Oscillator

```
BRIDGE OSC. ANALYSIS, Fosc = 9,999,967.9 Hz
Limiting signal levels in circuit analysis branches .
-------------------------------------------------------------
Iout = 1.269 mA.rms.:   Angle of Vf = - 6.3 deg
-------------------------------------------------------------
V1=+40.4e+01 mV,@ - 16 deg.     I1=+38.1e-01 mA,@ + 74 deg
V2=+50.2e+01 mV,@ - 12 deg.     I2=+37.7e-01 mA,@ +259 deg
V3=+10.3e+01 mV,@ +184 deg.     I3=+36.2e-01 mA,@ - 86 deg
V4=+65.0e+00 mV,@ +179 deg.     I4=+92.7e-02 mA,@ +185 deg
V5=+38.5e+00 mV,@ +191 deg.     I5=+10.9e-03 mA,@ +101 deg
V6=+26.8e-01 mV,@ +238 deg.     I6=+57.0e-03 mA,@ +238 deg
V7=+36.7e+00 mV,@ +188 deg.     I7=+49.0e-04 mA,@ +188 deg
V8=+43.8e+01 mV,@ - 14 deg.     I8=+89.2e-02 mA,@ +  3 deg
-------------------------------------------------------------
Icom=+12.4e-01 mA,@ -  2 deg.   Iin=+53.0e-03 mA,@ +241 deg
```

Figure 6. Analysis of Bridge Oscillator

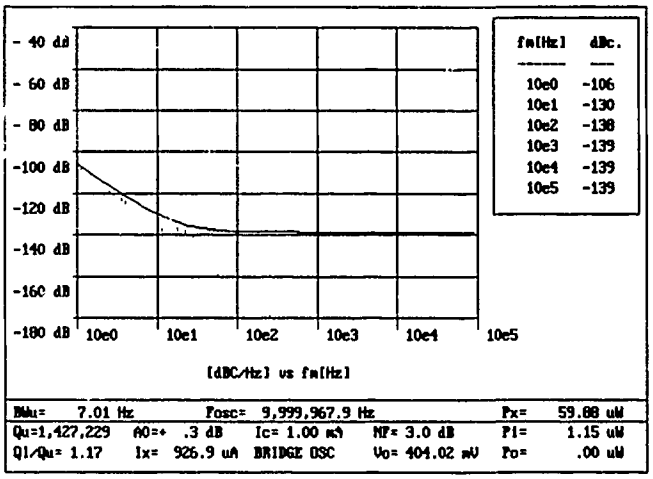


Figure 8. Noise Characteristics of Bridge Oscillator

With ALC, the overtone oscillator turns on in less than 20 milliseconds. The ALC unbalances the bridge in favor of fast oscillation buildup, until the controlled level is reached and the bridge stabilizes.

A micropower voltage regulator provides a regulated 4 volts DC to the two oscillators, the VCXO and their amplifiers. It provides sufficient voltage regulation to isolate the oscillators from the power source, while drawing only about 75 microamperes of quiescent current at a supply voltage of 5 ±0.5 volts DC.

Oscillator Stability Considerations

These new circuits have been designed very carefully in an attempt to meet the conflicting MCXO goals of low oscillator power consumption and maximum stability over a wide temperature range, while running on two modes simultaneously.

Much of the analysis was done using CODA,[6] a new software program now available for *C*rystal *O*scillator *D*esign and *A*nalysis. Examples of CODA output are shown in Figures 5 through 10. CODA allows the designer to evaluate quickly and accurately the effects of component changes on oscillator performance.

High stability requires the use of moderately high transistor current to increase transconductance and permit the use of larger capacitors that can swamp out circuit strays and raise operating Q. Large capacitors also allow the necessary inductors to be smaller and sometimes to be made without iron cores.

Lowering oscillator current reverses the situation forcing the use of iron-cored coils, particularly at lower frequencies. Minimizing total current also forces a reduction in the number of stages employed, and this too affects stability by reducing load isolation.

Dual-mode operation requires the use of inductors to aid in separating the modes. The number in use has been reduced by careful design, but three critical inductors remain that cannot be eliminated without serious compromises in circuit design.

Oscillator frequency variations with temperature are also induced by changes in the values of inductance and capacitance in the oscillator circuits. The capacitors used in critical areas have a TC (temperature coefficient) of ±30 ppm per degree Celcius, and the inductors have a TC ranging from +20 to +50 ppm per degree, with air-core inductors having a lower TC than the iron-core units. These variations are not serious as long as they are repeatable, since the calibration run takes them into account.

While iron-core coils might seem to pose a greater problem than air-core coils, in practice it is not their TC itself that causes trouble but rather hysteresis in the TC curve. Both air-core and iron-core coils have been measured. The air-core coils do exhibit lower TC but no less hysteresis than the iron-core units. In this regard the iron-core units seem to perform as well as the air-core units.

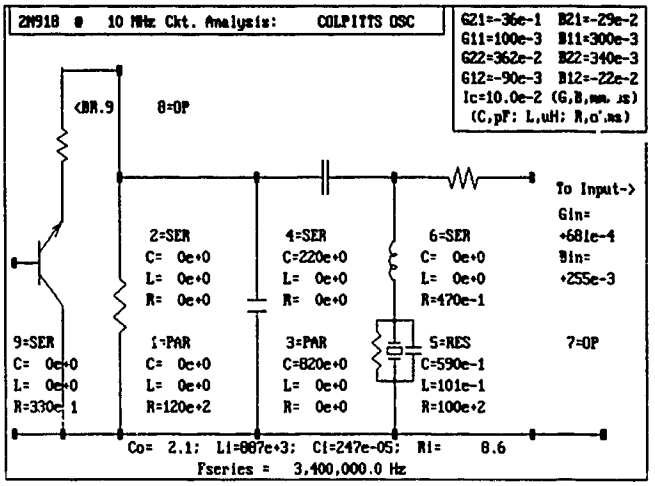


Figure 9. Analysis Circuit for Fundamental Oscillator

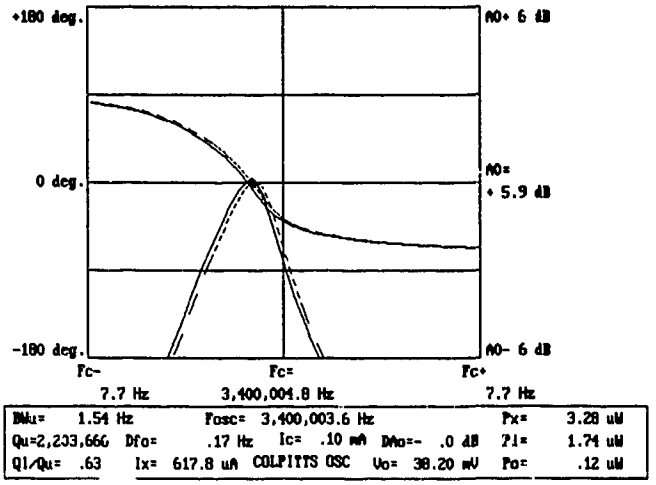


Figure 10. Open-loop Response of Colpitts Oscillator

Hysteresis, in either these reactances or the resonator, is a limitation on repeatability. Total circuit hysteresis (excluding the resonator) has been less than 400 ppm per degree (measured at 25 degrees Celcius).

CODA analysis of the oscillators has shown that a one percent change in inductance will change the bridge oscillator frequency by $1*10^{-8}$ (see Figure 7), and the Colpitts oscillator by less than $7*10^{-8}$ (see Figure 10). This suggests that a 400 ppm change in inductance will pull the overtone frequency by $4*10^{-10}$, and the fundamental frequency by about $2.5*10^{-9}$.

The FXO is less stable because it operates on the fundamental mode and at a lower collector current than the OXO.

Much more data will be needed to show the extent of component aging in these circuits, and how it might compare with resonator aging over this wide temperature range. Hysteresis, however, does not seem any worse in iron-cored coils than in air-cored units. Keep in mind that copper wire is common to all the coils, and that it has a positive TC. It may very well be the dominant factor, since it is very difficult to suppress the expansion and contraction of the copper. A solenoid that is wound on a stable form, such as ceramic, should have the wire stretched nearly to its elastic limit to force the wire to conform to the form. In a toroid, however, it is very difficult to stretch the wire and this may result in hysteresis and aging due to movement of the wire itself.

The Phase-Locked-Loop

The PLL circuit uses an edge-triggered digital phase-frequency detector followed by a very low-powered BIMOS Op-Amp. The Op-Amp acts as an integrator and lag-lead filter creating a Type-2 third-order control loop. It also includes an additional R-C network to increase sideband suppression in the PLL.

The integrating action of the PLL forces the VCXO to track the sum of the third overtone oscillator and synthesizer frequencies without error to produce the temperature-corrected output.

An algorithm was developed to enable the optimization of the loop for maximum suppression along with good loop stability. A unity-gain frequency of 10 Hz was selected as a compromise between sideband reduction and a low PLL noise floor.

An Isolated-Colpitts circuit was selected for the VCXO, since the fundamental mode AT-cut resonator also has a low crystal resistance. The impedance-inverting design used in the FXO was not used here since the VCXO does not operate at its series-resonant frequency but requires a pulling capacitor for frequency control.

The combination of Isolated-Colpitts and Grounded-Base amplifier is similar to a Cascode configuration and as such provides good isolation from changes in loading effects, using only two transistors. Since the AC collector current flows directly through the grounded base amplifier to the high impedance L-C circuit and inverter, high voltage and power gains are obtained. The isolation also reduces crystal loading so that all that remains is the loading produced by the input and output impedances of the sustaining amplifier.

Since the VCXO is phase-locked to the sum of f_o and Df_o by the digital phase-frequency detector (PFD), it will assume the stability of these signals within the 10 Hz bandwidth of the PLL filter.

At about 10 Hz there is a transition from the spectrum of the VCXO to that of the third overtone oscillator whose short-term stability dominates the close-in phase noise through the operation of the PLL. The noise floor for the system is about 140 dBc.

The AT-cut crystal used in the VCXO has the high value of motional capacitance (12 fFd) and the low resistance (less than 10 ohms) needed for a VCXO. These values permit operation with the limited capacitance range of the varicaps available for low-voltage operation (about 15 to 22 pFd). The pulling range must be large enough to compensate for the TC of the AT-cut crystal only, since the output frequency is fixed at 10 MHz. It must include about 400 Hz for deviation in frequency due to the temperature coefficient of the AT-cut, plus 200 Hz for initial frequency tolerance, plus a reasonable margin for reliable operation.

The Direct Digital Synthesizer and Phase Noise

A Direct Digital Synthesizer (DDS) is used to generate the correction frequency required to offset the temperature variations of the oscillators.[5] The circuit elements can be seen in Figure 3. The DDS is composed of a 24-bit adder coupled with a 24-bit register. It adds the frequency word N_2 to the register output R once each clock cycle so that $R+N_2$ approaches R in modulo 2 arithmetic. The register will overflow, on average, once every $2^{24} / N_2$ clock periods, thus producing an output whose frequency is:

$$f_d = N_2 * f_{in} / 2^{24}$$

The output is available as a square wave at the most significant bit (MSB) of the register. Its frequency is controlled by the processor to precisely match Df_o, the temperature offset of the third overtone oscillator. The required value of N_2 is obtained from the following equation.

$$N_2 = f_d (2^{29}) / (10^7 - Df_o)$$

The resolution is constant and defined by $f_{in} / 2^{24}$ which, with the values shown in the figure, is 0.0186 Hz, or less than $2 / 10^9$ at 10 MHz. The maximum value required of N_2 will be less than 2^{16} in the frequency mode and 2^{18} in the clock mode.

The DDS generates spurious sidebands, however, whenever the number N_2 is other than a simple power of 2. For all other numbers the circuit behaves like a fractional divider with a sequence of non-uniform periods, which repeat with a definable pattern to give an average period equal to the desired period. The level of the spurious sidebands that are created by this periodic phase modulation can be calculated by:

$$A_S = 20 \log (Df / f_{in}) \quad dBc$$

and at the maximum correction frequency of 1 KHz, the sideband level will be less than –50 dBc at the output of the DDS. At the minimum correction frequency of about 200 Hz it falls to –64 dBc.

Additional filtering in the phase detector and in the following active lead-lag filter reduces the spurious levels to those levels predicted below for three sideband frequencies:

FM	Midrange	Extremes
10 Hz	–90 dBc	–76 dBc
100 Hz	–96 dBc	–82 dBc
1 KHz	–116 dBc	–102 dBc

Frequency Accuracy and Temperature Tracking Performance

While the FXO is less stable than the OXO, it must provide high stability at the temperature extremes to accurately predict the temperature. It is most critical at –55 degrees Celcius, where the slope of the stable mode is typically about 20 Hz per degree. This indicates that the temperature signal must be repeatable to 0.01 degree to maintain the required accuracy of $2*10^{-5}$. This in turn leads to a stability requirement for the FXO of $1*10^{-8}$ at –55 degrees Celcius (assuming the OXO is absolutely stable).

Since both oscillators are involved in the determination of temperature, there will be no temperature hysteresis error if they deviate at the same rate. Temporal drift, however, will produce an aging error. If a differential drift occurs, then the temperature deviation error curve will rotate about the turning point of the resonator. The problem also exists at +85 degrees Celcius but is less severe by a factor of about two.

Digital Circuit Implementation

The MCXO system under development will have all of its digital circuitry contained in three chips: an Intel 87C51 microcomputer; an ICT 93C46 serial EEPROM used as non-volatile RAM; and a custom ASIC (application-specific integrated circuit) engineered to hold all the remaining digital functions.

The Intel 87C51 microcomputer was chosen because it offers the following benefits:

- Small 44 pin surface mount device has sufficient EPROM and RAM on board to implement the MCXO software.
- Instruction set includes a multiply instruction, which is heavily used in solving polynomials.
- CMOS technology and "sleep mode" decrease power needs.
- Military temperature and package version is available.

The software which runs on the microcomputer is written in the "C" language supplemented with function calls written in the 87C51's native assembly language.

In addition to the EPROM and RAM on the 87C51, the MCXO also takes advantage of the on-board interrupt controller, counter/timer unit and UART serial port. The ICT 93C46 was used as a convenient way to store the polynomial coefficients. It allows the coefficients to be electronically stored and altered after the unit is assembled and yet not invalidate them if power is lost. The 93C46 is an eight pin surface mount CMOS device which consumes very little power. Radiation hardness may pose a problem for this type of part in military applications. In such cases a "one time" PROM or battery backed-up RAM would have to be substituted.

The ASIC will contain all the digital control logic, the divider chains, mixers, period measurement counters, and the DDS. It will be built from a gate array utilizing 1.5 micron CMOS technology. Power consumption is minimized by the small geometry of the process technology and by the fact that having all the logic in one chip eliminates printed wiring board capacitances which must be overcome when signals change state.

The ASIC will be in a 44 pin surface mount package similar to the one for the microcomputer. This one package replaces the equivalent of an entire circuit board full of discrete logic chips. The demonstration version will use a commercial version of the gate array, however, military versions up to and including full radiation hardened devices can be fabricated using the same design inputs.

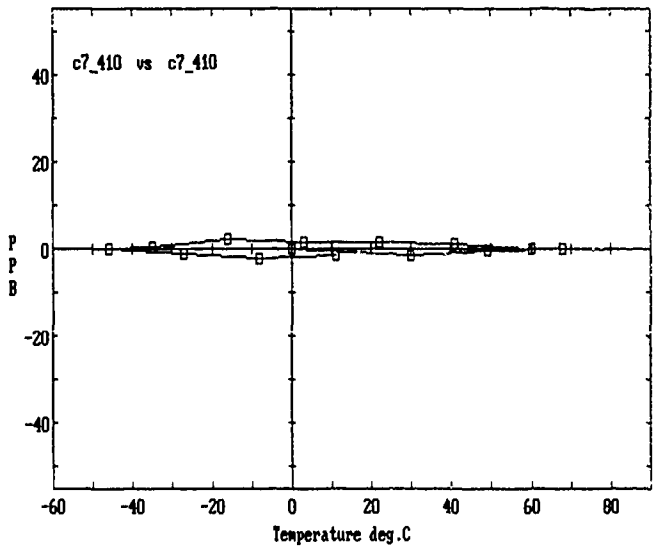


Figure 11. Data point deviation versus polynomial

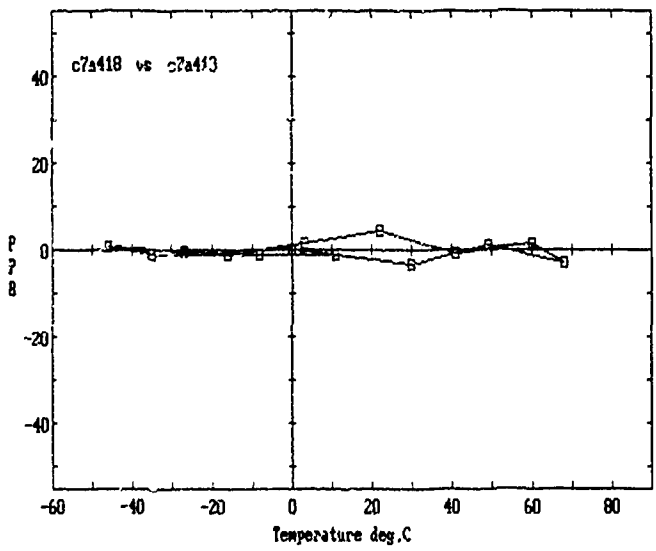


Figure 12. Data point deviation versus polynomial

Power Projections

The MCXO uses an assortment of techniques including low-power oscillators, CMOS circuits, an ASIC, and power sequencing to reduce the overall power consumption. Using measurements where available and projections in situations like the gate array ASIC, it is estimated that the MCXO will operate on approximately 62 milliwatts in Frequency Mode and 21 milliwatts in Clock Mode. These estimates assume that the temperature is stable or changing slowly.

The MCXO Interim Prototype System

The MCXO test system (see Figure 13), currently operating, does not yet have a custom gate array ASIC. In its place, a Logic Cell Array (LCA) has been used. This device mimics a gate array in that it allows large amounts of custom logic to be implemented in a single chip. Additionally, the user can program and re-program it without spending time or money on outside vendors. This allows logic circuits to be changed and tested without even the cost of altering the printed wiring board. This encourages circuit improvements and allows extensive testing before committing the design to a gate array. While the LCA is a dream come true for testing, it is both larger physically and much more power-hungry than a true gate array of similar complexity.

The interim prototype also differs from the final unit in that it is built with through-hole rather than surface mount components and in that there is more circuitry on the board. The additional circuits to provide signal level conversion to RS-232 and to aid in debugging in general have been added to the interim prototype.

Summary of Performance Achieved

A group of crystals are currently under test using two of the interim prototype systems. The aim is to determine the long-term effects of repeated temperature cycling on overall performance.

The systems are cycled daily over the temperature range of -55 to +85 degrees Celcius. Resonators are being replaced after about two weeks of cycling. Daily comparisons are made between the new data and the original polynomial fit. Typical results are shown in Figures 11 and 12. The last three digits in the name refer to the date of calibration.

The width of the hysteresis loop, including a degree of random variation, is usually less than 10 ppb.

Obviously, much more data must be accumulated on different systems and different resonators to achieve reliable performance specifications. It is our aim to do this.

The Future of the MCXO Project

Now that the interim prototype has been tested, the development of a gate array has begun. Upon completion it will be integrated into a final surface mount version of the MCXO. The interim prototype system will be used with the calibration software to continue crystal calibration and test efforts. The crystals on hand, manufactured by Piezo Crystal Co. and Colorado Crystal Co., need to be temperature-cycled long enough to determine their aging and hysteresis characteristics. Since the resolution of the digital system is better than the overall specification by a factor of 10, the stability of the crystals is expected to limit the system's accuracy.

Research must also be continued into the design of simpler and more stable dual-mode oscillator circuits. Smaller resonator packages and resonators with reduced hysteresis should also be pursued. The impact of resonator frequency on size, power consumption, and aging will need to be considered as different applications for MCXOs develop.

Acknowledgement

The authors wish to acknowledge the support of the U.S. Army's Electronic Technology and Devices Laboratory under Contract Number DAAL01-88-C-0804.

References

1. S. Schodowski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," in Proceedings of the 43rd Frequency Control Symposium, 1989.

2. R. Filler and J. Vig, "Resonators for the Microcomputer-Compensated Crystal Oscillator," in Proceedings of the 43rd Frequency Control Symposium, 1989.

3. M. M. Driscoll, "Low Noise Microwave Signal Generation Using Bulk and Surface Acoustic Wave Resonators," in Proceedings of the 42nd Frequency Control Symposium, 1988, pp. 369-379.

4. A. Benjaminson, "A Crystal Oscillator With Bidirectional Frequency Control and Feedback ALC," in Proceedings of the 40th Frequency Control Symposium, 1986, pp. 344-349.

5. A. L. Bramble, "Direct Digital Frequency Synthesis," in Proceedings of the 35th Frequency Control Symposium, 1981, pp. 406-414.

6. The CODA software program was developed under U.S. Army Contract Number DAAL01-88-C-0804. For additional information, contact the Frequency Control & Timing Branch of the U.S. Army Laboratory Command, Ft. Monmouth, NJ 07703-5000, or General Technical Services, 105 Clermont Avenue, Alexandria, VA 22304.

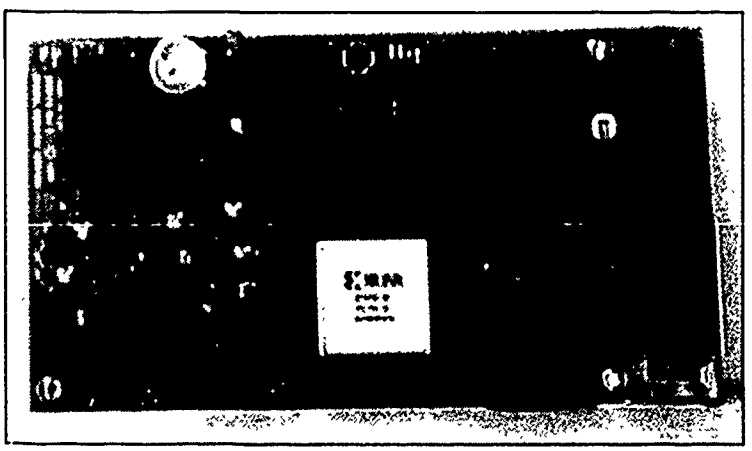


Figure 13. The MCXO Interim Prototype System

43rd Annual Symposium on Frequency Control - 1989

FREQUENCY-TEMPERATURE and AGING PERFORMANCE of MICROCOMPUTER COMPENSATED CRYSTAL OSCILLATORS

R. L. Filler, J. A. Messina, and V. J. Rosati
US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

SUMMARY

Evaluation of five (5) samples of the Microcomputer-Compensated Crystal Oscillator built during a development program has verified the feasibility of achieving frequency vs. temperature stability of $\pm 5 \times 10^{-8}$, aging of 1×10^{-10}/day, and hysteresis of no worse than $\pm 3 \times 10^{-8}$, with an input power of 45 milliwatts.

In addition, a simulated compensation was performed on several samples of each of two dual-mode oscillator designs. One design uses a lateral-field resonator and the other uses a conventional, thickness-field resonator. The frequency could be compensated to better than $\pm 2 \times 10^{-8}$ for a -55°C to +85°C temperature range for both designs.

INTRODUCTION

The previous four papers in these proceedings describe the concept of resonator self-temperature sensing,[1] and the practical implementation of both the dual-mode oscillator and the Microcomputer-Compensated Crystal Oscillator (MCXO).[2-4] In this paper we describe the frequency vs. temperature (f-T) and the frequency vs. time (aging) performance of one version of the MCXO, as well as the f-T performance of two versions of the dual-mode oscillator. Aging and f-T are probably the two most important performance parameters for a clock driver, which is the intended application of the present MCXO.

CONFIGURATIONS TESTED

The three configurations tested were:

1. A packaged MCXO manufactured by Frequency Electronics, Inc, consisting of a dual-mode oscillator, a microcomputer, and a gate array. A complete description can be found elsewhere in these proceedings.[3] We shall refer to this device as the MCXO. We will report results for five samples.

2. A dual-mode oscillator design using a lateral field resonator and hybrid circuit construction. The lateral field resonator was designed to eliminate the requirement for the b-mode suppression circuit in the oscillator.[5] We will refer to this device as the LFR dual-mode oscillator. The LFR dual-mode oscillator is a component in the MCXO described above.

3. A dual-mode oscillator design using a thickness-field resonator.[4] We will refer to this oscillator as the TFR dual-mode oscillator. We tested two samples of the TFR dual-mode oscillator. Each sample was tested with two resonators, one each from two manufacturers.

f-T PERFORMANCE

Temperature Profile

All of the f-T data in the paper were obtained using a quasi-static "step and soak" temperature-time profile. The test was always initiated with storage at the low temperature extreme. After the first measurement, the temperature was increased by the preselected step. After the selected period of time (the "soak" time), the next measurement was made. This procedure was repeated until the high temperature extreme at which time the direction of the step was reversed and the process was continued until the low temperature extreme was reached. Figure 1 is a graphic representation of the f-T profile.

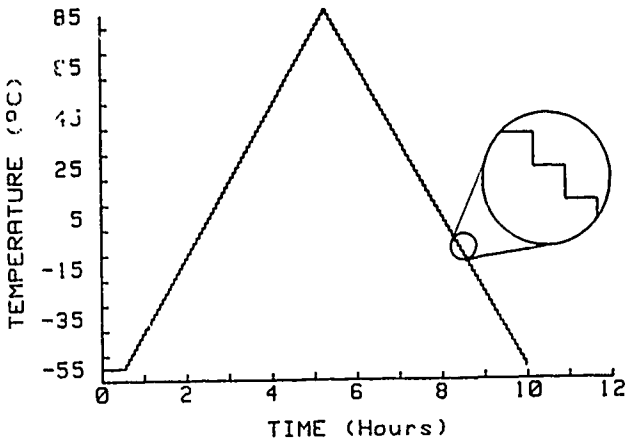


Figure 1 - Temperature vs. time for f-T measurement.

MCXO f-T

Figure 2 shows the experimental setup for measurement of the f-T performance of the MCXO. One to five devices were measured during each temperature run.

Figures 3-7 show typical results for the five MCXO's. Each device was tested between 11 and 17 times with various step and soak parameters. These curves, and all that follow, depict f-T stability.[6] The center frequency can be adjusted by an auto calibration circuit without any trim effect.[6]

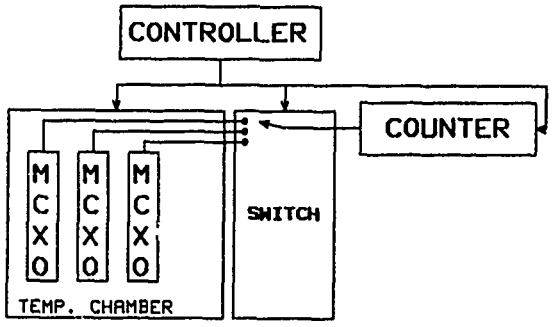


Figure 2 - Experimental setup for measuring the f-T of the MCXO.

Figure 5 - f-T characteristic of MCXO 16725.

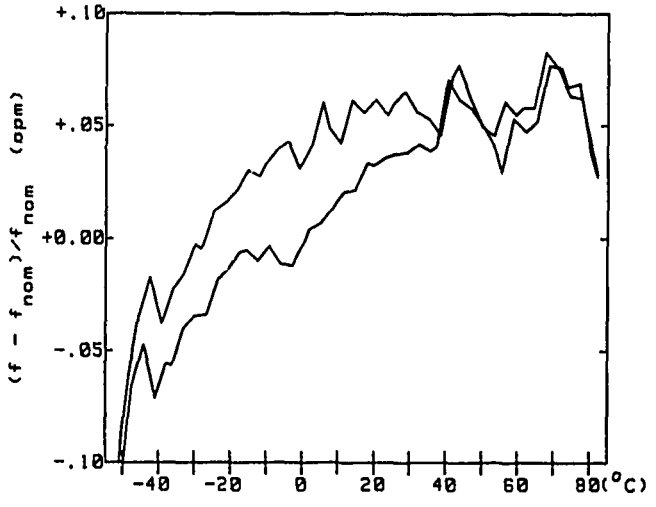


Figure 3 - f-T characteristic of MCXO 16723.

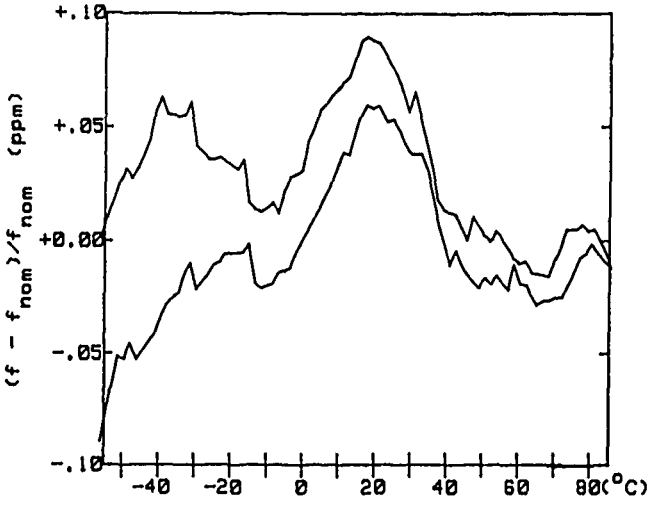


Figure 6 - f-T characteristic of MCXO 18236.

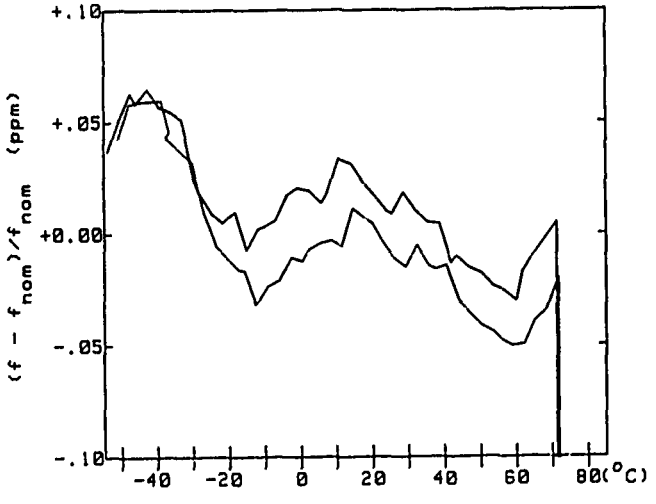


Figure 4 - f-T characteristic of MCXO 16724.

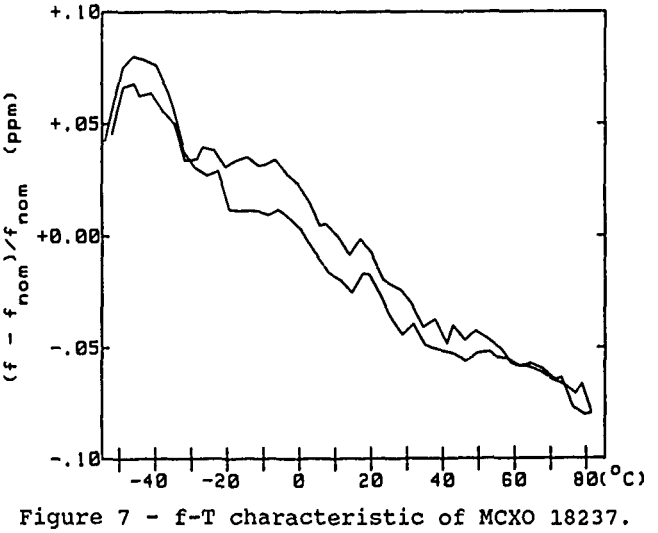


Figure 7 - f-T characteristic of MCXO 18237.

Dual-Mode Oscillator f-T

It can be seen, especially from Figure 7, that the compensation of most of the samples can be improved with a simple "rotation," that is, a change in the first order temperature coefficient in the compensation algorithm. The necessary frequencies are available from the MCXO. Therefore, it was decided to attempt to simulate compensation of the devices externally on a desktop computer. Observation of the output of the dual-mode oscillator enabled us to determine that the high temperature anomaly in SN 16724 was in the digital circuitry, not the analog circuitry.

Figure 8 shows the experimental setup for the collection of data for external compensation of the dual-mode oscillator. A true MCXO has one counter with one of the dual-mode frequencies as the time-base and the other as the counter input. We used two counters in this measurement, with a cesium standard as the time-base. We converted the data to be equivalent to the true MCXO by software (the variation of the time-base frequency is small enough so that the results are identical with or without the conversion).

It was important that both counters be triggered simultaneously. This eliminated any error due to temperature fluctuations which change the frequency of one mode while the other mode is being measured.

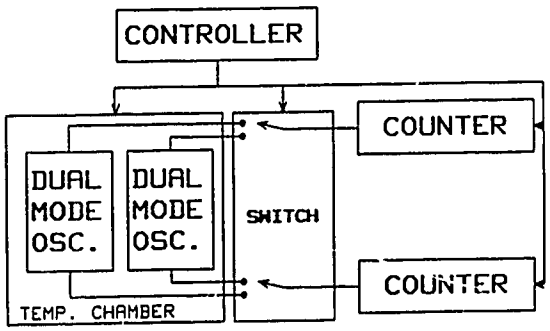


Figure 8 - Experimental setup for data collection for external compensation of dual-mode oscillators.

Figure 9 is a typical f-T curve for the fundamental mode (f_1) of a dual-mode oscillator. Figure 10 is a plot of f_β vs. temperature, where f_β is the difference frequency between three times the fundamental mode and the third overtone.[1] There is an obvious temperature offset between the "up" and "down" segments of the cycle, i.e., "apparent hysteresis."[7]

Figure 11 is a plot of f_1 vs. f_β. Note that both the "up" and "down" segments are superimposed, i.e., the "apparent hysteresis" has vanished.

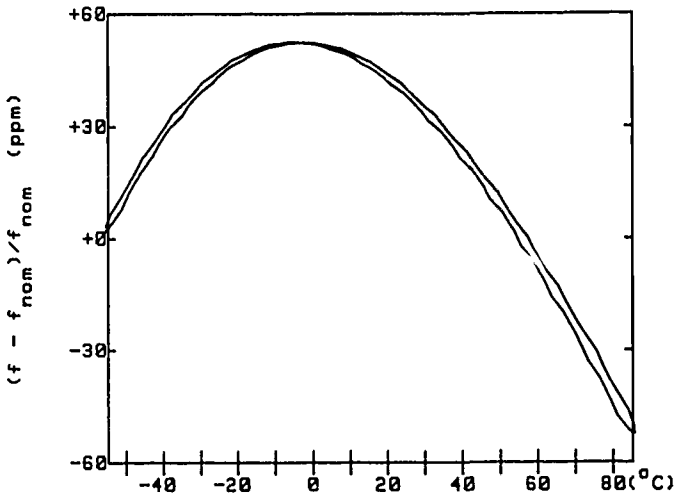


Figure 9 - f-T of the fundamental mode of a dual-mode oscillator.

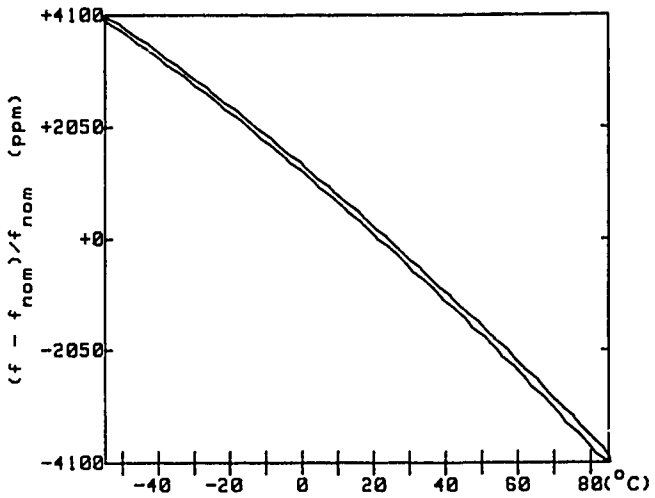


Figure 10- f_β vs. temperature.

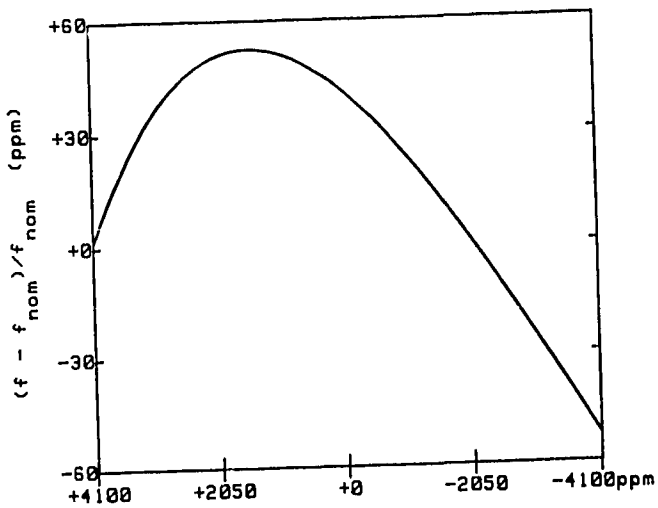


Figure 11 - f_1 vs. f_β.

To perform the "simulated" compensation we made a least square fit of the set of data (f_β, f_1) to a 6th order polynomial.

$$F(f_{\beta i}) = \sum_{p=0}^{6} C_p \, (f_{\beta i})^p \qquad (1)$$

We then computed the residuals, R_i,

$$R_i = f_{1i} - F(f_{\beta i}) \qquad (2)$$

A plot of the residuals vs. temperature is, assuming a perfect technique for applying the calculated correction, e.g., pulse deletion or numerical synthesis, equivalent to an MCXO f-T stability plot.

LFR Dual-Mode Oscillator f-T

Figures 11-15 are plots of the residuals from a 6th order polynomial for the 5 LFR dual-mode oscillators. These can be directly compared to Figures 3-7, respectively. The hysteresis is identical in each corresponding pair of curves.

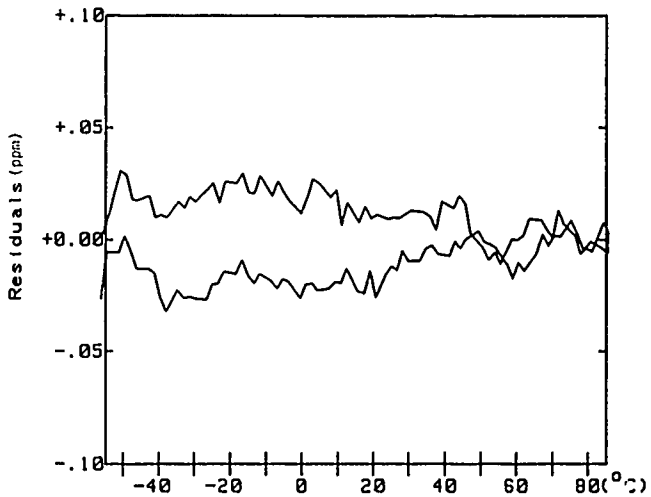


Figure 12 -Residuals vs. temperature for the dual-mode oscillator in MCXO 16723.

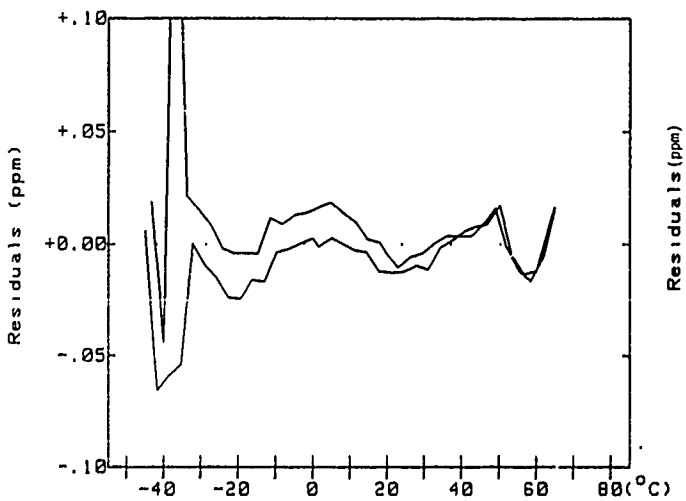


Figure 13 - Residuals vs. temperature for the dual-mode oscillator in MCXO 16724.

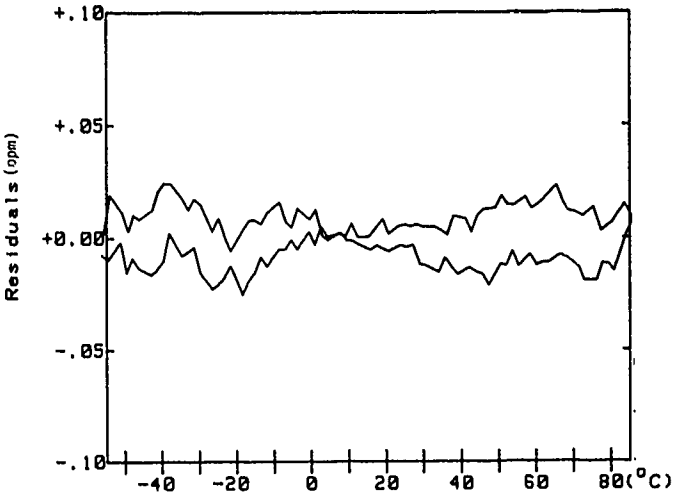


Figure 14 - Residuals vs. temperature for the dual-mode oscillator in MCXO 16725.

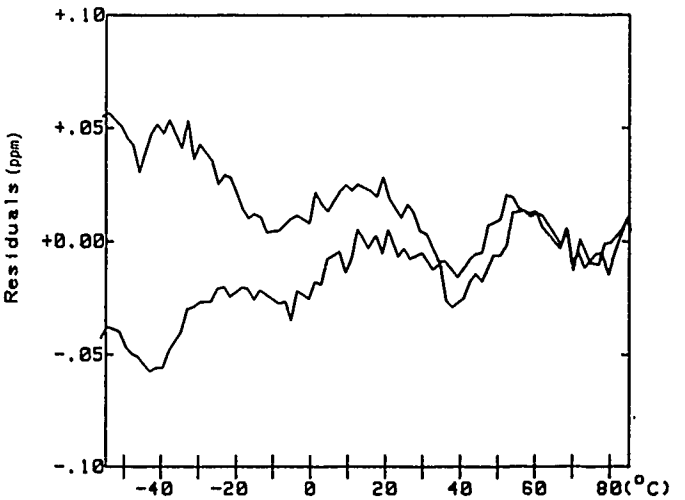


Figure 15 - Residuals vs. temperature for the dual-mode oscillator in MCXO 18236.

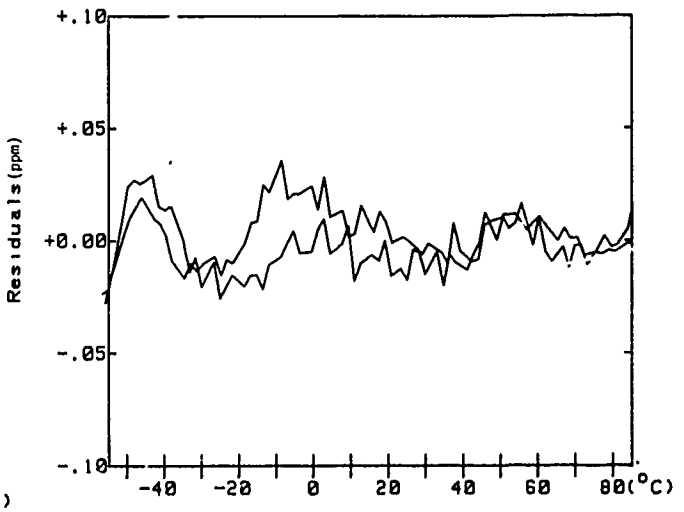


Figure 16- Residuals vs. temperature for the dual-mode oscillator in MCXO 18237.

The algorithm does not take into account thermal history, therefore, it cannot eliminate hysteresis.

Even including the hysteresis, the f-T stability is $\pm 2 \times 10^{-8}$. The probable cause for the difference between the MCXO results and the "simulated" compensation is that the MCXO was compensated midway in the manufacturing process. Subsequent processing steps caused a change in the required compensation coefficients, possibly by changing circuit strays. In the future, improved manufacturing techniques should allow realization of the "simulated" compensation.

TFR Dual-Mode Oscillator f-T

Figures 17-18 are plots of the residuals of the two TFR dual-mode oscillators. The large difference in behavior between the two samples prompted us to interchange the resonators between the two circuit boards in order to determine if the resonator or some circuit component was the cause of the lesser performance of SN A36. When the resonators were interchanged, the results followed the resonators. The excellent results shown in Figure 17 indicate that it is possible to reduce hysteresis to an extremely low value.

The function used to calculate the residuals plotted in Figure 17 was a three-segment polynomial with free boundary conditions on each segment. The three segments were used to better accommodate the anomaly at -30°C.

AGING

All five MCXO's were placed into a temperature chamber and put through the sequence depicted in Figure 19. The duration of the aging segments was approximately 30 days. Figures 20 - 25 show the aging data. The dashed line is a linear approximation of the last several days of data. The value of the slope of the dashed line is given in each figure. (The μP in 16723 was reset at day 20)

f-T Repeatability

Figure 25 is the superposition of 13 f-T runs on MCXO 16723 taken over the course of six months. Two different temperature ranges (-55°C to +85°C and -46°C to +68°C), two step sizes (2K and 3K) and three soak times (4, 12, and 30 minutes) were used. The performance of the MCXO appears to be independent of temperature range, soak time, and step size. Most importantly, there is no evidence of any f-T aging (which would occur e.g., if the aging of the two modes differed significantly).

PERFORMANCE SUMMARY

A performance summary is given in Figure 26.

CONCLUSIONS

A new era in temperature compensation is a reality. Whereas analog TCXO accuracy has remained at about 1 ppm for the past 30 years, the MCXO now allows a temperature stability of 0.02 ppm and an overall accuracy of $<5 \times 10^{-8}$. Future effort should be expended on improving the modeling of the f-T function and further reducing thermal hysteresis in the resonator.

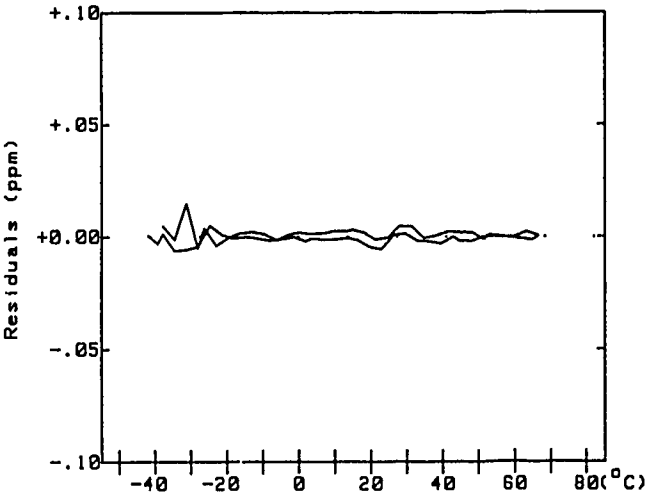


Figure 17 - Residuals vs. temperature for the TFR dual-mode oscillator SN AP7.

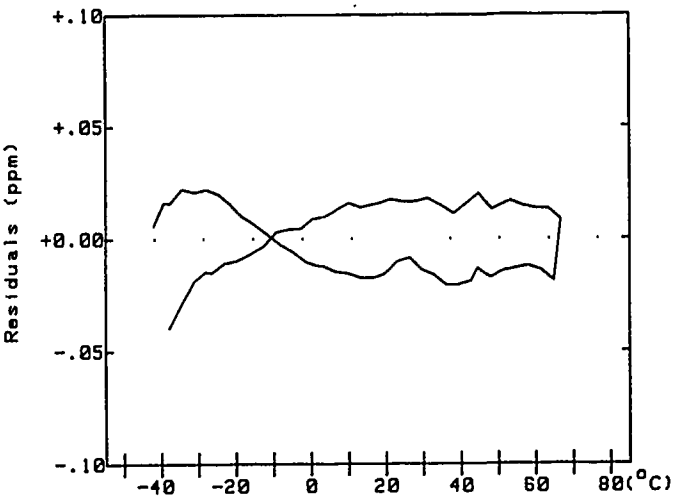


Figure 18 - Residuals vs. temperature for the TFR dual-mode oscillator SN A36.

Figure 19 - Aging sequence.

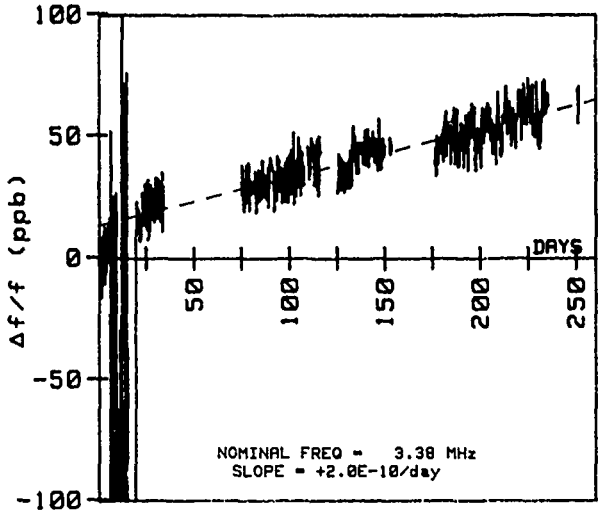


Figure 20 - Aging data for MCXO _6723.

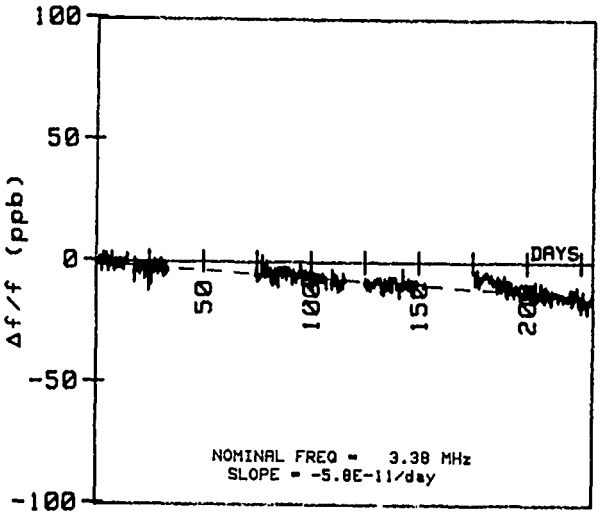


Figure 21 - Aging data for MCXO 16724.

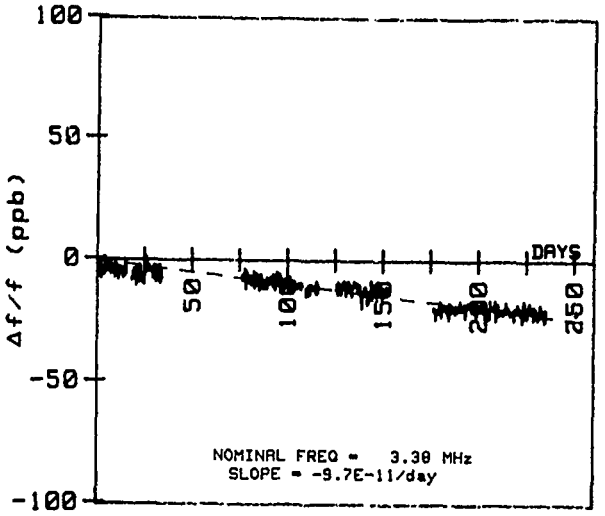


Figure 22 - Aging data for MCXO 16725.

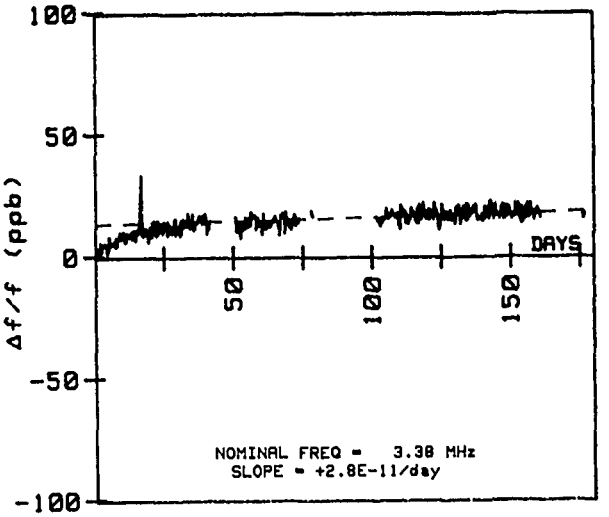


Figure 23 - Aging data for MCXO 18236.

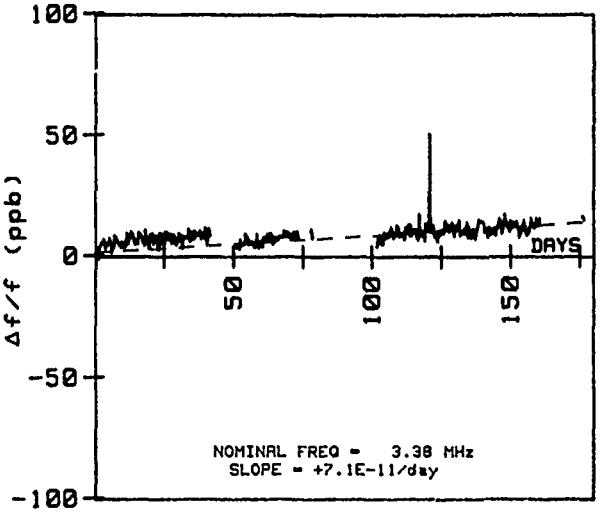


Figure 24 - Aging data for MCXO 18237.

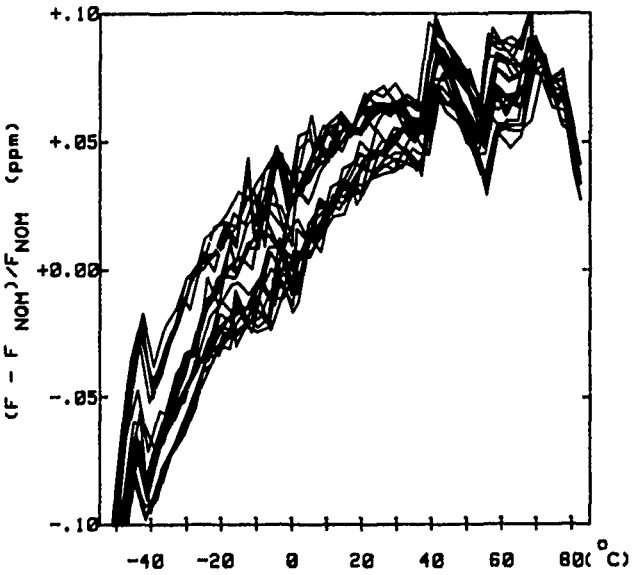


Figure 25 - Superposition of 13 f-T runs with various parameters for MCXO 16723.

MCXO PERFORMANCE SUMMARY

Manu.	Res.	# of f-T	f-T (int) 10^{-8}	f-T (ext) 10^{-8}	Hyst 10^{-8}	Aging 10^{-10}/d
FEI	LFR	17	12.5	3.2	2	2.0
FEI		16	5.5	-	1	0.6
FEI		13	5.0	2.5	1	1.0
FEI		12	11.0	5.7	4.5	0.3
FEI		12	8.5	3.2	1.5	0.7
GTS	TFR	11	-	1.1	0.5	-
GTS		11	-	3.1	2	-

Figure 26 - Performance Summary.

REFERENCES

1. S. S. Schodowski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator" elsewhere in these Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

2. R. L. Filler and J. R. Vig, "Resonators for the Microcomputer-Compensated Crystal Oscillator" elsewhere in these Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

3. M. Bloch, M. Meirs and J. Ho, "The Microcomputer Compensated Crystal Oscillator" elsewhere in these Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

4. A. Benjaminson and S. Stallings, "A Microprocessor-Compensated Crystal Oscillator Using an SC-cut Third Overtone Resonator" elsewhere in these Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

5. A. W. Warner, Jr. and B. Goldfrank, "Lateral Field Resonators" in the Proceedings of the 39th Annual Symposium on Frequency Control, 1985, pp 473-474.

6. R. L. Filler, V. J. Rosati, S. S. Schodowski and J. R. Vig, "Specification and Measurement of the Frequency vs. Temperature Characteristics of Crystal Oscillators" elsewhere in these Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

7. R. L. Filler, "Measurement and Analysis of Thermal Hysteresis in Resonators and TCXO's" in the Proceedings of the 42nd Annual Symposium on Frequency Control, 1988, pp 380-387.

8. W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, Numerical Recipes: The Art of Scientific Computing. Cambridge, Cambridge University Press, 1986, pp 498-546.

43rd Annual Symposium on Frequency Control - 1989

LOW POWER TIMEKEEPING

MARTIN BLOCH, MARVIN MEIRS, AND JOHN HO
Frequency Electronics, Inc.
55 Charles Lindbergh Blvd.
Mitchel Field, NY 11553

JOHN R. VIG AND STANLEY S. SCHODOWSKI
U.S. Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

ABSTRACT

A low power timing system has been developed that can achieve the following performance:

Power Consumption	From 5 to 41 mW
Timing Accuracy	Within 5 to 70 milliseconds in 24 hours
Size	Less than 3 cubic inches
Operating Temperature Range	-55 to +85°C.

The system achieves this performance by using a Microcomputer Compensated Crystal Oscillator (MCXO) to periodically update a low power oscillator clock system. This timing system was developed by FEI in conjunction with the technical staff at LABCOM. A breadboard model of the system has been built and demonstrates the capability.

Clocks are presently available that operate at very low power (less than a few milliwatts) and maintain good accuracies (less than 100 milliseconds per day) and at relatively low cost. The disadvantage of these clocks is that they can maintain their timing accuracy only over a very narrow temperature range. Over any substantial temperature variation, especially those that can be expected in a military environment, the frequency variation becomes quite large and the timing error increases considerably. Improvements can be made to the frequency accuracies by using a higher precision oscillator with better temperature coefficient such as TCXO or OCXO, but these have the disadvantages of requiring considerably more power, size and weight than the original simple clock.

These disadvantages are overcome by combining a simple clock oscillator with the MCXO to periodically correct both frequency and time within the digital clock. This technique does not substantially increase the power dissipation because the MCXO is turned on only for a few seconds each time to recalibrate the clock oscillator frequency. Using this technique, a clock oscillator of $+/-50 \times 10^{-6}$ over the temperature range can be made to approach the $+/-5 \times 10^{-6}$ accuracy of the MCXO.

INTRODUCTION

The quartz wristwatches that are common these days represent accurate very-low-power timekeeping. The temperature range over which watches would remain accurate however is very limited.

Several designs could be considered for extending the temperature range while approaching the power consumption of an ordinary watch. For comparison, the desired accuracy could be achieved over the temperature range by placing critical circuit elements in an oven-controlled environment isolating those elements from temperature variations. Of course, the oven could not be considered in many applications because it requires a large power source. See Figure 1.

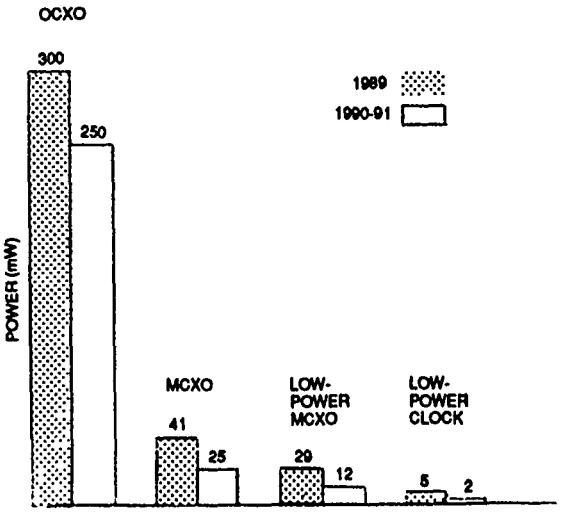


Figure 1. Power Consumption in Different Approaches to a Low-Power Clock

The next possibility is the use of an MCXO to control the clock. It would provide the accuracy over the temperature range without the need for an oven. As pointed out in a previous paper in these Proceedings (ref. 1), the development of the MCXO has reduced its power to 41 mW. This is a few orders of magnitude greater than the watch, but has the advantage of much improved accuracy across the much wider temperature range.

The third approach indicated in Figure 1, low-power MCXO, extends the time between frequency measurements and has the MCXO operating in idle (standby) between calibration times. With extended intervals between calibrations, the minimum power can be reduced to the idle power of the MCXO.

The fourth approach, the low-power clock, keeps the MCXO turned off except when it is performing calibration. If the duty cycle is sufficiently low, the total power consumed will be almost independent of MCXO power. In this approach as in the previous one, temperature variations between calibrations will affect the frequency, reducing the accuracy of the clock. Thus, there will be a trade-off between power consumption and accuracy.

OPERATION OF THE LOW-POWER MCXO

The low-power MCXO approach is illustrated in the block diagram (Figure 2). In this application, the MCXO pulse-train frequency output is divided to provide a 1 pps signal.

To keep the power consumption in the low-power MCXO as low as possible, the microprocessor and the read-only memory (which are the circuits consuming most of the power) are operated in standby (idle) except during the time required to perform a measurement. While these circuits are in standby, the correction factor is held constant. This permits the designer some flexibility in trading off the duty cycle for these circuits against the clock accuracy.

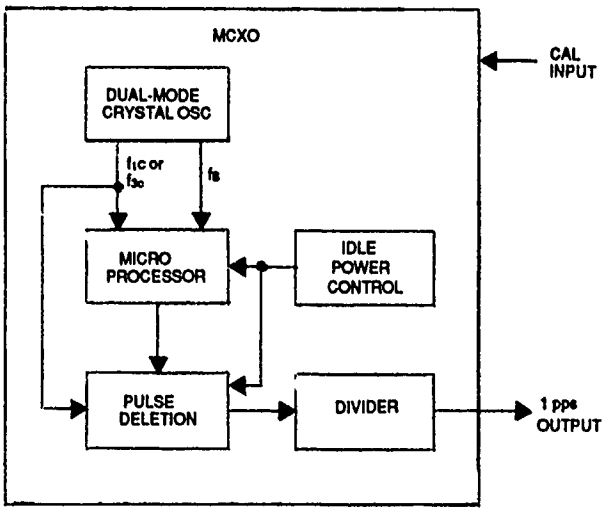


Figure 2. Low-Power MCXO Block Diagram

The power consumed by each circuit element in idle and when operating are compared in Table 1. The total power during the idle periods is less than one-fourth of the total when all circuits are operating.

The effect of changing the duty cycle of the idled circuits can also be seen in Table 1, which compares the power consumed with a 20% duty cycle (normal MCXO operation) with that consumed with a 5% duty cycle. The effect on power consumption of extending the interval between up-dates is shown in Figure 3. At long update intervals, the power approaches the power consumed in the idle mode.

OPERATION OF THE LOW-POWER CLOCK

The low-power clock approach is illustrated in Figure 4. The low-power clock advances one step beyond the low-power MCXO. Instead of keeping the MCXO in idle in the time between frequency measurement and updates, the entire MCXO and additional calibration circuits of the low-power clock are deenergized between updates, and the low-power clock oscillator is used as the flywheel.

TABLE 1. STANDBY AND OPERATING POWER CALCULATIONS FOR LOW-POWER MCXO

Circuit	Average Power (mW)			
	Idle	Peak	20% Duty Cycle	5% Duty Cycle
Dual-mode crystal oscillator	6.5	6.5	6.5	6.5
ASIC	15.0	15.0	15.0	15.0
Microprocessor	2.5	50.0	12.0	4.85
Read only memory	1.0	35.0	7.8	2.7
Total Power	25.0	106.5	41.30	29.05

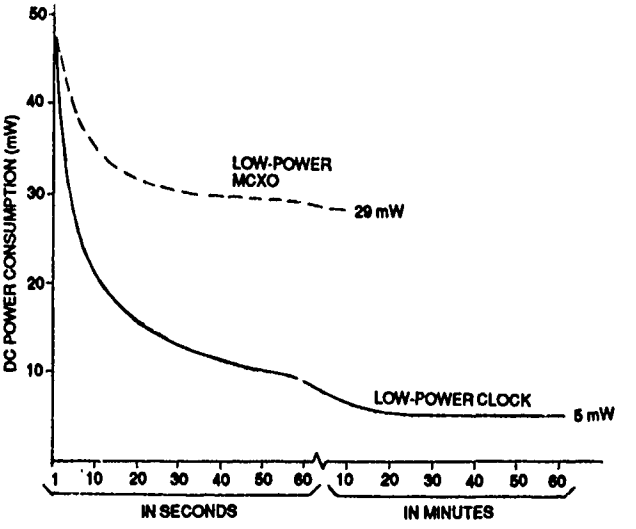


Figure 3. DC Power Consumption vs. Up-date Interval

The 1 pps output of the MCXO determines the time interval for the counter that counts the output of the low-power uncompensated crystal oscillator. The digital count from the counter is compared with the desired count, stored in PROM. The comparator provides an error signal whose amplitude is proportional to the difference

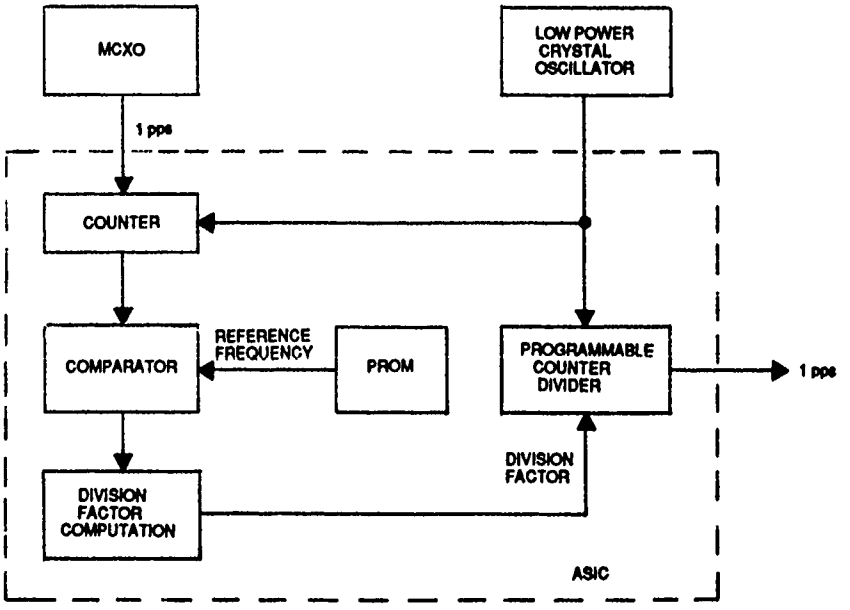


Figure 4. Low-Power Clock Block Diagram

between the actual oscillator frequency and the desired one and whose sign depends on whether the actual frequency is above or below the desired one.

The error signal is used to determine the division that is required to convert the actual uncompensated crystal oscillator frequency to a 1 pps output. This calculated division factor is applied to the programmable counter divider, where the frequency division is accomplished. The duty cycle for the calibration circuits (MCXO, etc.) determines the average power consumption of the low-power clock.

The lower curve in Figure 3 is a plot of the average power in the low-power clock as a function of the time between up-dates. Here, the limiting value is the power consumed in the low-power crystal oscillator and divider circuit; the MCXO and other circuits in the ASIC and microcomputer, when off for prolonged periods, do not contribute significant amounts to the limiting value of average power consumption.

ACCURACY

The frequency change that can occur between periodic updates in the low-power MCXO and in the low-power clock is a function of the duty cycle. Thus the timing error can be plotted directly against power consumption (Figure 5). As seen in Figure 3, the low-power MCXO approaches a lower limit of power of 29 mW; in Figure 5, this corresponds to a timing error of approximately 12 milliseconds in a 24-hour period. The low-power clock can reduce the power consumption to 5 mW, but accuracy continues to decrease as shown in Figure 5, reaching 70 milliseconds in 24 hours.

PACKAGING DESIGN

The production low-power clock will be packaged by adding a low-power (32 kHz) clock oscillator to the existing low-power MCXO package and using a modified ASIC. Thus the components of the low-power clock will be housed in a compact package (Figure 6) that is separated into three pc boards.

The flatpack crystal on the upper board in Figure 6 will be connected to the dual-mode oscillator (hybrid

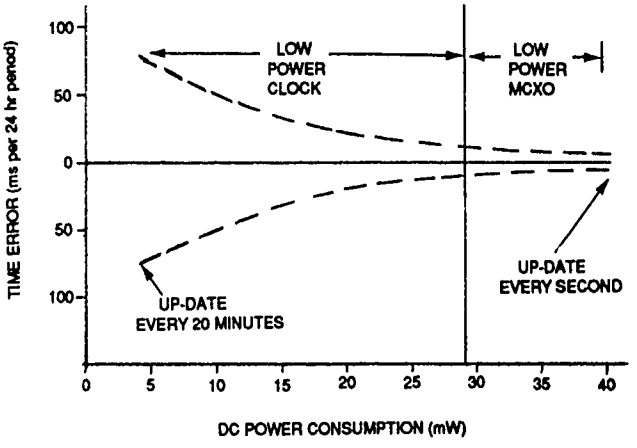


Figure 5. Timing Error vs. Power Consumption

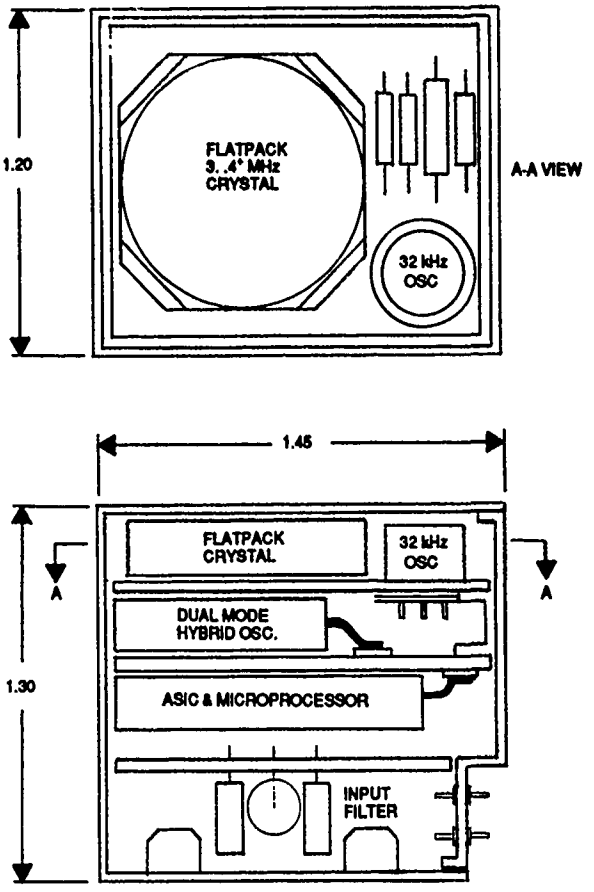


Figure 6. Low-Power Clock Internal Layout

circuit); the 32 kHz oscillator is the low power clock oscillator.

The dual-mode oscillator and the ASIC (application-specific integrated circuit) and microcomputer hybrid will be mounted on the second pc board.

The power conditioning filter and input/output connections will be located on the last board.

SUMMARY

Low-power timekeeping methods have been developed that provide improved accuracy over the full military temperature range. These methods provide trade-offs of accuracy vs. power consumption dependent upon the application.

ACKNOWLEDGMENT

Partial support was provided by the U.S. Army Electronics Technology and Devices Laboratory under Contract DAAL01-85-C-0403.

REFERENCES

[1] M. Block, M. Meirs, and J. Ho "The Microcomputer Compensated Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

43rd Annual Symposium on Frequency Control - 1989

TRANSIMPEDANCE CRYSTAL OSCILLATOR

L. S. Ferriss
Kearfott Guidance & Navigation Corporation
1225 McBride Avenue, Little Falls, NJ 07424

Abstract

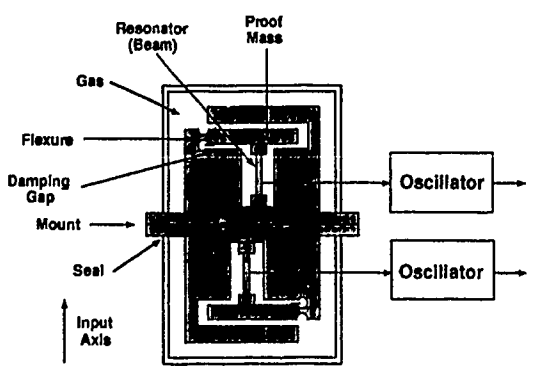
Adapting quartz beams to force sensors with coupling structures unavoidably degrades Q requiring careful electronics design to preserve transducer performance. In this application, an inertial quality accelerometer, restructuring the bridge-stabilized oscillator [1] reduces oscillation frequency sensitivity to circuit element variations and changes in the motional resistance, R_m, of the resonator. The transimpedance oscillator [2] is developed in steps beginning with the bridge oscillator, through transposition of bridge elements, frequency compensation of the resonator arm, and then mechanization of the transimpedance configuration. The analytical basis for performance improvement is shown along with the results of simulations of modeled oscillator circuits and actual circuit tests. In conclusion a means to compensate for the effect of crystal shunt or "clamp" capacitance variation is suggested.

Introduction

The Vibrating Beam Accelerometer [3], shown schematically in Figure 1, exercises the force-frequency properties of a pair of quartz beam resonators mechanically attached to a pair of proof masses producing reaction forces due to acceleration of the instrument. The resonators, coerced in opposite sense, are connected with oscillator electronics producing output frequencies related to the stress along the major axes of the vibrating beams. For the sense of the input acceleration shown, the lower resonator, in tension, produces an oscillator output frequency increase due to alteration of the motional capacitance, C_m, in the series branch of the electrical equivalent circuit of the crystal [4]. The upper resonator, in compression, produces a decrease in frequency. The operational output frequency of the accelerometer is

$$f_d = f_b (1 + SA) \qquad (1)$$

where f_b is the so-called bias frequency, S is the scale factor, and A is the input acceleration [5]. In operation, A is estimated from models of the systematic effects of environmental influences on f_b and S.



Pictorial of the Vibrating Beam Accelerometer
Figure 1

The role of the oscillator electronics in this process is to vibrate the resonators at frequencies that would be mapped by force changes without introducing frequency changes due to the circuit components or variations. Since C_m is the dominant force-varied element of the resonator, it makes sense to operate the crystal right on series resonance, f_o, defined by L_m and C_m only, rather than some other frequency such as the antiresonance f_a, influenced also by C_o and R_m. Therefore the choice of a series mode oscillator is most appropriate in this case. However C_o and other circuit capacitances do affect the frequency of oscillation, as does R_m in many configurations. Frequency stability is further exacerbated by lower Q and lower figure of merit, M [6], of the resonator brought on not only by the attachment of force couplers but also from operation of the beams in the presence of a viscous gas introduced into the interior of the accelerometer cavity. This is purposely done to dampen resonances associated with the proof mass, its flexure hinge, and coupling structures.

Resonator Characterization

The change in properties of the resonator due to fettering and gas damping can be appreciated from a comparison of two resonators, shown in Table 1 [7]. The crystals are x-cut tuning forks identical in construction except that the "V" resonator is housed in an evacuated, hermetically sealed ceramic package whereas the "H" resonator is sealed at 1 atmosphere.

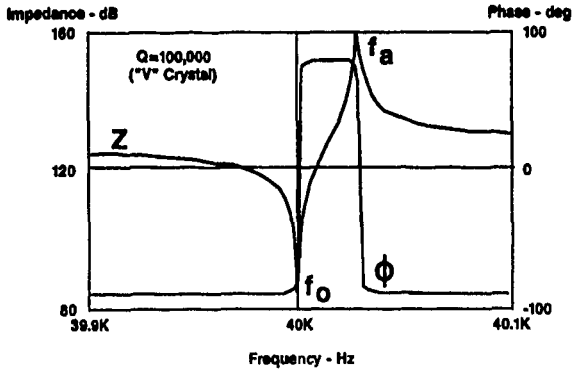
TABLE 1. COMPARISON OF PROPERTIES OF LOW-FREQUENCY QUARTZ CRYSTALS

V - Hermetic Seal, Evacuated H - Sealed at 1 Atmosphere			
Property	Value		Units
	V	H	
Frequency	40	40	kHz
R_m	20	100	Kohm
C_m	2.0	2.0	fF
Q	100,000	20,000	-
C_o	1.4	1.4	pF
M	140	28	-

As shown above, Q suffers a 5-to-1 reduction for the nonevacuated device. Although it is possible to achieve the required amount of damping of vibrational resonances of the accelerometer with gas pressures less than 1 atmosphere, it is reasonable to expect the degradation of Q, due to the effects of both mechanical losses and viscous damping in the accelerometer, can be represented by that factor.

Figure 2a is a plot of the impedance of the V crystal showing the series and parallel resonances. The plots were made from circuit models of the resonators constructed from the values listed in Table 1.

CH2690-6/89/0000-037 $1.00 C 1989 IEEE

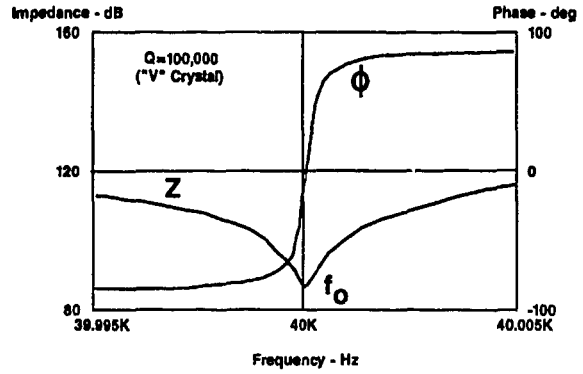


Impedance of High-Q Resonator
Figure 2a

In Figure 2b, expanded about series resonance, the slope of the phase is steep implying good frequency stability for circuit phase variations as would be expected in applying the well-known relationship

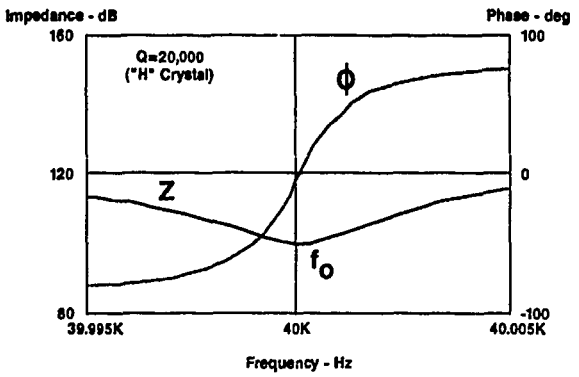
$$\frac{d\omega}{\omega_o} = \frac{1}{2Q} \, d\phi \qquad (2)$$

This equation, valid for resonators having moderate figures of merit, is a result of approximations. A derivation that includes M and limits for approximation is contained in the Appendix.



High-Q Resonator With Frequency Scale Expanded
Figure 2b

Figure 3 is the corresponding plot for the H crystal. The slope of the phase at resonance is one-fifth that of the V resonator as would be expected from the ratio of Q's.

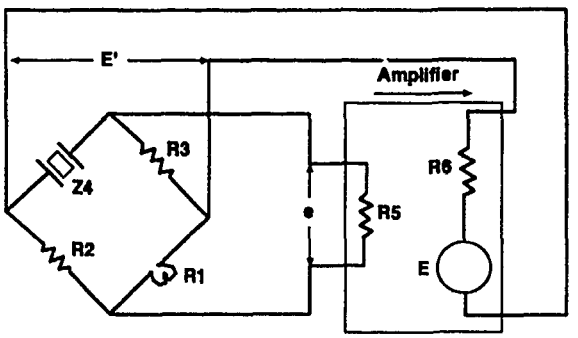


Impedance of Lower Q Resonator
Figure 3

As a result of lowered Q and less certainty of the stability of its value the requirements for the oscillator electronics become more stringent. Not only must the design exhibit reduced sensitivity to R_m but also greater care must be exercised in the design to minimize phase shifts to reduce resultant frequency sensitivity.

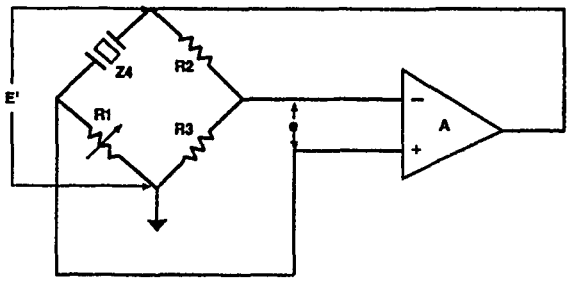
Transimpedance Oscillator

The circuit topography of the transimpedance configuration which addresses the extended performance requirements for the electronics can be developed from the Meacham bridge-stabilized oscillator circuit, shown in Figure 4 with the same circuit element designations as in the original paper [1].



Bridge-Stabilized Oscillator
Figure 4

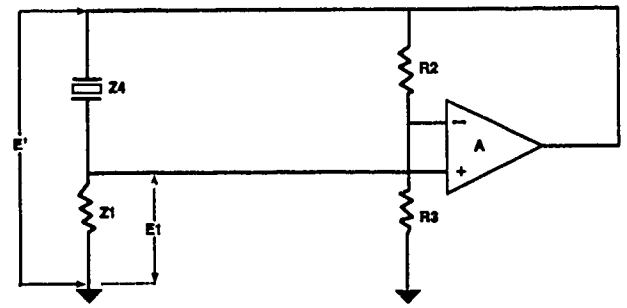
By rotating the bridge one-quarter turn to the right and reorienting it in mirror image , replacing the full differential-input differential-output amplifier with a high-gain wide-band amplifier of the "op-amp" variety, and exchanging the locations of the amplitude control element R1 and fixed resistor R3, the schematic diagram of Figure 5 is obtained. The exchange of these circuit elements follows Meacham's own assertion of that point. The location of the amplitude control element, the drain-source resistance of a field effect transistor in this application, in series with the crystal, brings about one of the desireable performance attributes of the circuit: reduced sensitivity to Rm variation.



Oscillator of Figure 4 With Control Element R1
Exchanged With R3
Figure 5

Figure 6 is merely an associative redrawing of Figure 5 clearly showing the half-bridge comprising the resonator and control element, and a noninverting feedback amplifier with gain-determining resistors R2 and R3. The component identifier for the amplitude control device has been changed to Z1 in anticipation of adding a shunt capacitor

to frequency-compensate that arm. The value of that capacitor can be determined by imposing Barkhausen's criterion [8] on the circuit along with the further constraint that oscillation take place at the series resonance of the crystal.



An Association of R2 and R3 With Amplifier A
Figure 6

The closed-loop gain of amplifier A, applying the usual assumptions of very high gain and input impedance, and low-output resistance is

$$\frac{E'}{E1} = \frac{R2 + R3}{R3} \qquad (3)$$

The gain of the half-bridge is

$$\frac{E1}{E'} = \frac{Z1}{Z4 + Z1} \qquad (4)$$

Equating the gain around the loop to 1,

$$\left(\frac{Z1}{Z4 + Z1}\right)\left(\frac{R2 + R3}{R3}\right) = 1 \qquad (5a)$$

This reduces to

$$Z1 = \left(\frac{R3}{R2}\right) Z4 \qquad (5b)$$

Now, at series resonance the crystal, Z4, appears as a resistor R_m with capacitance C_o in parallel. By placing a fixed capacitor C1 = $(R2/R3)C_o$ in parallel with control element R1, the requirements for steady-state sinusoidal oscillation at series resonance are satisfied when R1 takes on a steady-state value R1 = $(R3/R2)R_m$. Moreover, as Rm varies due to Q changing as a result of a variety of environmental and mechanical effects, R1 tracks R_m proportionately through the action of the amplitude control loop of the oscillator. The time constants of the upper and lower legs of the half-bridge remain equal to one another because of the fixed capacitance ratio producing no phase shift from the half-bridge. This permits the oscillator to operate at the crystal's series resonant frequency even as it is varied by force-producing acceleration inputs.

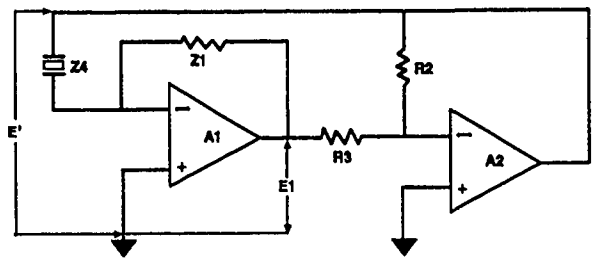
The last source of frequency error to be addressed is that produced by circuit capacitance changes. The greatest sensitivity is at the crystal and half-bridge of Figure 6. The magnitude of the frequency sensitivity is the same to fractional capacitance changes whether or not the changes occur across the crystal or across the lower arm. The equation, also derived in the Appendix, is

$$\frac{d\omega}{\omega_o} = \frac{1}{2QM}\,\frac{dC}{C} \qquad (6)$$

In practice the crystal is usually shielded, and the bridge is scaled to reduce phase shifts produced by absolute capacitance changes across the lower arm. In a like manner the gain-determining resistor values for R2 and R3 are chosen two orders of magnitude or less than bridge impedances reducing phase-induced frequency errors to levels of insignificance. Taking the last step to the transimpedance configuration substantially reduces the effect of circuit capacitances with respect to ground or shield ground. Rearranging Equation 5b results in

$$\left(-\frac{Z1}{Z4}\right)\left(-\frac{R2}{R3}\right) = 1 \qquad (7)$$

Equation 7 suggests the direct implementation of the oscillator with a pair of cascaded inverting amplifiers, shown in Figure 7.



Transimpedance Oscillator
Figure 7

Two advantages are derived from this configuration. The most dramatic is that capacitance changes introduced at the current summing point of A1 with respect to ground are in parallel with the small impedance at that node, thus producing little phase shift. The other advantage is that A2 and A1, configured as inverting amplifiers, suffer no common mode errors in that one input terminal is at ground potential.

TABLE 2. CIRCUIT SIMULATIONS

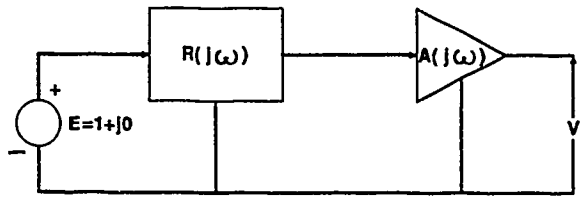
Model and Component Values (Figures 4, 6 and 7)	
Description	**Value**
Crystal Z4	L_m = 7915.72 h C_m = 2.0 fF R_m = 100 K ohm C_o = 1.4 pF
Bridge Leg Z1	R1 = 10.0 K ohm* C1 = 14.0 pF
Components	R2 = 100 K ohm R3 = 10 K ohm
Amplifier(s) A	A = 150x10⁹ ∠90°
Frequency	f = 40 KHz*

*Initial Values of Program-Iterated Variables

Circuit Simulations

To aid in analyzing oscillator performance, each of the oscillator types was modeled to the configurations of Figures 4, 6 and 7. The circuits have been entitled

"Meacham Bridge", "Transposed Bridge" and "Transimpedance", respectively, following the evolution of these configurations. The parts list for each oscillator is identical. The equivalent circuit values for the resonator were computed from the properties for the H crystal as listed in Table 1. The op. amp. is based on VTC's wideband 700-series amplifiers. The gain-bandwidth model was simplified to a simple 1/f gain slope crossing unity gain at 150 MHz. Model and component values are listed in Table 2, with initial conditions for the program-iterated variables. The circuits were solved for the frequency of oscillation for changes in the values of C_o, R_m, and C_s, the latter being circuit capacitance introduced between the output end of the crystal and ground. Solutions were obtained by opening the oscillator loop at the output of the crystal drive amplifier, introducing a voltage source having an output of 1+j0, then measuring the output of the drive amplifier, much in the same way one would bench test an oscillator open-loop as in Figure 8.



Oscillator Loop Open to Measure Return Ratio
Figure 8

Barkhausen's criterion was then forced by simultaneously iterating frequency to null the quadrature component of V, and the value of the control element to force the in-phase component of V to unity, using a convergence algorithm under program control. The results of these simulations are contained in Table 3.

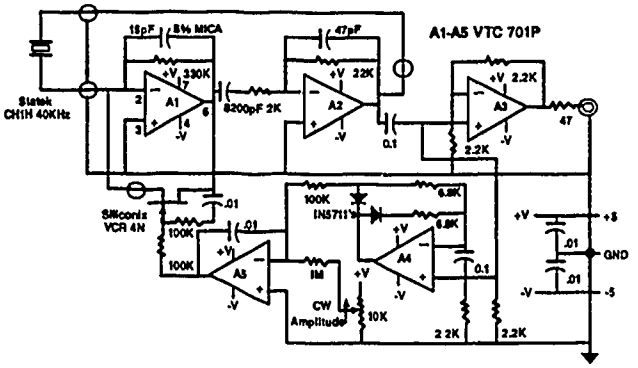
TABLE 3. SIMULATION RESULTS

Circuit Configuration			Sensitivity Factor
Meacham Bridge	Transposed Bridge	Transimpedance	
8.8×10^{-7}	8.8×10^{-7}	8.8×10^{-7}	$\frac{df}{f_o}/\frac{dC_o}{C_o}$
-8.1×10^{-7}	-8.1×10^{-8}	-8.1×10^{-8}	$\frac{df}{f_o}/\frac{dR_m}{R_m}$
-8.8×10^{-7}	-8.8×10^{-7}	-6.2×10^{-14}	$\frac{df}{f_o}/\frac{dC_s}{C_s}$

As is expected, sensitivity to C_o is identical for each of the circuits. For the Transposed Bridge and Transimpedance circuits, sensitivity to R_m variation is reduced by an order of magnitude. For changes in C_s, the Transimpedance circuit is highly immune.

Circuit Tests

A simple circuit was designed to the transimpedance configuration and operated closed-loop with an H crystal at 40 KHz. A schematic diagram of the breadboard is shown in Figure 9. Amplifiers A1, A2 and the crystal form the oscillator loop with A3 serving as a buffer stage to a frequency counter. A4 is configured as a precision half-wave rectifier with A5 functioning as a filter, integrator and amplitude control stage, operating the gate of an N-channel FET control device. The ratio of the neutralizing capacitor (15 pF) to C_o (1.4 pF) is about 10:1 after including FET drain-source capacitance and strays across the crystal. Gain stage A2 is configured as a lead-lag broad-band filter (to suppress crystal overtones) with a gain of -10 at 40 KHz. These scalings, by Equation 7, result in operation at the series resonance of the crystal. Because the input impedance of A1 and the output impedance of A2 are comparatively small due to heavy negative feedback, the effect of capacitive loading to ground on frequency of oscillation is also small.



Schematic Diagram of the Transimpedance Oscillator
Figure 9

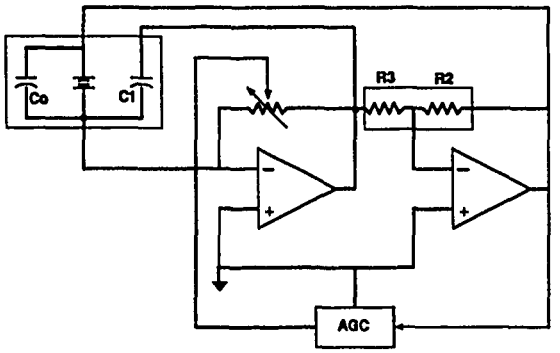
Two tests were performed with the crystal running at 1 microwatt. In the first a 10% increase in the value of the compensation capacitor was introduced and the sensitivity factor was computed from the test data; it compares well with the simulation and theory. In the second experiment a capacitance of 14pF was introduced at the output side of the crystal to ground. That value, equal the value of the neutralization capacitance C1, had little perturbing effect on the output frequency, also in agreement with the simulation results and theory.

Conclusion

The transimpedance configuration, an adaptation of the bridge-stabilized oscillator, exhibits properties suited to operating crystals having low Q's and M's, conditions under which environmental influences on the electronics often compromise transducer performance. By the design of the the circuit the crystal can be operated accurately at series resonance over a wide range of variability of motional resistance, in general not achievable with parallel-mode oscillators. That condition is forced by equalizing the ratio of the neutralizing capacitor to crystal clamp capacitance with the ratio of the gain-determining resistors, stated succinctly in Equation 8.

$$\frac{C1}{C_o} = \frac{R2}{R3} \qquad (8)$$

A simplified schematic diagram of the transimpedance oscillator is shown in Figure 10 with the components whose ratios are important shown enclosed within boxes.



Transimpedance Oscillator With Component Compensation
Figure 10

By fabricating C1 as a fixed, quartz-dielectric capacitor on a section of the transducer structure intended for that purpose and depositing neutralization and resonator electrodes simultaneously, virtual similarity between C_o and C1 could be expected from the monolithic capacitors in close thermal proximity with resultant good tracking characteristics. Ratio-matched resistors with close-tracking temperature coefficients are available, which can also be laser trimmed to achieve the equality of ratios.

Because of insensitivity to capacitive loadings, transimpedance oscillators can operate with the crystal connected to the electronics through several feet of a pair of coaxial cables with shields grounded, effectively placing an electrostatic shield between the terminals of the crystal, with little effect on frequency or performance. This has permitted independent temperature testing of the electronics and resonator while functioning as an oscillator.

For applications where the crystal's series resonant frequency is the desired frequency of oscillation with small errors contributed by the electronics, the transimpedance configuration offers attractive performance advantages.

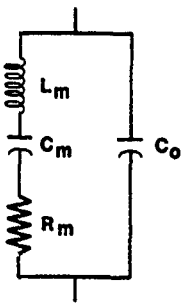
Appendix

Fractional Frequency Stability of the Quartz Resonator, Including the Figure of Merit, M

The equation,

$$\frac{d\omega}{\omega_o} = \frac{1}{2Q}\, d\phi \qquad (A1)$$

is well known and used to estimate the effect of phase variation on the oscillator frequency. It results from assuming a moderate figure of merit, M. Here we derive an expression that includes M.



Equivalent Electrical Circuit of Crystal
Figure A1

For the resonator equivalent circuit the impedance of the series arm can be written

$$Z_s = R_m \left[1 + jQ\left(\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right)\right] \qquad (A2)$$

where

$$\omega_o = (L_m C_m)^{-1/2} \qquad (A2a)$$

and

$$Q = \frac{\omega_o L_m}{R_m} = \frac{1}{\omega_o R_m C_m} \qquad (A2b)$$

Define $$\delta = \frac{\omega - \omega_o}{\omega_o} \qquad (A3)$$

Now $$\frac{\omega}{\omega_o} = \delta + 1 \qquad (A3a)$$

and $$\frac{\omega_o}{\omega} = \frac{1}{\delta + 1} \qquad (A3b)$$

Near ω_o where $(\omega - \omega_o) \ll \omega_o$ and $\delta \ll 1$

$$\frac{1}{\delta + 1} \approx 1 - \delta \qquad (A3c)$$

Using that approximation,

$$Z_s = R_m (1 + j2Q\delta) \qquad (A4)$$

The other arm of the resonator, Co, has a reactance

$$X_o = \frac{1}{j\omega C_o} = \frac{1}{j\omega_o C_o (1 + \delta)} \qquad (A5)$$

A definition of the figure of merit is

$$M = Q \frac{C_m}{C_o} \qquad (A6)$$

and substituting for Q from Equation A2b, M can be expressed

$$M = \frac{1}{\omega_o R_m C_o} \qquad (A7)$$

The quantity $1/R_m C_o$ has units of radian frequency. Thus M can be interpreted as the ratio of two frequencies.

$$M = \frac{\omega_m}{\omega_o} \qquad \text{(A8)}$$

where

$$\omega_m = \frac{1}{R_m C_o} \qquad \text{(A9)}$$

is the corner frequency of R_m in parallel with C_o and ω_o is the frequency of series resonance. By combining Equations A5 and A3 the reactance X_o becomes

$$X_o = \frac{MR_m}{j(1+\delta)} \qquad \text{(A10)}$$

The current flowing in the resonator due to voltage V across it is

(A11)

$$i = \frac{V}{Z_s} + \frac{V}{X_o} = I_o \left[\frac{1}{1 + j2Q\delta} + j \frac{(1+\delta)}{M} \right]$$

where I_o is the current that would flow in the series branch at resonance. Multiplying the first term of A11 by the complex conjugate and expressing terms over the common denominator,

(A12)

$$i = \left[\frac{1 + j\left(\frac{1 + \delta[1 - 2QM] + \delta^2 4Q^2 + \delta^3 4Q^2}{M} \right)}{1 + \delta^2 4Q^2} \right]$$

Neglecting terms containing squared and higher-order powers of δ,

$$i = \frac{I_o}{M} \left[M + j(1 + \delta[1 - 2QM]) \right] \qquad \text{(A13a)}$$

$$\phi = \tan^{-1} \left(\frac{1 + \delta[1 - 2QM]}{M} \right) \qquad \text{(A13b)}$$

and at series resonance where $\delta = 0$

$$i_o = \left(1 + j\frac{1}{M}\right) \qquad \text{(A14a)}$$

$$\phi_o = \tan^{-1}\left(\frac{1}{M}\right) \qquad \text{(A14b)}$$

Equation A14a suggests another interpretation of M as a ratio of currents in the two branches at series resonance. That is

$$M = \frac{I_o}{I_{X_o}} \qquad \text{(A14c)}$$

The frequency sensitivity to phase shift is the slope of ϕ

$$\frac{d\phi}{d\delta} = \frac{d}{d\delta}\left(\tan^{-1}\left[\frac{1 + \delta(1 - 2QM)}{M} \right] \right) \qquad \text{(A15a)}$$

Performing the differentiation and evaluating at $\delta = o$

$$\left.\frac{d\phi}{d\delta}\right|_{\delta = o} = \frac{1 - 2QM}{M + \frac{1}{M}} \qquad \text{(A15b)}$$

Now $\delta = \frac{\omega - \omega_o}{\omega_o}$ and $d\delta = \frac{d\omega}{\omega_o}$

Solving A15b for $\frac{d\omega}{\omega_o}$

$$\frac{d\omega}{\omega_o} = \left(\frac{M + \frac{1}{M}}{1 - 2QM} \right) d\phi \qquad \text{(A15c)}$$

Arranging A15a in the form of Equation A1 but with a multiplier containing M, for Q >> M,

$$\frac{d\omega}{\omega_o} = \left(\frac{M}{M + \frac{1}{M}} \right) \left(-\frac{1}{2Q} \right) d\phi \qquad \text{(A16)}$$

The multiplying factor has a value of 0.9 for M=3, and is greater than 0.99 for M=10. Hence the approximation

$$\frac{d\omega}{\omega_o} = \frac{1}{2Q} d\phi \qquad \text{(A1)}$$

is in error by less than 1% for M > 10. Also A1 has been signed positive, reflecting the general application of the expression to the phase of loop voltages or impedances. Since the derivation of A16 was performed in terms of current and admittance, the sign is negative.

<u>Fractional Frequency Stability of the Quartz Resonator to Series Arm Capacitance Changes in Neutralized Applications</u>

Consider the arm Z1 of the half-bridge of Figure 6 or of Figure 7. The phase of Z1 is

$$\phi = \tan^{-1}(-\omega R1C1) \qquad \text{(A17)}$$

and the differential of phase with respect to the fractional capacitance C1

$$d\phi = -\left(\frac{\omega R1C1}{1 + \omega^2 R1^2 C1^2} \right) \left(\frac{dC1}{C1} \right) \qquad \text{(A18)}$$

With the circuits neutralized and operation at series resonance, $R1C1 = R_m C_o$, $\omega = \omega_o$, and therefore

$$d\phi = -\left(\frac{\frac{1}{M}}{1 + \frac{1}{M^2}} \right) \frac{dC1}{C1} \qquad \text{(A19)}$$

Substituting A19 into A16

$$\frac{d\omega}{\omega_o} = \left(\frac{1}{M + \frac{2}{M} + \frac{1}{M^3}} \right) \left(\frac{1}{2Q} \right) \frac{dC1}{C1} \qquad \text{(A20)}$$

The approximation

$$\frac{d\omega}{\omega_o} = -\frac{1}{2QM}\frac{dC1}{C1} \qquad (A21)$$

is in error by less than 10% for M > 5 and by less than 1% for M > 15. Derived for capacitance changes in Z1, Equation A21 produces a decrease in the frequency of oscillation for an increase in C1, for as the incremental phase of the impedance of Z1 becomes more negative (as C1 increases) the phase of the current in the series arm of the crystal increases in a positive direction so as to produce a total resonator current having a phase equal in magnitude but opposite in sign to the phase of Z1 to satisfy the zero phase criterion around the loop. This corresponds to a lower crystal operating frequency.

For the case of change of capacitance C_o where an increase in C_o increases the angle of lead of total resonator current, the series arm must develop a lag angle to retain the total angle. This corresponds to a frequency increase signing Equation A22 positive.

$$\frac{d\omega}{\omega_o} = \frac{1}{2QM}\frac{dC_o}{C_o} \qquad (A22)$$

Acknowledgments

The author thanks Lawrence W. Van Der Hey for data reduction, Francis X. Buran for computer circuit simulations and Frank Castronuovo for oscillator construction.

References

[1] L. A. Meacham, "The Bridge-Stabilized Oscillator," Proc. of the IRE, vol. 26, no. 10, pp 1278-1295, Oct. 1938.

[2] L. S. Ferriss, "Transimpedance Oscillator Having High Gain Amplifier," U.S. Patent No. 4,358,742, Nov. 1982.

[3] W. C. Albert, "Vibrating Quartz Crystal Beam Accelerometer," Proc. of the 38th International Instrumentation Symposium, vol. 38, no. 1, 1982, pp.33-44.

[4] W. C. Albert, "Force Sensing Using Quartz Crystal Flexure Resonators," Proc. of the 38th Annual Symposium on Frequency Control, 1984, pp. 233-239.

[5] W. J. Kass and G. S. Snow, "Double-Ended Tuning Fork Quartz Accelerometer," Proc. of the 40th Annual Symposium on Frequency Control, 1986, pp.230-236.

[6] I. E. Fair, "Piezoelectric Crystals in Oscillator Circuits," The Bell System Technical Journal, vol.24, no. 2, pp. 161-201, April, 1945.

[7] Statek Corporation, Short Form Catalog for Miniature Low-Frequency Quartz Crystals, Jan. 1987, p.2.

[8] W. A. Edson, Vacuum Tube Oscillators, New York: John Wiley and Sons, 1953, introduction, p.7.

43rd Annual Symposium on Frequency Control - 1989

THERMOELECTRIC COOLER/HEATER CONTROLLED CRYSTAL OSCILLATOR (TECXO)

M.ZELITZKI, D.PINCU and I.EDRY
TFL, Time and Frequency Ltd
POB 1792, Holon 58117, Israel

Abstract

It is well known that the performance of the conventional Oven Controlled Crystal Oscillator (OCXO) degrades significantly when operated at ambient temperatures above 70°C since the internal oven temperature, which is limited by the reliability of the electrical components to about 80°C, must be at least 10°C higher than the external ambient temperature.
This paper describes a new development that enables the operation of OCXO's for military or special applications over the extended temperature range of -55°C up to 120°C in a moderately sized case, and with moderate power consumption.
The implementation of a device with the above features is achieved by using a Thermoelectric (TE) Cooler component which is a plate, one side of which is cold and the other hot. The cold side is attached to the device to be cooled, and the hot side to a heatsink. The amount of absorbed heat or the cooling level are proportional to the direct current passing through the component. Changing the current polarity causes an inversion of the hot and the cold sides, in this manner enabling heating of the attached device. Hence, it becomes possible to stabilize the internal temperature of the Thermoelectric Cooler/Heater Controlled Crystal Oscillator (TECXO) at any level independently of the external ambient temperature.
In order to evaluate the potential advantages of the TECXO, a basic prototype has been developed in the laboratories of TFL. The performance of the device has been measured and analyzed. The conclusions drawn are detailed in the paper.

Introduction

The conventional Oven Controlled Crystal Oscillator (OCXO) operates with poor performance at ambient temperatures above 70°C. This limitation is due to the fact that the reliability of the electrical components used in the internal oven degrades significantly at temperatures higher than 80°C, while this internal oven temperature must be at least 10°C above the external ambient temperature in order to enable proper oven operation.
A new development, using a Thermoelectric (TE) Cooler component, permits operation of OCXO's, for military or other special applications, over the extended temperature range of -55°C to 120°C, with moderate power consumption, and in a moderately sized case.

The principle of operation of the TE cooler is based on the Peltier effect which involves heat absorption or emission when direct current flows through a PN junction. The thermal nature of the effect stems from the transition of electrons from a specific energy level in one material to a different energy level in another material. Heat is absorbed by the junction whenever electrons transition occurs from a lower energy level to a higher one, and is emitted by the junction whenever the transition occurs in the opposite direction. The amount of heat absorbed or emitted is proportional to the level of direct current passing through the junction. As a result, hot and cold junctions are obtained, and these can be attached on both sides to substrates. A diagram of a TE cooler junction is shown in Figure 1.

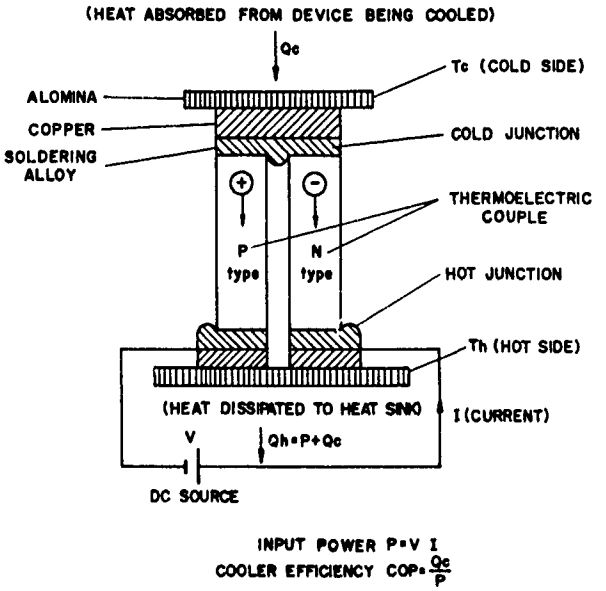


Figure 1. Diagram of a Thermoelectric Cooler Junction.

An array of several junctions will produce a plate, one side of which is cold and the other hot. An array of junctions is shown in Figure 2. The cold side is attached to the device to be cooled, while the hot side is attached to a heatsink.

CH2690-6/89/0000-044 $1.00 C 1989 IEEE

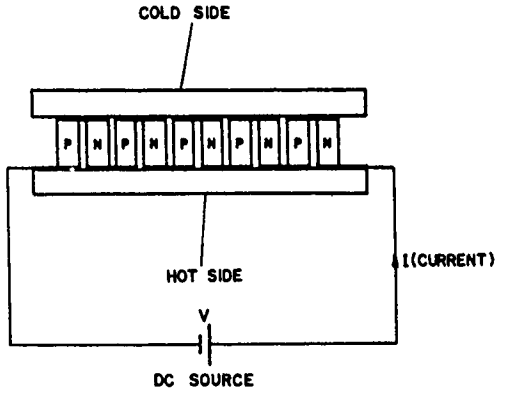


Figure 2. Thermoelectric Cooler Comprising an Array of Junctions.

A typical curve, describing the variation of the cold side temperature versus current when no heat is pumped from the cold side (Qc=0), is shown in Figure 3.

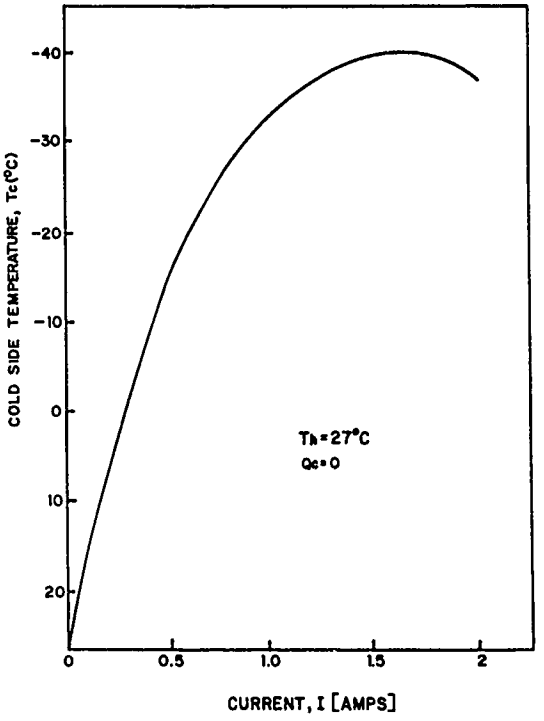


Figure 3. Cold Side Temperature T_c versus Current for constant Hot Side Temperature T_H and $Q_c=0$

Changing the current polarity causes an inversion of the hot and cold sides, thus enabling heating of the attached device. TE coolers, which are relatively cheap, are produced by several manufacturers for various uses.
Hence, using a thermoelectric heater/cooler creates the possibility of stabilizing the internal temperature of an oscillator at any level independently of the external ambient temperature. We have called this device a Thermoelectric Cooler/Heater Controlled Crystal Oscillator (TECXO).

Design Approach

The potential advantages to be attained by this novel type of oscillator over the conventional OCXO seemed to be as follows:
a. Operation at particularly high ambient temperatures and over a wide temperature range - due to the above explained operation principle.
b. Shorter warm-up periods feasible - due to the fact that the internal temperature stabilization is carried out using both heating and cooling processes which reduces the thermal overshoot damping period.
c. Better aging characteristics - due to the fact that the oscillator is operated at lower temperatures and since higher operating temperatures accelerate the detrimental aging process.
d. Lower phase noise - due to the fact that the oscillator is operated at lower temperatures and since higher operating temperatures lead to increase the phase noise.
e. Lower power consumption (despite the fact that the Thermoelectric Cooler is not particularly efficient) - due to the fact that it is possible to set the internal temperature in the middle of the range of external temperature variations, and then the heating or cooling are carried out over only half of the range, while in the conventional OCXO only heating over the whole range may be performed.

In order to test the validity of the above assumptions concerning the potential advantages of the TECXO and confirm them, a basic prototype oscillator of this type has been developed in the laboratories of TFL. The oscillator contained an Sc-cut type crystal that was designed especially to have a relatively low turnover temperature (the temperature at which the frequency is the most stable and at which the oven is generally kept), a proportional temperature controller which controls the current fed into the TE cooler, as used in the conventional OCXO's, and a temperature sensor connected to the temperature controller. Since the internal temperature was lower than the external ambient temperature, it was necessary to be particular about the correct insulation of the oscillator and the crystal from the surroundings. Factors that were liable to adversely affect the insulation included the connecting wires to the TE cooler, the fastening screws and even the TE cooler itself that come between the internal oscillator (the cold side) and the external surroundings (the hot side). It is possible to overcome the first two factors but not the last, which reduces the cooling efficiency.

Results and Performance

The TECXO prototype was constructed with a mechanical structure and associated circuitry similar to those used with an existing OCXO having an excellent known performance (built in a double oven configuration and consisting of an Sc-cut type crystal). The performance of both devices has been measured and a comparative analysis has been performed.

There was no degradation as compared with the OCXO and some slight improvement in

performance was noted. Some of the measurements made to assess performane of the TECXO compared to the OCXO performance are given below :

Stability over temperature range

TECXO (-30°C to 85°C)	OCXO (-30°C to 60°C)
$\pm 2 \cdot 10^{-9}$	$\pm 2 \cdot 10^{-9}$

Warm-up time (at 25°C)

	TECXO	OCXO
within $\pm 1 \cdot 10^{-7}$	5 minutes	8 minutes
within $\pm 3 \cdot 10^{-9}$	15 minutes	30 minutes

Long-term stability

TECXO	OCXO
$\pm 5 \cdot 10^{-11}$/day	$\pm 5 \cdot 10^{-11}$/day

Short-term stability

Averaging Time	Allan Deviation TECXO	OCXO
1msec	$1 \cdot 10^{-10}$	$2 \cdot 10^{-10}$
10msec	$5 \cdot 10^{-11}$	$7 \cdot 10^{-11}$
100msec	$1 \cdot 10^{-11}$	$1 \cdot 10^{-11}$
1 sec	$5 \cdot 10^{-12}$	$5 \cdot 10^{-12}$

Phase noise

Offset from carrier	Phase Noise TECXO	OCXO
10Hz	-130dBc/Hz	-125dBc/Hz
100Hz	-140dBc/Hz	-135dBc/Hz
1kHz	-150dBc/Hz	-150dBc/Hz
10kHz	-160dBc/Hz	-160dBc/Hz

Maximum power consumption (after warm up)

TECXO (at 85°C)	OCXO (at -30°C)
8.0 watts	8.1 watts

Conclusions

The conclusions drawn concerning the achievement of the potential advantages of the Thermoelectric Cooler/Heater Controlled Crystal Oscillator mentined above, are as follows :

a. The TECXO can operate at an unprecedented high temperature of 120°C, and over a very wide temperature range, -55°C to +120°C.

b. The warm-up time obetained is indeed shorter but the improvement is not of a significant order.

c. The aging characteristics were found to be without major change.
This may be attributed to the fact that the aging is expected to improve at very low operating internal temperatures, which after all were not found feasible for the oscillator, for the reasons mentioned in e. below.

d. The phase noise and the short term stability were both improved but not by a significant order, for the reasons mentioned in c. above.

e. The power consumption was found to be similar to the conventional OCXO. Moreover, the Thermoelectric Cooler is not efficient enough to enable setting the internal temperature in the middle of the external temperature range; in fact it is required to set it closer to the upper range limit in order to save the cooling energy. For example an internal temperature of 40°C is recommended for an external temperature range of -40°C to 85°C. In order to heat the oscillator in the low ambient temperatures, a heating element appears to be required which would be more efficient than the TE heating (despite the fact that TE heating is more efficient than TE cooling). Such an additional heating element also saves the need for switching the TE current polarity. Some advantage may be the fact, that at 25°C external ambient temperature (the normal operating temperature) the power consumption of the TEXCO is minimal.

The following additional conclusions were drawn:

f. It is required to minimize the volume being cooled in order to reduce the amount of heat to be pumped and so decrease the power consumption.

g. Several TE elements in parallel are required, distributed around the volume to be cooled to achieve the cooling continuity and uniformity similar to a spiral heating element wound around the volume.

h. The TECXO requires the attachment of a heatsink to dissipate the heat pumped away.

References

[1] A.D.Kraus and A.Bar-Cohen, "Thermal Analysis Control of Electronic Equipment". ch. 18, pp. 435-465.

[2] E.S.Rittner, "On the Theory of the Peltier Heat Pump". J.Appl. Phys, vol.30, PP. 702-707.

[3] P.Kartaschoff, "Frequency and Time". Academic Press, 1978, ch. 3, PP. 37-43.

[4] "Detector Cooling Systems". ch. 12.2.4, pp. 544-557.

[5] R.Marlow, "Estimating Heat Loads for Thermoelectric Coolers". Marlow Industries.

[6] "Guide to Thermoelectric Heat Pumps". Marlow Industries.

[7] "Solid State Cooling with Thermoelectrics". Melcor.

43rd Annual Symposium on Frequency Control - 1989

RESULTS OF LONG TERM TESTING OF TACTICAL MINIATURE CRYSTAL OSCILLATORS

J. MESSINA, D. BOWMAN, R. FILLER, R. LINDENMUTH, V. ROSATI, S. SCHODOWSKI
U.S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

Summary

This paper describes the results of long-term testing of the Tactical Miniature Crystal Oscillator. The TMXO employs high-vacuum thermal insulation, microelectronic packaging, and a precision ceramic-flatpack-enclosed SC-cut crystal.

Engineering models and pilot-production models, delivered under a Manufacturing Methods and Technology effort with the Bendix Division of Allied Corporation, were subjected to a variety of tests that included turn-on/turn-off at temperature extremes, input-power aging, and frequency aging. One of the primary goals of the testing was to verify the ability of the TMXO to maintain vacuum integrity over a reasonable lifetime (10 years).

Twenty-two of 25 oscillators tested for input-power aging showed less than an eight percent increase over the one-year period during which input power was monitored. It is concluded that TMXO-type devices can be manufactured such that the high vacuum so critical to their performance can be maintained over a long period of time.

Introduction

The Tactical Miniature Crystal Oscillator is a high performance quartz oscillator designed for use as a frequency/time reference in advanced communication and navigation systems. It contains a ceramic-flatpack-enclosed 10 MHz, 3rd overtone, SC-cut crystal resonator, and it employs microelectronic packaging and high vacuum thermal insulation. It is the smallest low-power oven-controlled crystal oscillator known to the authors. Maintenance of the high vacuum is provided by a nonevaporable, refireable, zirconium-graphite getter.[1] The getter is intended to remove gases that result from outgassing of components in the interior of the TMXO enclosure, or from fine leaks from the ambient atmosphere.

Vacuum integrity is crucial to the performance of the TMXO. Verification of the ability to maintain the vacuum is necessary for the TMXO to be accepted as a viable, manufacturable design.

Over the past three years, 80 pilot-line and engineering-model TMXOs were subjected to long-term testing. These tests focused on input power aging, frequency aging, turn-on/turn-off testing, and temperature stressing. The experimental methods,[2] the ceramic-flatpack-enclosed resonators,[3] and the TMXOs[4] used in these tests have been described previously.

Input Power Aging

Twenty-five TMXOs were monitored for changes in input power. Five units were engineering samples (made in 1985) that had undergone several temperature cycles (-55°C to 75°C). Ten of the units were pilot-line models (made in 1986) that had previously been tested for retrace (off/on) at -55°C and 70°C.[5] The remaining ten units were pilot line models (also made in 1986) that had been tested only during manufacture and acceptance.

Heater current and ambient temperature were measured once a week for 52 weeks. The current was measured through a one-ohm resistor. Temperature was measured at the chassis holding the 25 units under test.

Figure 1 is a representative plot of input current versus time for a 52 week period. The curve at the bottom of Figure 1 is the recorded temperature. The bold curve at the top is the recorded input current vs. time, and the third curve is the current corrected for deviations in the ambient temperature.

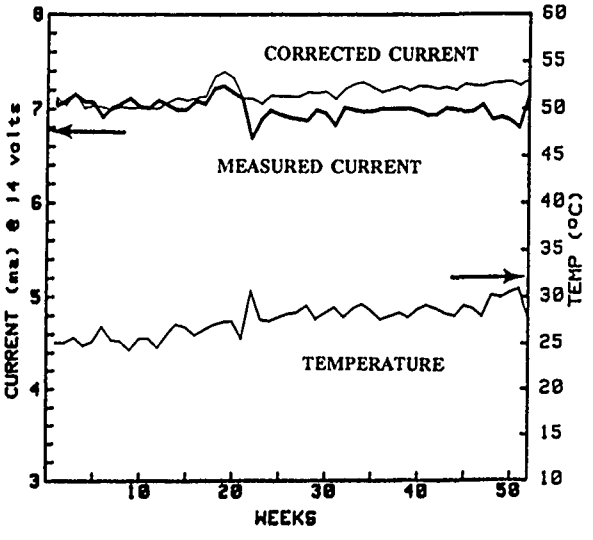


Figure 1 - Input Current vs. Time

Results of these tests on the 25 TMXOs reveal that all but three of the oscillators experienced a change in input power of less than 8 percent after 52 weeks (Figure 2). The average power increase at room temperature was about seven milliwatts. There was one notable exception to the average behavior of the oscillators. One unit doubled its input power over a period of only five weeks which indicates a gross leak that allowed the interior of the oscillator enclosure to reach atmospheric pressure. Prior to mounting in the test chassis, that unit showed normal input power, which suggests that a leak was caused by over-tightening of the mounting studs.

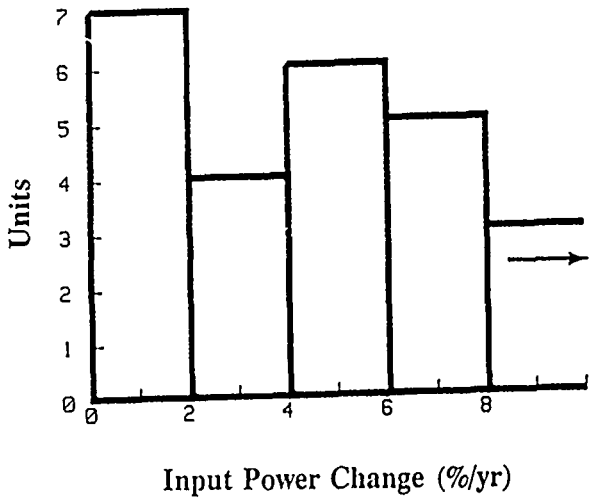


Figure 2 - Distribution of input power aging for 25 oscillators

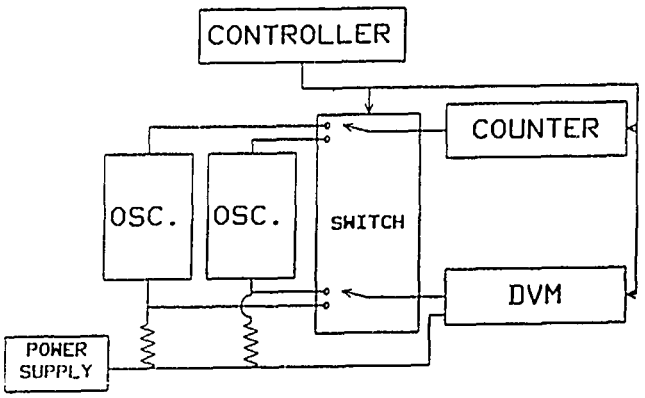


Figure 3 - Automated test setup used to measure aging

Frequency Aging

Only the twenty pilot-line TMXOs that were measured for input-power aging were measured for frequency aging. (The five engineering models were undergoing other tests, which precluded their inclusion in the frequency-measuring routine.) The automated test setup in Figure 3 was used to measure and record the frequency of each oscillator four times per day. About 600 days of data was collected.

In Figure 4, frequency aging data is presented for a typical TMXO. The dashed line is a linear approximation of the slope over the last third of the data. The best TMXO shows an aging rate of parts in 10^{13} per day (Fig. 5). Figure 6 is a histogram of the aging of all the tested units.

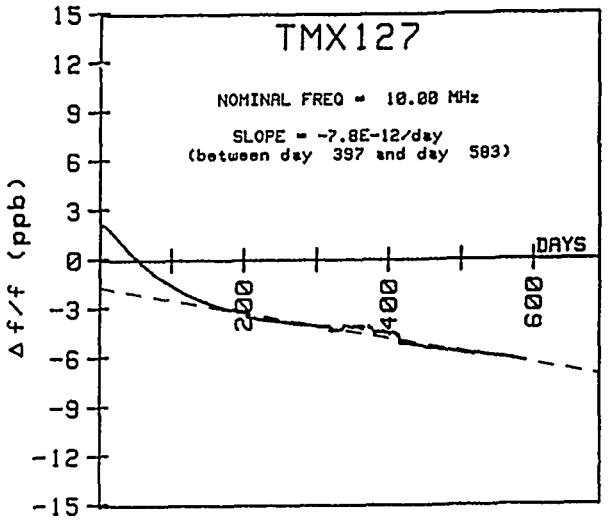


Figure 4 - Frequency aging for a typical TMXO

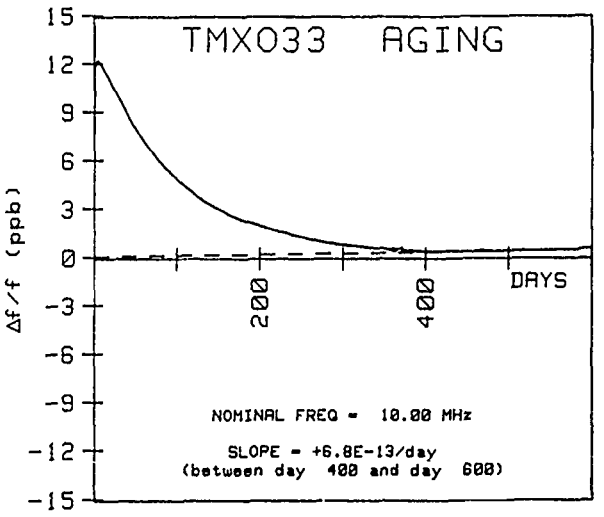


Figure 5 - Parts in 10^{13} aging for a TMXO

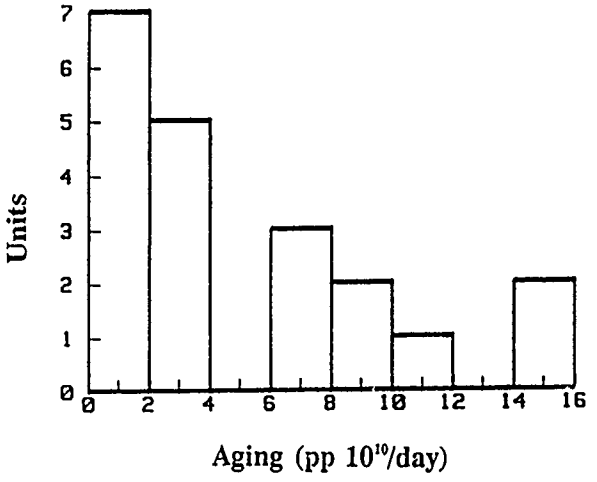


Figure 6 - Distribution of frequency aging for 20 TMXOs at day 500

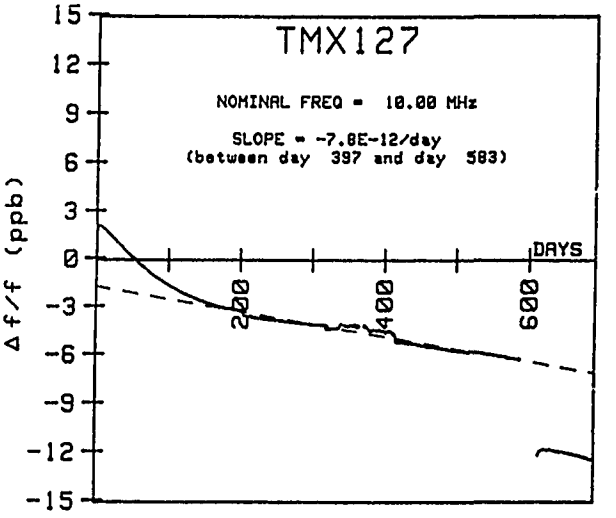


Figure 7 - TMXO frequency aging after temperature cycling

Temperature Cycling During Storage

The final test performed on the TMXOs was designed to study the effect of temperature cycling on the TMXO while the oscillator was powered down. Eighteen TMXOs were removed from the input power/frequency aging setup and placed in a temperature chamber. The oscillators were stored for 24-hour periods at -45°C, 85°C, -45°C... for one month, after which the TMXOs were returned to the aging set-up where the measurement and recording of frequency and input power were resumed.

Input power measurements recorded after the temperature cycling indicate that this stress had little effect on the input power. There was virtually no change in the input current or the input current aging rate after the temperature cycling. The temperature cycling did, however, have an effect on the frequency. Data recorded after the storage test reveals an offset in the frequency. Figure 7 presents the aging data for the TMXO shown previously in Figure 1 and includes data taken after the temperature cycling. The temperature cycling test was performed near day 600. The offset which occurred after the temperature cycling was about 6 ppb for this TMXO. A distribution of the change in frequency for all 18 TMXOs is shown in Figure 8. Eight units changed less than 10 ppb, but several demonstrated much larger changes.

Frequency-temperature performance

The frequency vs. temperature characteristics of five TMXO are plotted in Figure 9. The worst unit shows less than 3 X 10^{-9} fractional frequency deviation over the entire range of the test (-45°C to 75°C), and the best unit in this group shows a deviation of only 3 X 10^{-10}. This TMXO compares favorably with rubidium frequency standards, which show a frequency deviation of 3 X 10^{-10} over a much narrower temperature range.

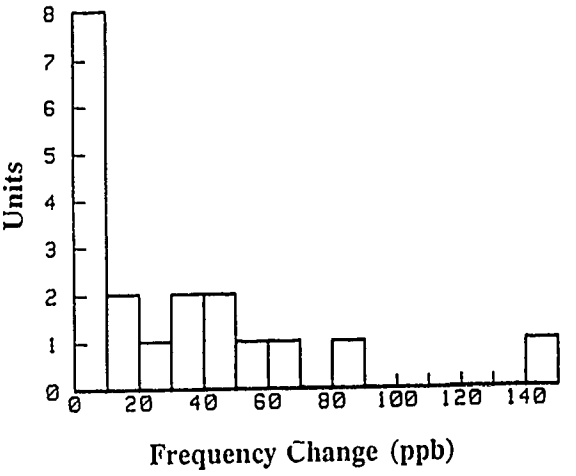


Figure 8 - Distribution of the change in frequency after temperature cycling for 18 oscillators

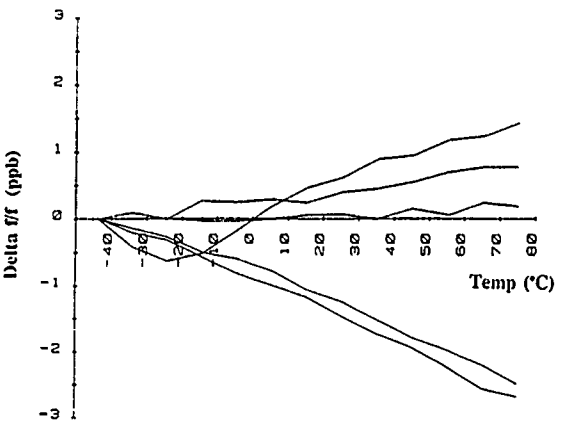


Figure 9. Frequency vs. temperature of five representative TMXOs

Conclusions

We conclude from the results of long-term input-power measurements that the concept of using vacuum insulation is viable. An effort to lower the cost of the TMXO by improving producibility is currently ongoing.[6] The crystal package, hybrid circuit, and outer enclosure have been redesigned to lower the cost of components and improve the production yield.

References

1.Saes Getters/USA Inc. Colorado Springs, Colorado, Part No. St 172/LHI/4-7/200.

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. D. Brown, E. Laszlo, R. McGill, and P. Stoermar, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillators," Proc. 38th Ann.Symp. on Freq. Control, pp. 380-386, 1984.

5. R. Filler, R. Lindenmuth, J. Messina, V. Rosati, and J. Vig, "The Aging of Resonators and Oscillators Under Various Conditions," Proc. 41st Ann. Symp. on Freq. Control, pp.444-451, 1987.

6. DAAL01-86-C-0011 with Piezo Technology Inc. Improved Tactical Miniature Crystal Oscillator.

43rd Annual Symposium on Frequency Control - 1989

LOW PROFILE HIGH STABILITY DIGITAL TCXO ;
ULTRA LOW POWER CONSUMPTION TCXO

V. CANDELIER, G. CARET and A. DEBAISIEUX
CEPE - 44, Avenue de la Glacière 95105 ARGENTEUIL CEDEX FRANCE

ABSTRACT

ATCXO

CEPE has recently developped a new analog TCXO range. These designs are based on the special thermal compensation in order to achieve a frequency stability of ± 2.10^{-6} over all operating conditions (temperature, power supply, load, 1 year aging) with an ultra low power consumption (300 µA, 3V). This new generation of oscillators correspond exactl. to the key requierements of the new telecommunication systemes (Manpack, Portable...).

DTCXO

System requirements and also analogical temperature compensated crystal oscillator (TCXO) capabilities have conducted to develop a new generation of TCXO with a thermal compensation by numerical method, called digital temperature compensated crystal oscillator (DTCXO).

This paper describes the method that has been used in order to realize the digital compensation and presents results abtained on a first batch of DTCXO.

The principle of the digital compensation is to store in a memory the curve of frequency drift versus temperature of the resonator. There fore a thermometer is used in order to indicate the real value of the re onator temperature this method uses a piezoelectric resonator as a sensor, its frequency variation versus temperature is quite linear. Count of the frequency gives temperature information with addresses directly the memory. AD/A converter is required to apply the corrected value to a varactor.

This configuration is used to achieve a frequency stability of 3.10^{-7} over a vide temperature range (- 40°C to 85°C). In order to reduce the volume and thickness (0,4 inche), we have developed a gate array circuit which integrates the counting system and the thermal oscillator.

1/ INTRODUCTION

For the last 10 years, the TCXO requirements in terms of power consumption, frequency stability etc... for modern navigation GPS and telecommunications systems have required new generations of oscillators for which CEPE has developed two oscillators with different thermal compensation concepts :

- Analog TCXO with ultra low power consumption and a frequency stability under overall conditions (temperature, ageing ...) better than +/- 2.10^{-6}.
- Digital TCXO with a thermal stability better than +/- 3.10^{-7} over a large temperature range.

The purpose of this paper is to present the experimental results of these two types of TCXO's.

2/ ANALOG TCXO : ATCXO

This TCXO has been developed in order to get a frequency and time standard for frequency-hopping transceivers for vehicular and portable requirement.

2.1/ Definition

The electronic design of this temperature compensated crystal oscillator uses very common temperature sensitive components and analogic compensation network (fig.1)

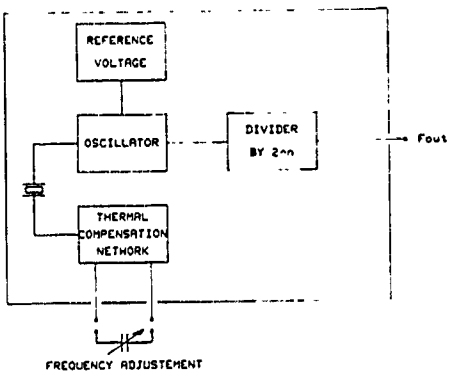


Fig. 1 - Analog TCXO block diagram

We have in the same package one HCMOS divider (Fout = $Fo/2^n$) and one voltage supply regulator in order to assure the frequency stability for different supply voltages between 2.8 V and 5.2 V. The frequency adjustment is done by the variation of an external capacitor. This oscillator is enclosed in a small sealed metal package.

CH2690-6/89/0000-051 $1.00 C 1989 IEEE

2.2/ Experimental results

After spice modelling of the oscillator concept, we have selected the best circuit compromise between thermal compensation definition and resonator parameters in order to guarantee the frequency stability under all environmental conditions at lowest possible power consumption. For this type oscillator, we use a fundamental AT cut resonator. The table 1 hereunder summarizes key experimental results for different lots of oscillators.

Frequencies	2^n Hz	
	$0 \leq n \leq 15$	$15 \leq n \leq 22$
Output level	HCMOS	
on load	15 pF	
Supplies Voltage	2.8 V to 5.2 V	
Supplies Current	$\leq$300 µA/3 V $\leq$600 µA /5V	$\leq$600 µA/3V $\leq$1 mA/5V
Overall stability versus - Temperature range - Frequency adjustment - Load - Supply voltage variation - Ageing first year	$\leq \pm 2.10^{-6}$	
Temperature range	- 40°C 85°C	
Adjustment range	$\geq \pm 5.10^{-6}$ with CVext between 0 and 5.6 pF	
Adjustment guaranteed for	7 years	

Table I - ATCXO BASIC RESULTS

As you will notice the power consumption is depending on the output frequency. The lowest current supply (300 µA) is achieved when the output frequency range is between 1 Hz to 32 KHz.

The frequency stability with in the temperature range of - 40 to + 85°C and with the supply voltage variations (2.8 V to 5.2 V) is expressed in fig. 2. We achieve the stability +/- 2.10^{-6} with the frequency adjustment (Fig. 3). The ageing for the first months is shown in fig. 4.

Fo = 4 Hz

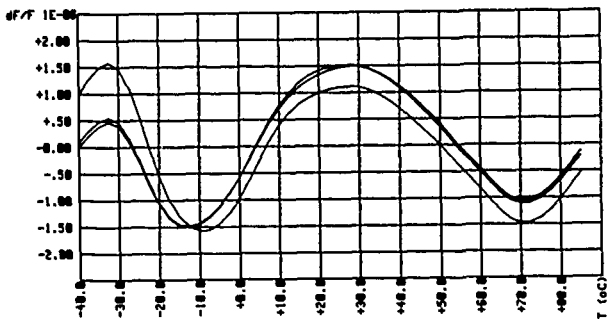
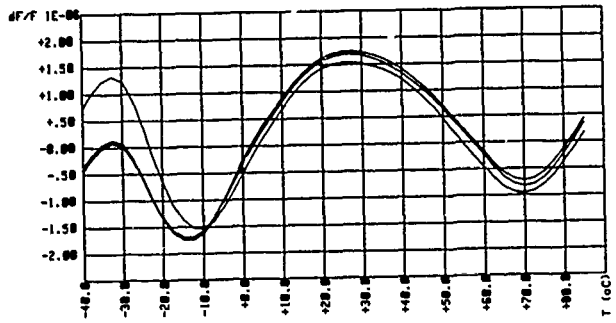


Fig. 2 - ATCXO frequency stability in the temperature range (-40° + 85°C) at three supply voltages 2.8V, 4V, 5.2V. The measurements are made at the reference frequency Fo.

Fo x $(1 + 5.10^{-6})$



Fo x $(1 - 5.10^{-6})$

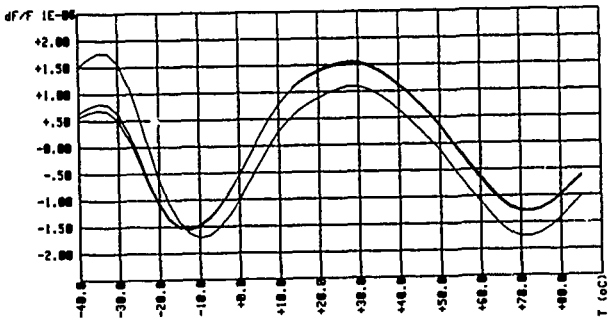


Fig. 3 - ATCXO frequency stability in the temperature range (-40° + 85°C) at three supply voltages 2.8V,4V,5.2V. The measurements are made with the frequency adjusted at Fo (1 + 5.10-6) and Fo (1 - 5.10-6).

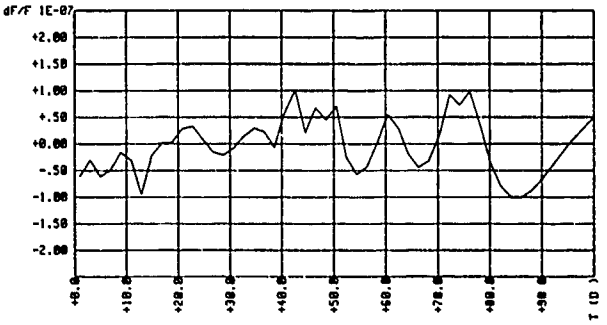


Fig. 4 - ATCXO ageing for the first months

3/ DIGITAL TCXO : DTCXO

3.1/ Principe

The principe of the digital compensation is to store in a memory the curve of frequency drift versus temperature of the resonator. Therefore a thermometer is used in order to indicate the real value of the resonator temperature. We use a piezoelectric resonator as a sensor in order to assure the calibration during the life of this DTCXO(1).Counting of the frequency of the thermal sensor gives the temperature information which adresses directly the EPROM memory. AD/A converter is required to apply the corrected value to the varactor of the reference oscillator To miniaturize the package we developed two gate array circuits which integrate, one for the counting function and one for the reference oscillator The block diagram is shown in fig. 5. We have separeted the digital and the analogic electronics in order to suppress spurious and ripple phenomenon.

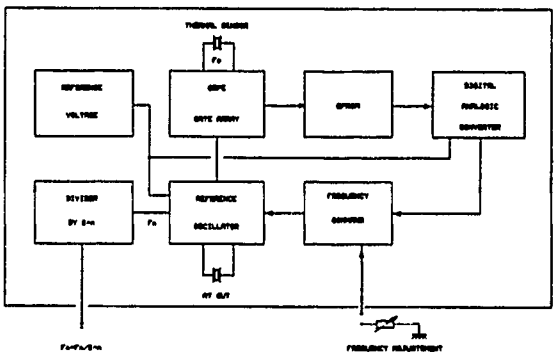


Fig. 5 - DTCXO block diagram

3.2/ EXPERIMENTAL RESULTS

During the experimental phase, we verified the thermal behavior. The improvement of the DTCXO concept required that we focused on the following points :

- The thermal coupling between the thermal sensor resonator and the reference resonator. Indeed the frequency stability versus thermal gradient is directly depending on this parameter (fig. 6).

- In order to guarantee the thermal behavior of the DTCXO for several control voltages years, we hadto compensate the non linearity of the frequency control varactor of the oscillator after recalibration. Regarding the frequency adjustment, we show that we maintain stability over the temperature range (fig. 7).

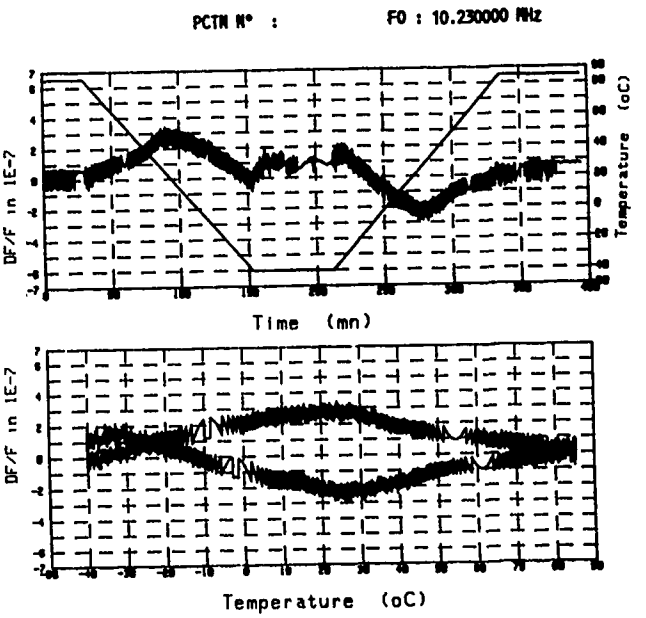


Fig. 6 - DTCXO frequency stability under a thermal gradient of +/- 1°C/min between (-40° + 85°C).The frequency hysteresis is smaller than+/- 3.10^{-7}. The measurements are made at the reference frequency, Fo, 10,23 MHz.

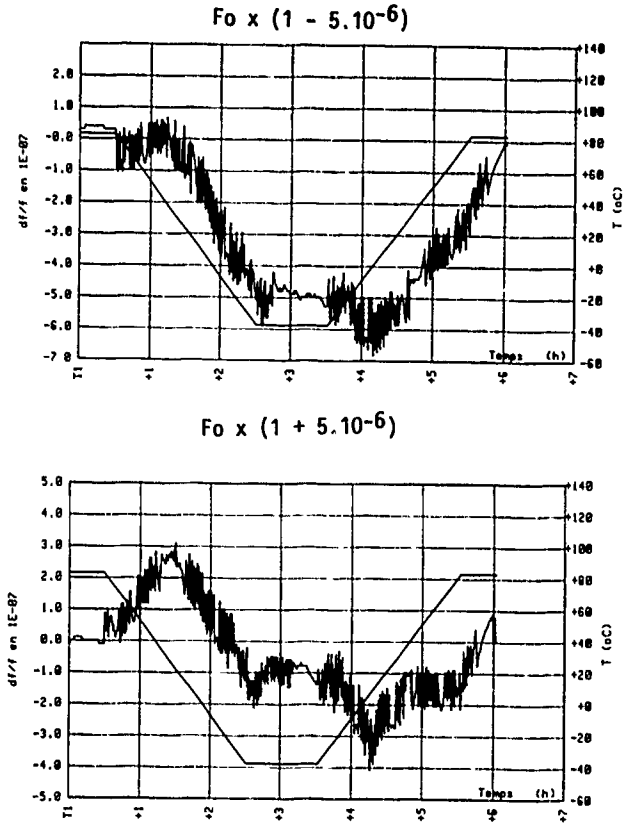


Fig 7 - DTCXO frequency stability under a thermal gradient of +/- 1°C/min between (-40° + 85°C). The measurements are made at Fo(1+5.10-6) and Fo(1-5.10-6). The frequency hysteresis stays smaller than +/- 3.10-7

The table II hereunder summarizes the basic experimental results.

Frequency range		2 MHz to 25 MHz		
Standard frequency		5 MHz - 10 MHz		
Temperature range		C -40° + 70°C	F - 40° + 85°C	D - 55° + 105°C
Worst case stability		≤ ± 3.10-7	≤ ± 5.10-7	≤ ± 8.10-7
For combined effects of vibration of :				
. Temperature		Over all temp. range		
. Supply voltage		± 5 % of nominal voltage		
. Load		1 to 2 TTL LS Loads		
Frequency stability for stable conditions	per day month year	≤ ± 1.10-8 ≤ ± 2.10-7 ≤ ± 1.10-6		
Freq. adjust. range	Total	≥ 5.10-6		
Insured for		7 YEARS		

Table II - DTCXO BASIC RESULTS

Concerning the phase noise, we have exactly the same performances of the analogic TCXO between 1 KHz and 100 KHz (fig. 8).

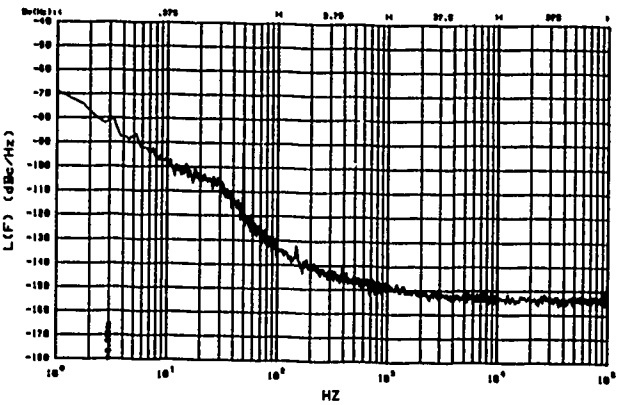


Fig. 8 - Phase noise measurement of DTCXO L(f)

4/ CONCLUSION

This data shows that CEPE has developed two new generations of TCXO. The new oscillators ATCXO and DTCXO have been in production since the beginning of 1989. For the ATCXO the pin out is compatible with the European Standard. The low current consumption (300 µA) of this ATCXO fully corresponds to the portable systems requirements.

Reference

1. G. CARET, E. GERARD, G. MAROTEL

"Oscillateur à quartz compensé en température numérique (PCTN)".

Revue Technique Thomson/csf
Vol. 18 - n° 2 - Juin 86

43rd Annual Symposium on Frequency Control - 1989

A TEMPERATURE-CONTROLLED CRYSTAL OSCILLATOR

Les Hurley
Piezo Crystal Company
100 K Street
Carlisle, Pa 17013

INTRODUCTION

The electronics industry continues to press for more and more performance from smaller and smaller packages. These pressures are felt more strongly, perhaps, in the modular-oscillator business than anywhere else. Piezo Crystal Company, like most others in our segment of the industry, has responded with a steady stream of innovative ideas, refined by state-of-the-art engineering practices. The final analysis is always a comprehensive evaluation by the market, measured by sales.

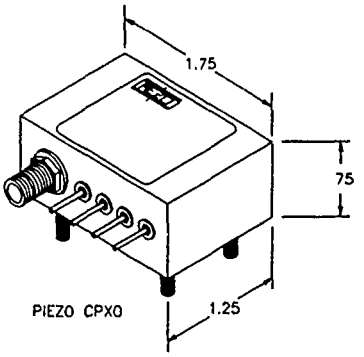
The latest contribution of Piezo Crystal Company to the advancing art and science of precision frequency generating is a co-planar temperature controlled crystal oscillator. The term, "ovenized" doesn't seem to quite fit. We have chosen to refer to it as a "CPXO". All of the oscillator and temperature-control electronics is contained on one double-sided Beryllium Oxide substrate. This construction allows a more accurate control of crystal temperature than the classical ovenized approach. At the same time the package size is reduced by eliminating the need for a separate "Oven". Smaller heated mass combined with greatly shortened thermal paths results in much shortened warmup time. Smaller surface area results in reduced heater power needed to sustain operating temperature. These enhancement features of the CPXO are obtained without sacrificing the desirable features of the much larger "OCXO".

PHYSICAL DIMENSIONS

This first generation Piezo CPXO, shown in Figure 1 and Figure 2, measures 1.75" long, 1.25" wide, and 0.75" high for a displacement of 1.64 cubic inches. Although our ultimate goal is a much smaller package than this, we chose it to accommodate our first effort internal construction.

Figure 1: Piezo CPXO Unit

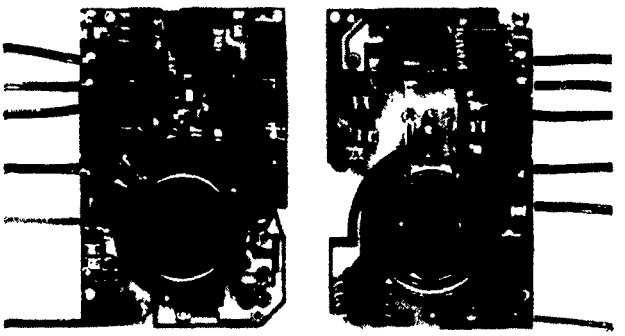
Figure 2: Piezo CPXO Diagram



INTERNAL DESIGN AND CONSTRUCTION

Ordinary thick film hybrid technology has been combined with the thermal properties of a Beryllium Oxide substrate to produce an ovenless temperature controlled crystal oscillator. Figure 3 shows the construction.

Figure 3:View of Both Sides of Assembled Substrate



The heater consists of two power FET's, each in series with a series current limiting resistor. These are arranged in a semi-circle around the crystal. Midway around the semi-circle, on the opposite side of the substrate, is a temperature sensing chip thermistor. After initial warmup, very little power is dissipated by the thick film resistors. During normal operation, nearly all of the temperature sustaining heat energy comes from the two power FET's. These may be considered to be point-sources for purposes of analysis. The flow of heat from the power FET's to the crystal is through short, wide paths in the BeO substrate. The crystal is mounted inside a close fitting copper ring which is soldered to substrate

CH2690-6/89/0000-055 $1.00 C 1989 IEEE

metalization. Void spaces between the crystal and the copper ring is filled with a thermally conductive RTV. The crystal leads are also soldered direct to substrate metalization. Because Beryllium Oxide has thermal properties nearly equal to those of metallic aluminum, and because the substrate area is small, the temperature gradients over the entire circuit are small.

Transient and steady state thermal analyses were performed on the CPXO using TAK-1, a computer software product of K&K Associates of Lyons, CO. to find the optimum location for the power FET's and to predict temperature drift versus ambient. Transient analysis was also used to predict control loop stability and to optimize loop gain. Excessive gain in the control loop results in oscillation of the substrate temperature around it's set point. The steady state analysis predicted thermistor temperature stability of about ± 0.1 degrees C. and crystal temperature remaining within about 0.1 degrees of the thermistor. Physical measurements confirmed the crystal-to-thermistor temperature gradient.

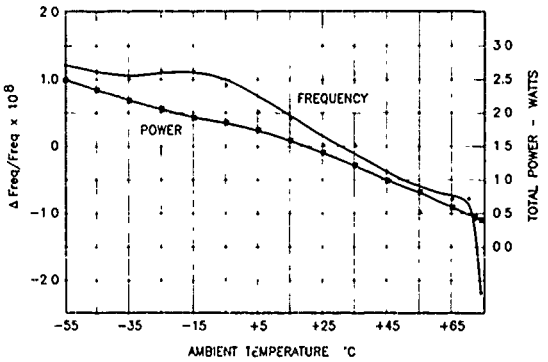
One problem, not revealed by these analyses is that the temperature control circuit over controls i.e., as ambient temperature is lowered the substrate temperature rises, and vice-versa. The effect is monotonic and nearly linear.

The crystal is a doubly-rotated stress-compensated unit made at Piezo. The oscillator is a Colpits modified for B-mode suppression of the SC cut crystal. There is also a grounded base buffer amplifier following the oscillator. Package insulation is Hydra-Matic PF-105 aircraft insulation.

PERFORMANCE

Frequency stability versus temperature: Figure 4 shows measured frequency versus ambient temperature. The frequency instability, though small, is due to instability of the crystal temperature. When the problem mentioned above is solved, the frequency curve of Figure 4 should improve significantly. The sixth order variation in the curve has not been investigated and is not considered significant at this time.

Figure 4: Frequency versus Temperature

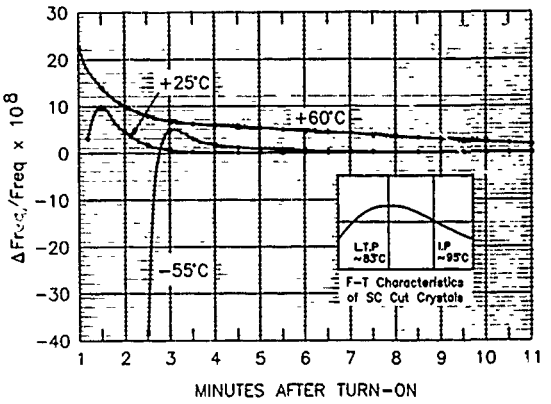


Heater sustaining power: Figure 4 also shows the variation of heater power versus ambient temperature.

Warmup time: Figure 5 shows frequency versus elapsed time after turn-on from "cold" start with ambient temperature as a parameter. In all cases the frequency rises rapidly after turnon. A graph of frequency versus temperature for an SC cut crystal is given as an inset to figure 5. As may be expected, the higher the ambient temperature, the shorter the length of time required for the frequency to pass through it's final value. However, there is in all cases a frequency overshoot. The figure shows that the frequency settling time is considerably longer than this.

Starting from room temperature, the CPXO warms up to within 3 parts in 10^8 of it's final value in about 2 1/4 minutes, and to within 1 part in 10^8 in about 2 3/4 minutes. Corresponding times for -55 degree C and +60 degree C ambient may be read from figure 5. A curve for -10 degree C ambient is too close to the 25 degree curve to show with clarity.

Figure 5: Piezo CPXO Warm-up Time



The frequency overshoot and exponential fall back were unexpected. Care was taken to set the steady state crystal temperature precisely to it's turnover point of about 83 degrees. If the **temperature** of the crystal is overshooting it's set point, one should expect the **frequency** to first rise tangentially to it's steady state value as the temperature goes through the turning point, fall off rapidly as the temperature continues to rise, and then to rise again asymptotically to it's steady state value as the excess heat energy drains off and the temperature falls again to it's final value. This obviously is not the performance observed in figure 5.

Several "explanations" have been offered to account for the observed performance:

1. The frequency overshoot is due to transient thermal gradients in the crystal - probably due to the rapid rate of heating during initial warmup and the difference of finite thermal resistances in the several heat flow paths.

2. The frequency overshoot is due to transient temperature differences between warmup and steady state conditions in the sustaining amplifier; particularly in a varactor used for frequency adjustment.

3. The steady state temperature is in reality set to some value below the crystal lower turning point.

While one or more of these factors may be

contributing to the observed performance, none of them satisfactorily account for the observed difference between the three curves of figure 5. In all of the above cases, the initial heating rate is the same for each ambient condition while this high rate is applied longer for low ambient temperatures than for high. One should expect steeper temporal gradients and possibly steeper spacial gradients for low ambient than for high. Therefore one should expect a greater effect from these factors at -55 degrees than at +60 degrees, for example. Obviously, the reverse of this is true.

Static phase noise: Figure 6 shows single-sideband random phase noise versus offset from the carrier.

Other performance characteristics: Figure 7 lists most of the other pertinent characteristics that could be specified for this oscillator.

Figure 6: Piezo CPXO Phase Noise Curve

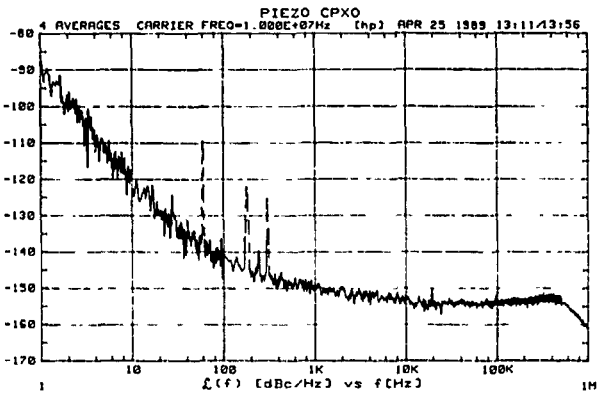


Figure 7: CPXO Specifications

Size: 1.75" x 1.25" x .75" (1.64 cu. in.)
Operating Temperature: –10°C to +62°C for full specified stability
Frequency Stability Over Temperature: $\pm 3 \times 10^{-6}$
Supply Voltage: +15 Vdc Oscillator;
+15 Vdc Heater
Warm-up Heater Current: 500 mA max.
Idling Heater Current at R.T.: 100 mA max.
Oscillator Current: 15 mA max.
Warm-up Time@25°C: 1 min to within 3×10^{-6}
15 min. to within 1×10^{-8}
Vibrational Sensitivity: 1×10^{-9} per g
Phase Noise Floor: –155 dBc max. @ 10 KHz
Aging: 5×10^{-10}/day, 1×10^{-7}/year

43rd Annual Symposium on Frequency Control - 1989

DESIGN AND PERFORMANCE OF VOLTAGE-CONTROLLED OSCILLATORS USING TFR STACKED-CRYSTAL FILTERS

P.H. Thompson, S.G. Burns, G.R. Kline, R.J. Weber
Microelectronics Research Center
1925 Scholl Road
Iowa State University
Ames, Iowa 50011

ABSTRACT

This paper presents two unique applications of thin-film resonator (TFR) technology. These designs include a 1 GHz VCO with a 0.5% tuning range and a frequency-agile 1 GHz oscillator which provides discrete frequency shifts of 4 MHz. The basic TFR frequency-control structure is a stacked-crystal filter (SCF) synthesized using multiple layers of AlN for piezoelectric transduction sandwiched between Al electrodes. Filters may be synthesized with single-mode bandpass or multiple (over-moded) bandpass characteristics. Typical performance for a single-mode double-stacked filter, when used in a 1 GHz VCO, provides an insertion loss as low as 1.5 dB, a linear phase extending over a 20 MHz passband, and an open-loop Q = 80.

The VCO was designed using a large-signal perturbational S-parameter technique. Measured SCF S-parameters and optimization using eeSOF's Touchstone circuit analysis routine were used to obtain a Butterworth VanDyke-based equivalent circuit embedded in an overall model incorporating package parasitics. VCO operation was obtained using a series connected hyperabrupt junction varactor diode. The use of this diode in series to provide the required variable loop phase does not significantly decrease Q_L and therefore does not degrade phase noise performance.

The frequency-agile oscillator employs an overmoded SCF filter structure which provides a rich comb filter band of high-Q (Q_{UL} > 2000) responses separated by 4 MHz in cascade with a second SCF to provide additional linear phase delay in the feedback path. Individual mode selection is accomplished by using a varactor diode in series with the second linear-phase SCF to provide additional voltage-tuneable linear phase.

Supporting data provided includes S-parameter filter characterization and resultant models, VCO tuning curves, and phase-noise measurements. Open-loop flicker noise measurements using an HP 3047 were conducted on the filters and amplifiers and these data were substituted into Leeson's model and there was close agreement with closed-loop slope.

Specifically, the VCO exhibited -70dBc/Hz at 1 kHz offset with an f^{-3} behavior. The frequency-agile oscillator exhibited > -100dBc/Hz at 1 kHz offset, again with f^{-3} behavior.

These topologies utilizing the TFR have the unique advantage of potential monolithic realization on the same substrate as used for the active device circuitry.

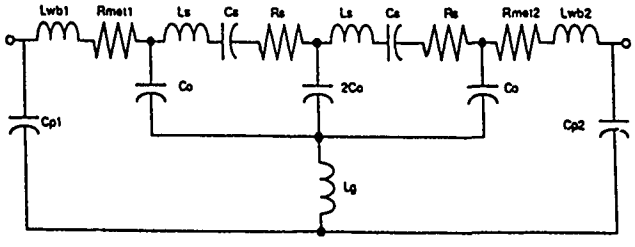
BACKGROUND

Bulk-acoustic-wave thin-film resonators (TFR), TFR-based stacked crystal and monolithic filters and TFR transduced semiconductor delay lines have all been used in single-line spectra or combline UHF and L-band oscillators.[1,2,3,4] We now extend this work by reporting on the design, realization, and performance of two types of VCOs using the stacked-crystal filter (SCF)[5,6] in the feedback loop. The first of these is a 0.5%, 1 GHz VCO. Its design has been extended to include demonstrated applications of direct FM and phase-locking at fundamental frequencies over 1 GHz. The second type of 1 GHz oscillator uses an overmoded filter structure, essentially a filter bank with 4 MHz mode spacing, to allow selectable tuning between modes.

METHOD

The VCO design includes a forward gain element, stripline matching sections, the SCF in the feedback loop, and a voltage-variable phase shift provided by a hyperabrupt junction varactor diode.

The SCF is synthesized using multiple layers of AlN, for piezoelectric transduction, sandwiched between Al electrodes. Although the SCFs used in this work are fabricated on a Si substrate, similar topology devices have also been built on GaAs.[7] Insertion loss for a double-stacked crystal filter (DSCF) at 1.1 GHz is as low as 1.5 dB. The linear phase characteristic extends over a 20 MHz passband with an open-loop Q = 80. The SCF is packaged in a ceramic chip carrier. Oscillator design requires measuring S-parameters of the SCF as a function of frequency in and near the passband and converting these data to a Butterworth VanDyke equivalent circuit, derived from the Mason model, as shown in Figure 1. The SCF dimensions exhibit a large lateral dimension to thickness ratio, consequently we use a one-dimensional plane wave analysis.



L_{wb1} = 1.712μH L_{s1} = 290.99μH L_G = 0.0377μH
L_{wb2} = 2.085μH L_{s2} = 291.71μH C_{p1} = 0.0507pF
R_{met1} = 2.689Ω R_{s1} = 12.22Ω C_{p2} = 0.0509pF
R_{met2} = 2.691Ω R_{s2} = 12.15Ω
C_s = 0.08316pF C_o = 3.8376pF

Figure 1 DSCF Butterworth VanDyke equivalent circuit with packaging parasitics. Typical values used for Touchstone and SPICE simulations are included.

The interdependent model parameters are computed using

$$C_s = C_o/(2k/\pi)^2 \quad (1)$$

$$R_s = 1/\{\omega_p C_o (2k/\pi)^2\} \quad (2)$$

$$L_s = \frac{1}{\omega_s^2 C_s} \quad (3)$$

CH2690-6/89/0000-058 $1.00 C 1989 IEEE

where

C_O = the capacitance across the electrodes determined from device electrode geometry and AlN dielectric properties
R_S = non-ohmic loss
C_S = series branch capacitance dependent upon k, coupling (stiffness) coefficient, and C_O
L_S = series inductance needed for series resonance, f_S, with C_S.

In addition, package parasitics represented by R_{MET}, L_{WB}, and C_P are extracted from the data set and included as shown in Figure 1. The equivalent circuit parameters depend strongly on k and Q. These values are difficult to measure directly when TFRs are coupled in an SCF configuration. It is also likely, that k will differ somewhat between film layers. Model parameters are obtained by varying k and Q values using eeSOF's Touchstone circuit analysis and optimization routines to provide a best fit to measured data.

The forward gain element consists of a commercial hybrid broadband, 2 GHz, 10 dB gain, impedance-matched amplifier. A perturbational method of measuring large-signal full-matrix S-parameters was used to characterize the amplifier operating at the compression level present in the closed-loop configuration. The large-signal S-parameters for the amplifier were combined with linear S-parameters for the remainder of the oscillator circuit and the net two-port S-parameters were computed using Touchstone. The net two-port oscillator is shown in Figure 2.

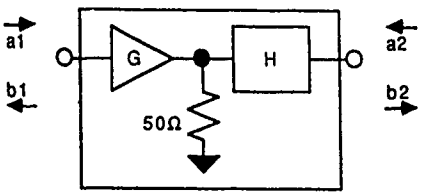


Figure 2 Open loop oscillator used for S-parameter where G is the Amplifier gain element and H represents composite of feedback network elements.

For the circuit, we require

$$a_1 = b_2 \quad (4a)$$

$$b_1 = a_2 \quad (4b)$$

$$1 = \Delta S + S_{12} + S_{21} \quad (4c)$$

$$\Delta S = S_{11} S_{22} - S_{12} S_{21} \quad (4d)$$

Equations 4a and 4b are the conditions necessary to sustain oscillation upon closing the loop. Equations 4c and 4d are derived from these conditions and are used to determine if the loop will oscillate at the predicted frequency upon closing the loop.

Since we have a discrete equivalent circuit for the SCF and can obtain an amplifier schematic diagram with active device model parameters the frequency at onset of oscillation can also be predicted using a closed-loop SPICE circuit simulation and observing the frequency at which the Barkhausen magnitude ($|G|\,|H| = 1$) and phase ($\phi_H + \phi_G = 0°$) criteria are satisfied. A swept-frequency series-connected signal is used for excitation to simulate a noise generator.

VCO operation is achieved by adjusting the reverse bias voltage of a hyperabrupt junction diode which is placed in series with the frequency selective feedback network. For a single diode, a 0 to 20 volt change in reverse bias resulted in a 32° change in open-loop phase.

The selectable-mode oscillator employs an overmoded SCF in cascade with a single-mode conventional SCF in the feedback network as shown in Figure 3. This configuration allows for a single filter mode of the overmoded SCF to be selected within the passband of the delay SCF. In particular, the 3 dB bandwidth of the single-mode SCF envelopes four of the narrow bandwidth modes. Figure 4 illustrates the amplitude and linear phase response of the cascaded filters. Frequency selections are accomplished using an adjustable line stretcher for coarse loop-phase adjustments and a series varactor for fine tuning.

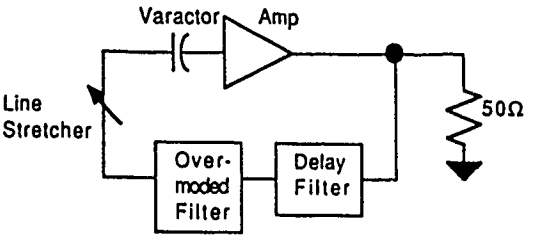


Figure 3 Overmoded SCF based oscillator with mode selection.

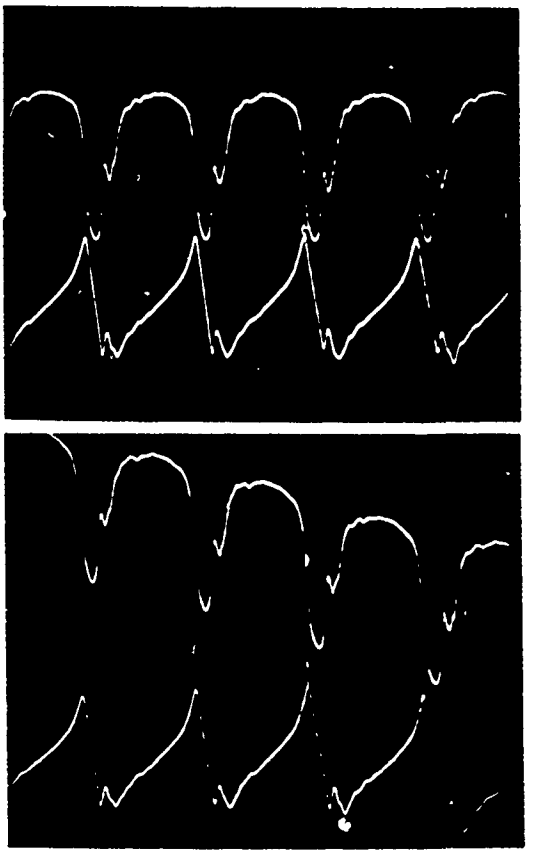


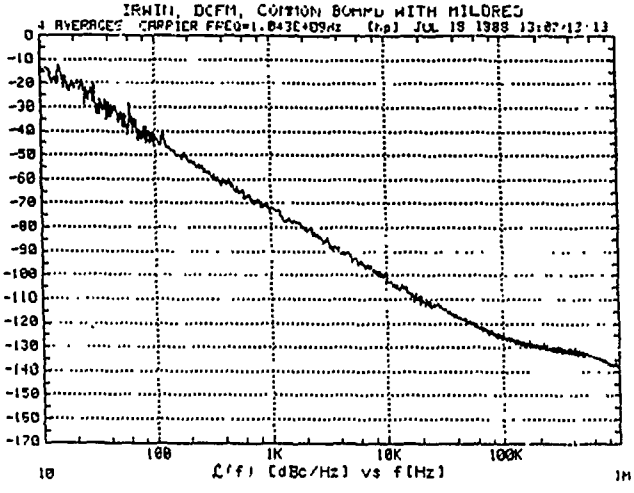
Figure 4 Overmoded SCF in Cascade with a single-mode SCF. f_c(marker) = 1034.6 MHz ± 10 MHz.
Top: Composite phase characteristic showing overall linear phase characteristic, 45°/div.
Bottom: Composite amplitude characteristic, 10 dB/div.

PERFORMANCE

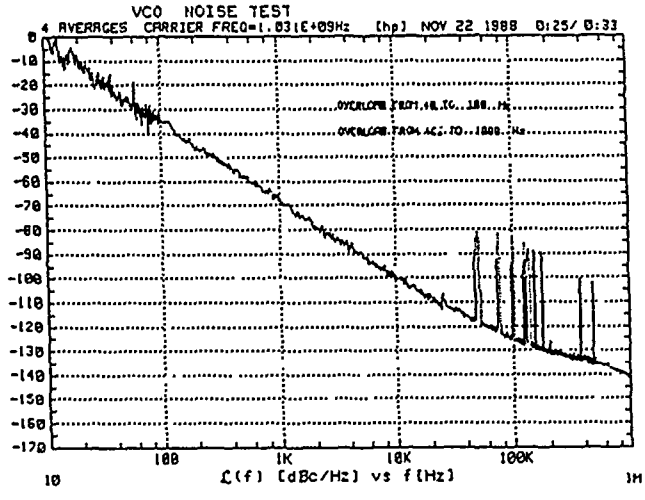
Operation of an SCF-controlled oscillator as a VCO or TCXO requires that the total phase shift, ϕ_H, in the feedback path be altered. For low noise operation, ϕ_H must change to a different operating point without changing the composite feedback network slope, $d\phi_H/df$. In simple resonator feedback circuits, the oscillator can be "pulled" by adding a parallel capacitance but the Q related to the resonator phase slope by

$$Q = \frac{f_r}{2} \frac{d\phi_H}{df} \Bigg|_{f_r} \qquad (5)$$

is a rapid function of frequency and consequently noise performance will degrade. However, the inclusion of the varactor diode in series with the VCO feedback path does not significantly effect the SCF Q with the result that there is no noticeable degradation of the spectral purity of the VCO as compared with earlier fixed frequency SCF-based oscillators. This is illustrated in Figures 5a and 5b which are plots of the SSB phase noise for a typical SCF-based oscillator and the varactor tuned VCO respectively.



(a) Fixed Frequency SCF-Controlled Oscillator



(b) VCO, Varactor-Tuned, SCF-Controlled Oscillator

Figure 5 SSB phase noise comparison.

The phase noise was dominated by an f^{-3} slope with a nominal value of -70 dBc/Hz at 1 kHz offset. Open-loop phase-noise measurements using the SCF were conducted to obtain an empirical flicker noise coefficient, $\alpha = 1.6 \times 10^{-12}$ sec^{-1}. The SCF flicker noise contribution was considered dominant since silicon BJTs were used in the amplifier. The open-loop measurements yielded values from -115 dBc/Hz to -128 dBc/Hz extrapolated to a 1 Hz offset. A typical open-loop phase-noise plot for an SCF is shown in Figure 6.

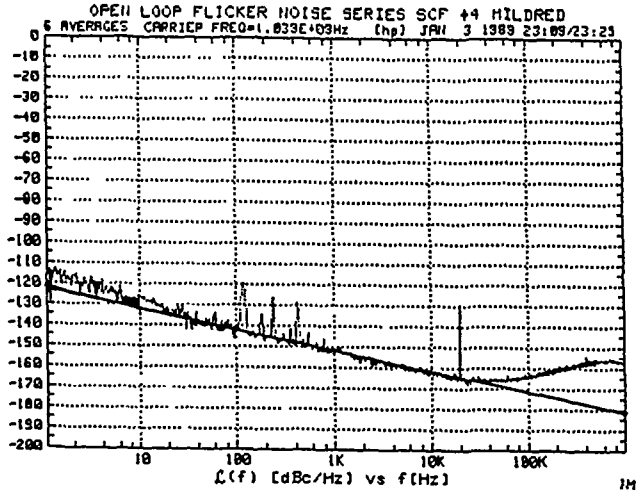


Figure 6 Typical SCF flicker noise characteristic used to obtain α.

Observing that the varactor did not significantly degrade the phase noise allows for an indirect determination of the closed-loop phase slope by plotting the VCO phase slope shown in Figure 7 and computing a value of 0.875 MHz/volt in the linear region.

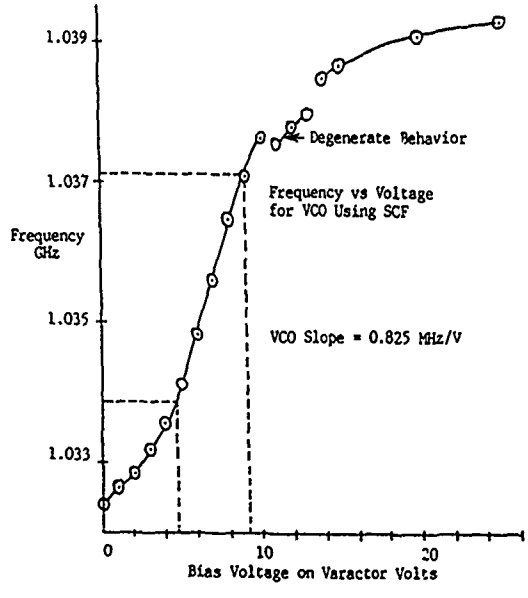


Figure 7 VCO tuning characteristic in series with a single-mode SCF.

The linear portion of the varactor's phase vs. voltage slope was selected to coincide with the SCF phase slope. Equation 6 was then used to estimate the closed-loop phase slope. Equation 5 was used to estimate the loaded Q of the oscillator to be 35.8.

$$\frac{d\phi_H}{df} = \left(\frac{d\phi}{dv}\right)\left(\frac{dv}{df}\right) \qquad (6)$$

The single-sideband phase-noise in dBc/Hz can be described by[8]

$$\frac{P_{sb}}{P_c} = 10 \log \left[N^2\left(\frac{\alpha}{\omega_m{}^3\tau^2} + \frac{GFkT}{P_c\omega_m{}^2\tau^2} + \frac{\alpha}{\omega_m} + \frac{GFkT}{P_c}\right)\right] \qquad (7)$$

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

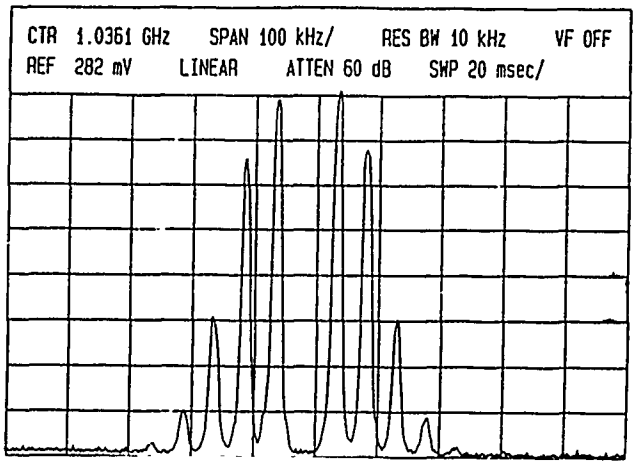
$\omega_m = 2\pi f_m$ = Fourier modulation frequency, frequency offset from carrier frequency

kT = $(1.38\times10^{-23}$ J/°K)(300°K) = 4.14×10^{-21} Joules = 4.14×10^{-21} watts/Hz

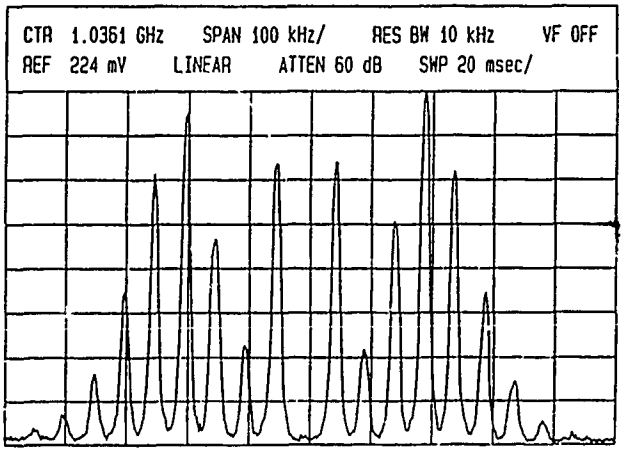
α = flicker noise constant in sec^{-1}

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.

Using $\alpha = 1.6 \times 10^{-12}$ sec^{-1} and Q_L = 35.8 in Equation 7, Leeson's equation predicts a dominant f^{-3} behavior



(a) Modulation Index x = 2.4, $J_0(2.4) = 0$, $0 \leq n \leq 9$

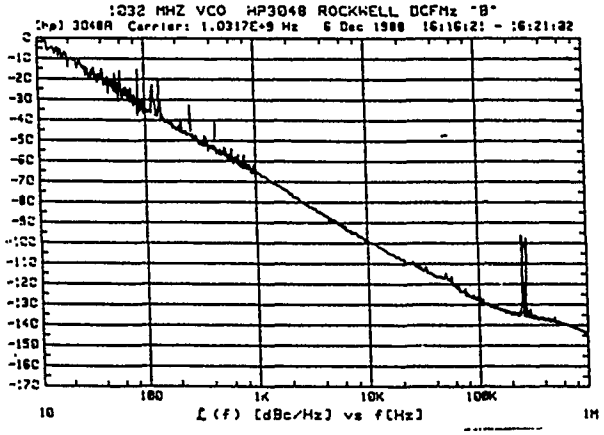


(b) Modulatio Index x = 5.5, $J_0(5.5) = 0$, $0 \leq n \leq 9$

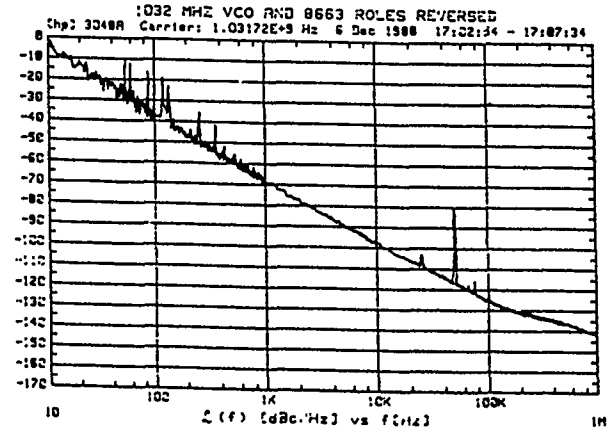
Figure 8 Direct FM at 1.03GHz spectral plots with f_m = 50 kHz illustrating agreement with $J_n(x)$ theoretical analysis.

with a value of -73.8 dBc/Hz which agrees well with measured values.

Direct FM at > 1 GHz was implemented by modifying the varactor bias network to incorporate a modulating signal centered along the linear portion of the tuning curve. Resultant spectral plots for $J_0(2.4) = J_0(5.5) = 0$ at f_o = 1.03 GHz with f_m = 50 kHz are shown in Figures 8a and 8b. The VCO was then incorporated in a PLL by phase locking to a more stable source. This was demonstrated by interchanging the role of the reference oscillator (normally the HP 8663) and the SCF-controlled VCO when performing SSB phase-noise measurements. Essentially identical results were obtained when observing the SSB phase noise of the unlocked SCF VCO and the SCF VCO serving as a reference for the HP 8663 as shown in Figure 9.



(a) HP 3048 Using HP 8663 Synthesizer Measurement of SCF = VCO



(b) SCF-VCO Using HP 3048 System Voltage-Control VCO. Roles of VCO and Hp 8663 Synthesizer Reversed From (a).

Figure 9 Phase locking performance comparison.

The mode selectable oscillator was mechanically tuned with a line stretcher and electrically tuned with a series varactor diode. Thus, there were three elements contributing to phase delay in the feedback network, the overmoded SCF operating as a comb frequency filter, a DSCF whose bandpass encompassed at least 4 modes, and the series varactor used to select between one of the four modes. Figure 4 illustrates the overall phase shift provided by the cascaded filters and how the varactor can then select the appropriate phase condition to satisfy the Barkhausen criteria for any of the modes within the somewhat limited tuning range of the varactor. The

SSB phase noise performance is illustrated in Figure 10.

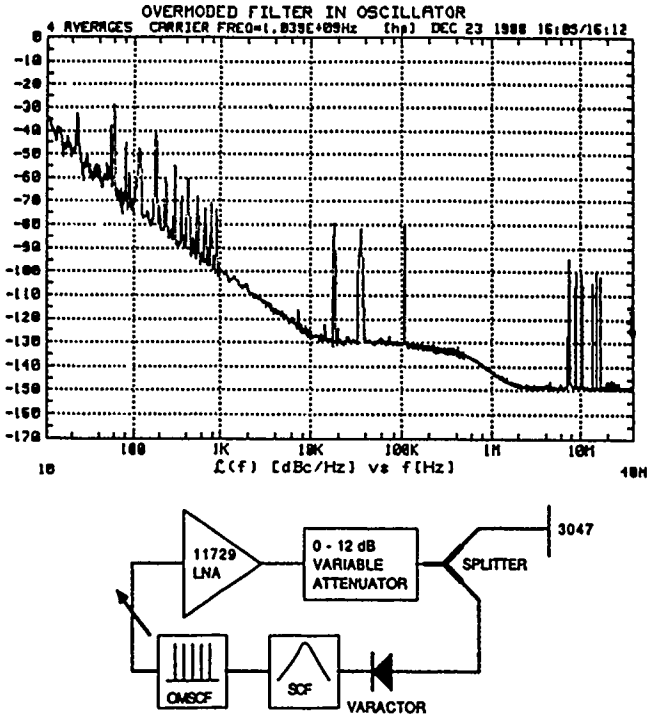


Figure 10 Mode-selectable SCF oscillator test configuration and SSB phase noise performance.

The varactor was also used to indirectly measure the closed-loop phase slope and Q_L = 3025 for the individual modes of the overmoded SCF. In addition, the oscillator could be phase locked within a 50 kHz range at any of the four modes.

DIRECTIONS OF FUTURE WORK

The single-mode VCO performance, including FM operation and the operation in a phase-locked loop, show promise as not only a stand-alone circuit but as a component in communications systems architectures. The oscillator circuit using an overmoded SCF offers an interesting alternative to synthesized frequency hopping signal generation. Since mode tuning occurs within a single feedback loop, spurious responses associated with frequency synthesis techniques are non-existent. The response time would be a function of delay circuitry and not be limited by external loop bandwidths. The design presented used a single section overmoded filter. Multiple section filters with steeper phase slopes could be used to achieve a higher degree of spectral purity. Work continues on reducing SCF insertion loss and the open loop characterization.

Although both these circuits were constructed using external circuitry and matching networks, the TFR has the unique advantage over other technologies of monolithic realization on the same substrate as active elements. A single-mode TFR-controlled UHF oscillator has been reported.[9] It is expected that these hybrid configurations will be fabricated on a single die leading to fully integrated communications systems.

This work was conducted at and supported by the Iowa State University Microelectronics Research Center.

REFERENCES

1. R.A. Moore, J.T. Haynes, B.R. McAvoy, "High Overtone Bulk Resonator Stabilized Microwave Sources," Proceedings of the 1981 IEEE Ultrasonics Symposium, pp. 414-424.

2. S.G. Burns, G.R. Kline, K.M. Lakin, "UHF Oscillator Performance Using Thin Film Resonator Based Topologies," Proceedings of the IEEE 41st Annual Symposium on Frequency Control, May 1987, pp. 382-387.

3. M.M. Driscoll, S.V. Krishnaswamy, R.A. Moore, 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.

4. S.G. Burns, G.R. Kline, K.M. Lakin, "Design and Performance of Oscillators Using Semiconductor Delay Lines," IEEE Proceedings of the 1987 Ultrasonics Symposium.

5. A. Ballato, T. Lukaszek, "Stacked Crystal Filters," Proceedings IEEE, Vol. 51, Oct. 1973, pp. 1495-1496.

6. G.R. Kline, R.S. Ketcham, K.M. Lakin, "Low Insertion Loss Filters Synthesized With Thin Film Resonators," Proceedings of the 1987 Ultrasonics Symposium.

7. G.R. Kline, R.S. Ketcham, K.M. Lakin, "Thin Film Microwave Acoustic Filters on GaAs," 1988 IEEE GaAs IC Symposium.

8. E.A. Gerber and A. Ballato, Editors, Precision Frequency Control: Oscillators and Standards, Vol. 1, Academic Press, 1985.

9. W. A. Burkland, A.R. Landin, G.R. Kline, R.S. Ketcham, "A Thin-Film Bulk-Acoustic-Wave Resonator-Controlled Oscillator on Silicon," IEEE Electron Device Letters, Nov. 1987.

43rd Annual Symposium on Frequency Control - 1989

NON-LINEAR ANALYSIS OF HIGH Q OSCILLATOR PHASE NOISE

by Jules Goldberg--Silicon Valley Engineering
2129 Hoover Court Santa Clara, Ca., 95051 (408)296-8872

1.0 Abstract

This paper uses non-linear analysis to translate between circuit noise and high Q oscillator phase noise. The osci-llator[6] concept is used to separate the resonant circuit from the rest of the system. Analysis is based on the llator characteristics.

It is shown that:

o The non-linearities interact by mixing all noise frequencies near an integral multiple of the oscillation frequency (including zero and 1) into the oscillator pass band. (This is mentioned in [1] and [2].) Since all such components are additive, AGC controlled linear oscillators are needed in critical applications.[2,3,5].

o Noise far from an integral multiple of the oscillation frequency produces the "noise floor" effect shown in [1] and [5].

o The final result reduces to Leeson's model[1] for zero non-linearity.

The results of AGC are also analyzed and it is shown how our formulas relate to the "noise amplifier" model discussed in [10], [11], and [2]. If AGC is not practical, this paper can be used to estimate the contribution of each circuit component to the total oscillator noise.

In the past, non-linear analysis was either neglected [2,4,10,11]; rejected as being too complicated for practical use[3]; or limited to a specific differential equation. [6,7,8] This paper applies to any crystal or high Q tuned circuit oscillator regardless of circuit configuration. Fourier analysis is used to focus on one noise frequency and then integrate the result over the entire spectrum. (as is done with the inverse Fourier transform[9]) Euler's identity for sinusoids simplifies the mathematics.

Because the noise is small compared to the oscillation, all equations are linearized. Finally, the relationship between circuit noise, phase noise and amplitude noise is derived without using Bessel functions.

An analysis program such as Spice is used to observe the "hidden non-linearities" of a particular llator. The results are then plugged into the theoretical formulas to create a noise spectrum. Since the actual computations are complicated, it is recommended that the entire procedure be computerized.

A numerical example for the distortion limited llator is included for clarity and completeness. The AGC case is discussed in detail.

Outline

2.0 General Theory of Non-Linear N-ports

Consider a complicated non-linear N-port as shown in figure 1, where a large

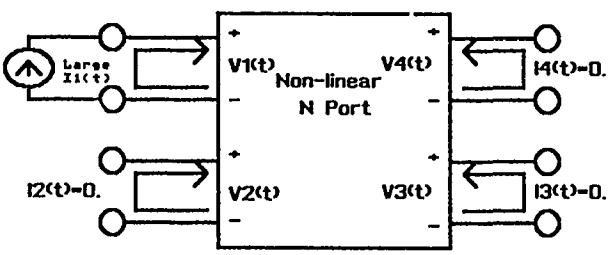


Figure 1 Co.nplicated Non-Linear N port driven by a large Current at port 1

signal current is flowing through port 1 and all other ports are open circuited. (Similar to a noiseless oscillator)

This network is assumed to be asymptotically stable and quiescent if all ports are open. We may not be able to solve for the port voltages in terms of the port currents, however, we assume all voltages and currents are related by a continuous, differentiable frequency dependent function. This is expressed by (1).

$$G\left\{\ldots,\frac{d^{-2}I1}{dt^{-2}},\frac{d^{-1}I1}{dt^{-1}},I1,\frac{dI1}{dt},\frac{d^{2}I1}{dt^{2}},\frac{d^{-2}I2}{dt^{-2}},\frac{d^{-1}I2}{dt^{-1}},I2,\frac{dI2}{dt},\frac{d^{2}I2}{dt^{2}},\ldots,\ldots,\frac{d^{-2}I3}{dt^{-2}},\frac{d^{-1}I3}{dt^{-1}},I3,\frac{dI3}{dt},\frac{d^{2}I3}{dt^{2}},\frac{d^{-2}I4}{dt^{-2}},\frac{d^{-1}I4}{dt^{-1}},I4,\frac{dI4}{dt},\frac{d^{2}I4}{dt^{2}},\ldots,\ldots,I5..,I6..,I7.....ect.\ldots,\frac{d^{-2}V1}{dt^{-2}},\frac{d^{-1}V1}{dt^{-1}},V1,\frac{dV1}{dt},\frac{d^{2}V1}{dt^{2}},\frac{d^{-2}V2}{dt^{-2}},\frac{d^{-1}V2}{dt^{-1}},V2,\frac{dV2}{dt},\frac{d^{2}V2}{dt^{2}},\ldots,\ldots,\frac{d^{-2}V3}{dt^{-2}},\frac{d^{-1}V3}{dt^{-1}},V3,\frac{dV3}{dt},\frac{d^{2}V3}{dt^{2}},\frac{d^{-2}V4}{dt^{-2}},\frac{d^{-1}V4}{dt^{-1}},V4,\frac{dV4}{dt},\frac{d^{2}V4}{dt^{2}},\ldots,\ldots,V5..,V6..,V7.....ect\right\}\equiv G_o=0.,\quad (1)$$

where the negative order derivatives represent time integrals from 0 to t, and G_o is used to represent the noiseless case. (1) is analogous to a Z parameter equation at port 1. The function equals 0. to relate the voltages and currents to each other. It is not an identity.

2.1 Superposition of Small and Large Signals

If all ports are driven by a small signal current disturbance in addition to the main source (see figure 2), a new relationship among the port currents and voltages follows from (1). If the signals are small enough to not generate there own harmonics, then their effects are linear, as described by (2).

A look at (2) shows that the required partial derivatives cannot be calculated unless the circuit details are known. This is because the function G always equals zero in a physical circuit. Later on in this paper, the need for these derivatives cancels out. Since G_o=0., (3) is the final form of (2).

The partial derivatives in (3) are functions of time, unless the N-port is linear.

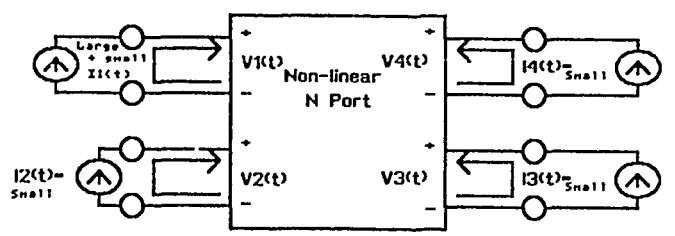


Figure 2 Complicated Non-Linear N port driven by a large Current at port 1 plus additional small currents at all ports.

$$G\left\{\ldots,\frac{d^{-2}(I1+\Delta I1)}{dt^{-2}},\frac{d^{-1}(I1+\Delta I1)}{dt^{-1}},I1+\Delta I1,\frac{d(I1+\Delta I1)}{dt},\frac{d^{2}(I1+\Delta I1)}{dt^{2}},\ldots,\ldots,\frac{d^{-2}(I2+\Delta I2)}{dt^{-2}},\frac{d^{-1}(I2+\Delta I2)}{dt^{-1}},I2+\Delta I2,\frac{d(I2+\Delta I2)}{dt},\frac{d^{2}(I2+\Delta I2)}{dt^{2}},\ldots,I3+\Delta I3,\ldots,I4+\Delta I4,\ldots,I5+\Delta I5,\ldots.ect.,\ldots,\frac{d^{-2}(V1+\Delta V1)}{dt^{-2}},\frac{d^{-1}(V1+\Delta V1)}{dt^{-1}},V1+\Delta V1,\frac{d(V1+\Delta V1)}{dt},\frac{d^{2}(V1+\Delta V1)}{dt^{2}},\ldots,\ldots,\frac{d^{-2}(V2+\Delta V2)}{dt^{-2}},\frac{d^{-1}(V2+\Delta V2)}{dt^{-1}},V2+\Delta V2,\frac{d(V2+\Delta V2)}{dt},\frac{d^{2}(V2+\Delta V2)}{dt^{2}},\ldots,V3+\Delta V3,\ldots,V4+\Delta V4,\ldots,V5+\Delta V5,\ldots.ect.,\right\}=0.$$

$$=G_o+\sum_k\frac{\partial G}{\partial\left(\frac{d^k V1}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{are unchanged.}}}\times\left(\frac{d^k\Delta V1}{dt^k}\right)$$

$$+\sum_i\sum_k\frac{\partial G}{\partial\left(\frac{d^k I_i}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{except that of port i}\\ \text{are unchanged.}\\ \text{All voltages except}\\ \text{that of port i vary in}\\ \text{response to the disturb}\\ \text{at port i. V1 is unchanged.}}}\times\left(\frac{d^k\Delta I_i}{dt^k}\right)\quad (2)$$

$$\sum_k\frac{\partial G}{\partial\left(\frac{d^k V1}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{are unchanged.}}}\times\left(\frac{d^k\Delta V1}{dt^k}\right)+\quad (3)$$

$$\sum_i\sum_k\frac{\partial G}{\partial\left(\frac{d^k I_i}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{except that of port i}\\ \text{are unchanged.}\\ \text{All voltages except}\\ \text{that of port i vary in}\\ \text{response to the disturb}\\ \text{at port i. V1 is unchanged.}}}\times\left(\frac{d^k\Delta I_i}{dt^k}\right)=0.$$

2.2 Case of a Periodic Large Signal at Port 1

If the N-port is a llator, then its response to a periodic I1 should have the same period as I1 and contain no sub-harmonic or non-periodic components. Unfortunately, this must be guaranteed by design. The engineer must watch out for the effects of:

a)Non-linear Components with Hysteresis

b)Digital logic that causes crosstalk or severe VCC and ground spiking in response to the llator's output.

This paper assumes that the above effects are small enough to generate no additional harmonics and that they can be lumped with the disturbances of figure 2. The partial derivatives within (3) can therefore be expressed as Fourier series.

$$\frac{\partial G}{\partial\left(\frac{d^k V1}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{are unchanged.}}}=\sum_{l=-\infty}^{\infty}h_{l,k,1}e^{jlw_o t}$$

$$\frac{\partial G}{\partial\left(\frac{d^k I_i}{dt^k}\right)}\Bigg|_{\substack{\text{All port currents}\\ \text{except that of port i}\\ \text{are unchanged.}\\ \text{All voltages except}\\ \text{that of port i vary in}\\ \text{response to the disturb}\\ \text{at port i. V1 is unchanged}}}=\sum_{l=-\infty}^{\infty}s_{i,k,l}e^{jlw_o t},\quad (4)$$

where

i = The port number
k = The derivative order
l = The Fourier Harmonic Number
w_o = The fundamental large signal radian frequency of I1.

The Fourier coefficients in (4) are difficult if not impossible to calculate.

(3) is a complicated integro-differential equation like that explored by Hafner.[10] For this paper, the forcing function in (3) is current.

Using the Fourier transform[9], it is true that

$$I(t) = \frac{\int_{w=-\infty}^{\infty} I_f(jw) e^{jwt} dw}{2\pi j}, \quad (5)$$

where $I_f(jw)$ is the Fourier transform of I(t).

2.3 Solution of (3) for a Small Signal Sinusoidal Forcing Function

Theoretically, I(t) can be broken up into an infinite number of forcing functions of the form

$$\frac{I_f(jw) e^{jwt} dw}{2\pi j} \quad (6)$$

Since the the circuit is linear for small signals, the network can be solved for the driving function $I=e^{jwt}$ and then integrated over all jw as is done in (5) to get the full response.

Plugging (4) into (3) produces

$$\sum_k \frac{d^k \Delta v_1}{dt^k} \sum_{l=-\infty}^{\infty} h_{1,k,l} e^{jlw_0 t} + \sum_i \sum_k \frac{d^k \Delta i_i}{dt^k} \sum_{l=-\infty}^{\infty} s_{i,k,l} e^{jlw_0 t} = 0, \quad (7)$$

where

i = The port number
k = The derivative order
l = The Fourier Harmonic Number
w_o = The large signal radian frequency of I1.

Let us assume that all ports are driven by a small signal current $I_i e^{jwt}$ and that and that we are trying to find V1(t). (7) becomes

$$\sum_k \frac{d^k \Delta v_1}{dt^k} \sum_{l=-\infty}^{\infty} h_{1,k,l} e^{jlw_0 t} + \sum_i I_i \sum_k (jw)^k e^{jwt} \sum_{l=-\infty}^{\infty} s_{i,k,l} e^{jlw_0 t} = 0. \quad (8)$$

It is proposed that V1 be of the form

$$V1(t) = e^{jwt} \sum_{m=-\infty}^{\infty} V_m(w) e^{jmw_0 t}. \quad (9)$$

where
m is a Fourier harmonic number

w_o = the fundamental large signal radian frequency of I1

w = the radian frequency of the disturbance

The response to $I_i e^{-jwt}$ is the complex conjugate of (9).

$$V1^*(t) = e^{-jwt} \sum_{m=-\infty}^{\infty} V_m^*(w) e^{-jmw_0 t} = e^{-jwt} \sum_{m=-\infty}^{\infty} V_{-m}(-w) e^{-jmw_0 t} \quad (10)$$

If the $h_{1,k,l}$ and $s_{i,k,l}$ are known, plugging (9) into (8) produces

$$\sum_k e^{jwt} \sum_{m=-\infty}^{\infty} V_m(w) (jw+mw_0)^k e^{jmw_0 t} \sum_{l=-\infty}^{\infty} h_{1,k,l} e^{jlw_0 t} + \sum_i I_i \sum_k (jw)^k e^{jwt} \sum_{l=-\infty}^{\infty} s_{i,k,l} e^{jlw_0 t} = 0. \quad (11)$$

After rearrangement, (11) becomes

$$e^{jwt} \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} e^{j(l+m)w_0 t} V_m(w) \sum_k h_{1,k,l} (jw+mw_0)^k + \sum_i I_i e^{jwt} \sum_{l=-\infty}^{\infty} e^{jlw_0 t} \sum_k s_{i,k,l} (jw)^k =$$

(Letting c=l+m)

$$e^{jwt} \sum_{c=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} e^{jcw_0 t} V_m(w) \sum_k h_{1,k,(c-m)} (jw+mw_0)^k + \sum_i I_i e^{jwt} \sum_{l=-\infty}^{\infty} e^{jlw_0 t} \sum_k s_{i,k,l} (jw)^k = 0. \quad (12)$$

Replacing the symbol c by l plus a small re-arrangement produces (13).

$$e^{jwt} \sum_{l=-\infty}^{\infty} e^{jlw_0 t} \sum_{m=-\infty}^{\infty} V_m(w) \sum_k h_{1,k,(l-m)} (jw+mw_0)^k + \sum_i I_i e^{jwt} \sum_{l=-\infty}^{\infty} e^{jlw_0 t} \sum_k s_{i,k,l} (jw)^k = 0. \quad (13)$$

(13) represents a system of equations in $V_m(w)$. The $V_m(w)$'s can be measured on a wave analyzer or with a Spice Fourier analysis. Spice is most practical when $w=m_o w_o$. (m_o is an integer.)

3.0 Application to a High Q Oscillator

3.1 Interaction of the Non-Linear N-Port and a Linear 1 Port.

An oscillator is shown in figure 3.

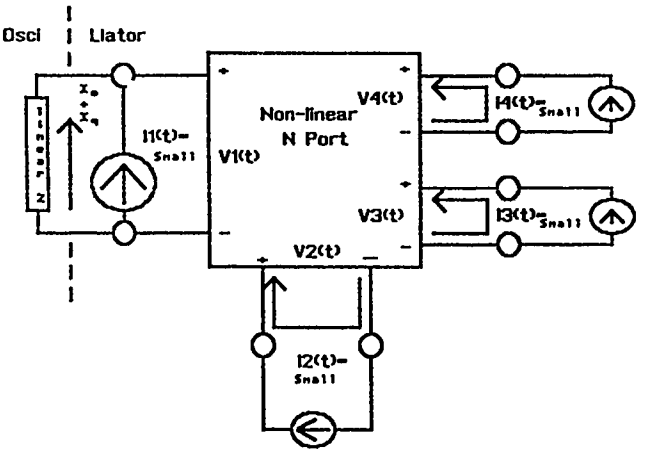


Figure 3 High Q Oscillator with Extraneous Current Inputs.

I_o = the large signal oscillation current
I_1-I_n = small signal noise stimuli
I_q = The deviation in Oscillation current due to I_1-I_n

Applying (7) to figure 3 produces (14)

$$\sum_k \frac{d^k \Delta v_1}{dt^k} \sum_{i=-\infty}^{\infty} h_{1,k,i} e^{jiw_o t} + \sum_k \frac{d^k \Delta i_n}{dt^k} \sum_{i=-\infty}^{\infty} s_{1,k,i} e^{jiw_o t} + \sum_l \sum_k \frac{d^k \Delta i_l}{dt^k} \sum_{i=-\infty}^{\infty} s_{1,k,i} e^{jiw_o t} = 0. \quad (14)$$

and the linear Z requires that for any w

$$V1(jw) e^{jwt} + Z(jw) I_m^q(w) e^{jwt} = 0. \quad (15)$$

Following the procedure in sect. 2.3, assume that

$$I_l(t) = I_l e^{jwt}.$$
$$V1(t) = e^{jwt} \sum_{m=-\infty}^{\infty} V_m(w)\, e^{jmw_o t}$$
and
$$I_q(t) = e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w)\, e^{jmw_o t} \quad (16)$$

Plugging (16) into (15), rearranging and taking advantage of the orthogonality of the functions $e^{jmw_o t}$ produces

$$e^{jwt} \sum_{m=-\infty}^{\infty} V_m(w)\, e^{jmw_o t} + e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w) Z(j(w+mw_o))\, e^{jmw_o t} = 0. \text{ or}$$
$$V_m(w) = -I_m^q(w) Z(j(w+mw_o)) \quad (17)$$

Plug (16) and (17) into (14) to get

$$-\sum_k e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w) Z(j(w+mw_o)) (j(w+mw_o))^k e^{jmw_o t} \sum_{i=-\infty}^{\infty} h_{1,k,i} e^{jiw_o t}$$

$$+\sum_k e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w) (j(w+mw_o))^k e^{jmw_o t} \sum_{i=-\infty}^{\infty} s_{1,k,i} e^{jiw_o t}$$

$$+\sum_l \sum_k (jw)^k e^{jwt} \sum_{i=-\infty}^{\infty} s_{i,k,l} e^{jiw_o t} = 0. \quad (18)$$

The $I_m^q(w)$'s are the unknowns in (18). Process and rearrange (18) the way it was done with (11)-(13) in sect. 2.3 to get

$$-e^{jwt} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_{m=-\infty}^{\infty} I_m^q(w) Z(j(w+mw_o)) \sum_k h_{1,k,i-m} (j(w+mw_o))^k$$

$$+\sum_{m=-\infty}^{\infty} I_m^q(w) e^{j(w+mw_o)} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_k s_{1,k,i} (j(w+mw_o))^k$$

$$+\sum_l (jw)^k e^{jwt} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_k s_{i,k,l} (jw)^k = 0. \quad (19)$$

$$I_q(t) = \sum_{m=-\infty}^{\infty} I_m^q(w) e^{j(w+mw_o)t} \quad (20)$$

Term 2 of (19) can now be rewritten by applying (13) as if there were no crystal and the only noise is the sum of currents given by (20) occurring at port 1. Using superposition, the result is

$$\text{term 2} = -\sum_{m=-\infty}^{\infty} (I_m^q(w) e^{j(w+mw_o)t}) \times \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_{g=-\infty}^{\infty} V1_g(w+mw_o) \sum_k h_{1,k,(i-g)} (j(w+(m+g)w_o))^k. \quad (21)$$

where g corresponds to m, and $V1_g(w+mw_o)$ corresponds to $V_m(w)$ in (13) for the normalized driving current $i=e^{j(w+mw_o)t}$. $V1_g(w+mw_o)$ has the dimensions of impedance. In particular, $V1_o(w)$ is the small signal impedance of port 1 at the disturb frequency, w. All values of $V1_g(w+mw_o)$ can be obtained by SPICE transient circuit simulations.

We will now make term 2 of (19) look like term 1 of (19). Rearrange (21) to get

$$\text{term 2} = -e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w) \times \sum_{l=-\infty}^{\infty} e^{j(l+m)w_o t} \sum_{g=-\infty}^{\infty} V1_{g+m-m}(w+mw_o) \times \sum_k h_{l,k,(l+m-g-m)} (j(w+(m+g)w_o))^k \quad (22)$$

In (22), replace (l+m) by c, and (g+m) by d, changing summation limits accordingly.

$$\text{term 2} = -e^{jwt} \sum_{m=-\infty}^{\infty} I_m^q(w) \times \sum_{c=-\infty}^{\infty} e^{jcw_o t} \sum_{d=-\infty}^{\infty} V1_{d-m}(w+mw_o) \sum_k h_{l,k,(c-d)} (j(w+dw_o))^k \quad (23)$$

Change symbols in (23). Replace m by g, c by l, and d by m to get

$$\text{term 2} = -e^{jwt} \sum_{l=-\infty}^{\infty} e^{jlw_o t} \sum_{g=-\infty}^{\infty} I_g^q(w) \times \sum_{m=-\infty}^{\infty} V1_{m-g}(w+gw_o) \sum_k h_{l,k,(l-m)} (j(w+mw_o))^k \quad (24)$$

Plug back into (19) to get

$$-e^{jwt} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_{m=-\infty}^{\infty} I_m^q(w) Z(j(w+mw_o)) \sum_k h_{1,k,i-m} (j(w+mw_o))^k$$

$$-e^{jwt} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_{g=-\infty}^{\infty} I_g^q(w) \sum_{m=-\infty}^{\infty} V1_{m-g}(w+gw_o) \sum_k h_{l,k,(l-m)} (j(w+mw_o))^k$$

$$+\sum_l (jw)^k e^{jwt} \sum_{i=-\infty}^{\infty} e^{jiw_o t} \sum_k s_{i,k,l} (jw)^k = 0. \quad (25)$$

Collecting terms produces

$$-e^{jwt}\sum_{l=-\infty}^{\infty} e^{jlw_0t} \times \sum_{m=-\infty}^{\infty}\left(I^q_m(w)Z(j(w+mw_0)) + \sum_{g=-\infty}^{\infty} I^q_g(w)V1_{m-g}(w+gw_0)\right) \times \sum_k h_{1,k,(1-n)}(j(w+mw_0))^k$$

Term 1 and Term 2

$$+\sum_i I_i e^{jwt}\sum_{l=-\infty}^{\infty} e^{jlw_0t}\sum_k s_{i,k,l}(jw)^k = 0. \quad (26)$$

Term 3

(26) is the analog of (13). For both of them to be true

$$V_m(w) = -I^q_m(w)Z(j(w+mw_0)) - \sum_{g=-\infty}^{\infty} I^q_g(w)V1_{m-g}(w+gw_0) \quad (27)$$

for each value of m.

Recall that the $V_m(w)$ voltages are produced by removing the linear 1 port from the circuit without changing the stray noise currents. (as in fig. 2) $V1_{(m-g)}(w+gw_0)$ values come from driving the the N-port at port 1 at various frequencies, $(w+gw_0)$ with a normalized current source.

(27) represents a solvable set of simultaneous equations in the unknowns $I_m(w)$ and $I_g(w)$. As was pointed out in sect. 2.2, the values of $h_{1,k,1}$ and $s_{i,k,1}$ have been eliminated.

3.2 Application to the High Q Oscillator Case

If $Z(w+mw_0)$ is a high Q resonant circuit,

$$Z(w+mw_0) = \frac{RQw_1}{j(w+mw_0)} + R + \frac{jRQ(w+mw_0)}{w_1}, \quad (28)$$

where w_1 is the resonant frequency, Q is the Quality factor and R is the series resistance. Substitution into (27) produces

$$V_m(w) = -I^q_m(w)\left(\frac{RQw_1}{j(w+mw_0)} + R + \frac{jRQ(w+mw_0)}{w_1} + V1_0(w+mw_0)\right) - \sum_{g=-\infty,\, g\neq m}^{\infty} I^q_g(w)V1_{m-g}(w+gw_0) \quad (29)$$

(29) can be solved for each value of $I^q_m(w)$ in terms of everything else.

$$I^q_m(w) = \frac{-V_m(w) - \sum_{g=-\infty,\, g\neq m}^{\infty} I^q_g(w)V1_{m-g}(w+gw_0)}{\left(\frac{RQw_1}{j(w+mw_0)} + R + \frac{jRQ(w+mw_0)}{w_1} + V1_0(w+mw_0)\right)} \quad (30)$$

The thing to note here is that as long as $(w+mw_0)$ is not close to $\pm w_1$, $|I^q_m(w)|$ is very small.

When $\pm w_1 \sim w+mw_0$, (29) reduces to two simultaneous equations. Let

$$w = m_0w_0 + \delta. \quad (30a)$$

where $\delta \ll w_0$ and m_0 is an integer. w,m and m_0 can be positive, negative or zero. The important values of m are (solving for m)

$$m = \frac{\pm w_1 - w}{w_0} = \frac{\pm w_1 - m_0w_0 - \delta}{w_0} \approx \pm 1 - m_0. \quad (31)$$

since $w_1 \sim w_0$.

Let us rewrite (29), plugging in (30a) and (31), and setting all values of $I^q_g(w)$ to zero except those with $g = \mp 1 - m_0$. (When $m = 1 - m_0$, $g = -1 - m_0$ and when $m = -1 - m_0$, $g = 1 - m_0$.) If

$$|\delta + w_1 - w_0| \ll w_0. \quad (32)$$

we can write (29) as

$$V_{(\pm 1-m_0)}(m_0w_0+\delta) = -I^q_{(\pm 1-m_0)}(m_0w_0+\delta)\left(\frac{RQw_1}{j(\pm w_0+\delta)} + R + \frac{jRQ(\pm w_0+\delta)}{w_1} + V1_0(\pm w_0+\delta)\right) - I^q_{(\mp 1-m_0)}(m_0w_0+\delta)V1_{\pm 2}(\mp w_0+\delta) \quad (33)$$

where $\mp = -(\pm)$.

(33) represents two equations and can be simplified by writing

$$R + \frac{RQw_1}{j(\pm w_0+\delta)} + \frac{jRQ(\pm w_0+\delta)}{w_1} = R + \frac{RQ(w_1^2 - w_0^2 \mp 2w_0\delta - \delta^2)}{j(\pm w_0+\delta)w_1}$$

$$\approx R + \frac{2RQw_0(w_1 - w_0 \mp \delta)}{\pm jw_0w_1} = R + \frac{2jRQ(\delta \pm (w_0 - w_1))}{w_1}. \quad (34)$$

Since the circuit obeys Kirchoff's current law at the steady state oscillation frequency ($\delta = 0.$),

$$R + \frac{2jRQ(w_0 - w_1)}{w_1} \approx R - \mathrm{Im}(Z_0(w_0)) = -Z_0(w_0). \quad (35)$$

(34) can now be written as

$$R + \frac{RQw_1}{j(\pm w_0+\delta)} + \frac{jRQ(\pm w_0+\delta)}{w_1} \approx \frac{2jRQ\delta}{w_1} + R \mp \mathrm{Im}(Z_0(w_0)). \quad (36)$$

Before applying Cramer's rule to (33), both equations will now be written out in full including all updates from (34) to (36).

$$-V_{(1-m_0)}(m_0w_0+\delta) \approx I^q_{(1-m_0)}(m_0w_0+\delta)\left(\frac{2jRQ\delta}{w_1} - Z_0(w_0) + V1_0(w_0+\delta)\right) + I^q_{(-1-m_0)}(m_0w_0+\delta)V1_2(-w_0+\delta)$$

$$-V_{(-1-m_0)}(m_0w_0+\delta) \approx I^q_{(-1-m_0)}(m_0w_0+\delta)\left(\frac{2jRQ\delta}{w_1} - Z_0^*(w_0) + V1_0(-w_0+\delta)\right) + I^q_{(1-m_0)}(m_0w_0+\delta)V1_{-2}(-w_0+\delta) \quad (37)$$

Application of Cramer's rule to (37) gives the solution for $I_q(m_o w_o+\delta)$, the crystal current response.

$$I^{q}_{(-1-m_o)}(m_o w_o+\delta) =$$

$$-\left(\frac{2j\pi Q\delta}{w_1} - Z_o(w_o) + V1_o(w_o+\delta)\right) V_{(-1-m_o)}(m_o w_o+\delta) +$$

$$V1_{-2}(w_o+\delta) V_{(1-m_o)}(m_o w_o+\delta)$$

$$\left(\frac{2j\pi Q\delta}{w_1} - Z_o(w_o) + V1_o(w_o+\delta)\right)\left(\frac{2j\pi Q\delta}{w_1} - Z_o^*(w_o) + V1_o(-w_o+\delta)\right)$$

$$-V1_{-2}(w_o+\delta) V1_2(-w_o+\delta)$$

$$I^{q}_{(1-m_o)}(m_o w_o+\delta) =$$

$$\left(\frac{2j\pi Q\delta}{w_1} - Z_o^*(w_o) + V1_o(-w_o+\delta)\right) V_{(1-m_o)}(m_o w_o+\delta) +$$

$$V1_2(-w_o+\delta) V_{(-1-m_o)}(m_o w_o+\delta)$$

$$\left(\frac{2j\pi Q\delta}{w_1} - Z_o(w_o) + V1_o(w_o+\delta)\right)\left(\frac{2j\pi Q\delta}{w_1} - Z_o^*(w_o) + V1_o(-w_o+\delta)\right)$$

$$-V1_{-2}(w_o+\delta) V1_2(-w_o+\delta) \qquad (38)$$

Inspection shows that for large δ, (38) is similar to Leeson's equation. However, for non-linear llators, the magnitudes are clipped and distorted near δ=0.

The dimensions of (38) are (VZ/Z^2)=I as expected. It can be seen that the denominator in (38) does not vary with the value of m_o. Using (16) and summing over all mixing currents, I_q can be written as

$$I_q(t) = \sum_{m_o} e^{j(m_o w_o+\delta)t}\left[I^{q}_{(1-m_o)}(m_o w_o+\delta)\, e^{jw_o(1-m_o)t} + I^{q}_{(-1-m_o)}(m_o w_o+\delta)\, e^{jw_o(-1-m_o)t}\right] =$$

$$e^{j\delta t}\left[e^{jw_o t}\sum_{m_o} I^{q}_{(1-m_o)}(m_o w_o+\delta) + e^{-jw_o t}\sum_{m_o} I^{q}_{(-1-m_o)}(m_o w_o+\delta)\right] \qquad (39)$$

3.3 The Linear Case and its Relationship to Leeson's and the Noise Amplifier Models

(39) represents a non-linear mixing process. In a linear N-port, the only value of m_o is 1.

The quantities $V1_o(w_o+\delta)$ and $V1_o(-w_o+\delta)$, in (38) represent ((9), (10), (21)) normalized fundamental network responses to the small signal driving currents $e^{j(w_o+\delta)t}$ and $e^{-j(w_o+\delta)t}$ at port 1. These are essentially small signal driving point impedances the crystal 'sees' at w_o and $-w_o$. If the N-port is linear, these small signal impedances must equal the corresponding large signal values. Therefore, in the steady state

$$V1_o(w_o)-Z_o(w_o)=ar=0., \text{ and}$$
$$V1_o(-w_o)-Z_o{*}(w_o)=ar{*}=0. \qquad (40)$$

Since a linear llator can have no response at any frequency other than that which is applied,

$$V_m(w_o+\delta)=0., \quad V1_m(w_o+\delta)=0.$$
and
$$V_m(-w_o+\delta)=0., \quad V1_m(-w_o+\delta)=0.,$$
for $m \neq 0$. (41)

It follows from (38)-(41) that I_q is very sensitive to circuit non-linearities in addition to disturbance magnitudes.

To relate (38) to Leeson's model, we can calculate the noise near w_o caused by the sum of I_q and I_i by applying (37) to the linear case. (m_o=1)

$$V_{across\ crystal}(w_o+\delta) = -I^{q}_{o}(w_o+\delta)V1_o(w_o+\delta) - V_o(w_o+\delta) \qquad (42)$$

For a linear llator, (38) reduces to

$$I^{q}_{(-1-m_o)}(m_o w_o+\delta)=0. \qquad I^{q}_{(1-m_o)}(m_o w_o+\delta) = -\frac{w_1}{2j\pi Q\delta} V_{(1-m_o)}(m_o w_o+\delta) \quad (m_o=1) \qquad (43)$$

Applying (42) and taking magnitudes produces Leeson's equation for circuit noise.

$$V_o^2(w_o+\delta)\left(\left(\frac{w_1}{2\pi Q\delta} V1_o(w_o+\delta)\right)^2+1.\right) = V^2_{across\ crystal}(w_o+\delta) \qquad (44)$$

$$'Q_{op}' = RQ/V1_o(w_o+\delta) \qquad (44a)$$

If, however, the circuit is linear enough so that (41) is true and ar in (40) is not 0 but a small real number, then (38) becomes

$$I^{q}_{(-1-m_o)}(m_o w_o+\delta)=0. \qquad I^{q}_{(1-m_o)}(m_o w_o+\delta) = \frac{-V_{(1-m_o)}(m_o w_o+\delta)}{\frac{2j\pi Q\delta}{w_1} + ar} \quad (m_o=1) \qquad (45)$$

Plugging (45) into (42) and taking magnitudes produces

$$V_o^2(w_o+\delta)\left(\frac{V1_o^2(w_o+\delta)}{\left(\frac{2\pi Q\delta}{w_1}\right)^2 + ar^2}+1.\right) = V^2_{across\ crystal}(w_o+\delta) \qquad (46)$$

(46) is identical to (17) in [12] and to (15) in [11] when I_q is observed alone. (46) is a key result of the 'noise amplifier model' published by Robbins, Parzen and Hillstrom, and is used to handle the well known dilemma in Leeson's model of white noise going to infinity when δ goes to 0.

3.4 Relationship of This Paper to Circuit Simulations

To find the response of distortion limited oscillators to $e^{jM_o w_o t}$, run spice with the main oscillation current= $I_o\cos(w_o t)$ plus a noise current $=I_n\cos(m_o w_o t)$. Repeat with a noise current=$I_n\cos(m_o w_o t-\pi/2)$= $I_n\sin(m_o w_o t)$. (See figure 4.) I_n

should be small enough to run distortion free when applied alone. Spice accuracy parameters such as TIME STEP and RELTOL should be tough enough to handle the small noise current. To measure the noise effects, take the Fourier series of the differences between each noisy waveform and the noiseless oscillation. (As is done in fig. 4) Because the results represent 'steady state', simulated time should be long enough for all initial transients to die out.

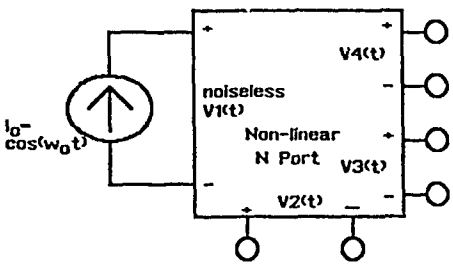


Figure 4a Simulation of a Noiseless Ilator.

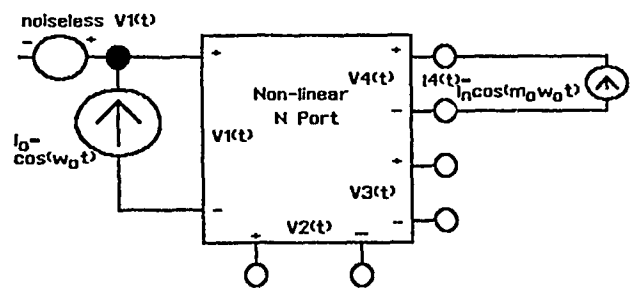


Figure 4b Simulation of a noisy Ilator driven by a cosin function at port 4. Take Fourier series of voltage between the negative terminals of "noiseless V1(t)" and V1(t).

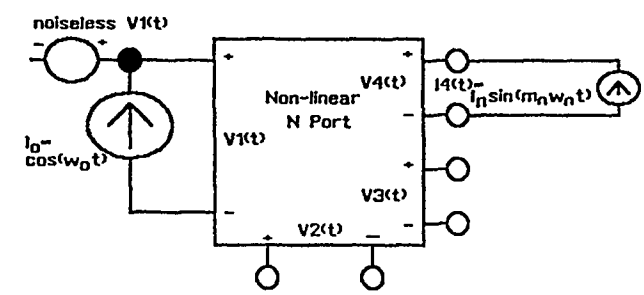


Figure 4c Simulation of noisy Ilator driven by a sin function at port 4. Take Fourier series of voltage between the negative terminals of "noiseless V1(t)" and V1(t).

The values printed out of Spice can be written as follows:

$$\text{response to } \cos(m_0w_0t) = \sum_{n=0}^{\infty} |\text{mag}_n| \cos(nw_0t+\beta_n) = \sum_{n=0}^{\infty} \frac{|\text{mag}_n|}{2}\left(e^{j(nw_0t+\beta_n)}+e^{-j(nw_0t+\beta_n)}\right) = \sum_{n=-\infty}^{\infty} Z_n e^{jnw_0t}(Z_{-n}=Z_n^*) \tag{47}$$

where n is the harmonic number in the Spice output. Similarly,

$$\text{response to } \sin(m_0w_0t) = \sum_{n=0}^{\infty} |\text{mag}'_n| \cos(nw_0t+\beta'_n) = \sum_{n=-\infty}^{\infty} Z'_n e^{jnw_0t}(Z'_{-n}=Z'^*_n) \tag{48}$$

The voltage coefficients required by (38) can now be found. Euler's identity can be used to get

$$\text{response to } e^{jm_0w_0t} = \text{response to } \cos(m_0w_0t)+j\times\text{response to } \sin(m_0w_0t) = \sum_{n=-\infty}^{\infty}(Z_n+jZ'_n)e^{jnw_0t} = e^{jm_0w_0t}\sum_{n=-\infty}^{\infty}(Z_n+jZ'_n)e^{j(n-m_0)w_0t} \tag{49}$$

Similarly,

$$\text{response to } e^{-jm_0w_0t} = \text{response to } \cos(m_0w_0t)-j\times\text{response to } \sin(m_0w_0t) = \sum_{n=-\infty}^{\infty}(Z_n-jZ'_n)e^{jnw_0t} = \sum_{n=-\infty}^{\infty}(Z_{-n}-jZ'_{-n})e^{-jnw_0t} = \sum_{n=-\infty}^{\infty}(Z_n^*-jZ_n'^*)e^{-jnw_0t} = \text{conjugate of response to } e^{jm_0w_0t} \tag{50}$$

To relate (49) to (16) and (9), define

$$w=m_0w_0 \text{ and } m=n-m_0 \tag{51}$$

The following examples follow from (51). (see (38))

m	m_0	$n=m+m_0$
$1-m_0$	m_0	1
$-1-m_0$	m_0	-1
$1+m_0$	$-m_0$	1
$-1+m_0$	$-m_0$	-1
-2	1	-1
2	-1	1
0	1	1
0	-1	-1

The important values of n are always ±1.

To take meaningful data, one must simulate the individual effects of each noise source (a very time consuming job that can be automated) and then calculate the total RMS effect near w_0. A composite simulation will give errors because fixed phase relations within the N-port can cause fictitious signal cancellations. Each source, including the one across the crystal, should be simulated for m_0=0 through 9. (see 30a) (38), the noise spectral density near w_0, was rearranged to look like (52).

$$I_{(\pm 1-m_0)}(m_0w_0+\delta) = \frac{-A\left(\left(\frac{2jRQ\delta}{w_1QQ}\right)+QNR+jxQNI\right)+B}{\left(\frac{2jRQ\delta}{w_1QQ}\right)^2+2xQNR\left(\frac{2jRQ\delta}{w_1QQ}\right)+1} \tag{52}$$

where

$$QQ = \sqrt{|V1_0(w_0+\delta)-Z_0(w_0)|^2-|V1_{-2}(-w_0+\delta)|^2}$$

$$QNR = \frac{Re(V1_0(w_0+\delta))+B}{QQ}$$

$$QNI = \frac{\pm Im(Z_0(w_0)+Im(V1_0(\pm w_0+\delta))}{QQ}$$

$$A = \frac{V_{(\pm 1-m_0)}^{(m_0 w_0+\delta)}}{QQ}$$

$$B = \frac{V1_{\pm 2}^{(\mp w_0+\delta)} V_{(\mp 1-m_0)}^{(m_0 w_0+\delta)}}{QQ^2}$$

A and B are to be multiplied by the magnitude of the noise currents when calculating the actual noise. (52) is normalized so that A and B represent noise multipliers for $V_{(1-m_0)}$ and $V_{(-1-m_0)}$ respectively respectively at δ=0. QQ is a derating factor on the crystal Q, affecting the 'sharpness' of (52). The $\pm$ sign follows that of the '1' in '$\pm$1-m_0.

4.0 Application to an Ordinary Non-Linear Oscillator

To illustrate the process described above, clamp diodes were added to the inverter of [12] to produce figure 5. Stray capacitance was adjusted so that the small signal properties of the resulting CMOS llator matched that described in [13]. At a simulated oscillation current of 1.9ma, the llators produced the results of table 1.

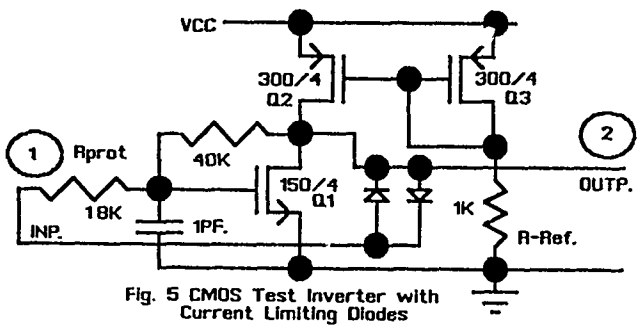


Fig. 5 CMOS Test Inverter with Current Limiting Diodes

Table 1

	Clamped	Unclamped
$Z_o(w_o)$(ohms)	-57.36-j353.8	-139.8-j315.
% Distortion in crystal voltage	3.98	.976
$\lvert V1_o(w_o) / -Z_o(w_o) \rvert$(ohms)	1191.	1.41

The third item in table 1 is the measure of the oscillator's linearity given in (40). Clearly the diodes affect both the linearity and steady state current without generating highly distorted waveforms.

Spice simulations were run for separate current noise inputs across the crystal (nodes 1 and 2), and the output (nodes 2 and ground) for the circuit of figure 5. Table 2 gives the processed Spice Fourier coefficients required by (52).

Table 2

Simulated Results for the Clamped LLator

QQ=.2322E+03 2QNR=.9543E+01

QNR=.4771E+01 QNI=-.1874E+01

NOISE IN PARALLEL WITH CRYSTAL

COEFFICIENTS OF EXP(-JWOT) ($-1-m_o$)

Mo	!A!	!B!
0	.5215E-01	.2622E+00
1	.5658E+01	.2528E+02
2	.1584E+00	.7731E+00
3	.1051E+01	.4090E+01
4	.5546E-01	.2178E+00
5	.5822E+01	.2344E+01
6	.4266E-01	.1869E+00
7	.3029E+00	.1343E+01
8	.3183E-01	.1483E+00
9	.1045E+00	.5292E+00

COEFFICIENTS OF EXP(JWOT) ($1-m_o$)

Mo	!A!	!B!
0	.5809E-01	.2921E+00
1	.5446E+00	.2792E+01
2	.6373E-02	.3267E-01
3	.1948E+00	.9986E+00
4	.7828E-02	.4013E-01
5	.9245E-01	.4739E+00
6	.6561E-02	.3363E-01
7	.3700E-01	.1897E+00
8	.3709E-02	.1902E-01
9	.1024E-01	.5247E-01

Mo	!A!x!QNR+jQNI!	!B!
0	.2673E+00	.2577E+02
1	.2900E+02	.2845E+02
2	.8122E+00	.7882E+00
3	.5389E+01	.4170E+01
4	.2843E+00	.2221E+00
5	.2985E+01	.2389E+01
6	.2187E+00	.1906E+00
7	.1553E+01	.1369E+01
8	.1632E+00	.1512E+00
9	.5356E+00	.5395E+00

NOISE ORIGINATING FROM THE INVERTER OUTPUT

COEFFICIENTS OF EXP(-JWOT) ($-1-m_o$)

Mo	!A!	!B!
0	.5809E-01	.2978E+00
1	.5446E+00	.2792E+01
2	.6373E-02	.3267E-01
3	.1948E+00	.9986E+00
4	.7828E-02	.4013E-01
5	.9245E-01	.4739E+00
6	.6561E-02	.3363E-01
7	.3700E-01	.1897E+00
8	.3709E-02	.1902E-01
9	.1024E-01	.5247E-01

COEFFICIENTS OF EXP(JWOT) ($1-m_o$)

Mo	!A!	!B!
0	.5809E-01	.2921E+00
1	.5446E+00	.2425E+01
2	.1640E-01	.8403E-01
3	.7571E+00	.3875E+01
4	.3275E-01	.1678E+00
5	.3726E+00	.1909E+01
6	.2851E-01	.1461E+00
7	.1641E+00	.8406E+00
8	.1700E-01	.8708E-01
9	.5179E-01	.2653E+00

Mo	!A!x!QNR+jQNI!	!B!
0	.2978E+00	.2921E+00
1	.2425E+01	.2472E+01
2	.3263E-02	.1672E-01
3	.1506E+00	.7719E+00
4	.6515E-02	.3340E-01
5	.7412E-01	.3799E+00
6	.5671E-02	.2907E-01
7	.3264E-01	.1673E+00
8	.3381E-02	.1733E-01
9	.1030E-01	.5280E-01

Mo	!A!x!QNR+jQNI!	!B!
0	.2673E+00	.2622E+00
1	.2577E+02	.2845E+02
2	.7882E+00	.7966E+00
3	.4170E+01	.5285E+01
4	.2221E+00	.2788E+00
5	.2389E+01	.2927E+01
6	.1906E+00	.2145E+00
7	.1369E+01	.1523E+01
8	.1512E+00	.1600E+00
9	.5395E+00	.5253E+00

The total noise current I_q is arrived at by a 'square root of the sum of squares' process over all values of m_o and all ports. According to the table, noise directly across the crystal near w_o is the largest contributor.

A major drawback of Spice is its long run time when simulating circuits containing large and small currents. Harmonic Balance programs should speed things up. However, as of March, 1989 the latest Berkley program 'Specter 1a1' does not handle MOSFETS and will not run on a standard IBM PC.[13]

4.1 Relationship Between Circuit, Phase, and Amplitude Noise.

This section translates between Circuit noise; and Amplitude and Phase noise. Single and double sideband effects will also be looked at in relation to the translation.

To start out, assume that the main oscillation current is $\cos(w_o t)$ and that the small noise current, n(t) is given by (53), where $e(w) \ll 1$. Trigonometric identities are used to break the noise up into two components. The sum of the two is given by (54). (55) identifies the phase and amplitude noise components.

$$n(t) = \int_0^\infty e(w)\cos(wt + \phi(w))dw =$$
$$\int_0^\infty e(w)\cos((w-w_o+w_o)t+\phi(w))dw =$$
$$\cos(w_o t)\int_0^\infty e(w)\cos((w-w_o)t+\phi(w))dw$$
$$-\sin(w_o t)\int_0^\infty e(w)\sin((w-w_o)t+\phi(w))dw \quad (53)$$

$$\cos(w_o t)+n(t) = \cos(w_o t)\left(1+\int_0^\infty e(w)\cos((w-w_o)t+\phi(w))dw\right) \quad (54)$$
$$-\sin(w_o t)\int_0^\infty e(w)\sin((w-w_o)t+\phi(w))dw$$

$$A_n(t) = \int_0^\infty e(w)\cos((w-w_o)t+\phi(w))dw \text{ , and}$$
$$\phi_n(t) = \int_0^\infty e(w)\sin((w-w_o)t+\phi(w))dw \quad (55)$$

If $e(w) \ll 1$, then $A_n(t) \ll 1$, then $\phi_n \ll 1$, $\phi_n \sim \sin(\phi_n)$ and $\cos(\phi_n) \sim 1$. (56) uses these approximations to get its final form.

$$\cos(w_o t)+n(t) = \cos(w_o t)(1+A_n(t)) - \sin(w_o t)\phi_n(t) \approx$$
$$(1+A_n(t))\{\cos(w_o t) - \sin(w_o t)\phi_n(t)\} \approx$$
$$(1+A_n(t))\{\cos(w_o t)\cos(\phi_n(t))-\sin(w_o t)\sin(\phi_n(t))\} \approx$$
$$(1+A_n(t))\{\cos(w_o t+\phi_n(t))\} \quad (56)$$

The main point of (53)-(56) is that one noise signal affects both the phase and amplitude of the carrier, and that the effects are related to each other by (55).

By letting $\phi(w)$ equal 0. and $-\pi/2$ in two applications of (53)-(56), Euler's identity and superposition, can be used to show that the 'amplitude' noise due to he^{jwt} is
$he^{j(w-w_o)}$.

(56a)

The corresponding 'phase modulation' is $-jhe^{j(w-w_o)}$.

(56b)

A similar exercise proves that the 'amplitude' noise due to $he^{j(w-2w_o)t}$ is

$he^{j(w-w_o)}$,

(56c)

and the 'phase modulation' is $jhe^{j(w-w_o)}$.

(56d)

(56c) and (56d) can be combined to get (56e) a signal with phase noise and no amplitude noise.

$$\text{noise} = h(e^{jwt} - e^{j(w-2w_o)t}) \quad (56e)$$

(56e) will be used later on in this paper.

(57)-(58) reverses the logic of (53)-(56) to calculate the circuit noise in terms of the phase and amplitude noises. A trigonometric identity is used to transform (58) into its final form. (58) states that circuit noise has two sidebands unless the phase and amplitude noises have equal magnitudes and are 90° out of phase.

$$A_n(t) = \int_0^\infty e(w)\cos(\delta t+\phi(w))d\delta$$
$$\phi_n(t) = \int_0^\infty e_1(w)\sin(\delta t+\phi_1(w))d\delta \quad (57)$$

From (56)

$$n(t) = \cos(w_o t)A_n(t)-\sin(w_o t)\phi_n(t)$$
$$-\cos(w_o t)\int_0^\infty e(w)\cos(\delta t+\phi(w))d\delta$$
$$-\sin(w_o t)\int_0^\infty e_1(w)\sin(\delta t+\phi_1(w))d\delta$$
$$= .5\int_0^\infty e(w)\{\cos((\delta+w_o)t+\phi(w)) + \cos((w_o-\delta)t-\phi(w))\}d\delta$$
$$+.5\int_0^\infty e_1(w)\{\cos((\delta+w_o)t+\phi_1(w)) - \cos((w_o-\delta)t-\phi_1(w))\}d\delta \quad (58)$$

4.2 Application to an AGC Oscillator

The AGC oscillator is an important circuit in use today. No paper is complete without a discussion of the relation between the AGC circuit and phase noise.

The AGC controls the oscillation level by varying the large signal impedance ($Z_o = Z_o(w_o)$) in response to a filtered rectified oscillation waveform. (See figure 6. for a possible circuit.) Noise near D.C. (w near 0.) and slow amplitude modulation (w near w_o) produce similar effects. (See figure 7 and (56).) All other noise is ignored. In a good oscillator, the limiter operates close to linearity under AGC control. Since distortion harmonics are reduced, $w_o+\delta$ and δ are the only serious noise sources.

If, however, the oscillations are AC coupled to the rectifier filter, (through C1 in figure 6, for example) then limiter noise near DC is blocked. (Since the AGC can generate it's own noise, modulation of Zo near DC and w_o still occurs.) This effect can be calculated in a manner similar to distortion limiting. Details will be given below. For now, let us concentrate on current noise near w_o injected into port 1. (across the crystal.)

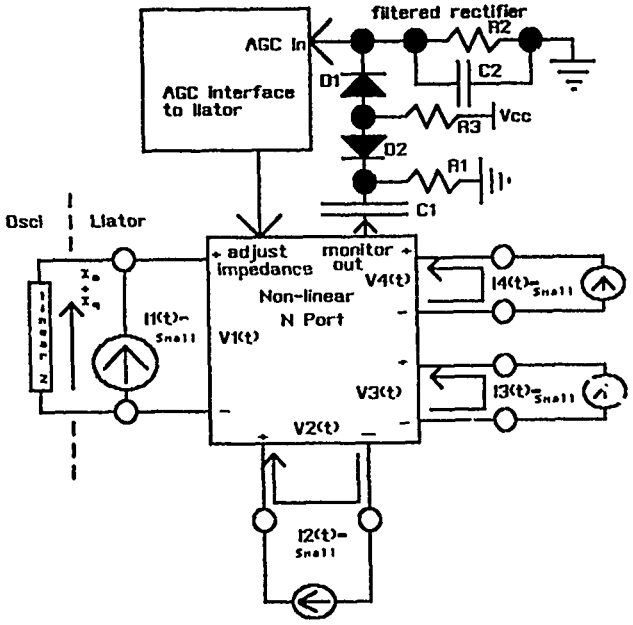


Figure 6 High Q Oscillator with Extraneous Current Inputs and AGC Circuitry

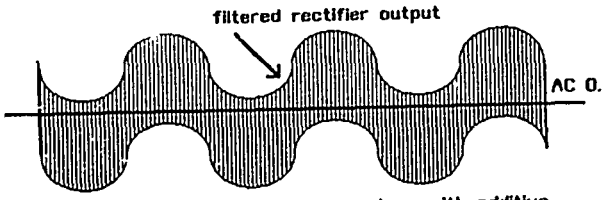


Figure 7a Exaggerated oscillation envelope with additive low frequency noise

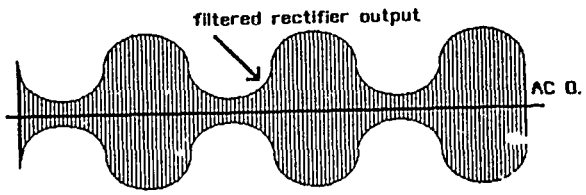


Figure 7b Exaggerated oscillation envelope with additive noise near w_0

The relatively long time constant in the filter makes it inconvenient to simulate the entire system as is done with the distortion limited case. Instead, one should tabulate Zo as a function of Vc. (the control voltage applied to the 'AGC in' terminal in figure 6) This voltage varies with the magnitude of the oscillation. Phase changes are not detected.

$$dVc/di_o = d|Vc|/di_o = kd|V|/di_o, \qquad (59)$$

where i_o is the magnitude of the oscillation current at port 1, V is the oscillation voltage amplitude across the crystal, and k is the translation constant.

The control loop can be investigated to find the magnitude of the small signal AC impedance 'seen' by the crystal. (60)-(61) expresses |V| as a function of i_o taking the $|Z_o|$ variation into account.

$$|V| = |Z_o|\, i_o$$

$$\Delta|V| = |Z_o|\Delta i_o + i_o \frac{d|Z_o|}{d|V|}\Delta|V| \qquad (60)$$

Solving for $\Delta|V|$ produces,

$$\Delta|V| = \frac{|Z_o|\Delta i_o}{1 - i_o \frac{d|Z_o|}{d|V|}} \qquad (61)$$

The change in the magnitude of i_o referred to in (60)-(61) is related to noise by the A_n term in (54). $d|Zo|/d|V|$ follows from the tabulation of Zo vs. Vc and (59). Residual non-linear effects that add to this value are neglected.

To apply this paper to the AGC oscillator, the coefficients of $\Delta V1(t)$ will now be calculated in terms of the noise, assuming that

$$\text{noise} = i_n e^{jwt}. \qquad (62)$$

The main oscillation is a cosin function, and the analysis is in the time domain. The solution should have the form given by (63). (63) is (16) rewritten with the symbols of (38).

$$\Delta V1(t) = e^{jwt} \sum_{m=-\infty}^{\infty} V1_m(w)\, e^{jmw_0 t} \qquad (63)$$

The expression for V1(t) and $\Delta V1(t)$ is

$$V1(t) = .5\,(Z_o i_o e^{jw_0 t} + Z_o^* i_o e^{-jw_0 t}). \qquad (64)$$

$$\Delta V1(t) = Z_o i_n e^{jwt} + .5\,(\Delta Z_o i_o e^{jw_0 t} + \Delta Z_o^* i_o e^{-jw_0 t}) + h(e^{jwt} - e^{j(w-2w_0)t}) \qquad (65)$$

h represents small non-linear effects in the Ilator which only generate phase noise. (64) is actually a form of (1) and (65) is (3) solved for $\Delta V1(t)$. It plays an important part in the following discussion.

From the tabulation of Zo vs. Vc, (59) and (61), we can calculate

$$\Delta Z_o = \frac{dZ_o}{d|V|}\Delta|V| = \frac{dZ_o}{d|V|}\frac{|Z_o|\Delta i_o}{1 - i_o \frac{d|Z_o|}{d|V|}}$$

$$\Delta Z_o^* = \frac{dZ_o^*}{d|V|}\Delta|V| = \frac{dZ_o^*}{d|V|}\frac{|Z_o|\Delta i_o}{1 - i_o \frac{d|Z_o|}{d|V|}} \qquad (66)$$

From (56a),

$$\Delta|i_o| = i_n e^{j(w-w_c)t} \qquad (67)$$

Substitution of (66)-(67) into (65) produces (68).

$$\Delta V1(t) = Z_0 i_n e^{jwt} + \frac{.5\,|Z_0|^2 i_0 i_n}{1 - i_0 \frac{d|Z_0|}{dVl}} e^{j(w-w_0)t} \left\{ \frac{dZ_0}{dVl} e^{jw_0 t} + \frac{dZ_0^*}{dVl} e^{-jw_0 t} \right\} + h(e^{jwt} - e^{j(w-2w_0)t}) -$$

$$i_n e^{jwt} \left\{ Z_0 + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} + h + \left(\frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} - h \right) e^{-j2w_0 t} \right\} \quad (68)$$

A similar analysis assuming that

$$\text{noise} = i_n e^{-jwt} \quad (69)$$

produces a corresponding expression.

$$\Delta V1(t) = i_n e^{-jwt} \left\{ Z_0^* + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} + h^* + \left(\frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} - h^* \right) e^{j2w_0 t} \right\} \quad (70)$$

The coefficients required by the denominator of (38) follow from (68) and (70).

Letting $w = w_0 + \delta$, and $-w = \delta - w_0$,

$$V1_0(w_0+\delta) = Z_0 + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} + h$$

$$V1_{-2}(w_0+\delta) = \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} - h$$

$$V1_0(\delta - w_0) = Z_0^* + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} + h^*$$

$$V1_2(\delta - w_0) = \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} - h^* \quad (71)$$

Up to now, we have measured noise effects near w_0 by forcing a current somewhere in the circuit and measuring a voltage at port 1. Since the AGC response is slow, this not is practical. When designing a real oscillator, a way must be found to incorporate noise originating at any port. Another way of doing this is to use a controlled voltage source to force port 1 to its 'noiseless' waveform, while forcing noise current into the other ports one at a time. (See figure 8.) This shorts out the AGC, eliminating its slow response. The difference between the controlled source current and I_0, the main oscillation current represents the effect of a particular port. (27) describes this method if $Z(j(w+mw_0))$ is set to zero, $I_g{}^q(w)$ represents the simulated current difference, and the V1's are given by (71). $V_m(w)$ is the calculated open circuit voltage at port 1 required by (38).

For the almost linear case, small signal AC analysis near w_0 can be substituted for the transient simulations without much loss in accuracy. (72) is a rewrite of (27) including (71) and the substitution of the symbol I1 for Iq. I1 represents a statistical sum over all noise sources. The almost linear case simplifies things by causing $I1_2$ and $I1_{-2}$ to be ~ 0.

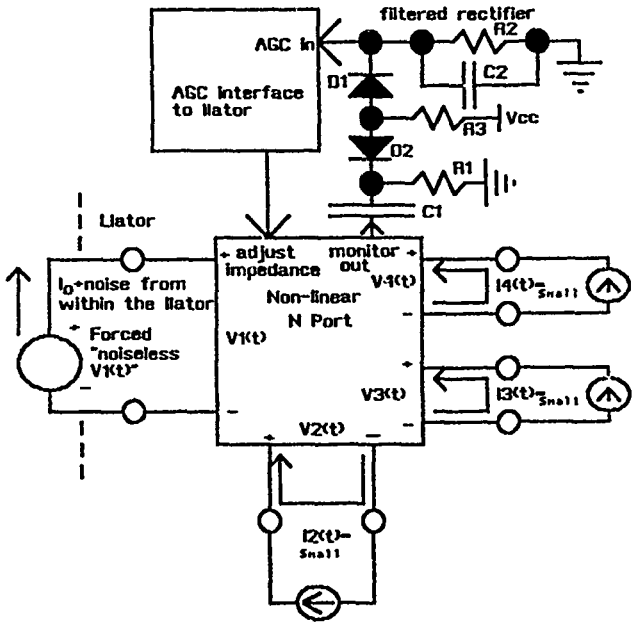


Figure 8 Measurement of noise effects by forcing a "noiseless" voltage waveform on port 1.

$$V_0(w_0+\delta) = I1_0(w_0+\delta) \left\{ Z_0 + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} + h \right\}$$

$$V_{-2}(w_0+\delta) = I1_{-2}(w_0+\delta) \left\{ \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} - h \right\} \quad (72)$$

$$V_0(\delta - w_0) = I1_0(\delta - w_0) \left\{ Z_0^* + \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0^*}{dVl} + h^* \right\}$$

$$V_2(\delta - w_0) = I1_2(\delta - w_0) \left\{ \frac{.5\,|Z_0| i_0}{1 - i_0 \frac{d|Z_0|}{dVl}} \frac{dZ_0}{dVl} - h^* \right\}$$

Noise originating anywhere in the Ilator and AGC is now taken into account. Due to its random nature, only the current magnitudes are important.

Low frequency noise in the AGC will modulate Z_0. (Because C1 in figure 5 does not block it.) The quickest way to simulate this is to use small signal analysis to find the change in Z_0 caused by a small current added to a port in the AGC. All other ports should be left open. The AGC should be embedded in the Ilator and forced to its operating level. The analysis of low frequency noise given in (73)-(76) follows from (65) with i_n and h set to 0.

$$\Delta V(t) = .5\left(\Delta Z_0 i_0 e^{jw_0 t} + \Delta Z_0^* i_0 e^{-jw_0 t} \right) \quad (73)$$

$$\Delta Z_0 = \sum_i \frac{dZ_0}{dI_i} \Delta I_i e^{j\delta t} \qquad \Delta Z_0^* = \sum_i \frac{dZ_0^*}{dI_i} \Delta I_i e^{j\delta t} \quad (74)$$

where δ is the noise frequency and $\sum_i \frac{dZ_0}{dI_i} \Delta I_i$ represents a sum over all low frequency noise sources.

Substituting (73) into (72) produces

$$\Delta V(t) = .5 i_0 e^{j\delta t} \left(e^{jw_0 t} \sum_i \frac{dZ_0}{dI_i} \Delta I_i + e^{-jw_0 t} \sum_i \frac{dZ_0^*}{dI_i} \Delta I_i \right) \quad (75)$$

In terms of (38),

$$V_1(\delta) = .5 i_0 \sum_i \frac{dZ_0}{dI_i} \Delta I_i(\delta) \qquad V_1(\delta) = .5 i_0 \sum_i \frac{dZ_0^*}{dI_i} \Delta I_i(\delta) \quad (76)$$

The final task for this section is to write (38) for an AGC llator. (71) and (38) were used to calculate the denominator of (77). The numerator terms are given in (72) and (76). It is assumed that the llator is linear enough to have significant noise sources only near DC and $\pm w_o$ (m_o=-1,0,1). h, a measure of the latent non-linearity, determines the noise density at δ =0. As h decreases in value, (77) resembles the 'noise amplifier' equation [2,10,11] discussed earlier. h determines the value of ar in (40), (45) and (46).

Use (39) to sum (77) over all values of m_o when observing the composite AGC noise spectrum.

$$I^q_{(m_o w_o+\delta)} \cdot \left(\frac{2jRQ\delta}{w_1} + \frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{dZ_0}{dV} + h \right) V_{(m_o w_o+\delta)} + \left(\frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{dZ_0^*}{dV} - h \right) V^*_{(m_o w_o+\delta)}$$

$$\left(\frac{2jRQ\delta}{w_1} + \frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{d(Z_0+Z_0^*)}{dV} \right)\left(\frac{2jRQ\delta}{w_1} + h + h^* \right)$$

$$I^q_{(m_o w_o+\delta)} \cdot \left(\frac{2jRQ\delta}{w_1} + \frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{dZ_0^*}{dV} + h^* \right) V_{(m_o w_o+\delta)} + \left(\frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{dZ_0}{dV} - h^* \right) V^*_{(m_o w_o+\delta)}$$

$$\left(\frac{2jRQ\delta}{w_1} + \frac{.5\,|Z_0|\,|i_0}{1-i_0\frac{d|Z_0|}{dV}}\frac{d(Z_0+Z_0^*)}{dV} \right)\left(\frac{2jRQ\delta}{w_1} + h + h^* \right) \qquad (77)$$

5.0 Conclusions

The effects of circuit noise on the high Q oscillator were analyzed using a non-measurable function with non-measurable parameters. These parameters were eliminated, yielding a practical solution. This result acts like Leeson's equation and reduces to it for the linear case. Noise effects are exacerbated by the interaction of the crystal and the circuit non-linearities. The 'Noise Amplifier' model proposed by various authors is shown to apply for almost linear circuits.

Phase, amplitude and circuit noise are related by a simple formula. Double side band effects are due to the randomness of circuit noise.

Examples of diode limited and AGC limited llators are given. It is shown that for the both llators, the main effects are near w_o. Spice was used to calculate coefficients for diode limited case.

For the diode limited case, it is shown that circuit noise near DC can be reduced if the oscillations are AC coupled to the control circuit. For 'almost linear' AGC llators, the 'noise amplifier' is fairly accurate. The noise density very close to w_o is a funccion of the llator's small deviation from linearity.

Since the analysis is so complicated and time consuming, the computations must be automated for practical use.

References

[1] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise", Proceedings of the IEEE, Feb. 1966, PP. 329-330

[2]W. F. Robbins, "Phase noise in Signal Sources", Peter Peregrinus Ltd., 1982 PP. 47-60, 68-69

[3]M. J. Buckingham, "Noise in Electronic Devices and Systems", Ellis Horwood Ltd., 1983, PP. 216-217.

[4]G. Sauvage "Phase Noise in Oscillators: A Mathematical Analysis of Leeson's Model" IEEE Transactions on Inst. and Meas. Vol. IM-26, No.4 PP. 408-410.

[5]B. Parzen, A. Balleto "Design of Crystal and Other Harmonic Oscillators", Wiley - Interscience, 1983, Chapter 14.

[6]B. Van Der Pol "The Non Linear Theory of Electric Oscillations" Proceedings of the IRE, Sept. 1934, Vol. 22, No. 9 PP. 1051-1086

[7]Y. Sakuta, Y. Sekine, and M. Suyama "Short Term Frequency Stability of Crystal Oscillator Circuits With the Current Noise in Transistor", Electronics and Communications in Japan, Part 2 Vol. 69, No. 2, 1986.

[8]P. Franklin "An introduction to Fourier Methods and the Laplace Transform" Dover Publications, 1948 PP. 91-94 and PP. 1-21.

[9]E. Hafner, "The Effects of Noise in Oscillators", Proceedings of the IEEE, Feb., 1966, Vol. 54, No. 2.

[10]T. L. Hillstrom, "A Unified Approach to Understanding Oscillator Noise and Injection Locking" Santa Clara R.F.Technology Expo '89 2/14/89-2/16/89 PP. 247-258.

[11]B. Parzen "A Clarification and Generalized Restatement of Leeson's Oscillator Noise Model" Proceedings of the 42'd Annual Symposium on Frequency Control 6/88, PP. 348-351

[12]J. Goldberg "A Simple Way of Characterizing High Q Oscillators" Proceedings of the 42'd Annual Symposium on Frequency Control 6/88, PP. 304-326.

[13]Industrial Liaison Program, "Public Domain Software" Dept. of Electrical Engineering and Computer Sciences Electronics Research Laboratory. University of California at Berkeley. March,1989

43rd Annual Symposium on Frequency Control - 1989

A LOW-NOISE, MULTIPLE-FREQUENCY OSCILLATOR UTILIZING LARGE NUMBERS OF ELECTRONIC SWITCH-SELECTABLE CRYSTALS

M. M. Driscoll and N. Matthews

Westinghouse Electric Corporation
Electronic Systems Croup
Baltimore, Maryland 21043

Summary

In many applications requiring low-noise, selectable, multiple frequency sources, frequency/phase coherency is not required. It is possible in these cases to utilize a single oscillator sustaining stage in combination with a number of (PIN or Schottky diode) switch-selectable quartz crystal resonators. Normally, the maximum number of resonators utilized is small due to the additive effects of off-state switch circuit component and printed wiring board track inductance and capacitance. This is especially true for oscillator operation in the VHF range.

In 1986, one of the authors described a quartz crystal oscillator design using a low 1/f noise, modular amplifier and a lumped element approximation, quarter wave transmission line for transforming/inverting the impedance of the quartz crystal resonator [1].

This paper is a report on the design and performance of an oscillator using the same design techniques but incorporating 16 PIN diode switch-selectable, fifth overtone crystals operating in the 100-MHz range.

Use of the lumped element, quarter wavelength line provides a means for absorption of the printed wiring board track connections and switch parasitic reactances for the 15 "switched-off" crystals. PIN diodes were utilized (rather than Schottky) due to lower off-state capacitance, and PIN diode use has no measurable degrading influence on oscillator signal phase noise performance.

Oscillator flicker of frequency noise is characterized by Sy (100 Hz) = 3×10^{-25} and is due to short-term instability in the resonators themselves. The phase noise floor level is -169 dB/Hz. Using modest levels of excess gain, interchannel frequency switching time is on the order of 2 msec, compared to 9 msec turn-on time.

The same design technique should be equally applicable in the low UHF band using, for example, SAW resonators.

Introduction

Figure 1 shows a crystal oscillator circuit based on a design approach described in a previous paper presented at the 1986 Frequency Control Symposium [1].

As shown in the figure, the oscillator incorporates a self contained modular amplifier and other 50-ohm subcircuits and simple AGC (signal limiting diode) mechanization to ensure linear amplifier operation. In addition, crystal resonator connection to the sustaining stage is made via a quarter wave (lumped element approximation) transmission line impedance transformer that inverts the series resonant crystal impedance to a low value parallel resonant impedance. An advantage associated with use of the circuit is that printed wiring board microstrip connections to the crystal (nodes x to y in figure 1) can be absorbed into the circuit as part of the quarter wave line.

This fact has greatly eased circuit modification to incorporate 16, switch-selectable crystal resonators requiring reasonably long printed wiring branch connections to the crystals.

Design of the Multiple Crystal Oscillator

Figure 2 shows a drawing of the artwork for the crystal connection portion of the circuit. The artwork is arranged so that the track length to any crystal (from figure 1, node y) is identical, and the printed wiring connections to the "switched-off" crystals constitute an array of open circuit stubs whose impedance, along with that of the "on" crystal connecting track and the crystal static capacitance, can be absorbed as part of the overall, quarter wave line by making the lumped element portion of the line slightly shorter.

The success achieved with regard to identical impedance transformation of each crystal is demonstrated in figure 3, which shows Smith chart impedance plots (measured at figure 1, node "c" with nodes "a" and "b" disconnected) for crystals switched "on" from four quadrants of the figure 2 layout. As shown in the figure, nominal node "c" parallel resonant resistance is on the order of 22 ohms. The transformation of the 50-ohm crystal series resistance to this value is accomplished using an effective quarter wave line characteristic impedance of 33 ohms.

Figure 4 shows plots of the small and large (quiescent operating point) signal magnitude and phase responses of the sustaining stage circuit (figure 1, node "a" to node "b") with a 22-ohm resistor substituted for the crystal-plus-quarter wave line. The figure indicates the gain variation of the AGC circuit and the tuned circuit selectivity required to assure oscillation only at the desired crystal (fifth overtone) resonance.

As shown in figure 4, small signal excess gain is approximately 5 dB (slightly larger value than normally utilized), selected on the basis of achieving the requisite interchannel switching time performance.

Figure 5 shows a schematic diagram of the PIN diode, crystal switching circuit. Use of PIN diodes (CR1 and CR2 in figure 5) provided low capacitance off-state impedance. Use of diode off-state reverse bias was not necessary, and use of a series feed diode in addition to the shunt diode assures the requisite, high crystal-to-ground off-state impedances are not degraded by the impedance of the printed wiring track connections to the switching circuits.

The switching circuit was designed so as not to utilize transistor saturation in the PIN diode on state based on the assumption that saturated transistor operation would likely result in higher diode bias current noise content.

CH2690-6/89/0000-075 $1.00 C 1989 IEEE

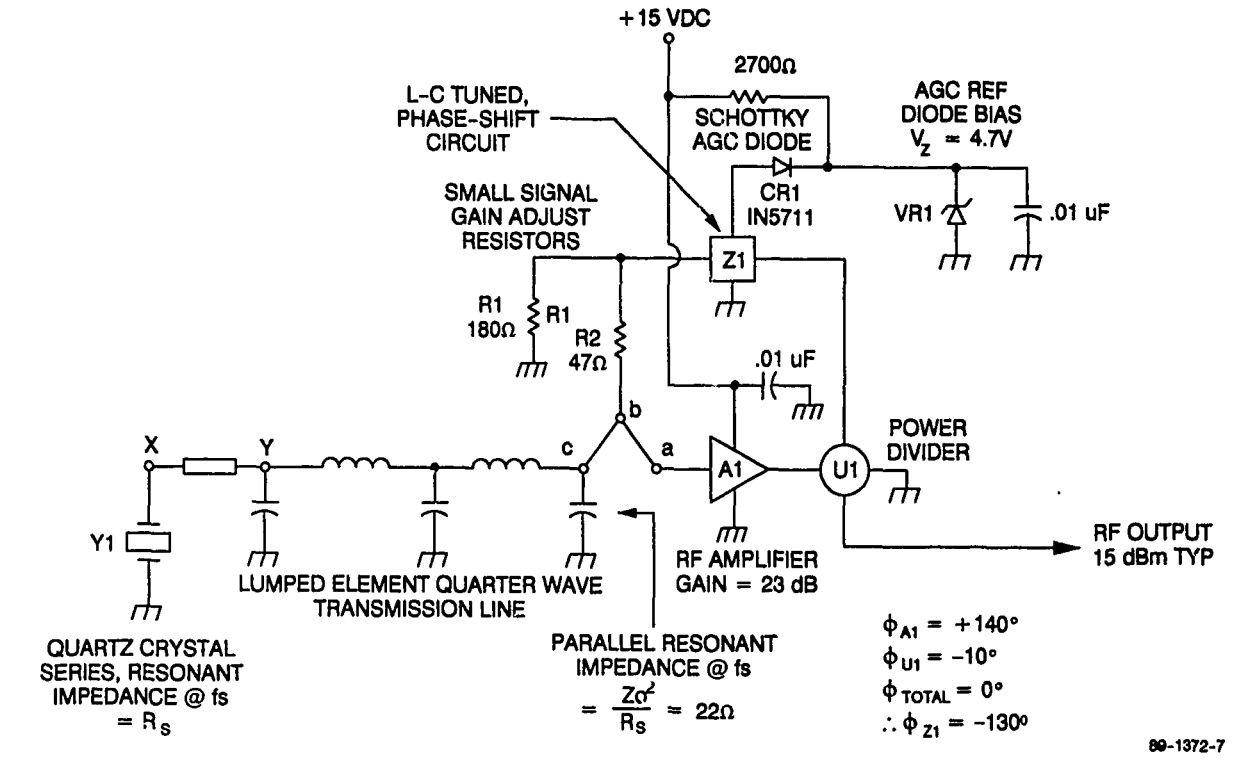


Figure 1. Basic Oscillator Circuit Employing Quarter-Wave Transformation of Quartz Crystal Impedance

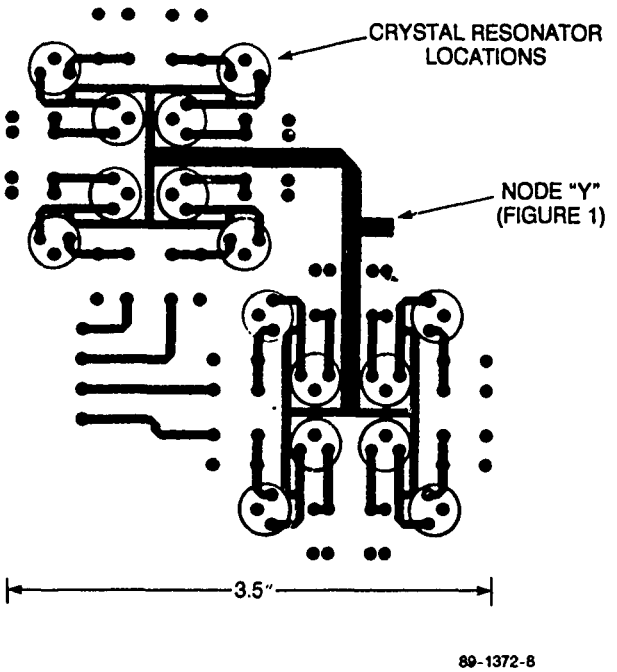


Figure 2. Crystal Connection Printed Board Layout

Phase Noise and Switching Time Performance

Figure 6 shows the results of measurement of oscillator phase noise, made by phase locking two oscillators, one of which has been modified for voltage tuning. The -165 dB/Hz noise floor level (-168 dB/Hz per oscillator) is simply determined by sustaining stage input drive level and noise figure.

Near carrier, flicker of frequency noise is characterized by $\mathcal{L}$ (100 Hz) = -125 dB/Hz (-128 dB/Hz per oscillator), corresponding to Sy (100 Hz) = 3×10^{-25}. Independent measurement of quartz resonator flicker noise confirms the fifth overtone, AT-cut crystals are the primary contributor to oscillator near-carrier spectral performance.

Figures 7 and 8 show the results of measurement of oscillator frequency switching time made using an HP5371A Frequency and Time Interval Analyzer. As shown in figure 8 (an expanded view of switching between channels) interchannel switching time is on the order of 2 to 5 milliseconds, including switching between largest frequency separation channels.

Conclusions

Utilization, in a multiple-crystal oscillator circuit, of a quarter wavelength (lumped element approximation) transmission line connection to the crystal resonators allows absorption of the parasitic effects of printed board crystal connection lines.

Because the microstrip connection lines can be absorbed into the circuit, operation at VHF, using large numbers of switch-selectable crystals is possible.

Utilization of series-shunt PIN diodes provides requisite, high off-state impedance with no measurable effect on output signal phase noise performance.

Reference

[1] M. M. Driscoll, "Low Noise Crystal Oscillators Using -50-Ohm Modular Amplifier Sustaining Stages," Proc. 40th Annual Frequency Control Symposium, May 1986, pp. 329-335.

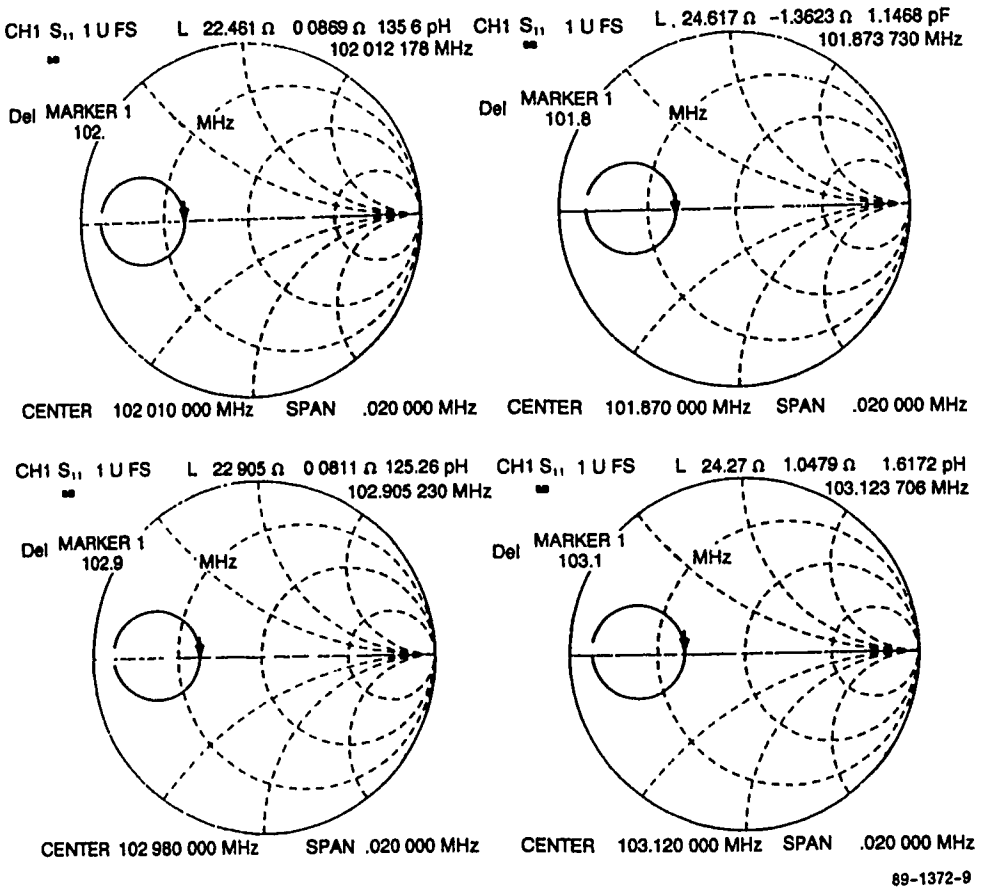


Figure 3. Measured Inverted Crystal Impedances for Four Crystals

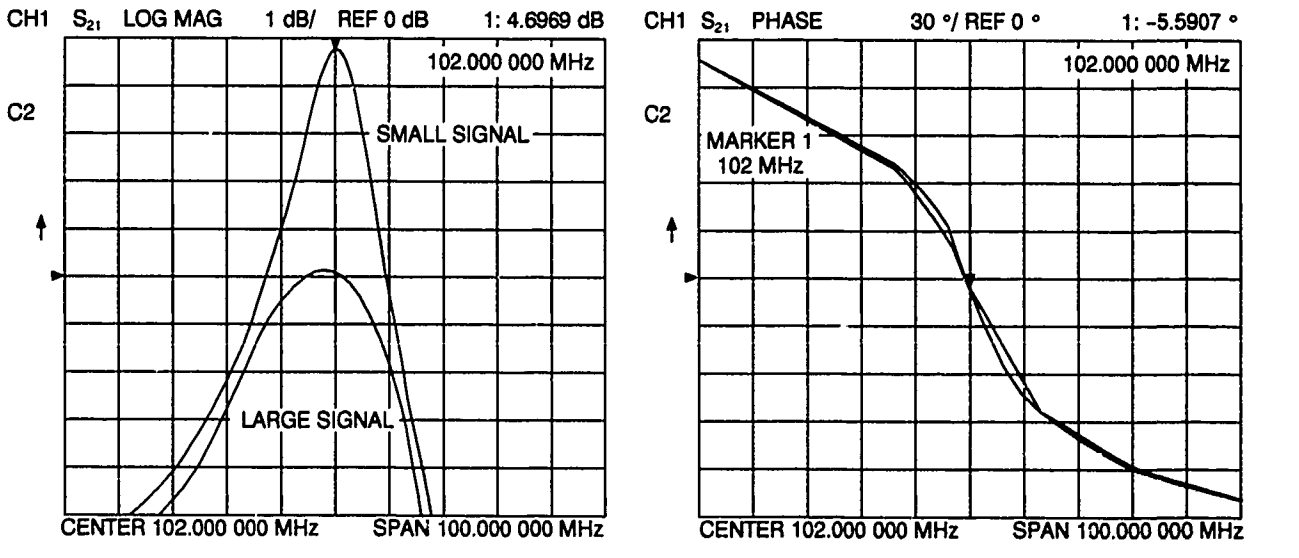


Figure 4. Measured Sustaining Stage Open Loop, Small and Large Signal Transmission Response

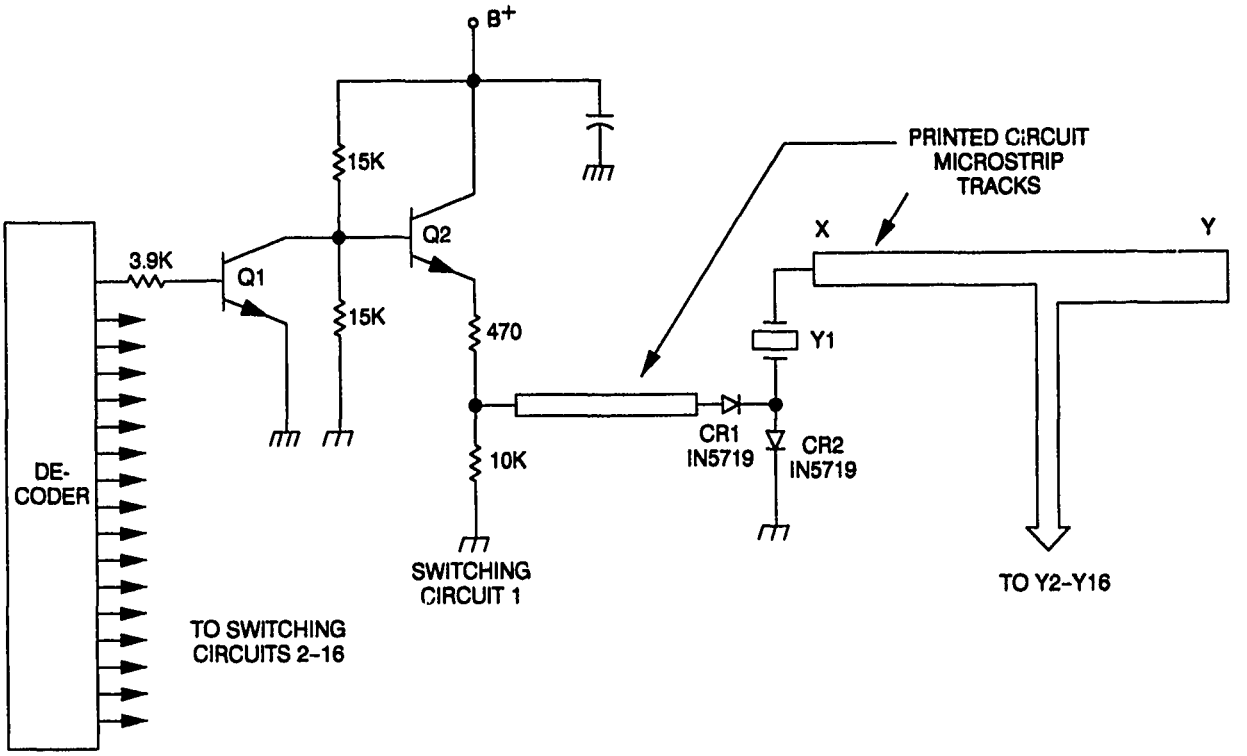


Figure 5. Crystal Resonator PIN Diode Switching Circuit

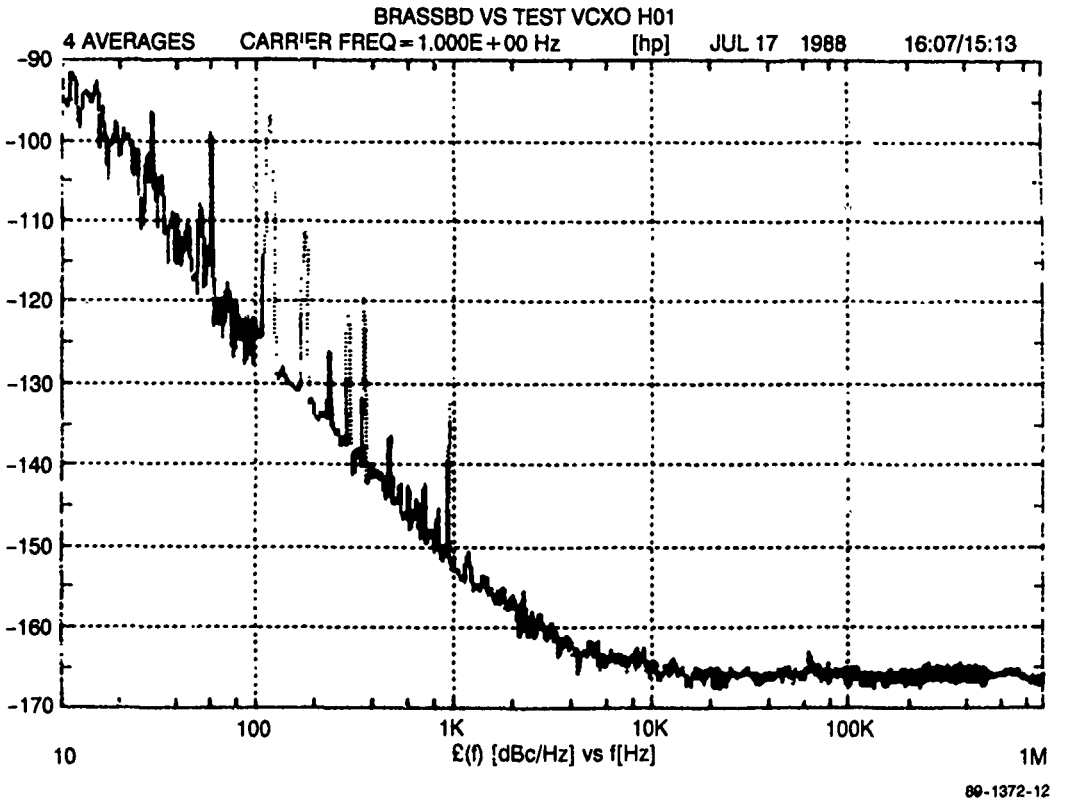


Figure 6. Measured (Two-Oscillator) Phase Noise Spectra

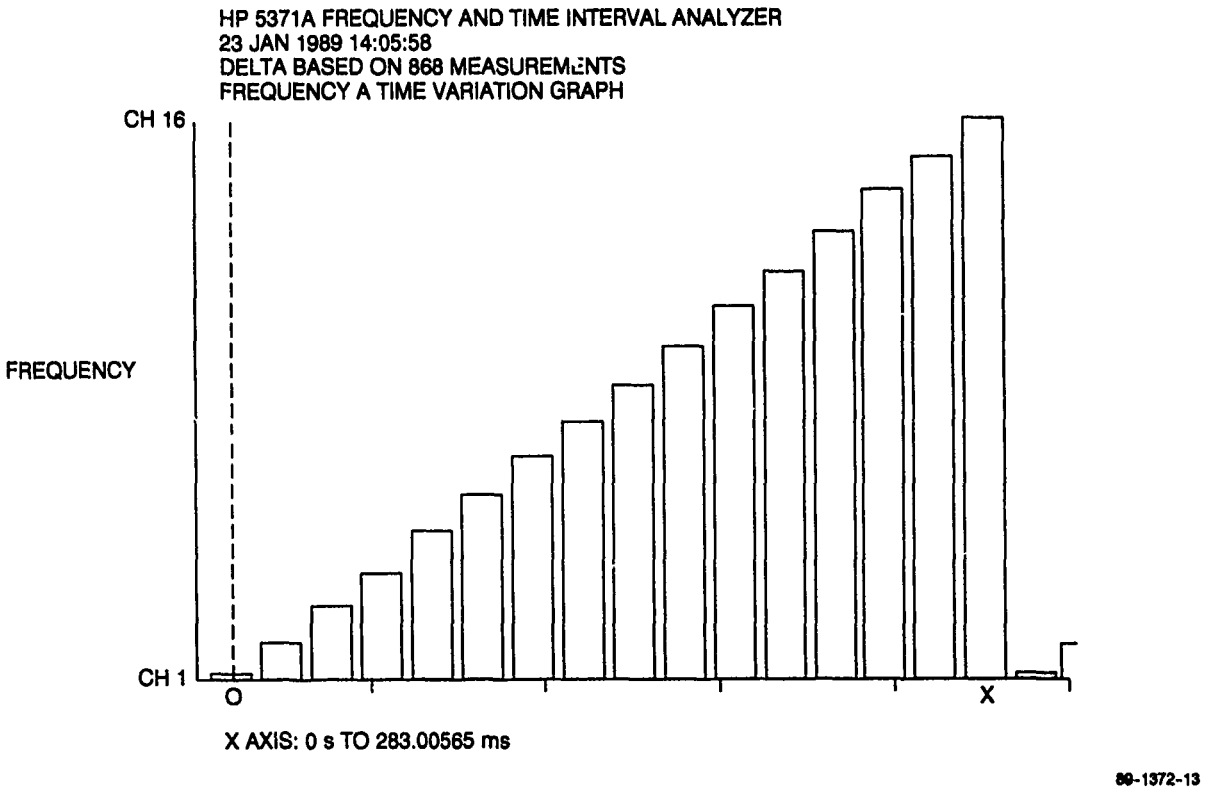


Figure 7. Measured Oscillator Switching Time (16 Channels)

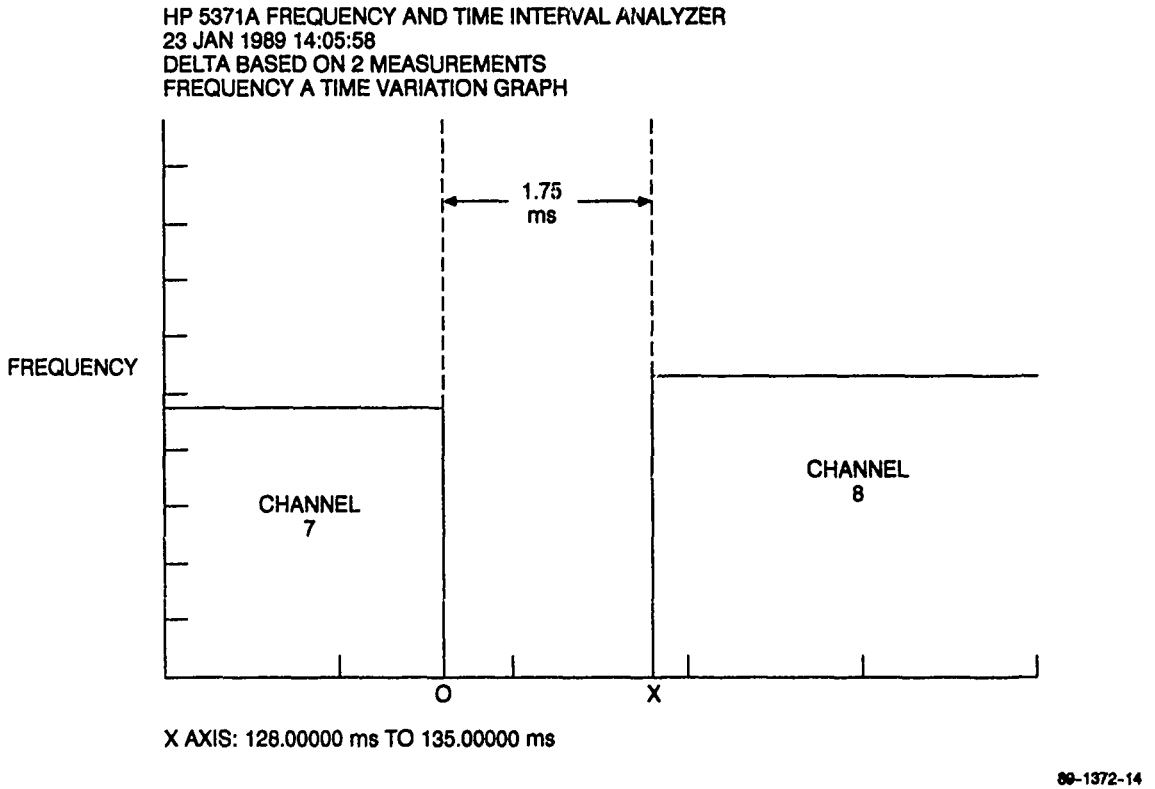


Figure 8. Measured Oscillator Switching Time Between Channels

43rd Annual Symposium on Frequency Control - 1989

STABILITY OF HIGH QUALITY QUARTZ CRYSTAL OSCILLATORS: AN UPDATE

M.B. BLOCH and J.C. HO, Frequency Electronics, Inc., Long Island, NY 11553

C.S. STONE and A. SYED, Brightline Corporation, Cedar Park, TX 78613

F.L. WALLS, National Institute of Standards and Technology, Boulder, CO 80303

ABSTRACT

Historically, quartz oscillator stability has benefited from ever-better understanding of oscillator circuit noise mechanisms, while quartz resonator design and fabrication have also made important advances, notably as represented in the contemporary, high-quality, low-anomaly, SC-cut, overtone-mode resonator units. Recently, two specially modified low-level, high-quality 5 MHz oscillators were tested for spectral purity and stability at the National Institute of Standards and Technology. Using a third, high-quality, prior-technology oscillator for triangulation the individual phase-noise power spectral density (PSD) of one of the oscillators was determined to be $S_\phi(f)$ = -133 db ±2 dB below 1 rad^2/Hz at a Fourier frequency of 1 Hz, while the second oscillator exhibited -125 dB ±2 dB at 1 Hz. Such oscillators can exhibit parts-in-10^{14} flicker floor stability in high- precision quartz, frequency-source applications. Extensive details of measurement methodology will be given.

PART I-PROGRESS IN QUARTZ OSCILLATOR STABILITY

Twenty-five years ago quartz-crystal-oscillator short-term stability near 4×10^{-12} [1] was exciting performance! In retrospect, however, we know that even the finest crystal frequency standards of that era were designed without regard to two important--but at the time unidentified--noise processes. The first of these came to light with the discovery that significant, direct, intrinsic, phase-noise modulation having a 1/f- or "flicker"-distributed power spectral density (PSD) signature is a universal property of transistors and other active devices---and that something can be done about it [2,3]. With this realization, the stage was set for a new generation of quartz frequency standards: by 1971, short-term stability improved by a hundredfold (in mean square terms) over the earlier figure was reported at that year's Frequency Control Symposium [4] as shown in Figure 1.

The next milestone was set in place when phase-bridge measurements showed that the phase of a carrier passing through a crystal resonator acquires significant random noise-modulation in transit; further, with the benefit of today's perspective we can say that this modulation appears at a level that assures that it will dominate the flicker-floor stability region of well-designed quartz oscillators [5-9]. The PSD signature suggests that it may be ascribed to flicker-noise modulation of the crystal's resonant frequency, which in turn may be modeled as microflicker of either of the crystal's motional reactances. By 1978 the phase bridge method had demonstrated flicker-floor stability

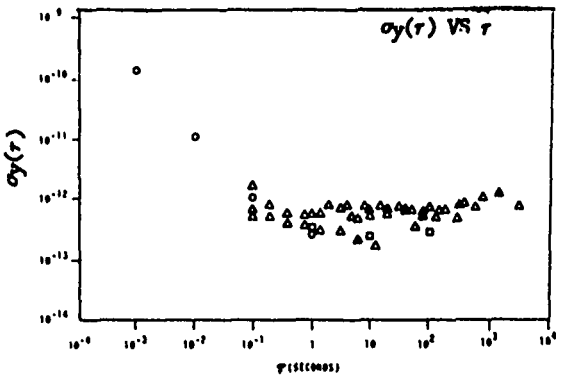


Fig. 1 Fractional frequency stability, $\sigma_y(\tau)$, of high-quality quartz oscillators from Brandenberger et. al. 1971 [4]. The measurement bandwidth was 1 kHz for the squares and circles and about 25 Hz for the triangles. The squares and circles data have some dead time.

slightly below the 10^{-13} level in a commercial 5 MHz resonator. However, achieving the same performance in an active oscillator continued to be elusive [6]. It is easy to believe that this is due, at least in part, to fleeting, transient activity-dips [10-12] that are known to infect virtually [10-12] all earlier designs of AT-crystal resonators. Such resonators include those high-precision overtone types designed for the lower end of the HF spectrum which have long been associated with superior stability, both long-term and short-term. It may be noted that activity-dips, or "band-breaks", are usually believed to be the result of intrinsically noisy partitioning of energy, via nonlinear elastic pumping, between the desired mode and a nearby inadvertent secondary mode of the resonator. This harmonic syntonization often occurs only briefly as a complex accident of temperature, excitation level, tuning, stress, etc. The effects of activity-dips can range from totally disabling the oscillator to having a barely perceptible effect on frequency, crystal resistance, etc. over a few hundredths of a degree span of resonator temperature---except where resonator noise is concerned. Even virtually imperceptible activity-dips may degrade crystal noise by more than 3 dB. Activity-dip noise provides the most likely explanation for the non-stationary quality of Allan variance flicker-floor that is sometimes observed in otherwise well-behaved oscillators.

In this context the prediction of the Stress-Compensated (SC) resonator in the middle seventies and its later debut as a high-stability, high-precision, quantity production component must be considered a pivotal development [13-15]. With the Q of SC-cut resonators generically improved over the Q [13-15] of the workhorse AT-cut devices by approximately 15%, the SC crystal could promise a modest phase-noise superiority of about 2.5 dB. This expected advantage (not at all unique to the SC) as it has turned out has been far overshadowed

by other predicted--and realized--qualities of SC-cut behavior: some of these, e.g. insensitivity to thermal shock, often make the SC-cut devices an attractive choice for reasons not immediately related to phase noise. However, it is the relative freedom of the practical SC-cut resonator from activity-dips at high excitation level that differentiates the SC from the AT, BT, and virtually all other quartz resonator types, and raises the hope of realizing the full potential of quartz crystals for uniformly reproducible short-term stability.

Oscillator Improvement Program

As part of a continuing program of improving products through in-house technology insertion, Frequency Electronics, Inc. recently tasked Brightline Corp. to upgrade the high-precision quartz oscillator used as the time base generator subassembly of a disciplined time-frequency standard. This 5 MHz source represented the state-of-the-art at the beginning of its model life, which was well before the advent of SC resonators. Accordingly, the improvements were to include retrofitting the oscillator with a premiere-grade-production, 5-MHz, fifth-overtone, SC resonator plus installation of other modifications required to bring the oscillator system to a fully contemporary configuration.

Phase-Noise Results

While not all contemplated improvements have yet been made, preliminary phase-noise testing of modified high performance oscillator units indicates that a high order of short-term stability has been attained [16]. The best phase-noise results obtained thus far at Brightline are shown in Figure 2. Converting the flicker-frequency component of phase-noise density ($\mathcal{L}(f)$ ~ [-134 - 30log(f)] dBc) obtained from Figure 2 to the corresponding two-sample, time-domain-stability, yields flicker-floor $\sigma_y(\tau)$ near 6.6×10^{-14}. However, we have also observed that a random-walk component is evident in some (but by no means all) of the phase-noise records extending to f=0.1 Hz and involving several oscillators and crystals. At its typical level ($\mathcal{L}(f)$ approximately [-140 - 40log(f)] dBc), the random-walk contribution would match the flicker-floor at a sampling time of only 3.6 s, and parts in 10^{14} stability could be expected in direct measurements only for a narrow range of sampling intervals near 1 s, if at all. The origin of this random-walk behavior has not yet been identified; however, we do not believe that it is due to excessive crystal current or any intrinsic property of the crystal. Hopefully, further investigation will lead to extending the flicker-floor dominance, that we now see in the vicinity of 1 Hz, out to the equivalent of multi-day sampling, which recently has been demonstrated by other high-quality oscillators employing AT-crystals [17].

Future Developments

The very short-term time-domain performance can readily be enhanced by using an existing-art, low-noise, crystal postfilter. In this case the wide-band noise would be expected to be reduced by approximately 20 dB. The fractional, time-domain frequency stability under these conditions should be approximately $\sigma_y(\tau) \sim 1.6 \times 1^{-15}\ \tau^{-1}$. If the reduction in random-walk frequency modulation can be achieved and the postfilter added without degrading

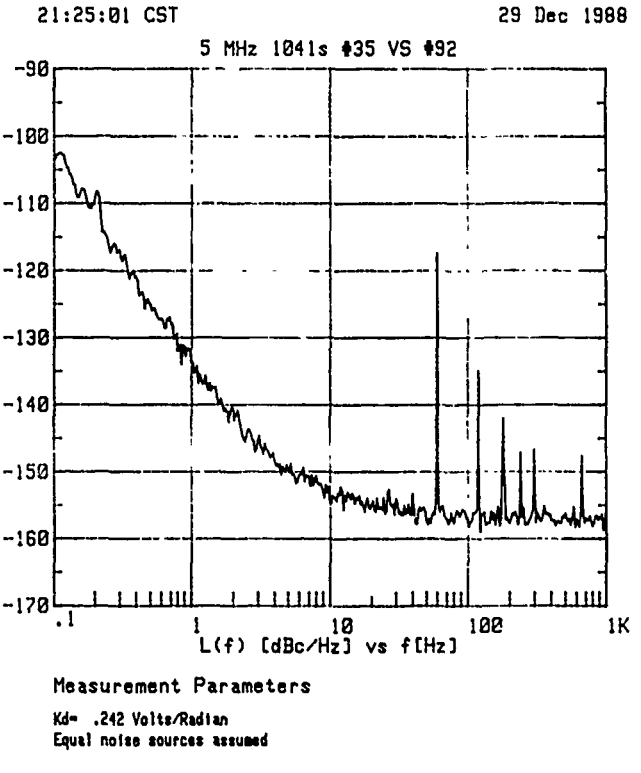


Fig. 2 Phase-noise PSD of improved oscillators [16].

the present flicker performance, parts in 10^{14} stability could be realized from approximately 20 ms to several days. Given the experience of data compiled on a pilot quantity of crystal units showing frequency domain performance that implies flicker-floor performance of parts in 10^{14} in a number of units, we can look forward with cautious optimism to the production of quartz oscillators with stability of a few parts in 10^{14}.

Two selected high-performance oscillators [16] were furnished to NIST to assist in a calibration; Part II of this paper describes experimental results obtained, and the methodology of the NIST measurements.

PART II- DESCRIPTION OF PHASE NOISE MEASUREMENTS

Very precise phase-noise measurements between three different oscillators to obtain unbiased estimates of the phase noise of each oscillator were made at NIST. Relative precision of ± 0.2 dB and accuracies of ± 0.6 dB were necessary to obtain reliable results because one oscillator (92) had substantially better phase noise than the others. The general block diagram of the measurement system is shown in figure 3. First, a precision NIST noise standard was used to check the spectral density function and the internal voltage reference of the FFT [18]. The difference between the voltage PSD measured on the FFT and the value independently determined from first principles with our noise standard was less than 0.01 dB from 20 Hz to 20 kHz (see figure 3).

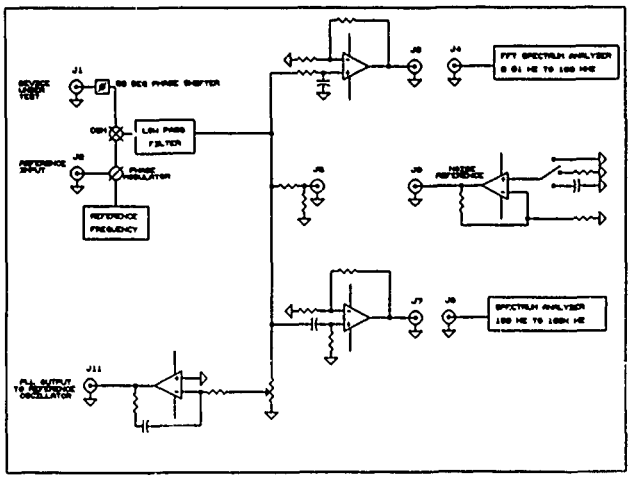


Fig. 3 Block diagram of NIST phase noise measurement system [16].

Second, the sensitivity of the mixer and amplifier for converting small phase deviations from phase quadrature to voltage was determined for each oscillator pair at a beat frequency of about 0.12 Hz. The FFT was used to digitize and analyze the beat waveform. The time base was set to 10 s to accurately determine the beat period and then reduced to 0.2 s to determine the slope through the zero crossing, GK_d, in volts per radian. Both the negative- and positive-going zero crossings were measured. The difference between the two slopes was less than 10%, which indicated that there was no significant injecting locking at this frequency and that the mixer was symmetric, that is, that there were no bad diodes. The accuracy for the determination of the mixer sensitivity and amplifier gain was typically ± 1%, which corresponds to ± 0.1 dB in the measurement of $S_\phi(f)$, the spectral density of phase fluctuations for the pair.

Third, the frequency dependency of both the amplifier following the mixer and the action of the phase lock loop was determined using an ultra-flat phase modulator. This modulator produces phase-modulation side bands on the carrier which are demodulated by the mixer to produce a reference signal which is constant in amplitude to better than 0.2 dB from dc to well beyond 200 kHz as shown in figure 4 [19]. By sweeping the frequency of the

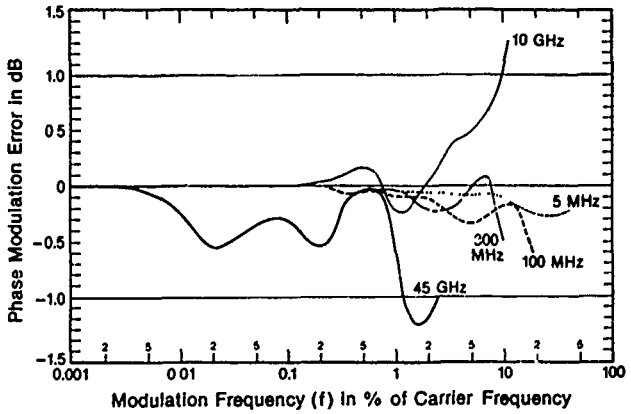


Fig. 4 Error in the phase modulation reference signal as a function of the modulation frequency [16].

modulation over the entire range of measurements, one can correct the measurement system for all baseband frequency dependent effects [19]. The corrections due to the phase lock loop were typically -1.3 dB at 0.2 Hz and -0.24 dB at 1 Hz. Corrections at all higher frequencies in these measurements were small compared to 0.1 dB and neglected.

Fourth, the phase noise was measured with the FFT using the Hanning window. Generally, 1000 samples were taken. The rms (67%) fractional statistical confidence interval of these spectral estimates can be calculated using Table 1 to be approximately 1 ± 0.032 which corresponds to ± 0.14 dB [17].

Table 1. Confidence Intervals for FFT Spectral Estimates

power law noise type	window uniform	Hanning	flattened peak
f^0	$1.02/\sqrt{N}$	$0.98/\sqrt{N}$	$0.98/\sqrt{N}$
f^{-2}	$1.02/\sqrt{N}$	$1.04/\sqrt{N}$	$1.04/\sqrt{N}$
f^{-3}	unusable	$1.04/\sqrt{N}$	$1.04/\sqrt{N}$
f^{-4}	unusable	$1.04/\sqrt{N}$	$1.04/\sqrt{N}$

Table 1. Fractional confidence intervals for several FFT spectral density estimators as a function of noise type and N, the number of independent samples. $S = S_m(1 \pm \frac{V(w,f}{\sqrt{N}})$ where S is the time spectral density, S_m is the measured spectral density and $\frac{V(w,f)}{\sqrt{N}}$ is the appropriate table entry.

Figure 5 shows $S_\phi(f)$, for oscillator pair 483 and 92 with all corrections applied. Additional measurements against oscillator 8600 are shown in figure 6. These and other measurements were used to obtain the following estimates of $S\phi(f)$ for oscillators 92 and 483:

$$S\phi_{92}(f) = 10^{-13.31}f^{-3} + 10^{-13.6}f^{-1} + 10^{-15.46}$$

$$S_{\phi 483}(f) = 10^{-12.46}f^{-3} + 10^{-13.6}f^{-1} + 10^{-15.46}$$

The coefficients for the power law spectra of $S_\phi(f)$ for oscillator 92 where then used in a new NIST software program SIGINT, which calculates the fractional frequency stability, $\sigma_y(\tau)$ (also Mod $\sigma_y(\tau)$), using the primary definition

$$\sigma_y(\tau) = \left[\int_0^{f_h} 2 \frac{f^2}{\nu^2} S_\phi(f) \frac{\sin^4 \pi f\tau}{(\pi f\tau)^2} \, df \right]^{1/2}.$$

Here, f_h is the noise bandwidth of the $\sigma_y(\tau)$ measurement system, and ν is the carrier frequency, i.e., 5 MHz. The accuracy of the SIGINT software has been extensively checked and the errors are less than 3%. The results for oscillator 92 are shown in figure 7 for measurement bandwidths of 100 Hz and 1 kHz. Also shown is time-domain stability of oscillator 92 derived from direct measurements of $\sigma_y(\tau)$ made between the three oscillators. Unfortunately, the noise floor of our present time-domain measurement system precludes verifying the time domain stability of oscillator 92 at 1 s. By the time the measurement floor is low enough, random-walk has degraded the stability above

1×10^{-13}. Oscillators 92 and 483 were installed in separate closed containers to stabilize the humidity and pressure. The random-walk appeared to decrease by about a factor of 2. Earlier work on oscillator 8600 had also demonstrated significant improvements in the random-walk performance when the humidity was controlled [17]. These results suggest that humidity and possibly temperature play a role in the random walk process.

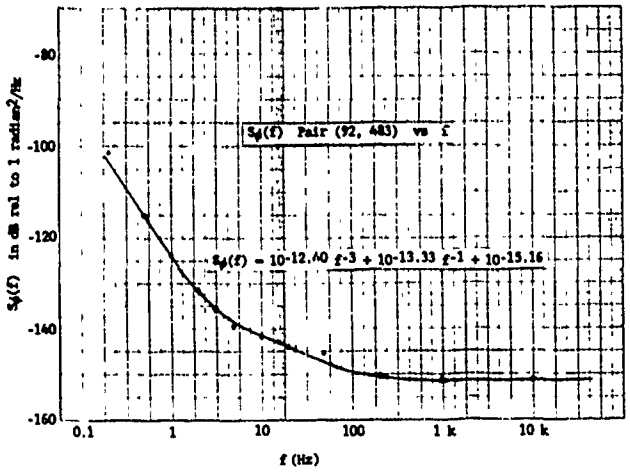


Fig. 5 Spectral density of phase fluctuations, $S_\phi(f)$ of oscillator pair 483-92 as a function of Fourier frequency offset.

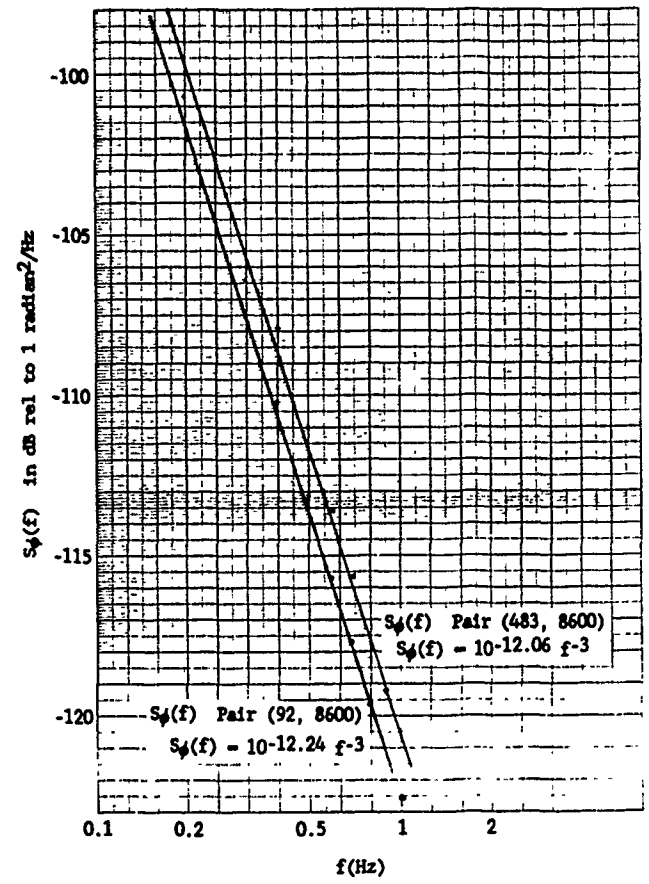


Fig. 6 Spectral density of phase fluctuations, $S_\phi(f)$ of oscillator pairs 483-8100 and 92-8600 as a function of Fourier frequency offset.

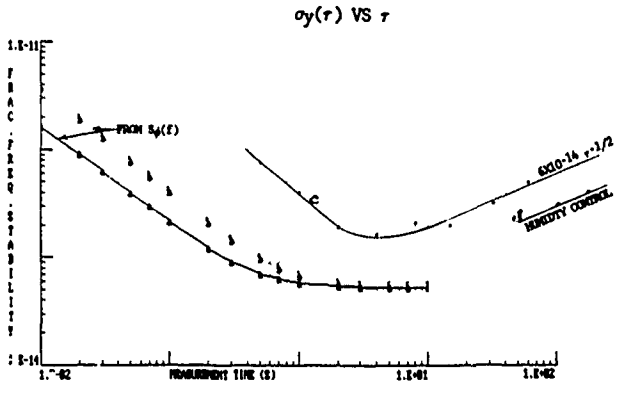


Fig. 7 Fractional time domain frequency stability, $\sigma_y(\tau)$, of oscillator 92 derived from the phase noise, curves a and b. Curve c shows the initial measured time domain stability. Curve d shows the approximate effect of stabilization the ambient pressure and humidity on the logner-term frequency stability.

CONCLUSION

We have described a little of the history of ultra-low flicker phase noise in quartz crystal oscillators. We have also described measurement techniques for measuring the phase noise of oscillators pairs with a relative precision of ± 0.2 dB and accuracy of ± 0.6 dB. These techniques were then applied to the measurement of a set of three oscillators. By unfolding the results of the measurements, we have demonstrated that one of them (oscillator 92) had exceptionally low flicker noise, $S_\phi(1\ \text{Hz}) = 133 \pm 2$ dB. The fractional, time-domain frequency stability, $\sigma_y(\tau)$, calculated from the measured phase noise yields $\sigma_y(1\ \text{s}) = 5.2 \pm 1.3 \times 10^{-14}$. The relatively high level of random-walk frequency modulation (6×10^{-14}) thwarted our attempts to directly confirm this low flicker level.

REFERENCES

[1] H.P. Stratemeyer, "The Stability of Standard-Frequency Oscillators," The General Radio Experimenter, 38, 1-16 (1964).

[2] D. Halford, A.E. Wainwright, and J.A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure," Proc. of the 22nd Annual Symposium on Frequency Control, 340-341, 1968.

[3] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum, " Proc. of the IEEE, 54, 329-330, February 1966.

[4] H. Brandenberger, F. Hadorn, D. Halford, and J. H. Shoaf, "High Quality Quartz Crystal Oscillators: Frequency Domain and Time Domain Stability," Proc. of the 25th Annual Symposium on frequency Control, pp. 226, 1971.

[5] A.E. Wainwright, F.L.Walls, and W.D. McCaa, "Direct Measurements of the Inherent

Frequency Stability of Quartz crystal Resonators," Proc. of the 28th Annual Symposium on Frequency Control, pp. 177, 1974.

[6] F.L. Walls and S.R. Stein, "A Frequency-Lock System for Improved Quartz Crystal Oscillator Performance," IEEE Transactions on Instrumentation and Measurement, IM-27, pp. 249-252, September, 1978.

[7] F.L Walls and A.E. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications," IEEE Transactions on Instrumentation and Measurement, IM-24, pp. 15-20, March 1975.

[8] 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," Proc. of th 35th Annual Symposium on Frequency Control, pp. 476-483, 1981.

[9] T. E. Parker, "Characteristics of Phase Noise in Stable Oscillator," Proc. 41st Annual Symposium on Frequency Control, 1987, pp. 99-110.

[10] J.J. Gagnepain, M. Oliver, and F.L. Walls, "Excess Noise in Quartz Crystal Resonators," Proc. of the 37th Annual Symposium on Frequency Control, pp. 218-225, 1983.

[11] A. Ballato and R. Tilton, "Electronic Activity Dip Measurement, IEEE Transactions on Instrumentation and Measurement, IM-27, pp. 59-65 March 1978.

[12] A. Ballato, T.J. Lukaszek, and G.J. Iafrate, "Subtle Effects in High Stability Vibrators," Proc. of the 34th Annual Symposium on Frequency Control, pp. 431, 1980.

[13] E.P. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proc. of the 29th Annual Sympoisum on Frequency Control, pp. 1-4, 1975.

[14] E.P. EerNisse, "Calculations on the Strees Compensated (SC-Cut) Quartz Resonator," Proc. of the 30th Annual Symposium on Frequency Control, pp. 8, 1976.

[15] J.A. Kusters and C.A. Adams, "Production Statistics of SC (or TTC) Crystals, Proc. of the 34th Annual Symposium on Frequency Control, pp. 167, 1980.

[16] These are FF1041 units with FEI 5th-overtone SC-Cut resonators.

[17] F.L. Walls, "The Influence of Pressure and Humidity on the Medium and Long Term Frequency Stability of Quartz Oscillators," Proc. of the 42nd Annual Symposium on Frequency Control, pp. 279-283, 1988.

[18] F. L. Walls, D. B. Percival, and W. R. Irelan, "Biases and Variances of Several FFT Spectral Estimates as a Function of Noise Type and Number of Samples," these proceedings.

[19] F. L. Walls, A. J. D. Clements, C. M. Felton, M. A. Lombardi, and M. D. Vanek, "Extending the Range and Accuracy of Phase Noise Measurements," Proc. of the 42nd Annual Symposium on Frequency Control, 1988, pp. 432-441.

PULSED, COHERENT LOW POWER VHF/UHF TRANSMITTERS - OF MINIMAL COMPONENT COUNT & WEIGHT

Ian J. Dilworth B.Sc. Ph.D.,
Department of Electronic Systems Engineering,
University of Essex, Wivenhoe Park,
Colchester, CO4 3SQ, Essex, England.
Fax: +44 206 873598 Telex: 98440 (UNILIB G)

Abstract

The parameters which directly influence the design of stable, coherent and pulsed *VHF/UHF* sources are identified. Assuming currently available commercial technology, an attempt is made at identifying the optimum frequency , for maximum range, at which systems used for animal tracking should operate. The current state of the art of pulsed *VHF/UHF* transmitters, using Quartz crystal resonators, of minimum weight and volume is reviewed together with measurements made on simple oscillator and frequency multiplier circuitry. Commercial chemically etched fundamental and overtone quartz crystal oscillators in the *VHF* frequency band have only recently become available and the performance of these units are compared with those available from low effective series resistance *(ESR)* quartz crystal oscillators operating at third overtone in the 60-100MHz region and frequency multiplication factors of 3 - 8. Finally areas are identified where further research and development are required to meet the requirements of this application.

Introduction

In many animal conservation studies there is a requirement for some form of location and tracking system, for example see reference (1) and (2) For roaming terrestrial vertebrates this can conveniently and practically take the form of a radio frequency transmitter attached to the animal Ideally the transmission needs to be powerful, continuous and of indefinite longevity. For large vertebrates, such as Bears and Rhinoceroses there is no problem posed by the weight of the transmitter or with size requirements and therefore to a first approximation the aforementioned goals can be met. This is not the case with smaller animals. In particular Birds present a difficult problem and a radio frequency transmitter for almost all species of bird needs to be very lightweight - of the order of a few grams, including the power source - and small in size - including the radiating element.

Therefore this particular application requires a highly efficient (i.e. conversion of DC power to RF power) *VHF/UHF* stable frequency source whose short term stability needs to be of the order of a few tens of Hertz, together with an equally efficient radiating structure and a power source, most likely a chemical battery, of high power to weight (and volume). There exist applications in tracking fish (2), mammals which live underground and indeed insects (3). In addition the telemetry of biological and/or environmental data may be required. Such applications are not directly addressed in this paper although some of the basic conclusions remain valid.

Systems considerations

From the operational view point the most important consideration is the maximisation of the range over which the transmitted signal may be detected, or put another way a large received carrier to noise (C/N) which will allow rapid unambiguous location of the transmitter. This in turn depends on all of the following:

(i) *frequency of operation i.e. the inverse square law.*
(ii) *size and efficiency of the transmitter and receiver antennas.*
(iii) *power output of the transmitter.*
(iv) *longevity and power output of the transmitter power source e.g. battery and /or a solar panel.*
(v) *detection bandwidth - which depends on stability of transmitter (and receiver) oscillators.*

Practical constraints impose limits on the size of the transmitter (and receiver) antennas. The transmitter antenna may be a short quarter wave monopole for example - to give quasi omnidirectional coverage or indeed reduced in length somewhat by loading(4). For birds such an antenna limits the minimum usable frequency to about 200MHz because of its length. Conversely the upper frequency limit depends on (iii) the realisable power output of the transmitter. In practice, using present technology (including power sources), frequencies between 200 and 800MHz are feasible with the 'optimum' frequency being around 400MHz when judged from practical tracking considerations. Therefore in this transmitter application, for stable operation, high frequency overtone quartz crystals are an obvious candidate when used in association with frequency multiplication stage(s). A brief review of (iv), battery power sources appears later.

The portable receive antenna is limited in aperture size by practical considerations and by the fact that since the transmitter employs a slow - pulsed format an excessively narrow azimuthal beamwidth can make searching for a weak transmission very difficult and time consuming. More important is a high front to back ratio and low sidelobe levels.

In a terrestrial system the minimum possible detectable signal level is theoretically limited to -174dBm/Hz by the black-body background radiation of approximately 290 Kelvin corresponding to the average temperature of the earth. Very low noise receivers can be produced which hardly degrade the minimum theoretical achievable sensitivity. However in order to produce a transmitting source which remains energetic for several days (if not weeks) a low duty cycle has to be employed whereby the transmitter is *pulsed* once every few seconds, thus reducing battery drain. This is necessary because the available battery technologies can only provide limited power /weight ratios and indeed fundamentally limit the transmitter output power since the maximum weight of the whole transmitter assembly must not exceed a few grams. e.g. even 'large' birds of prey are limited to approximately 6-10 grams. In addition the transmitter requires a stable oscillator source in order that the receiver bandwidth can be made as small as possible thereby allowing narrow noise bandwidth reception and increased sensitivity. Typically short-term drift limits the extent to which the bandwidth can be reduced while long-term temperature drift is only a problem if it is not predictable. The relationship between achievable sensitivity, receiver noise figure and detector noise bandwidth are summarised in figure (1). The aim is to achieve sensitivities below about -140dBm. This is most readily achieved using heterodyne demodulation and narrowband audio phase locked detector. In most operational instances the transmitter and (portable) receiver will experience approximately the same external temperature and environment. The effects of long term i.e. diurnal transmitter frequency drift can be partially cancelled by employing 'conjugate' temperature characteristics in the receiver local oscillator assuming some form of superheterodyne receiver system is employed. Longer drifts due to aging etc. have to be compensated by receiver local oscillator adjustments.

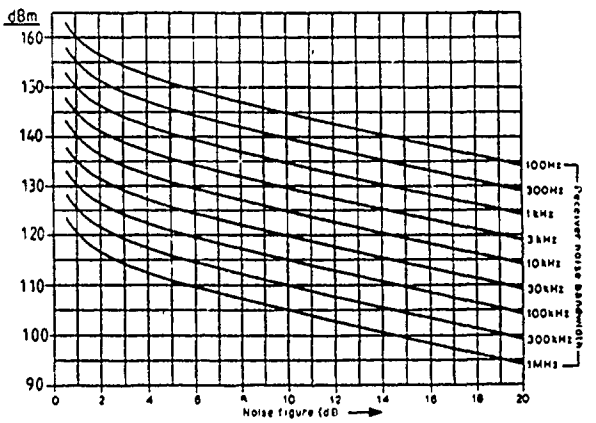


Figure 1 - Receiver sensitivity vs noise figure and bandwidth

The 'optimum' frequency to use?

A 'parametric' study was performed in order to try and identify the 'optimum' frequency to use in a *terrestrial* animal tracking system assuming current readily available technologies for the transmitter (and receiver) are employed. In most areas of the world discreet bands of frequencies starting at 27MHz and upward are available for the use of 'telemetry' systems so a relatively free choice of frequency is assumed. The results of calculations *and* measurements are shown in table(1) below. In this study, the path is fixed at 10 miles (16km), the receive antenna gain is based on a equispaced

CH2690-6/89/0000-085 $1.00 C 1989 IEEE

element Yagi design of fixed boom length, the transmitter power is that measured using oscillator-multiplier and direct generation using the circuitry described in this paper, the receive antenna height reduction is based on measurements made and assuming that the maximum height of the antenna above dampened ground is about 5 feet (1.52 metres) and finally the recovered C/N depends on the realisable receiver bandwidth, here assumed to be -134dBm. Line of sight is assumed and the effects of propagation, antenna cross-polarisation the height and orientation of the transmitter assembly are ignored although all of these factors serve to reduce the available C/N.

Table 1

Frequency MHz.....	200	300	400	500
Free space loss (dB over 10 miles).	-103	-107	-109	-111
Receiver Antenna gain (dB).....	+5	+6	+8	+9
Effective transmit power (dBm)....	-2	-8	-10	-13
Antenna height reduction (dB).....	-10	-8	-5	-4
Total C/N (dB).....	24	17	18	15

Discussion

If the frequency of operation doubles the inverse square law introduces 6dB further loss, on the plus side a Yagi antenna (of fixed boom length) produces a realisable 3dB gain improvement assuming the boom is filled with elements and receiver antenna height reduction effects result in a 3dB improvement with doubling of frequency. Conversely, for the same frequency increase, current realisable oscillator/multiplier systems produce approximately an 8dB reduction in power level. The overall result is that a relatively broad peak in the realisable carrier to noise (C/N) is available if frequencies between about 250-450MHz are employed. At frequencies below 250MHz higher C/N's are available, essentially because of improved transmitter efficiencies, but useful directive and efficient receive antennas become rather cumbersome. Above 500MHz the available C/N drops off fairly rapidly.

Crystal and electronics technology

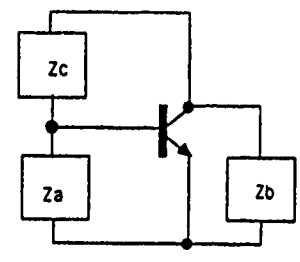
Because available chemical battery power sources all suffer from a significant diminishing voltage during their useful life, the essential requirements of the transmitter are that only minimal or negligible change in output frequency and power should occur with diminishing supply voltage. Similarly 'chirp '(or dynamic frequency excursions) should not increase and temperature induced frequency change should be smooth and predictable. Significant crystal aging is to be avoided but long term drifts in frequency can be adjusted in the receiver. For this application an integrated quartz and oscillator assembly plus frequency multiplier is an obvious choice. However some form of hybrid will be required because printed inductors at the frequencies of interest are best realised using lumped elements. These should be realised using toroidally wound components in order to minimise the effects of stray fields which interact with a metal container/housing, reducing the effective inductance, and make tuning and testing difficult. Unfortunately the quantity required to economically realise such a product is not required and therefore this technology is not presently a viable proposition. In this application the conformality of the electronics, with the crystal and the power source, in order to realise the smallest possible transmitter is of paramount importance. In fact the battery technology virtually dictates the shape and size of the transmitter assembly together with the available quartz crystal packages i.e. presently limited to HC44 and HC45. Surface mounted crystals may help in this respect but current packaging for this construction actually increase the transmitter overall size compared to that using conventional discreet quartz packaging and the associated (surface mounted) components. *VHF* quartz crystal units (16) with blanks of only 5mm diameter (5) are now becoming available which could significantly reduce volume requirements for the transmitters. However some manufacturers are reluctant to produce these units because of poor aging performance. The 'inverted mesa' process used to produce the *AT* cut *VHF* crystals, used in this investigation, have a very small volume in which the drive level energy is dissipated If the drive level exceeds 0.3mW then a 'pseudo drive level effect ' occurs heating up and expanding the resonant volume in the ratio 1 11 140 respectively for a 70MHz 3rd overtone, 70MHz fundamental and a 210MHz fundamental. The *VHF* crystal may also manifest a secondary effect due to the anisotropic heat transfer resulting in internal stresses because of the irregular shape of the heated volume (16). It is therefore to be expected that oscillator design using these VHF crystals needs special care.

Battery chemistries

Large power-weight ratios are required of the power sources used for this application. The maximum power-weight ratio should also coincide with minimum volume. Physically small power sources are available in many chemistries (10,11,12) such as: Alkaline, Silver Oxide, Mercury oxide, Zinc-Air, Lithium-Manganese dioxide and Lithium-carbonmonoflouride, Lithium-thionylchloride, rechargeable Nickel-Cadmium and Solar cells . In most applications Solar cells have only limited usefulness, their size and operating requirements being a major constraint. Battery chemistry is improving all the time although not all types (including the most attractive) are readily available. Some chemistries, such as Zinc-Air, do not work at low temperatures and require air to so they are ruled out for some applications although there energy/weight ratio is the best. Lithium chemistries offer good performance except in respect to volume. Silver oxide or perhaps mercury based cells however represent a good working compromise (13). All the batteries exhibit similar voltage versus current drain characteristics. Initially all cells produce a voltage which rapidly decreases, until reaching a plateau region, here the voltage remains essentially constant until the end of useful life when it drops off very rapidly. Some chemistries are better than others in this respect. Perhaps the most stable is silver oxide (14) nevertheless the transmitter electronics needs to be able to cope with beginning of life voltages of approximately 4.9Volts and at the end of life 2.5Volts or less (assuming three cells in series). A measured current, voltage and power output characteristic, using three Silver Oxide batteries is shown in figure (5). As mentioned conformality with the electronics (including the quartz crystal package) and the animal is required. For available battery shapes see (15). In short no ideal solution exists. In practice present battery technology allows transmitters of about 5-8 gram weights to operate for 5-7 days at usable power levels.

Oscillator configuration and operation

The oscillator chosen for this application is essentially a modified Pierce (harmonic) configuration (6,7,17). There are several reasons for this choice, the overriding one being governed by the necessity to operate at low Vcc which, in this pulsed mode of operation, requires the maximum circuit Q a condition provided by the Pierce configuration because it presents a reactive load to the crystal (8). The feedback configuration, depicted in figure (2), and operation of this oscillator is well known and fully discussed in the quoted references.



PIERCE AC REPRESENTATION

Figure 2 - Oscillator feedback

Referring to the diagram Za and Zb should normally be minimised, however for maximum output power, operating Q and power this condition cannot be met. In our case this implies that the load cannot be exactly resonant. In terms of the *DC* conditions: increasing the supply (Vcc) and the base bias increase stability but we require that these parameters be minimised in this application. The requirements of the *AC* oscillator operating conditions, to achieve the required coherence are relatively critical and they depend to a large extent on the characteristics of the quartz crystal. A minimum component count *R-C* base pumping network produces a periodic

base current surge which must be sufficient to produce a burst of oscillation completely controlled by the (overtone) crystal. The basic oscillator with simple frequency multiplier employed is shown in figure (3).

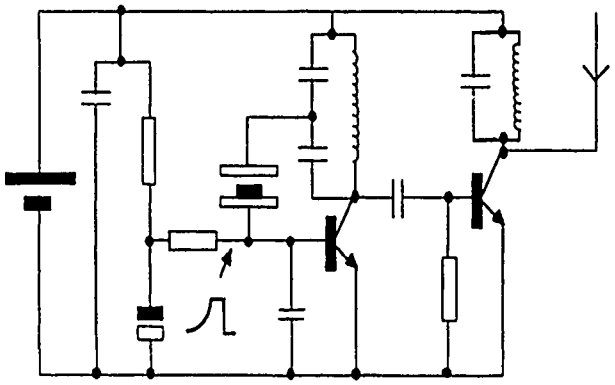


Figure 3 - Pulsed oscillator schematic

The oscillator feedback network is tuned to near resonance (see below) at the overtone for which the crystal is designed, the feedback is reduced by a capacitive divider in order to maximise the oscillator 'Q'. In the harmonic Pierce configuration the collector tuned circuit is resonated slightly below the crystal resonance in order to present an equivalent capacitive reactance to the crystal. This combined with the series capacitive reactance and the *BJT* phase shift provide the required phase conditions for oscillation - the well known Barkhausen criterion. However, in this pulsed implementation the dynamic characteristics of the *BJT* change as conduction ensues resulting in a changing capacitive load presented to the crystal the result is a frequency (and output power) change i.e. a 'chirp' which is highly undesirable because such an effect disperses the transmit pulse energy over a band of frequencies requiring an increase in the receiver detector bandwidth and a decrease in sensitivity (and range obtainable).

Parametric analysis *and* computer simulations of the *empirical* equivalent circuit of the oscillator show that in order to approach the operating requirements a low effective series resistance (*ESR*) is required from the crystal in conjunction with its matching to the tuned circuit. The base spreading resistance *Rbb'* of the *BJT* is effectively in series with the feedback path and therefore a device which possesses a small *Rbb'* in addition to a large *DC Hfe* and an *Ft* of a few times the required operating frequency is required. Fortunately medium power *RF* devices can exhibit low base spreading resistance, 4 - 30 ohms being common (9) and therefore such devices are an obvious candidate although choice is difficult because manufacturers do not usually supply information on this parameter. Although operation at a crystal overtone helps reduce unwanted 'chirp' effects, by virtue of an increased crystal equivalent inductance, it is associated with an increase in the *ESR* which has a deleterious effect on the oscillator starting characteristic in the pulse mode i.e. it can be responsible for 'chirp' effects.

Measurements

Performance variations of the configuration shown in figure (3) have been characterised using *AT* cut conventional 70MHz third overtone crystals (having low *ESR's*) and also with the oscillator alone using fundamental 210MHz chemically etched crystals (16) as well as 210MHz third overtone types. The measured collector current for the 70MHz oscillator alone are shown in figure (4).

Measured characteristics of the oscillator at and around the correct crystal operating point, are shown in figure (5). It is important to note that the 'no-chirp' condition corresponds to a detuning of the parallel resonant circuit in the collector. Consequently the available output voltage swing (and power) is reduced compared to that available at resonance, (about 3-5dB reduction in relative output level has been measured). Although reliable oscillation will occur at the 'resonant' operating point, this condition is accompanied by a significant frequency chirp - when the oscillator is pulsed. Two further observed effects are also illustrated in figure (5).

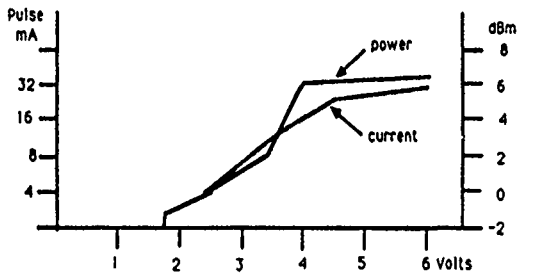


Figure 4 - Output power vs Vcc and pulse current

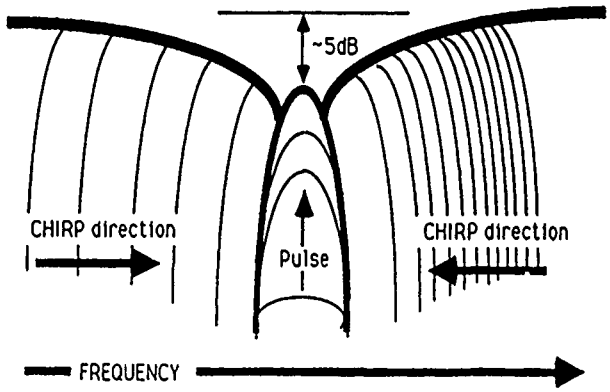


Figure 5 - Coherence vs frequency and output amplitude

These are that the pulse repetition rate (*PRF*) of the oscillator changes with the operating point. On the low frequency side the *PRF* is lengthened and on the high frequency side the *PRF* increases and in addition the chirp sweeps from low to high when the operating point is low in frequency and high to low when operating high of the coherent operating point. Therefore it is important, in this application, that on each periodic current impulse, from the base driving *R-C* network, the crystal immediately governs the frequency of oscillation (within a few 10's of Hertz) without excursions in frequency and amplitude. Relative output levels, including harmonic energy, for the oscillator (and multiplier) driven by 70MHz third overtone and 210MHz fundamental crystals are shown in figures (6) and (7). In the latter case the crystal was deliberately overdriven, the recommended drive level being 0.3mW. The triangular envelopes depict the measured *far-field* 'filtering' effect of a quarter wave radiator on the unwanted products. Simple radiated power measurements, based in the laboratory, are generally inadequate because they can be misleading. The transmitter-antenna field strength needs to be measured in the *far field* under suitable controlled conditions to accurately obtain usable figures. This is because of the interactions of the transmitter housing with the simple circuitry and the external antenna not to mention those with the animal to which the transmitter is attached. The latter two effects are extremely difficult to usefully theoretically model and near field measurements (in the laboratory) have been found to be only comparatively useful. Using a 70MHz third overtone crystal and multiplication factors of 4,5 and 6 times produce maximum outputs of -6, -5 and -8dBm respectively. A 210MHz fundamental unit produced a relative output of +1dBm and times by two (420MHz) tuned output of -8dBm. This was produced without a separate multiplier and therefore indicates a performance equivalent to that obtained using a 70MHz 3rd overtone crystal oscillator plus multiplier. However it should be noted that the sensitivity of the 210MHz oscillator to circuit values (and tolerances) is increased by nine times compared to the 70MHz oscillator. Hence it is not surprising that it was found necessary to employ some temperature compensation in the 210MHz oscillator in order to eliminate frequency 'chirp' induced as a result of operational temperature variations.

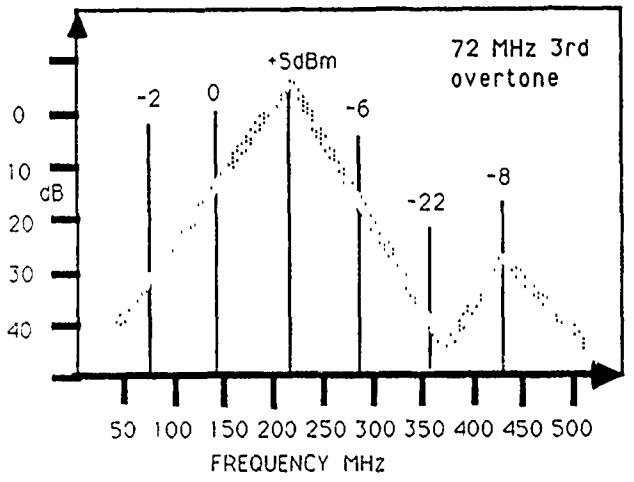


Figure 6 - Output vs frequency

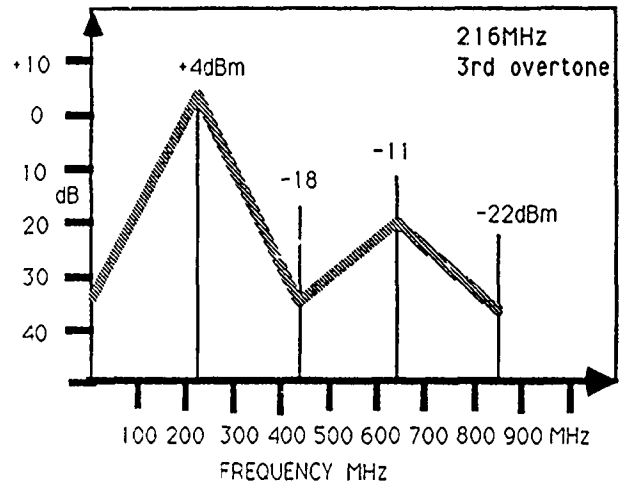


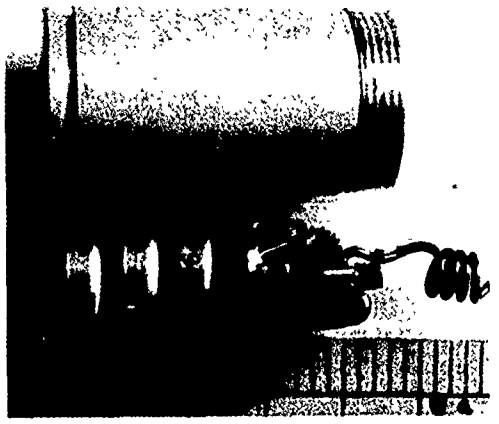
Figure 7 - Output vs frequency

Conclusions

Attempts to realise *coherent* pulsed oscillators of minimum component count, weight and volume which produces a few dBm of power output at spot frequencies in the frequency range 200-800MHz have been described. Oscillators and multipliers capable of operating below 1Volt have been developed and characterised. Systems design considerations, using current technology, indicate that an 'optimum frequency', based on maximisation of the operating range, are frequencies in the region of 350-450MHz. Such frequencies are also attractive because they offer the possibility of relatively small, efficient antennas both for transmission and reception.

The apparent resurgence of interest in quartz crystals and the new commercially available *VHF* crystals produced by chemical milling and similar processes need to advance a little more (and become commonplace) before they become serious commercial contenders for this or similar applications. Their present (AT cut) sub-milliwatt drive levels combined with standard crystal package sizes means they offer little operational advantage although we have demonstrated that it is possible to slightly 'overdrive' such devices when they are used in a pulse mode of low duty cycle.

Applications in the 800-1600MHz region exist but these must await suitable quartz technology. In particular the production of stress compensated (*SC*) cut *VHF* fundamental crystals (5) possessing relatively low *ESR* will allow the development of miniature transmission sources of sufficient frequency stability and, in particular, realisable output power. However the required suppression of the unwanted B and C modes in *SC* cut crystals will add to the circuit complexity and resultant component count.



Typical hardware realisation -
the small increments of the ruler are millimetres

Areas requiring further investigation and development

- Overdriving of *VHF* crystals fundamentally cut for frequencies above 200MHz. This may be possible because of pulsed operation and a duty cycle very much less than one.
- The use of *SC* cut *VHF* crystals - if and when they become commercially available could alleviate drive constraints and improve frequency stability problems at higher frequencies. However mode suppression could require a large component count.
- Low component count temperature compensation for frequency drift in *VHF* oscillators. (Assuming industry being able to readily handle 0805 and smaller surface mount components).
- *VHF/UHF* fundamental oscillators based on *SAW* devices.
- Advances in chemical batteries of high power to weight ratio and of small volume.
- Efficient, electrically small radiating antennas are a key area requiring further investigation.

References

[1] G.Ferry, 'The understanding of animals', Basil Blackwell Ltd., Oxford, 1984.

[2] 'The Bonytail Chub being monitored in Dinosaur, Colorado', Echoes, Summer 1989, 2nd edition, National Park Service, USA.

[3] 'Investigating Bees', (Miniature LED's on Bees allows detection upto one mile) Nature -Times news service (UK) 1988.

[4] K Fujimoto, Henderson, Hirasawa, and James 'Small Antennas', Research Studies Press, UK, 1987.

[5] F. Stern, (STC Ltd,UK), private communication, May 1989 (at this conference!).

[6] B.Parzen and A. Ballato, 'Design of Crystal Oscillators and Other Harmonic Oscillators', Wiley, 1983.

[7] D.K. Belcher, 'A high stability VHF oscillator', RF Design, Vol 6, No 1, pp 26-36, Jan/Feb 1983.

[8] Bruce Long, 'Quartz Crystals and Oscillators', Piezo Crystal Company, Proc. of the RF Technology Conf., Feb. 1989, Santa Clara, CA., pp 113 - 122.

[9] E.H.Nordholt and R.M.van Vierzen, 'Ultra low noise preamplifier for moving coil phone cartridges', Journal of the Audio Engineering Society, Vol 28, No 4, pp 219-223, April 1980.

[10] 'Duracell Guide for designers', Duracell Europe 1985.

[11] 'Sanyo lithium batteries' Sanyo Ltd, Japan.

[12] 'Varta technical catalogue', Ni-Cad batteries, Spring 1987, UK.

[13] R F.Sands, (Rayovac (UK) Ltd), private communication, April 1989.

[14] 'Hawk Chalk' Vol XXIV, No 2, August 1985, Tom Munson (USA).

[15] 'Panasonic micro batteries technical handbook', 1988.

[16] Innovative Frequency Control Products, Plainfield, PA 17081, USA, over period Nov-May 1989. (Samples of chemically etched fundamental and third overtone crystals were provided operating at approx. 200MHz)

[17] M E Frerking, 'Crystal oscillator design and temperature compensation', Van Nostrand 1978.

43rd Annual Symposium on Frequency Control - 1989

A HIGH STABILITY MICROWAVE OSCILLATOR BASED ON A SAPPHIRE LOADED SUPERCONDUCTING CAVITY

A.J.Giles, S.K.Jones, D.G.Blair, and M.J.Buckingham
Department of Physics, The University of Western Australia,
Nedlands, Western Australia, 6009

Abstract

By combining the high mechanical rigidity and low loss tangent of cryogenic sapphire with the excellent shielding and low loss properties of a superconducting cavity we have developed a microwave resonator with both high electrical quality factor and very high intrinsic stability even at relatively high power. We have implemented the sapphire loaded superconducting cavity resonator in a novel phase stabilized loop oscillator circuit and achieved Allan Variances of around 10^{-14} for 1 to 1000 seconds integrating time. This level of stability is competitive with that of the best hydrogen masers and in fact superior for integrating times less than a few tens of seconds. We present here an overview of the oscillator system and discuss applications and prospects for further improvement in performance.

I. Introduction

The ultra-high Q and low temperature coefficient obtainable in superconducting cavities at low temperatures has enabled the construction of microwave oscillators of exceptional short term frequency stability. Using a niobium cavity with a loaded Q of over 10^{10}, Stein and Turneaure achieved a short term performance of $\sigma_y(2,\tau) = 2.1\times10^{-16}$ in their superconducting cavity stabilised oscillator (SCSO) [1]. Despite this performance it is recognised [2] that superconducting cavity stabilised oscillators suffer from stability limitations set by mechanical deformation of the cavity due to vibration and tilt causing degradation to short and long term frequency stability respectively. These limitations, combined with the relative difficulty in obtaining ultra-high Q cavities, have restricted the practical development of this type of oscillator.

Research by Braginskii et al [3] has shown that high purity sapphire is an excellent dielectric for high stability resonator applications because of its high mechanical rigidity, good thermal properties and low dielectric loss at cryogenic temperatures. In order to take advantage of the low internal loss, the radiation loss from a dielectric resonator must be prevented (by using a low loss electromagnetic shield) or made sufficiently small by using a high order mode . Superconductor-on-sapphire resonators, in which a thin superconducting film is deposited onto the sapphire surface, have been extensively studied [3,4] to determine their suitability as high stability microwave resonators.

At the University of Western Australia we have developed a sapphire loaded superconducting cavity (SLOSC) resonator in which only a fraction (~20%) of the cavity volume is occupied by the dielectric. It combines the advantages and avoids the disadvantages of its two principle components, namely the superconducting cavity and the sapphire dielectric. For some resonant electromagnetic modes most of the energy is confined to the dielectric (which has ε ~10) resulting in a much reduced sensitivity to the effects of cavity perturbations. Furthermore, certain convenient thermal properties enable relegation of the effects of temperature fluctuations effectively to second order.

We have been developing a high stability microwave oscillator based on the SLOSC. The oscillator configuration is novel in that it uses the same resonator both as the frequency determining element in a loop oscillator and as the dispersive element in a Pound type discriminator to stabilize the oscillator against frequency fluctuations due to phase noise introduced in the loop circuit. This actively stabilized, partly cryogenic oscillator achieves stability superior to an all-cryogenic loop oscillator based on the SLOSC. Present performance is $\sigma_y(2,\tau) \sim 10^{-14}$ for $1<\tau<1000$ secs with a loaded Q of 3×10^8 and represents the present noise floor of the stabilization circuit.

In this paper we review the features of the SLOSC resonator and describe the design and performance of the stabilized oscillator based on the SLOSC.

CH2690-6/89/0000-089 $1.00 C 1989 IEEE

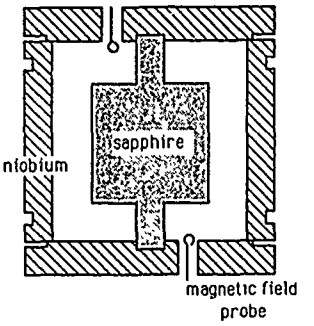


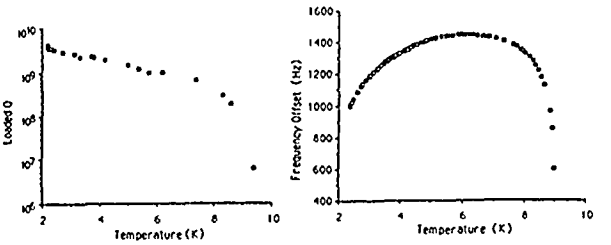
Fig.1 Cross-sectional view of the sapphire loaded superconducting cavity resonator (SLOSC)

II. Sapphire Loaded Superconducting Cavity Resonator

Design considerations and stability estimates for the SLOSC resonator (Fig. 1) have been previously communicated [5] so the description of the SLOSC here will be brief and review the features relevant to the high stability oscillator.

Quality factor

Most of the research has concentrated on the properties of a mode in the SLOSC at 9.73 GHz . Measurements of Q as a function of temperature for this mode [6] (Fig. 2a) reveal that it remains relatively high until near the superconductor's transition temperature. The effect of the resistive loss of the superconductor, which causes rapid degradation of Q with temperature in a conventional superconducting cavity resonator, is clearly reduced in the SLOSC; this is due to the high confinement of the mode to the sapphire.



(a) (b)

Fig. 2 Measurement of (a) Q as a function of temperature, and, (b) frequency as a function of temperature, for the SLOSC 9.73 GHz mode.

Temperature dependence

Measurements of the temperature dependence of the frequency of the 9.73GHz mode (Fig. 2b) reveal a turning point at about 6K. This has been interpreted [7] to be the result of the competing effects of the temperature dependent superconducting penetration depth and of a Curie law dependence of the susceptibility of paramagnetic impurities in the sapphire. From the data it was possible to determine the concentration of Cr cations to be about 1ppm and determine a geometric factor of about 28 000 for this mode.

Temperature control / cryogenic support

Operation near the frequency maximum at about 6K significantly relaxes the temperature control requirements otherwise necessary for high stability. Consideration of the curvature at the maximum implies that even for a relatively modest level of temperature control stability of 100 microK at a temperature offset of 1mK the resulting fractional frequency fluctuations would be only ~ 3×10^{-16}. Temperature control of the SLOSC is currently performed

by a room temperature AC resistance bridge type controller incorporating a Ge cryogenic resistance thermometer in good thermal contact with the SLOSC.

To ensure that the SLOSC attains the highest possible Q factor during cryogenic operation the sapphire is cleaned ultrasonically in a bath of hydroflouric and nitric acid followed by an alcohol bath and then finally rinsed in triply distilled deionized water and left to dry. The niobium cavity receives the same treatment except that no hydroflouric acid is used. After assembly in a dust free environment the SLOSC is mounted in a small, stainless steel vacuum pot which has been designed to minimize the effects of mechanical stress and to allow for effective thermal regulation of the SLOSC. Short pieces of coaxial transmission line connect the SLOSC coupling probes to microwave feedthroughs in the lid of the vacuum pot so that once evacuated and sealed it can be stored at room temperature until required for cryogenic operation; this has proved to be a very robust design. This pot is, in turn, mounted inside a larger vacuum can which is then lowered into a dewar and cooled with liquid helium. This second stage of vacuum helps to decouple the SLOSC from any mechanical vibrations associated with the boiling helium or pressure changes above the helium bath and is also required for proper operation of the temperature control system.

III. Oscillator Configuration

In this section we derive from the oscillation conditions for a generalized oscillator, the dependence of oscillator frequency stability on critical circuit parameters such as the Q of the resonator and noise in the active device. These relations are then used to form guidelines for the design of high stability oscillator configurations for the SLOSC resonator. We discuss first an all-cryogenic loop oscillator and then describe and analyse our phase stabilized system.

General oscillator model

A general oscillator model is obtained by considering the effect of a feedback network, with (complex) transfer function β, on the gain of an amplifier, with transfer function A (Fig. 3). The gain of the amplifier is modified in the presence of the feedback loop to become $A' = A / (1 - \beta A)$.
The condition for oscillation is then $\beta A = 1$.

It is convenient to write this (complex) oscillation condition as the amplitude and phase conditions:

I $\quad |\beta A| = 1$

II $\quad \arg(\beta A) = 0$

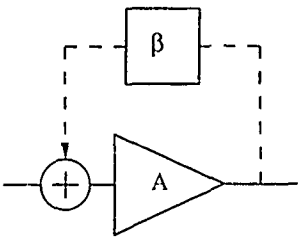


Fig. 3 General oscillator model as a feedback loop.

Loop oscillator

The oscillation conditions are now used to determine the operational parameters for a loop oscillator consisting of amplifier, transmission line, and resonator with transfer functions A, L and S_{12} respectively. The transfer function of the resonator, in magnitude and phase representation, is

$$|S_{12}| \exp(-j\phi_R) = \frac{2(\beta_1\beta_2)^{1/2}}{(1+\delta^2)^{1/2}} \exp(-j\arctan(-\delta)) \qquad (1)$$

where $\delta = (f-fo)(2Q_L/fo)$ is the normalised offset frequency and β_1, β_2 are the resonator input and output couplings, respectively.

The oscillation conditions now become, in obvious notation,

I $\quad |A L S_{12}| = 1$

II $\quad \phi_A + \phi_L + \phi_R = 0 \pmod{2\pi}$.

The first condition requires that the amplifier gain equals the total loop losses. This is met in practice by having more than sufficient small signal gain and exploiting the saturation characteristics of the amplifier to achieve equality at some input power.

The second condition can be rewritten using (1) to give

$$\delta = -\tan(\phi_L + \phi_A)$$

i.e. $\quad f - fo = -(fo/2Q_L)\tan(\phi_L + \phi_A)$. $\qquad (2)$

This equation shows that an oscillator that is configured around the resonator will have a frequency f that is not necessarily the same nor as stable as fo. The extent of this difference is ultimately determined by the noise of the active device and the quality factor of the resonator.

For small phase deviations and constant fo, the fractional frequency fluctuation is expressed as

$$\Delta f/fo = -(\Delta\phi_L + \Delta\phi_A)/2Q_L \,. \qquad (3)$$

Equation (3) can be expressed in terms of spectral densities of fractional frequency fluctuations to obtain the relation first derived heuristically by Leeson [8]

$$S_y(f_m) = S_\phi(f_m)/(2Q_L)^2. \qquad (f_m < fo/2Q_L) \qquad (4)$$

All-Cryogenic loop oscillator

Equation (2) above relates how the loop oscillator frequency stability is dependent on the phase stability of the amplifier and loop components. Our initial configuration of a high stability oscillator based on the SLOSC was as an all cryogenic loop oscillator using GaAsFET amplifiers. The rationale behind this approach was to exploit the stable cryogenic environment to minimise thermally and mechanically induced changes in loop phase ϕ_L. The amplifiers were 2 stage units using GaAs MESFETS and were developed by the CSIRO Division of Radiophysics for use as extremely low noise amplifiers on the front end of radiotelescopes with typical input power of -170dBm. The small signal white noise temperature for these devices was about 25K at physical temperatures of 4.2K. In our application the amplifiers were (necessarily) operating in saturation with input powers of about 0dBm . Amplifier bias supplies with microvolt stability were required to reduce the observed bias dependent phase shifts in the amplifier.

Cryogenic loop oscillator performance

The all-cryogenic loop oscillator system achieved a best stability of $\sigma_y(2,\tau) = 6\times10^{-14}$ at 10 seconds integrating time as determined from a comparison between three independent systems.

We concluded that the performance of the all-cryogenic loop oscillator was limited by the flicker noise in the GaAsFET amplifiers. We can estimate the flicker noise level by using the observed Allan variance to calculate an equivalent spectral density of frequency fluctuations by using the well known relation for flicker noise

$$\sigma_y^2(2,\tau) = 2\ln 2 . S_y(f_m) . f_m$$

and then, using equation (4) , $Q_L = 3\times10^8$ and $\sigma_y(2,\tau) = 6\times10^{-14}$, we obtain

$$S_\phi(f_m) = 1.8\times10^{-9}/f_m \ \text{(/Hz)}$$

$$= -87\text{dBc}/f_m \ \text{(/Hz)}.$$

es with $S_\phi(f_m) \sim -90dBc/f_m$ (/Hz) as measured by Lusher and Hardy [9] for the 1/f noise in an 8.5 GHz GaAs MESFET amplifier.

In summary, although the all-cryogenic loop oscillator based on the SLOSC was perhaps capable of better performance, it was eventually abandoned because the relatively high level of flicker noise in the loop amplifiers would ultimately determine the best performance, or "flicker floor", in the oscillator frequency stability. Furthermore, experimentation was made difficult as all the components were at cryogenic temperatures and relatively inaccessible. There was also a reluctance to operate the amplifiers at input powers 17 orders of magnitude greater than their nominal input power!

IV. Phase Stabilised Loop Oscillator

The desire for greater stability over that obtained with the cryogenic loop oscillator has led to the conception of a phase stabilised loop oscillator. This configuration, shown schematically in Fig. 4, can be regarded as a novel synthesis of loop oscillator and the form of Pound stabilization developed to a high degree by Stein and Turneaure [1]. The circuit is somewhat similar to that of Galani et al [10] which actively degenerates the phase noise of a loop oscillator, but differs in its intrinsically better dc performance owing to the use of an IF frequency in the stabilisation servo.

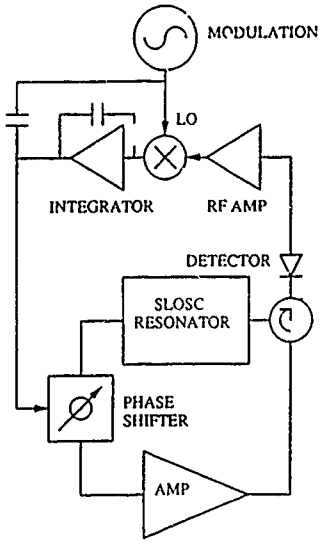


Fig. 4 Block diagram of the phase stabilized loop oscillator circuit.

Operational description

The phase of the loop is modulated at a rate much larger than the resonator bandwidth. The resultant PM sidebands undergo PM to AM conversion at the cavity with a magnitude and phase dependant on the difference between the frequency of the loop oscillator and that of the resonator. The reflected AM signal is envelope detected, amplified and then demodulated with the original modulation frequency to give a dc error signal proportional to the frequency offset. The discriminator output is then integrated and fed back to the phase shifter to correct the offset so that the integrator input is zero. Functionally, the stabilisation circuit corrects for any phase shift within its bandwidth and locks the oscillator frequency to that of the cavity.

Noise analysis

To consider the effectiveness of the stabilization servo, the circuit is modelled (Fig. 5) as an oscillator with pushing figure $[df/dV]_{osc}$ ($=fo/2Q_L \cdot d\phi/dV$) and equivalent input noise $V_n(f_m)$ at Fourier frequency f_m; the discriminator is characterised by its sensitivity $[dV/df]_{disc}$ and output noise $V_{os}(f_m)$; the integrator is described by C/jf_m, where C is a constant.

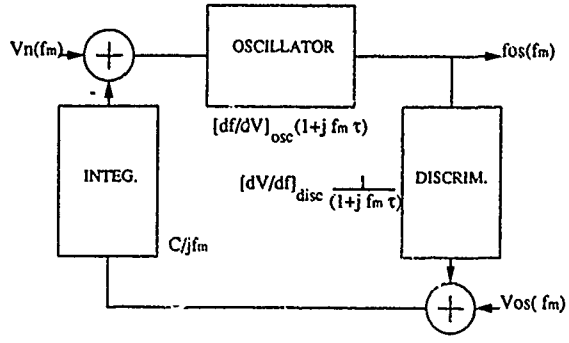


Fig. 5 Representation of the phase stabilized oscillator, including noise sources.

The stabilized oscillator output is then written as (rms sum implied)

$$f_{os}(f_m) = \frac{V_n(f_m).[df/dV]_{osc}}{(1+B/jf_m)} + \frac{V_{os}(f_m)}{[dV/df]_{disc}} \quad (f_m < B) \qquad (5)$$

where $B = [df/dV]_{osc}.[dV/df]_{disc}.C$ is the stabilization loop bandwidth.

The first term represents the degeneration of the (open) loop oscillator frequency noise by the stabilization servo. The second term represents the effect of noise introduced by the servo loop (usually from the detector diodes) and it is this term that determines the stabilized loop oscillator noise floor.

It is instructive to rewrite (5) in terms of spectral densities

$$S_y(f_m) = \frac{S_\phi(f_m)/(2Q_L)^2}{|1 + B/jf_m|^2} + \frac{(1/fo)^2\, S_v(f_m)}{|K_{det}|^2} \qquad (6)$$

where K_{det} represents the discriminator sensitivity referred to the output of the detector and $S_v(f_m)$ the voltage noise of the detector. If detector noise is the dominant contribution to the discriminator noise then K_{det} must be increased for better stabilized loop oscillator performance. The dependence of K_{det} on system parameters is given by

$$K_{det} = 4\, s\, \frac{2\beta_1}{(1+\beta_1+\beta_2)}\, m\, P_i\, \frac{2Q_L}{f_0}. \qquad (7)$$

In this expression, s represents the sensitivity of the detector, P_i is the incident power and m is the phase modulation index.

Practical implementation

In the present configuration of the phase stabilised loop oscillator, all of the loop oscillator components are at room temperature except for the circulator and other isolators that are placed near the SLOSC to minimise frequency pulling effects resulting from changes in cable VSWR. Microwave signal transmission to and from the SLOSC is via low thermal conductivity coaxial transmission lines. Typically 2mW of microwave power is dissipated in the SLOSC. We have variously used cooled, biased GaAs detectors as well as standard zero bias Schottky detectors for the envelope detection. To achieve maximum discriminator sensitivity, the coupling on the discriminator port is adjusted to unity for maximum carrier suppression to permit square law operation of the diodes. The voltage controlled phase shifter consists essentially of two varactor diodes terminating a hybrid tee. The device has a dynamic range of about 100 degrees and a typical phase coefficient of 12 degrees/volt. The phase modulation index is about 0.02 in the present system. The transfer function of the stabilization loop is essentially that of a single integration and has a gain crossover point at about 3kHz.

V. Oscillator Performance

Measurement system

To evaluate the performance of the phase stabilized loop oscillator based on the SLOSC resonator we have constructed two nominally identical systems, mounted in separate cryostats, and performed a standard two oscillator comparison in the time domain. As shown schematically in Fig. 6 this involves a double heterodyne configuration. The first difference frequency of approximately 709kHz (which is due to the slight difference in dimensions of the two sapphires) is mixed with a similar frequency from a quartz locked synthesiser (HP3325A) which has less relative instability than the first difference frequency. The output signal mixer is typically made to be about 130Hz and is low pass filtered (f_{LP}=1kHz) prior to being counted (HP5316A). The counter is usually gated every 10 seconds and the data is read over the IEEE bus by an IBM PC-AT compatible computer for later analysis. For integrating times less than about 1 second we are limited by the measurement system noise which we believe is mainly due to the synthesiser. For the same reason we have so far been unable to directly measure the phase noise (in the frequency domain) of the oscillator.

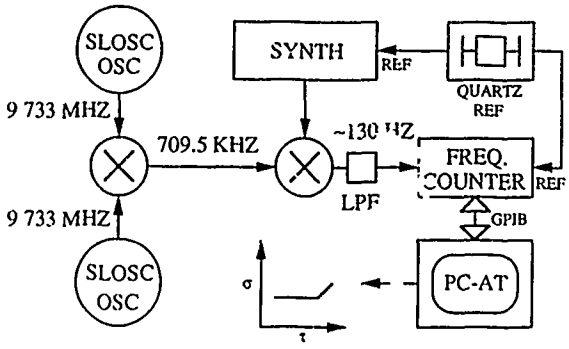


Fig 6 Schematic diagram of the frequency stability measurement system.

Short term stability

The square-root Allan Variance (SRAV) determined from the two oscillator comparison is shown in Fig. 7. As is standard it. this type of comparison a factor of $\sqrt{2}$ is removed to permit interpretation as SRAV of one system. The phase stabilized loop oscillator displays a minimum SRAV of 9×10^{-15} at 100 seconds integrating time. Also shown is the typical unstabilised loop oscillator (open loop) performance and the best performance of the all-cryogenic loop oscillator.

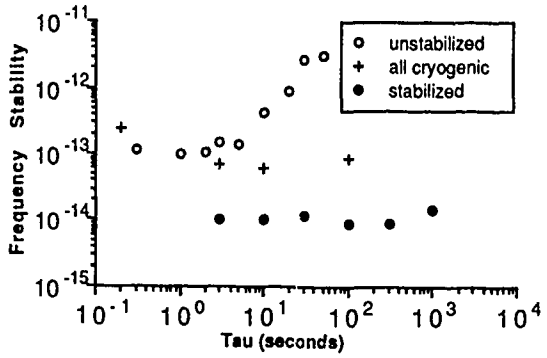


Fig 7 Measured square root Allan variances for the various oscillators discussed in this paper.

The remarkably flat portion in the phase stabilized loop oscillator performance from 3 seconds to about 300 seconds is suggestive of a dominant flicker noise source in the system. We are currently investigating the origins of this and believe it is probably due to multiplicative 1/f noise in the detector.

It can be seen from equations (6) and (7) that as well as using lower noise detectors, the signal to noise of the discriminator may also be improved by using a larger modulation index or by improving the loaded Q of the SLOSC resonator.

The result 9×10^{-15} achieved with a Q_L of 3×10^8 is interesting in that it yields the value 3×10^{-6} for $\sigma_y.Q_L$ which is the same as that for the SCSO of Stein and Turneaure.

Long term stability

For integrating times longer than about 1000 seconds, the performance begins to degrade. We have identified systematic correlations of the beat frequency with room temperature as well as liquid helium level. Fig. 8 shows the difference frequency between two systems over approximately 2 days data collection. There is a strong diurnal component correlated with the room temperature as well as a characteristic signature toward the end of the collection period where one system was running low on liquid helium. The mechanisms here are probably residual temperature dependent servo errors (because of offsets) and residual frequency pulling of the SLOSC.

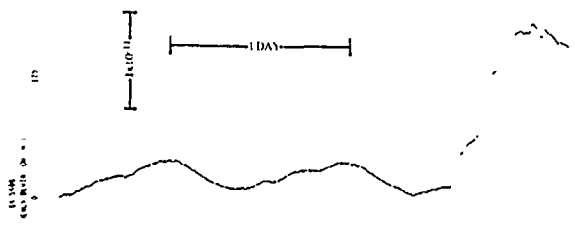


Fig. 8 Measurements of the relative frequency difference between two phase stabilized loop oscillators over a collection period of 2 days showing relative variations of $\sim5\times10^{-12}$ with a day long cycle.

As yet there is no evidence of mechanical sensitivity but further tests are required to fully characterise the system. We are confident of being able to significantly improve the long term performance.

VI. Applications

The wide ranging applications and potentials for high stability resonators and oscillators in fundamental physics experiments and other high precision measurements are well known (see, for example, the review by Stein [11]). We confine this brief discussion to application of the SLOSC oscillator as the local oscillator in Australia-wide Very Long Baseline Interferometry (VLBI). This application calls for a local oscillator of high stability to perform the dual functions of downconverting the the received signal and recording phase information for later correlation and analysis. For this role it was necessary to convert the high stability 9 733 MHz output to a standard 5 MHz output. Several trials in conjunction with receivers in the East of Australia have already been made but the data has not yet been analyzed. We are currently developing a stand alone oscillator system that incorporates a long hold time liquid helium dewar so as to lessen the inconvenience of liquid helium refills. This system is intended for VLBI use as well as other future applications requiring high performance, low maintenance and continuous operation.

VII. Conclusions

We have described a sapphire loaded superconducting cavity and its use as the resonator in a novel phase stabilized loop oscillator. Demonstrated performance is very encouraging and is comparable to that of hydrogen maser systems. The short term

stability appears to be limited by the present noise floor of the stabilization circuit. We have observed some frequency dependence on microwave power level and this is currently under investigation. Long term performance is at this time degraded by some sensitivity to ambient temperature and helium level. As environmental sensitivities, power and temperature dependent effects etc become better clarified we believe significant further improvement in performance will be achieved.

Acknowledgements

We are indebted to Dr Alan Young of the CSIRO Division of Radiophysics for the design of the frequency control electronics and the design of the 5MHz receiver. This research was funded by the Australian Research Council.

References

[1] S.R.Stein and J.P.Turneaure, "Superconducting Resonators: High Stability Oscillators and Applications to Fundamental Physics and Metrology," in Future Trends in Superconductive Electronics, ed.B.S.Deaver et al (American Inst. Phys., New York, 1978), pp.192-213.

[2] B.Komiyama, "A 9.2 GHz superconducting cavity stabilised oscillator," IEEE Trans.Instrum.Meas., IM-36, pp.2-8, March 1987.

[3] V.B.Braginskii, V.I.Panov, S.I. Vasiliev "The properties of superconducting resonators on sapphire," IEEE Trans Magn, MAG-17, pp.955-957, Jan. 1981.

[4] S.Thakoor, D.M.Strayer, G.J.Dick and J.E.Mercereau, "A lead-on-sapphire superconducting cavity of superior quality," J.Appl.Phys., vol. 59, pp.854-858, Feb. 1986.

[5] D.G.Blair and S.K.Jones, "High Q sapphire loaded superconducting cavities and application to ultrastable clocks," IEEE Trans Magn, MAG-21, pp.142-145, Mar. 1985.

[6] S.K.Jones and D.G.Blair, Electron. Lett., "High-quality factors for a sapphire-loaded superconducting cavity resonator," vol.23, pp. 817-818, July 1987.

[7] S.K.Jones, D.G.Blair and M.J.Buckingham, "Effect of paramagnetic impurities on frequency of a sapphire-loaded superconducting cavity resonators," Electron Lett, vol.24, pp.346-347, Mar. 1988.

[8] D.B.Leeson, "A simple model of feedback oscillator noise spectrum," Proc. IEEE, vol.54, pp.329-330, Feb. 1966.

[9] C.P.Lusher and W.N.Hardy, "Effects of gain compression, Bias conditions, and temperature on the flicker phase noise of an 8.5 GHz GaAs MESFET amplifier," IEEE Trans. Microwave Theory Tech., MTT-37, pp.643-646, Mar.1989.

[10] Z.Galani, M.J.Bianchini, R.C.Waterman Jr., R.W.Laton and J.B.Cole, "Analysis and design of a single-resonator GaAs FET oscillator with noise degeneration," IEEE Trans. Microwave Theory Tech., MTT-32, pp. 1556-1565, Dec. 1984.

[11] S.R.Stein, "Space applications of superconductivity: Resonators for high stability oscillators and other applications," Cryogenics, pp.363-371, July 1980.

43rd Annual Symposium on Frequency Control - 1989

L-BAND DIELECTRIC RESONATOR FILTERS AND OSCILLATORS WITH LOW VIBRATION SENSITIVITY AND ULTRA LOW NOISE

S.M. SPARAGNA

Raytheon Company
Equipment Division
Radar Systems Laboratory
1415 Boston Providence Turnpike
Norwood, Massachusetts 02062

ABSTRACT

Because of the increasing requirements of signal sources used in military radar applications to have improved noise and spurious performance there has been a push to develop stable oscillators that operate at, or closer to, the radar operating frequency. These sources do not have the phase noise or spurious enhancement caused by frequency multiplication which would be required for a lower frequency source. Presently, Dielectric Resonators offer an effective way to produce such sources by providing a high-Q frequency selective device for determining the operating frequency of an oscillator. However, the loaded Q of the Dielectric Resonator Filter (D.R.F.) is not high enough, when used in an oscillator, to compete with a crystal oscillator in close-to-carrier noise performance even though the crystal source must be multiplied. Typically the loaded Q is reduced even further when the D.R.F. is designed to meet the operational vibration requirements found in a military environment. The optimum source would have the close-to-carrier noise performance of a multiplied crystal source and the far from carrier noise performance of a Dielectric Resonator Oscillator (D.R.O.).

The focus of the design was to develop a low loss, high Q D.R.F. with good vibration performance to use in a low noise D.R.O. at 1.5 GHz. The D.R.O. had to have very good noise performance while maintaining a low sensitivity to vibration in a military environment. Also the D.R.O. must have the ability to be phased-locked to a multiplied crystal source to improve the close-to-carrier noise and vibration performance. The resulting development produced a D.R.F. at 1.538 GHz with a vibration sensitivity (γ) of $\leq 2.2 \times 10^{-8}$ G^{-1} in all planes and a loaded Q of 7000 at 75°C with 6 dB insertion loss. The oscillator developed using this D.R.F. has an output power of +23 dBm at 1.538 GHz and measured phase noise which is -135 dBc @ f_m = 2 kHz, -175 dBc @ f_m = 100 kHz, and $\leq$ -185 dBc @ $f_m \geq 1.0$ MHz.

This paper will describe the evolution of the D.R.F. design and the incorporation of this design into a low noise D.R.O. The results of each intermediate step will be included along with the final results.

INTRODUCTION

The contents of this paper are separated into three discrete sections which comes about naturally from the development process. The first section contains the design and development procedure used to create a D.R.F. which would meet the requirements stated above. Several hurdles were identified at the beginning of the D.R.F. development and each one is considered separately in this section. Results are given within the context being developed. The second section describes the design criteria and procedure used to develop an " Ultra-Low Noise D.R.O. " using the above D.R.F. in a external feedback oscillator. Each component of the oscillator was designed to operate with a 50 ohm (Ω) source and load impedance to allow for direct characterization of each component. This was done to improve the predictability of the D.R.O. performance and also to allow problems to be isolated if they occurred. A description of each component is included along with their measured parameters. The final section compares the measured results with the predicted performance of the entire D.R.O.

CH2690-6/89/0000-094 $1.00 C 1989 IEEE

D.R.F. DESCRIPTION

The D.R.F. uses a high dielectric material to load a cavity thus reducing the size of the cavity and also stabilizing the cavity since 90% of the electric field and 65% of the magnetic field exists within the dielectric resonator. This helps to reduce the effect that the boundary conditions (i.e. the cavity) have on the operating frequency.[1] To date, to obtain a high loaded Q in the D.R.F. the resonator was placed on a low loss, low dielectric pedestal in the center of the cavity. This was done to approach a free space mounting of the resonator which keeps the gradients in the fields outside the resonator small thus reducing the losses encountered by high circulating currents in the conducting boundaries of the cavity.[2] However this approach causes the D.R.F. to have poorer vibration sensitivity due to the increased size of the cavity and the relative movement between the resonator and the cavity.[3] When a very high Q is not required the resonator is attached directly to the bottom of the cavity on a dielectric substrate. This causes a large gradient in the electric field with high circulating currents at the surface of the conductor and introduces dielectric losses but offers much improved vibration performance. Figures 1 and 2 shows the calculated field distribution for the $TE_{01\delta}$ mode relative to the Z axis for a resonator centered in a cavity and for the final D.R.F. design found in this paper.[4]

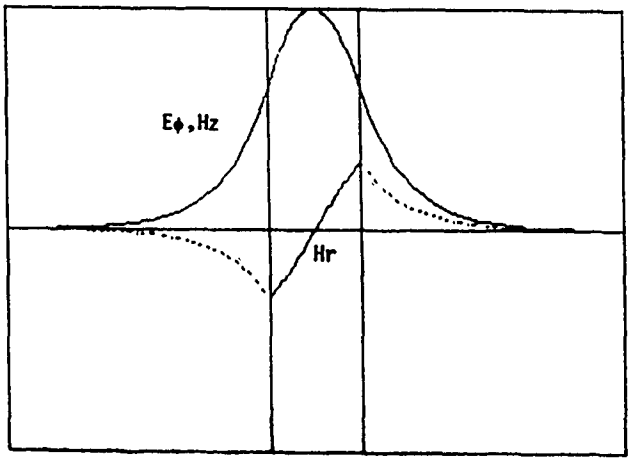


Figure 1: D.R.F. field distribution (with no substrate) as a function of position along the Z axis. The left and right regions represent the air space in the cavity and the middle region is the Dielectric Resonator (D.R.).

To solve the problem of both high Q and vibration sensitivity there were six major hurdles which were identified and addressed. These six problems were the dielectric resonator, the cavity, the substrate material, the epoxy and bonding which are considered together, and finally the coupling structure. Because the thermal mass of the D.R.O. is mostly determined by the cavity the thermal properties of the D.R.F. are discussed as part of the cavity problem.

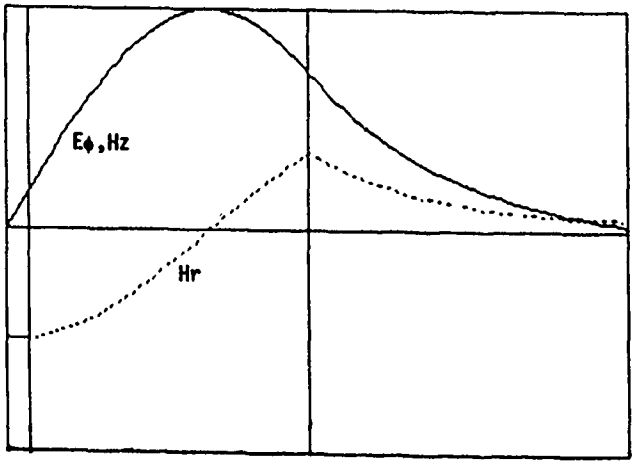


Figure 2: D.R.F. field distribution (with substrate) as a function of position along the Z axis. The left region is the substrate, the middle region is the D.R., and the right region is the air space in the cavity.

Dielectric Resonator

The first item addressed was the dielectric resonator itself. Since the resonator is the main contributor to the loaded Q of the filter a search was done to obtain a dielectric resonator with both a very high unloaded Q and a relatively high dielectric constant (to reduce the size of the D.R.F.). A device available from Murata Erie (" M " material) has an unloaded Q of > 8000 at 7.0 GHz and > 23000 at 1.5 GHz. The part number is DRT377M060C167. The dielectric constant of this material is 37.64 with a chemical composition of $(ZrSn)TiO_4$. The resonator has the shape of a puck with a 6mm hole down the center of it, an outside diameter of 37.7mm and a height of 16.7mm (this height is roughly half a wavelength with this dielectric at 1.5 GHz). The hole in the center of the resonator does not effect the performance since the electric field strength here is almost zero and the presence or absence of

the dielectric plug has little direct effect on the magnetic field.[5] The hole allows for a low loss plastic screw to be added to center the resonator in the cavity during assembly and adds to the higher mode suppression of the filter.[6]

Cavity

The second factor addressed is the cavity. The inside dimensions of the cavity are 38.1mm in height and 76.2mm in diameter. These dimensions are roughly twice the diameter and three times the height of the resonator. These dimensions seem to give the highest loaded Q while still maintaining good mode separation. At this frequency the $TE_{01\delta}$ mode allows for the smallest physical dimensions (i.e. lowest mass) with a reasonably high loaded Q.[7] Because the cavity supplies part of the boundary conditions which determine the resonator fi quency, any modulation of these boundary conditions will manifest itself as an undesirable frequency modulation of the D.R.F. and thus a phase modulation of the completed oscillator. Although the dielectric resonator reduces this effect as compared to a regular cavity, the effect is still significant in this application. The major source of such modulation is environmental vibration found in military applications. Figure 3 shows the vibration spectral density in which the D.R.F. was designed to operate.

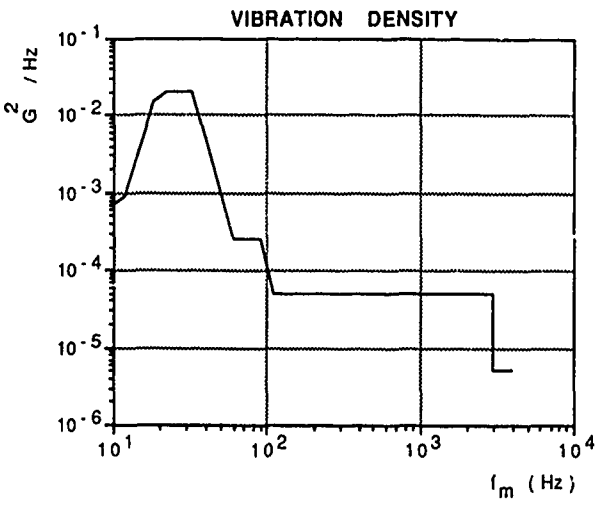


Figure 3 : Operational vibration spectral density for the D.R.F.

An investigation was done into a cavity design which would keep the inside dimensions stated but offered the most rigid structure possible while still keeping within reasonable size and weight bounds. As shown in Figure 3 the vibration spectral density is greatly reduced above 3.0 kHz so the major guide line for the analysis was to keep any resonances as far above 3.0 kHz as possible. Not only is the vibration density reduced in this region but it would make it easier if any mechanical isolation is needed. The analysis produced a cavity with 9.525mm walls and a tongue-and-grove joint at the middle of the cavity. A tuning screw is used to make small adjustments in frequency by modifying the boundary condition of the top of the cavity.[8]

Since the cavity dimensions were now determined the thermal properties of the D.R.F. could be considered. Work done at Raytheon's Research Division[9] shows that the frequency vs temperature response of a D.R.O. can be represented as a second order equation of the form:

$$\Delta f/f_o \text{ (ppm)} = C_o + C_1 \times T + C_2 \times T^2 \qquad (1)$$

Here the coefficient C_o is an arbitrary constant related to the reference frequency (1538.103 MHz) used to define the fractional frequency change. C_1 is the linear coefficient which is defined as:

$$C_1 = \tau_\varepsilon + \eta_\alpha \qquad (2)$$

τ_ε is the linear coefficient of fractional frequency change given by the D.R. vendor in units of ppm/°C and η_α represents any additional linear shifts in the frequency vs temperature characteristics due to those components external to the D.R. such as the cavity, D.R. support, the oscillator electronics, etc. The coefficient C_2 is the second order coefficient and has units ppm/°C² and is a function of the type and source of material used for the D.R.

For the D.R. bought from Murata/Erie which is their " M " material with a temperature coefficient of +0.3 ppm/°C, and using a copper cavity, the coefficients C_1 and C_2 are:

$$C_1 \simeq 0.3 \text{ ppm/°C} - 2 \text{ ppm/°C} = -1.7 \text{ ppm/°C} \qquad (3)$$

$$C_2 \simeq -0.025 \text{ ppm/°C}^2 \qquad (4)$$

By taking the derivative of equation (1) and setting it equal to zero, the turn over point can be calculated from the following equation:

$$T_o = -C_1 / (2 \times C_2) \qquad (5)$$

Software was developed to measure the D.R.F. parameters over a long interval of time and at discrete temperatures. Figure 4 shows the measured $(\Delta f/f_o)/\Delta T$ for the D.R.F.

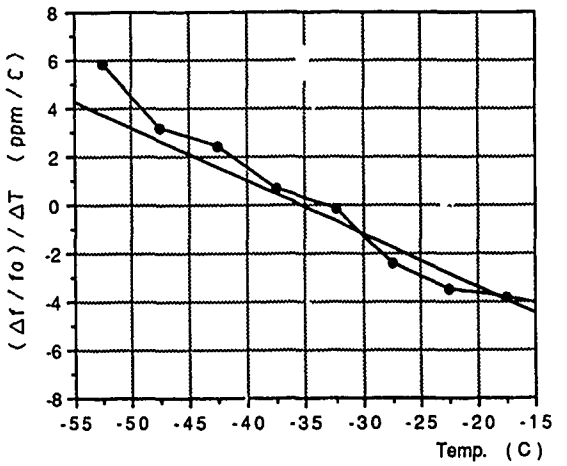


Figure 4 : Fractional change in frequency of the D.R.F. per degree change in temperature (PPM/°C) vs temperature from -55°C to -15°C.

The measured turn over point for the present D.R.F. is –35°C and the calculated turn over point is –34°C which is in good agreement. Using equation (5) the following table for a 75°C turnover point (which is the desired operating point) can be generated for a copper cavity or a Kovar cavity with $C_2 \simeq -0.025$ ppm/°C² which is the constant required for Murata/Erie " M " material:

cavity material	π_ϵ (ppm/°C)	n_ϵ (ppm/°C)
copper	5.75	-2.0
Kovar	4.45	-0.7

Murata/Erie also has temperature coefficient of ~ +4.0 ppm/°C and ~ +6.0 ppm/°C available in their " M " material. Using the +4.0 ppm/°C temperature coefficient with a Kovar cavity would yield a nominal T_o of 66°C and using the +6.0 ppm/°C temperature coefficient with a copper cavity would yield a nominal T_o of 80°C. As of this publication these cavities have not been assembled and tested to verify these predictions.

Another thermal problem to consider is the stabilization time required for the D.R.F. at turn-on. Presently two 15 watt proportional-control monolithic heaters from Micropac are attached to the cavity with the goal being to be within ± 9 PPM of stable frequency in 30 min. with an ambient from 0°C to 65°C. The cavity was insulated and placed in an oven to soak at 0°C while the oscillator electronics were placed external to the oven and kept heated to 75°C. At τ (time) = 0 the oscillator was turned on along with one 15 watt heater on the D.R.F. The cavity temperature and oscillator frequency were measured in one minute intervals. Then the test was repeated using two 15 watt heaters on the D.R.F. Figures 5 and 6 show the results of both test for oscillator frequency (PPM) and D.R.F. temperature vs τ. If used in the present configuration these tests predict that ~60 watts will be required to meet a 30 minute turn on time required. However by moving T_o to the desired operating frequency the required power would be reduced.

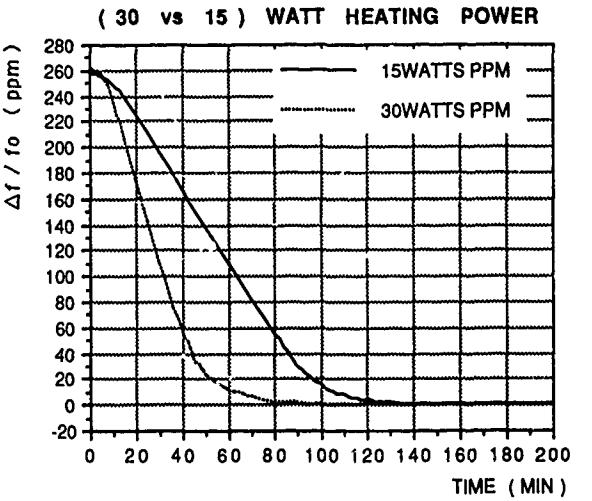


Figure 5 : Fractional change in frequency (ppm) vs time (min.) for the insulated D.R.F. soaking at 0°C for both 15 and 30 watts of heater power.

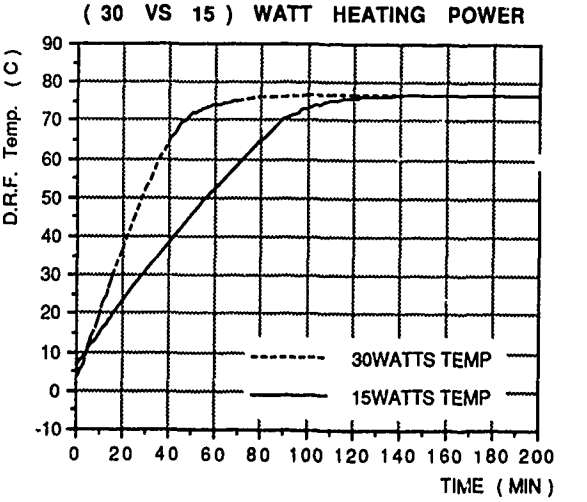


Figure 6 : Temperature of the insulated D.R.F. soaking at 0°C (°C) vs time (min.) with both 15 and 30 watts of heater power.

Substrate

The third element investigated was the dielectric material which holds the dielectric resonator to the bottom of the cavity and also has the coupling structures etched onto it. A material was needed which had a very low loss tangent (high Q), a reasonably high dielectric constant (however not so high as to compete with the resonator), a coefficient of expansion which is close to that of the dielectric resonator, and the ability to be easily machined and polished. A material produced by Trans-Tech, Inc. called D-MAT was used. This material has a thermal coefficient of expansion of 7.5 PPM/°C while the resonator has a thermal coefficient of expansion of 6.5 PPM/°C. At the D.R.F. operating frequency the loss tangent for this material is < 0.0002 with a dielectric constant (ϵ_r) of 10.2. The material is in the form of a disk with the following dimensions:

D-MAT 10.2 (Grade A)

O.D. = 76.073 ± 0.127 (mm)
I.D. = 5.969 ± 0.0254 (mm)
Thickness = 5.08 ± 0.0254 (mm)

The hole in the center is to allow for the low loss plastic alignment screw mentioned earlier. The thickness of 5.08 mm has a free space electrical length at 1.538 GHz of roughly 16.22 mm due to its dielectric constant of 10.2. This allows the resonator to be electrically centered in the cavity while being physically attached to one end. The obvious disadvantage is that the D-MAT has a higher loss tangent than air thus reducing the loaded Q of the filter for a given coupling (insertion loss). However the D-MAT is good enough (i.e. loss tangent < 0.0002) to allow for a loaded Q of > 6500 at 1.538 GHz with an insertion loss of ~6 dB. Therefore a high Q has been attained with a good vibration insensitive configuration. Figures 7 through 10 show the measured S parameters for D.R.F. #3 at an operating temperature of75°C. The measurements were made using a Hewlett Packard 8510 Network Analyzer after performing a full two-port calibration. The analyzer was configured using a 401 point sweep with a 1 MHz span, and a measurement center frequency of 1.538103 GHz.

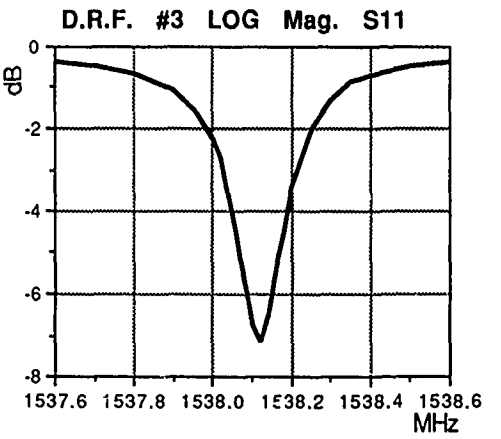


Figure 7 : D.R.F. #3 S11 log magnitude. The peak return loss is -7.17 dB at 75°C and 1.538123 GHz.

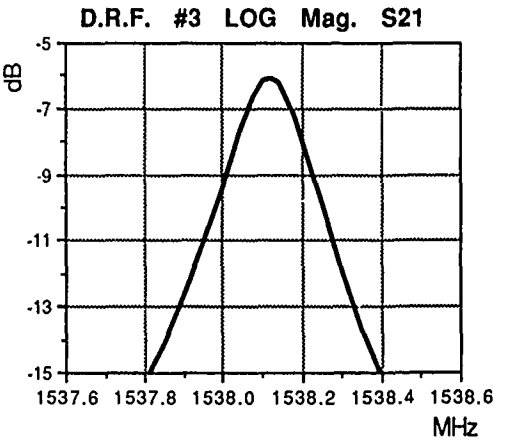


Figure 8 : D.R.F. #3 S21 log magnitude. The minimum filter insertion loss is -6.124 dB at 75°C and 1.538123 GHz.

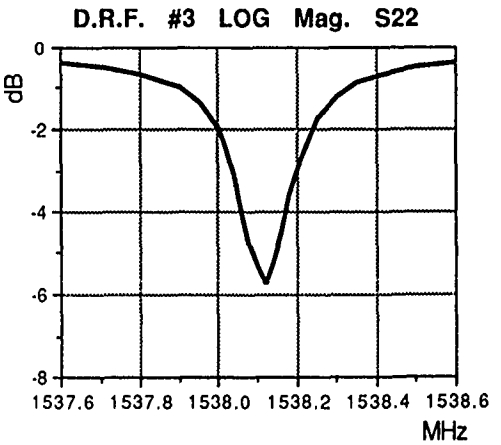


Figure 9 : D.R.F. #3 S22 log magnitude. The peak return loss is -5.74 dB at 75°C and 1.538123 GHz.

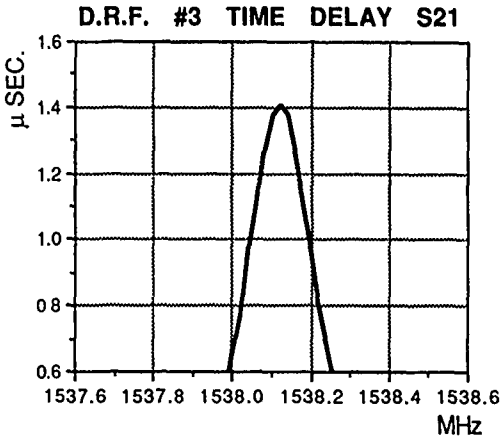


Figure 10 : D.R.F. #3 S21 time delay. The maximum delay is 1.404 µS at 75°C and 1.538123 GHz.

Epoxy and Bonding

The fourth and fifth factors are tightly coupled and are considered together. The fourth is what epoxy or bonding agent to use to bond the dielectric resonator to the substrate, and the fifth is how to bond all the pieces into the cavity. A low loss, high strength and very rigid epoxy was needed to bond the dielectric resonator to the dielectric substrate. Inorganic compounds were considered as was high temperature bonding (" fritting ") but both require extremely high temperatures. The fritting process takes temperatures in excess of 850°C and the lowest temperature inorganic epoxy I could find cured at 425°C. Also even though the dielectric resonator and the dielectric substrate have coefficients of expansion which are close (6.5 PPM/°C and 7.5~8.2 PPM/°C respectively), over this temperature range the bond would experience a large shear and could fatigue or break. A high oxide organic epoxy from Amicon Co. (ME-868) was found. This epoxy cures in an hour at 180°C. The Young's Modulus of the epoxy is estimated at $0.4 \sim 0.6 \times 10^6$ PSI. The loss of the epoxy was unknown so a layer of the epoxy was cured on a piece of Teflon and then peeled off. The cured piece of epoxy was then placed into a test D.R.F. on top of the resonator to see how much degradation it caused. The effects were very slight even though this was an extreme case since the thickness of an actual bond would be much less.

The second half of the problem was how to bond the dielectric substrate to the cavity. The first and most obvious solution was to plate the bottom of the substrate and solder it to the bottom of the cavity. However three problems occurred with this approach. First, because of the well-like construction of the cavity it was not possible to verify if there were any voids in the solder or trapped flux between the substrate and the cavity. Second the cavity expansion coefficient required a 90 Mil. Kovar barrier plate to avoid cracking the substrate under the temperature required to flow the solder. The third and most important problem was all the added loss incurred by each additional conducting surface. Since the gradient in the electric field is large here small resistances are significant due to the high circulating currents. As a solution to all three problems the same epoxy used to bond the resonator to the substrate was used to bond the non-plated substrate to the cavity, which was milled to be very smooth and flat at the bonding surface. This worked very well since now the only conducting surface is now the cavity itself (i.e. no Kovar barrier plate or plating on the substrate) which is where the circulating currents have to exist due to the boundary conditions presented by the conducting walls of the cavity. The dielectric losses of a thin bond of this high oxide (non-conductive) epoxy are much less than the conducting losses in the substrate and Kovar plating. Also since the curing temperature is only 180°C the epoxy does not see a very large shear due to expansion and is pliable enough to withstand the shear that does exist. However there was a question as to whether either bond can with-stand a severe shock, which is a requirement for military applications, and how the epoxy pliability will effect vibration.

The vibration sensitivity of the D.R.F. was tested first using a sinusoidal drive function out to 1 kHz that was varied in small increments to determine the vibration sensitivity as a function of frequency. The D.R.F. was part of the oscillator and the vibration sensitivity was measured on a delay line discriminator with only the D.R.F. being vibrated. Figure 11 shows the test setup used to make the vibration sensitivity measurements. Using cylindrical coordinates let the Z axis be the direction normal to the bottom of the cavity and substrate and the R (radial) axis parallel to these surfaces. Because of the cylindrical symmetry there is no θ dependance in the vibration sensitivity.

VIBRATION SENSITIVITY MEASUREMENT SETUP

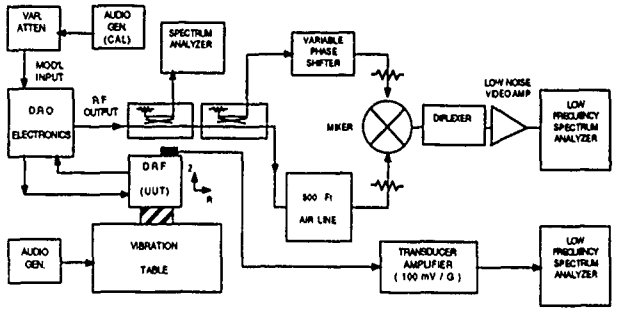


Figure 11 : Vibration sensitivity test measurement setup used to measure the vibration sensitivity of the D.R.F. in an oscillator with the D.R.O. electronics isolated.

The actual vibration sensitivity γ (G^{-1}) is calculated from the measured RMS frequency deviation Δf (Hz RMS):

$$\Delta f \text{ (Hz RMS)} = F_o \times \gamma \times \sqrt{(D \times BW)} \qquad (6)$$

Here D is the vibration drive spectral density in G^2/Hz, BW is the measurement bandwidth in Hz and, F_o is the operating frequency in Hz. If a sinusoidal drive function is used to generate the vibration, as in this case, then the vibration sensitivity is calculated from:

$$\Delta f \text{ (Hz RMS)} = F_o \times \gamma \times \Phi \qquad (7)$$

Where Φ is the measured vibration in G RMS.[10]

The first unit was bonded together in one process with the substrate to cavity and resonator to substrate bond both cured at the same time with no pressure. The Z axis vibration sensitivity was measured as 4.2×10^{-7} (G^{-1}) for this unit. This seemed much worse than anticipated by analysis. After some investigation it was postulated that the epoxy bond itself was deforming allowing for relative movement between the cavity, the substrate and the resonator. To test this, a second unit was built in two separate curing cycles with ~2.5 PSI on the substrate to cavity surface during the first cycle and ~10.0 PSI on the resonator to substrate surface during the second cycle. It was reasoned that a thinner bond line would allow less deformation under vibration. This unit had a measured vibration sensitivity of 5.4×10^{-8} (G^{-1}) which is an ~18 dB improvement over the first unit. A third unit was built, again using two curing cycles but this time a heated hydraulic press was used to yield a curing pressure of 100 PSI during both cycles. This unit had a measured vibration sensitivity of 2.2×10^{-8} (G^{-1}) which is an ~8 dB improvement over unit two and ~26 dB improvement over the first unit. This sensitivity is good enough that both electrical isolation (e.g. phase locking) and mechanical isolation systems become practical to meet the projected noise requirements under operational vibration. Figure 12 shows the Z axis vibration sensitivity vs frequency for all three units.

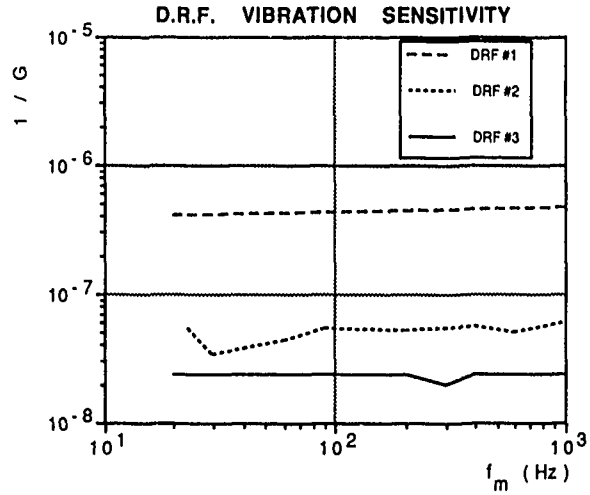


Figure 12 : D.R.F. vibration sensitivity for the three units built with different bonding pressure at assembly.

Since 100 PSI is equivalent to 700 Lbs of force on the substrate and 200 Lbs of force on the resonator no further increases were attempted because of feared breakage but also because of demonstrated diminishing returns. D.R.F. # 3 was also tested without the tuning screw and showed no discernible difference in its vibration performance. The R axis vibration sensitivity was measured on the first D.R.F. only (no pressure bond). The vibration sensitivity in this plane measured 1.0×10^{-8} (G^{-1}) which is ~32.5 dB lower than the sensitivity in the Z axis for the same unit. R axis vibration sensitivity was not measured on the other units since the vibration sensitivity of the original unit was so low therefore, the improvement in the R axis vibration sensitivity with the high pressure bond is not directly known. However,

measurements on the vibration sensitivity of the whole oscillator using the second D.R.F. (low pressure bond) show that the R axis sensitivity is at least as low as the Z axis sensitivity and is probably much better since relative movements in this direction are less and their effect on frequency is less due to the field configuration in the D.R.F.

The loaded Q and insertion loss were measured on each unit. The data for each D.R.F. is given below.

DRF	S21 (dB)	Loaded Q	Freq. (GHz)	Temp (°C)
#1	5.9	7440	1.622	25
#2	7.0	7600	1.498	75
#3	6.1	7000	1.538	75

The second part of this problem was the survivability of the D.R.F. under shock especially with the thin bond line made by the high pressure curing. A test unit was made with the 100 PSI curing pressure for both cycles just as with the third D.R.F. discussed above. The shock requirement specifies a 75 G shock having the form of a 6 mS half sine. The unit was subjected to 15 shocks with the first five at 50 G, the second five at 100 G, and the last five at 150 G. The test unit showed no visible signs of damage or fatigue. From this test it was determined that the high pressure bonding technique was more than strong enough to survive the shock requirements. A question was raised as to the aging properties of the cured epoxy as to its strength under shock. The vendor had no data when asked but predicted that the cured epoxy would show no degradation at all for the first 5 to 10 years and this was a conservative estimate.

Coupling Structure

A coupling structure needed to be developed that would not adversely affect the vibration sensitivity, the D.R.F. loaded Q and insertion loss, and had the ability to be adjusted to compensate for variations in production. Since 90% of the electric field and 65% of the magnetic field is confined to the resonator, a coupling structure that coupled into the external 35% of the magnetic field was developed.[11] The top side of the 5.08 mm substrate (resonator side not the cavity side) was metalized with 250 angstroms of Cr and 381 μm of Cu. The Cr is needed as a barrier metal. Figure 13 is a copy of the art work (not to scale) that is etched on to the substrate such that the only metallization left corresponds to the lands in the art work.

D.R.F. COUPLING STRUCTURE

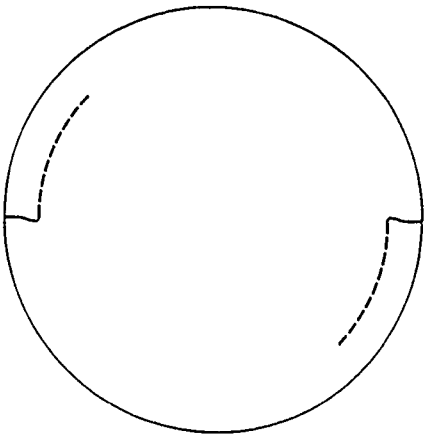


Figure 13 : Copy of the art work (not to scale) for the coupling structure used in the D.R.F. The line around the coupling structure shows the edge of the substrate and is not part of the art work.

It is important that all unwanted metallization including the Cr must be completely removed as this caused a problem in the initial unit with the loaded Q of the D.R.F. The segmented line or tracks form an arc with a radius of 31.75 mm with each track having 4 degrees of arc with a 1 degree space between tracks. There are 9 tracks which have a total of 45 degrees of arc. The dielectric substrate allowed for very strong coupling to the resonator for small coupling structures. To lower the coupling to the proper level (i.e. 6 dB insertion loss) and to keep the structure from getting too close to the cavity wall, the width of the coupling structure was reduced to 0.508 mm. By raising the impedance of the line the current and thus the magnetic field intensity of the coupling signal was reduced. To increase the coupling, successive tracks are connected on both sides (i.e. symmetry is maintained) by soldering in tinned copper jumpers (0.508 mm wide) till the insertion loss is close to 6 dB. The two units built with this structure have both had three tracks connected.

Figure 14 shows a cross section of the D.R.F. with the individual components labeled.

D.R.F. CROSS SECTION

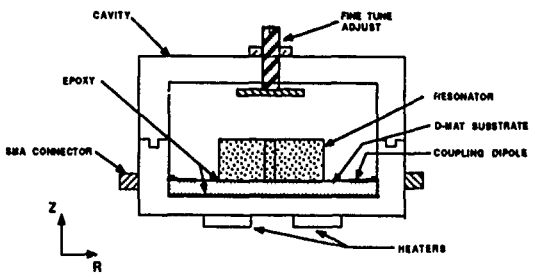


Figure 14 : Cross section of the D.R.F. showing the critical elements of the filter and the cavity wall structure.

D.R.O. Description

The electronics built around the D.R.F. were developed to have an operating band from 1.47 GHz to 1.57 GHz. This would allow the frequency of the oscillator to be changed by substituting a different D.R.F. into the oscillator and resetting the phase of the feedback. The oscillator was also required to supply an output of +24 dBm and to have the ability to be phase locked. Figure 15 shows a block diagram of the D.R.O.

D.R.O. BLOCK DIAGRAM

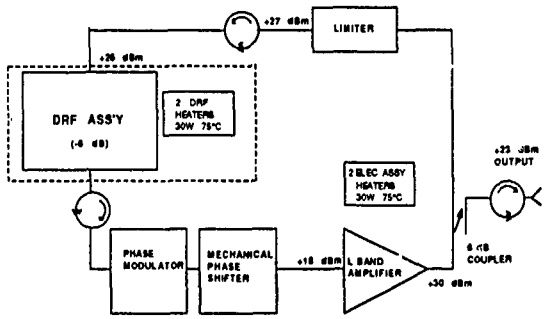


Figure 15 : Block diagram of the D.R.O. The power levels and position of each assembly are shown.

All the electronics are connected to a heater plate and held at 75°C using two 15 watt heaters. These heaters are the same type as the ones used on the D.R.F. The two isolators were added to the feedback to isolate the D.R.F. from the rest of the oscillator. This was done due to the poor out-of-band return loss of the D.R.F. All the isolators used in the design are specified to have an insertion loss < -0.4 dB over the operating band at 75°C. To help eliminate any dynamic effects in the oscillator amplifier and make the D.R.O. design more predictable, an external limiter is used as the non-linear device in the oscillator and the amplifier is operated at less than 1 dB compression.

The levels shown in Figure 15 were determined by noise considerations, power requirements, and loop gain requirements. The following equation was used to determine the oscillator configuration and to predict the noise performance of the D.R.O. given that a fixed gain was required to overcome losses in the feedback.

$$\rho_n(\text{dBc SSB}) = N_p(f_m)(\text{dBc}) + 20\times\text{Log}(F_o \div (2\times Q_l \times f_m))(\text{dB}) + 20\times\text{Log}(M)(\text{dB}) \quad (8)$$

This equation is for the single-sideband (SSB) phase noise of a source operating at F_o, with a modulation frequency (f_m)inside the oscillator resonator bandwidth, where the loaded Q (Q_l) is constant. Here $N_p(f_m)$ is the open-loop single-sideband phase noise of the oscillator in dBc, and M is the required multiplication factor. ρ_n is in the same bandwidth as $N_p(f_m)$.[12] From equation (8) the best noise performance is obtained by having the smallest $N_p(f_m)$ and largest Q_l possible for a given operating frequency and multiplication factor. Since Q_l is set by the D.R.F. and M = 1, $N_p(f_m)$ is the only variable.

$N_p(f_m)$ is determined by the noise figure and 1/f noise of the oscillator sustaining stage, the 1/f noise of the other components in the oscillator, and any dynamic noise caused by operating in or close to saturation.

Amplifier

The amplifier used in the oscillator is a hybrid-pair configuration using two Hewlett Packard (HP) bipolar transistors (HP 5102) that are selected. Each amplifier in the hybrid pair is an identical mirror image of the other. The bias is applied to the device using an active bias configuration which holds the collector current constant independent of temperature and drive. The

collector is biased at +18 V_{dc} with a current of 110 mA I_{dc}. Figure 16 shows a block schematic of the amplifier and the active bias.

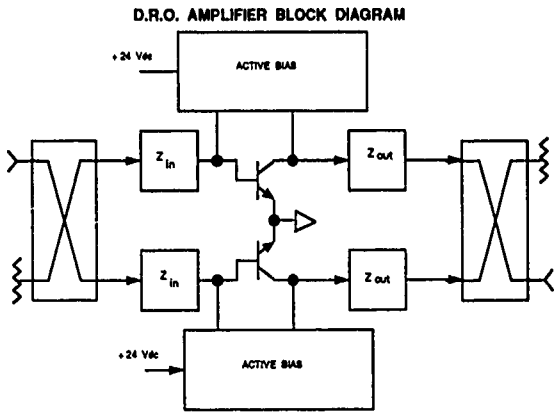


Figure 16 : Block schematic of the hybrid-pair amplifier used in the D.R.O.

A hybrid pair was used to allow for the required one watt output and a gain of 12 dB when in 1 dB compression. Also this arrangement offers a good return loss on both ports. The hybrids are branch line couplers[13] and are printed on the same substrate as the amplifier forming one circuit card (CCA). The CCA is made from 1.27 mm (50 mils) thick RT/duroid 6010.5 which measures 104.14 mm by 99.06 mm. All the components are either chip or surface mount except for the HP 5102 which is a flange mount. All the components are derated to NAVAL AIR SYSTEMS COMMAND specification AS-4613 Class B.[14] Bipolar transistors are used to reduce the 1/f noise effects in the amplifier while still having sufficient output power and gain at these frequencies and a reasonably-low noise figure. After balancing the noise effects of using a high power device with the advantages gained in signal-to-noise, the final amplifier design results were:

Gain = 12.5 dB $\pm$0.5 dB @ 1 dB comp.
Input Power = +18 dBm
Output Power = +30 dBm min. $\leq$ 1 dB comp.
Noise Figure = 7 dB max
1/f noise = (135 - 140) dBc SSB/Hz
@ (1 Hz, 1 dB compression)
Bandwidth = (1.47 to 1.57) GHz
Bias = +24 Vdc @ 360 mA
Temperature = 75°C

Coupler / Limiter

The 6 dB coupler and the passive limiter are both printed on one CCA using the same material as the amplifier. The coupler is also a branch line coupler[13] with the coupled path connected to a an isolator which is the oscillator output. The through path is connected to the passive limiter which is also connected to an isolator. This isolator output feeds the D.R.F. Both isolators have tab inputs and SMA outputs. This was done to accommodate direct connection to the CCA and coaxial connections to the D.R.F. and the output. The limiter diode is a PIN diode from HP (HP 5082-3141). Figure 17 shows a block schematic of the assembly and the measured results are given below:

Input Power = +28.5 dBm nom.
Output Power = +23 dBm min. (coupled path)
+27 dBm nom. (through path)
Bandwidth = (1.47 to 1.57) GHz
Bias = -14 Vdc @ < 1 mA
Temperature = 75°C

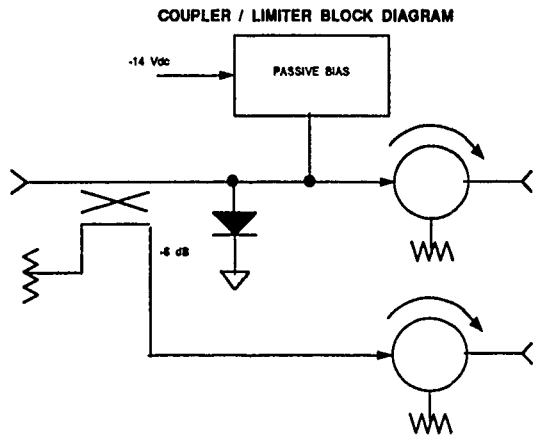


Figure 17 : Block schematic of the branch line coupler / limiter assemble used in the D.R.O. The limiter is a passive using a PIN limiter diode. The limit level is set for an oscillator output power of +24 dBm nominally.

Phase Modulator

The phase modulator design was taken from Garver[16] using a circulator and a varactor in a tuned circuit to change the phase of the reflected signal from the tuned port to the circulator isolated port. The varactor is an abrupt junction GaAs device from M/A-Com (MA46951-186) with the following characteristics:

Capacitance = (18-20)pF [f=1 MHz; V_R=2 Vdc]
(7-8)pF [f=1 MHz; V_R=20 Vdc]
Q = 1500-2200 [f=50 MHz; V_R=4 Vdc]
Gamma = .49@.54 (measured in package)
Breakdown = > 30 Vdc @ I_R = 10 μA

Figure 18 shows a block schematic of the assembly. Again the circulator port that attaches to the CCA has a tab while the input and output ports are SMA. The CCA is made from the same material as all the other assemblies. A fixed bias of + 12 Vdc is applied to the cathode of the varactor and the modulation signal is applied to the anode. The anode is by-passed to the R.F. signal with a 100 pF chip capacitor. The linearity of the phase modulator was measured with a signal generator. A signal at 1.5 GHz and +20 dBm was injected at the input to the phase modulator and the anode of the varactor was modulated at 50 kHz with a peak deviation of 700 Hz. The measured harmonic of the 50 kHz signal (100 kHz) was down 36 dB relative to the 50 kHz sideband at the output of the phase modulator. The measured results of the phase modulator are given below:

PHASE MODULATOR BLOCK DIAGRAM

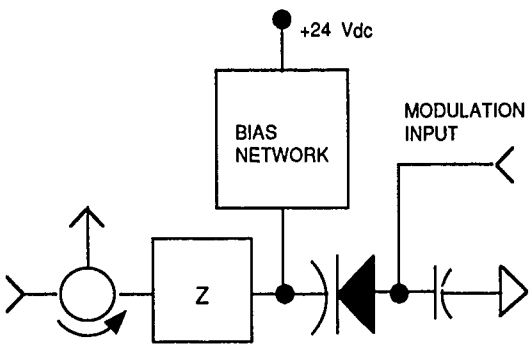


Figure 18 : Block schematic of the phase modulator used in the D.R.O. The phase modulator uses a GaAs abrupt junction varactor in a tuned circuit to load the output port of a circulator. The phase of the signal reflected to the isolated port of the circulator is adjusted by changing the bias on the varactor.

Bandwidth = (1.47 to 1.57) GHz
S21 (dB) = -1.0 dB max.
Bias = 24 Vdc @ < 10 mA
Temperature = 75°C
Sensitivity = (see following table)

Freq. (GHz)	Bias (Vdc)	Sensitivity (Rad/V)
1.478	24.0	12.24 x 10^{-3}
1.478	12.0	10.78 x 10^{-3}
1.478	2.0	13.56 x 10^{-3}
1.518	24.0	9.80 x 10^{-3}
1.518	12.0	7.55 x 10^{-3}
1.518	2.0	11.34 x 10^{-3}
1.568	24.0	7.35 x 10^{-3}
1.568	12.0	7.16 x 10^{-3}
1.568	2.0	9.37 x 10^{-3}

Manual Phase Shifter

The manual phase shifter is used to compensate the loop through phase to satisfy the zero loop phase required by the oscillator. To accommodate different D.R.F. and variation from unit to unit the phase shifter needed an adjustment range of 360 degrees over the entire operating band. Figure 19 is a block diagram of the Manual Phase Shifter.

MANUAL PHASE SHIFTER BLOCK DIAGRAM

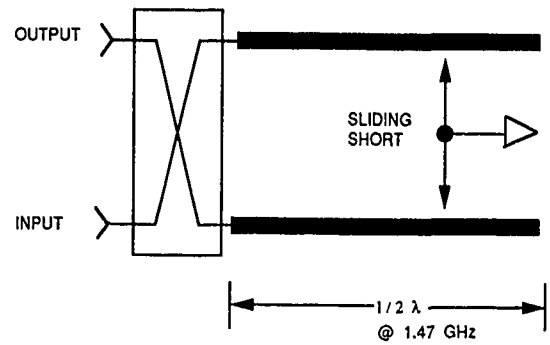


Figure 19 : Block schematic of the manual phase shifter used in the D.R.O. The manual phase shifter uses a sliding short on a quadrature hybrid to allow for a minimum of 360 degrees of adjustment from 1.47 GHz to 1.57 GHz.

A quadrature hybrid, of the same design used in the amplifier, has both of its outputs connected to a sliding short which can be moved up and down the output lines to change the phase of the reflection which comes out of the isolated port of the hybrid. To ensure 360 degrees of adjustment range the sliding short must be able to move 0.5 λ at the lowest frequency (1.47 GHz). Once the proper adjustment is made the sliding short is clamped into place so that there are no

adverse effects under vibration. The measured results are given below:

Bandwidth = (1.47 to 1.57) GHz
S21 (dB) = -0.8 dB max.
Temperature = 75°C
Adjustment Range = (see following table)

Frequency (GHz)	Adjustment Range (Degrees)
1.470	375
1.515	386
1.570	400

Dielectric Resonator Oscillator

The D.R.O. noise results are given in Figure 20. The predicted noise is plotted along with the measured oscillator noise and the measured noise floor. The FM noise of the oscillator was measured using a delay line discriminator. To increase the sensitivity of the discriminator a second D.R.F. was built to use as the delay element in the discriminator. Figure 4 shows the delay to be 1.4 μS and to be a constant over ±50 kHz. Outside of 50 kHz the cavity used for the delay element in the discriminator was replaced with 500 feet of air line. The noise measurement setup is shown in Figure 21.

The discrepancy between the measured and predicted noise is the weighting due to the floor of the discriminator. The predicted D.R.O. noise lies almost on top of the measured noise floor of the discriminator in the areas of the largest discrepancies. An equal contribution by the measurement floor would cause a 3 dB increase in the apparent measured noise. Only measured data is shown since compensating the data for the effects of the noise floor becomes prone to large errors in the unweighted data due to small errors in reading the data. The D.R.O. performance parameters using D.R.F. #3 are listed below:

D.R.O. PERFORMANCE PARAMETERS

Noise = f_m < 1 MHz (see Figure 20)
$f_m \geq$ 1 MHz 185 dBc/Hz
Output Power = +23 dBm min.
Freq. Sens. = 1 kHz/V (for phase locking)
Bias = +24 Vdc @ 360 mA
-14 Vdc @ 25 mA
Heater Bias = 28 Vdc (floated) @ 2.15 A
(for four 15 watt heaters)

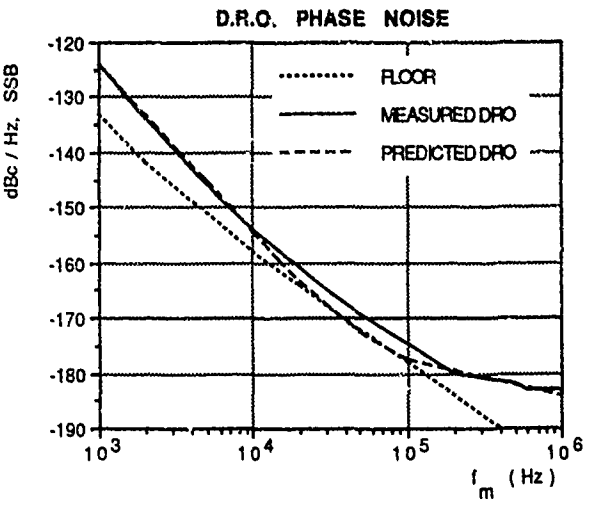


Figure 20 : Noise results for the D.R.O. using D.R.F. #3 at the operating temperature of 75°C. The data is plotted in dBc single side band in a 1 Hz bandwidth.

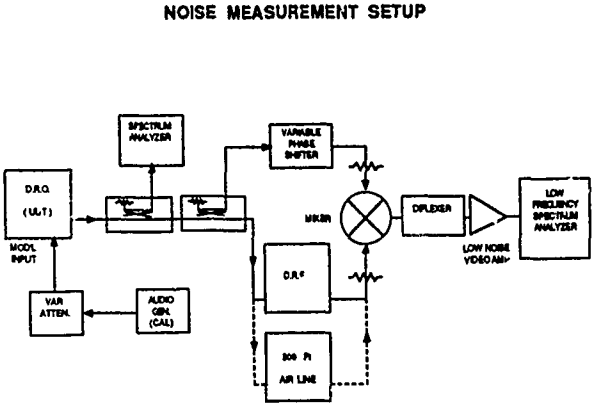


Figure 21 : Test setup used to measure the noise of the D.R.O. The equipment is configured as a delay line discriminator using either an other D.R.F. or 500 feet of air line for the delay element.

Summary

The unique properties of the D.R.F. allows it to have a loaded Q of > 6500 at 1.538 GHz with an insertion loss of ~6 dB while maintaining a vibration sensitivity 2.2×10^{-8} (G^{-1}). The use of high pressure bonding and the elimination of the metallization at the cavity and substrate interface allows this configuration to work. The high Q is attained by using a low loss dielectric substrate with a high enough dielectric constant (ϵ_r = 10.2) such that

the electrical dimensions of the cavity approach the free space or optimum model with the puck floating at the center of the cavity. The use of the coupling tracks allows for variation at the production level while not degrading the loaded Q or the vibration performance which is unique to this design.

The high loaded Q and low insertion loss of the D.R.F. allow the oscillator to have a good signal to noise ratio at the input of the amplifier without using very high power devices which have poorer noise figures. This coupled with the good 1/f noise of the open loop oscillator results in the phase noise performance which was achieved. The use of individual components in the oscillator caused the D.R.O. to be bigger than if a different design were used but has allowed for very good agreement between measured and predicted noise results of the Oscillator.

Acknowledgments

This paper would not have been possible had it not been for the support and cooperation of all the people in the Receiver Department of Raytheon Company Equipment Division, Radar Systems Laboratory. I would like to give special thanks to the Richard Kessler for all his support and guidance during this development and to Brian Fishwick and Tom Parker for their comments and suggestions.

References

1.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch. 1, pp. 2.

2.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.4, pp. 139-141.

3.) M.J. Loboda, T.E. Parker, and G.K. Montress, " Frequency Stability of L-Band, Two-Port Dielectric Resonator Oscillators, " IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-35, No. 12, pp. 1334-1339, December 1987.

4.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 165-176.

5.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 301-302.

6.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 305.

7.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 108.

8.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 161-164.

9.) M.J. Loboda, T.E. Parker, and G.K. Montress, " Temperature Sensitivity Of Dielectric Resonators And Dielectric Resonator Oscillators, " in Proceedings Of The 42ND Annual Symposium On Frequency Control, 1988, pp. 263-271.

10.) M.J. Loboda, T.E. Parker, and G.K. Montress, " Frequency Stability of L-Band, Two-Port Dielectric Resonator Oscillators, " IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-35, No. 12, pp. 1334-1339, December 1987.

11.) D. Kajfez and P. Guillon, Dielectric Resonators. Norwood, Mass.: Artech House, Inc., 1986, ch.1, pp. 479.

12.) G.D. Vendelin, Design of Amplifiers and Oscillators by the S-Parameter Method. New York, N.Y.: John Wiley & Sons, 1982, ch. 4, pp. 145-162.

13.) G.L. Matthaei, L. Young, and E.M.T. Jones, Microwave Filters, Impedance-Matching Networks, and Coupling Structures New York, N.Y.: McGraw-Hill, 1964, ch. 13, pp. 809-815.

14.) NAVAL AIR SYSTEMS COMMAND, Dept. of The Navy, " Application and Derating Requirements For Electronic Components General Specification For ", AS-4613, 30 July 1976.

15.) G.L. Matthaei, L. Young, and E.M.T. Jones, Microwave Filters, Impedance-Matching Networks, and Coupling Structures New York, N.Y.: McGraw-Hill, 1964, ch. 13, pp. 809-815.

16.) R.V. Garver, Microwave Diode Control Devices Dedham, Mass.: Artech House Inc., 1986, pp. 280-281.

MEASUREMENT AND ANALYSIS OF A MICROWAVE OSCILLATOR STABILIZED BY A SAPPHIRE DIELECTRIC RING RESONATOR FOR ULTRA-LOW NOISE*

John Dick and Jon Saunders

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

Abstract

Phase noise measurements are presented for a microwave oscillator whose frequency is stabilized by a "whispering gallery" mode sapphire ring resonator with Q of $2 \cdot 10^5$. Isolation of RF fields by the special nature of the electromagnetic mode allows the very low loss of the sapphire itself to be realized. Several mode families have been identified with fairly good agreement with calculated frequency predictions. Waveguide coupling parameters have been characterized for the principal (lowest frequency) mode family, for $n = 5$ to $n = 10$ full waves around the perimeter. For a 5 cm wheel resonator in a 7.6 cm container, Q's above 10^5 were found at room temperature for all of the modes in this sequence. Coupling Q's for these same modes ranged from 10^4 ($n = 5$) to 10^5 ($n = 10$) for a WR112 waveguide port at the center of the cylindrical wall of the containing can. Phase noise measurements for a transistor oscillator locked to the $n = 10$ ($7.84GHz$) mode showed a $1/f^3$ dependence for low offset frequencies, and a value of $\mathcal{L}(f) = -55db/Hz$ at an offset of 10 Hz from the carrier. We believe this value to be lower than any previously published for a non-cryogenic X-band oscillator. Based on these measurements and on the performance of commercially available phase detectors, we project performance for a cooled resonator operating at 77°K with a Q of $3 \cdot 10^7$ to be -85dB/Hz at an offset of 1Hz. This value is 30dB below that of the best X-band source presently available, a frequency-multiplied quartz crystal oscillator.

*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.

Introduction

Until recently the only technology suitable for meeting the spectral purity (or equivalently, phase noise) requirements for NASA's radio science experiments has been based on crystal quartz oscillators. The lowest phase noise quartz oscillators have fundamental or overtone oscillation frequencies of 5 or 10 MHz. For radio science experiments requiring S or X-band signals with high spectral purity at frequency offsets less than about 1 KHz, the signal is obtained by frequency multiplying the output of such an oscillator. The advantage of this method derives from the high Q (quality factor) of the quartz crystal (the resonator element) and low 1/f noise in transistors at these lower rf frequencies. This approach nevertheless produces multiplicative phase noise that becomes unacceptably large at higher frequencies such as 10 or 32 GHz. In fact, the noise in the microwave signal derived from the multiplication of 5 or 10 MHz references to higher frequencies has established a limitation on the sensitivity with which radio science observations could be made.

The use of currently available oscillators operating at GHz fundamental frequencies has also proved prohibitive. Where quartz crystals have Q's of $\approx 2 \cdot 10^6$ microwave Q's are two orders of magnitude lower; and where amplifiers in the MHz range have 1/f noise of −150dB at 1 Hz, amplifiers at X-band show 1/f noise of −90dB to −120dB. These two factors conspire to increase the close-in noise of microwave oscillators many orders of magnitude above that available from a multiplied quartz crystal oscillator.

Recent developments now make possible a practical oscillator with the operating frequency at X-band, Ka-band and higher frequencies with phase noise below that of the best available multiplied sources. The first of these is the development of practical resonators at X-band with Q's of 10^6 to 10^7 and even higher. These resonators have Q's of $2 \cdot 10^5$ at room temperature; further-

CH2690-6/89/0000-107 $1.00 C 1989 IEEE

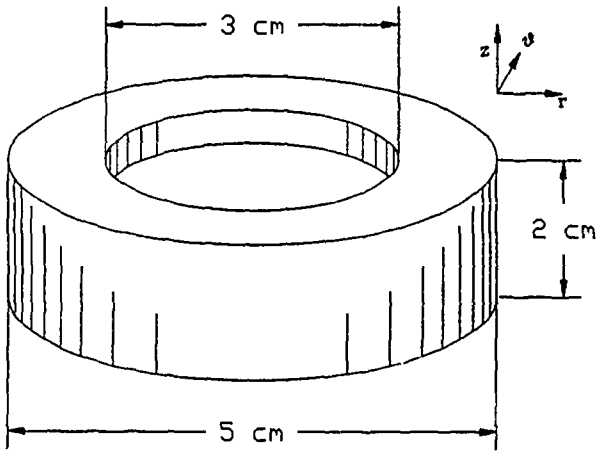


Figure 1: Configuration of the sapphire element in the "whispering gallery" mode resonator under test showing directional axis identification at ring perimeter. This element was placed in an aluminum containing can large enough to provide a 1.25 cm gap on all sides.

more the Q increases approximately as T^{-5} as the resonator is cooled, where T represents temperature above absolute zero. Thus a Q of $3 \cdot 10^7$ can be achieved at liquid nitrogen temperature (77°K), and quartz crystal Q's at temperatures practical with thermoelectric coolers. This performance is made possible by the character of the "whispering gallery" mode allows the intrinsic Q of sapphire material to be expressed without the addition of metallic dissipation.

In addition the use of a stabilized local oscillator (STALO) configuration, allows a passive mixer to take the place of the RF amplifier used in simpler oscillator configurations. While somewhat more complicated due to the use of feedback at baseband (frequencies nominally zero) rather than at RF, the low noise of the mixer ($\approx -135dB$ at 1Hz offset frequency) essentially eliminates the disadvantage of high frequency operation.

In the following sections we present results of tests on a STALO with a room-temperature sapphire resonator. This oscillator shows phase noise below that previously reported for an X-band source.

Background

Resonator and Mode Description

The whispering gallery mode sapphire resonator[1–8] isolates electromagnetic energy to the sapphire element itself and away from metallic supports and containers by means of a phenomenon similar to total internal reflection in optical devices. In such a resonator, evanescent fields decay exponentially with distance away from the sapphire element. Even so, practical resonators are composite structures in which the containing can plays a significant role in performance. This is because isolation is incomplete. With a significant resonator curvature in the space of a wavelength, the optical analogue breaks down to give a finite radiation-limited Q_r for an isolated sapphire element. For any given frequency, as the resonator size and mode number are increased, Q_r increases. Conversely, the required size increases with increasing Q_r. However, the necessary size for any given Q can be reduced by placing a metallic wall within the evanescent field region. The rate of field decay in this region is large, and so even a small gap between sapphire and wall can give a large Q enhancement.

Figure 1 shows the ring-shaped sapphire element used for our oscillator tests. It has a thin central web to allow support at its center. Various field regions associated with the ring resonator are identified in Figure 2. Here outer and inner caustic surfaces divide regions of travelling or standing waves, from regions of evanescent fields. Simply put, if the local phase velocity for the mode is less than that of the local medium, no combination of propagating waves in the medium can match that of the mode, leaving only non-propagating solutions.

Modes in the ring resonator have been calculated by Guillon et al[8], with good agreement to experimental data for relatively high mode numbers ($n > 10$). One of the authors[3] has also published calculations based on a simple "bent optical waveguide" approach. These calculations also give good qualitative agreement with experimental results and are shown in a later section.

Material Parameters

Sapphire resonators are capable of higher Q's than any other dielectric resonator, equaling those of quartz crystals ($Q = 2 \cdot 10^6$) at 170 Kelvin, a temperature achievable by means of thermoelectric cooling. As shown in Figure 3, the material-limited Q for these resonators increases as T^{-5} from their room-temperature value of about $2 \cdot 10^5$. At lower temperatures between 77 and 10 Kelvins, experiments at JPL have demonstrated Q's from $3 \cdot 10^7$ to $2 \cdot 10^9$ in X-band sapphire resonators.

The temperature coefficient of frequency has been measured by Blair et al[1] and confirmed by our own measurements. Temperature sensitivity is 8 to 10 times as large as the physical expansion coefficient and is due to variation in the dielectric constant ϵ with temperature. Values range from $2 \cdot 10^{-6}/°K$ at a temperature of $77°K$ to $6 \cdot 10^{-5}/°K$ at room temperature. Such large values make the resonator unsuitable for providing long-term stability without compensation unless the temperature is reduced to $20°K$ or below.

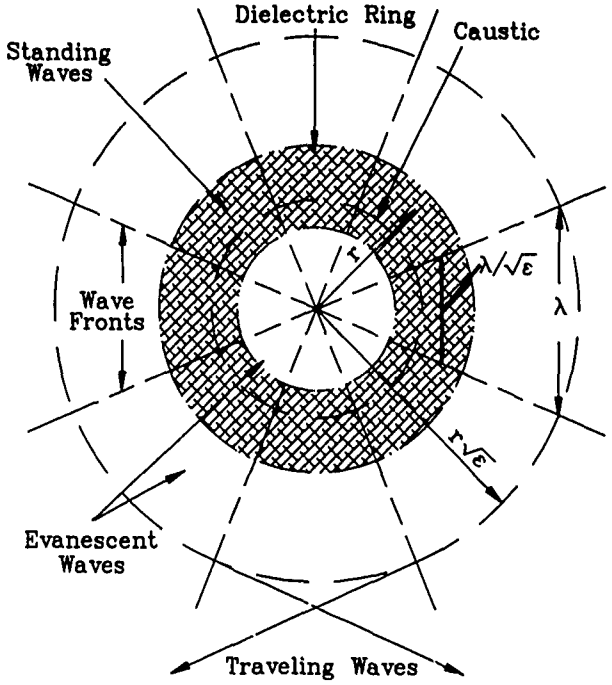


Figure 2: Diagram showing the electromagnetic field's character in the dielectric ring's vicinity for an 8-fold cylindrically symmetric mode.

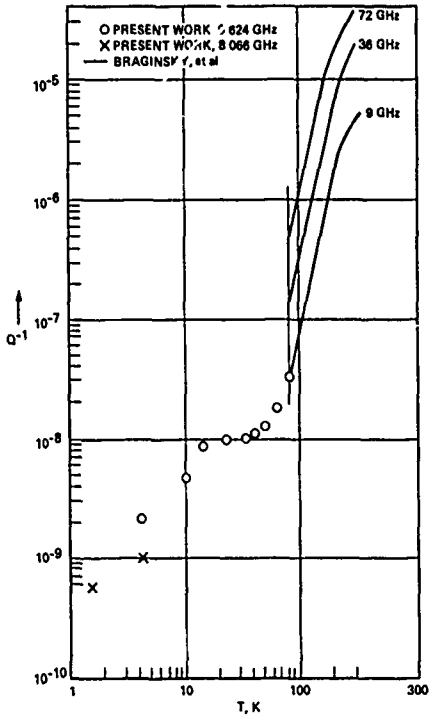


Figure 3: Q measurements for a sapphire cylinder 5 cm diameter, 5 cm high contained in a lead-plated shielding can approximately 1 cm away. Also shown are higher-temperature data by Braginsky, et al, (reference 4).

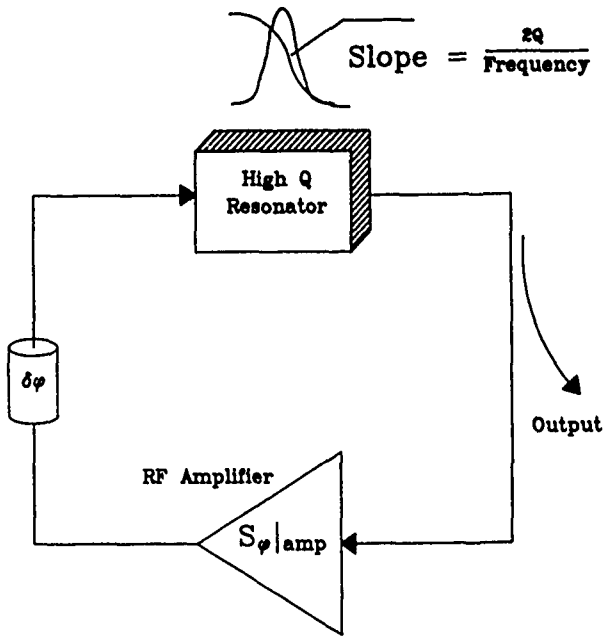


Figure 4: Block diagram of simple oscillator with direct RF feedback. Output phase noise is derived from amplifier noise together with phase slope of resonator. Phase is adjusted to give $2n\pi$ radians around the loop at the center of the passband.

Microwave Oscillator Design

The second part of a quiet microwave oscillator is an excitation mechanism with very low 1/f phase noise. Passive phase detectors show a substantial advantage over presently available active devices. This means that for any given resonator Q, it is possible to construct a quieter frequency discriminator than the corresponding active oscillator. Thus a stabilized oscillator, in which the frequency of a microwave oscillator is sensed by means of a whispering gallery-type discriminator, allows the lowest possible phase noise.

A stabilized oscillator frequency source is also advantageous in that it allows very effective mode selection (see figure 7). For this reason it is certainly the application of first choice for practical application and lowest phase noise.

Figures 4 and 5 show conventional microwave self-excited oscillator and STALO configurations together with an identification of the in-oscillator and oscillator output noise spectral densities. In the self-excited oscillator shown in Fig. 4, the oscillation condition requires that the phase shift around the complete feedback loop comprising the amplifier, resonator, and interconnections be a multiple of 2π. With this condition satisfied, any phase fluctuation in the microwave amplifier must be accompanied by an opposite shift of equal magnitude in the resonator. For slow phase fluctuations

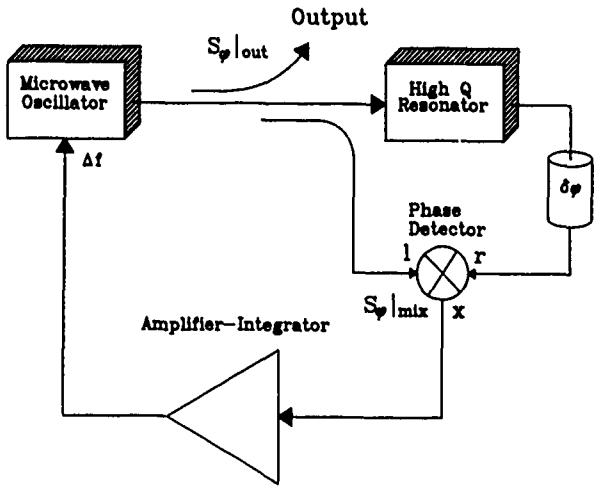


Figure 5: Block diagram of stabilized local oscillator (STALO) with double-balanced mixer type phase detector. Mixer noise plays the same role as amplifier noise in Fig. 4. Phase is adjusted to give l and r signals in quadrature.

($f \ll \nu/Q$), the characteristic phase slope of the resonator $\partial\phi/\partial\nu = 2Q/\nu$ implies a corresponding slow fluctuation in the frequency of the oscillator. In this way, a power spectral density of phase fluctuations for the amplifier $S_\phi(f)|_{amp}$ results in oscillator output frequency noise

$$S_y(f)|_{out} = (2Q)^{-2} S_\phi(f)|_{amp} \qquad (1)$$

or the mathematically equivalent output phase fluctuations

$$S_\phi(f)|_{out} = \frac{(2Q)^{-2} S_\phi(f)|_{amp}}{f^2}. \qquad (2)$$

Here f represents the fluctuation frequency, ν the microwave frequency, Q the quality factor of the resonator and $y \equiv \delta\nu/\nu$ is the fractional frequency deviation.

Figure 5 shows the schematic diagram for a STALO in which the frequency variations of a noisy microwave source are cancelled by a feedback loop that detects the consequent phase shifts across a high-Q resonator to generate a frequency correction voltage. In the limit of large loop gain the oscillation condition requires that the phases at the two input ports of the mixer are in quadrature (mixer output = zero). A significant advantage of the STALO is that the properties of the feedback loop are particularly easy to control since the signal is mixed down to baseband (near zero frequency). This allows the use of active filters with narrow bandwidths and sophisticated response shapes which are not possible at microwave frequencies. It is easy to see that the form of the analysis for the STALO is identical to that above except that the mixer noise $S_\phi(f)|_{mix}$ replaces $S_\phi(f)|_{amp}$ as the noise source.

The consequence is that performance of a direct feedback oscillator with an amplifier with 1/f noise of -120 db/Hz at 1 Hz ($S_\phi(f)|_{amp} = 10^{-12}/f$ $radians^2/Hz$) is

$$S_\phi(f)|_{out} = 10^{-12}(2Q)^{-2}\left(\frac{\nu^2}{f^3}\right). \qquad (3)$$

A STALO using a mixer with 1/f noise of -135 db/Hz at 1 Hz will be 15 db quieter.

Oscillator Considerations

In order to be of practical use, output of the reference oscillator must not only have low noise, but also good accuracy and long term stability. Since the $10^{-5}/°K$ thermal expansion coefficient of the sapphire resonator presently considered is uncompensated, it will be designed as a "clean up" oscillator, stabilized by a quartz crystal oscillator. Several methods are available for compensation of the sapphire resonator's thermal frequency coefficient, including one (incidental impurity doping) which has been used successfully in the temperature range 5°K - 10°K[3]. Successful compensation of the linear thermal coefficient would allow stability of better than 1 part in 10^9 in the temperature range 77°K - 100°K.

We expect that phase locking the sapphire STALO to a stabilizing quartz oscillator would be effected in two stages. A fast loop would be implemented with a tuning range of a fraction of a cavity bandwidth (say 1/3Q). For a Q of 10^6 at 170°K, for example, temperature stability of 0.01°K would be required to stay within this range. A second loop would sense the stress in the fast loop and adjust the temperature of the resonator to keep it within the required range.

It is possible that extension of the temperature range of impurity doping-type compensation may allow compensation to be effected at higher temperatures. Although the compensated temperature varies only as the 1/4 power of impurity concentration,[3] operation very near 10.5 GHz chromium ion resonance frequency is expected to enhance the effect.

Experimental

Mode Frequencies and Coupling Constants

Figure 6 shows the resonator test configuration including input and output WR112 waveguide ports which are

Figure 6: Resonator test configuration.

centered on the axis of the sapphire element and container. The cross-section of the sapphire ring was chosen to be twice as tall (z direction) as wide, (r direction) for a principal (lowest frequency) mode family with electric field polarized in the z direction. This corresponds to an $H_{n,1,1}$ mode family with H essentially transverse to the z axis. This choice was made to allow effective coupling to the principle modes by virtue of similar field configurations in the waveguide and resonator.

Figure 7 shows the frequency and Q for modes of the ring resonator between 5GHz and 9GHz. These modes are all doubly degenerate, a fact which was noted for many of them during the measurement process. The two modes correspond to physical orientation of the fields rotated by $\pi/(2n)$. We confirmed that mode splitting was due to the coupling ports by observing that, for strongly coupled modes, they invariably resolved into a pair, one of which was so weakly coupled as to require mode spoiling in order to observe its presence. This means that the modes are accurately oriented to the ports' position, since only then would rotation by 1/4 wave result in exact cancellation of the signal at the coupling ports. Typical splitting of the modes was observed to be 10^{-5} to 10^{-6} fractional frequency deviation.

The modes typically show $n = 3$ to $n = 10$ full waves around the perimeter of the ring. Measurement of n was accomplished for many modes by rotating a tee shaped mode spoiler inside the containing can, and observing the changing interference pattern in the due to simultaneous excitation of both modes for each pair of modes. The observed values for n are also shown in figure 7.

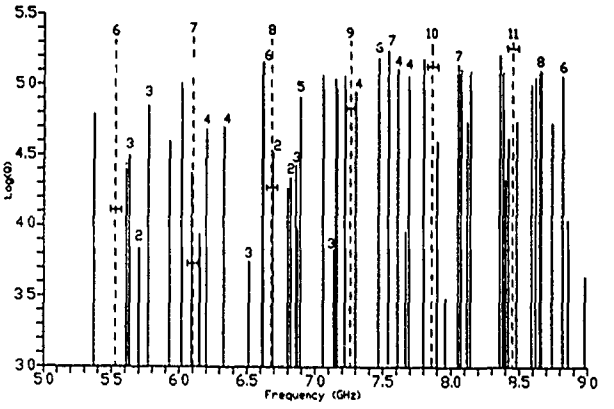


Figure 7: Sapphire ring resonator mode frequencies and Q's at room temperature. The principal mode family is identified by dashed lines with coupling Q's indicated by cross-bars.

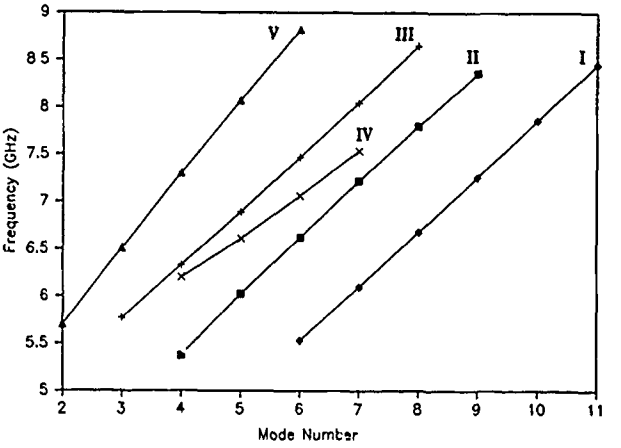


Figure 8: Mode family identification for selected modes from Figure 7.

The micrometer-driven waveguide shunts shown in figure 6 provide some degree of variable coupling, but interference effects in the distance between the shunts and the port openings limit their utility. Q values plotted in Fig. 7 all represent the weakly coupled element of the mode pair whose Q is not spoiled by the presence of the coupling ports. Loaded values Q_l were measured for the more strongly coupled of the pair in those cases where significant coupling obtained. Measured values for the $H_{n,1,1}$ family are indicated by cross-bars. Because the $H_{10,1,1}$ was almost exactly critically coupled, we chose that mode for further study.

We have identified 4 or 5 mode families among the many modes in Fig. 7. They were identified primarily by their frequency progression with increasing n, and also by other characteristics. Five tentatively identified mode families are shown in Figure 8. The familics have qualitative features that distinguish them from each other and help to allow their identification. Generally, the following

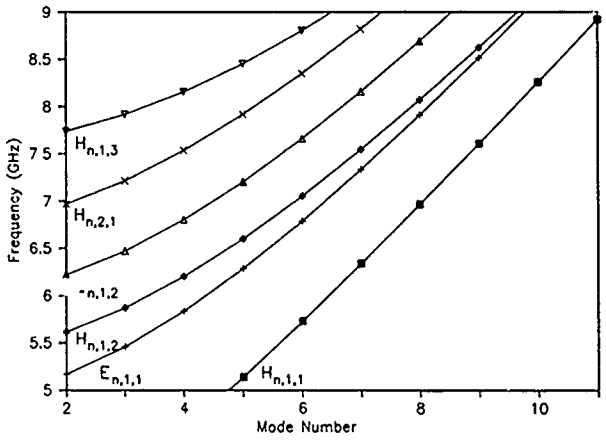


Figure 9: Calculated modes from dielectric waveguide model. Notation is similar to that used by Guillon et al[8].

considerations hold true.

- Families II, IV, and V show vanishingly small coupling to the waveguide ports, indicating a mode configuration orthogonal in some way to the waveguide fields.
- Families I, II, and to a lesser extent III show only slightly diminished Q when the top of the containing can is opened.
- The frequencies of families I and II decrease as a circular metal plate approaches the top; those of family IV increase.

We identify modes as $E_{n,m,l}$ and $H_{n,m,l}$, where E and H refer to modes with electric and magnetic fields principally transverse to the z direction. Here n refers to the number of full waves in the azimuthal direction, and m and l count the field maxima in the radial and axial directions, respectively.

An earlier paper has presented the results of calculations of mode frequencies and of Q's due to wall losses based on a curved dielectric waveguide model.[6,9,10] Figure 9 shows a plot of modes based on this calculation over the same range of frequency and mode number used in Fig. 8. All the lowest lying modes are shown. The notation has been changed from [6] to be more in line with that used by Guillon.[8]

Comparing Figures 8 and 7, together with the considerations listed above lead us to the following mode identifications. Modes I and II are surely the $H_{n,1,1}$ and $H_{n,1,2}$ families, respectively. The qualitative features of I and II are identical, except for the coupling, which would vanish for the $H_{n,1,2}$ family due to its antisymmetric electric fields about an imaginary mid-plane divider. Family IV is probably $E_{n,1,1}$ which should also show no coupling, due to symmetry. Family III may be either $E_{n,1,2}$ or

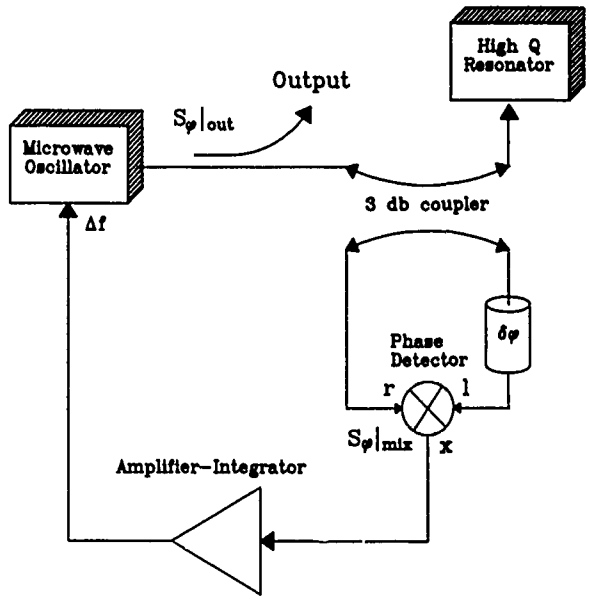


Figure 10: STALO with different configuration but functionally identical to that of Fig. 5. Signal returned from resonator is superposition of resonator signal and (constant) reflected signal.

$H_{n,1,3}$, neither of which would give an exact cancellation of the coupling; and V may be either of these, or $H_{n,2,1}$.

Qualitatively, the very simple model gives excellent predictions, the most apparent difference between calculated and experimental results being a more nearly straight-line performance of the experimental data. This may be explained qualitatively as due to an increase in effective diameter for very low mode numbers, where fields are not so tightly bound to the sapphire. No explicitly cylindrical considerations were used in our model, except for the use of a periodic boundary condition.

Oscillator Measurements

The configuration of the STALO used for our tests is shown in Figure 10. While mixer performance is identical that shown in Figure 5, use of a single port avoids loading the resonator twice, thus reducing its Q. The actual signal processed by the mixer is very different from that for the configuration in Fig. 5, where a signal of nearly constant amplitude, but varying phase is applied to the mixer. Figure 11 shows the mixer signal which "reflects" from the resonator as the source is tuned through resonance. This signal is the sum of two parts, one of which is the phase-varying signal just discussed, and the other is a signal of constant amplitude and phase. At exactly critical coupling, no signal is returned on resonance. However, the added signal is in quadrature with that applied to the mixer's other port, and so has only a small effect on mixer output. Figure 12 shows phase noise calcula-

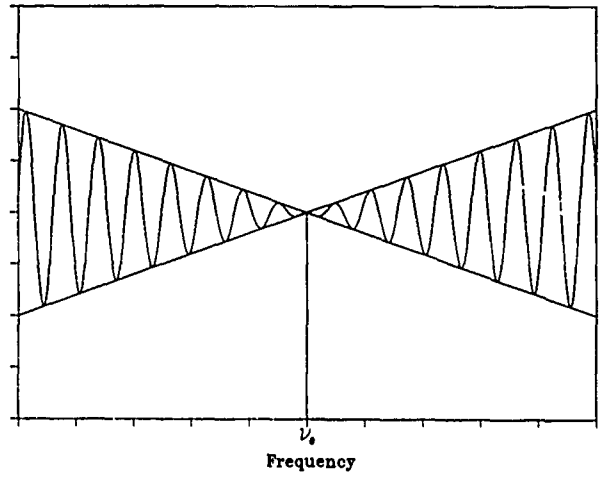


Figure 11: RF envelope of returned signal for critical coupling as frequency ν is varied. Phase inversion at center allows linear dependence in mixer output with frequency for arbitrarily small frequency errors.

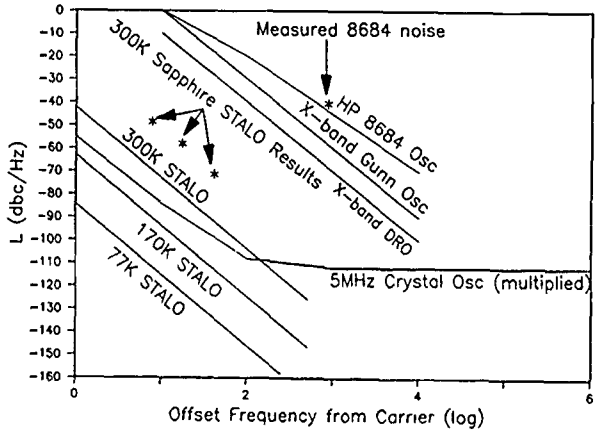


Figure 12: Phase noise calculations and measurements for sapphire "whispering gallery" mode STALO shown in Figures 10 and 6. Use of a cryogenic (170°K to 77°K) sapphire resonator allows further improvement by 20 to 43 db. Noise plots for various conventional X-band frequency sources are also shown. The multiplied 5 MHz crystal oscillator presently represents the best performance available at X-band.

tions for a sapphire "whispering gallery" mode STALO at various temperatures as well as experimental data for the STALO just described. These results are the lowest to our knowledge for an X-band source,[11,12] The difference between experimental and calculated results is inferred to be due to mixer noise of −115dB/Hz compared to the best available value of −135dB/Hz at 1 Hz offset. Quality factors are assumed to be $Q_i = 2 \cdot 10^5$ at room temperature, $Q_i = 2 \cdot 10^6$ at 170 Kelvin, and $Q_i = 3 \cdot 10^7$ at 77 Kelvin. Noise plots for various conventional X-band frequency sources are also shown. The multiplied 5 MHz crystal oscillator presently represents the best performance available at X-band.

Applications

Spacecraft radio science experiments examine fluctuations of phase, amplitude and frequency of radio signals propagating through planetary and inter-planetary atmospheres, ionospheres and rings before reception on Earth. The resolution and accuracy of their data depend fundamentally on the stability and spectral purity of the reference oscillator. The reference oscillator primarily impacts four areas:

- The resolution of ring experiments for the determination of the size of constituent particles is directly related to the spectral purity of the oscillator. The proposed oscillator will enable improvement of resolution of a factor of 20 compared with a conventional USO.
- Atmospheric occultation experiments would also benefit, e.g. measurement of the ionosphere of Titan.
- The range of solar wind measurements is presently limited by the bandwidth of the received signal. The effect on bandwidth due to the solar wind varies as r^{-3}; an 8 times reduction in source bandwidth would double the range of the experiments.
- S/N and reliability for the data down-link may be improved.

Power consumption of the sapphire-stabilized USO would be roughly the same as for the USO and multiplier chain without stabilization. Two additional stages of RF buffering are required, as well as baseband feedback, and thermal stabilization circuitry. Cooling requirements are estimated to be 15 mW at 77°K of which 1 mW is RF dissipation in the resonator. Mass of the sapphire element itself is 100 g, and total added mass except for the required radiative cooler is expected to be less than 500g.

Acknowledgements

The authors would like to thank L. Maleki and L. Primas for their helpful suggestions.

References

[1] D. G. Blair and S. K. Jones, IEEE Trans. Magnetics **MAG-21**, 142, 1985.

[2] D. G. Blair and I. N. Evans, J. Phys. D: Appl. Phys., **15**, 1651-1656 (1982).

[3] A. Giles, S. Jones and D. Blair, "A High Stability Microwave Oscillator Based on a Sapphire Loaded Superconducting Cavity", This Conference.

[4] V. B. Braginsky, V.P. Mitrofanov and V.I. Panov, *Systems with Small Dissipation*, Univ. of Chicago Press, Chicago, 1985, 85-89;

[5] "Frequency Stabilization of oscillators with high-Q leucosapphire dielectric resonators", V. I. Panov and P. R. Stankov, *Radiotekhnika i Electronika* **31**, 213, 1986,(In Russian).

[6] G.J. Dick and D.M. Strayer, "Measurements and Analysis of Cryogenic Sapphire Dielectric Resonators and DRO's," *Proc. 41st Ann. Frequency Control Symposium*, 487 (1987).

[7] S. Thakoor, D. M. Strayer, G. J. Dick and J. E. Mercereau, "A Lead-on-Sapphire Superconducting Cavity of Superior Quality," *J. Appl. Phys.* **59**, 854, 1986.

[8] X. H. Jiao, P. Guillon, and L. A. Bermudez, "Resonant frequencies of whispering-gallery dielectric resonator modes", *IEE Proceedings* **134**, Pt. H, 497 (1987).

[9] M. Gastine, L Courtois, J. L. Dormann, IEEE Trans. *Microwave Theory and Techniques*, **MTT-15**, 694 (1967).

[10] E. A. J. Marcatili, *Bell System Technical Journal*, **48**, 2071(1969).

[11] R. G. Rogers, "Theory and Design of Low Phase Noise Microwave Oscillators," *Proc. 42nd Ann. Frequency Control Symposium*, 301 (1988).

[12] This compares to the best cavity oscillator so far discovered by the authors; F. Walls, private communication.

[13] 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.

43rd Annual Symposium on Frequency Control - 1989

MICROWAVE PROPERTIES OF THIN FILMS OF HIGH-T_c SUPERCONDUCTOR Y-Ba-Cu-O

M. Golosovsky+*, D. Davidov+, C. Rettori+* and A. Stern*

*TIME and FREQUENCY LTD., POB 1972, Holon 58117, Israel

+Racah Institute of Physics, Hebrew University, Jerusalem, Israel

*Lady Davis Fellow. Permanent address:
Instituto de Fisica, Universidade Estadual UNICAMP 13081
Campinas, Sao Paulo, Brasil

ABSTRACT

We have measured the magnitude and phase of the microwave transmission through thin films of Y-Ba-Cu-O on MgO substrate in the presence of a small magnetic field. Such films will be used for the construction of high-Q superconducting cavity. The microwave transmission through these films is due as to the leakage through microcracks, as much as to the finite value of complex conductivity. The real and imaginary parts of the complex conductivity and their dependence on temperature and magnetic field are evaluated.

The presence of a constant magnetic field appreciably changes the magnitude and phase of the microwave transmission through superconducting films, while the presence of oscillating field modulates it with the generation of many high order harmonics. This strong nonlinear behavior may be explained by a model which assumes an array of Josephson junctions between superconducting grains.

****

The high-T_c oxide superconductors offer a possibility to build a high-Q microwave cavity operating above the liquid nitrogen temperature for applications in frequency control as a secondary source of a stable frequency.

Presently, there are oxide superconductors with the transition temperatures in the range 95-125K, well above the liquid nitrogen temperature. In view of the further progress in synthesis of materials with high T_c to build high-Q superconducting cavity operating above 77K from oxide superconductors seems realistic. Recently, the superconducting cavity from ceramic Y-Ba-Cu-O was built[1] but its Q-factor at 77K is $\sim 10^4$ which is not sufficiently high for frequency control applications. Only the superconducting cavity with Q-factor better than 10^6 may offer some advantages over crystal oscillators currently in use.[2]

Low Q of this cavity[1] arises from the great microwave absorption in ceramic superconductors[3]. It seems that epitaxial films could be better for construction of superconducting cavity. Indeed, absorption in best epitaxial films of Y-Ba-Cu-O prepared by laser ablation[3] is much more lower than in ceramics.

However, this films are very thin and microwave leakage through such a film may be sufficiently high that wouldn't permit to obtain high Q of the cavity. Hence, to find the optimal way of design of high Q cavity from oxide superconductors, all factors that determine the performance of such a cavity should be thoroughly analyzed. These are:

1. Absorption in the walls.

2. Leakage through the walls and coupling holes.

3. Frequency change due to variation of temperature, magnetic field, etc.

The losses in the cavity, Q^{-1}, and the frequency shift, $\Delta\omega/\omega$, arising from the material of the walls, can be expressed as follows:

$$\frac{\Delta\omega}{\omega} = \frac{Q^{-1}}{Xs/G} = Q^{-1}_{a} + Q^{-1}_{r}, \quad Q^{-1}_{a} = Rs/G, \tag{1}$$

where Q^{-1}_{a} - absorption losses in the walls, Q^{-1}_{r} - radiation losses, Rs - surface resistance, Xs - surface reactance, G-geometrical factor of the mode.

Rs and Xs can be expressed through the complex conductivity of superconductor, $\sigma_1 - j\sigma_2$.

$$R_s + jX_s = \left(\frac{j\omega\mu_0}{\sigma_1 - j\sigma_2}\right)^{1/2} \tag{2}$$

where ω - microwave frequency, μ_0 -magnetic permeability .
Radiation losses depend on the mechanical design of the cavity. Material parameter that determines the leakage through the walls is penetration depth, λ (the length on which electromagnetic field penetrates into superconductor).

$$\lambda = (\mu_0 \omega \sigma_2)^{-1/2} \tag{3}$$

Hence, the material parameter that determines the performance of superconducting cavity, e.g. $Q^{-1}, \lambda, \Delta\omega/\omega,$ is the complex conductivity of superconductor.

The aim of this work is to measure the complex conductivity of thin films of Y-Ba-Cu-O and its dependence on temperature and magnetic field.

CH2690-6/89/0000-115 $1.00 C 1989 IEEE

The experiments were carried out on thin films (~1 μm thick) on MgO substrate with a thickness of 1-2 mm. The substrates were prepared by the cleavage of large MgO single crystals originating from the Dead Sea. The superconducting films were prepared by the spray pyrolysis method[4]. At least six different films (all exhibiting superconducting transition temperature around T ≈ 85K) were studied with consistent results. These films retain their superconducting microwave properties for at least a year after preparation.

The transmission measurements were carried out at X-band frequency (9.5 GHz). The experimental set-up is shown in Fig. 1. The film-substrate combination is held in a frame across the waveguide flanges or across a small iris in the center of the waveguide. Special precautions were taken to eliminate any microwave leakage except through the film. The incident power is ~ 100mW. The transmitted power is picked up electrically by a coaxial cable and measured with a crystal detector. The phase shift of the transmitted wave through the sample is determined by balancing the transmitted microwave with a reference arm using a precision attenuator and a reference phase shifter for a null readout. The sample is situated in a vacuum can immersed in liquid helium or nitrogen. The ac magnetic field perpendicular to the film is produced by a small coil around the vacuum can.

The dc magnetic field is parallel or perpendicular to the film and is produced by independent coils (Fig. 1). The modulated microwave transmission is analyzed using a spectrum analyzer or a lock-in amplifier.

The experimental results can be summarized as follows:

The temperature dependence of the intensity and phase shift of the transmitted wave in the presence and in the absence of the dc magnetic field is shown in Fig. 2. It is readily apparent that the microwave transmission properties are almost temperature independent in the normal state, in the range 85-100 K, but dramatically change upon transition to the superconducting state.

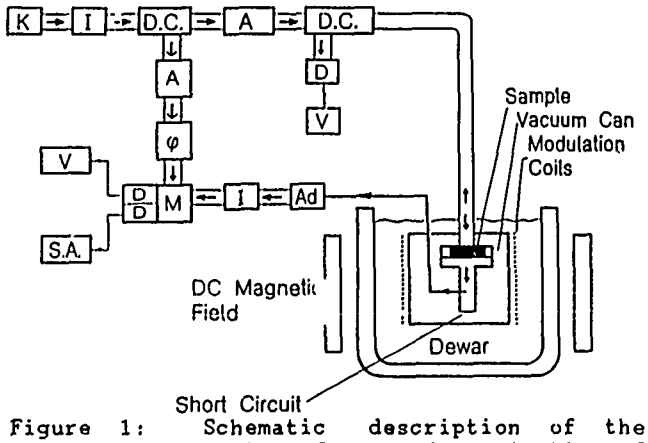


Figure 1: Schematic description of the experimental system for the determination of the transmission (both magnitude and phase) and reflection. K - klystron; I - isolator; D.C. - directional coupler; A - attenuator; V - digital voltmeter; φ - phase shifter; Ad-adapter; M - mixer; S.A. - spectral analyzer.

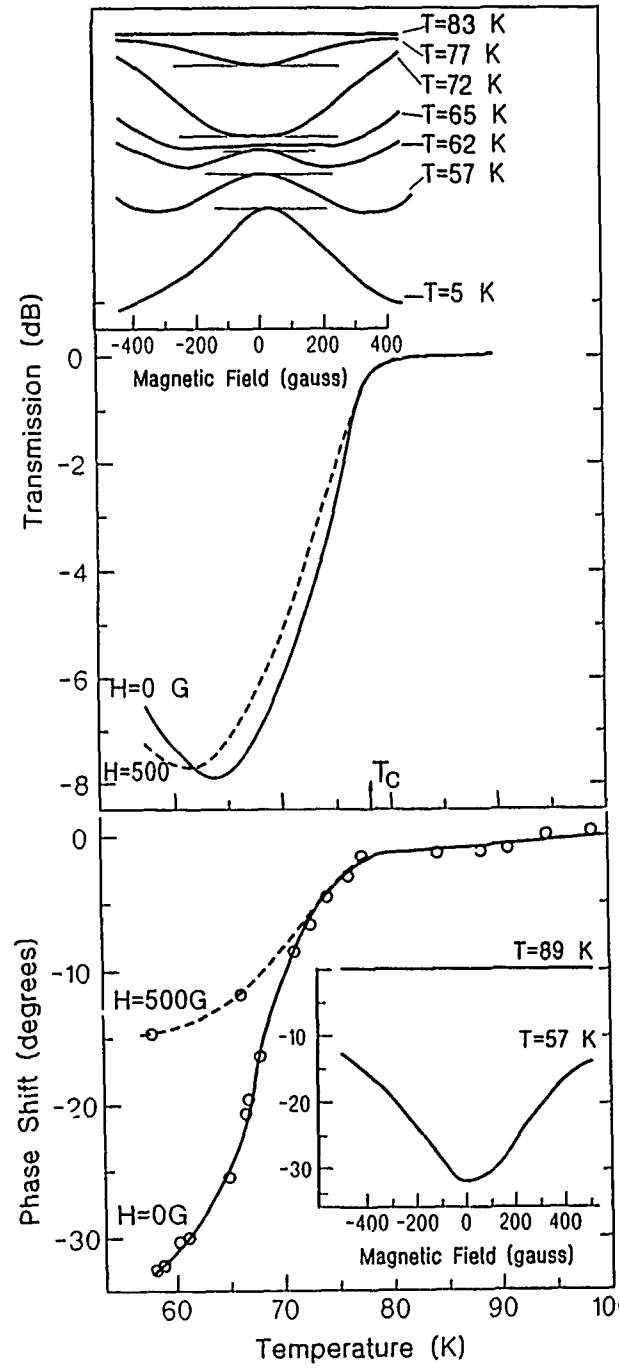


Figure 2: Upper panel Transmission amplitude versus temperature at H_o - OG and H_o = 500G. Insert: the transmission versus dc magnetic field at different temperatures. The level of the transmission at zero field is denoted by horizontal lines.

Lower panel Temperature dependence of the phase shift at Ho = OG and Ho = 500 G. Insert: the phase shift versus dc magnetic field in normal and superconducting states.

Particularly, the transmission sharply decreases, the reflection increases, and a negative phase shift appears upon transition to the superconducting state. The transmission goes through a minimum at about T = 65 K (Fig. 2, upper panel).

Following ref. 5, we have recalculated the transmission coefficient, T, and phase shift, $\Delta\varphi$, for a film of thickness, d, on a substrate with thickness, l , and refraction index, n, placed across a TE waveguide with impedance Z. In our calculations we have assumed that the film thickness, d, is significantly smaller than the normal state skin depth, δ , and the penetration depth in the superconducting state, λ.
These assumptions are justified in our case: δ is estimated to be 80 μm from dc conductivity measurements, and in films λ is estimated to be between 1 μm and 3 μm at zero temperature[6,7]. However, λ should be significantly larger in the vicinity of T_c, where most of our experiments were carried out. One obtains for the transmission coefficient, T, and the phase shift, $\Delta\varphi$, the following expressions:

$$T=4n^2\{[(2+\sigma_1 dZ)^2 + (\sigma_2 dZ)^2]n^2\cos^2 kl + \{(n^2+1+\sigma_1 dZ)^2 + (\sigma_2 dZ)^2]\sin^2 kl - n(n^2-1)\sigma_2 dZ\sin 2kl\}^{-1} \quad (4A)$$

$$\Delta\varphi = \text{arctg}\{[(n^2+1+\sigma_1 dZ)\sin kl - n\sigma_2 dZ\cos kl]/[(2+\sigma_1 dZ)n\cos kl + \sigma_2 dZ\sin kl]\} \quad (4B)$$

Here k is the propagating wavevector in the substrate. Using equations (4A) and (4B) as well as the experimental data (Fig. 2) we have evaluated the real and imaginary part of the complex conductivity σ_1 and σ_2. It is clearly seen that both σ_1 and σ_2 increase upon going to the superconducting state; σ_1 goes through a maximum at 65K, while σ_2 saturates at low temperatures. The levelling off of σ_2 at low temperatures can be attributed to microwave leakage through microcracks and normal regions[7].

Penetration depth, λ , calculated from (3) and extrapolated to zero temperature is 3μm. It is consistent with other microwave measurements[6,7], but is an order of magnitude higher than the value obtained by other methods[8]. The discrepancy is due to the granular structure of oxide superconductors[9]. Indeed, the microwave penetrates into granular superconductor mostly through grain boundaries which are regions of weak superconductivity and the penetration depth obtained in microwave transmission experiments is really Josephson junction penetration depth[9].

Thus, the length on which microwave penetrates into oxide superconductor is unusually high, 3μm at zero temperature and even higher at 77K. Then the walls of superconducting cavity should be sufficiently thick to prevent leakage. Presently it is difficult to prepare films of high quality with the thickness ~ 10 μm. The use of dielectric loading or special container should decrease radiation losses due to the leakage through the walls but these losses remain still a serious problem.

Consider now the effect of magnetic field on microwave transmission and phase shift.

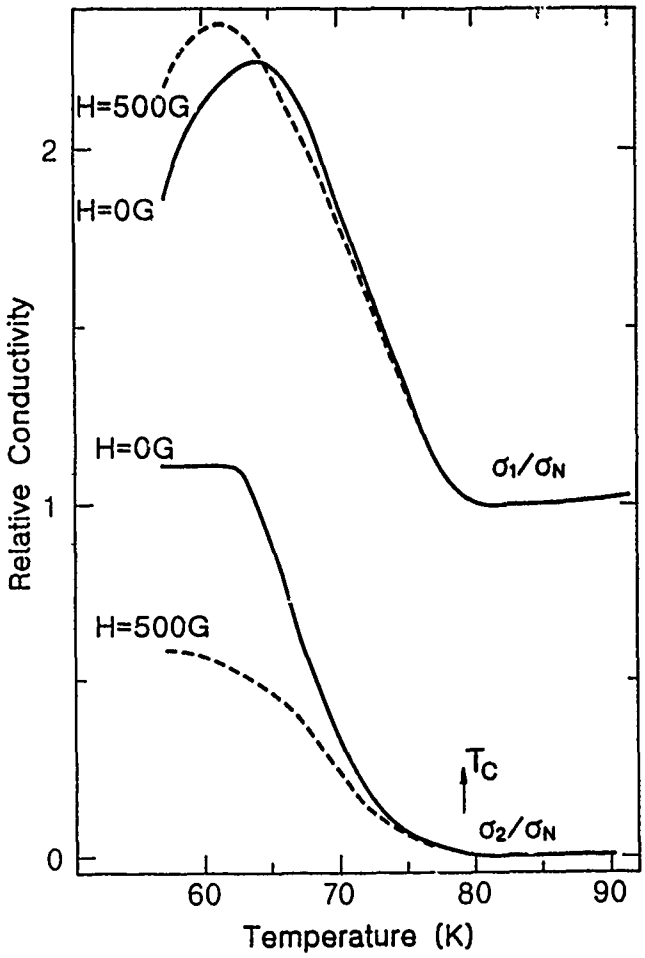


Figure 3: The temperature dependence of σ_1 and σ_2 calculated using the data in Fig. 2 and equations 4A and 4B.

a) Applying a dc magnetic field does not affect the microwave properties in the normal state but has a significant effect in the superconducting state. Particularly, a dc magnetic field enhances the microwave transmission just below the transition temperature in the range 80K-65K (Fig. 2, upper panel) but decreases the transmission for T <65K. The insert of Fig. 2 (upper panel) shows the dependence of the transmission on the magnetic field, H_o, at different temperatures. Note that the transmitted power versus H_o exhibits a clear minimum at zero field above 65 K, but a clear maximum below 65 K. At T = 65K the transmission is almost independent of H_o. The phase shift always decreases when a magnetic field is applied (Fig. 2, lower panel). Fig. 3 shows change of complex conductivity in magnetic field.

b) Applying an audio frequency magnetic field, $H_1 \sin 2\pi ft$, has no effect on the transmitted power in the normal state (Fig. 4). However, an ac magnetic field modulates the transmitted power in the superconducting state and generates even higher harmonics (2f, 4f, 6f etc.) of considerable intensity.

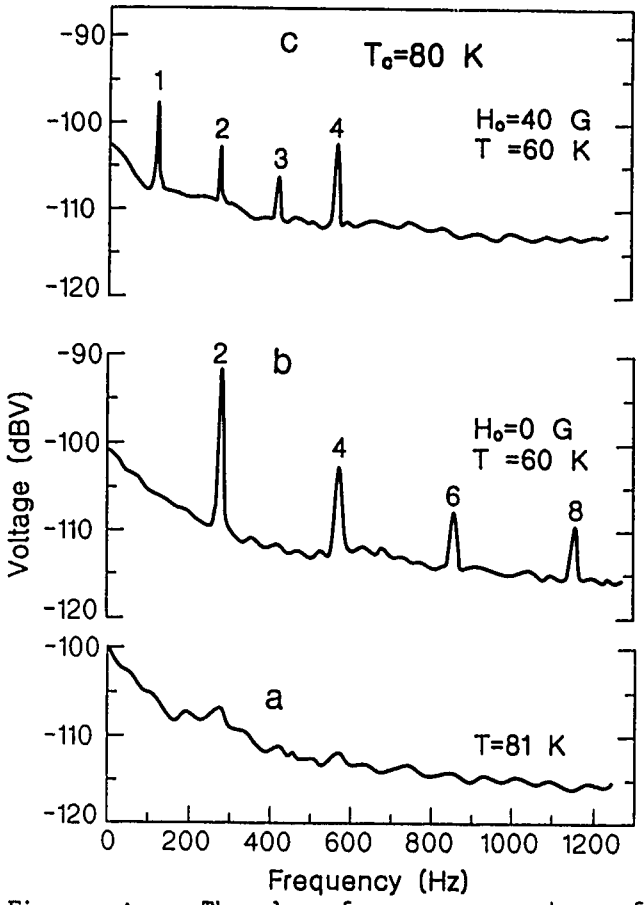


Figure 4: The low frequency spectrum of transmitted power through a film (T_c = 80 K) in the presence of an a.c. magnetic field perpendicular to the film. (a) Normal state, H_1 = 15G, f = 132 Hz, T = 81K. (b) Superconducting state, H_1 = 15G, f = 132 Hz, Ho = 0, T = 60 K. (c) Superconducting state, H_1 - 15G, f = 132 Hz, Ho = 40G, T = 60 K. Note the even harmonics in (b) but the generation of additional odd harmonics in (c).

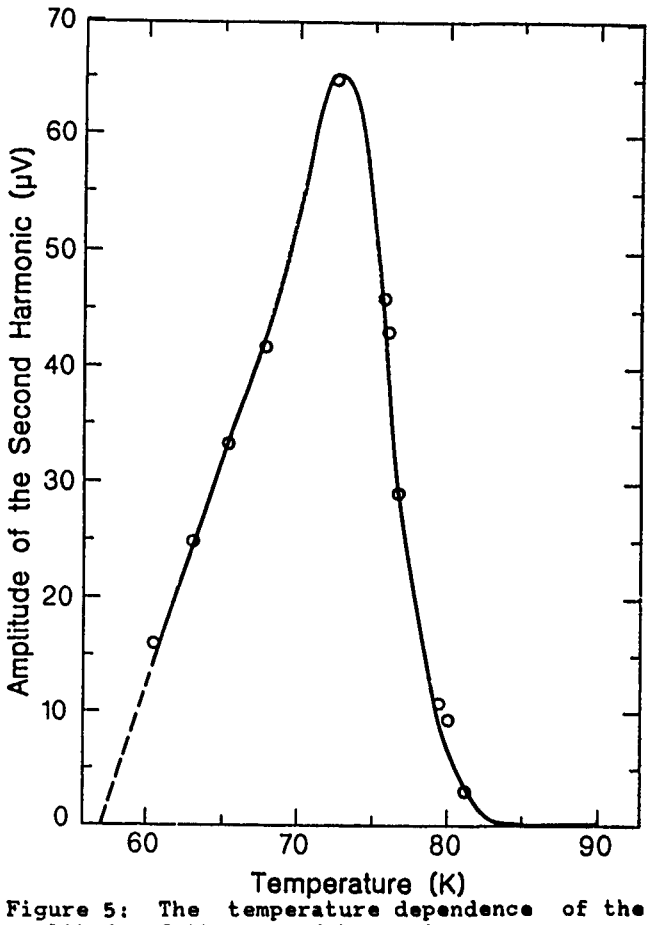


Figure 5: The temperature dependence of the amplitude of the second harmonic.

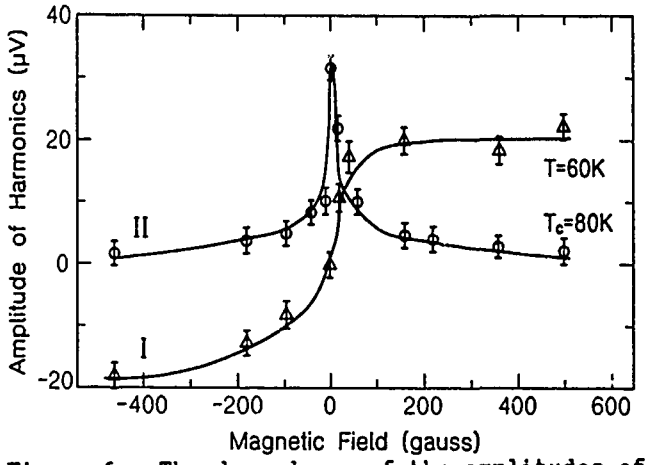


Figure 6: The dependence of the amplitudes of the first and second harmonics on the magnitude of the dc magnetic field parallel to the film. H_1 = 15G; f - 132 Hz.

Fig. 4 exhibits the lower frequency power spectrum of the transmitted wave in a small oscillating magnetic field. Fig. 5 exhibits the temperature dependence of the amplitude of the second harmonic. As can clearly be seen, the intensity of the second harmonic dramatically increases below T_c and exhibits a clear maximum. It almost vanishes around T = 60 K (corresponding to the field-independent transmission, see insert of Fig. 2, upper panel). Although not shown here, the amplitude of the second harmonic increases again at lower temperatures, changes phase and goes to saturation. These features are consistent with the field dependence of the transmission (insert, Fig. 2, upper panel).

c) Applying an additional dc magnetic field, H_o, (either parallel or perpendicular to the film) results in symmetry breaking and in the generation of odd harmonics in addition to the even harmonics (Fig. 4). Fig. 6 shows the intensity of the first and second harmonic as a function of H_o. Clearly, a dc magnetic field of the order of 20 G significantly suppresses the second harmonic (Fig. 6).

We attribute the change of σ_1 and σ_2 in magnetic field to the appearance of vortices.

The motion of vortices in the microwave field produces additional dissipation and affects σ_1. In these granular superconductors σ_2 to a large extent is determined by superconducting currents flowing through weak links. The appearance of vortices reduces the critical current in these links and thus decreases σ_2.

The modulation of the microwave by an oscillating magnetic field and the observation of high order harmonics can be interpreted, at least in the vicinity of T_c, in terms of the network of Josephson junctions. Following ref. 10, 11 the amplitude of the microwave current, I_c, flowing through a junction can be expressed as follows:

$$I_c = (\Delta (T)/eR_N) \tanh[(\Delta(T)/2k_BT)] \times (\sin(HS/\Phi_o)/(HS/\Phi_o) \quad (5)$$

where R_N is the resistance in the normal state, S is the junction area, Φ_o is the flux quantum, and H is the magnetic field which can be expressed as

$$H = H_o + H_1 \sin 2\pi ft \quad (6)$$

Thus, I_c is time dependent and the microwave current is amplitude modulated. Using (5) and (6) together with a Fourier analysis, it is easy to demonstrate that in the absence of a dc field ($H_o = 0$) only even harmonics of the ac frequency appear; while, if $H_o \neq 0$, additional odd harmonics are generated in agreement with experimental data (Fig. 4). From (5) it is seen that with increasing H_o the second harmonic is significantly suppressed in good agreement with the data in Fig. 6. From the field dependence of the second harmonic amplitude (Fig. 6) it is possible to estimate the junction area, S. Taking SH = 1 for a field of 20G at which the amplitude of second harmonic is 0.7 of its maximum value, one obtains $S = 10^{-8}cm^2$. Certainly the assumption of "single" Josephson junction is only an approximation and a complete analysis of the dependence of the intensity of the various harmonics on field and temperature requires proper "averaging" over many junctions[11,12].

The temperature dependence of the second harmonic amplitude (Fig. 5) can be explained, however, by equation (5) only at T> 70K. This may suggest an additional mechanism for the harmonic generation at T <70K. We attribute the modulation of the microwave at lower temperatures to the motion of vortices under the action of the applied oscillating magnetic field[13]. Note that at these temperatures the change of transmission in magnetic field is dominated by energy dissipation in vortex motion (Fig. 3).

It is seen that magnetic field considerably changes microwave properties of oxide superconductor. From the data of Fig. 3 and (1) - (3) it can be calculated that magnetic field 1 G increases Q^{-1} for 2.10^{-4} and produces a frequency change $\sim 10^{-8}$ in the TE_{011} cylindrical cavity made from Y-Ba-Cu-O with conductivity in the normal state $\sigma_N = 10^4$ Ω^{-1} cm^{-1}.

Hence, a high Q superconducting cavity should be isolated from external magnetic field but a small field $\sim$ 100mG can be used for monitoring the resonant frequency of the cavity.

Conclusions:

1. The complex conductivity of Y-Ba-Cu-O thin films and its dependence on temperature and magnetic field is measured. The penetration depth is determined. It is shown that the leakage through the walls of superconducting cavity produces serious problems and special precautions should be made to reduce the leakage.
2. The microwave properties of Y-Ba-Cu-O are highly sensitive to magnetic field. To obtain high Q the proper shielding from the external magnetic field should be made.
3. Small magnetic field may be used for tuning the resonant frequency of the cavity.

Acknowledgments

The authors wish to thank Prof. J. Imry for interesting discussions and Prof. W. Low for kindly providing us with parts for the experimental system. This work was supported by a special fund of the Ministry of Industry and Commerce, Israel. Special thanks to Dr. H. Arbeli (deceased) for her advice and encouragement.

References

1. C. Zahopoulos, W.L. Kennedy, S. Sridhar, Appl. Phys. Lett. 52, 2168 (1988)
2. G.J. Iafrate, Proc. of the 42nd Annual Symp. on Frequency control, 1988, p. 540.
3. N. Klein, G. Muller, H. Piel, B. Roas L. Schultz, U. Klein, M. Peiniger, Appl. Phys. Lett. 54,757 (1989)
4. A. Gupta, G. Koren, E. A. Giess, N. Moore, E.J.M. O'Sullivan, E.I. Cooper, Appl. Phys. Lett. 52, 163 (1988).
5. N.M. Rugheimer, A. Lehoczky, C.V. Briscoe, Phys Rev. 154, 414 (1967).
6. W. Ho, P.J. Hood, W. E. Hall, P Kobrin, A.B. Harker, R.E.DeWames, Phy. Rev. B37, 7029 (1988).
7. C.S. Nichols, N.S. Shiren, R.B. Laibovitz, T.G. Kazyaka, Phys. Rev. B38, 11971 (1988).
8. A. T. Fiory, A. F. Hebard, P. M. Mamkiewich, R. E. Howard, Phys Rev. Lett. 61, 1419 (1988) and references therein.
9. J. R. Clem. Physica C 153, 50 (1988).
10. C. Jeffries, Q.H. Lam, Y, Kim, L.C. Bourne, A. Zettl, Phys. Rev. B37, 9840 (1988).
11. T.K. Xia, D. Stroud, Phys. Rev. B39, 4792 (1989).
12. Q.H. Lam, C.D. Jeffries, Phys. Rev. B39, 4772 (1989).
13. K. H. Muller, J. C. MacFarlane, R. Driver, Physica C 158, 69 (1989).

43rd Annual Symposium on Frequency Control - 1989

W BAND DIELECTRIC RESONATOR POWER COMBINER

D. CROS - P. GUILLON - XIAO JIAO
I.R.C.O.M. / C.N.R.S.
UFR Sciences - 123 Avenue Albert Thomas
87060 LIMOGES Cédex - FRANCE

Introduction

With encreasing frequency, the power output from a single solid-state device decreases rapidly and is limited to milliwatts of power.

It is therefore desirable to extend the power level in order to take advantage of the many desirable features of solid-state devices such as small size and weight, reliability and performance in a broader range of applications.

Although there are fondamental limitations to the power that can be generated from a single device, the achievable power level can significantly increased by combining a number of devices operating coherently.

The device presented here combines the output power of several millimeter wavelength devices in a single step by means of a whispering gallery modes dielectric resonator.

At millimeter wavelengths cylindrical dielectric resonators used in microwave integrated circuits become impractical. In effect, if the conventional TE, TM or hybrid modes are used D.R.'s will have dimensions that are impractically small. When used on their whispering gallery modes these cylindrical D.R. have dimensions larger than normal for millimeter wavelength. [1]→[8]

So, in this paper, after a brief description of the whispering gallery modes phenomena, we define both electromagnetic and electrical parameters of these resonators when they are coupled with transmission lines. Then, the design and the test of W band whispering gallery modes D.R. power combiner are given.

Whispering gallery modes (W.G. modes)

In a dielectric rod, the W.G. modes are described as comprising waves running against the concave side of the cylindrical boundary. The wave move essentially in the plane of the circular cross section. Most of the modal energy is confined between the cylindrical boundary and an inner modal caustic a_c.

The energy confinement can be explained from a ray optics point of view. In this way, it is said that a ray is totally reflected at the dielectric-air interface, it is then tangent to an inner circle called a caustic. Thus, the ray moves merely within a small region near the boundary (figure 1 a). From an analytical point of view, it is well known that waves guided in a dielectric rod can be described by BESSEL functions Jn(kr). The W.G. modes correspond to those for which the argument kr is of the order of n which is a large integer. In this way, it can be shown that the field is oscillatory between the rod boundary and a slightly small radius called a caustic, while it exponentially decays elsewhere (figure 1-b).

Whispering gallery modes have been excited in thin dielectric disks, the diameter D of which is greater than thickness 2d (D >> 2d). These planar resonators are compatible with the realization of millimeter wavelength integrated devices. [9]→[11]

Whispering gallery modes are classified as $WGE_{n,m,l}$ and $WGH_{n,m,l}$ modes : WGE modes have a large radial electrical field, while WGH modes have a large radial magnetic field, n,m,l, characterizes respectively the azimuthal, radial and axial variations.

Electromagnetic parameters analysis

As mentionned above WG modes move essentially in the plane of a circular cross section and most of the modal energy is confined between the resonator boundary and the inner modal caustic. As a result, the WG modes dielectric resonator is analogous to a travelling wave ring resonator, an analogy which is justified by the energy confinement, the existence of the modal caustic, and the fact that the field of the WG modes decays exponentially outside. For that dielectric ring the external radius coincides with the radius of the real resonator, while the interior one takes the place of the modal caustic.

This model can be used to define approximatively electromagnetic parameters : resonant frequency, electromagnetic field, unloaded Q factor of WG modes D.R. For better accuracy, it will be necessary to used a more sophisticated analysis and we have choosen finite element method. [4]

The finite element used are triangular and are associated with the two degree Lagrange polynomials. A large number of elements is necessary where there is an important field concentration. The number of triangular elements used determines the computing time requires to use the finite element program. The numerical analysis permits the calculation of the resonant frequency, the unloaded Q factor and also to draw the field cartography associated with each modes (figures 2-a and b).

Electrical analysis of WG modes

Before discussing the applications of WG modes D.R. in circuits, the structure of such a resonator coupled to transmission lines must be examined.

As seen previously, the resonator under consideration can be viewed as a travelling wave ring resonator. This analogy can help in the analysis of the coupling mechanism.

Let us consider a WG mode resonator coupled to two transmission lines (figure 3). The transmission lines are general, they may be metallic or dielectric waveguides, or microstrip lines. The system which forms

CH2690-6/89/0000-120 $1.00 C 1989 IEEE

a directional filter, consists of two directional couplers formed by the two transmission lines and the dielectric resonator. When the circuit is fed at port A, the transmission power from port A to port B is absorbed into the resonator because of the existence of coupler 1 and there is a travelling wave circulating within the closed loop structure. By the coupling 2 of coupler 2 we can get power at port C but nothing is coupled to port D. For the same reason the power obtained at port B is due to two parts : the power transmitted from port A and the power coupled from the resonator.

The scattering parameters of this directional filter satisfied the following relations :

$$|S_{BA}| = (1-k_1^2)^{1/2} + \frac{k_1^2 \sqrt{1-k_2^2}\, e^{-\alpha l}}{1 - \sqrt{(1-k_1^2)(1-k_2^2)}\, e^{-\alpha l}}$$

$$|S_{CA}| = \frac{k_1 k_2 e^{-\alpha/2\, l}}{1 - \sqrt{(1-k_1^2)(1-k_2^2)}\, e^{-\alpha l}} \qquad (1)$$

$$|S_{DA}| = 0$$

k_1 and k_2 are respectively the coupling coefficients of coupler 1 and 2, α is the total attenuation of the wave travelling in the resonator.

A lumped equivalent network of this directional coupler is given in figure 4. It consists to two resonant circuits, the coupling of those with transmission lines is achieved respectively by mean of a mutual inducance M, and by mean of a mutual capacity C_c.

W band power combiner

The configuration of the device is that presented in figure 3. For the combination on 2 sources, solid-state devices are applied respectively at ports (1), (3) for example, and the added power is obtained at port (2). Let be P_i the image power respectively at access (i=1,3 and P_2 the output power.

$$P_{2_i} = \left[\frac{\sqrt{1-k_1^2} - \sqrt{1-k_2^2}\, e^{-\alpha} + k_1 k_2 e^{-\alpha/2}}{1 - \sqrt{1-k_1^2}\sqrt{1-k_2^2}\, e^{-\alpha}} \right]^2 P_i \qquad (2)$$

$$P_2 = P_{21} + P_{23}$$

From (2) using :

- the lumped equivalent circuit (figure 4)
- the finite element method to compute electrical and EM parameters of the device, we obtain the variations of P_2 as function of the frequencies (for frequencies closed the resonant frequency of the WG D.R. modes - figure 5). We can also evaluate the efficient combination defined by (3).

$$\eta = \frac{1}{2} \left[\frac{1-k_1^2 - 1-k_2^2\, e^{-\alpha} + k_1 k_2 e^{-\alpha/2}}{1 - 1-k_1^2 \quad 1-k_2^2\, e^{-\alpha}} \right]^2 \qquad (3)$$

From (2) and (3), the conditions for maximum power combination are :

- the frequencies of single solid-state devices must be closed the resonant frequency of the W.G. modes
- the relation (4) between k_1 and k_2 must be satisfied

$$k_1 = k_2 \frac{2 - 1-k_2^2}{1 + k_2^2} \qquad (4)$$

A power combiner using 2 Gunn diodes and a planar dielectric resonator has been realized and tested in the W band.

- dielectric resonator : mode WGH
  diameter D = 6 mm
  thickness 2d = 0.7 mm
  relative permittivity $\varepsilon_r = 9.6$
- transmission lines : microstrip on alumina substrate
- transition between microstrip lines and metallic waveguides : ridge, VSWR 10 db between 90 and 100 GHz

The response of the power combiner near 94 GHz is given in figure 6. The efficient combination in the coupling plane is closed to 85 %.

Conclusion

Whispering gallery modes of dielectric resonators are suitable for millimeter wave integrated circuits.

Their high Q value and oversized dimensions permit designers to use them in both passive and active devices.

References

[1] D. KAJFEZ, P. GUILLON
"Dielectric resonators"
Artech House - 1986

[2] J. ARNAUD
"Note on the use of W.G. modes in communications"
Bell Labs. Memorandum - Sept. 1971

[3] J. ARNAUD
"Beam and fiber optics"
Academic Press - 1976

[4] MODULEF Computer Club
INRIA Domaine de Rocquencourt - FRANCE

[5] J.R. WAIT
"Electromagnetic W.G. modes in a dielectric rod"
Radio Science - vol.2 - n°89 - Sept. 1967- pp.1005-1017

[6] *L.B. FELSEN, T. ISHIHARA*
"High frequency surface field excited by a point source on a concave perfectly conducting cylindrical boundary"
Radio Science - vol.14 - n°2 - 1979 - pp.205-216

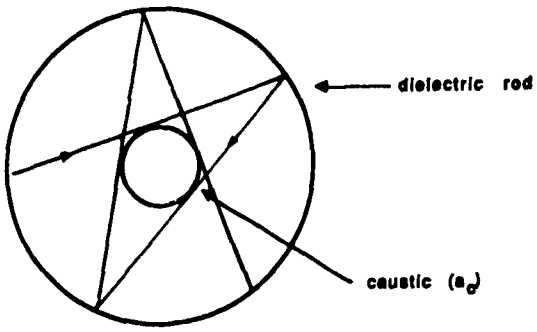
[7] *WEINSTEIN*
"Open resonators and open waveguides"
Crulem Press - 1969

[8] *C. VEDRENNE, J. ARNAUD*
"Whispering gallery modes of dielectric resonators"
IEEE Proceed. H - vol.129 - August 1982 - pp.183-187

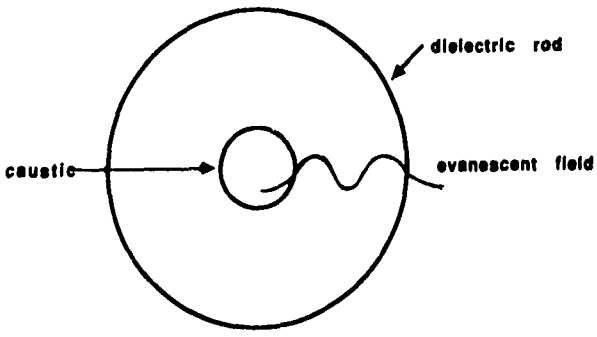
[9] *P. GUILLON, X.H. JIAO, Ph. AUXEMERY, L. BERMUDEZ*
"Whispering gallery modes herald D.R. millimeter wave use"
Microwaves and R.F. - September 1987 - p.85

[10] *L. BERMUDEZ, P. GUILLON, J. OBREGON, A. BERT*
"A 94 GHz low noise GaAs FET oscillator using W.G. D.R. modes and an new push-push configuration reducing 1/f converted noise"
1988 IEEE International Microwave Symposium - New-York - Juin 1988 - pp.481-484

[11] *D. CROS, Ph. AUXEMERY, P. GUILLON et al...*
"A W band D.R. power combiner"
1989 International Microwave Symposium - Long Beach - Juin 1989

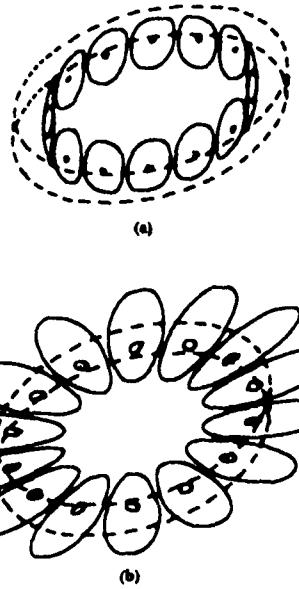


(a) **ray optics representation**



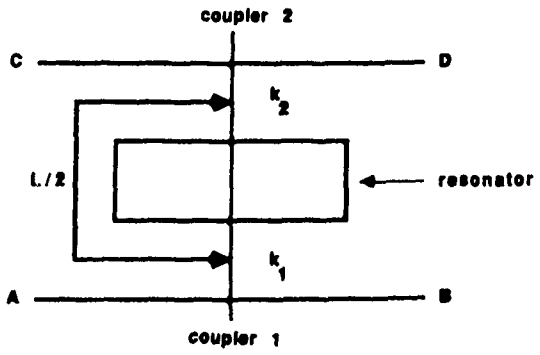
(b) **analytical representation**

- Figure 1 -
Whispering gallery modes

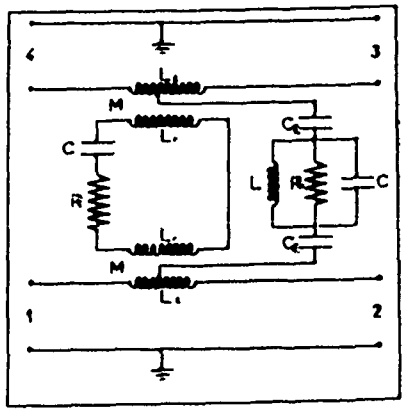


Magnetic field lines of W.G. modes of a D.R.
a) mode $WGE_{7,0,0}$ b) mode $WGH_{7,0,0}$

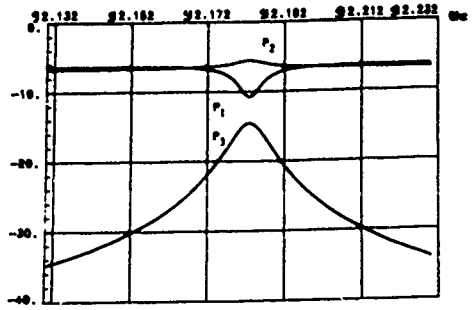
- Figure 2 -



- Figure 3 -
Coupling configuration



- Figure 4 -
Lumped circuit

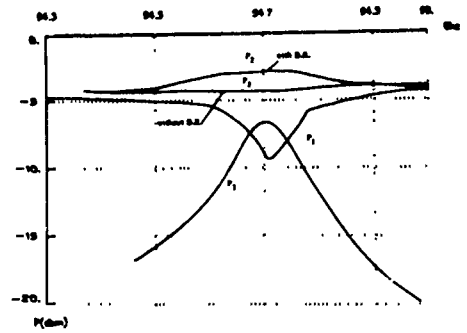


WGE MODE

$$k_1 \rightarrow \frac{M_1^2}{L_T} = 1,22\ 10^{-2}\ pH$$

$$k_2 \rightarrow \frac{M_2^2}{L_T} = 1,22\ pH$$

- Figure 5 -
Simulation



- Figure 6 -
Experimental results

43rd Annual Symposium on Frequency Control - 1989

TF-4000A, TFL'S HIGH PERFORMANCE RUGGEDIZED RUBIDIUM FREQUENCY STANDARD

A. Stern, M. Golosovsky, Y. Elbaz, A. Hertz and A. Lepek
TFL. Time & Frequency Ltd. P.O.B 1792, Holon 58117. Israel

Abstract

We report the progress achieved with the development of TFL's rubidium frequency standard, TF-4000A. Highlighted are short-term-stability ($5x10^{-12}/\sqrt{tau}$), temperature low sensitivity ($1x10^{-10}$/-55°C to +71°C) and low magnetic sensitivity (< $1x10^{-12}$/gauss) which is obtained with an ac "C" field method*. These are analyzed and discussed within a framework of models that are presented. Preliminary results of a study relating to the application of the ac "C" field are presented. We observe transient signals at the zero crossing of the "C" field. These are related to adiabatic and non-adiabatic processes that occur at a low and changing magnetic field.

Introduction and General Description

In the previous Symposium on Frequency Control we published a preliminary report on the development of TFL's TF-4000A - a compact and ruggedized Rubidium Frequency Standard [1]. The paper presented the design and preliminary performance results measured on an engineering model. In particular we described novel schemes aimed at improving stability under environmental disturbances such as magnetic field and vibration*. For the magnetic stabilization we introduced the ac "C" field instead of the conventional dc "C" field. The scheme for reducing the vibration sensitivity used an adaptive servo-loop time constant.

The present paper reports additional greatly improved results that were obtained. It also describes a more basic study relating to the ac "C" field method.

Table 1 highlights the key performance levels that have been achieved. Notable levels are short-term-stability of $5x10^{-12}/\sqrt{tau}$, very low temperature sensitivity of $1x10^{-10}$ over the temperature range, -55°C to +71°C and extremely low magnetic sensitivity of

< $1x10^{-12}$/gauss.

This low magnetic sensitivity is obtained using the ac "C" field method, discussed below.

The program of this paper is as follows. In sections, 2 and 3 we analyze in depth the short-term-term stability and the temperature sensitivity and present the data that was measured. The ac "C" field method was described at first time in [1]. Then it was shown that the sensitivity to external magnetic field can be reduced by 2-3 orders of magnitude or more. In section 4 we describe a further study aimed at solving phase-noise problems. we find that the oscillator phase-noise increases when applying the ac "C" field. We study the cause for this and find a solution to the problem.

SHORT-TERM-STABILITY	: $5x10^{-12}$ to $1.5x10^{-11}\sqrt{\text{tau}}$
DRIFT	: 1 to $4x10^{-10}$/month
TEMPERATURE	: 1 to $3x10^{-10}$ stability over -55°C to 71°C
VIBRATION	: 1 to $7x10^{-11}$ for 1/2 hr 5 g rms, random vibration
MAGNETIC	: < $1x10^{-12}$/gs (ac "C" field)
WARM-UP	: $5X10^{-10}$ within 2 min
SIZE	: 3.25x3.25x4.00 inch

Table 1: key performance levels and features for TF-4000A.

Short-Term-Stability

Model

Transfer of stability from the rubidium to the crystal oscillator and the theory of operation of the servo-loop was analyzed in many papers [2-5]. In the following analysis we reformulate some of the servo-loop equations to be applied to our system.

Consider the servo-loop block diagram shown in figure 1.

We write the current, i(t), at the photo-detector:

$$i = i_{DC} + i_{AC} + \delta i \qquad (1)$$

Where i_{DC} is the dc current due to background light, $i_{AC}(t)$ is the signal current and $\delta i(t)$ is the current noise.

$$i_{AC} = p\cdot(f_1-f_0)\sin 2\pi f_m t \qquad (2)$$

+ higher harmonics

Here f_1 is the input interrogating microwave frequency, f_0 is the Rb resonance frequency, f_m is the modulation frequency and p is the so called "slope" of the discriminator pattern which is defined by equation (2).
For $\delta i(t)$ we write,

$$\delta i(t) = \delta i_{DC}(t) + \delta i_{AC}(t) \qquad (3)$$

CH2690-6/89/0000-124 $1.00 C 1989 IEEE

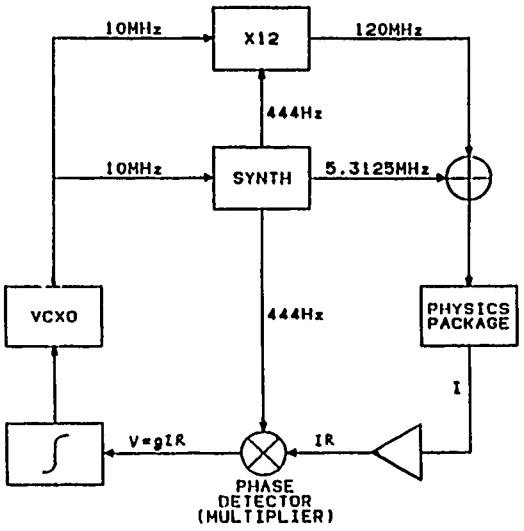


Figure 1: Servo-loop diagram for a rubidium frequency standard.

where $\delta i_{DC}(t)$ relates to direct current fluctuations induced by i_{DC} (shot-noise), and $\delta i_{AC}(t)$ arises from fluctuations in f and f_o. Here, we concentrate on fluctuations of f_o due to the light-shift effect:

$$\delta i_{AC} = p \cdot \delta f_o \cdot \sin 2\pi f_m t \qquad (4)$$

$$+ \text{ higher harmonics}$$

$$\delta f_o = f_o k_L \, \Delta I/I \qquad (5)$$

where k_L is the light shift coefficient and $\Delta I/I$ is the relative light intensity fluctuations.

The voltage V(t) after the phase-detector is given by

$$V(t) = g(t) \cdot R \cdot i(t) \qquad (6)$$

where R is the preamp response and q(t) is the demodulation function of the phase detector.
detector. When the loop is closed, $\overline{V(t)}=0$, where the average is calculated for frequencies lower then the loop bandwidth, f_L.

Calculating $\overline{V(t)}$ and equating to zero we obtain for the spectral density, $S_y(f)$, at low frequencies ($< f_L$),

$$S_y(f) = e \cdot i_{DC}/p \cdot f_o^2 + k_L \cdot S_{\Delta I/I}(f) \qquad (7)$$

The first term is due to shot-noise, with e, the electron charge (a reduction in the noise power by a factor of 2 is included due to the phase detector). The second term arises from light fluctuations which induce frequency fluctuations via the light-shift effect. $S_{\Delta I/I}$ the spectral density power of light fluctuations has the form, [5],

$$S_{\Delta I/I}(f) = a_L^2/f^2 + b_L^2 \qquad (8)$$

The transformation to the "Allan Variance" $\sigma_y(\tau)$ is done through, [6],

$$\sigma_y^2(\tau) = 2\int_0^\infty df \; S_y(f) \; \sin^4 x \; / \; x^2 \qquad (9)$$
$$x = \pi f \tau$$

Using the transformation formulas in [6], we obtain,

$$\sigma^2_y(\tau) = e \cdot i_{DC}/2(pf_o)^2 \cdot \tau^{-2} + (2\pi)^2/6 \; a_L^2 \cdot \tau \qquad (11)$$

where the second term in (8) is neglected.

Considering now environmental effects such as temperature, vibration, and external magnetic field. Each effect contributes a term to σ_y which includes a sensitivity coefficient and the measured fluctuations of the environmental effect. Here we consider only temperature fluctuations which are modeled as temperature drift (Due to thermal time constant of the unit, $\approx 10^3$ sec, the temperature fluctuations with time constant $< 10^3$ sec are insignificant. The period of long term temperature fluctuation is $\approx$24 hours, or 10^6 sec. Then for 10^3 sec $< t < 10^6$ sec temperature fluctuations can be treated as drift). i.e., $T = a_T \cdot t$. In addition, we include a direct drift term in σ_y. Hence.

$$\sigma_y^2(\tau) = e i_{DC}/2(pf_o)^2 \tau^{-2} + (2\pi)^2/6 \; a_L^2 \cdot \tau + (k_T^2 a_T^2/2) \cdot \tau^2 + (d^2/2) \cdot \tau^2 \qquad (12)$$

where $k_T = \Delta f/\Delta T$ is the temperature coefficient and d is the drift.

Optimizing the unit performance the following points were attended:
a. Slope maximizing. This is a part of physics-package parameters optimization. Relevant parameters are cavity and lamp temperatures, buffer gas pressures, lengths of resonance and filter cells, microwave power and light optics. Since the previous report a substantial improvement which is related to microwave power and light optics has been achieved.
b. Photo current fluctuations are limited by the shot-noise effect. This is not yet the case for our unit as evident from figure 1 blow.
c. Light shift coefficient should be minimized.

Table 2 lists several key parameters and measured values that were obtained and are relevant to the above model.

In figure 2 we plot the various contributions to σ_y, calculated from equation (12), together with measured data for two units, #1 and #2 Data for unit 1 was taken at room temperature with no temperature stabilization while data for unit 2 was taken with temperature stabilization of $\emptyset.2^oC$.

The following points are noted:
a. It is seen that at short averaging time ($< 10^2$sec) measured value of σ_y is larger than the calculated value derived from shot-noise Therefor a further improvement can be achieved.

PARAMETER		VALUE
dc photo current	i_{dc}	220 μA
slope	p	200 to 260 pA/Hz (rms)
light shift coef.	k_L	3×10^{-11}/10% change in intensity
temp coef.	k_T	2×10^{-12}/°C
drift	d	4×10^{-11}/month
relative light fluctuations	a_L^2	1.9×10^{-12} Hz
temperature drift	a_T	0.1 to 1°C/hour
loop time const		4 to 8 msec

Table 2: Key parameters and measured values used for model calculations.

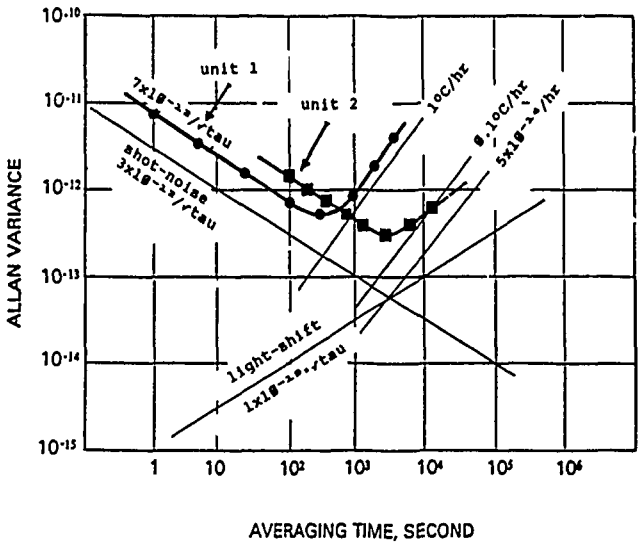


Figure 2: Short-term stability, $\sigma_y(\tau)$. Model calculations compared with measurements.

b. The minimum (floor) in the curves is dependent on environmental disturbances. With a good temperature stabilization we expect to reach the floor at $\approx1\times10^{-13}$ and around 10^4 sec.

Temperature Sensitivity

Frequency dependence on temperature is most important and quite complicated. However, one can identify and study the major effects that contribute to the frequency shift which is observed under temperature change. Furthermore these effects can be minimized and controlled so that they will compensate one another. These effects are summarized in table 3 and table 4. Table 3 lists the magnetic effects and table 4 lists non-magnetic effects. Consider first the magnetic effects.

The clock transition depends on magnetic field, H, via,

$$f_e = f_{oo} + a\cdot H^2 \qquad (13)$$

Where f_{oo} is the frequency at zero field and $a \approx 575$ Hz/gs² is a constant. We write H as a sum:

$$H = H_c + H_h + H_r \qquad (14)$$

Here H_c is the field induced by the "C" field coils, H_h is the field induced by the current in the heaters, and H_r is a residual field that emanates from magnetic materials (shields) and currents around the rubidium resonance cell.

We write

$$H_c = k_c\cdot I_c \qquad (15)$$

where I_c is the current through the "C" field coil. k_c is coil constant, which depends on the magnetic susceptibility of the various parts that are located inside and outside the coil and therefor is temperature dependents. The stability of I_c is determined by the stability of the resistors in the current supply.
Differentiating equation (13) we obtain,

$$\Delta f/f_e \approx 2aH_c/f_e\cdot \qquad (16)$$
$$(I_c\cdot\Delta k_c + k_c\cdot\Delta I_c + \Delta H_h + \Delta H_r)$$

In order to determine the various contributions to equation (16) we have measured the magnetic field by mean of the hyperfine transition, $(F,m_F)=(2,1)$ to $(F,m_F)=(1,1)$. This is a σ polarized transition, same as the clock transition, and its dependence on H is,

$$f_{11} = f_{11,o} + b\cdot H \qquad b \approx 1.4 \text{ MHz/gs} \quad (17)$$

An example is illustrated in figure 3 where the magnetic field was measured for the two directions of the current in the heaters.

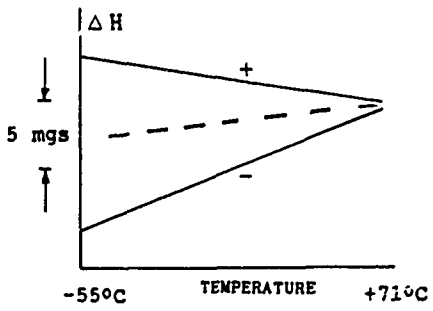


Figure 3: Magnetic field vs temperature for + and - directions of the heaters current derived from (2,1)->(1,1) transition.

From these measurement an others we have collected the data which is summarized in table 3. It is seen that major contributions arise from the residual magnetic field, presumably from magnetic materials, and from the heaters current. However, the later can be controlled as described below.

PARAMETER	CHANGE -55 TO 71°C	f/f
k_m≈ 70 gs/A	0.2 gs/A	8xE-11
I_m=6 mA	20 mA	1xE-10
H_z=20 mgs	4 mgs	2.8xE-10
H_h=0 to 10 mgs	0 to ±10mgs	0 to ±7xE-10

Table 3: Temperature sensitivities: magnetic effects

PARAMETER	PARAMETER CHANGE -55°C to +71°C (MEASURED)	SENSITIVITY (MEASURED)	Δf/f (CALCULATED
Lamp Temp	0.1°C	5xE-11 /°C	5xE-12
Cavity Temp	0.5°C	8xE-11 /°C	4xE-11
Microwave Pwr	0.5 dB	2xE-11 /dB	1xE-12
Phase Shift	5° to 20°	8xE-12 /°	4 to 16xE-11
Cross Talk	0 to 200 pA	5xE-13 /pA	0 to 1xE-10
Integ Offset	2 mV	4xE-13 /mV	8xE-13

Table 4: Temperature sensitivities : non-magnetic effects.

The major non-magnetic effects that contribute to the temperature sensitivity are assembled in table 4. Parameter changes over the temperature range and sensitivities are measured values while Δf/f is calculated. Modulation phase-shift is the sum of all contributions from the various elements in the multiplication chain, which is included in the servo-loop, and from the physics-package. In cases where a range of parameters is indicated the lower value can be obtain by a careful tuning.

A Compensation Scheme

Heaters current effect can be controlled, sign and magnitude, to a large extent. Also, lamp and cavity sensitivities can be controlled to some extent. This way, a compensation procedure can be defined so as to minimize the overall temperature sensitivity.
Doing so a very low temperature sensitivities with has been achieved some of the unites. An example is illustrated in figure 4, where the frequency change over the temperature range, -55°C to +71°C, is less then 1 part per 10^{12}.

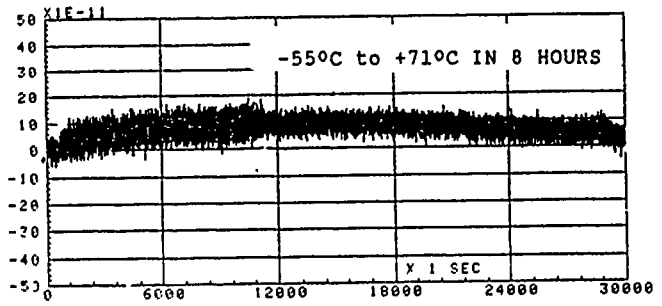


Figure 4: Temperature run for a selected unit.

Magnetic Sensitivity: The AC "C" Field Method

The ac "C" field scheme was described in [1]. The idea is to alternate the "C" field between + and - and to average out the result. The alternating field can take the form of a square wave or other periodic function such as sine wave. The alternating frequency should be smaller than the rubidium linewidth (≈1200 Hz) but higher than the servo-loop bandwidth (≈20 Hz)

It was shown in [1] that, using this method, the magnetic sensitivity is reduced by a factor of $2\alpha H_m/H_e$, where H_e is the external disturbance field which is being reduced by a shielding factor of α.
Figure 5, which was presented in [1], demonstrates the application of the method to a rubidium clock and proves its validity. As seen from the figure, the noise is not increased when applying the ac "C" field. However, this frequency noise, in figure 5, is related to a time constant of a few seconds. Observing the phase-noise (figure 6) we find

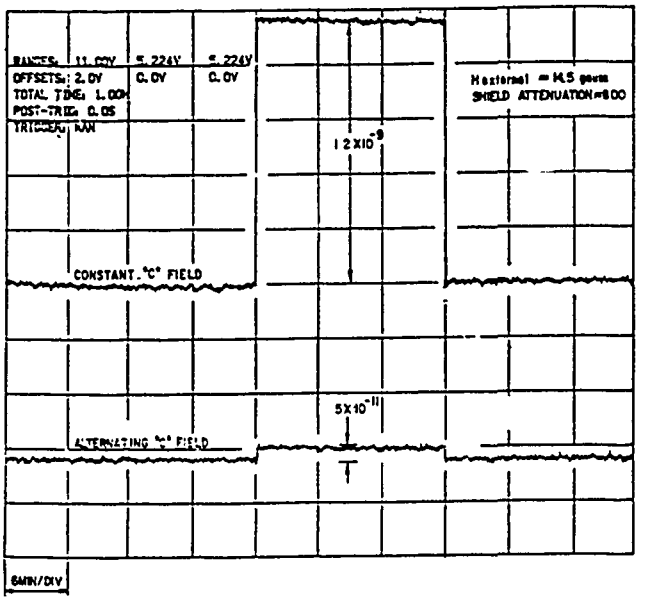


Figure 5: Frequency shift measured with an alternating "C" field (lower curve), compared with the shift observed in the conventional non-alternating "C" field (upper curve).

a considerable degradation of 30 to 40 dB. We observe peaks at frequencies that are related to the "C" field ac frequency (30Hz). The cause for these peaks is found to be transient signals that occur wnen the "C" field crosses zero (figure 7).
We have undertaken a thorough study of this phenomenon. Here we give only a preliminary short report whereas a more detailed description will be published elsewhere.

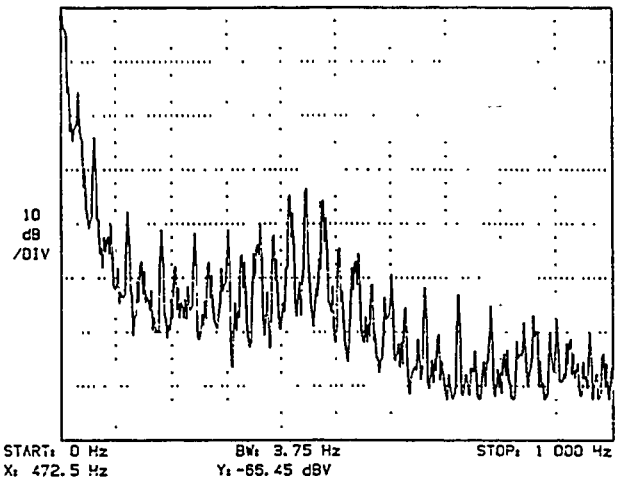


Figure 6: Degradation of the phase-noise due to the ac "C" field.

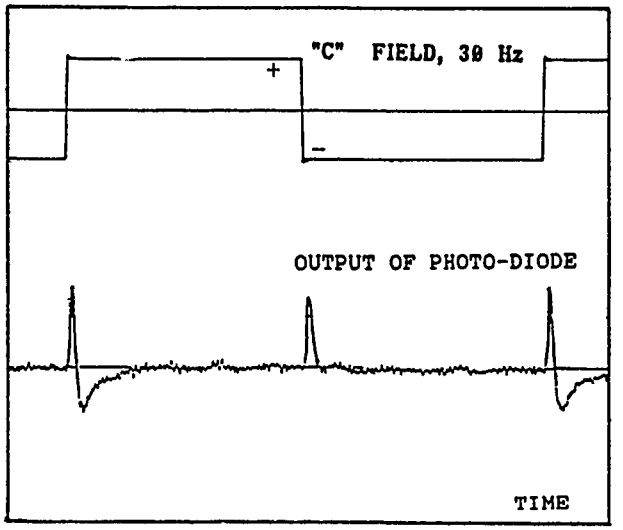


Figure 7: Output of photo-diode in time domain (lower curve) for a square wave "C" field modulation, 30 Hz, (upper curve).

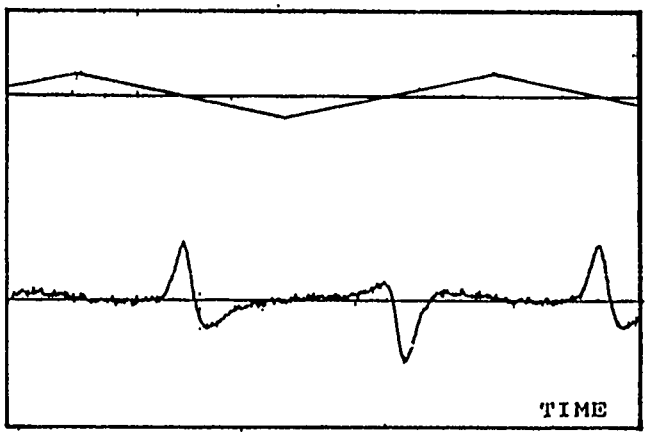


Figure 8: Output of photo-diode in time domain (lower curve) for a triangle wave "C" field modulation, slope=1.6 gs/sec (upper curve).

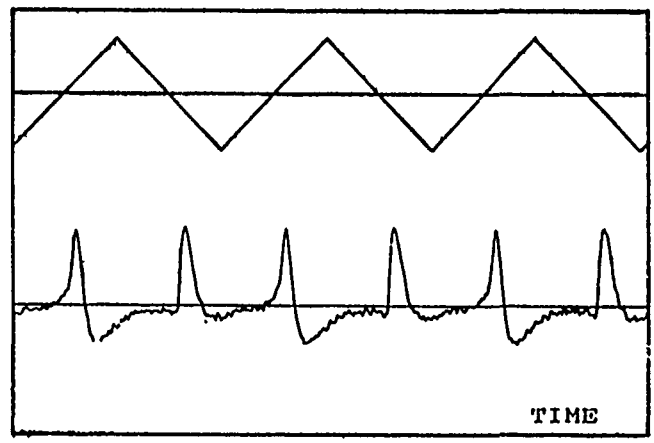


Figure 9: Output of photo-diode in time domain (lower curve) for a triangle wave "C" field modulation, slope=20 gs/sec (upper curve).

The response of the system was studied for various modulations of the "C" field, where the microwave frequency was roughly set to the clock transition at zero field. Figures 8 and 9 exhibit the signals observed for a triangle wave modulation. For a "slow" modulation, (slope=1.6 gs/sec, figure 8) we observe positive and negative signals while for a "fast" modulation (slope=20 gs/sec, figure 9) the dominant signals are positive.

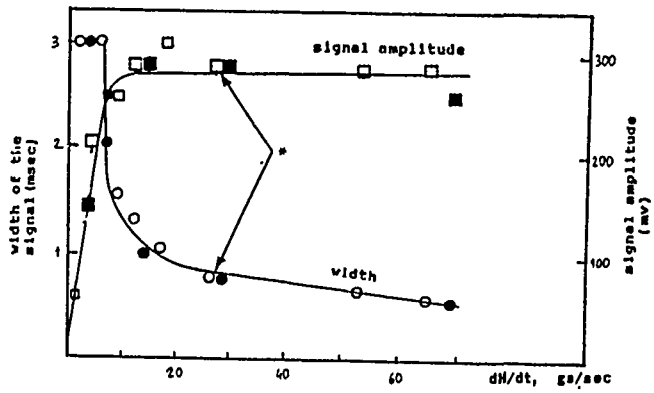


Figure 10: Dependence of the amplitude and width of transient signal on the slope of a triangle wave.
o □ - the frequency of triangle wave is 48 Hz the amplitude is changed.
● ■ - the amplitude of triangle wave is 0.35 gs the frequency is changed.
* - points where the amplitude and the frequency of triangle wave are changed but the slope remains constant.

Figure 10 brings accumulative data of signals observed with various triangle wave modulations of the "C" field. It exhibits the signal amplitude and signal "width" as a function of the slope dH/dt. One observes two regions: for dH/dt < 6 gs/sec signal amplitude is linearly dependent on the slope and the width is constant.
For dH/dt > 6 gs/sec signal amplitude is constant while the width is decreasing.

Discussion

The intensity of light transmitted by optically pumped rubidium vapor in magnetic field depends on the populations of hyperfine levels $(F,m_F)=(1,0)$ and $(F,m_F)=(2,0)$. These populations are determined by temperature, microwave power and magnetic field and optical pumping.
When the magnetic field changed rapidly the Rb spin system is not being able to follow the magnetic field variation, the populations of hyperfine levels change and a transient in the transmitted light occurs.

The Rb spin system follows the magnetic field variations by means of spin precession as well as by spin relaxation. The condition of adiabaticity is that the relative field variation is slower than the angular velocity of spin precession i.e,

$$(1/H)\cdot dH/dt << \gamma H \qquad (18)$$

where H is the magnetic field, and γ is the gyromagnetic ratio of the electron .
The condition of thermal equilibrium is that the relative field variation is slower than the rate of thermal relaxation, i.e.

$$(1/H)\cdot dH/dt << \tau_{rel}^{-1} \qquad (19)$$

When the magnetic field approaches zero the conditions (18) and (19) are violated. (18) is violated for $H \approx H_1 = [(1/\gamma)\cdot dH/dt]^{1/2}$.
(19) is violated for $H \approx dH/dt\cdot \tau_{rel}$.
If dH/dt > 6 gs/sec (empirical value obtained from Figure 10) then $H_1 < H_2$ and the transient is due to the nonadiabatic transitions between Zeeman sublevels. This corresponds to the behavior observed in figure 9. If dH/dt < 6 gs/sec then $H_2 < H_1$ and the transient is due to thermal inequilibrium in spin system.
This corresponds to figure 8. As mentioned before this is only a preliminary interpretation and a more thorough discussion will be published elsewhere.

Filtering Circuit

Our understanding of these phenomena is not yet complete. However, from a practical point of view, we have solved the phase-noise degradation by developing a simple filtering circuit.
Figure 11a shows the VCXO input with "C" field modulation, without filter, while figure 11b exhibits the same VCXO input, with the filter. (the phase-noise images to large extent the VCXO input). The noise level in figure 11b is the normal level observed without "C" field modulation. Therefor the ac "C" field method can be applied with no degradation in performance.

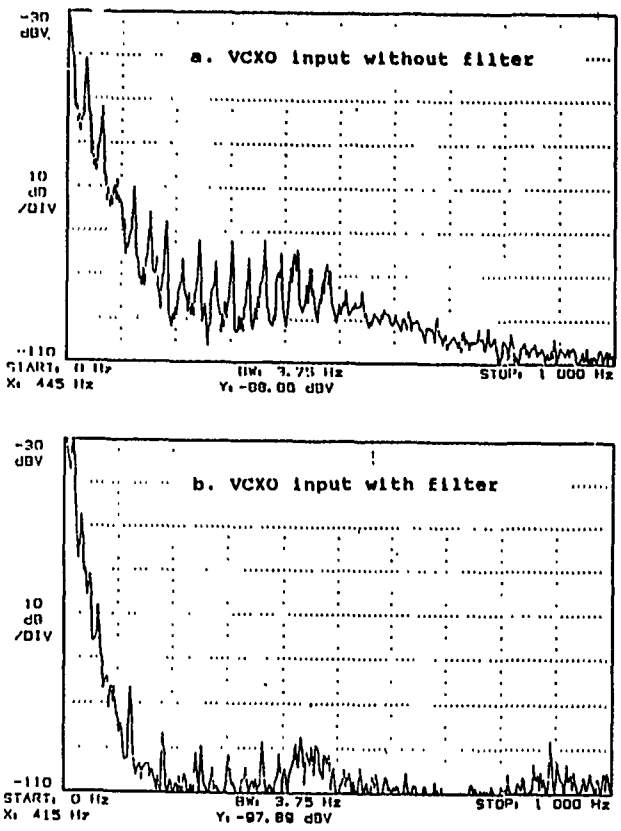


Figure 11: a. VCXO input without filter, "C" field modulation
b. VCXO input with filter, "C" field modulation.

Acknowledgment:

We acknowledge T. McClelland for helpful discussions.
This work was partially supported by the Israel Ministry of Commerce and Industry, office of chief scientist.

References

1. A. Stern, A. Hertz, Y. Zarfaty and A. Lepek, Proc. 42nd Annual Frequency Control Symposium, 1988, p. 519.
2. J. Vanier, M. Tetu and L.G. Bernier, IEEE Trans. Instrum. Meas. 1,28,188, 1979.
3. J.Vanier and L.G. Bernier, IEEE Trans Imstrum. Meas. 30,277, 1981.
4. J.C. Camparo and R.P. Frueholz, J.Appl. Phys. 59,301,1986.
5. C.H. Volk and R.P. Frueholz, J.Appl. Phys. 57, 980, 1985.
6. J.A. Barnes et al, IEEE Trans. Instum. Meas. 20, 105,1971.

* Pantent applications pendings.

43rd Annual Symposium on Frequency Control - 1989

NEW DESIGN FOR AN EFFICIENT OPTICALLY PUMPED CESIUM BEAM TUBE

V. Giordano, A. Hamel, P. Petit, G. Théobald, N. Dimarcq,
P. Cérez, C. Audoin.

Laboratoire de l'Horloge Atomique
Unité Propre de Recherche du CNRS associée à l'Université Paris-Sud
Bât. 221 - Université Paris-Sud
91405 ORSAY - France

Abstract

The use of optical methods to replace state selecting magnets in a cesium beam resonator is thought to improve the clock frequency stability and accuracy. Especially the Rabi pulling frequency shift disappears as a consequence of the microwave spectrum symmetry. It then brings the possibility of decreasing the amplitude of the static magnetic induction in the microwave interrogation region to values less than the conventional one (60 milliGauss).

Nevertheless, we have pointed out that coherent population trapping may occur in the two optical interaction regions of the resonator at low static magnetic field. In order to avoid the resulting decrease of the clock signal amplitude a magnetic induction of the order of 3×10^{-5} Tesla (300 mGauss) must be produced in the light-atoms interaction regions.

We have thus designed a new optically pumped cesium beam tube including three magnetic field regions which will be described. We present the preliminary results obtained from this machine.

I. Introduction

For a few years, reliable monomode semi-conductor lasers emitting at 852 nm have been available. These laser sources allow to apply optical pumping techniques to atomic beam preparation and clock signal detection in cesium beam frequency standards. A number of laboratories have started research work in this field. They built some experimental set-ups in which the magnet state selectors of the conventional cesium beam resonator were replaced by two light-atoms interaction regions [1,2,3]. This substitution procure a lot of advantages [4]. The loss of atoms by velocity dispersive deflection is greatly reduced and, moreover the beam cross-section is only limited by the microwave cavity hole diameter. It follows that the atoms flux in the detection region is increased by a factor of several hundreds. Furthermore, under appropriate optical pumping conditions, which will be discussed in this paper, one can transfer all atoms of one of the ground state hyperfine levels F = 3 of F = 4 to the other one. The beam preparation is thus realized without atoms loss whereas the magnetic state selection eliminates 50% of the atoms.

In a previous work [5] we have demonstrated the interest of the optically pumped cesium beam frequency standard. Our experimental resonator Cs II provides a clock signal with a signal to noise ratio equal tà 10,000 in a 1 Hz noise bandwidth. The short term frequency stability of the 10 MHz quartz resonator frequency locked on this atomic reference was : $\sigma(\tau) = 2.\ 10^{-12}\ \tau^{-1/2}$ for $1\ s < \tau < 500\ s$. This result was obtained using a single laser diode. The laser diode was tuned to the optical transition that connects the F = 3 ground state level to the $6^2P_{3/2}$ F' = 3 excited state level, and the linear light beam polarization was orthogonal to the static magnetic field direction. In this configuration the amplitude of the seven $\Delta F = \pm 1$; $\Delta m_F = 0$ lines of the microwave spectrum of the cesium atom varies symmetrically around the central $|F = 4, m_F = 0\rangle \rightarrow |F = 3, m_F = 0\rangle$ line. Indeed, the optical pumping performed with linearly polarized light populates equally the hyperfine Zeeman sublevels having opposite values of the magnetic quantum number m_F. The Rabi pulling frequency shift must then disappear. In this case, it is possible to decrease the magnitude of the magnetic induction in the microwave interaction region in order to limit both the Zeeman shift effect and the sensitivity to magnetic field fluctuations. Thus the accuracy and the long term frequency stability can be improved.

However, the transition that gives the highest clock signal, i.e. the $3 \rightarrow 3'$ transition excited by σ polarized light [6], is sensitive to the non-linear Hanle effect (NLHE). When the magnetic induction in the light-atoms interaction region becomes smaller that $3\ 10^{-5}$ Tesla the pumping efficiency and the fluorescence intensity rapidly decrease. This is detrimental to clock signal detection. G. Théobald et al made a detailed study of this effect [7]. We shall remind their principal results in Section II.

To take full benefit of optical pumping methods we have designed a new cesium beam resonator with three magnetic field regions called Cs III [1]. Our experimental set-up is described in Section III. Then preliminary results obtained with this set-up are given in Section IV. The theoretical predictions about NLHE are verified and we deduce the working conditions that optimize optical interaction processes. Additional features connected with homogeneity of the magnetic field in the microwave interaction region are presented.

II. Short Recall about non linear Hanle effect

Figure 1 shows the energy level diagram of the cesium atom. In our experiments we usually lize the laser frequency to one of the pumping : itions.

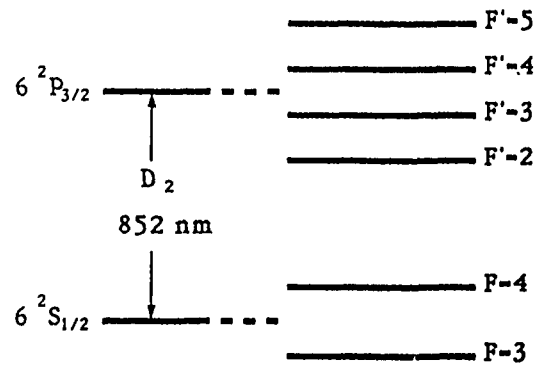


Figure 1. Energy level diagram corresponding to the D_2 line of cesium atom (not to scale)

Cesium atoms are brought to the $6^2P_{3/2}$ excited state. Then they decay to the ground state emitting fluorescence photons. After a number of cycles, depending on the choosen transition, all atoms populate one of the ground state hyperfine levels F = 3 or F = 4. When the optical pumping techniques are applied to the cesium beam frequency standard, we are mainly interes-

CH2690-6/89/0000-130 $1.00 C 1989 IEEE

ted in two physical quantities [7] which determine the clock signal amplitude :

- the fractional population difference Δn between the $m_F = 0$ ground state hyperfine sublevels obtained after optical preparation. It is defined as :

$$\Delta n = \frac{n(|F = 4, m_F = 0>) - n(|F = 3, m_F = 0>)}{n}$$

where n is the total population. Δn specifies the efficiency of the light to create a population difference.

- The yield L of fluorescence photons emitted by a cesium atom under the action of the light beam. These fluorescence photons are collected in the detection region to give the clock signal.

It has been shown [6] that the $3 \rightarrow 3'\sigma$ scheme is the most interesting one. Indeed it allows to maximize the product $\Delta n \times L$. However, when a σ-polarized laser light beam is used in a low magnetic field region, successive absorption and stimulated emission of photons create $\Delta m_F = \pm 2$ Zeeman coherences in the ground state of cesium atoms. Since these coherences are coupled by the laser to the population of the excited state, it then results a decrease in Δn and L when the magnetic induction goes to zero. Consequently, the amplitude of the clock signal decreases, whereas the continuous background due to the unpumped atoms increases. Figures 2a and 2b report predicted variations of Δn and L versus static magnetic induction B for the four cesium pumping transitions. The light-atom interaction time and the laser intensity were assumed equal to 12 µs and 2mW.cm^{-2} respectively.

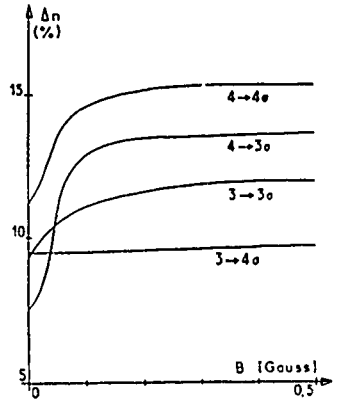


Figure 2a. Population difference Δn versus the amplitude of magnetic induction

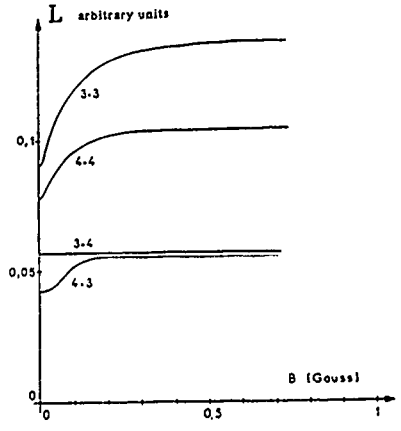


Figure 2b. Fluorescence yield L versus the amplitude of magnetic induction

One can see on these figures that, for the $3 \rightarrow 3'\sigma$ scheme, B must be set above 3. 10^{-5} Tesla (300 milli-Gauss) in order to optimize the light-atoms interaction process.

III. Cs III Design

The vacuum chamber of the Cs III resonator is an amagnetic stainless steel cylinder, its total length and its diameter are 800 mm and 170 mm respectively. A 200 l/s ion-pump maintains a residual pressure smaller than 10^{-5} Pa. A set of two pairs of optical windows allows to inject the pump and detection laser light beams perpendicularly to the atoms path. The separation between the two optical interaction regions is 450 mm.

In the following experiments a 10 mW laser diode is used. Its emission frequency is stabilized to the selected cesium optical transition using the fluorescence signal provided by a photodetector in the pumping region.

The collimator of the cesium oven consists of a stack of corrugated foils. Its emissive area is 2 x 4 mm^2. A set of graphite diaphragms limits the beam divergence. The atomic beam cross section is constant along the resonator axis, it is equal to 2 x 4 mm^2.

- Magnetic field

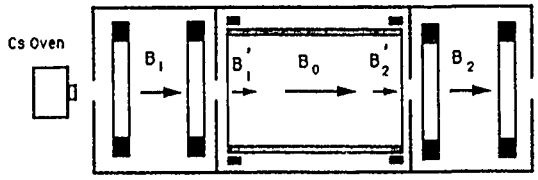


Figure 3. Magnetic field generation in Cs III

The Cs III static magnetic field design is entirely innovative as shown in figure 3. The atomic beam passes through three distinct regions. The inner cylindrical magnetic shield shows two septa which provides the screening between the three regions. It is surrounded by two other mumetal cylinders. The three cylinders are closed by end-caps. Each region includes its own internal field generation set-up. In order to easily obtain smooth magnetic field variation between two successive regions the magnetic field direction was choosen parallel to the atomic beam axis. In the first region, where atomic preparation is performed, we use a set of twin flat coils to generate the magnetic induction B_1. A solenoïd surrounding the Ramsey cavity allows to obtain a uniform magnetic induction B_0 between the two cavity arms. In the detection region, B_2 is generated with coils similar to those in the pumping region. Furthermore, to improve the uniformity of the magnetic induction in the microwave interaction region two compensation coils are added at the ends of the solenoïd. They can generate two weak magnetic induction B'_1 and B'_2 that cancel the influence of B_1 and B_2 on the field pattern in the RF interaction regions. The currents producing these five magnetic inductions can be adjusted independently. Figure 4 shows an example of static magnetic induction pattern recorded along the atomic path.

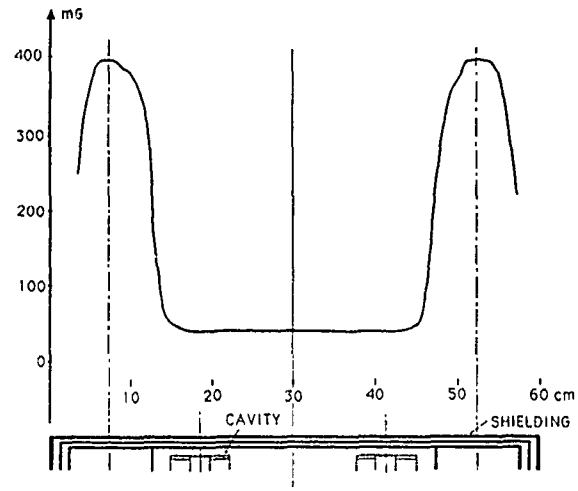


Figure 4. Magnetic induction profile in Cs III

- Microwave cavity

Our cavity shown in figure 5 is home-machined and is made of hard aluminium alloy 6061. We estimate that the electrical length of the two arms are identical within 0,01 mm. Due to the longitudinal magnetic field structure, the cavity is bent in the H-plane. It operates on the $TE_{0,1,13}$ mode. The inner drift distance is 21 cm. This yields a linewidth of about 500 Hz.

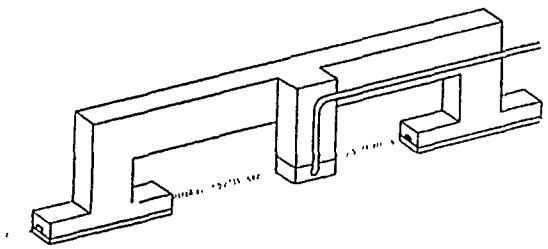


Figure 5. Cs III microwave cavity

In order to eliminate the usual hole in the magnetic shields and to avoid a field perturbation along the atomic path, we have developped a coaxial feeding of the cavity. Great care has been taken to accurately tune the cavity since once it is in the vacuum tube no further frequency adjustement can be made. The loaded Q of our cavity is about 800 which minimizes the cavity pulling effect.

- Optical collecting device

Recently a very efficient photon collecting device was implemented on Cs III. It includes two mirrors and a 100 mm^2 area photodetector. The photon collection efficiency has been measured by comparing the beam absorption and its fluorescence yield. This efficiency reaches 60 %.

IV. Experimental results

Experimental results obtained with Cs III are reported in the single laser diode configuration. The $3 \rightarrow 3' \sigma$ scheme is used to prepare the atomic beam and to detect the clock signal.

- Optimization of optical interaction processes

As previously mentioned in Section II, NLHE reduces the optical pumping efficiency. Due to its three independent magnetic field regions, Cs III is well suited to investigate this phenomenon. When the magnetic induction in the pumping region, i.e. B_1, is weak, a number of cesium atoms remain trapped in the F = 3 ground state hyperfine level. In the detection region these unpumped atoms may give a background on which the clock signal would be superposed. We recorded the variation of the fluorescence signal F in the detection region versus B_1. Let be F_o the fluorescence of the beam when no pumping occurs in the preparation region. Figure 6 shows the ratio F/F_o, which accounts directly for the percentage of unpumped atoms.

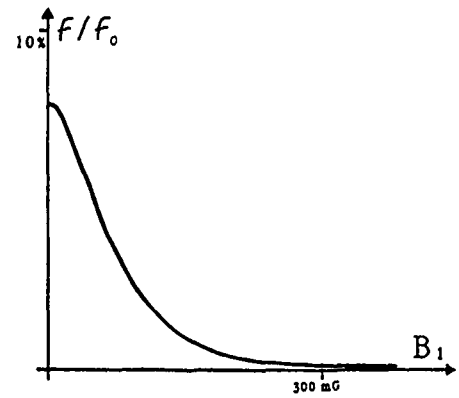


Figure 6. Percentage of unpumped atoms versus B_1

Its clearly appears that when B_1 is very small, F/F_o is about 8 % and that complete preparation is only achieved when B_1 is increased above 3×10^{-5} T (300 mG). In this experiment B_0 and B_2 are set respectively to 4×10^{-6} T and 5×10^{-5} T.

Furthermore the incomplete atomic beam preparation induces a dramatic enhancement of the fluctuations in the fluorescence signal F. Figure 7 shows the spectrum of the detection signal fluctuations for several B_1 values. This effect can be explained by noting that when the pumping is no longer complete, the fluorescence yield of the atomic beam depends on the laser frequency fluctuations.

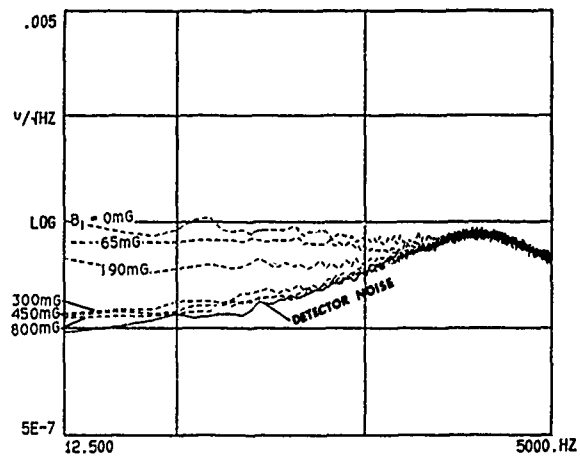


Figure 7. Spectrum of the detection signal for several B_1 values

It then results that B_1 must be set above 3.10^{-5} T according to the theoretical predictions. In the following optical pumping and detection are performed within a magnetic induction of 5.10^{-5} T (500 mG) to achieve full efficiency of these two processes.

- Central Ramsey pattern

Figure 8 shows the resonance pattern of the $|F = 4, m_F = 0\rangle \rightarrow |F = 3, m_F = 0\rangle$ transition obtained with Cs III. The central fringe is 500 Hz wide. The oven temperature was set about 115° C, which means that the atomic flux in the detection region was about 8×10^{10}. s^{-1}.

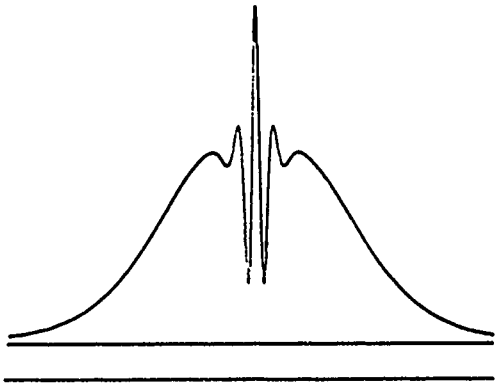


Figure 8. The Cs III clock signal

We note that the background signal, due to the stray light is very small. The amplitude signal to noise ratio is defined as the peak to valley photocurrent difference divided by the standard deviation of the peak current. It reaches 20,000 in a 1Hz noise bandwidth. This result should lead to a short term frequency stability better than $1 \times 10^{-12}\ \tau^{-1/2}$ [8].

- Microwave spectrum

Figure 9 shows the microwave spectrum of the seven σ resonances. The value of the magnetic induction B_o is 4×10^{-6} T. We note that the Ramsey patterns of all the microwave transitions are very well resolved. Furthermore they are centered on their Rabi pedestal. This reflects directly the magnetic field uniformity along the atomic beam path in the whole microwave interaction region.

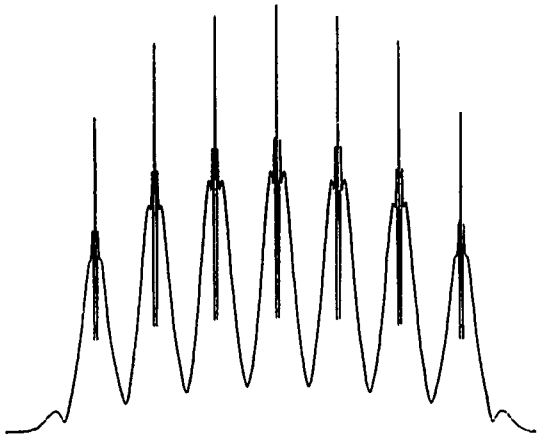


Figure 9. Microwave spectrum $B_o = 4 \times 10^{-6}$ T (40 mG)

- Effect of compensation coils

The influence of the two compensation inductions, B'_1 and B'_2, may be illustrated by monitoring the magnetic field dependent microwave transition patterns. It is known that when the average strength of the static magnetic induction is not the same in the microwave interaction regions and in the drift space between them, the Ramsey pattern of the field dependent microwave transition is no longer centered on the Rabi pedestal [9]. Figure .0a shows the resonance pattern of the $|F = 4, m_F = -3\rangle \rightarrow |F = 3, m_F = -3\rangle$ transition when $B'_1 = B'_2 = 0$ and when the average value of B_o in the drift region was set to 10^{-5} T. The frequency of the Ramsey pattern maxima is about at a frequency 650 Hz higher than the centre of the Rabi pedestal. It is due to the influence of B_1 and B_2 on the magnetic induction in the microwave interaction regions, which is thus a little more larger than 10^{-5} T. Owing to the compensation coils, B'_1 and B'_2 can be adjusted in order to centre the Ramsey pattern as it is shown in figure 10b.

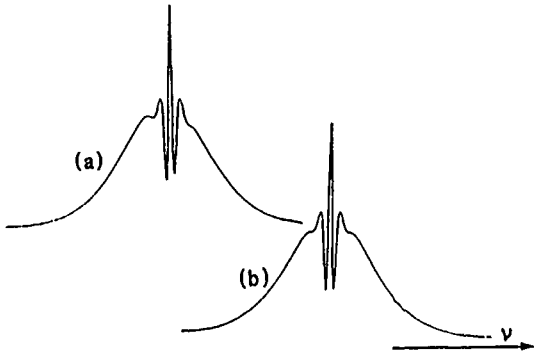


Figure 10. Effect of compensation inductions on the field dependent transitions

- Ramsey pattern at very low magnetic induction

The three separate fields configuration of Cs III allows to set the longitudinal magnetic induction in the RF interaction region to very small values, while keeping the magnitude of the induction at 5×10^{-5} T (500 mG) in the optical interaction regions. The two compensation coils are used to correct the field near each end of the solenoïd. As results, the mean value of the magnetic induction B_o can be closely adjusted to zero, while its deviation from the mean value can be made very small. In such conditions there is no Zeeman splitting any more. All the atoms of the beam undergo the hyperfine transition. As shown on figure 11, the Ramsey pattern obtained is about 6 times higher with a slightly increased width. The usual fringe is displayed for comparison purpose. This demonstrates that the magnetic field can be made very homogeneous in this configuration.

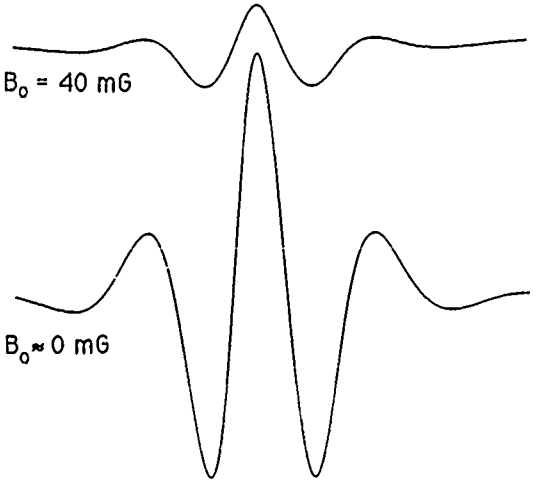


Figure 11. Ramsey pattern at 4×10^{-6} T (a) and at very low magnetic induction (b)

Conclusion

A new optically pumped cesium beam resonator which allows to optimize the three interaction processes has been designed.

It has been demonstrated that an optically pumped cesium beam resonator can be operated at low C-field value without any loss of optical pumping efficiency. This shows that the optical pumping technique is able to improve the long term frequency stability as well as the short term frequency stability.

Acknowledgements

We are grateful to S.T.E.I. for its financial support.

References

1 A. Hamel, P. Petit, G. Théobald, P. Cérez, C. Audoin, "Progress in an optically pumped cesium beam resonator operating with a longitudinal magnetic field", presented at the 3rd European Frequency and Time Forum, Besançon, France, March 1989

2 S. Oshima, Y. Nakadan and Y. Koga, "Development of an optically pumped Cs frequency standard at the NRLM", IEEE Trans. on Instrum and Meas., Vol. 37, n° 3, pp. 409-413, Sept. 1988

3 R.E. Drullinger, "Frequency standards based on optically pumped cesium" in Proceedings of the IEEE, Vol. 74, n° 1, pp. 140-142, Jan. 1986

4 G. Théobald, V. Giordano, M. de Labachelerie, A. Hamel, N. Dimarcq, P. Cérez and C. Audoin, "Research on the optically pumped cesium beam frequency standards", presented at the 42th Symposium on Frequency Control, Baltimore, USA, May 1988

5 V. Candelier, V. Giordano, A. Hamel, G. Théobald, P. Cérez and C. Audoin, "Frequency stability of an optically pumped cesium beam frequency standard", Applied Physics, to be published

6 G. Avila, V. Giordano, V. Candelier, E. de Clercq, G. Théobald and P. Cérez, "State selection in a cesium beam by laser diode optical pumping", Phys. Rev. A, Vol. 36, n° 8, pp. 3719-3728, Oct. 1987

7 G. Theobald, N. Dimarcq, V. Giordano and P. Cérez, "Ground state Zeeman coherences effects in an optically pumped cesium beam", Optics Communications, to be published

8 V. Giordano, A. Hamel, G. Théobald, P. Cérez, C. Audoin and V. Candelier, "Measurement of the signal-to-noise ratio in an optically pumped cesium-beam frequency standard", Metrologia, Vol. 25, pp. 17-20, 1988

9 J. Vanier and C. Audoin, The Quantum of Atomic Frequency Standards, Bristol : Adam Hilger, 1989, p. 804

43rd Annual Symposium on Frequency Control - 1989

LINEAR ION TRAP FOR SECOND ORDER DOPPLER SHIFT REDUCTION IN FREQUENCY STANDARD APPLICATIONS*

J. D. Prestage, G. R. Janik, G. J. Dick, T. K. Tucker and L. Maleki
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

We have designed and are presently testing a novel linear ion trap which permits storage of a large number of ions with reduced susceptibility to the second order Doppler effect caused by the RF confining fields. This new trap should store about 20 times the number of ions as a conventional RF trap with no corresponding increase in second order Doppler shift from the confining field. In addition the sensitivity of this shift to trapping parameters, i.e., RF voltage, RF frequency and trap size, is greatly reduced.

We have succeeded in trapping mercury ions and xenon ions in the presence of helium buffer gas. Trap times as long as $2 \cdot 10^3$ seconds have been measured.

Introduction

There has been much recent activity directed toward the development of trapped ion frequency standards, in part because ions confined in an electromagnetic trap are subjected to very small perturbations of their atomic energy levels. The inherent immunity to environmental changes that is afforded by suitably chosen ions suspended in DC or RF quadrupole traps has led to the development of frequency standards with very good long term stability[1]. Indeed, the trapped $^{199}Hg^+$ ion clock of ref[2] is the most stable clock yet developed for averaging times $\geq 10^6$ seconds. However, certain applications such as millisecond pulsar timing[3] and low frequency gravity wave detection across the solar system[4] require stabilities beyond that of present day standards.

While the basic performance of the ion frequency source depends fundamentally on the number of ions in the trap, the largest source of frequency offset stems from the motion of the atoms caused by the trapping fields via the second order Doppler or relativistic time dilation effect[5]. Moreover, instability in certain trapping parameters, e.g., trap field strength, temperature, and the actual number of trapped particles will influence the frequency shift and lead to frequency instabilities. Since this offset also depends strongly on the number of ions, a trade–off situation results, where fewer ions are trapped in order to reduce the (relatively) large frequency offset which would otherwise result.

We have designed and constructed a hybrid RF/DC linear ion trap which should allow an increase in the stored ion number with no corresponding increase in second order Doppler instabilities. The 20 times larger ion storage capacity should improve clock performance substantially. Alternatively, the Doppler shift from the trapping fields may be reduced by a factor of 10 below comparably loaded hyperbolic traps.

Second Order Doppler Shift for Ions in an RF Trap

Trapping forces in an RF ion trap are due to time–varying electric fields which increase in every direction from the trap's center. A single particle at rest in such a trap at its very center (where, in an ideal trap, these fields are zero) would have no velocity and thus no second order Doppler shift.

A very different condition holds for many particles in such a trap. In this case, electrostatic repulsion tends to keep the ions away from each other and from the center of the trap. As the number of ions increases, the size of the cloud also increases, pushing the ions into regions of larger and larger RF fields. The resultant velocity gives rise to a downward shift of atomic transition frequencies with increasing ion number.

Calculation of the second order Doppler shift requires a detailed knowledge of the ionic distribution density which

*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.

CH2690-6/89/0000-135 $1.00 C 1989 IEEE

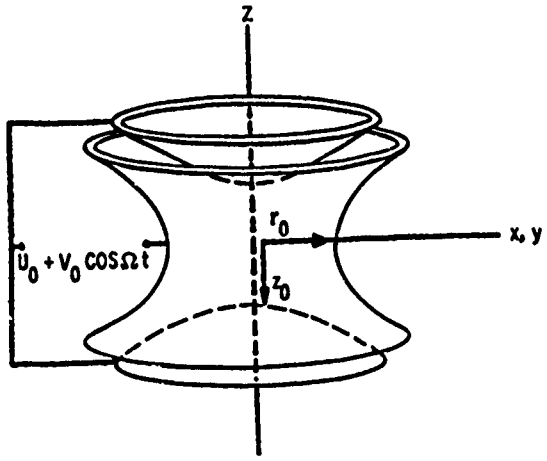


Figure 1: A conventional hyperbolic RF ion trap. A node of the RF and DC fields is produced at the origin of the coordinate system shown.

results from the balance between trapping and (repulsive) Coulomb forces. A method has been developed in which an average over one cycle of the RF field reduces its effect to that of a pseudopotential acting on the charge of the particle[6]. The effect is subsequently further reduced to that of a pseudocharge distribution which produces the equivalent effective potential. Ionic distribution density is then calculated by considering the response of charges to this resultant "background" pseudocharge.

This method has been previously applied[7] to the trap shown in Fig. 1. Here, calculation shows a spatially uniform pseudocharge giving rise to a spherical ion cloud, also with uniform density. The resulting average frequency shift can be expressed in terms of the total number of trapped ions, together with a trap strength parameter ω, ion mass m and charge q.

In the following sections we perform a similar calculation for a cylindrical geometry, a case not previously examined. The cloud forms a cylinder of uniform density[8], in a manner analogous to that of the spherical trap. Comparison between the consequences of the two geometries shows a very different story. While physically similar in overall size, the linear trap can hold many more ions than the spherical one with no increase in the second order Doppler shift, or conversely, the shift can be greatly reduced. Furthermore, its dependence on trap parameters is qualitatively different, allowing miniaturization of the transverse trap dimensions without penalty in performance.

Calculation for a Spherical Trap

Figure 1 shows a conventional RF ion trap along with the applied voltages. Trapping forces are generated by the driven motion of the ions (at frequency Ω) in an inhomogeneous RF electric field created by hyperbolic trap electrodes[6]. The motion in each of three directions for a single ion in an RF trap is characterized by two frequencies, the fast driving frequency Ω and a slower secular frequency ω which is related to the harmonic force binding the particle to the trap center. An exact solution to the equations of motion shows that frequencies $k \cdot \Omega \pm \omega$, k = 0, 1, 2 ... are also present. However, in the limit $\omega/\Omega \ll 1$ (which is the primary condition for stability of the ion orbits) the ω and $\Omega \pm \omega$ frequencies dominate and the kinetic energy (K.E.) of a particle, averaged over one cycle of Ω, separates into the kinetic energy of the secular motion and the kinetic energy of the driven motion. The average kinetic energy is transferred from the secular to the driven motion and back while the sum remains constant just as a harmonic oscillator transfers energy from kinetic to potential and back.

We consider two cases. A hot ion cloud, or one containing a very few ions where interactions between ions are negligible, shows a second order Doppler shift given by:

$$\left(\frac{\Delta f}{f}\right)_{hot} = -\frac{1}{2}\frac{< v^2 >}{c^2} = -\frac{< \text{total K.E.} >}{mc^2} \quad (1)$$

$$= -\frac{< \text{secular K.E.} + \text{driven K.E.} >}{mc^2} \quad (2)$$

$$= -2\frac{< \text{secular K.E.} >}{mc^2} \quad (3)$$

$$= -\frac{3k_BT}{mc^2}, \quad (4)$$

where m is the ionic mass, T the temperature, and <> indicates a time average over one cycle of Ω. We have also averaged over one cycle of ω to equate the secular and driven K.E. This is analogous to a simple harmonic oscillator where the average K.E. is equal to the average potential energy. The consequence is a frequency shift that is twice as large as that due to thermal motion alone.

Of greater interest is the case where many ions are contained in a trap and interactions between ions dominate. In this cold cloud model[7] displacements of individual ions from the trap center are primarily due to electrostatic repulsion between the ions, and random thermal velocities associated with temperature can be assumed to be small compared to driven motion due to the trap fields.

The electric potential inside the trap of Fig. 1 is

$$\phi(\rho, z) = \frac{U_o + V_o \cos(\Omega t)}{\epsilon^2}\left(\rho^2 - 2z^2\right), \quad (5)$$

where $\epsilon^2 = r_o^2 + 2z_o^2$ describes the trap size, and U_o and V_o represent the amplitudes of DC and AC trap voltages, respectively.

The trapping force generated by the RF field alone can be described by an electric pseudopotential[6]:

$$\Psi(\rho,z) \equiv \frac{q(\mathrm{E}_o(\rho,z))^2}{4m\Omega^2}, \tag{6}$$

where q is the ionic charge and E_o is the peak local RF field. This becomes

$$\Psi(\rho,z) = \frac{qV_o^2}{m\Omega^2\epsilon^4}\left(\rho^2 + 4z^2\right) \tag{7}$$

for the effect of the RF part of Eq. (5). Adding the DC potential from Eq. (5) gives the total potential energy for an ion in the trap of Fig. 1:

$$\Phi(\rho,z) = \frac{1}{2}\left(m\omega_\rho{}^2\rho^2 + m\omega_z{}^2 z^2\right), \tag{8}$$

where

$$\omega_\rho{}^2 = \frac{2q^2V_o^2}{m^2\Omega^2\epsilon^4} + \frac{2qU_o}{m\epsilon^2} \tag{9}$$

and

$$\omega_z{}^2 = \frac{8q^2V_o^2}{m^2\Omega^2\epsilon^4} - \frac{4qU_o}{m\epsilon^2} \tag{10}$$

describe secular frequencies for radial and longitudinal ion motion.

The pseudopotential can be further analyzed in terms of an effective pseudocharge by applying Poisson's equation to equation (7) or (8). The result of this calculation is a uniform background charge density throughout the trap region which is given by

$$Q_b = -\frac{\epsilon_o m(2\omega_\rho{}^2 + \omega_z{}^2)}{q}. \tag{11}$$

An easy solution for the charge configuration can be found if we assume that the DC and RF voltages are adjusted to make the trapping forces spherical, i.e., $\omega_\rho = \omega_z \equiv \omega$. In this case the ion cloud is also spherical and trapped positive ions exactly neutralize the negative background of charge, matching its density out to a radius where the supply of ions is depleted. Ion density is given by

$$n_o = \frac{3\epsilon_o m\omega^2}{q^2}, \tag{12}$$

and the total number of ions by

$$N = n_o\frac{4\pi}{3}R_{sph}{}^3, \tag{13}$$

where R_{sph} is the radius of the sphere of trapped ions.

The oscillating electric field which generates the trapping force grows linearly with distance from the trap center. The corresponding amplitude of any ion's driven oscillation is proportional to the strength of the driving field, i.e., also increasing linearly with the distance from the trap center. The average square velocity of the driven motion for an ion at position (ρ,z) is

$$< v^2 >= \frac{1}{2}\omega^2\left(\rho^2 + 4z^2\right). \tag{14}$$

For a given trapping strength, reflected in force constant ω^2, the density is fixed by Eq. (12) while the radius of the spherical cloud is determined once the ion number N has been specified. The second order Doppler shift due to the micro-motion is found by taking a spatial average of Eq. (1) over the spherical ion cloud. Using Eq. (14) for the spatial dependence of the micromotion:

$$\left(\frac{\Delta f}{f}\right)_{sph} = -\frac{1}{2}\frac{\overline{< v^2 >}}{c^2} \tag{15}$$

$$= -\frac{3}{10}\frac{\omega^2 R_{sph}{}^2}{c^2} \tag{16}$$

$$= -\frac{3}{10c^2}\left(\frac{N\omega q^2}{4\pi\epsilon_o m}\right)^{2/3}. \tag{17}$$

For typical operating conditions[7], $N = 2{\cdot}10^6$ and $\omega = (2\pi){\cdot}50\mathrm{kHz}$, $\Delta f/f = 2{\cdot}10^{-12}$. This second order Doppler shift is about 10 times larger than the shift for free ^{199}Hg ions at room temperature, $\Delta f/f = 3k_BT/2mc^2 = 2\cdot 10^{-13}$.

If the temperature is not too high, its effect on the ion cloud is to broaden the sharp edge at its outside radius. In this case the plasma density falls off in a distance characterized by the Debye length[8]:

$$\lambda_D = \sqrt{\frac{k_BT\epsilon_o}{n_oq^2}}. \tag{18}$$

The cold cloud model should be useful provided the ion cloud size is large compared to the Debye length. This ratio is given by

$$\left(\frac{\lambda_D}{R_{sph}}\right)^2 = \frac{1}{30}\frac{\left(\frac{\Delta f}{f}\right)_{hot}}{\left(\frac{\Delta f}{f}\right)_{sph}}. \tag{19}$$

This indicates a relatively small fractional Debye length throughout the regime of interest. For the typical conditions indicated above, the Debye length is about 1/5 mm in comparison to a spherical cloud diameter of 2.5 mm.

Calculation for a Linear Trap

For increased signal to noise in the measured atomic resonance used in frequency standard applications, it is desirable to have as many trapped ions as possible. However, as we have just seen, larger ion clouds have larger second order Doppler shifts. This frequency offset must

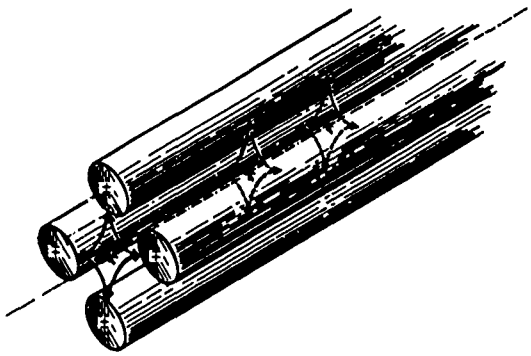


Figure 2: The RF electrodes for a linear ion trap. Ions are trapped around the line of nodes of the RF field with reduced susceptibility to second order Doppler frequency shift.

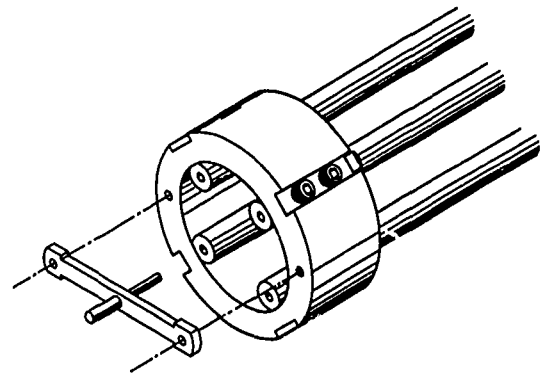


Figure 3: The details of the DC endcap needle electrodes used to prevent ions from escaping along the longitudinal axis.

be stabilized to a high degree in order to prevent degradation of long term performance.

To reduce this susceptibility to 2nd order Doppler shift we now propose a hybrid RF/DC ion trap which replaces the single field node of the hyperbolic trap with a line of nodes. The RF electrode structure producing this line of nodes of the RF field is shown in Figure 2. Ions are trapped in the radial direction by the same type of RF trapping forces used in the previously discussed hyperbolic trap and we follow a similar analysis in terms of an equivalent pseudopotential and background pseudocharge.

Ions are prevented from escaping along the axis of the trap by DC biased "endcap" needle electrodes mounted on each end as shown in Figure 3. These electrodes approximate the electrostatic effect of the missing parts of an infinitely long ion cloud. Their diameter is the same as the ion cloud to be trapped and is small compared to the trap diameter so that the RF trapping field is perturbed only slightly. Because these endcaps reach well inside the RF electrodes any end effect of the RF fields on the ion cloud should be small. Unlike conventional RF traps this linear trap will hold positive or negative ions but not both simultaneously.

Near the central axis of the trap we assume a quadrupolar RF electric potential:

$$\phi = \frac{V_o\,(x^2 - y^2)\cos(\Omega t)}{2R^2}, \tag{20}$$

from which, as in the previous section, we derive a corresponding pseudopotential:

$$\Psi = \frac{qV_o^{\,2}}{4m\Omega^2R^4}\left(x^2 + y^2\right), \tag{21}$$

for a total ionic potential given by

$$\Phi = q\Psi = m\frac{\omega^2}{2}\rho^2; \tag{22}$$

where, for the cylindrical electrodes of Fig. 2 R is an approximate distance from the trap center to an electrode's surface, and

$$\omega^2 = \frac{q^2V_o^{\,2}}{2m^2\Omega^2R^4}. \tag{23}$$

Here ω is the characteristic frequency for transverse or radial motion in the trap. Longitudinal motion is described in terms of motion at thermal velocities between the trap ends.

Application of Poisson's equation shows Eq. (21) to be equivalent to a uniform background pseudocharge with density:

$$Q_b = -\frac{2\epsilon_o m\omega^2}{q}. \tag{24}$$

Solving for the charge configuration for an infinitely long trap follows a nearly identical process to that of the preceding section since, from Gauss's law, cylindrical or spherical surfaces of charge induce no fields in their interior. Thus we find a uniform cylinder of ions just cancelling the background pseudocharge out to a radius R_c with density:

$$n_o = \frac{2\epsilon_o m\omega^2}{q^2}, \tag{25}$$

with ion number per unit length of

$$N/L = n_o\pi R_c^{\,2}. \tag{26}$$

The motion induced by the RF trapping field is purely transverse and is given by

$$< v^2 >= \omega^2 \rho^2. \quad (27)$$

As before we average this quantity over the ion cloud to find the second order Doppler shift:

$$\left(\frac{\Delta f}{f}\right)_{lin} = -\frac{1}{2}\frac{\overline{< v^2 >}}{c^2} = -\frac{\omega^2 R_c^2}{4c^2}. \quad (28)$$

We assume for simplicity a cylindrical ion cloud of radius R_c and length L. Equation (28) can be written in terms of total ion number N, and trap length L, as

$$\left(\frac{\Delta f}{f}\right)_{lin} = -\left(\frac{q^2}{8\pi\epsilon_o mc^2}\right)\frac{N}{L}. \quad (29)$$

In contrast to the spherical case, this expression contains no dependence on trap parameters except for the linear ion density N/L. This is also true for the relative Debye length:

$$\left(\frac{\lambda_D}{R_c}\right)^2 = \frac{1}{24}\frac{\left(\frac{\Delta f}{f}\right)_{hot}}{\left(\frac{\Delta f}{f}\right)_{lin}}, \quad (30)$$

which must be small to insure the validity of our "cold cloud" model.

From this it is seen that the transverse dimension R of the trap may be reduced without penalty of performance, providing that operational parameters are appropriately scaled. This requires ω and Ω to vary as R^{-1}, and the applied voltage V_o to be held constant.

Comparison

We can compare the second order Doppler shift for the two traps assuming both hold the same number of ions by

$$\left(\frac{\Delta f}{f}\right)_{lin} = \frac{5}{3}\frac{R_{sph}}{L}\left(\frac{\Delta f}{f}\right)_{sph}. \quad (31)$$

As more ions are added to the linear trap their average second order Doppler shift will increase. It will equal that of the spherical ion cloud in the hyperbolic trap when

$$N_{lin} = \frac{3}{5}\frac{L}{R_{sph}}N_{sph}. \quad (32)$$

A linear trap can thus store $(3/5)(L/R_{sph})$ times the ion number as a conventional RF trap with no increase in average second order Doppler shift. For the trap we have designed, L is 75 mm. Taking $R_{sph} = 2.5$ mm for $2 \cdot 10^6$ $^{199}Hg^+$ ions in a spherical trap with similar over-all size, we find that the linear trap capacity is about 18 times larger. Furthermore, it seems likely that the transverse dimension of the linear trap can be reduced to a value 100 or more times smaller than its length while maintaining constant ion number and second order Doppler shift. This corresponds to a reduction in volume of 10,000 times.

Sensitivity to Trapping Parameters

An important consideration in the design of any frequency standard is the sensitivity of its output frequency to changes in its operating parameters. In the present case we consider how variations in the trapping voltages, ion number, trap size and trapping field frequency will influence the output frequency. For the large ion clouds discussed in this paper we will use eq. (17) for the hyperbolic trap (spherical cloud) and eq. (29) for the linear trap (cylindrical cloud) for this analysis.

For variations in ion number in the spherical cloud there is a corresponding variation

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{sph}}{\left(\frac{\Delta f}{f}\right)_{sph}} = -\frac{2}{3}\frac{\delta N}{N} \quad (33)$$

in second order Doppler shift and a similar variation for the linear trap

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = -\frac{\delta N}{N}. \quad (34)$$

For example, if

$$\left(\frac{\Delta f}{f}\right)_{sph} = 2 \cdot 10^{-12}$$

and we want total frequency instability to be 10^{-15} or better then

$$\frac{2}{3}\frac{\delta N}{N}\left(\frac{\Delta f}{f}\right)_{sph} < 10^{-15}.$$

That is, we must maintain $\delta N/N < 7.5 \cdot 10^{-4}$ over the integration time required to achieve the 10^{-15} performance. Similarly,

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{sph}}{\left(\frac{\Delta f}{f}\right)_{sph}} = -\frac{2}{3}\frac{\delta\omega}{\omega}. \quad (35)$$

For the linear trap we see from eq(29) that only the linear charge density determines the second order Doppler shift. For an infinitely long trap (or a closed circuit "racetrack" configuration with a large radius) this shift from the trapping fields is independent of ω. That is, the second order Doppler shift is independent of the transverse trapping fields and thus independent of the applied RF voltage and trap size. As the RF trapping force changes, the ion density and ion cloud radius change together so that the second order Doppler shift from the trapping field is constant (see eqs. (28) and (29)).

For a finite length trap we now estimate the residual dependence on changes in the trapping fields. The charged ion column is contained inside the conducting 4

bar linear trap which will shield axial electric fields to a range R_t which is approximately equal to the transverse radius of the trap.Thus the electric field at each end of the plasma column is not generated from the full length L of the ion cloud but only from those charges within the shielding length R_t of each end. We approximate this electric field to be that of a cylinder of charge of density n_o, radius R_c and length R_t, where $R_c \ll R_t$. In this approximation the field at the surface

$$E = 2\pi n_o R_c$$

is equal to the field produced by the end electrode ,assumed to vary on axis as

$$E_e e^{-(H-L)/2R_t}$$

where H is the end electrode separation and L is the ion cloud length. $(H-L)/2$ is the distance between one end electrode and the near en.' of the ion cloud. Equating these gives

$$2\pi n_o R_c = E_e e^{-(H-L)/2R_t}. \tag{36}$$

From eq. (29)

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = -\frac{\delta L}{L}. \tag{37}$$

We consider two cases. The first is a change in the RF confining force, $\delta\omega/\omega$. This is done without changing the end electric field, E_e, and without changing the ion number $N = \pi {R_c}^2 L n_o$. Holding constant the ion number leads to

$$2\cdot\frac{\delta\omega}{\omega} + 2\cdot\frac{\delta R_c}{R_c} + \frac{\delta L}{L} = 0. \tag{38}$$

where we have used

$$\frac{\delta n_o}{n_o} = \frac{2\delta\omega}{\omega}.$$

From equation (36) we find

$$\frac{\delta R_c}{R_c} + 2\cdot\frac{\delta\omega}{\omega} = \frac{\delta L}{L}\frac{L}{2R_t}. \tag{39}$$

Using equation (38) to eliminate $\delta R_c/R_c$ we find

$$\frac{\delta L}{L} = 2\cdot\frac{R_t}{L}\frac{\delta\omega}{\omega} \tag{40}$$

where we have neglected terms of higher order than R_t/L and R_c/R_t.

Considering now variations in the endcap voltage δV_e or electric field $\delta E_e/E_e = \delta V_e/V_e$. With constant ion number and $\delta\omega/\omega = 0$ we find

$$\frac{\delta L}{L} = -2\cdot\frac{R_t}{L}\frac{\delta V_e}{V_e}. \tag{41}$$

The linear trap with finite length has some interaction between second order Doppler shift and trapping fields but that interaction is reduced by R_t/L from that of the hyperbolic trap.

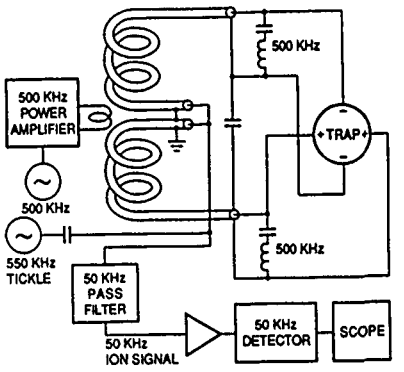


Figure 4: RF trapping and ion detection setup.

Linear Ion Trap Operation

The linear ion trap we have built consists of four molybdenum rods equally spaced on an approximately 10 mm radius. Ions are prevented from escaping along the axis by means of OFHC copper pins located at each end of the trap as shown in Fig. 3. The 75 mm separation of these pins defines the trap length.

The trap structure is housed inside a 3.375" ultra high vacuum cube which is evacuated to about 10^{-9} torr by a turbo-molecular pump with 40 l/sec pumping speed. Isotopic ^{199}Hg vapor is released from HgO powder when heated to about 200°C. Electrons from a LaB_6 filament are pulsed along the axis of the trap ionizing some of the ^{199}Hg atoms inside the trap. A background gas of helium (10^{-5} torr is introduced into the vacuum through a temperature controlled quartz leak. This cooling gas substantially increases the number of trapped ions. Alternatively, xenon atoms are leaked into the vacuum system through a high vacuum leak valve.

We have trapped Hg and Xe ions in this linear trap for confinement times up to 40 minutes. The ions in the trap are detected by exciting their motion transverse to the trap axis by injecting 550 kHz on one trap rod and measuring the image currents induced at 50kHz on the same trap rod. The RF system for trapping and detecting the ions is shown in Fig. 4. An ion signal derived from the image currents is shown in Fig. 5.

The input optical system which performs state selection and also determines which hyperfine state the ions are in has been modified from the previous system[9]. The present system illuminates about 1/3 of the 75 mm long cylindrical ion cloud. An ion's room temperature thermal motion along the axis of the trap will give an average round trip time of 1.4 msec, a value which is much smaller than our optical pumping and interrogation times. Thus, during the time of the optical pulse all

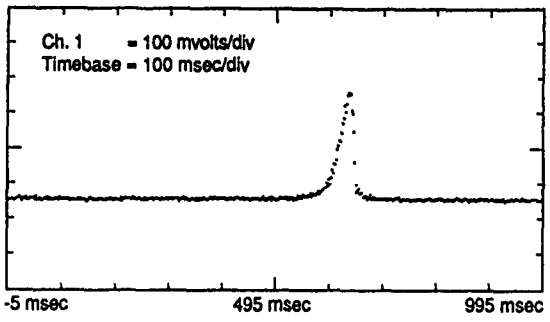


Figure 5: Ion signal obtained by ramping the trapping field amplitude to bring the ion secular frequency ω through the 50 kHz pass band.

ions will be illuminated, and pumping and interrogation completed. The only change is that a somewhat longer optical pulse is required.

In order to operate within the Lambe-Dicke regime[10] the 40.5 GHz microwave resonance radiation will be propagated perpendicular to the line of ions. The ions should then all experience phase variations of this radiation which are less than π so that the 1st order Doppler absorption in sidebands induced by an ions motion will not degrade the 40.5 GHz fundamental.

The optical axis of the fluorescence collection system is perpendicular to the axis of the input optical system as in the previous system. There is one difference, however. In the hyperbolic trap the collection has in its field of view the ion cloud and the semi-transparent mesh of both endcap trap electrodes. This mesh can scatter stray light into the collection system which will degrade the signal to noise ratio in the clock resonance. This linear trap has no trap electrodes, mesh or otherwise, in its field of view and, consequently, should have less detected stray light, allowing further performance improvement over the spherical trap.

Conclusions

Trapped ion frequency standards eliminate containing walls and their associated perturbations of the atomic transition frequencies by using electromagnetic fields alone to confine the particles. For any given trap, however, there exists a trade-off between the number of ions in the trap and a frequency shift due to second-order Doppler effects. This trade-off directly affects performance of the standard since the frequency shift is typically very much larger than the ultimate stability required and since the statistical limit to performance is directly related to ion number. We have calculated this performance trade-off for an RF trap with cylindrical geometry, a case not previously considered for a trapped ion frequency source.

By replacing the single node in the RF trapping field for a spherical trap by a line of nodes, a cylindrical trap effectively increases effective volume without increasing overall size. Furthermore, this performance is found to be independent of its transverse dimensions, as long as the driving frequency is scaled appropriately, with the driving voltage unchanged. More specifically, for the same frequency shift, we find that a linear trap with length L can hold as many ions as a spherical trap with diameter 6L/5. In addition to the practical advantage of greatly reduced overall volume, a fundamental advantage is also allowed since operation within the Lambe-Dicke regime places a limit on the size of the ion cloud, a requirement which may be met for a cylindrical trap by irradiating the microwave atomic transition in a direction perpendicular to the trap's longitudinal axis.

We have designed a trapped ion frequency source in which a cylindrical trap is implemented with a combination of RF and DC electric fields. With similar overall size and improved optical performance, this trap has 15 to 20 times the ion storage volume as conventional RF traps with no increase in second order Doppler shift.

Acknowledgements

We wish to thank Dave Seidel for assisting in the design of the linear trap described here.

References

[1] D. W. Allan, Proc. 19th Ann. PTTI App. Plan. Meeting, Ed. R. L. Sydnor, 375 (1987).

[2] L. S. Cutler, R. P. Giffard, P. J. Wheeler, and G. M. R. Winkler, Proc. 41st Ann. Symp. Freq. Control, IEEE Cat. No. 87CH2427-3, 12 (1987).

[3] L. A. Rawley, J. H. Taylor, M. M. Davis, and D. W. Allan, Science 238, 761 (1987).

[4] J. W. Armstrong, F. B. Estabrook, and H. D. Wahlquist, Ap. J. 318, 536 (1987).

[5] RF and Penning traps are both subject to second order Doppler problems related to the trapping fields. In this paper we confine our attention specifically to the case of RF traps. We simply note that, for the same number of stored ions, magnetron rotation of

the ion cloud in a Penning trap leads to a second order Doppler shift of comparable magnitude to that for an RF trap. A Penning trap based clock is described in: J. J. Bollinger, J. D. Prestage, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. **54**, 1000 (1985).

[6] H. G. Dehmelt, Adv. At. Mol. Phys. **3**, 53 (1967).

[7] L. S. Cutler, R. P. Giffard, and M. D. McGuire, Appl. Phys. B **36**, 137 (1985).

[8] S. A. Prasad, and T. M. O'Neil, Phys. Fluids **22** (2), 278 (1979).

[9] J. D. Prestage, G. J. Dick, and L. Maleki, Proc. 19th Ann. PTTI App. Plan. Meeting, Ed. R. L. Sydnor, 285 (1987).

[10] R. H. Dicke, Phys. Rev. **89**, 472 (1953).

43rd Annual Symposium on Frequency Control - 1989

PROGRESS AT NIST TOWARD ABSOLUTE FREQUENCY STANDARDS USING STORED IONS*

D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano,
D. J. Heinzen, S. L. Gilbert, C. H. Manney, and C. S. Weimer
Time and Frequency Division
National Institute of Standards and Technology
Boulder, Colorado 80303

Abstract

Experiments at NIST, whose goal is to realize frequency standards of high accuracy using stored ions, are briefly summarized. In one experiment, an rf oscillator is locked to a nuclear spin-flip hyperfine transition (frequency $\simeq 3.03 \times 10^8$ Hz) in $^9Be^+$ ions which are stored in a Penning trap and sympathetically laser-cooled. Stability is better than $3 \times 10^{-12}\tau^{-\frac{1}{2}}$ and uncertainty in Doppler shifts is estimated to be less than 5×10^{-15}. In a second experiment, a stable laser is used to probe an electric quadrupole transition (frequency $\simeq$ 1.07×10^{15} Hz) in a single laser-cooled $^{199}Hg^+$ ion stored in a Paul trap. The measured Q value of this transition is approximately 10^{13}. Future possible experiments are also discussed.

Introduction

One goal of the ion storage group at NIST has been to realize a stored ion frequency standard whose absolute inaccuracy is less than 1 part in 10^{14}. Primary design considerations in these experiments have been influenced by the assumptions that (1) for many ions in the trap, the uncertainty in the measurement of the second-order Doppler shift will be the largest contribution to inaccuracy and (2) the magnitude of the second-order Doppler shift decreases as the number of ions in the trap decreases.

These assumptions are supported both by experiments and theoretical analysis [1-5]. For a given number of ions N in the trap the second-order Doppler shift is minimized when the secular motion in the Paul trap [1-5] or the cyclotron and axial motion in the Penning trap [1-3] are reduced by some means to negligible values (for example, by the use of buffer gas collisions [4] or laser cooling [6]). For the Paul trap, the second-order Doppler shift is then dominated by the velocity in the rf micromotion [1-5]. The basic idea is as follows: in the quadrupole Paul trap the rf micromotion velocity increases with the distance the ion is from the center of the trap. As the number of trapped ions increases, space charge repulsion holds ions farther from the center of the trap thereby increasing the ion's micromotion speed and second-order Doppler shift. For a spherical cloud of ions, the number of ions N and the fractional second-order Doppler shift averaged over the cloud $\langle\Delta\nu_{D2}/\nu_o\rangle$ are proportional [2-4]:

$$N = -2.16 \times 10^{16} r_{c\ell} M \langle\Delta\nu_{D2}/\nu_o\rangle / Z^2 \quad \text{(rf trap)} \qquad (1)$$

where $r_{c\ell}$ is the cloud radius in centimeters, M is the ion mass in atomic mass units and Z is ion charge in units of the proton charge. The number of ions for a given second-order Doppler shift can be increased by using a nonspherical ion cloud geometry in an elongated trap [5]. However, the basic idea still holds that as the number of ions increases so does $|\langle\Delta\nu_{D2}/\nu_o\rangle|$.

In the Penning trap, when the cyclotron and axial kinetic energies are reduced to small values, the second order Doppler shift is dominated by the velocity in the rotation of the ion cloud. In this case, N and $\langle\Delta\nu_{D2}/\nu_o\rangle$ are related by [1-3]

$$N = 3.10 \times 10^{13}\, B \langle -\Delta\nu_{D2}/\nu_o\rangle^{\frac{1}{2}}\, z_{c\ell}(r_{c\ell} - r_c')/Z \quad \text{(Penning trap)} \qquad (2)$$

where B is the trap magnetic field strength in teslas, $2z_{c\ell}$ and $r_{c\ell}$ are the cloud height and radius in centimeters, $r_c' = (5\langle -\Delta\nu_{D2}/\nu_o\rangle)^{\frac{1}{2}} c/\Omega_c$ (Ω_c is the ion's cyclotron frequency and c is the speed of light) and Z is the ion charge in units of proton charge. Equations (1) and (2) are valid for one species of ion in the trap.

Therefore, for both traps, $|\langle\Delta\nu_{D2}/\nu_o\rangle|$ increases as N increases and we are faced with a compromise in design. For good signal-to-noise ratio and therefore good frequency stability, we desire large N. However this increases $|\langle\Delta\nu_{D2}/\nu_o\rangle|$ and therefore decreases accuracy because of our inability to measure precisely the velocity distributions needed to determine $\langle\Delta\nu_{D2}/\nu_o\rangle$. This trade-off between stability and accuracy has resulted in different approaches. In the work of one group[4,7], a stored $^{199}Hg^+$ ion standard with excellent stability has been realized. In these experiments $N \simeq 2 \times 10^6$ and $\langle\Delta\nu_{D2}/\nu_o\rangle \simeq -2 \times 10^{-12}$, so an accuracy of 10^{-14} would require a knowledge of $\langle v^2\rangle$ to 0.5% precision. This problem can be reduced in the elongated trap geometry of [5] but independent of the trap geometry, the accuracy can generally be improved by using smaller numbers.

At NIST the primary goal of stored-ion frequency standard work has been high accuracy. The preceding arguments have forced us to use small numbers of ions (approximately 10^4 or less) which are laser-cooled. The loss in short-term stability due to reduced numbers can be partially regained by going to very long interrogation times. These trade-offs are apparent from the expression for stability if we assume the Ramsey method of interrogation with 100% detection efficiency. For these conditions [3],

$$\sigma_y(\tau) = (\tau N T_R \omega_o^2)^{-\frac{1}{2}} \quad (\tau > T_R) \qquad (3)$$

where τ is the averaging time, T_R is the Ramsey interrogation time and ω_o is the clock transition frequency (in radians per second). From this expression, we also see the importance of making ω_o large.

Of course, to achieve high accuracy, we must also account for the perturbations due to static and time-varying multipole interactions for electric,

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

magnetic, and gravitational fields. These include atom-trap field interactions, collisions, shifts due to uncontrolled electric and magnetic fields, and gravitational red shifts. However for more than one ion in the trap, the dominant uncertainty appears to be caused by the uncertainty in the second-order Doppler shift.

Below, we briefly discuss experiments at NIST on stored-ion frequency standards. More detailed accounts are forthcoming.

$^9Be^+$ Hyperfine Clock

In this experiment, an oscillator has been locked to the (m_I = -1/2, m_J = 1/2) ↔ (-3/2, 1/2) nuclear spin-flip hyperfine "clock" transition ($\omega_o/2\pi \simeq 303$ MHz) in the ground state of $^9Be^+$ (Fig. 1). The

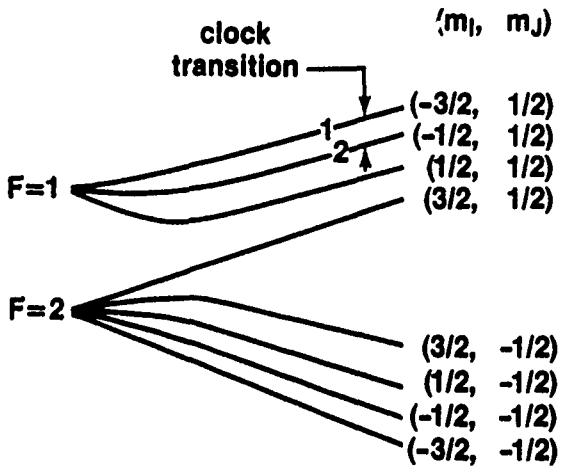


Fig. 1. Hyperfine energy levels (not drawn to scale) of the $^9Be^+$ 2s $^2S_{1/2}$ ground state as a function of magnetic field. At B = 0.8194 T the 303 MHz clock transition is independent of magnetic field to first order.

basic idea of this experiment has been described previously [8-10]; the current experiment works as follows [11]: Between 5000 and 10 000 $^9Be^+$ ions and 50 000 to 150 000 $^{26}Mg^+$ ions were simultaneously stored in a cylindrical Penning trap [8] with B ≃ 0.8194 T under conditions of high vacuum (≲ 10^{-8} Pa). At a magnetic field B of 0.8194 T the clock transition depends only quadratically on magnetic field fluctuations, and therefore the accuracy is not limited by field fluctuations. To minimize second order Doppler shifts of the clock transition, the $^9Be^+$ ions were cooled to less than 250 mK in the following manner: The $^{26}Mg^+$ ions were directly laser-cooled and compressed by a narrow-band (~ 1 MHz) laser radiation source at 280 nm. The $^9Be^+$ ions were then sympathetically cooled [12] by their Coulomb interaction with the cold Mg^+ ions (see Appendix A). A narrow-band 313 nm radiation source was used to optically pump and detect the $^9Be^+$ ions [8-10]. With the 313 nm source tuned to the 2s $^2S_{1/2}$ (m_I = 3/2, m_J = 1/2) to 2p $^2P_{3/2}$(3/2, 3/2) transition, 94% of the $^9Be^+$ ions were optically pumped into the 2s $^2S_{1/2}$(3/2, 1/2) ground state. The 313 nm source was then turned off to avoid optical pumping and ac Stark shifts while the clock transition was driven. The sympathetic cooling of the $^9Be^+$ ions by the continuously cooled Mg^+ ions is necessary if long interrogation times are to be used, since otherwise the $^9Be^+$ ions would slowly heat up while the 313 nm laser is off.

The clock transition was detected by the following method: After the 313 nm source was turned off, the ions in the (3/2, 1/2) state were transferred to the (1/2, 1/2) state and then to the (-1/2, 1/2) state by two successive rf π pulses. Each pulse was 0.2 s long and was resonant with the appropriate transition frequency (around 321 MHz and 311 MHz respectively). The clock transition was then driven by Ramsey's method of separated oscillatory fields in the time domain with rf pulses of about 1 s duration and a free precession time on the order of 100 s. This transferred some of the ions from the (-1/2, 1/2) state to the (-3/2, 1/2) state. Those ions remaining in the (-1/2, 1/2) state were then transferred back to the (3/2, 1/2) state by reversing the order of the two rf π pulses. The 313 nm source was then turned back on, and the population of ions in the (-3/2, 1/2) state was registered as a decrease in the $^9Be^+$ fluorescence, relative to the steady-state fluorescence, during the first second that the 313 nm source was on. (The optical repumping time of the ions from the (-3/2, 1/2) state to the (3/2, 1/2) state was about 10 s.)

The Ramsey signal was used to steer the frequency of a synthesized rf source [8-10]. Ramsey signal measurements were taken near the frequencies corresponding to the half minimum points on both sides of the center frequency. The difference in the measured signal strengths on either side of the line center was used to electronically steer the average frequency of the synthesizer to ω_o. Most runs were taken with a commercial cesium beam clock (fractional frequency stability $\sigma_y(\tau) \simeq 6 \times 10^{-12}\tau^{-1/2}$ for measurement time τ in seconds) as the reference oscillator, but a few runs were taken with a passive hydrogen maser ($\sigma_y(\tau) \simeq 2\text{-}3 \times 10^{-12}\tau^{-1/2}$) as the reference oscillator. Stabilities of the $^9Be^+$ clock are measured to be better than $3 \times 10^{-12}\tau^{-1/2}$ for the number of ions used, which is within a factor of 4 of the theoretical maximum stability given by Eq.3. The systematic error of our measurement of ω_o due to uncertainty in the second-order Doppler frequency shift is about 5×10^{-15}. We are continuing to search for other causes of systematic errors such as electronic offsets and pressure shifts due to collisions with background gas; however, we feel the accuracy could be improved beyond 5 parts in 10^{15} in the future.

$^{199}Hg^+$ Optical Clock

The velocity in the micromotion for an ion in a quadrupole rf trap is proportional to the distance of the ion from the center of the trap. For two or more laser-cooled ions in the trap, the Coulomb repulsion between ions holds them away from the trap center, and the second-order Doppler shift is dominated by the velocity of micromotion. However, a single ion can be held near the trap center if sufficiently cooled. In this case the kinetic energy in the micromotion can be equal to that of the secular motion [2,3,13,14]. If the ion is laser-cooled, resulting Doppler shifts can be

extremely small; uncertainties can be less than 1 part in 10^{20} in some cases [15]. However with N = 1, stability is marginal unless we make ω_o high enough (see Eq.3). One way to accomplish this is to let ω_o correspond to an optical transition. Dehmelt suggested this idea in 1973 [15]. The reasons that a clock based on an optical transition in an ion has not been realized yet are: (1) it took several years to isolate and manipulate single ions in the traps, (2) local oscillators (lasers) with the desired spectral purity are still not available, and (3) accurate comparison of laser and microwave frequencies is extremely difficult and remains an important problem. Nevertheless the potential accuracy of single-ion optical frequency standards is extremely high [16,17] ($\propto 10^{-18}$) so it is important to pursue this research.

At NIST we have investigated the use of the $5d^{10}6s\ ^2S_{1/2} \rightarrow 5d^9 6s^2\ ^2D_{5/2}$ electric quadrupole transition ($\omega_o/2\pi \propto 1.07 \times 10^{15}$ Hz) in $^{198}Hg^+$ (see Fig. 2) as

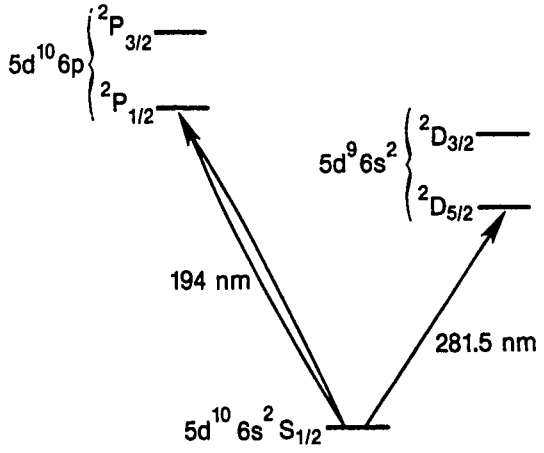


Fig. 2. Simplified optical energy-level diagram for Hg^+. The lifetime of the $^2D_{5/2}$ level is about 0.1 s which would give a linewidth of 2 Hz on the electric quadrupole $^2S_{1/2} \rightarrow {}^2D_{5/2}$ transition. By observing the presence of or lack of fluorescence from the $^2S_{1/2} \rightarrow {}^2P_{1/2}$ transition, the quadrupole "clock" transition can be detected with 100% efficiency.

an optical frequency standard [18]. The single mercury ion is confined in a miniature rf trap that has internal dimensions of $r_o \propto 466$ μm and $z_o \propto$ 330 μm [18,19]. The amplitude of the trapping field (frequency $\Omega/2\pi \propto$ 21-23 MHz) can be varied to a peak of 1.2 kV. The ion is laser-cooled to a few millikelvins by a few microwatts of cw laser radiation that is tuned below the $^2S_{1/2} - {}^2P_{1/2}$ first resonance line near 194 nm. In order to cool all motional degrees of freedom to near the Doppler cooling limit [6] ($T = \hbar\gamma/2k_B \propto 1.7$ mK) the 194 nm radiation irradiates the ion from 2 orthogonal directions, both of which are at an angle of 55° with respect to the symmetry (z) axis of the trap. The 282 nm radiation that drives the narrow $^2S_{1/2} - {}^2D_{5/2}$ transition is obtained by frequency-doubling the radiation from a narrowband cw ring dye laser. The frequency of the laser is stabilized by locking it to a stable Fabry-Perot cavity. The frequency of the laser is scanned by an acousto-optic modulator that is driven by a computer controlled synthesizer. Up to a few microwatts of 282 nm radiation could be focussed onto the ion in a direction counterpropagating with one of the 194 nm light beams.

Optical-optical double resonance was used to detect transitions driven by the 282 nm laser to the metastable $^2D_{5/2}$ state [19]. The 194 nm fluorescence rate from the laser-cooled ion is high when the ion is cycling between the $^2S_{1/2}$ and $^2P_{1/2}$ states (Fig. 2) and nearly zero when the ion is in the metastable $^2D_{5/2}$ state. The $^2S_{1/2} - {}^2D_{5/2}$ resonance spectrum was obtained by probing the S-D transition at a particular frequency for the 282 nm radiation for 20 ms, then turning off the 282 nm radiation and turning on the 194 nm radiation to look for the presence or absence of scattered photons at 194 nm (the two radiation fields are alternately applied to avoid light shifts and broadening of the narrow S-D transition). If there was no fluorescence at 194 nm, a transition into the metastable D state had occurred; the presence of 194 nm fluorescence indicated that the ion was in the ground state and no transition was recorded for this frequency of the 282 nm laser. The frequency of the 282 nm radiation was then stepped and the measurement cycle repeated. As the frequency was swept back and forth each new result at a particular frequency of the 282 nm radiation was averaged with the previous measurements at that frequency. Normalization (or digitization) of the signal was obtained by assigning a 1 to each measurement of high fluorescence and a 0 to each measurement of no fluorescence. The high fluorescence level made it possible to determine the state of the atom with almost no ambiguity in a few milliseconds. Thus, it is possible to reach the shot noise limit imposed by the single atomic absorber [19].

The quantized fluorescence signal obtained from an 8 MHz scan of the 282 nm laser through the $^2S_{1/2}(m_J = -1/2) \rightarrow {}^2D_{5/2}(m_J = 1/2)$ Zeeman component of the electric quadrupole transition in $^{198}Hg^+$ is shown in Fig. 3. The recoilless-absorption resonance (carrier) and the motional sidebands due to the secular motion in the harmonic well of the rf trap are completely resolved [19].

To avoid broadening of the quadrupole transition due to magnetic field fluctuations, we have recently performed the same experiment on the $^2S_{1/2}(F=0, m_F=0) \rightarrow {}^2D_{5/2}(F=2, M_F=0)$ transition in $^{198}Hg^+$ which becomes field independent as $B \rightarrow 0$. The carrier is now observed with a linewidth $\Delta\nu \leq 100$ Hz (limited by laser spectral purity), which gives a line Q of about 10^{13}, the highest reported in atomic or molecular spectroscopy. Current efforts are devoted to improving the 282 nm laser spectral purity by locking it to a more stable reference cavity. If the laser's spectral purity can be made high enough, then when the laser is locked to the ion transition, stabilities are anticipated to be better than $10^{-15}\tau^{-1/2}$ and accuracies could be 1 part in 10^{18} or better [1,3,14-19].

Future Penning Trap Experiments

The $^9Be^+$ ion experiments have the primary disadvantage that ω_o is relatively low and the resulting frequency stabilities are modest. We

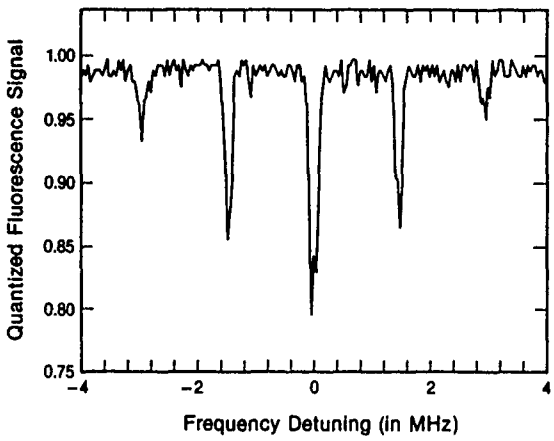


Fig. 3. Quantized signal showing the electric-quadrupole-allowed $5d^{10}6s\ ^2S_{1/2}(m_J=-1/2)$ - $5d^9 6s^2\ ^2D_{5/2}(m_J=1/2)$ transition in a single, laser-cooled $^{198}Hg^+$ ion. On the horizontal axis is plotted the relative detuning from line center in frequency units at 282 nm. On the vertical axis is plotted the probability that the fluorescence from the 6s $^2S_{1/2}$ - 6p $^2P_{1/2}$ first resonance transition, excited by laser radiation at 194 nm, is on. The electric-quadrupole-allowed S-D transition and the first-resonance S-P transition are probed sequentially in order to avoid light shifts and broadening of the narrow S-D transition. Clearly resolved are the recoilless absorption resonance (carrier) and the Doppler sidebands due to the residual secular motion of the laser-cooled ion. Each point is the average of 230 measurement cycles (From ref. 19)

might hope to substitute $^{201}Hg^+$ ions in place of the $^9Be^+$ ions because $^{201}Hg^+$ has a higher frequency clock transition ($\omega_o/2\pi \simeq 26$ GHz) which is field independent at B = 0.534 T [1]. However two disadvantages compared to the $^9Be^+$ case arise: (1) If the Hg^+ is sympathetically cooled by lighter ions such as Mg^+ or Cd^+ it will reside in an annulus surrounding the lighter ions [12]; this tends to make the second-order Doppler shift larger for a given density and number of ions. (2) $^9Be^+$ and $^{25}Mg^+$ have simple optical pumping schemes whereby a single laser frequency can be used to optically pump into a single ground state sublevel [9]. For $^{201}Hg^+$ or $^{199}Hg^+$ in a strong field (required for a Penning trap) optical pumping schemes would require auxiliary laser lines at 194 nm and microwave oscillators to manipulate the ground state sublevels [1]; the simple optical pumping schemes as in the case of $^9Be^+$ and $^{25}Mg^+$ do not appear possible. $^{199}Hg^+$ in a Penning trap would provide a very interesting system when magnets of high enough field strength become available. For example, the (m_I = 1/2, m_J = 1/2) $\leftrightarrow$ (-1/2, 1/2) hyperfine transition in the ground state of $^{199}Hg^+$ ($\omega_o/2\pi \simeq 20.9$ GHz) is field independent at B = 43.9 T. At present, we must await the required magnet.

Within the limits imposed by today's technology, an experiment similar to the $^9Be^+$ experiment but with better expected performance might be provided by $^{87}Zn^+$ ions [20]. The clock transition could be the (m_I = 5/2, m_J = 1/2) $\leftrightarrow$ (3/2, 1/2) transition ($\omega_o/2\pi \simeq 1$ GHz) which is field independent at B $\simeq$ 8 T. Some other examples are summarized in Table 1.

Similar schemes are possible with ions such as Ba^+ but optical pumping becomes even more complicated.

Future Paul Trap Experiments

The advantage of the Paul trap is that a magnetic field is not required for trapping. This allows us to be rid of a cumbersome magnet and allows use of transitions which are field independent at B = 0. The primary disadvantage is that up to the present time it has been impossible to laser-cool very many ions (N > 100). As discussed above, the use of small numbers may not be a limitation if ω_o can be made big enough. This is the basic philosophy behind the single-ion optical frequency standards. Even for $\omega_o/2\pi$ = 40.5 GHz ($^{199}Hg^+$) and N = 1, from Eq. 3 we can expect [3] $\sigma_y(\tau) = 3.9 \times 10^{-13}\tau^{-1/2}$ when T_R = 100 s. Because the second-order Doppler shift is expected to be so small for single ions, it is perhaps useful to look at this case a little more closely.

The main advantage of using a single ion in an rf trap is that the kinetic energy of micromotion can be on the order of the secular motion energy. For a single $^{199}Hg^+$ ion cooled to the Doppler-cooling limit [.9], the second-order Doppler shift would be [3] $\langle\Delta\nu_{D2}/\nu_o\rangle$ = - 2.3 $\times$ 10^{-18}. In a quadrupole ion trap, two or more ions in the trap are pushed from the center of the trap by the mutual Coulomb repulsion and the second-order Doppler shift is higher [2]. Consider the trap shown in Fig. 4. This design is essentially the same as that described by Prestage et al. [5] and Dehmelt [21].

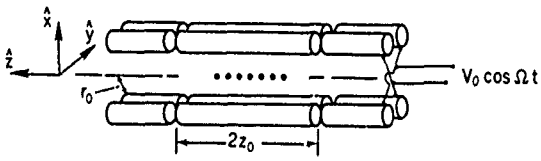


Fig. 4. Linear trap configuration. The alternating rf voltage $V_o\cos\Omega t$ is assumed to be applied to diagonally opposing electrodes as shown. We assume the end portions of the electrodes are long enough that the resulting rf potential at the position of the ions is independent of z, so that the rf electric fields are parallel to the x-y plane. To trap ions along z, we assume the center four electrodes are at static ground potential and the two sets of four electrodes on either end are at a static potential U_o ($U_o > 0$ to trap positive ions). The average position of the ions could be made to coincide with the rf electric field null by applying slightly different static potentials to the four central rods to correct for contact potential offsets etc.

In the trap of Fig. 4, the rf electric fields are transverse to the trap axis for the entire z extent of the trap. If a single string of ions is trapped along the z axis, then the kinetic energy of micromotion is about equal to the kinetic energy in the secular motion. Therefore, the fractional second-order Doppler shift could be as low as $5kT/2mc^2$. This is 5/6 of the value for a quadrupole

TABLE 1.

M_1(u)	M_2(u)	ω_o	B(T)	ξ	$\lvert\Delta\nu_{D2}/\nu_o\rvert$	b_1	$z_{c\ell}$(cm)	N_1	T^*	$\sigma_y(1s)$ (T_R=100 s)
9(Be^+)	26(Mg^+)	303 MHz[10]	0.8194	0.1	10^{-15}	125 μm	0.5	6630	65 mK	6.45×10^{-13}
9(Be^+)	26(Mg^+)	303 MHz[10]	0.8194	0.01	10^{-15}	1.25 mm	0.5	67350	65 mK	2.02×10^{-13}
67(Zn^+)	113(Cd^+)	1 GHz[20]	8	0.1	10^{-15}	56 μm	0.5	28450	484 mK	9.44×10^{-14}
67(Zn^+)	113(Cd^+)	1 GHz[20]	8	0.01	10^{-15}	560 μm	0.5	292400	484 mK	2.94×10^{-14}
199(Hg^+)	203(Tl^+)	19.8 GHz	43.5	0.01	10^{-16}	58 μm	0.5	52420	144 mK	3.51×10^{-15}
201(Hg^+)	203(Tl^+)	25.9 GHz[1]	0.534	0.1	10^{-15}	1.48 mm	0.5	50070	1.45 mK	2.75×10^{-15}
201(Hg^+)	203(Tl^+)	25.9 GHz[1]	0.534	0.01	10^{-15}	1.48 cm	0.5	524200	1.45 K	8.49×10^{-16}
201(Hg^+)	203(Tl^+)	7.72 GHz[1]	3.91	0.1	10^{-15}	204 μm	0.5	50070	1.45 K	9.21×10^{-15}
201(Hg^+)	203(Tl^+)	7.72 GHz[1]	3.91	0.01	10^{-15}	2.04 mm	0.5	524200	1.45 K	2.85×10^{-15}

Table 1. Expected frequency stabilities for various sympathetically cooled ions of interest for frequency standards. M_1 is the mass of the "clock" ion in atomic mass units (u), M_2 is the mass of the ion which is directly laser cooled and sympathetically cools M_1. For all cases $M_1 < M_2$ so that the cloud geometry of Fig. 5 applies. The clock frequency ω_o is assumed to be "field independent" as for the $^9Be^+$ case described in the text. This choice determines the magnetic field. From the magnetic field B and an assumed value of ξ (Eq. A3), the density of species 1 can be determined from Eq. A2. If we desire a certain value of the second order Doppler shift $\langle\Delta\nu_{D2}/\nu_o\rangle$, we then determine a value of the cloud radius b_1 and the ion number (Eq. A4) for an assumed value of $z_{c\ell}$. From Eq. 3, we then determine $\sigma_y(\tau)$ assuming T_R = 100 s for the Ramsey interrogation time. We refer to $\sigma_y(1s)$ for convenience. T^* is the temperature at which the second order Doppler shift from the axial and cyclotron motion would be equal to the value listed in the table. Therefore we have implicitly assumed the temperature of the cyclotron and axial motion is much less than T^*. From this, we see the importance of laser cooling.

rf trap [2] because of the absence of rf micromotion along the z direction. At the Doppler-cooling limit, this gives $\Delta\nu_{D2}/\nu_o \simeq -2\times10^{-18}$ for all Hg^+ ions in the string.

Use of the trap of Fig. 4 would allow N » 1 giving good stability and still yield a small second-order Doppler shift. For the experimentally achievable conditions of Appendix B, N = 50 ions could be stored along the z axis like "beads on a string" with ion spacings of approximately 5 μm. With this spacing each ion could be independently detected by using an image detector [22,23]. Therefore each ion could be treated as an independent atomic clock where the clock transition could be detected with 100% efficiency [19]. From Eq. 3, for T_R = 100 s and $\omega_o/2\pi$ = 40.5 GHz ($^{199}Hg^+$) the frequency stability of this clock "ensemble" would be $\sigma_y(\tau) = 5.5\times10^{-14}\tau^{-\frac{1}{2}}$. For these long interrogation times, sympathetic cooling might be required to avoid heating while the Hg^+ optical pumping laser was turned off to avoid light shifts during the Ramsey period. The ions used to cool the Hg^+ ions would also find a position in the string of ions along the z axis. Arrays of traps have also been proposed previously [24]. These trap arrays would have the same advantages as above for increasing N and minimizing the second order Doppler shift.

For optical spectroscopy, lasers with high spectral purity are required (see below). This is the current limitation to reaching the natural linewidth resolution in the NIST Hg^+ experiments. In future experiments, the laser can be locked to the ion resonance line in a manner similar to that described for the $^9Be^+$ hyperfine transition. A test of this lock can be made by briefly interrupting the lock, holding the servos while probing a portion of the clock resonance line, and then returning to servo operation. This process is then repeated until a resonance line similar to that of Fig. 3 is obtained. This, of course, does not test for a slow drift in the ion resonance frequency, but should be a good test of the ability to servo the laser to the clock transition.

The main systematic error for the Hg^+ optical experiment may ultimately be the uncertainty in the shift of clock transition from static electric fields of quadrupole symmetry such as those caused by static potentials applied to the trap electrodes. The basic idea is that the D state of the clock transition has a quadrupole moment that can interact with static fields of quadrupole symmetry to give a shift which must then be measured. Although troublesome, it should be possible to remove this offset from a clock calibration to about the same precision as the measurement precision [3]. Dehmelt has pointed out the advantage of the singly ionized group IIIA atoms in this respect [16]; the interesting clock transitions in these ions are the $^1S_o \rightarrow {}^3P_o$ intercombination lines which are not shifted by static quadrupole electric fields. However, at low magnetic field these transitions have magnetic field dependence on the order of the nuclear Zeeman effect. Therefore careful control of the magnetic field would be required [16]. At higher fields, field-independent transitions could be used to advantage [25]. A linear trap or trap arrays could be used to increase signal-to-noise ratio as described above. For clock transitions involving a state with a quadrupole moment, such as for Hg^+, the mutual Coulomb interaction between ions would cause an additional quadrupole frequency shift which must be taken into account [26]. For the group IIIA ions, this shift would be absent.

Lasers and Laser Metrology

For optical frequency standards and spectroscopy experiments, the most outstanding problem is that of local oscillator (laser) spectral purity. Lasers which are locked to reference cavities have been shown to track the cavity frequency to precisions of much less than 1 Hz [27]. The problem then remains to make this reference cavity stable enough over the attack time of the servo used to lock the laser to the ion line.

A more general problem is that of cost, power requirements, and complexity of the lasers required for optical pumping and laser cooling. At present, most experiments use near-ultraviolet laser lines provided by gas laser pumped dye lasers which are frequency doubled or summed to give a few microwatts at wavelengths corresponding to certain transitions in various ions. However, this technology may simplify in the coming years. Partly because of efforts by the opto-electronics industry, it is not unreasonable to think that such lasers will be replaced by cheap, high-efficiency solid-state lasers. This is already happening at near-infrared wavelengths where diode lasers are used to optically pump and laser-cool neutral cesium and rubidium atoms.

Finally, we call attention to the problem of frequency metrology which is important in the generation of time from optical frequency standards. The technology to connect the microwave region of the spectrum to the optical spectrum through the use of synthesis chains exists [28]. Other schemes have been proposed and are being worked on [29, 30]. However, this metrology problem is significant and simpler schemes would help realize the full benefits of optical frequency standards.

Acknowledgements

We gratefully acknowledge the continued support of the Office of Naval Research and the Air Force Office of Scientific Research. We thank R. Drullinger, H. Robinson and M. Young for helpful comments on the manuscript.

Appendix A: Sympathetically Cooled Ions in a Penning Trap

An approximate model of a two species ion plasma in a Penning trap is shown in Fig. 5. To make the problem tractable, we will assume that the species of spectroscopic interest is of lighter mass (species 1) and therefore resides inside species 2 which is directly laser-cooled. The case where species 2 is sympathetically cooled by species 1 [12] could be treated in a similar manner. If we assume $N_1 \ll N_2$, we can approximate the species 1 cloud by a cylinder of radius b_1, and height $2z_{c\ell}$. If we can assume that the cyclotron and axial motion of the ions have been cooled to negligible values, the second-order Doppler shift averaged over ions of species 1 is given by

$$\langle \Delta\nu_{D2}/\nu_o \rangle = - \omega^2 b_1^2/4c^2 , \qquad (A1)$$

where ω is the rotation frequency and c is the speed of light. For $N_1 \ll N_2$ the density of species 1 is

$$n_1 = m_1\omega(\Omega_1 - \omega)/2\pi q_1^2 \qquad (A2)$$

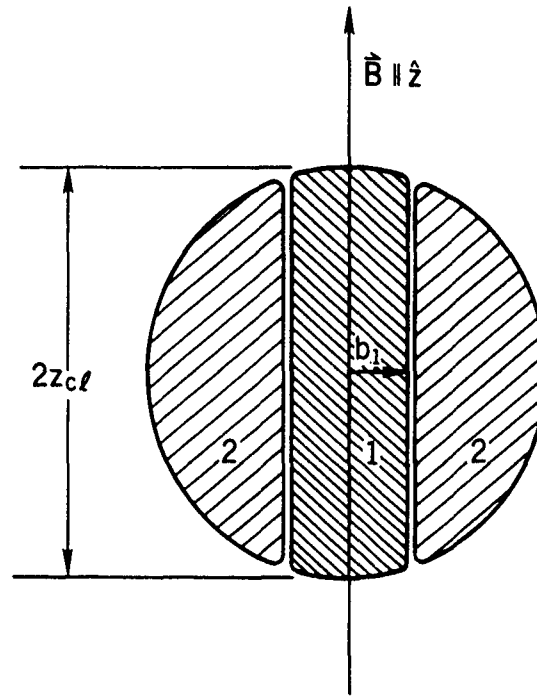


Fig. 5. Approximate model for a sympathetically cooled ion sample in a Penning trap. We assume $q_1/m_1 > q_2/m_2$ so that species 1 (which is the "clock" ion) is approximated by a uniform density column of height $2z_{c\ell}$ and radius b_1. Species 2 is assumed to be continuously laser cooled and by the Coulomb interaction continuously cools species 1. All ions are assumed to be in thermal equilibrium and therefore co-rotate about the z axis at frequency ω. The second order Doppler shift of species 1 is assumed to be dominated by the velocity in this rotation.

where m_1, Ω_1, and q_1, are the mass, cyclotron angular frequency and charge of species 1 [31,32]. An experimentally useful parameter to characterize the density of species 2 is the parameter

$$\xi \equiv 2\omega/\Omega_2 \qquad (A3)$$

where Ω_2 is the cyclotron frequency of species 2. ξ is the ratio of the rotation frequency ω to the maximum allowable rotation frequency $\Omega_2/2$ given by the Brillouin density [31,32]. A typical value of ξ for a laser compressed ion cloud is about 0.1.

From Eqs. A1 - A3, and the formula $N_1 = 2\pi b_1^2 z_{c\ell} n_1$, we have

$$N_1 = 8m_2c^2z_{c\ell}(1 - \xi m_1 q_2/2m_2 q_1)\langle -\Delta\nu_{D2}/\nu_o \rangle/\xi q_1 q_2 \qquad (A4)$$

where m_2 and q_2 are the mass and charge of ions of species 2. From these expressions and Eq. 3 applied to species 1, we can generate the parameters of Table 1. Accuracy of 10^{-15} with reasonable stability seems possible. Of course the local oscillator spectral purity must be good enough to reach these conditions. The $^{199}Hg^+$ example is not now realistic due to the lack of a suitable magnet.

Appendix B: Linear Paul Trap

To describe the trapping from ponderomotive fields we assume that near the trap axis, the time varying potential has the form

$$\phi = \frac{V_o(x^2-y^2)}{2R^2}\cos\Omega t \qquad (B1)$$

where R is approximately equal to the dimension r_o indicated in Fig. 4. ($R = r_o$ if the electrode surfaces lie along equipotentials of Eq. B1.) If Ω is high enough (defined below) this results in the harmonic pseudopotential well in the radial direction:

$$\phi_p = \frac{qV_o^2}{4m\Omega^2R^4}(x^2+y^2) = \frac{m}{2q}\omega_r^2(x^2+y^2) \qquad (B2)$$

where $\omega_r \equiv qV_o/(\sqrt{2}\, m\Omega R^2)$ is the single-ion oscillation frequency in the radial direction. Expression B2 is expected to be fairly accurate when $\Omega \gg \omega_r$ which is the condition for Ω being large enough.

For the static potential due to voltage U_o applied to the end sections, we will make the approximation that, near the center of the trap, this potential is given by

$$\phi_s = kU_o(z^2-(x^2+y^2)/2) = \frac{m}{2q}\omega_z^2(z^2-(x^2+y^2)/2) \qquad (B3)$$

where k is a geometric factor and ω_z is the z oscillation frequency for a single ion in the trap. This approximation should be reasonable if the extent of the ion sample is small compared to r_o and z_o. The addition of this static potential weakens the strength of the pseudopotential well in the radial direction so that the total potential in the radial direction is given by

$$\phi_r = \frac{m}{2q}(\omega_r^2-\omega_z^2/2)(x^2+y^2) = \frac{m}{2q}(\omega_r'^2)(x^2+y^2) \qquad (B4)$$

where ω_r' is the single ion radial oscillation frequency for the combined potentials.

A possible advantage of such a linear trap may be for laser-cooling large numbers of ions in an rf trap. As has been noted in recent papers [33,34], the rf micromotion in a quadrupole Paul trap perturbs the ion's spectrum and can inhibit laser-cooling. This may be one of the main reasons large numbers of ions in an rf trap have not been laser cooled. If a cooling laser is directed along the z axis of the trap in Fig. 4, the ion's spectrum is not perturbed by micromotion and laser cooling should occur according to the simple theory [35] if rf heating is not too severe.

"Beads on a String"

The most interesting case for accurate spectroscopy is when all ions are confined along the z axis. To analyze this case, first assume that ω_r' is large enough to make the ions lie along the z axis. Then, for a given N and ω_z, we minimize the total ion potential to obtain z_i, for i = (1,...,N). From the z_i we can then find a condition on ω_r' to insure all ions stay along the z axis.

For ions pinned on the z axis, the total ion potential energy is

$$q\phi_T = \frac{m}{2}\omega_z^2\sum_i^N z_i^2 + q^2\sum_{i<j}^N |z_i-z_j|^{-1} \qquad (B5)$$

Using a computer, we have minimized ϕ_T for various experimental parameters. Near z = 0 (trap center), the ions are closest together and fairly regularly spaced. The closer the spacing, the higher ω_r' must be to maintain all ions on the z axis.

To find the condition on ω_r' consider an infinite string of ions with equal spacing d along the z axis. If one ion is displaced a distance δx ($\ll$ d) from the z axis, it experiences a force away from the z axis of strength

$$F_x = \left(2\sum_{n=1}^{\infty} n^{-3}\right)\frac{q^2}{d^3}\delta x = (2.404)\frac{q^2}{d^3}\delta x \qquad (B6)$$

If we have only 5 ions spaced by d the analogous force on the center ion is given by Eq. B6 with the numerical factor equal to 2.250 rather than 2.404. Therefore even if the ions are only approximately evenly spaced, Eq. B6 will be a good estimate of F_x for a particular ion if we take d to be the mean of the spacing to the nearest neighbor ions. For the ions to be stably trapped on the z axis, we want the pseudopotential force inward to be larger than the force outward as expressed by Eq. B6. From Eqs. B4 and B6 we therefore require that

$$(\omega_r')^2 > 2.404\ q^2/md^3 \qquad (B7)$$

or $\omega_r'/2\pi > 91.9\ (M\ d^3)^{-\frac{1}{2}}$ where M is the ion mass in atomic mass units, d is the spacing in micrometers, and $\omega_r'/2\pi$ is in megahertz.

As an example, for N = 50, $\omega_z/2\pi$ = 50 kHz, and M = 199 ($^{199}Hg^+$ ions), the computer calculation gave an overall string length of 266 μm. The spacings of adjacent pairs of ions in μm, starting from the outermost pair of ions on one end and ending with the center pair are: 10.48, 8.46, 7.43, 6.77, 6.30, 5.95, 5.68, 5.46, 5.28, 5.12, 4.99, 4.88, 4.79, 4.71, 4.64, 4.58, 4.53, 4.48, 4.45, 4.42, 4.39, 4.37, 4.36, 4.35, 4.35. Using d = 4.35 μm, we find from Eq. B7, $\omega_r'/2\pi > 718$ kHz, $\omega_r/2\pi > 719$ kHz. If R = 0.75 mm and $\Omega/2\pi$ = 5 MHz, then we require $V_o > 233$ V.

References

[1] D. J. Wineland, W. M. Itano, J. C. Bergquist, and F. L. Walls, Proc. 35th Ann. Symp. Freq. Control, Philadelphia, PA, May 1981, p. 602-611. (Copies available from: Annual Frequency Control Symposium, c/o Electronic Industries Assoc., 2001 Eye St., Washington, D. C., 20006.)

[2] D. J. Wineland, Proc. Conf. on Precision Measurements and Fundamental Constants; Gaithersburg, MD, June 1981; Precision Measurement and Fundamental Constants II, B. N. Taylor and W. D. Phillips, eds. National Bureau of Standards (US) Spec. Publ. 617 (1984), pp. 83-92.

[3] D. J. Wineland, W. M. Itano, J. C. Bergquist, J. J. Bollinger, F. Diedrich, and S. L. Gilbert, in Frequency Standards and

Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 71.

[4] L. S. Cutler, R. P. Giffard and M. D. McGuire, Appl. Phys. B36, 137 (1985); L. S. Cutler, C. A. Flory, R. P. Giffard, and M. D. McGuire, Appl. Phys. B39, 251 (1986).

[5] J. Prestage, G. Janik, G. Dick and L. Maleki, J. Applied Physics, to be published. Also, these proceedings.

[6] D. J. Wineland and W. M. Itano, Physics Today, vol. 40, p. 34 (1987).

[7] L. S. Cutler, R. P. Giffard, P. J. Wheeler and G. M. R. Winkler, Proc. 41st Ann. Symp. Freq. Control, 1983, (copies available from IEEE, 445 Hoes Lane, Piscataway, NJ 08854), p. 12.

[8] J. J. Bollinger, S. L. Gilbert, W. M. Itano and D. J. Wineland, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 319.

[9] J. J. Bollinger, W. M. Itano, and D. J. Wineland, Proc. 37th Ann. Symp. Freq. Control, 1983, p. 37-41 (copies available from IEEE, 445 Hoes Lane, Piscataway, NJ 08854.)

[10] J. J. Bollinger, J. D. Prestage, W. M. Itano and D. J. Wineland, Phys. Rev. Lett. 54, 1000 (1985).

[11] J. J. Bollinger, D. J. Heinzen, W. M. Itano, S. L. Gilbert and D. J. Wineland, submitted for publication.

[12] D. J. Larson, J. C. Bergquist, J. J. Bollinger, W. M. Itano, and D. J. Wineland, Phys. Rev. Lett. 57, 70 (1986).

[13] H. G. Dehmelt, Adv. At. Mol. Phys. 3, 53 (1967).

[14] D. J. Wineland, W. M. Itano, J. C. Bergquist and R. G. Hulet, Phys. Rev. A36, 2220 (1987).

[15] F. Diedrich, J. C. Bergquist, W. M. Itano and D. J. Wineland, Phys. Rev. Lett. 62, 403 (1989).

[16] H. G. Dehmelt, IEEE Trans. Instrum. Meas. IM-31, 83 (1982), and references therein.

[17] D. J. Wineland, Science 226, 395 (1985).

[18] J. C. Bergquist, F. Diedrich, W. M. Itano, and D. J. Wineland, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy (Springer-Verlag, Berlin, Heidelberg) p. 287.

[19] J. C. Bergquist, W. M. Itano, and D. J. Wineland, Phys. Rev. A36, 428 (1987).

[20] W. M. Itano and D. J. Wineland, Bull. Am. Phys. Soc. 24, 1185 (1979).

[21] H. G. Dehmelt, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy (Springer-Verlag, Berlin, Heidelberg) p. 286.

[22] D. J. Wineland, W. M. Itano, J. C. Bergquist, S. L. Gilbert, J. J. Bollinger and F. Ascarrunz, in Non-neutral Plasma Physics, ed. by C. W. Roberson and C. F. Driscoll, A.I.P. Conf. Proc. 175, Amer. Inst. of Physics, New York, 1988. p. 93.

[23] W. M. Itano, J. C. Bergquist, and D. J. Wineland, in Proc. Workshop on Crystalline Ion Beams, Wertheim, W. Germany, ed. by R. W. Hasse, I. Hofmann, D. Liese (GSI report GSI-89-10, ISSN 0170-4546) p. 231.

[24] F. G. Major, J. de Physique (Lettres) 38, L-221 (1977).

[25] D. J. Wineland and W. M. Itano, Bull. Am. Phys. Soc. 27, 864 (1982).

[26] D. J. Wineland, W. M. Itano, J. C. Bergquist, J. J. Bollinger, and J. D. Prestage, Ann. Phys. Fr. 10, 737 (1985).

[27] D. Hils and J. L. Hall, in Frequency Standards and Metrology, Proc. 4th Symposium, ed. by A. DeMarchi, Ancona, Italy, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 162.

[28] See for example: Frequency Standards and Metrology, Proc. 4th Symposium, Ancona, Italy, ed. by A. DeMarchi (Springer-Verlag, Berlin, Heidelberg).

[29] D. J. Wineland, J. C. Bergquist, W. M. Itano, F. Diedrich and C. S. Weimer, in The Hydrogen Atom, ed. by G. F. Bassani, M. Inguscio, and T. W. Hänsch, (Springer-Verlag, Berlin, Heidelberg, 1989) p. 123.

[30] T. W. Hänsch, ibid. p. 93.

[31] T. M. O'Neil, Phys. Fluids 24, 1447 (1981).

[32] L. R. Brewer, J. D. Prestage, J. J. Bollinger, W. M. Itano, D. J. Larson, and D. J. Wineland, Phys. Rev. A38, 859 (1988).

[33] R. G. DeVoe, J. Hoffnagle, R. G. Brewer, Phys. Rev. A39, 4362 (1989).

[34] R. Blümel, C. Kappler, W. Quint, and H. Walther, Phys. Rev. A, to be published.

[35] D. J. Wineland and W. M. Itano, Phys. Rev. A20, 1521 (1979); W. M. Itano and D. J. Wineland, Phys. Rev. A36. 35 (1982).

43rd Annual Symposium on Frequency Control - 1989

LOW NOISE LASER FOR OPTICALLY PUMPED CESIUM STANDARDS

L. L. Lewis*
Ball Communication Systems Division
P.O. Box 1235
Broomfield, CO 80020

Abstract

We have developed a new external cavity diode laser for optical pumping and detection of a cesium atomic beam. The laser is about 2 × 4 × 5 cm in size, is tunable to any wavelength over a 40 nm range, and emits two collimated beams of light with power outputs to 6 mW each. The laser spectrum consists of a single longitudinal mode with very narrow linewidth. When locked to a single hyperfine transition in atomic cesium, the signal-to-noise ratio (SNR) for fluorescence detection of the transition far exceeds the SNR observed when an unmodified laser diode is used. When used in a miniature optically pumped cesium frequency standard, the laser should improve short term stability of the standard.

Introduction

A number of groups [1-10] have studied the use of optically pumped cesium beams for microwave frequency standards. The technique is of interest because it may improve the signal-to-noise ratio (SNR) on detection of the microwave transition, and reduce certain systematic frequency shifts associated with the use of state selection magnets. In some cases [11,12] however, the level of short term fractional frequency stability, $\sigma_y(\tau)$, has been limited by the noise properties of the lasers used to detect the cesium atoms. The lasers used for these studies have generally been monolithic laser diodes obtained from commercial sources. The line width of these devices are in the range of 10 to 100 MHz. Low frequency FM noise on the output of these lasers results in noise upon fluorescence detection of cesium in the atomic beam, which limits the performance of the frequency standard. In particular, the short term stability of the standard is given by [13]

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

where S/N is the power signal-to-noise ratio (SNR) in a one Hz bandwidth, Q is the microwave resonance quality factor, τ is the measurement interval, and K is a constant which depends upon the resonance lineshape and modulation/demodulation methods. Generally, K lies between 0.1 and 1.0. In the case of the NBS results [12], $S/N = 10^6$, $Q = 7 \times 10^6$, and $\sigma_y(\tau) = 10^{-11}\tau^{-\frac{1}{2}}$. It is desirable to improve the noise properties of the lasers so that the SNR and consequently the stability of the frequency standard is limited only by the 'shot-noise' contribution associated with the statistics of detecting a finite number of cesium atoms.

Various methods have been used to achieve this goal. If the microwave cavity length is made long enough, then the number of cesium atoms detected in the standard is reduced to a point where the 'shot-noise' dominates. In this case, the microwave Q increases with cavity length, and the short term stability of the standard improves until the laser noise is small compared with the 'shot-noise'. If a pumping transition such as the $F = 3 \rightarrow F' = 3$ transition (see Fig. 1) is used for detection of cesium, then the laser frequency noise becomes less important, because the transition is saturated. However, the laser frequency noise can still produce detection noise, because of sensitivity of the fluorescence detector to atomic position, and because of the short dwell time of atoms within the optical interaction region. In addition, it is difficult to obtain high collection efficiency in the fluorescence detector, which means that the quantum efficiency for detection with a pumping transition is less than one, which results in an increase in 'shot-noise.' Also, when a pumping transition is used for detection, the fluorescence signal may be so small that other noise sources, such as the Johnson noise of the detector amplifier, become significant.

When the detection laser is tuned to a cycling transition, however, such as the $F = 4 \rightarrow F' = 5$ transition in Fig.1, or when the microwave cavity length is short, the laser frequency noise becomes more important. One method of reducing this noise source is to lock the laser frequency to a reference using a fast servo loop [14-17]. This results in a narrow laser line width, and a consequent reduction of noise in the detected fluorescence. The disadvantages of the technique are that it requires fairly high speed electronics, a reference absorption cell or etalon, and the laser diode must be tunable to the desired wavelength. A second approach is to use optical feedback from a Fabry-Perot etalon [15,18,19]. This also greatly reduces the laser linewidth and removes laser frequency noise as a limitation to fluorescence detection. The disadvantages are that it requires an etalon mounted rigidly with respect to the laser diode, and the laser must tune to the exact wavelength of interest.

A third approach to reduction of laser noise is to place the laser diode within a relatively long cavity, with some means of wavelength selection [20,21]. This technique narrows the laser line width and provides a means of tuning the laser to any desired wavelength within a fairly broad range. The disadvantages are that it requires high quality antireflection coatings on the laser diode, and a stable mechanical structure for the external cavity. In this report we describe a laser structure which overcomes these difficulties. We also report measurements of the signal-to-noise ratio on fluorescence detection of a cesium atomic beam.

*Work supported by U.S. Army Laboratory Command, Contract No. DAAL01-88-C-0805

CH2690-6/89/0000-151 $1.00 C 1989 IEEE

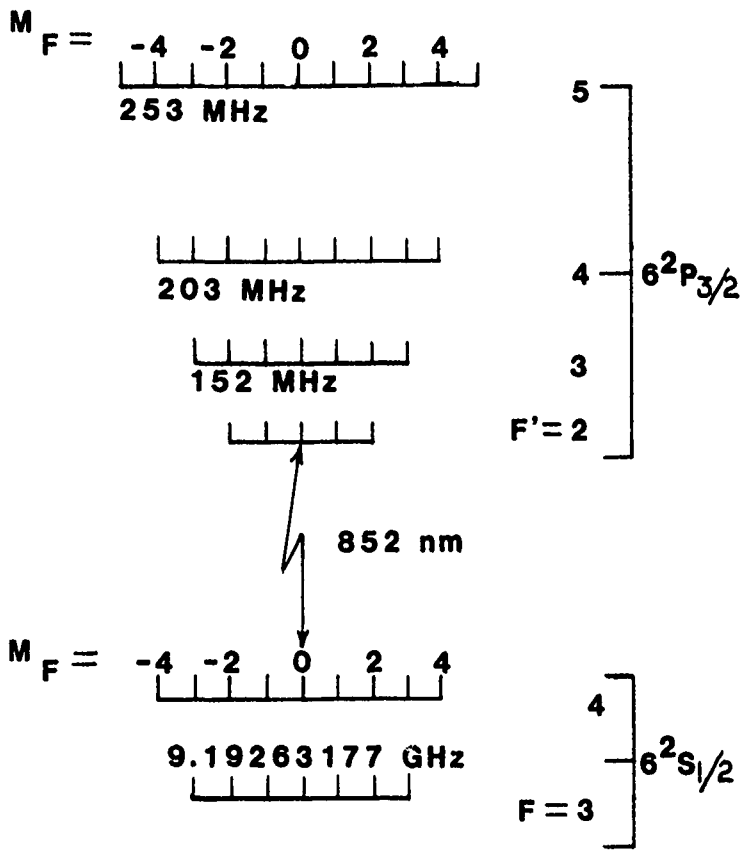


Figure 1: Energy level diagram for ^{133}Cs.

External Cavity Laser

A sketch of our method of forming an external cavity laser is given in Fig. 2. A drawing of a particular construction of this laser, called XLD-200, is given in Fig.3. The laser diode used in this device is an Hitachi HLP-1400. The back facet of the diode is gold coated, and the front facet has an anti-reflection coating, with reflectivity of less than one percent. The collimating lens is a 4.7 mm focal length, 0.50 numerical aperture lens which produces a 1 × 2 mm beam. The light is reflected by an output coupler mirror with reflectivity of 0.80 to a 1200 line/mm diffraction grating which is mounted on a piezoelectric transducer. This arrangement produces two output laser beams, with typical powers of 6 and 3 mW each. The free spectral range (FSR) associated with the external cavity length is 2.5 GHz. Residual reflectivity of the AR coated laser facet also affects the laser frequency, with a mode spacing of about 125 GHz. Coarse tuning is by adjusting the angle of the diffraction grating. This produces jumps in wavelength corresponding to the longitudinal mode spacing of the laser diode. Any specific wavelength may be reached by changing both the angle of the grating and the temperature of the laser diode, which shifts the modes of the diode. There are also mode hops corresponding to the longitudinal mode spacing of the external cavity. Any desired external cavity mode may be selected by changing the temperature of the diode and adjusting the injection current to the laser. Fine tuning is achieved by either changing the temperature of the external cavity, or applying a voltage to the PZT. In this manner, any exact wavelength may be selected and the laser held to that value indefinitely.

Laser Performance

The XLD-200 is tunable over a 40 nm wavelength range (16 THz), which greatly reduces the demands on diode selection. Previously, when monolithic laser diodes were used without benefit of the external cavity, the wavelength required selection within 2 or 3 nm. Then, with the limitations imposed by mode-hopping, approximately one-half of the lasers operated successfully at a desired wavelength. With the external cavity design, all laser diodes which operate within about 20 nm of a desired wavelength will tune to the exact wavelength.

The output of the XLD-200 is highly single mode, with less than -35 dBc of the power appearing on adjacent longitudinal modes. This fact is demonstrated in Fig. 4, where the spectrum obtained from a 0.5m monochromator is shown. Part (a) of the figure shows the carrier, and part (b) shows the adjacent spectrum with an expanded scale. The ripples in (b) are an artifact of the measurement apparatus, and there is no evidence of off-carrier light to this level.

Based on the work of other researchers [20,21], we expect the line width of the XLD-200 to be in the 10kHz range. We presently

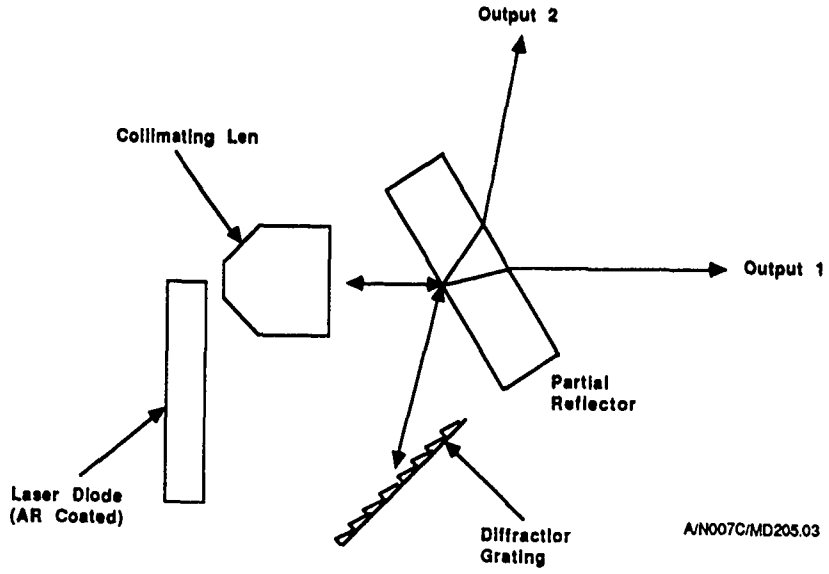


Figure 2: Sketch of external cavity diode laser.

are constructing a second laser in order to make heterodyne measurements of the line width. Using a fiber optic interferometer, we have set an upper limit on the line width of 70 kHz.

Cesium Beam Tube

In order to study the XLD-200 with fluorescence detection of cesium, we we built a single cesium beam tube, shown in Fig. 5. This beam tube used a cesium oven with a single 1 mm diameter collimator hole, and optical interaction region placed 7.6 cm from the cesium oven. We expect a cesium beam flux in excess of 6×10^{10} atoms per second at the detector. The detector consisted of two large area silicon photodetectors placed adjacent to the region of intersection of the laser beam with the cesium beam. Calculated collection efficiency of the combined detectors is 32 percent. This gives an effective quantum detection efficiency of about 50 percent for a pumping transition. If a cycling transition is used for detection, the quantum efficiency rises to 100 percent. No microwave interaction region was used in this beam tube, since the purpose of the measurement was to measure the noise properties of the laser.

The XLD-200 laser was swept across the $F = 3 \rightarrow F' = 2,3,4$ lines and also the $F = 4 \rightarrow F' = 3,4,5$ lines by sweeping the PZT voltage with the laser injection current and temperature set to appropriate values. Figure 6 shows the resultant fluorescence curve for the $F = 3$ transitions, and Fig. 7 shows the curve for the $F = 4$ transitions. Incident laser power was about 25×10^{-6} W in a 2mm diameter laser beam. About 2×10^{-8} W of fluorescent light was collected by each photodiode when the laser was tuned to the $F = 4 \rightarrow F' = 5$ transition. Increasing the laser beam intensity increases the size of the cycling transition fluorescence peak, but not the size of the pumping transition peaks indicating the optical pumping is fairly complete. Note that the width of the fluorescence peaks is an indication of the narrow line width of the XLD-200 laser.

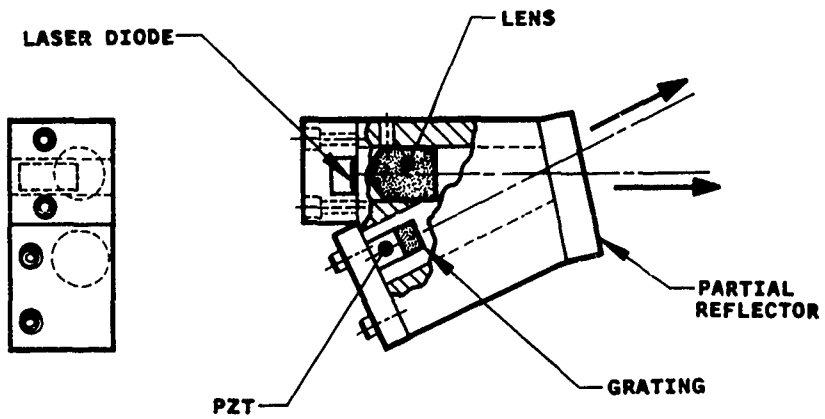


Figure 3: Line drawing of XLD-200.

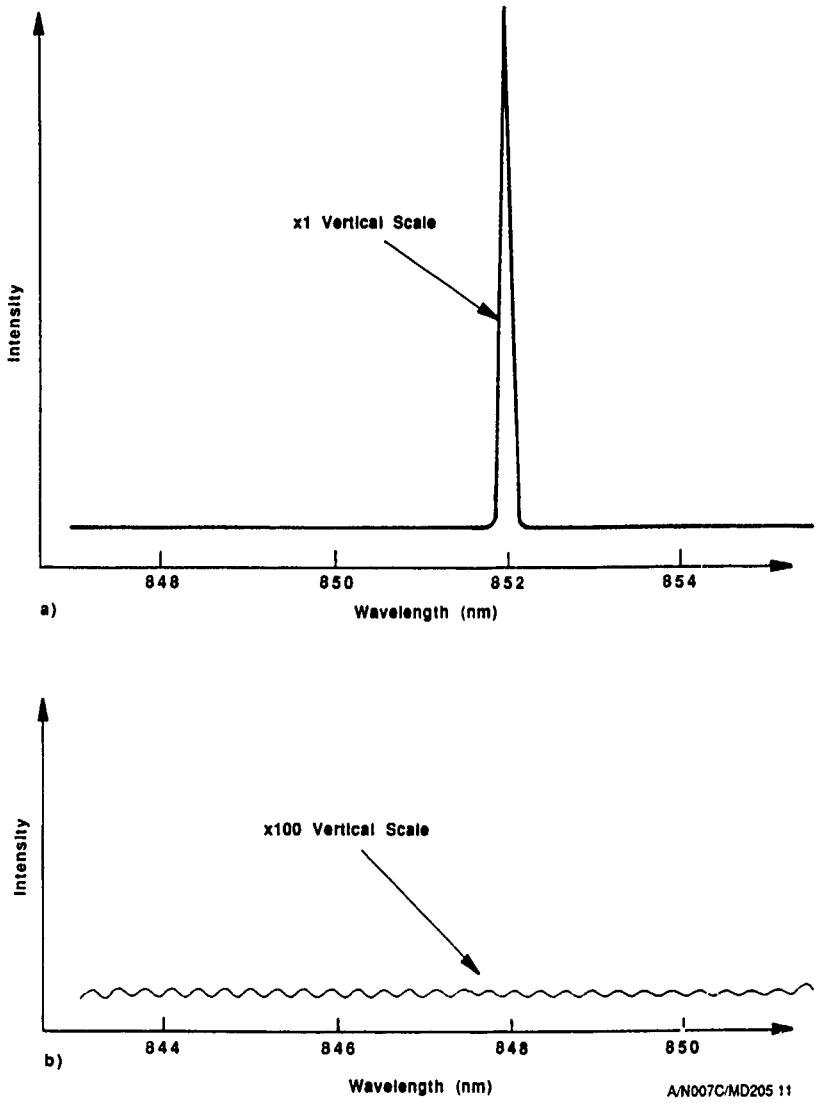


Figure 4: XLD-200 spectrum using 0.5 m monochromator.

Signal-to-Noise Ratio

In order to determine the SNR on fluorescence detection of the cesium atoms, we locked the XLD-200 frequency to the $F = 4 \rightarrow F' = 5$ cycling resonance in the atomic beam (Fig. 7). The laser frequency was tuned to this transition frequency by adjusting the laser injection current, laser and cavity temperatures, and PZT voltage. The laser frequency was modulated sinusoidally at about 680 Hz by means of the injection current, and the fluorescence signal fed to a lock in amplifier set to the same frequency. This generated a correction voltage which adjusted the PZT voltage to keep the laser tuned to the center of the atomic transition.

Under these conditions the fluorescence noise spectrum from 700 to 1000 Hz was limited by the noise floor of the photodetector and detector amplifier. The measured SNR on the $F = 4 \rightarrow F' = 5$ transition was 1.8×10^{10}/Hz. Assuming that approximately 25 photons/atoms are emitted for this transition (estimating from Fig. 7), a collection efficiency of 16 percent, and a collected power of 2.4×10^{-3} W, we predict a SNR for this situation of about 9.4×10^{9}. The discrepency between these numbers may be due to a less than ideal collection efficiency - an 8 percent collection efficiency would result in a SNR prediction of 1.8×10^{10}. More detailed measurements of these parameters are in progress.

Both the measured and calculated SNR result in an order of magnitude improvement of the short term stability limitation due to laser noise. When applied to a short microwave cavity, such as the NBS experimental beam tube [12], the noise properties of the XLD-200 laser would not limit short term fractional frequency stabilities larger than $\sigma_y(\tau) = 1 \times 10^{-12} \tau^{-\frac{1}{2}}$. Of course, in order to actually obtain such good stability values, all contributing noise sources in the frequency standard would need to be reduced below this level.

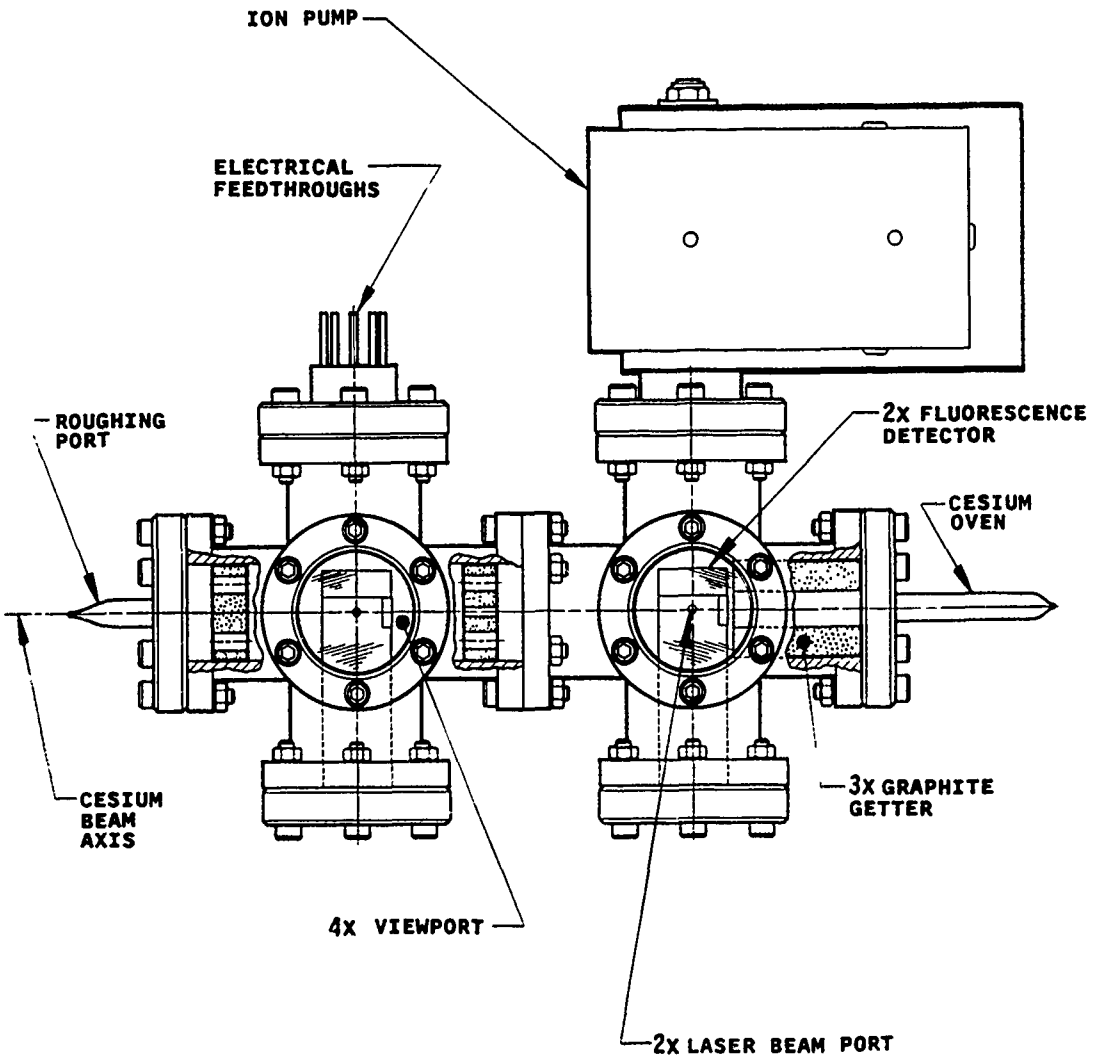


Figure 5: Line drawing of fluorescence beam tube.

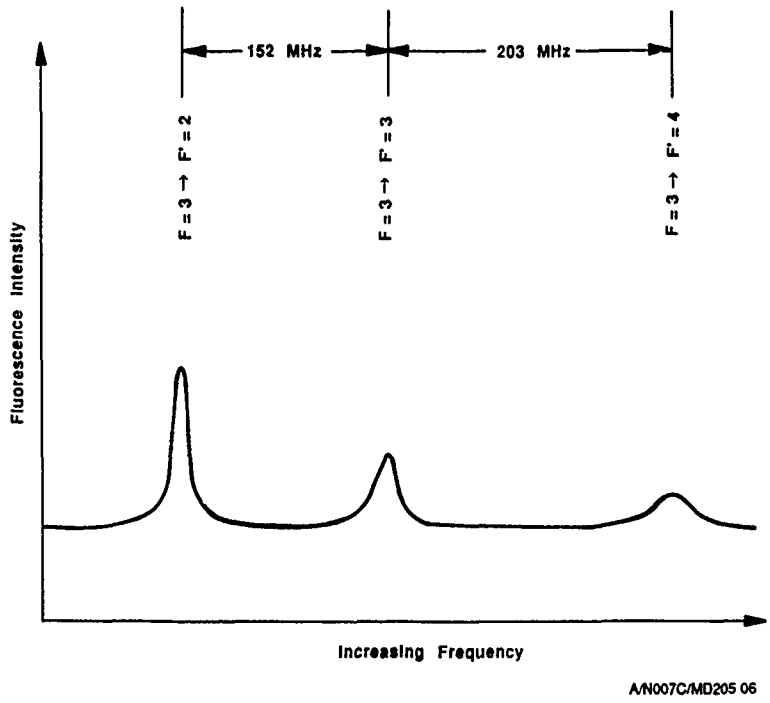


Figure 6: Cesium beam fluorescence, $F = 3 \rightarrow F' = 2, 3, 4$.

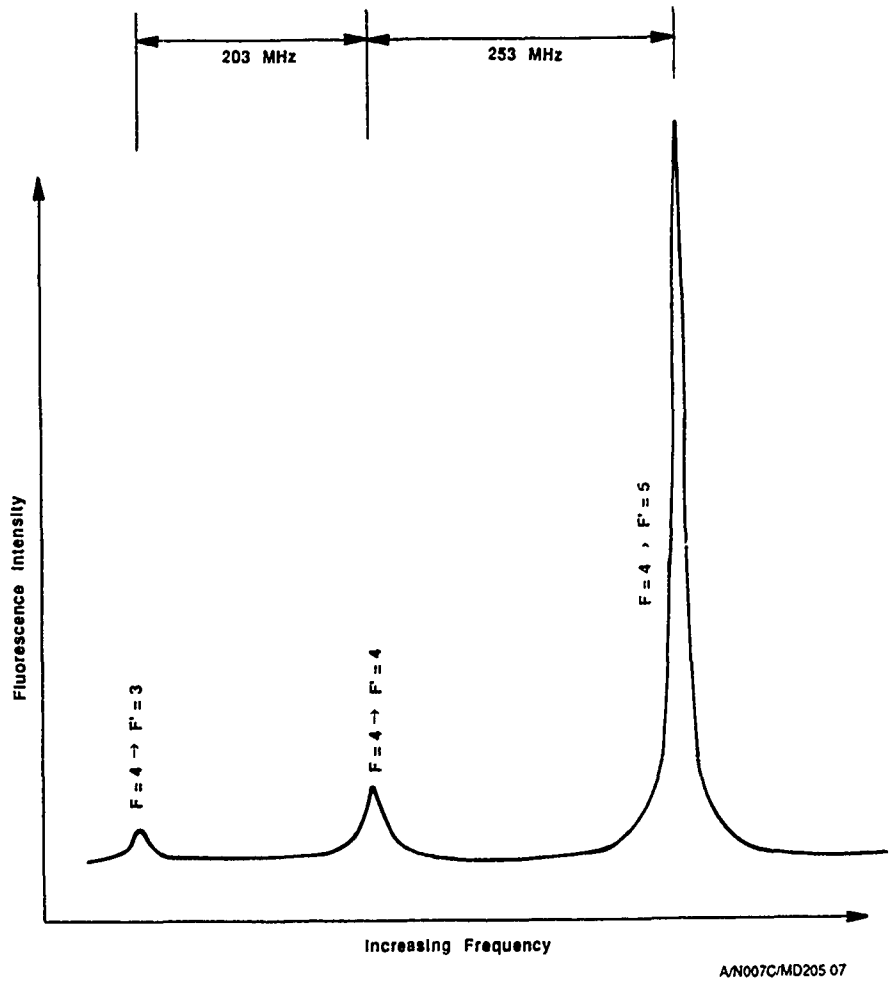


Figure 7: Cesium beam fluorescence, $F = 4 \rightarrow F' = 3, 4, 5$.

Conclusion

We have designed, constructed, and tested a compact, rugged external cavity diode laser which is tunable to the 852 nm optical transition in atomic cesium. The signal-to-noise ratio of fluorescence detection of a cesium beam using this laser is considerably larger than the SNR obtained by previous methods. This new laser may be used in a small optically pumped cesium beam standard to give improved short term frequency stability.

References

1. R.E. Drullinger, Jon Shirley, D.J. Glaze, and L.W. Hollberg, 'Progress toward an Optically Pumped Cesium Beam Frequency Standard,' Proceedings of the 40th Annual Symposium on Frequency Control, pp. 428-431, 1986.

2. A. Hamel, P. Petit, G. Theobald, P. Cerez, and C. Audoin, 'Progress in an Optically Pumped Cesium Beam Resonator Operating with a Longitudinal Magnetic Field,' Presented at the 3rd European Frequency and Time Forum, Besancon, France (March 1989). G. Theobald et al., 'Research on the Optically Pumped Cesium Beam Frequency Standards,' Proceedings of the 42nd Annual Frequency Control Symposium, pp. 496-504, 1988.

3. P. Thomann and F. Hadorn, 'Short-term stability of a Small Optically Pumped Cesium Resonator,'Presented at the 3rd European Frequency and Time Forum, Besancon, France (March 1989).

4. C. Jacques and P. Tremblay, 'Calculations on the Efficiency of Optical Pumping of a Cesium Atomic Beam by Lasers of Finite Linewidth,' Proceedings of the 42nd Annual Symposium on Frequency Control, pp. 505-509, 1988.

5. J.S. Boulanger and R.J. Douglas, 'Work on Cesium Standards at NRC,' in Frequency Standards and Metrology, ed. by A. DeMarchi (Springer-Verlag, 1989), pp. 379-381.

6. E. de Clerq, A. Clairon, B. Dahmani, A. Gerard, and P. Aynie, 'Design of an Optically Pumped Cs Laboratory Frequency Standard,' ibid., pp. 120-125.

7. S. Ohshima, Y.Nakadan, T. Ikegami, and Y. Koga, 'Experimental Results on an Optically Pumped Cs Frequency Standard,' ibid., pp. 132-136.

8. Y.G. Abashev, V.N. Baryshev, G.A. Yolkin, and Y.N. Yakovlev, 'Metrological Parameters of the Primary Cesium Beam Frequency Standards in the National Time and Frequency Standard of the USSR,' ibid., pp. 374-378.

9. Shang Song-quan, Wu Xin-xin, Yao Shu-tong, Xie Lin-Zhen, and Wang Yi-qiu, 'Microwave Transition between 'Clock Levels' of Cs Beam Pumped by a GaAlAs Diode Laser,' Chinese Physics Lett. 2, pp. 557-560, 1985.

10. Jun Umezu, Bokuji Komiyama, Haruo Saitho, and Yasusada Ohta, 'Experiments on Optically Pumped Cesium Beam Frequency Standard,' in Proceedings of the 2nd European Frequency and Time Forum, Neuchatel (March 1988) p.521.

11. V. Giordano, et al., 'New Design for an Efficient Optically Pumped Cesium Beam Tube,' presented at the 43rd Annual Symposium on Frequency Control, Denver, Colorado, 1989.

12. Amy Derbyshire, et al., 'Optically Pumped Small Cesium Beam Standards; A Status Report,' in Proceedings of the 39th Annual Symposium on Frequency Control, pp.18-21, 1985.

13. J. Vanier and L.-G. Bernier, 'On the Signal-to-Noise Ratio and Short-term Stability of Passive Rubidium Frequency Standards,' IEEE Trans. Instr. Meas. IM-30, pp. 277-282, 1981.

14. H.R. Telle, 'A FM-Stabilized Semiconductor Laser for Optical Pumping of Cesium Atoms and Length Measurements,' in Proceedings of the 2nd European Frequency and Time Forum, Neuchatel (March 1988) pp. 523-529.

15. S. Oshima, Y. Koga, Y. Nakadan, L. Hollberg, and R. Drullinger, 'The Effect of Laser Line-Narrowing on the Performance of Optically Pumped Cesium Atomic Beam Frequency Standards,' ibid., pp. 531-534.

16. Deng Jinquan, Tan Yongfang, Xiang Yuanhai, and Zhu Xiwen, 'Frequency-Stabilized Laser for Atomic Frequency Standards using Saturated Absorption by Magnetic Modulation,' ibid., p. 555.

17. M. Ohtsu and S. Kotajima, 'Linewidth Reduction of a Semiconductor Laser by Electrical Feedback,' IEEE J. Quantum Electronics, QE-21 pp.1905-1912, 1985.

18. A. Clairon, B. Dahmani, Ph. Laurent, and Ch. Breant, 'Ultra-narrow Semiconductor Laser Linewidth for Optically Pumped Cesium Clock,' in Proceedings of the 2nd European Frequency And Time Forum,' Neuchatel (March 1988) pp. 537-545.

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

20. M.W. Fleming and A. Mooradian, 'Spectral Characteristics of External Cavity Controlled Semiconductor Lasers,' IEEE J. Quantum Electronics, QE-17, pp. 44-59, 1981.

21. M. de Labachelerie, K. Diomande, and N. Dimarcq, 'Extended-Cavity Semiconductor Lasers for Optical Pumping,' in Proceedings of the 2nd European Frequency and Time Forum, Neuchatel (March 1988) pp. 547-554.

REPORT ON THE MASTER CLOCK UPGRADE PROGRAM AT USNO

GUY A. GIFFORD, Naval Research Laboratory (NRL)
Washington D.C, 20375-5000

PAUL WHEELER, U.S. Naval Observatory (USNO)
Washington D.C, 20392-5100

Abstract:

This paper will give a history of the Master Clock Upgrade Program at USNO and describe the current status of the program. Performance data on the Hydrogen Masers and Stored Ion Devices will be presented. Current projects will be described including a time difference measurement system , time scale algorithm test bed, and the new masers to be added over the next 18 months. A prediction of the status and performance of the program in two years will be made.

The Master Clock Upgrade Program was started in the late 1970's for two reasons. The first was to incorporate into the United States Naval Observatory (USNO) some of the clock technology that was being developed and in use at the Naval Research Laboratory. Of particular interest was the clock measurement technology and the hydrogen maser development for the Navy's participation in the Global Positioning System. The second reason for the creation of the Master Clock Upgrade Program was to address a more general requirement for the development of advanced frequency standards for DoD wide use. The Master Clock Upgrade Program was one of several programs funded for this purpose.

The Master Clock Upgrade Program is a SPAWAR and USNO funded project which tasks NRL and USNO to manage jointly some very specific projects. These tasks include the following:

1) Monitoring, evaluation and support of the development of the Trapped Mercury Ion Devices built by Hewlett Packard.

2) Procurement of advanced hydrogen masers for operational use.

3) Development of an improved time difference measurement system at USNO (for both short term and long term use).

4) Development of a time scale algorithm test bed.

5) Monitoring, evaluation, and support of the development of the Smithsonian Astrophysical Observatory's (SAO) advanced VLG12.

The Master Clock Upgrade Program should not be confused with the normal evolution of programs and systems at USNO. USNO is continuously developing and incorporating new technologies for the purpose of keeping and disseminating time throughout the world. The Master Clock Upgrade Program represents only a fraction of a broad spectrum of activity. These other activities include:

1) Upgrade of computing facility.

2) Upgrade of the clock set used for the calculation of the reference time scale.

3) Upgrade of OMEGA monitoring .

4) Upgrade of Loran monitoring.

5) Upgrade of GPS monitoring.

6) Upgrade of and operation of two way satellite time transfer.

7) Upgrade of information dissemination.

8) Continued research and development of time scale algorithms independent of the Master Clock Upgrade Program.

Task 1:

The first stage of development of the Mercury Ion Frequency Standard[1] is complete. The Stored Ion Devices (SID's for short) were developed in a joint technology program between the Navy and Hewlett Packard (HP). Three devices have been delivered and are in operational use by USNO while they are being evaluated These prototype units are the first of their type to be in an operational environment. Hewlett Packard is providing off line support for these devices.

These devices, serial numbers SID1, SID2, and SID3 represent an evolution of the design. SID3, the last designed, incorporates several design improvments developed with the experience and knowledge obtained from the development and operation of SID1 and SID2. SID1 was the prototype and was actually delivered last after enhancements were made. SID2 was the outgrowth of SID1 and was the first delivered. There is a known problem in the grounding of the trap in this device and it was corrected in SID3 and SID1. In time, the physics packages of SID1 and SID2 will be upgraded to that of SID3. It is not practical to upgrade the electronics of these earlier units. Figure 1 is a phase plot with the linear frequency term removed of SID3 against BIPM over a period of 300 days.

The long term performance of this device is outstanding. Additional information is available from a paper entitled "INITIAL OPERATIONAL EXPERIENCE WITH A MERCURY ION STORAGE FREQUENCY STANDARD"[2]. When the physics packages of SID1 and SID2 are completely upgraded the performance of all three is expected to be uniform.

Task 2:

Five of the fourteen masers under development have been delivered. Four of the masers are SAO VLG11's[3] and one is a Sigma Tau Maser[4]. A SAO maser has been used as one of the master clocks for five years. The Sigma Tau maser was delivered in March of 1989 and is the first of six masers of this type on order. These masers are operated and maintain by USNO personnel with minimal manufacturer support and have demonstrated that masers are reliable and can be operated easily in the field.

Two SAO masers, serial numbers P18 and P19, were delivered to USNO in October of 1983. These masers have been a part of the maser clock systems for five years. Over this period their performance has been outstanding, but recently P19 has demonstrated a frequency jump (the cause is apparently a bad electrical connection inside the vacuum system which will be repaired) and the newly installed receiver has problems common to all SAO masers at USNO. There have been five failures with four of the new receivers.

SAO masers P22 and P23 were delivered to USNO in June of 1987. The initial performance of P22 was poor. When delivered, P22 demonstrated two frequency states and was returned to SAO where a mechanical mounting problem was repaired. When it was returned it worked well for six months but degraded and the two frequency state problem reappeared. In addition, the phase noise of P22 is very poor and might be an indicator of a fundamental problem, perhaps the synthesizer. It should noted that both P22 and P23 are in a poor environment compared to that of P18 and P19.

P23 performed well in the beginning but frequency steps did appear and the retuning of the disassociator cured the problem. P23 is now performing well.

The first of six Sigma Tau Masers was delivered in March of 1989. The stability measurements are outstanding as shown by figure 2 which is an Allan variance plot of the stability between P19 and the Sigma Tau maser. The initial performance is equal to any of the VLG11 masers and perhaps better. In addition there is no measurable drift between the Sigma Tau maser and SID3.

Task 3:

A short term measurement system developed by NRL has been in use at USNO since April of 1987. The short term system is based on a dual mixer time difference sub-system developed by Dr S. R. Stein[5]. This sub-system is marketed by ERBTEC Engineering of Boulder Colorado. In addition, NRL developed the isolation amplifiers (120 dbm of isolation), the direct offset synthesizer, and the software. The system measures up to 24 clocks simultaneously and the software allows for easy comparisons of both short term and long term data. The low measurement noise of two picosecond is below the cesium frequency stability at one second and a hydrogen masers frequency stability at one hour. Identical systems are on loan to Hughes Research Labs and Dr. Carol Alley of the University of Maryland.

The long term system will be completed in early 1990. The layout of the system is seen in figure 3, it is actually two systems in mutual back-up. The long term system is based on the Hydrogen Maser Advanced Clock System (HMACS)[6] developed by NRL which is in operation at the Master Control Station of the Global Positioning System at Falcon Air Force Base. The USNO system has been significantly expanded in both the number of simultaneous measurements (48 instead of 12) and the functions. The long term system will collect frequency measurements from the Stored Ion Devices as well as the 20 second time difference measurements of the cesiums and masers. The control of the steered masers will be through the long term system.

The master clock configuration at USNO will change when the long term system is complete. Currently there are three master clocks. Master clock one is a cesium steered through a microstepper to the time scale. Master clock two is a system of clocks made up of one stored ion device and two masers. A maser steered through its internal synthesizer to the time scale and measured by a stored ion device while the second maser is slaved to the first maser. Master clock three is a maser steered through its internal synthesizer to the time scale.

In the new master clock configuration there will be three master clocks each of which will be a maser steered to a time scale algorithm. Two of the masers will be steered to the same algorithm while the third maser will be steered to an alternate algorithm under test and evaluation. It is anticipated that out of the algorithms under test and evaluation will come an improved USNO time scale algorithm that when combined with the new long term measurement system will optimize the reference clock ensemble made up of cesiums, masers, and stored ion devices.

Task 4:

As mention above, USNO is continuously studying new approaches to ensembling clocks. In addition, the Master Clock Upgrade Program is establishing a time scale test bed that will be integrated into the new measurement systems. Two algorithms will be included in the initial test bed. The first is an algorithm developed for the HMACS at Falcon AFB by Dr. S. R. Stein of Ball Aerospace Corporation[7] which has been modified for use at USNO by Dr. Stein. The features of this modified algorithm include:

1) Extension of the clock model to include white phase noise and random walk frequency aging noise in addition to the white FM and random walk FM included in the HMACS model.

2) Use of a mixed mode Kalman filter which permits the utilization of unequally spaced data and makes real time parameter estimation possible.

3) Improved updating of the covariance after each measurement for the case of clock deweighting.

4) Explicit inclusion of deterministic control terms in the state equation to allow the inclusion of steered clocks.

5) Simple but sensitive frequency step detection.

6) Real time parameter estimation.

The second algorithm to be included in the Master Clock Upgrade Program test bed is a new full Kalman in the process of being procured from Ball Aerospace which has been developed by Dr. Stein. This new Kalman approach includes:

1) Solves the problems of previous Kalman approaches of transients and incorrect weighting.

2) Allows the user to select weighting scheme.

3) Takes into account the correlations between all the relative states of the clocks with respect to the ensemble.

4) Extends the measurement model to include frequency measurements.

Task 5:

An additional task under the Master Clock Upgrade Program is the monitoring of development of the SAO advanced VLG12 maser. The enhancements over the VLG11 maser are:

1) Elimination of all elastomer seals and replacement with metallic seals.

2) Replacement of the 200L/sec, 8-element ion pump with four sorption cartridges and a small ion pump.

3) Addition of a new type of atomic hyperfine state selection system that removes atoms in undesired hyperfine magnetic sublevels from the beam of hydrogen atoms entering the bulb. This is referred to as an adiabatic fast passage (AFP).

The anticipated results are:

1) Short term stability of 1E-16 at 10000 seconds.

2) Reduced drift rate of a factor of 3.

3) Superior vacuum scavenging system.

4) Simplified vacuum maintenance.

5) Improved sensitivity to ambient magnetic field changes.

6) Improved reliability.

Program Goals for the Next Two Years

It is anticipated that in two years there will be twelve hydrogen masers in the time scale at USNO, six SAO's and six Sigma au's. Two of the SAO masers will be used for another purpose. The new measurement and control systems will have been in operation for some time as well as the time scale algorithm test bed in two years. The development of a forth stored ion device is planned as well as the continued procurement of additional hydrogen masers for replacement and upgrading of older units. In addition, NRL has been tasked under the Master Clock Upgrade Program to develop a spread spectrum modem for two-way commercial satellite time transfer and it is expected it will be in operation in two years.

References

[1] Leonard S. Cutler, "A trapped Mercury 199 Ion Frequency Standard." 13th PTTI Applications and Planning Meeting, 1981.

[2] Leonard S. Cutler, Robin P. Giffard, Paul J. Wheeler, and Gernot M. R. Winkler, "Initial operational Experience with a Mercury Ion Storage Frequency Standard." 41st Annual Symposium on Frequency Control.

[3] M. W. Levine, R. F. Vessot, and E. M. Mattison, "Performance Evaluation of the SAO VLG-11 Atomic Hydrogen Masers". 32nd Annual Frequency Control Symposium.

[4] H. Peters, A. Gifford, and J. White, "Hydrogen Maser Research and Development at Sigma Tau Standards Corporation and Tests of Sigma Tau Masers". 17th Annual Precise Time and Time Interval(PTTI), 1985.

[5] S. R. Stein, D. Glaze, J. Levine, J. Gray, D. Hilliard, D. Howe, and L. Erb, "Performance of an automated high accuracy phase measurement system". 36th Annual Symposium on Frequency Control, 1982.

[6] A. Gifford, F. Varnum, "The NRL Hydrogen Maser Ensemble and a GPS Time Steer Experiment". Third International Time Scale Algorithm Symposium, Turin, Italy 1988.

[7] S. R. Stein, "Kalman Ensembling Algorithm: Aiding Sources Approach". Third International Time Scale Algorithm Symposium, Turin, Italy 1988.

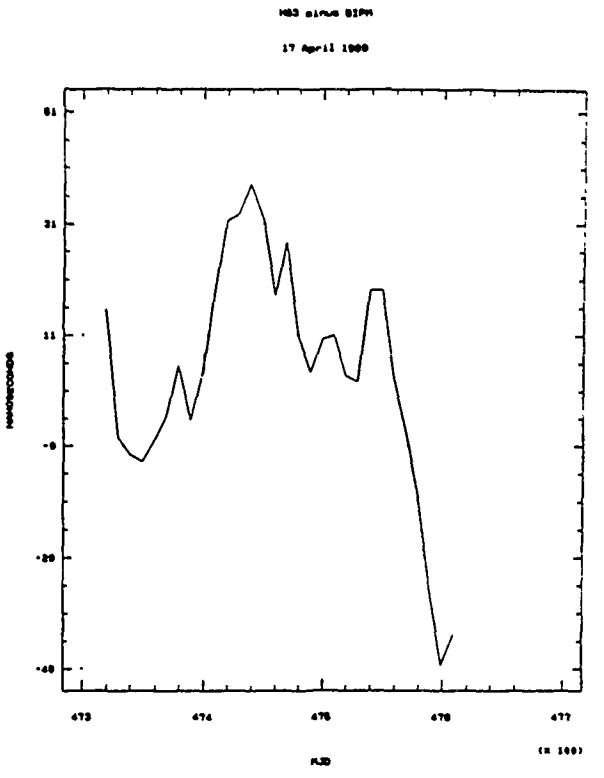


figure 1

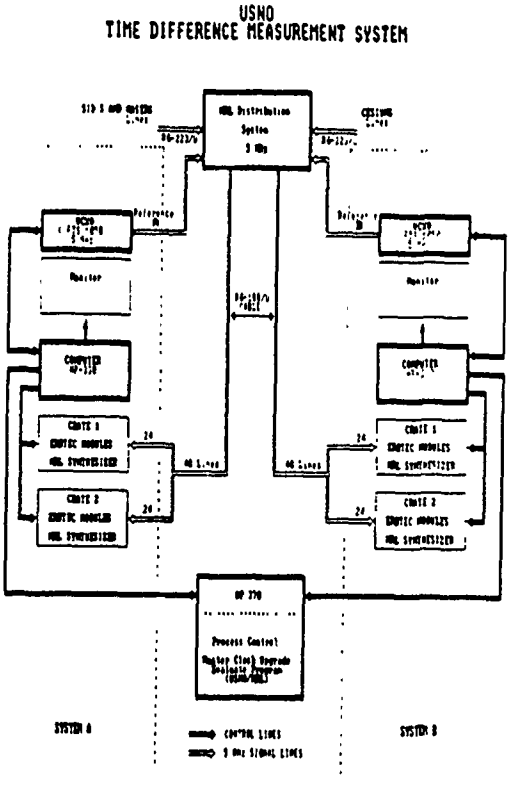


figure 3

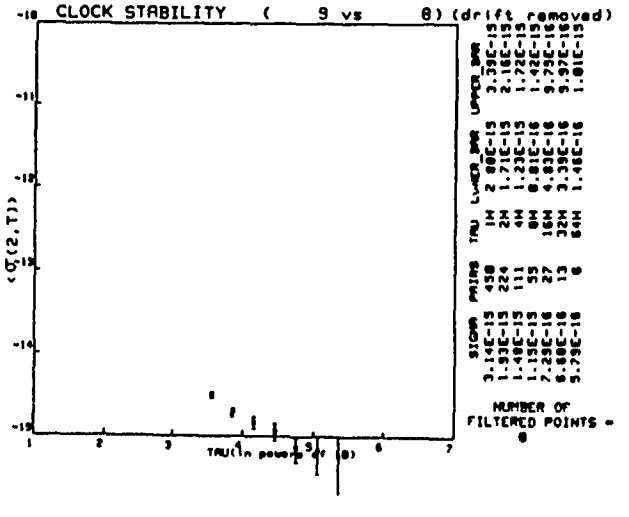


figure 2

43rd Annual Symposium on Frequency Control - 1989

VELOCITY DISTRIBUTIONS FROM THE FOURIER TRANSFORMS OF RAMSEY LINE SHAPES

Jon H. Shirley
Consultant to Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

A computerized method for finding velocity distributions from the Fourier transforms of Ramsey line shapes has been developed. Since atoms in certain velocity groups (those returning to their initial state) do not contribute to the lineshape, a single lineshape transform gives an incomplete picture of the velocity distribution. To bypass this problem we use Ramsey lineshape data taken at different excitation powers so that these velocity groups will contribute. A weighted average of data from three powers gives satisfactory results. The excitation amplitude parameter b is found by minimizing a quality-of-fit criterion. The method is limited to long standards by the assumption that the excitation length ℓ is much less than the drift region length L. However, the addition of first order ℓ/L corrections to the theory make the method usable for shorter standards. The method has been tested with lineshapes theoretically generated from known velocity distributions.

I. Introduction

The advent of optically pumped cesium beams [1] promises to improve the accuracy of primary frequency standards. But knowledge of the velocity distribution must also be improved to evaluate adequately systematic errors, particularly the second-order Doppler effect. Existing methods of obtaining velocity distributions based on pulsed excitation [2,3] or microwave power variation [4] have limitations in the precision of the experimental data. Measurement of Ramsey lineshapes can be done with greater precision, but the published method for extracting velocity distributions from lineshapes [5] is complex in theory and sometimes ambiguous in application.

We have developed an alternate method for finding velocity distributions that takes advantage of the large L/ℓ ratios found in primary standards. The idea for our method appears in a paper by Daams [6]. His analysis applies at one power. The need for lineshapes at different powers to obtain a complete velocity distribution was known by Jarvis [5] and will be further explained here.

II. Description of the Fourier Transform Method for $\ell \ll L$

A. Basic Theory

When the detuning is small enough to reveal only the Ramsey fringe pattern and not the shape of the underlying Rabi pedestal, the transition probability can be represented by [7, Eq. (V.38)]

$$P(\lambda) = \int_0^\infty \rho(T)\sin^2 2b\tau \cos^2 \tfrac{1}{2}\lambda T \, dT. \qquad (1)$$

Here $\tau = \ell/v$ and $T = L/v$ are the transit times for an atom of velocity v across the excitation and drift regions respectively, b is the Rabi frequency proportional to the microwave field strength, and $\lambda = \omega - \omega_0$ is the detuning of the exciting frequency ω from the atomic resonance frequency ω_0. The $\sin^2 2b\tau$ factor represents the Rabi excitation probability for the two excitation regions (without a drift region), while the $\cos^2 \tfrac{1}{2}\lambda T$ factor introduces the interference developed during the drift time by a mistuning of the excitation frequency. The atomic velocity average is represented by the integral over the distribution $\rho(T)$ of drift region transit times T. If we use the half-angle formula to expand the interference factor, we obtain

$$P(\lambda) = \tfrac{1}{4}R(0) + \tfrac{1}{2}R(\lambda), \qquad (2)$$

where

$$R(\lambda) = \int_0^\infty \rho(T) \sin^2 aT \cos \lambda T \, dT \qquad (3)$$

is the Ramsey fringe pattern and R(0) is the Rabi pedestal, constant over the range of detuning considered. We have also introduced $a = 2b\ell/L$ so that $2b\tau$ can be abbreviated by aT. For a distribution $\rho(T)$ of finite width, $R(\lambda)$ tends to zero for large λ.

From (3), $R(\lambda)$ is the Fourier cosine transform of the product of ρ and the transition probability. This product can be interpreted as the distribution of transit times for atoms which would make a transition were there no drift region. The transition probability factor depresses $\rho(T)$ a little where $\sin^2 aT$ is large, and a lot where $\sin^2 aT$ is small (see Fig. 1). The point T_m where the

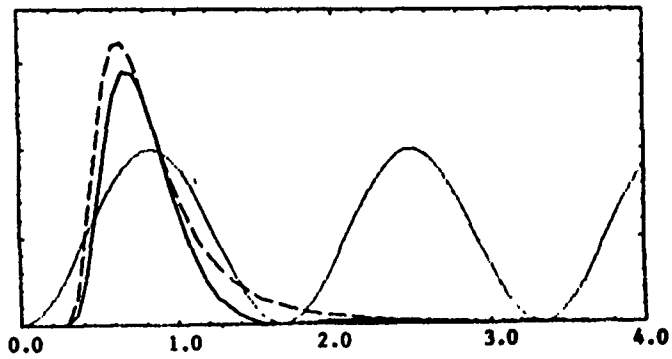


Fig. 1 The full thermal velocity distribution (dashed line), the transition probability factor $\sin^2 aT$ (dotted line), and their product (solid line). The horizontal scale is in units of $\alpha T/L$ where α is the most probable velocity in the source oven [7].

product first goes to zero provides a measure of the parameter a (or b) since $T_m = \pi/a$. Atoms having a transit time of T_m, $2T_m$, $3T_m$... have exactly returned to their initial state, hence do not contribute to the lineshape.

This Fourier transform picture helps explain the power dependence of Ramsey lineshapes. At lower powers the period of $\sin^2 aT$ increases and the product is peaked at higher T values than $\rho(T)$. The transform then oscillates more rapidly making a narrower lineshape. Similarly, at higher powers the product peaks at lower T values leading to a broader lineshape. At powers much above optimum $\sin^2 aT$ may divide $\rho(T)$ into a two-peaked function. The Ramsey fringe will then show the beating of two oscillation frequencies.

If we perform a numerical Fourier transform of a Ramsey fringe and attempt to divide out $\sin^2 aT$ to obtain $\rho(T)$, we encounter two problems. The first is illustrated in Fig. 2. The transform does

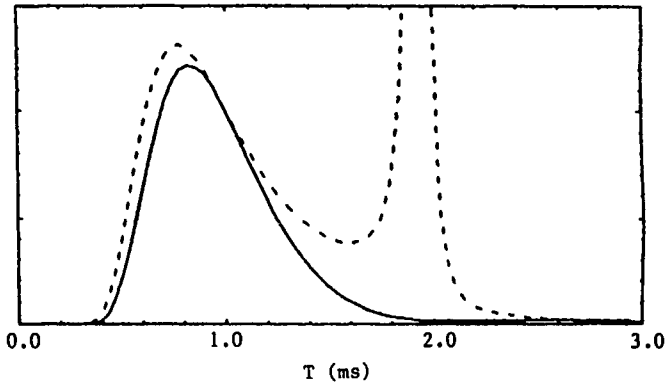


Fig. 2 Fourier transform of a Ramsey lineshape (solid curve) and trial value for $\rho(T)$ (broken curve) obtained by dividing out $\sin^2 aT$ everywhere.

not actually go to zero at T_m, $2T_m$,... Hence when we divide by small numbers we obtain large numbers. The second problem is that a (or T_m) is usually not well known, so $\sin^2 aT$ cannot be accurately evaluated. Both problems are solved in the next section.

B. Use of Multiple Ramsey Lineshapes

The transform of a single Ramsey pattern gives an incomplete picture of $\rho(T)$ since the $\sin^2 aT$ factor suppresses information about $\rho(T)$ near T_m, $2T_m$, $3T_m$.... To complete the picture we use Ramsey fringe data at different levels of microwave excitation power (a or b values). The transforms of this data then have different $\sin^2 aT$ factors corresponding to different regions of $\rho(T)$ being suppressed.

To visualize these regions suppose we reject data for which $\sin^2 aT$ is less than 0.25. Then a single Ramsey fringe pattern at optimum power gives us information about $\rho(T)$ only over those regions in T covered by the black bars in the first row of Fig. 3. The first bar includes the region around $\frac{1}{2}T_m$ where $\rho(T)$ has its maximum. But a broad distribution extends much further than the first bar. To fill in the gaps in the first row we try other power levels. Data taken at a power level 3 dB above optimum (b larger by a factor 1.41) covers the T ranges shown in the second row of Fig. 3. The low T limit is extended. The first gap in row 1 is nicely covered, but the second is only partially

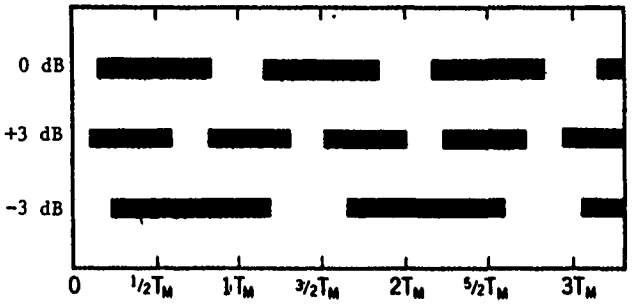


Fig. 3 Regions of T for which $\sin^2 aT > 0.25$ for three different a values. The scale is in units of T_m for optimum power.

covered. Data taken at a power level 3 dB below optimum (b smaller by a factor .71) covers the first two gaps in row 1. Data taken at all three powers gives multiple coverage at most T values. There will always be a gap at small values of T. But $\rho(T)$ vanishes there since small T values correspond to large velocities which are strongly cut off by a Maxwell distribution in the source. Three powers seem adequate to cover most distributions. Other choices of relative powers can also be used.

We can combine the data from several microwave powers as follows. Let the index j denote the power level. From each measured Ramsey fringe pattern $R_j(\lambda)$ we obtain a Fourier transform $F_j(T)$. We define trial distributions by

$$\rho_j(T) = F_j(T)/\sin^2 a_j T. \qquad (4)$$

These trial distributions usually differ from each other. For a best estimate of $\rho(T)$ we form a weighted average of the trial distributions:

$$\bar{\rho}(T) = \Sigma_j \rho_j(T) W_j(T) / \Sigma_j W_j(T). \qquad (5)$$

For a weight function we choose $W_j(T) = \sin^4 a_j T$, the square of the Rabi excitation probability. This weight function strongly suppresses contributions to $\bar{\rho}$ when $\sin^2 aT$ is not near unity. To avoid contributions to $\bar{\rho}$ from unphysical spikes in ρ_j like those shown in Fig. 2, we also set the weight function completely to zero whenever it is less than a cutoff value W_c. This zeroing of W_j implements the regions illustrated in Fig. 3. The $\sin^4 aT$ dependence softens the cutoff edge by making W_j small near the cutoff.

To get a measure of how closely the $\rho_j(T)$ do agree with each other, we introduce a quality-of-fit criterion

$$E(a) = \int_0^\infty D(T)\, dT, \qquad (6)$$

where $D(T) = \Sigma_j [\rho_j(T) - \bar{\rho}(T)]^2 W_j(T)$

is the mean square deviation of ρ_j from $\bar{\rho}$ at each T value, weighted by $W_j(T)$. Our choice of weight function makes E equivalent to the sum of the mean

square errors between the input lineshapes $R_j(\lambda)$ and lineshapes computed from the fitted distribution $\tilde{\rho}(T)$.

The quality-of-fit criterion provides the solution to our second problem, the unknown a_j. If the ratios of powers are known, we can assume the ratios of a_j's are known. Then only one absolute value, say a_1, need be determined. If we try different values of a_1, or $T_m = \pi/a_1$, adjust the other a_j's in their proper ratio, and compute E, we find the dependence shown in Fig. 4. The minimum is more than two orders of magnitude deep and defines T_m within 0.5%.

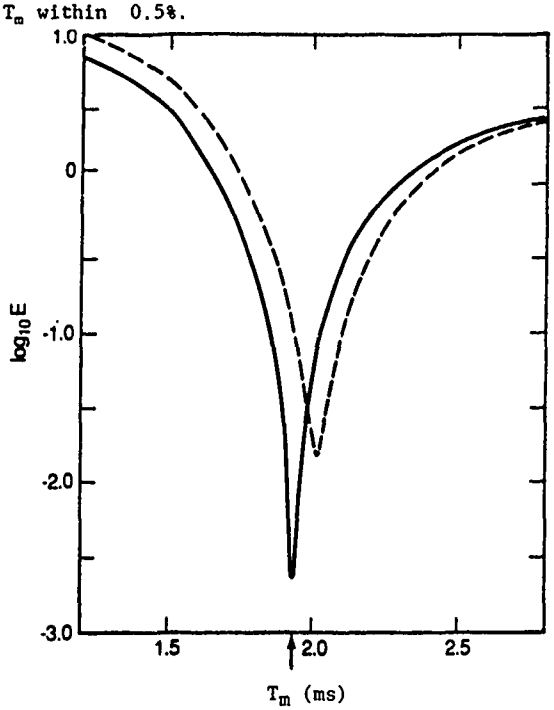


Fig. 4 Variation of E with trial values of T_m for theoretical lineshape data. The broken curve is without the ℓ/L correction and the solid curve is with the correction. The arrow points to the theoretical T_m value.

C. Summary of Method

To use our method we first obtain Ramsey fringe data $R_j(\lambda)$ at three or more power levels of known ratio ranging both above and below optimum power. Such a data set is shown in Fig. 5 for the full thermal distribution emerging from an oven. Numerical Fourier cosine transforms $F_j(T)$ are computed at each power level (Fig. 6). From the position of the first minimum in Fig. 6b a trial value for T_m is obtained. From it we obtain trial values for a_j and $W_j(T)$. We then evaluate $\tilde{\rho}$ and E according to Eqs (4), (5), and (6). The value of T_m is then varied and the computations repeated until a minimum in E is obtained (Fig. 4). Finally the distribution is normalized so that

$$\int_0^\infty \tilde{\rho}(T)\, dT = 1.$$

The fitted distribution $\tilde{\rho}$ thus obtained from the data in Fig. 5 is shown in Fig. 7, along with the actual distribution used to generate the data. The agreement is good except for a systematic shift of $\tilde{\rho}$ to higher T values. The minimum in Fig. 4 is also too high. Varying W_c from 0.025 to 0.250 produces little change in $\tilde{\rho}$, so the fit is insensitive to this cutoff.

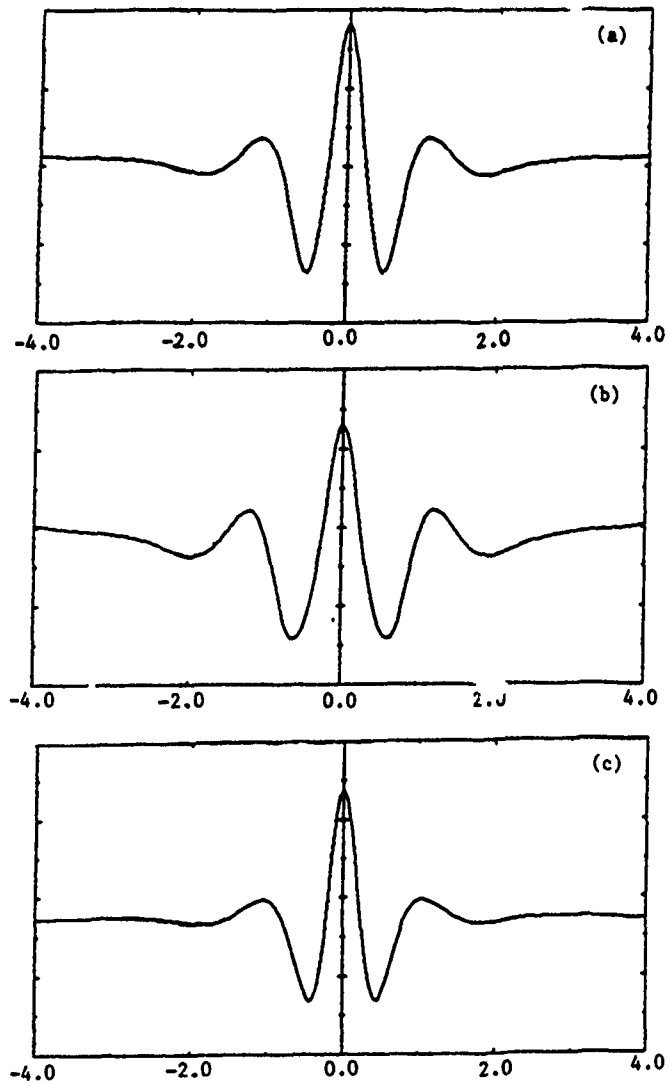


Fig. 5 Ramsey fringe lineshapes for full thermal distribution. (a) Optimum power. (b) 3 dB above optimum power. (c) 3 dB below optimum power. The horizontal scales are λ in kHz. L = 25 cm, ℓ = 1 cm and α = 251.3 m/sec.

III. First Order ℓ/L Corrections

Tests on data for other ℓ/L ratios show that the shifts in both Fig. 4 and Fig. 7 are proportional to ℓ/L. Daams's analysis [6] included first-order corrections in ℓ/L. We have done a similar analysis and find the following expression to replace (4):

$$\rho_j(T) = \frac{F_j(T) + b_j^{-1}\, G(b_j r)\, dF_j/dT}{\sin^2 a_j T[1-(\ell/L)G'(b_j r)]} \qquad (7)$$

To first order the numerator is equivalent to $F_j(T + G/b_j)$ explaining the shifts in Fig. 7. The function G depends on the microwave field profile seen by atoms as they traverse the excitation region. For a rectangular excitation field profile, found in most standards with transverse C-fields, we have

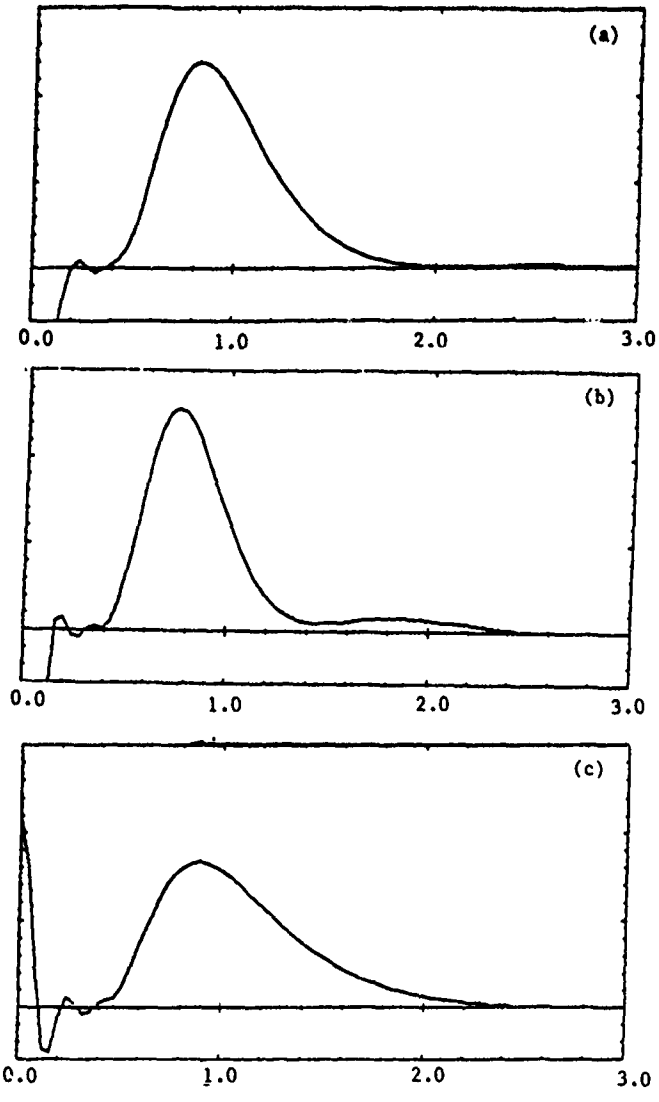


Fig. 6 Fourier transforms of the lineshapes in Fig. 5. (a) Optimum power. (b) 3 dB above optimum power. (c) 3 dB below optimum power. The horizontal scales are T in ms. The small T behavior arises from the Rabi pedestal.

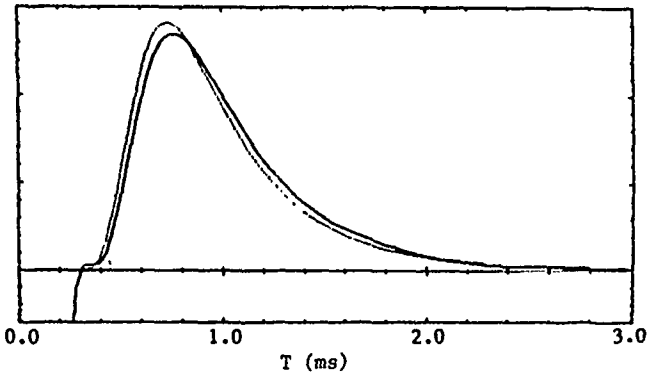


Fig. 7 Comparison of fitted distribution $\tilde{p}(T)$ (solid curve) with input thermal distribution (dotted curve) for $\ell/L = 0.04$ and no ℓ/L correction.

$$G_r(br) = \tan br.$$

For a half-sine-wave field profile, found in most standards with longitudinal C-fields, we have

$$G_s(br) = br \sec br\, J_0(br).$$

G' is the derivative of G with respect to its argument br. For $T \leq \frac{1}{4}T_m$ we have $G/b \approx r$ and $G' \approx 1$ for either profile.

When the first order ℓ/L corrections in (7) were included in the computer program, we obtained the fit shown in Fig. 8. The shift has disappeared and only small discrepancies remain. The quality-of-fit criterion, shown by the solid curve in Fig. 4, has a minimum smaller by a factor 8 and only 0.3% from the correct value of T_m.

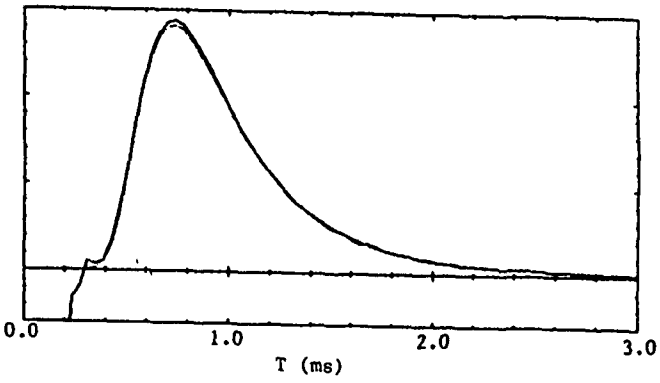


Fig. 8 Comparison of fitted distribution $\tilde{p}(T)$ (dotted curve) with input thermal distribution (solid curve) for $\ell/L = 0.04$ with ℓ/L correction.

IV. Test on Experimental Data

In Fig. 9 we show a set of experimental Ramsey curves for the primary frequency standard NBS-4 [8]. The wings of the curves show the shape of the underlying Rabi pedestal. The first two side lobes on each side are nearly the same height, while the next ones are weaker. Irregularities in the periodicity of the side lobes are also apparent in all three lineshapes. Such features hold information about the velocity distribution which a Fourier transform can reveal.

Figure 10 shows the transit-time distribution obtained by our method from the lineshapes in Fig. 9. The distribution is separated into two well-resolved parts. Each part produces a Ramsey fringe with its own oscillation frequency for the side lobes. The beating of these two frequencies accounts for most of the unusual features seen in Fig. 9, especially the phase change in the third side lobe (first side lobe at the +3 dB power level). The narrow dip separating the parts in Fig. 10 is responsible for the large number of weak side lobes. This dip is probably caused by a narrow obstruction in the beam tube which physically blocks a class of atomic trajectories spanning a limited velocity range. In Fig. 11 the transit-time distribution has been converted to a velocity distribution by the relation

$$\sigma(v) = (L/v^2)\rho(L/v).$$

This distribution is very similar to one obtained previously [8, Fig. 16] by a pulse method [3].

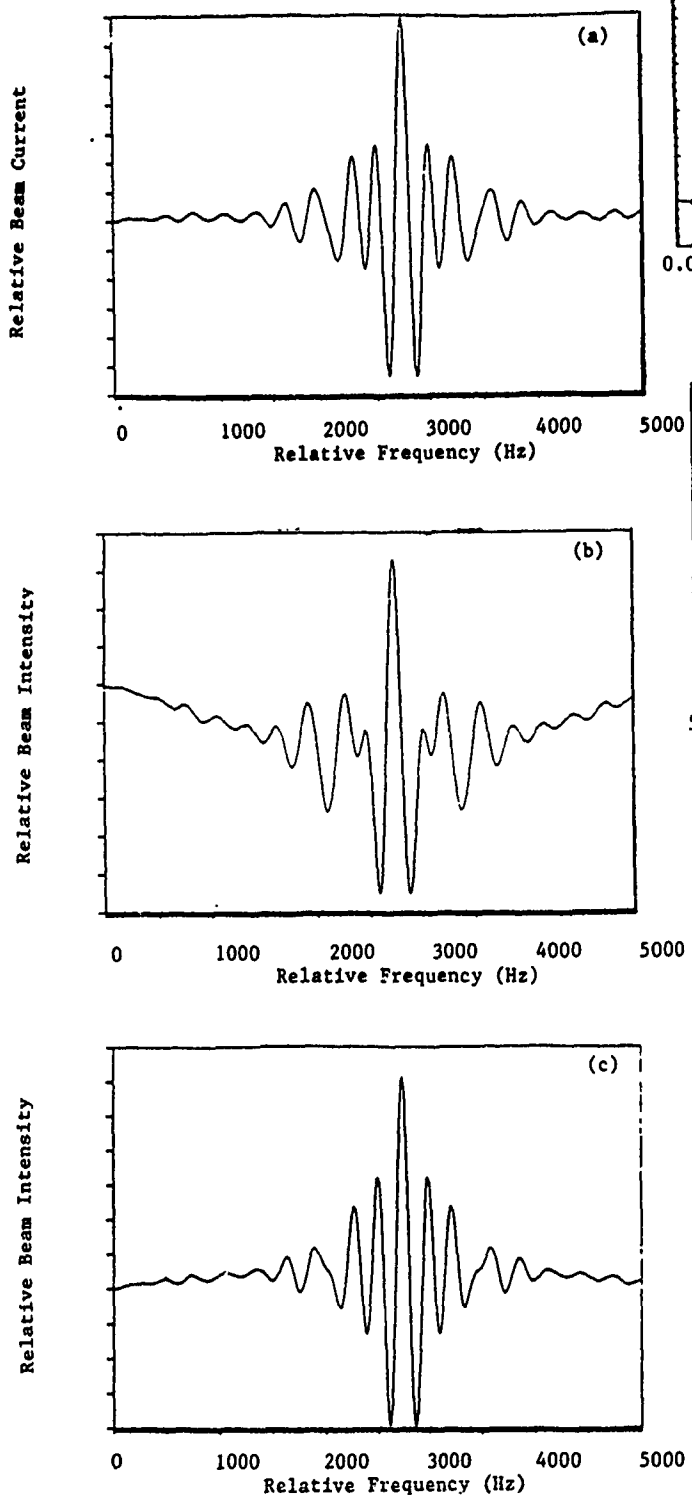


Fig. 9 Ramsey fringe lineshapes for the primary frequency standard NBS-4. (a) Optimum power. (b) 3dB above optimum power. (c) 3 dB below optimum power.

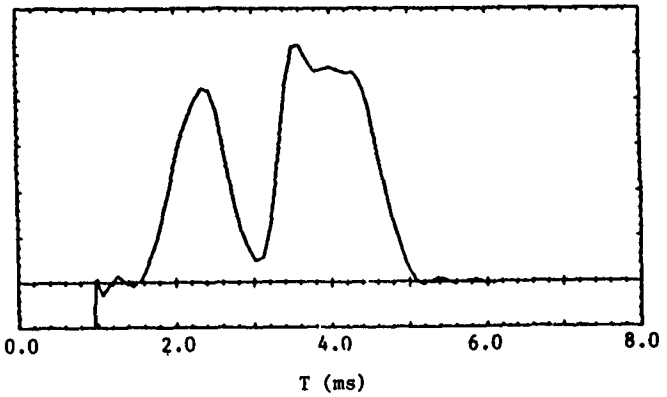


Fig. 10 Fitted transit-time distribution for the primary frequency standard NBS-4.

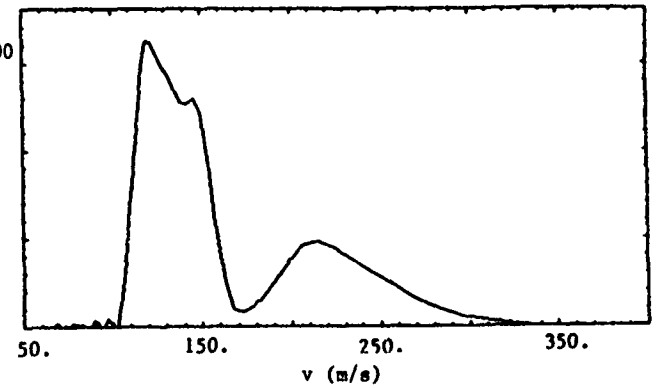


Fig. 11 Fitted velocity distribution for the primary frequency standard NBS-4.

V. Summary

We have developed a method that starts with the Fourier cosine transforms of Ramsey lineshapes and extracts from them a transit-time distribution $\rho(T)$. The method includes first order ℓ/L corrections and a quality-of-fit criterion that enables the power parameter b to be accurately determined. By testing the method on theoretically generated data we have shown it capable of finding distributions accurate within a few percent for $\ell/L = 0.04$. Analysis of theoretical lineshapes with $\ell/L = 0.01$ have produced fits that agree with the input distribution to 1% for a narrow, two-peaked distribution, and 0.1% for the broad thermal distribution. Hence our method can provide velocity distributions with accuracy sufficient to meet the anticipated needs of optically pumped primary standards.

References

[1] R. E. Drullinger, "Frequency Standards Based on Optically Pumped Cesium," Proc. IEEE 74, 140 (1986) and references therein.

[2] A. De Marchi, G. D. Rovera, and A. Premoli, "Pulling by Neighbouring Transitions and Its Effects on the Performance of Caesium-Beam Frequency Standards," Metrologia 20, 530 (1984).

[3] H. Hellwig, S. Jarvis, D. Halford, and H. E. Bell, "Evaluation and Operation of Atomic Beam Tube Frequency Standards Using Time

Domain Velocity Selection Modulation," Metrologia 9, 107 (1973).

[4] J.-S. Boulanger, "A New Method for the Determination of Velocity Distributions in Cesium Beam Clocks," Metrologia 23, 37 (1986).

[5] S. Jarvis, "Determination of Velocity Distributions in Molecular Beam Frequency Standards from Measured Resonance Curves," Metrologia 10, 87 (1974).

[6] H. Daams, "Corrections for Second-Order Doppler Shift and Cavity Phase Error in Cesium Atomic Beam Frequency Standards," IEEE Trans. Inst. Meas. IM-23, 509 (1974).

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

[8] D. J. Glaze, H. Hellwig, D. W. Allan, S. Jarvis, Jr., and A. E. Wainwright, "Accuracy Evaluation and Stability of the NBS Primary Frequency Standards," IEEE Trans. Inst. Meas. IM-23, 489 (1974).

43rd Annual Symposium on Frequency Control - 1989

ATOMIC TRAJECTORIES IN COMPACT CESIUM BEAM CLOCKS

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

Abstract-- Atoms travelling within conventional cesium beam tubes are deflected by inhomogeneous magnetic fields before and after interaction with the microwaves. The atomic trajectory deflections within the magnets impact clock performance by determining both the velocity distribution and the intensity of the detected atomic beam. By performing an accurate computation of atomic trajectories, we have calculated velocity distributions in beam tubes using the usual,"two-wire"-field magnets, and explored the role of oven and detector offsets, and magnetic field strength. We have also studied trajectories in a novel beam tube configuration using two orthogonal one-dimensional focussing dipoles as state selectors.

I. INTRODUCTION

Atoms travelling within a conventional cesium beam tube (CBT) are deflected by inhomogeneous magnetic fields before and after interacting with the microwaves. The first ("A") magnet selects the hyperfine atomic state which will enter the Ramsey cavity, while the second ("B") one steers atoms in the desired hyperfine state towards the detector. The atomic trajectory deflections within the magnets impact the performance of the clock in several ways. Firstly, the intensity of the atomic signal at the detector will depend strongly on how atoms are deflected by the magnets; this will in turn determine the signal-to-noise ratio at the CBT output, and thus the stability of the clock in the white-noise regime [1]. Conventional compact CBTs use the dispersive "two-wire" field configuration [2] for both magnets, but other combinations of magnetic state-selectors have been proposed [3]-[6], which in general incorporate two-dimensional (hexapole or quadrupole) or one-dimensional (dipole) focussing in an attempt to increase the detected beam signal to improve the clock stability in the white-noise regime. Also, since the deflections depend on atomic velocity, the magnets will, to a large extent, determine the velocity distribution of the detected atoms. This, in turn, will affect the clock's performance in several ways. The atomic velocity distribution determines the line-shape of the hyperfine transitions neighboring the clock transition, $(F,M)=(4,0) \leftrightarrow (3,0)$, and thus, through the phenomenon of Rabi-pulling [7], may impact the clock's long-term frequency stability. The spread in atomic arrival times at the detector will also be determined by the atomic velocity distribution, and, together with the detector response time, will affect the stability of the clock in the white-noise regime by smearing the information impressed on the atomic beam within the Ramsey cavity by the microwave modulation scheme [8],[9].

In this paper we analyze the problem of magnetic deflection of cesium atoms in compact CBTs . Atomic trajectories are computed accurately, using analytical expressions for the magnetic fields and field gradients; the effective magnetic moments of the atoms within the magnet are determined by the local field and allowed to vary as the atom moves; the transverse force on the atoms is calculated exactly, without making the usual constant-force approximation, and the atoms are allowed to fill the full magnet gap, modelling a commercial CBT in a realistic fashion. This analysis is carried out for CBTs having two conventional "two-wire" field magnets, as well as for a variation of the configuration proposed by Kartaschoff [4] in which both magnets are one-dimensional focussing dipoles. In both cases velocity distributions and detected beam intensities have been computed.

II. MAGNETIC DEFLECTION OF CESIUM ATOMS

The force acting on an atom having magnetic moment $\vec{\mu}$ immersed in an inhomogeneous magnetic field $\vec{H}$, is given by

$$\vec{F} = \text{grad}\,(\vec{\mu}\cdot\vec{H}) = \mu_{eff}\,\text{grad}\,(H) \qquad (1)$$

where μ_{eff} (effective magnetic moment) is the component of the magnetic moment along the direction of the field. Breit-Rabi's formula [10] gives μ_{eff}:

$$\mu_{eff} = \pm \frac{\varepsilon + 2M/(2I+1)}{[1 + 4M\varepsilon/(2I+1) + \varepsilon^2]^{1/2}}\,\mu_o$$

where I is the cesium nuclear spin, I=7/2; the (+) sign corrresponds to F=3 and (-) to F=4; ε is proportional to the ratio of magnetic-to-hyperfine energies, and for cesium is given by $\varepsilon = H/3268$, with H in gauss. μ_o is Bohr's magneton, and M the azimuthal quantum number. Inspection of Eq. (1) shows that atoms having $\mu_{eff} > 0$ will be deflected towards more intense fields, while atoms having $\mu_{eff} < 0$ will be deflected towards weaker fields. For H > 2451 gauss, all the F=3 atoms, as well as the F=4, M=-4 ones, have positive magnetic moments; atoms with F=4 and M between -3 and 4 have negative magnetic moments.

For the field configurations we will explore, $\text{grad}_y(H) \ll \text{grad}_x(H)$ and $\text{grad}_z(H) = 0$, where z is taken parallel to the atomic beam and y parallel to the ionizer wire. The equations of motion for the cesium atoms within the magnets are then

$$\ddot{x} = (\mu_{eff}/m)\,\text{grad}_x(H) \qquad (3a)$$

$$\ddot{y} = 0,\ \ddot{z} = 0 \qquad (3b,c)$$

where m is the mass of the cesium atom. Equations (3b,c) are trivial; Eq. (3a) describes the deflection of the cesium atoms by the state-selecting magnets. Since μ_{eff} depends on the field strength, which is a function of position, and the gradient of the field strength is also a function of position, the acceleration of the cesium atom will vary as the atom travels through the field. In many applications, by making the magnet gap substantially larger than the width of the atomic beam and the magnitude of the atomic deflections, it is possible to neglect that variation and solve Eq. (3a) for constant atomic acceleration [11]. But for compact CBTs, the magnet must be kept as small as possible, and then the constant-acceleration approximation is not valid. In that case, Eq. (3a) must be integrated numerically.

Figure 1 illustrates our trajectory calculations. An atom in the (F,M) hyperfine state

CH2690-6/89/0000-168 $1.00 C 1989 IEEE

leaves the oven slit from point (x_0, y_0) with velocity V. We assume the sixteen (F,M) sublevels of the cesium ground state to be equally populated, and the velocity distribution to be beam-maxwellian. The angular distribution is assumed to be proportional to $\cos^2 \theta$ where θ is the angle between V and the beam axis ($\theta \ll 1$). We solve the atomic equations of motion in each one of regions I-V; the position and velocity at the end-plane of each region are the initial conditions for the next one. In regions I, III and V the atomic motion is uniform; in regions II and IV, we integrate Eq. (3a) numerically. An atom is detected if it crosses the end plane of the CBT within the appropriate range of x. If the atom starts in the (3,0) state, in region III it is transferred to the (4,0) state; if it starts in the (4,0) state, it is transferred to the (3,0) one. Otherwise, it stays in its initial hyperfine state. This simulates the behavior of a perfect Ramsey cavity injected with resonant radiation at the clock transition frequency.

The velocity distribution of "signal" atoms is calculated by integrating over the set of initial coordinates and angles which lead to detection of an atom in the final state of the clock transition having a given speed V; integrating over the velocity distribution gives the detected signal strength. If atoms in other states are detected, they will contribute to the CBT background. We have developed a computer program which implements these calculations for arbitrary magnetic fields (prescriptions to calculate the field and its gradient have to be given).

III. ATOMIC VELOCITY DISTRIBUTIONS IN CONVENTIONAL CBTS

Atomic velocity distribution measurements can be performed by a pulsed-microwave technique, as shown, per example, by Howe [12]. These measurements show that the velocity distributions vary significantly from one CBT to anotner. Figure 2 shows two examples [12], one in which the velocity distribution is approximately symmetric about its peak, with the high-speed side dropping as steeply as the low-speed side, and another one with a pronounced high-speed tail. As it is well known, the shape of the Rabi pedestals of the microwave transitions occurring within the Ramsey cavity depends strongly on the shape of the velocity distribution [13]. Narrow velocity distributions introduce oscillations in the tails of the Rabi pedestals, which are smoothed over as the distribution is broadened. We have performed preliminary calculations which show that the observed variability in the CBT atomic velocity distributions is enough to introduce significant differences in the shape of the Rabi pedestals of the CBT microwave transitions; measurements of the microwave spectrum of the CBTs have confirmed this result. These differences may in turn lead to significantly different long-term frequency stabilities for the different CBTs through the phenomenon of Rabi-pulling [7].

We have used the computer code described in the preceding section to calculate the velocity distributions transmitted by the "A" and "B" magnets of compact CBTs. Conventional, "two-wire" field magnets are assumed. The magnet geometry is shown in Fig. 3; the field strength and its x-component of gradient are given by [14]

$$H = \frac{2a^2 H_o}{r_1 r_2} \qquad (4)$$

and

$$\mathrm{grad}_x(H) = -\frac{2a^2 H_o}{(r_1 r_2)^3}(r_1^2 + r_2^2)x \qquad (5)$$

where 2a is the distance between the "equivalent wires", H_o is the field strength at the convex poletip and r_1, r_2 are the distances from point (x,y) to the "equivalent wires" [14]. Realistic values were assigned to the geometric parameters of the CBT and the cesium oven temperature, and the effects of small variations in oven and detector offsets (δ_S and δ_D) and magnetic field strength on the velocity distribution of the detected cesium atoms were studied. Figures 4 and 5 show the effects of changes in δ_S and δ_D. These results show that the shape of the atomic velocity distribution (in particular the presence of high-velocity tails) is highly sensitive to the positioning of oven and detector. Figure 6 shows the effect of changes in magnetization. Clearly, the shape of the velocity distribution is quite insensitive to changes in magnetic field strength. As the field increases from 8 to 12 kilogauss (a quite significant variation), the distribution broadens somewhat, and the peaks shift towards higher speeds, but the rough shape of the distribution does not change significantly. This analysis shows that the observed variations in the shape of the velocity distribution from CBT to CBT can be understood in terms of small differences in detector or oven positioning. In our configuration the results are more sensitive to changes in the position of the oven, and differences of up to 0.3 mm suffice to explain the full range of observed variability.

IV. ONE-DIMENSIONAL FOCUSSING OPTICS

Kartaschoff [4] described a new design for a CBT using focussing, instead of dispersing, optics. He suggested the use of an hexapole magnet for the "A" magnet, providing two-dimensional focussing, and a novel one-dimensional focussing dipole for the "B" magnet, shown on Fig. 7. The polepieces are shaped like hyperbolic cylinders, giving hyperbolic- cylinder magnetic equipotential surfaces, while the field lines lie on the set of orthogonal elliptical-cylinder surfaces. The magnetic field strength and its x-component of gradient are given by

$$H = \frac{(d^2 - y_o^2)^{1/2} H_o}{(r_1 r_2)^{1/2}} \qquad (6)$$

and

$$\mathrm{grad}_x(H) = -\frac{(d^2 - y_o^2)^{1/2} H_o}{2(r_1 r_2)^{5/2}}(r_1^2 + r_2^2) \qquad (7)$$

where the polepieces have foci at x=0, y=±d and poletips at x=0, y=±y_0; r_1, r_2 are the distances from point (x,y) to the foci, and H_o is the poletip field. The field strength as a function of position is shown in Fig. 8, and it can be seen that in this field, atoms with $\mu_{eff} > 0$ will be deflected towards the central plane, i.e., will be focussed in one dimension, while atoms with $\mu_{eff} < 0$ will be deflected outwards.

In order to obtain a CBT configuration with flop-in optics, the second magnet must focus atoms with $\mu_{eff} < 0$. An hexapole magnet will indeed

accomplish that, but it presents some disadvantages: it is heavier and harder to manufacture than a dipole, and, since the on-axis field is zero, it is not that well suited to select between atoms with M=0. Instead, we used our computer code to explore the characteristics of a novel CBT configuration using two orthogonal one-dimensional focussing dipoles.

If the previously described magnet is turned on its side, so that the foci are at x=±d, y=0 and the poletips at x=±x_0, y=0, as shown in Fig. 9, the field strength and its x-component of gradient are given by

$$H = \frac{(d^2-x_o^2)^{1/2}H_o}{(r_1r_2)^{1/2}} \qquad (8)$$

and

$$\text{grad}_x(H) = -\frac{(d^2-x_o^2)^{1/2}H_o}{2(r_1r_2)^{5/2}}(r_1^2 + r_2^2-4d^2) \qquad (9)$$

The field strength as a function of position is shown on Fig. 10; those atoms which have $\mu_{eff} > 0$ will now be deflected outwards, while those having $\mu_{eff} < 0$ will be deflected towards the central plane. A combination of magnets in both orientations can be used to provide one-dimensional focussing flop-in optics for a CBT. Both magnets will introduce some defocussing in the y-direction, but for reasonably long slits and ionizer wire, the losses due to such defocussing can be kept low. The defocussing will be more serious for the "horizontal" magnet. A beam stop must be incorporated on the central plane, since on that plane $\text{grad}_x(H)=0$ for both magnets, and no state selection takes place; additional losses will be introduced by that stop.

We have calculated the velocity distributions and beam intensities expected with the arrangement described above for some combinations of field strengths and poletip separations. The overall geometry is similar to the one used with the dispersive magnets, in order to be able to make a fair comparison between both configurations. Of course, the oven and detector offsets, δ_S and δ_D, are both null. A beam stop of the same width as the detector is positioned on axis at the entrance to the "B" magnet. Figure 11 shows the velocity distribution obtained for a relatively low gradient-to-field ratio. The detected beam intensity is about the same as for the conventional CBT using "two-wire" magnets, but the detected atoms are much slower; the most probable velocity is two or three times lower than in a conventional CBT. In the white-noise regime, the two-sample standard deviation of fractional frequency fluctuation is related to the atomic lineshape quality factor Q and detected signal-to-noise ratio S/N by [1]

$$\sigma_y(\tau) \simeq \frac{K}{Q(S/N)}\frac{1}{\tau 1/2} \qquad (10)$$

where K depends on the microwave modulation scheme. Since Q is inversely proportional to the average atomic speed [1], this configuration could reduce $\sigma_y(\tau)$ by a factor 1/2 or 1/3.

Figure 12 shows the velocity distribution obtained for a higher gradient-to-field ratio. In this case, the detected intensity is about ten times higher than for the conventional CBT, while the mean velocity is about the same. If the detected signal is shot-noise limited, we can expect S/N to be about three times larger than in the conventional case, and $\sigma_y(\tau)$ will consequently be reduced by a factor 1/3.

Further optimization of the operating parameters may yield very significant improvements in clock performance over that achieved using conventional, "two-wire" magnets.

V. CONCLUSIONS

We have performed accurate calculations of atomic trajectories in conventional CBTs using dispersive, "two-wire" field magnets and also in a novel configuration using two orthogonal one-dimensional focussing dipoles of the type described by Kartaschoff [4]. In the first configuration, we were able to explain observed differences in atomic velocity distribution in terms of small variations in oven and/or detector offsets. The second configuration yields beam intensities which are significantly higher than the ones obtained with dispersive magnets and/or substantially lower atomic velocities, and thus could improve the frequency stability of compact cesium clocks in the white noise regime. Our design differs from Kartaschoff's prescription in that a one-dimensional focussing dipole, rather than a hexapole or quadrupole magnet, is used as the "A" magnet in the CBT, making the tube manufacturing simpler.

ACKNOWLEDGEMENTS

We would like to thank Dr. R. P. Frueholz for many stimulating discussions. This work was supported in part by the U.S. Air Force Space Division under contract No. F04701-86-CC-0087

REFERENCES

[1] C. Audoin and J. Vanier, "Atomic Frequency Standards and Clocks", J. Phys. E, vol. 9, pp. 697-720, 1976.

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

[3] P. Kartaschoff and P. E. Debely, "Resonateur a Cesium de Concepcion Nouvelle", in Proc. Colloque Internatl. de Chronometrie, Paris, 1969, p. A3-1.

[4] P. Kartaschoff, "Atomic Beam Resonator Having a Confocal Conics Field Geometry in the Second State Selector", U. S. Patent No. 3824394, July 16, 1974.

[5] D. Emmons and P. J. Rogers, "New Cesium Beam Tube Utilizing Hexapole/ Double-Diode Optics", in Proc. 33rd. Ann. Freq. Control Symp., Washington: EIA Press, 1979, pp. 490-493.

[6] C. Xue, "Newly Designed Dipole-Beam Optics System for Cesium-Beam Frequency Standards", IEEE Trans. IM, vol 29, pp. 301-304, 1980.

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

[8] C. Audoin, V. Candelier and J. Vanier, "Effect of the Atom Transit Time on the Frequency Stability of Cesium Beam Frequency Standards", in Proc. 40th. Freq. Control Symp., New York: IEEE Press, 1986, pp. 432-440.

[9] B. Jaduszliwer, "Atomic Transit Time and Delayed Ionization Effects on Cesium Beam Frequency Standards", IEEE Trans. UFFC, vol 34, pp. 690-694, 1987.

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

[11] N. Ramsey, Molecular Beams, New York: Oxford University Press, 1956, ch. IV, pp. 90-92.

[12] D. A. Howe, "Velocity Distribution Measurements of Cesium Beam Tubes," in Proc. 30th Ann. Freq. Control Symp., Washington: EIA Press, 1975, pp. 451-456.

[13] N. Ramsey, op. cit., ch. V, pp. 118-134.

[14] N. Ramsey, op. cit., ch. XIV, pp. 397-400.

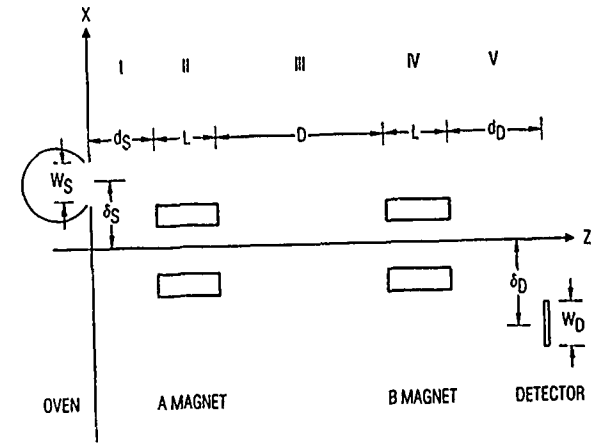


Fig. 1 Geometry of a compact CBT, used in our trajectory calculations. The oven slit (width W_S) is centered at δ_S; the detector wire (width W_D) is centered at δ_D. Both magnets have length L and they are separated by distance D. The distance between oven and "A" magnet is d_S, and between detector and "B" magnet, d_D. The trajectory is calculated separately in each one of the regions I-V.

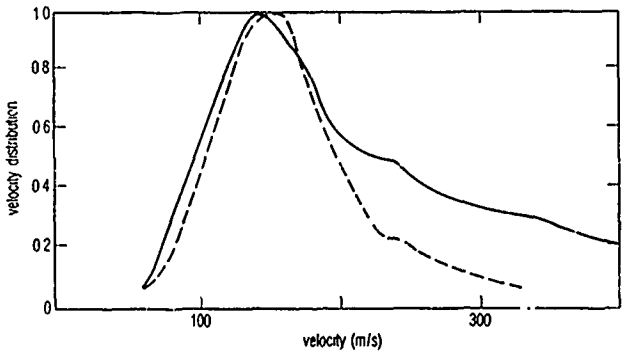


Fig. 2 Atomic velocity distributions measured for two different commercial CBTs [12].

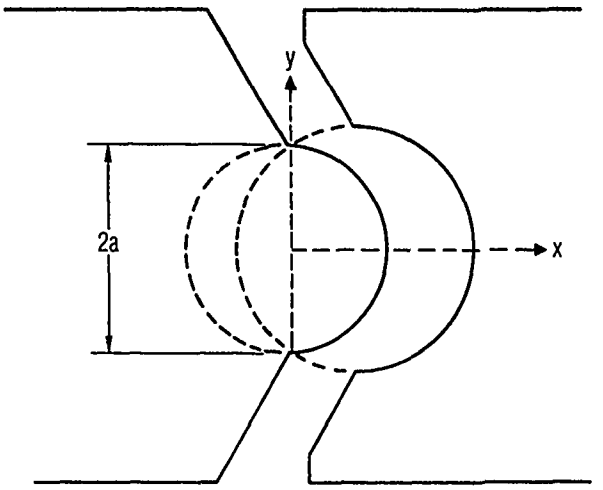


Fig. 3 Cross section of the "two-wire" field magnet. The distance between the "equivalent wires" is 2a.

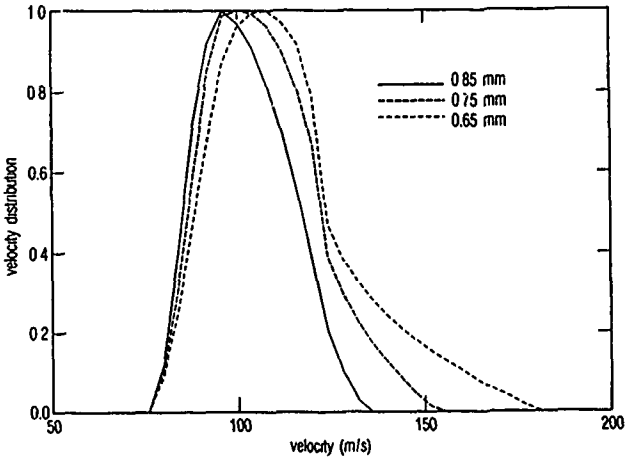


Fig. 4 Atomic velocity distributions calculated for a CBT using "two-wire" field magnets, for three different values of oven offset.

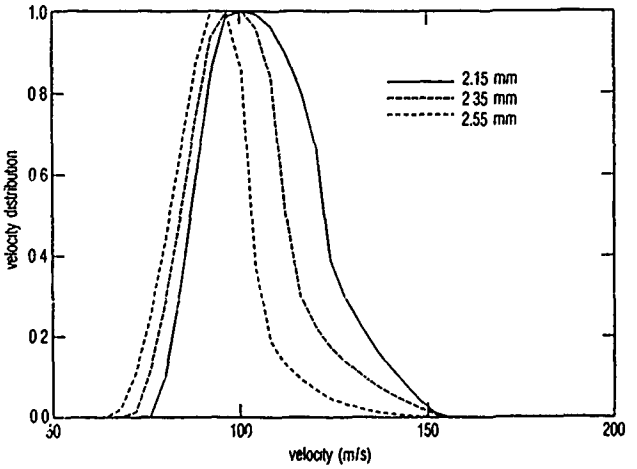


Fig. 5 Atomic velocity distributions calculated for a CBT using "two-wire" field magnets, for three different values of detector offset.

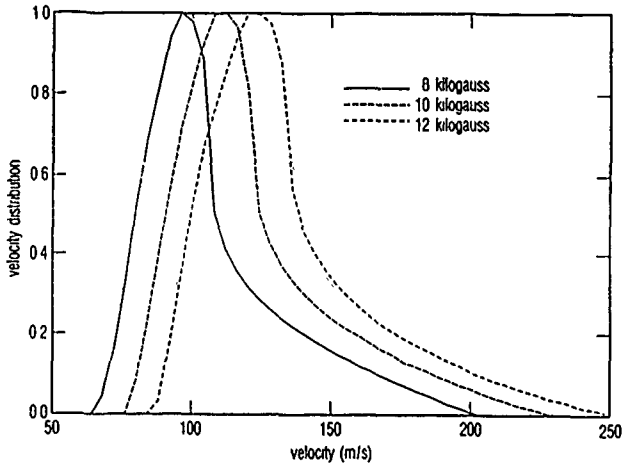


Fig. 6 Atomic velocity distributions calculated for a CBT using "two-wire" field magnets, for three different values of poletip magnetic field.

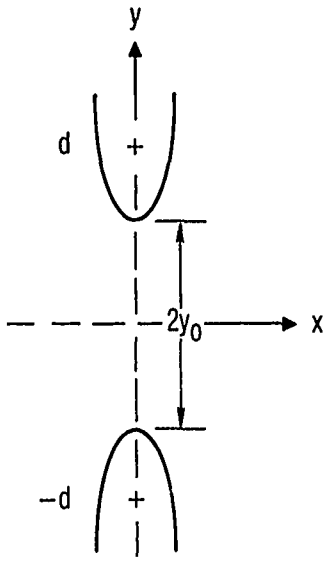


Fig. 7 Cross section of the one-dimensional focussing dipole magnet propossed by Kartaschoff [4]. The polepieces are shaped like hyperbolic cylinders, with foci at x=0, y=±d

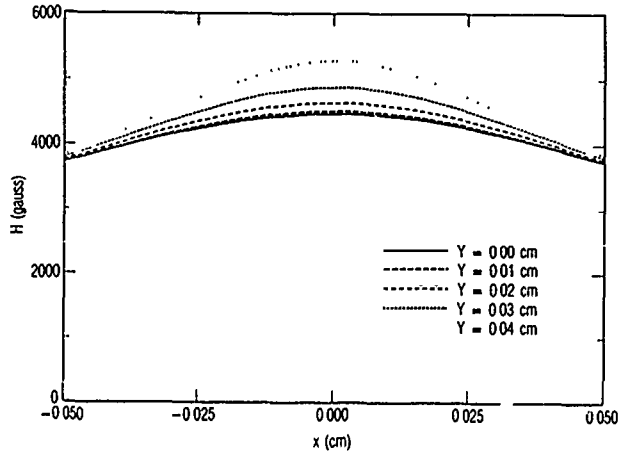


Fig. 8 Magnitude of the magnetic field in the magnet shown in Fig. 7.

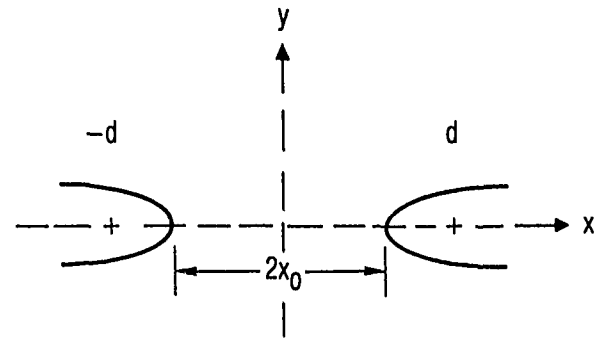


Fig. 9 Same as in Fig. 7, turned by 90°. The foci of the hyperbolic cylinder polepieces are at x=±d, y=0.

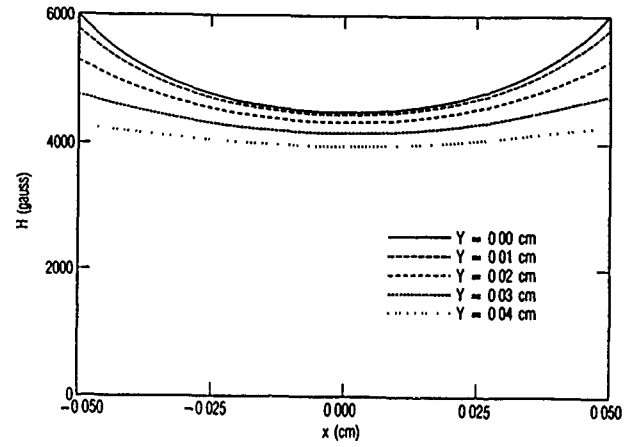


Fig. 10 Magnitude of the magnetic field in the magnet shown in Fig. 9.

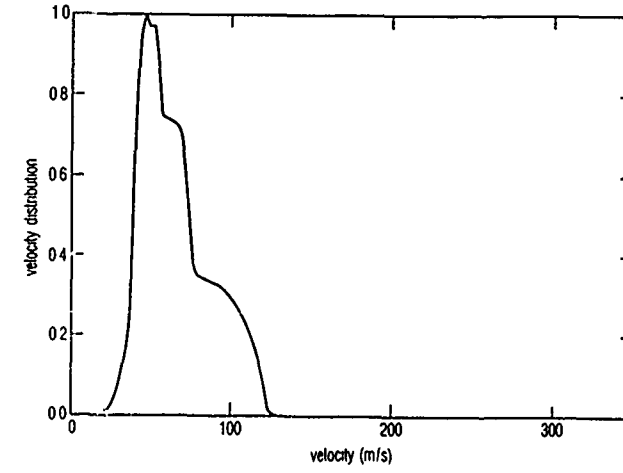


Fig. 11 Atomic velocity distribution calculated for a CBT using two orthogonal one-dimensional focussing dipoles. Low gradient-to-field ratio.

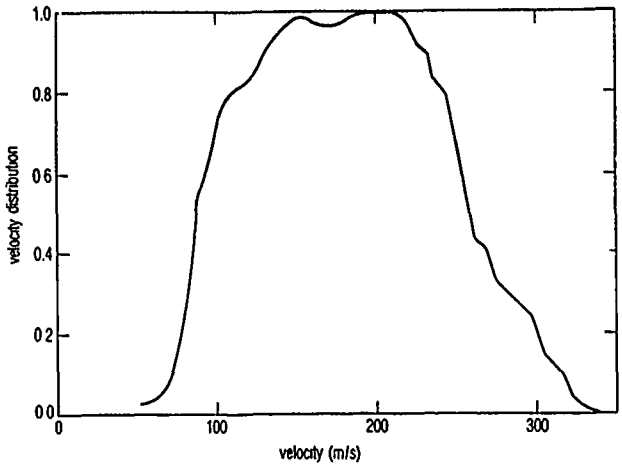


Fig. 12 Atomic velocity distribution calculated for a CBT using two orthogonal one-dimensional focussing dipoles. High gradient-to-field ratio.

43rd Annual Symposium on Frequency Control - 1989

FUNDAMENTALS OF TWO-WAY TIME TRANSFERS BY SATELLITE*

D. W. Hanson
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, Colorado 80303

Abstract

Two-way time transfer by satellite will eventually offer one of the highest accuracy systems at reasonable costs. Advantages of two-way transfer over other techniques include (1) leased space segment rather than specialized space hardware, (2) satellite location is required only to point antennas and not to compute distances, (3) effects on the accuracy of time transfer by the ionosphere and troposphere are sub-nanosecond without modeling, (4) locations of the clocks on the ground do not need to be known better than that provided by a geodetic map -- precise positions are not required, (5) simple averages of 100 one-second measurements yields about three hundred picoseconds resolution -- no elaborate data analysis is required, and (6) equipment delays are easily calibrated through the use of a standard earth station acting as a transfer standard.

This paper introduces the principles of two-way time transfer with emphasis on its use with commercial communications satellites. It discusses the limitations imposed by the atmosphere, the equipment, and the rotating, noninertial reference system on the assumption of reciprocal paths.

Introduction

Two-way time transfer is potentially one of the most accurate ways to compare clocks. The high accuracy is obtained by the users simultaneously exchanging signals via a communications satellite. If the paths between the clocks are reciprocal or very nearly so, the delays cancel. The difference between the clocks is then half the difference in time interval counter readings. The main advantage of the two-way technique is that no knowledge is required concerning the location of the clocks or satellite with a minor exception--the Sagnac effect, which is easily calculated. The main disadvantage of the system is the participants must be able to simultaneously transmit and receive signals.

Synchronization of clocks by the two-way method is not new. The technique coupled with communication satellites is also not new. Experiments have been conducted for more than 25 years using numerous different satellites, signal structures, modulation techniques, and carrier frequencies. A summary of these developments is given in Figure 1. Early experiments were characterized as using large satellite earth stations, inefficiently using the space segment (that is, unnecessarily large amounts of satellite power and bandwidth) earth stations that were located far from the laboratories housing time scales, and experimental satellites rather than operational systems with telecommunications traffic. Each subsequent experiment usually included improvements in one or more areas until today we have small on-site, user-owned earth stations The power and bandwidth demanded of the satellite are only a small percentage of the transponders' power-bandwidth product when operating in the code division multiple access (CDMA) mode using the highly efficient signal structure called spread spectrum.

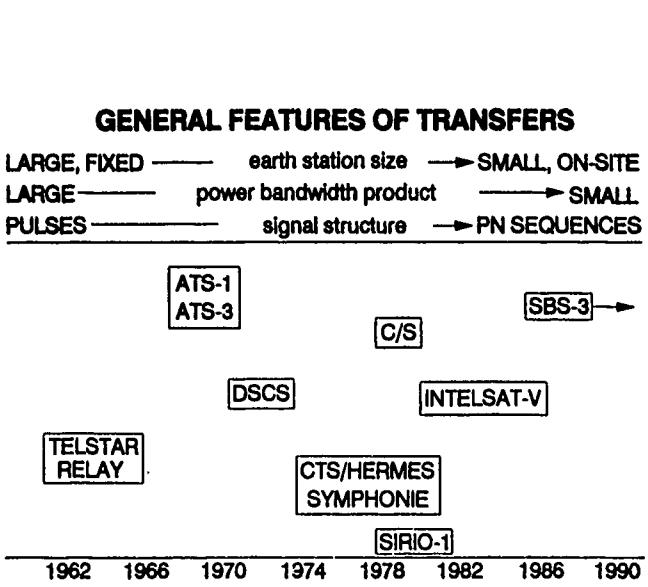
The first satellite-based, two-way time transfer took place between the United States and the United Kingdom in 1962 using the Telstar satellite, an early telecommunications satellite [1]. Later experiments used Telstar II and Relay and included participation by Japan [2]. Experiments during this period (1962-1965) employed large fixed earth stations, pulses as the signals, and frequency division multiple access (FDMA), and gave results of 0.1 to 20 μs accuracies. These experiments, while far from using facilities to the best advantage, did illustrate the potential of the technique for great improvements in time coordination on a global basis.

Between 1967 and 1975 the Application Technology Satellites (ATS) series operated by the National Aeronautics and Space Administration (NASA) supported a host of clock synchronization experiments. These experiments for the most part involved laboratories in the USA and Japan [3]-[7]. Some of these experiments for the first time involved small on-site earth stations and pseudo-noise sequences that allowed better use of the space segment. They possessed good time transfer qualities and allowed the use of code division multiple access (CDMA)--a technique that better satisfies reciprocity of path assumption which the two-way technique depends. These experiments led to a 5 μs accuracy and helped identify the Sagnac effect as significant to this application. Many improvements were made during this period but still with only experimental satellites, because commercial offerings were not fully suitable or affordable for time transfer.

One exception to the use of experimental facilities for two-way time transfer was an application using the Defense Satellite Communication System (DSCS) of the USA [8]. Here PN sequences, large earth stations, and CDMA combined to provide a 0.2 μs operational system that satisfied specific military requirements.

More experiments followed between 1976 and 1979; each used experimental satellites [9]-[13]. A high power communications satellite, called Communications Technology Satellite (CTS) by the US and Hermes by the Canadians, operated at Ku-band and provided, for the first time, long term comparison of time scales in Canada and the US. Small on-site earth stations located near the time scales were in use, but pulses and FDMA access remained. The European satellite called Symphonie provided time scale comparisons across the Atlantic, within Europe, and between India and Europe but was generally limited to large fixed earth stations located far from the time scales involved.

During the period of 1978 to 1980 an Italian experimental satellite called SIRIO [14] was used to



1. History of two-way time transfers using satellites.

transfer time between sites in Italy using a single communication channel, time-shared between the participating sites--a time division multiple access approach. By interpolation for satellite motion over periods of a few seconds, accuracies of a few nanoseconds were achieved.

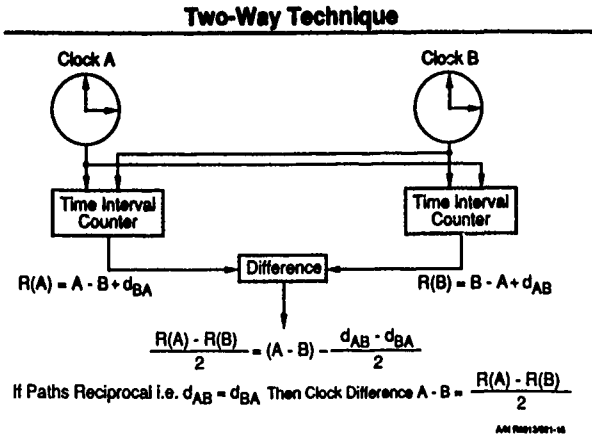
In 1981, the C/S satellite provided at 30/20 GHz two-way time transfers using earth stations with receivers as small as 1 m, PN sequences, and CDMA [15]. Accuracies of 13 ns were verified through terrestrial microwave links and portable clock transfers. The carrier frequencies were unusual in that they have not been available for use elsewhere.

In 1983 an Intelsat V satellite between the USA and Germany signified that the technique had finally matured to the point where operational use could be planned [16]. Small earth stations located on-site using PN sequences, CDMA, and commercially available modems routinely gave a 1 ns p ecision.

Today clocks at NIST in Boulder, Colorado and the United States Naval Observatory in Washington, D.C. are compared routinely using a commercial communications satellite in the two-way mode. A channel is leased for one half-hour three times a week. The earth stations are owned by the laboratories involved and connected directly to the time scales for best precision and accuracy of time transfer. The transfer uses spread spectrum signals, CDMA, and all commercial equipment. The uplinks are at 14 GHz and downlinks at 12 GHz, thereby providing a high degree of reciprocity through the ionosphere. Precision is less than 0.5 ns at all times. Accuracy will be determined following long term calibration exercises. Calibration is scheduled to begin in mid-1989. Two-way time transfers between North America and Europe are also scheduled to begin during late 1989 using an international communications satellite.

Theory

The two-way time transfer technique (without satellite) is illustrated in Figure 2 showing two clocks, clock A and clock B, and two time interval counters (TIC). The TIC on the left measures the difference between the 1 pps from clock A and the 1



2. Two-way time transfer basics.

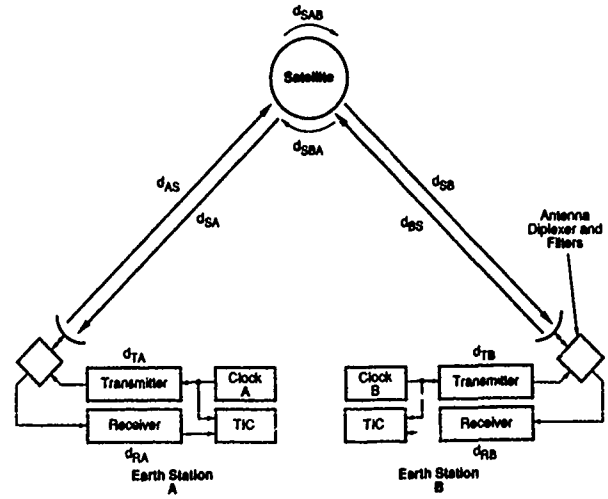
pps from clock B delayed by the cable delay d_{BA}. That measurement is given as R(A). Likewise, the TIC on the right measures R(B), the difference between clocks B and A with the cable delay d_{AB}. It is assumed that the delays in the cables connecting clocks to the start channels of the TIC's are very small or effectively 0. The readings on the two TIC's are subtracted to determine the time difference, A-B, in the two clocks as

$$A - B = [R(A) - R(B)]/2 + (d_{AB} - d_{BA})/2. \quad (1)$$

If the paths are reciprocal, that is, if the two cable delays are equal, then $d_{AB} = d_{BA}$, and the clock differences are just half the differences in the TIC readings, or

$$A - B = [R(A) - R(B)]/2. \quad (2)$$

If the two clocks are separated by large distances that cannot be spanned by cables they may be replaced by communication channels furnished by a communications satellite. In Figure 3 the cables are replaced by a transmitter and antenna, an uplink to the satellite, a path through the satellite, a downlink (at a different frequency from the uplink), and an antenna and receiver.



3. Two-way time transfer using a satellite.

Following the form of Equation 1 and selectively grouping the delays, we have the time difference between clock A and clock B as half the difference in the TIC readings, as was the case examined before, plus a number of new terms:

$$A - B = [R(A) - R(B)]/2 - (d_{TB} - d_{RB})/2 + (d_{TA} - d_{RA})/2 + (d_{AS} - d_{SA})/2 - (d_{BS} - d_{SB})/2 + (d_{SAB} - d_{SBA})/2 - 2\omega A/c^2. \quad (3)$$

Equation 3 is also shown in Figure 4 with additional comments added for clarity. The additional terms in Equation 3 relative to the simpler Equation 1 involve the differences in time delay in the transmitting and receiving equipment, differences in the uplink and downlink delays, differences in delays through the satellite, and one final term, $2\omega A/c^2$, which is due to a rotating system and finite signal velocity. These terms or pairs of terms are discussed in more detail below.

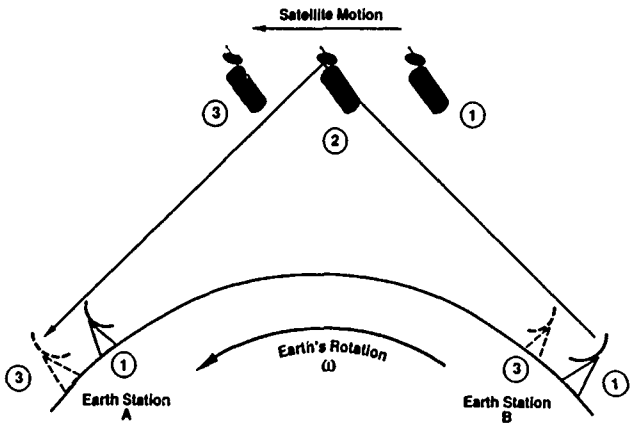
Two-Way Technique Using a Geostationary Satellite

$A-B = \frac{R(A)-R(B)}{2}$	Time Interval Counter Readings
$-\frac{d_{TB}\cdot d_{RB}}{2} + \frac{d_{TA}\,d_{RA}}{2}$	Earth Station Equipment
$+\frac{d_{AS}\cdot d_{SA}}{2} - \frac{d_{BS}\,d_{SB}}{2}$	Propagation
$+\frac{d_{SAB}\cdot d_{SBA}}{2}$	Delay in Satellite
$-\frac{2\omega A}{C^2}$	Earth's Rotation

4. Two-way time transfer equation.

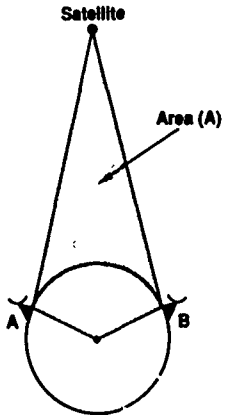
Sagnac Delay

The last term in Equation 3 is known as the Sagnac effect and can be derived using powerful theoretical tools. A simple physical understanding can be had, however, from consideration of Figure 5. Earth stations A and B, and the satellite are shown here



5. Earth's rotation introduces non-reciprocity.

at an instant of time "1" when pulses are sent to the satellite. The earth is rotating at the rate ω, causing earth station A to be in position "3" when the signal from earth station B arrives. The earth's rotation and the finite velocity of the signal have combined to increase the path length from B to A. Likewise, the signal from A to B experiences a corresponding decrease in path length. These changes in path length are equat to $2\omega A/c^2$ where c is the velocity of light, ω is the earth's rotation rate, and A is the area defined by the projections onto the equatorial plane by the line segments connecting the satellite and the earth's center to the two earth stations an shown in Figure 6. This term can be calculated with sufficient accuracy for nanosecond time transfers with only approximate knowledge of the earth station's and satellite's positions. For example, in the time transfer between NIST and the USNO using the geostationary satellite at 95° West longitude, a 1° error in satellite longitude produces less than 100 ps error in the calculation of the Sagnac effect. The same is true for a 300 m error in longitude or latitude for the location of NIST. The assumption of a perfectly circular geostationary orbit does not produce more than 50 ps error for the eccentricities usually associated with a typical "geostationary" orbit. This is a significant advantage over other methods of time transfer where very accurate knowledge of these positions is required. As is often the case an accurate satellite ephemeris is difficult to come by and limits any time transfer effort.



6. Area in the Sagnac equation.

Equipment Delays

The term

$$-(d_{TB}-d_{RB})/2 + (d_{TA}-d_{RA})/2,$$

from Equation 3 involves the differences in the signal delay in the transmitting and receiving sides of the satellite earth station. If the two earth stations could be collocated to assure cancelation of the propagation delays and a null Sagnac effect; if the clocks A and B were synchronized, as would be the case if clocks A and B were the same clock; if the delays through the satellite were equal, that is, $d_{SAB} = d_{SBA}$ (d_{SAB} denoting the delay through the satellite when the signal is moving from earth station A to earth station B), as would be the case for most satellite systems operating in the CDMA mode; then the difference of the TIC readings gives the desired equipment delays as

$$[R(A) - R(B)]/2 = -(d_{TA} - d_{RA})/2 + (d_{TB} - d_{RB})/2. \quad (4)$$

In most cases, however, where the two earth stations cannot be collocated even temporarily, a third or calibration earth station, a portable or mobile unit, can be used to make these measurements. With this calibration earth station collocated with earth station A the difference in the TIC readings, $[R(C)_A - R(A)]/2$, is given by

$$[R(C)_A - R(A)]/2 = -(d_{TC} - d_{RC})/2 + (d_{TA} - d_{RA})/2, \quad (5)$$

where $R(C)_A$ is the TIC reading at the calibration earth station when collocated with the earth station A. Similarly, when the calibration earth station is collocated with earth station B,

$$[R(C)_B - R(B)]/2 = -(d_{TC} - d_{RC})/2 + (d_{TB} - d_{RB})/2. \quad (6)$$

If delay, $d_{TC}-d_{RC}$, in the calibration earth station is constant the difference in these two measurements gives the desired quantity, or

$$[R(C)_B - R(B)]/2 - [R(C)_A - R(A)]/2 = -(d_{TA} - d_{RA})/2 + (d_{TB} - d_{RB})/2. \quad (7)$$

Therefore, what can be the most difficult nonreciprocal issue in the two-way time transfer is easily dealt with through the use of an earth station that may be collocated with the earth stations requiring calibration.

Propagation Delays

The signal delays associated with the up and down links are, from Equation 3, given as

$$-(d_{AS} - d_{SA})/2 + (d_{BS} - d_{SB})/2.$$

The uplink carrier frequencies for the Fixed Satellite Service (FSS) in use today are nominally 6 and 14 GHz paired with the downlinks of 4 and 12 GHz. The signal path is mainly free space with a small amount of ionosphere and troposphere. The path followed by the uplink and the downlink are essentially the same path. The free space part is reciprocal, as is the tropospherical part since, at these frequencies, the group or signal delays are frequency independent. The signal delay through the ionosphere has a $1/f^2$ frequency dependence. For typical elevation angles the differences in the delay terms are small and the total effect additionally reduced since the differences are halved. The reciprocity assumption is, therefore, likely to hold to less than 100 ps for operation in the Ku-band of the FSS.

Satellite Delays

The satellite time delay term represents the delays, d_{SAB}, to the signal by the satellite when going from earth station A to earth station B and those, d_{SBA}, going from B to A. For many satellites the equipment has a small delay variation across the transponder bandwidth so even in a FDMA mode the differences amount to only a few nanoseconds. When using the CDMA mode, made possible by the use of PN signals, signals from each earth station passes through the same satellite equipment at the same frequencies. The delays in each direction are therefore identical and the differences 0.

Data

Table 1 presents 30 s of data collected at the USNO and NIST during one of the now routine time transfers. The first column contains the time tag for each data point in hours, minutes, and seconds. The second and third columns are the 1 s averages of time interval measurements in units of seconds as shown in Figure 2. The fourth column is one-half the difference between the second and third columns also given in seconds. A simple analysis of these

Table 1. Example of two-way data.

HH:MM:SS	R(A)=A-B-d_{BA}	R(B)=B-A-d_{AB}	A-B=[R(A)-R(B)]/2
15:49:00	0.25103279152	0.25103074887	1.02133E-06
15:49:01	0.25103279322	0.25103075123	1.02099E-06
15:49:02	0.25103279394	0.25103075164	1.02115E-06
15:49:03	0.25103279506	0.25103075353	1.02076E-06
15:49:04	0.2510327965	0.25103075422	1.02114E-06
15:49:05	0.25103279824	0.25103075529	1.02147E-06
15:49:06	0.2510327992	0.25103075654	1.02133E-06
15:49:07	0.25103280086	0.25103075828	1.02129E-06
15:49:08	0.25103280127	0.25103075928	1.02099E-06
15:49:09	0.25103280291	0.25103076115	1.02088E-06
15:49:10	0.25103280471	0.25103076182	1.02144E-06
15:49:11	0.25103280502	0.25103076328	1.02087E-06
15:49:12	0.2510328066	0.25103076463	1.02099E-06
15:49:13	0.25103280807	0.25103076615	1.02096E-06
15:49:14	0.2510328085	0.2510307668	1.02085E-06
15:49:15	0.25103281102	0.2510307682	1.02141E-06
15:49:16	0.25103281041	0.25103076943	1.02049E-06
15:49:17	0.25103281232	0.25103077098	1.02067E-06
15:49:18	0.25103281311	0.25103077217	1.02047E-06
15:49:19	0.25103281557	0.25103077342	1.02107E-06
15:49:20	0.25103281648	0.51030774	1.02124E-06
15:49:21	0.25103281777	0.25103077648	1.02065E-06
15:49:22	0.25103281973	0.25103077697	1.02138E-06
15:49:23	0.25103282057	0.25103077754	1.02151E-06
15:49:24	0.25103282162	0.25103077922	1.0212E-06
15:49:25	0.25103282285	0.25103078064	1.0211E-06
15:49:26	0.25103282324	0.25103078221	1.02051E-06
15:49:27	0.25103282449	0.25103078348	1.0205E-06
15:49:28	0.25103282682	0.25103078521	1.02081E-06
15:49:29	0.25103282732	0.25103078568	1.02082E-06

NUMBER OF MEASUREMENTS = 30

MEAN = $\sum_{i=1}^{30} x_i/n$ = 1021.01 ns

σ = $\{ 1/(n-1) \sum_{i=1}^{30} (x_i-x)^2 \}^{1/2}$ = 0.32 ns

maximum = 1021.51 ns

minimum = 1020.47 ns

90% confidence level is 0.1 ns

30 measurements gives the mean as 1021.01 ns (± 0.1 ns for 90% confidence) and a standard deviation of 0.32 ns. All 30 data points are within ± 0.5 ns of the mean. This is a data set selected at random to illustrate the simplicity of the data analysis.

Summary

Two-way time transfer used with geostationary communication satellites can be a practical technique for the comparison or synchronization of clocks at the highest levels of precision and accuracy at reasonable costs. High accuracy is possible because (a) the use of a transfer or calibration earth station, provides the required measure of earth station delays, (b) the relativistic (Sagnac) effect may be accurately calculated with relatively imprecise information on the locations of the satellite and user's clocks, and (c) satellite and propagation path delays cancel to a large extent due to a high degree of path reciprocity.

Commercially available, low cost equipment and an abundance of satellite channels available for lease make implementation of the technique relatively simple. Reduction of data requires very little effort or special procedures. The technique has been used operationally in the United States for more than one year and has shown itself to be simple to operate and maintain. Space segment costs have been approximately $100 (U.S.) per week with enough excess capacity to support the synchronization of a network of at least ten clocks rather than the present two.

The only real barriers to further expansion involves the organization of participants and approval/licensing procedures for the earth stations.

References

[1] STEELE, J. McA, MARKOWITZ, W. and LIDBACK, C. A., "Telstar time synchronization," IEEE Trans. on Instrumentation and Measurement, Vol. IM-13, No. 4, pp. 164-170, December, 1964.

[2] MARKOWITZ, W., LIDBACK, C. A., UYEDA, H. and MURAMATSU, L., "Clock synchronization via Relay II satellite," IEEE Trans. on Instrumentation and Measurement, Vol. IM-15, No. 4, pp. 177-184, December 1966.

[3] JESPERSEN, J., KAMAS, G., GATTERER, L., and MACDORAN, P., "Satellite VHF transponder time synchronization," Proc. IEEE, Vol. 56, No. 7, pp. 1202-1206, July 1968.

[4] MAZUR, W. E. JR., "Dual transponder time synchronization at C-band using ATS-3," Proc. IEEE (Lett.), Vol. 60, No. 5, May, 1972, 633-634.

[5] RAMASASTRY, J., ROSENBAUM, B., MICHELINI, R., and KUEGLER, G., "Clock synchronization experiments performed via the ATS-1 and ATS-3 satellites," IEEE Trans. on Instrumentation and Measurement, Vol. IM-22, No. 1, pp. 9-12, March 1973.

[6] CHI, A. R. and BYRON, E., "Two-way time transfer experiment using a synchronous satellite," Proceedings of the Seventh Annual Precise Time and Time Interval Planning Meetings, Washington, D.C., December 2-4, 1975.

[7] SABURI, Y., YAMAMOTO, M. and HARADA, K., "High precision time comparisons via satellite and observed discrepancy of synchronization," IEEE Trans. on Instrumentation and Measurement, Vol. IM-25, No. 4, pp. 473-477, December, 1976.

[8] MURRAY, J. A., PRITT, D. L., BLOCKER, L. W., LEAVITT, W. E., HOOTEN, P. M., and GORING, W. D., "Time transfer by Defense Communications Satellite," Proc. of 25th Annual Frequency Control Symposium, Atlantic City, New Jersey, pp. 186-193, April, 1971.

[9] COSTAIN, C., DAAMS, H., BOULANGER, J., HANSON, D. W., and KLEPCZYNSKI, W.J., "Two-way time transfer between NRC/USNO via the Hermes(CTS) satellite," Proceedings of the Tenth Annual Precise Time and Time Interval Applications and Planning Meeting, Washington, D.C., November 28-30, 1978.

[10] BRUNET, M., "Synchronization of atomic clocks through the Symphonie Satellite," Radio Science, Vol. 14, No. 4, pp. 721-730, July-August, 1979.

[11] COSTAIN, C. C., BOULANGER, J. S., DAAMS, H., HANSON, D. W., BEEHLER, R. A., CLEMENTS, A. J., DAVIS, D. D., KLEPCZYNSKI, W. J., VEENSTRA, L., KAISER, J., GUINOT, B., AZOUBIB, J., PARCELIER, B., FREON, G., and BRUNET, M., "Two-way time transfer via geostationary satellites NRC/LPTF (France) via Symphonie," Proc. of 11th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 499-520, November 27-29, 1979, Washington, D.C.

[12] MATHUR, B. S., BANERJEE, P., SOOD, P. C., SAXENA, M., KUMAR, N., and SURI, A. K., "Precise time and frequency intercomparison between NPL, India and PTB, Federal Republic of Germany via satellite Symphonie-1," Proc. 12th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 863-875, December 2-4, 1980, Washington, D. C., or Journal of IETE, Vol. 27, No. 10, October, 1981.

[13] VEENSTRA, L., KAISER, J., COSTAIN, C., KLEPCZYNSKI, W., and ALLAN, D., "Frequency and time coordination via satellite," Comsat Technical Review, Vol. 11, No. 2, Fall 1981.

[14] DETOMA, E. and LESCHIUTTA, S., "Two-way sequential time synchronization: preliminary results from the SIRIO-1 experiment," Proc. 12th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 329-351, December 2-4, 1980, Washington, D.C.

[15] IMAE, M., OKAZAWA, H., SATO, T., URAZUKA, M., JOSHIMUTA, K., and YASUDA, Y., "Time comparison experiments with small k-band antennas and SSRA equipments via a domestic geostationary satellite," IEEE Trans. on Instrumentation and Measurement, Vol. 32, No. 1, 1983.

[16] HARTL, P., VEENSTRA, L., GIESCHEN, N., WENDE, C. M., KLEPCZYNSKI, W., NAU, H., and STOIBER, R., "Spread Spectrum time transfer experiment via INTELSAT," Proc. 15th Annual Precise Time and Time Interval Applications and Planning Meeting, Washington, D.C., Dec. 6-8, 1983.

43rd Annual Symposium on Frequency Control - 1989

TWO-WAY SATELLITE TIME TRANSFERS BETWEEN AND WITHIN NORTH AMERICA AND EUROPE*

L. B. Veenstra
International Telecommunications Satellite Organization
3400 International Drive, N.W.
Washington, D.C. 20008-3098
and
D. W. Hanson
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, Colorado 80303

Abstract

A satellite operated by the International Telecommunications Satellite Organization (INTELSAT) and located at 307° East longitude (53° West) provides new and unique capabilities for the coordination of time scales in North America and Europe using the two-way technique. A network of coordinated clocks using small satellite earth stations collocated with the time scales is possible. Antennas as small as 1.8 m at K-band and 3 m at C-band transmitting powers of less than 1 W will provide signals with timing jitters of less than 1 ns using the MITREX spread spectrum modems.

The technical details of the satellite and requirements on satellite earth stations are given. The resources required for a regular operational international time transfer service is analyzed with respect to the existing international digital service offerings of the INTELSAT Business Service (IBS) and INTELNET. Coverage areas, typical link budgets, and a summary of previous domestic and international work using this technique are provided. Administrative procedures for gaining access to the space segment are outlined. Contact information for local INTELSAT signatories is listed. A typical satellite earth station with time transfer capability is described in some detail.

Spread Spectrum Time Transfer via INTELSAT

The microwave time and ranging experiments (MITREX) modem is an efficient device for time transfer via satellite using spread spectrum techniques. International time transfer using satellite spread spectrum techniques regularly will probably require the use of an INTELSAT space segment. There are no substantial technical problems in such use. However, it is necessary to identify the operational issues of using this system in the INTELSAT environment. This paper describes the INTELSAT service compatible with spread spectrum time transfer and suggests how such a service could be implemented.

We use trade names and company names to specify the experimental procedure adequately and do not imply endorcement by the National Institute of Standards and Technology.

INTELSAT

INTELSAT, the International Telecommunications Satellite Organization, with headquarters in Washington, D.C., is an international cooperative of 117 member nations that owns and operates the global commercial communications satellite system used by countries around the world for international communications and by 35 countries for domestic communications. A fundamental characteristic of the system, from the point of view of time transfer is that access to the INTELSAT space segment is controlled by various national entities, usually the members of INTELSAT. These entities are responsible for the operation of earth stations accessing INTELSAT space segment. In many cases, these entities, the posts, telephones and telegraphs (PPT's), are part of their national government or are chartered by their governments to provide access to INTELSAT. The earth stations may be owned directly by the entities, by common carriers or by the end users, depending on national policy. INTELSAT operates only the space segment and has no direct role in the operation of the ground segment. Thus, time transfer users will need to arrange with their national entities for earth station operation and for the right to use INTELSAT space segment. Appendix 1 of this paper contains contact points for these entities in North America and Europe.

Services

Two specialized INTELSAT services are intended for digital links to small earth stations. INTELNET in particular, has enough flexibility in its technical description to allow the operation of a spread spectrum time transfer link.

INTELNET

INTELNET was designed to facilitate the operation of very small earth stations in one-way data broadcasting and two-way low speed data transfers. Spread spectrum operation is allowed, along with conventional BPSK or QPSK modulation. Operation under the INTELNET service description is specifically authorized for very small antennas.

Space segment is leased in "bulk" under the INTELNET service. This offering is defined in terms of specific transponder bandwidths with a corresponding allocation of power. Any required bandwidth may be used, with the resources scaled from the defined allocation. For example, at K-band, a lease of 1 MHz capacity would provide 22 dBW of transponder power. A lease of 2.25 MHz would provide 25.52 dBW. In general, the ratio of power to bandwidth is higher (excess power) than is needed for a single spread spectrum time transfer link. Both full time and occasional use service, with a minimum of 30 minutes per period, is available.

Listed below are the basic lease powers, referenced to a 1 MHz bandwidth, available on the INTELSAT V-A(F-13) spacecraft.

Global	C-Band	3.5 dBW
Hemi	C-Band	4.5 dBW
Zone	C-Band	4.5 dBW
West spot	K-Band	13.5 dBW
East spot	K-Band	16.0 dBW

INTELSAT Business Service

Space segment for digital communications links can also be obtained under the INTELSAT Business Service (IBS) tariff. In this service the power and bandwidth supplied for a channel are defined in terms of reference links between standard sized earth stations. The reference link assumes conventional QPSK data transmission with either rate 3/4 or rate 1/2 forward error correction (F.E.C.). Sufficient power is available to provide better than 10^{-8} bit error rate performance under clear sky conditions. It is available under a full-time, part-time (scheduled at least 1 hour per day, 7 days per week), or occasional-use tariff. While the spread spectrum nature of the MITREX modem is outside the IBS technical description, the IBS service should be considered when it is necessary to provide communication links between standards sites. On most transponders where IBS is used, INTELSAT has reserved occasional use capacity. This bandwidth could be used to accommodate the MITREX modem operating under the INTELNET service as described below.

Operational Matters

In all three ocean regions there are satellites with capacity reserved for occasional use. The reserved bandwidth is 3.173 MHz wide, sufficient for a 2.048 Mbits/s IBS carrier. These occasional use channels clearly would accommodate a spread spectrum link if the transmit spectrum were restricted by additional filtering. In the full-connectivity transponders, described below, the reserved capacity is in the form of two adjacent occasional-use channels, providing a bandwidth of 6.345 MHz. The operating frequencies for the occasional use channels are listed in Appendix 2.

Transponder Configurations

Most of the Europe-North American IBS traffic is carried on the INTELSAT VA(F-13) located at 307° East. The following configurations of transponders are currently available:

East K-band spot	to	West K-band spot
West K-band spot	to	East K-band spot
East K-band spot	to	West C-band hemi
West C-band hemi	to	East K-band spot
West K-band spot	to	East C-band hemi
East C-band hemi	to	West K-band spot
East K-band spot	to	West C-band hemi
West C-band hemi	to	East K-band spot
West C-band zone	to	East C-band zone
East C-band zone	to	West C-band zone
West C-band hemi	to	West C-band hemi
East C-band zone	to	East C-band zone

East K-band spot	full-connectivity	East K-band spot
West K-band spot		West K-band spot
East C-band hemi		East C-band hemi
West C-band hemi		West C-band hemi

The full-connectivity transponder configuration provides the most flexible environment for time transfer links. This consists of a set of four transponders, interconnected at the satellite so that a signal received on any of the four uplink beams is retransmitted on all four downlink beams simultaneously. Two of the beams operate at C-band. These are the West Hemi and the East Hemi. The coverages of these spacecraft antennas is shown in Figure 1. Interconnected with these two C-band hemi beams are two K-band beams. The K-band West spot covers the United States and southern Canada, as can be seen Figure 2. The East K-band beam, Figure 3, covers western Europe.

Full-connectivity operation can be used with spread spectrum code division multiple access (CDMA) to allow several time transfer links to be established simultaneously on the same frequency, with all carriers visible to each user. This means that in a two-station, two-way transfer, it is possible for each site to monitor its own signal while receiving from the remote site.

Additional capacity in the form of East spot-West spot K-band capacity is available at the 325.5° East and 335.5° East locations. C-band capacity is assigned on the 325.5, 335.5 and 341.5° East locations serving the Atlantic Ocean region, on the Indian Ocean satellite at 63° East and in the Pacific region at 174 and 180 degrees East. K-band capacity is also available in the Pacific Ocean region between Korea or Japan and the West coast of the United States.

Link Budgets for MITREX

The link budgets in Table 1 show the required power for a time transfer link operating in the full-connectivity transponder. The transmitted power has been set to produce at least 54 dB-Hz to a small K-band (1.8 m) station. This same power will also be sufficient for use with a C-band 4.5 m antenna. These link budgets show both the uplink and the downlink calculations for each of the four full-connectivity transponders, West spot, East spot, West hemi and East hemi. In the full-connectivity configuration, an uplink will produce different downlink powers in the four downlink beams, depending on the individual TWT characteristics and satellite antenna gains. The bottom section recalculates what the spread spectrum signal would look like at the largest earth stations using the transponder, standard C station K-band, and standard A at C-band, on a spectrum analyzer. This indicates what the interference potential of such a carrier would be to other users operating links in these transponders. Even to such large stations, the spread spectrum operation appears as a very low level carrier. At the most visible location, a standard C-band receiving earth station located in the West spot downlink, the received noise floor will be increased by 6.4 dB.

With the assumption of the use of 1.8 m K-band antennas and 4.5 m C-band antennas, the links can be summarized as follows:

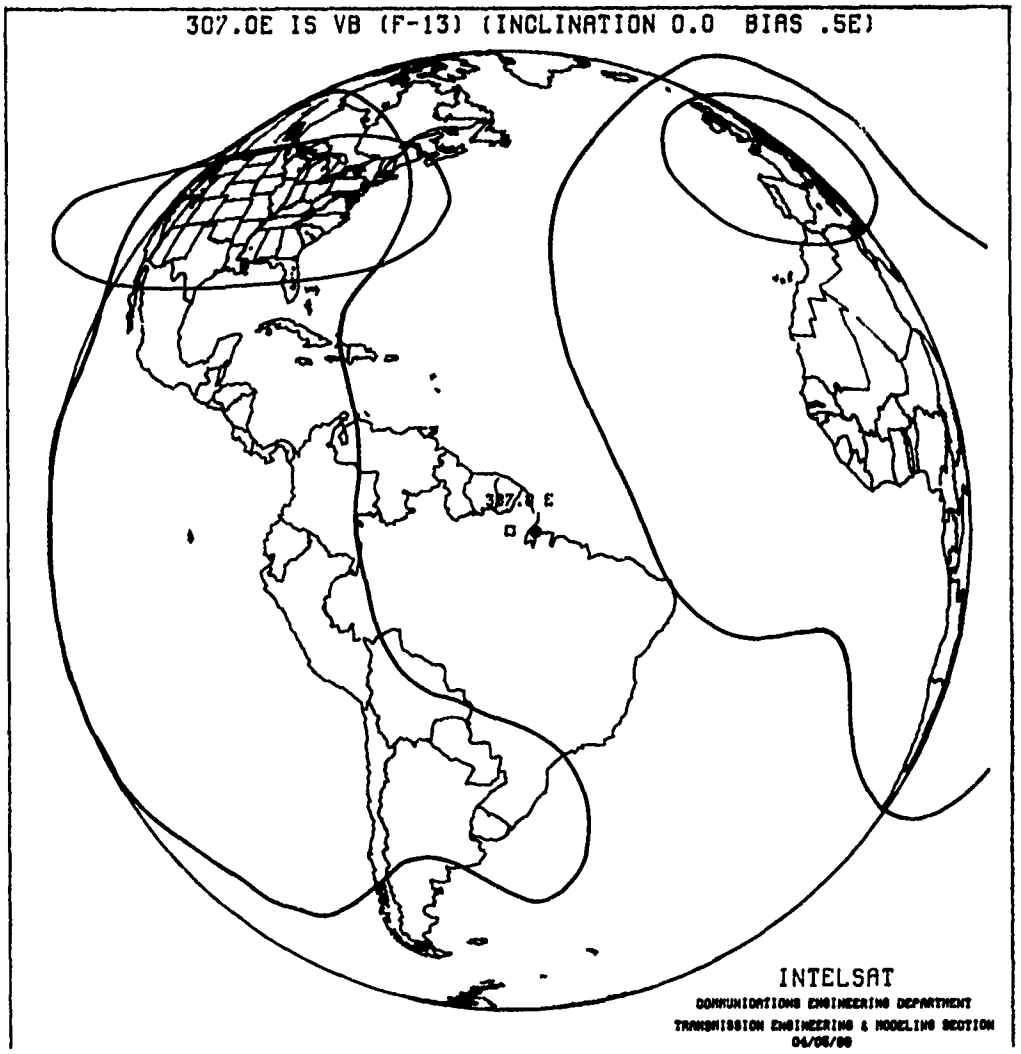


Figure 1. INTELSAT VA(IBS) Interconnections from 307° East

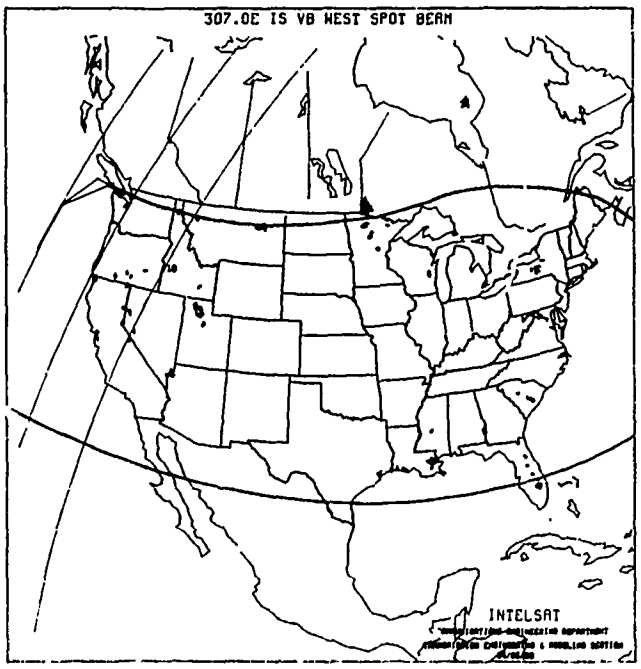


Figure 2. INTELSAT VA(IBS) West Spot Coverage from 307° East

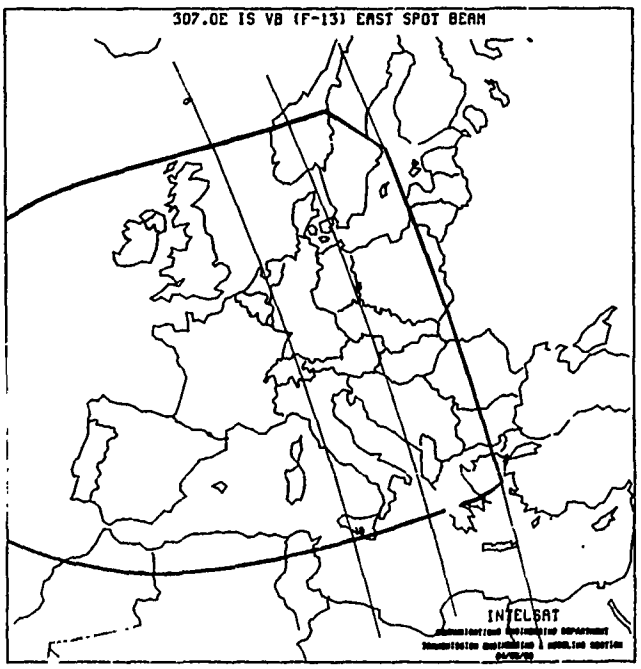


Figure 3. INTELSAT VA(IBS) East Spot Coverage from 307° East

Table 1

Project: MITREX Modem Link
Actual Transponder Operating Values
TX E/S: Std. G RX E/S: Std. G 1.8 m K-band 4.5 m C-band

SATELLITE & LOC. : INTELSAT V-A(F-13) @ 307° EAST
TRANSP.: WS/ES Full-connectivity High Gain
Carrier Type: Time Transfer Spread Spectrum
U/L CARR. FREQ.(MHz): 14092.16 14092.16 6017.16 6017.16
D/L CARR. FREQ.(MHz): 12597.16 11797.16 3792.16 3792.16

Co-channel Interference X Value (IESS-410 Table 1) 22 dB

Symbol Rate	2500.0 Kbit/s WS/WS	ES/ES	WH/WH	EH/EH	
Transmit e.i.r.p.	54.9	45.6	46.0	45.9	dBW
Tracking error	0.5	0.5	0.5	0.5	dB
U/L margin	0.0	0.0	0.0	0.0	dB
U/L path loss	207.3	207.4	197.7	197.7	dB
U/L asp. corr.	2.0	2.0	2.0	2.0	dB
Gain 1 m^2	44.4	44.4	37.0	37.0	db/m^2
Operating flux b.c.	- 110.4	-119.8	-117.1	-117.2	dBW/m^2
Sat. flux b.c.	-73.6	-83.0	-80.3	-80.4	dBW/m^2
Input back-off	-36.8	-36.8	-36.8	-36.8	dB
TWT I/O	6.0	6.0	4.3	4.3	dB
Output back-off	-30.8	-30.8	-32.5	-32.5	dB
saturated e.i.r.p. b.c.	46.5	44.6	32.4	33.0	dBW
D/L e.i.r.p. b.c.	15.7	13.8	-0.1	0.5	dBW
D/L e.i.r.p. beam edge	11.7	9.8	-4.1	-3.5	dBW
C/T & C/N CALCULATIONS:					
Transmit e.i.r.p.	54.9	45.6	46.0	45.9	dBW
Tracking error	0.5	0.5	0.5	0.5	dB
U/L margin	0.0	0.0	0.0	0.0	dB
U/L path loss	207.3	207.4	197.7	197.7	dB
U/L asp. corr.	2.0	2.0	2.0	2.0	dB
Satellite G/T b.c.	-1.9	0.5	-8.7	-8.7	dB/K
$(C/T)_{up}$	-156.8	-163.8	-162.9	-163.0	dBW/K
(C/T) intermod	-136.9	-138.8	-152.7	-152.1	dBW/K
D/L e.i.r.p. at b.c.	15.7	13.8	-0.1	0.5	dBW
Tracking error	0.5	0.5	0.5	0.5	dB
D/L margin	0.0	0.0	0.0	0.0	dB
D/L path loss	205.7	206.5	195.8	196.5	dB
D/L asp. corr.	2.0	2.0	2.0	2.0	dB
E/S G/T	21.0	21.0	22.7	22.7	dB/K
$(C/T)_{dn}$	-171.5	-174.2	-175.7	-175.8	dBW/K
(C/T) Downlink X-pol & ASI	-141.8	-141.8	-141.8	-141.8	dBW/K
$(C/T)_t$ link	-171.7	-174.6	-176.0	-176.1	dBW/K
Boltzmann's const.	-228.6	-228.6	-228.6	-228.6	dBW/Hz-K
C/N_o link	56.9	54.0	52.6	52.5	dB/Hz
Allocated BW	3500.0	3500.0	3500.0	3500.0	kHz
Occupied BW	3000.0	3000.0	3000.0	3000.0	kHz
10 log BW	64.8	64.8	64.8	64.8	dB-Hz
C/N	-7.8	-10.8	-12.1	-12.2	dB
Trans. rate E_b/N_o	-7.1	-10.0	-11.3	-11.4	dB
C_o/N_o	-7.1	-10.0	-11.3	-11.4	dB
Spectrum analyzer $(C_o+N_o)/N_o$	0.8	0.4	0.3	0.3	dB
Same links as received at large Standard G and A earth stations					
D/L e.i.r.p. at b.c.	15.7	13.8	-0.1	0.5	dBW
Tracking error	0.5	0.5	0.5	0.5	dB
D/L margin	0.0	0.0	0.0	0.0	dB
D/L path loss	205.7	206.5	195.8	196.5	dB
D/L asp. corr.	2.0	2.0	2.0	2.0	dB
E/S G/T	37.0	37.0	35.0	35.0	dB/K
$(C/T)_{dn}$	-155.5	-158.2	-163.4	-163.5	dBW/K
(C/T) Downlink X-pol & ASI	-141.8	-141.8	-141.8	-141.8	dBW/K
$(C/T)_t$ Link	-159.3	-164.9	-166.4	-166.4	dBW/K
Boltzmann's const.	-228.6	-228.6	-228.6	-228.6	dBW/Hz-K
C/N_o link	69.3	63.7	62.2	62.2	dB/Hz
Allocated BW	3500.0	3500.0	3500.0	3500.0	kHz
Occupied BW	3000.0	3000.0	3000.0	3000.0	kHz
10 log BW	64.8	64.8	64.8	64.8	dB-Hz
C/N	4.5	-1.0	-2.5	-2.6	dB
Trans. rate E_b/N_o	5.3	-0.2	-1.7	-1.8	dB
C_o/N_o	5.3	-0.2	-1.7	-1.8	dB
Spectrum analyzer $(C_o+N_o)/N_o$	6.4	2.9	2.2	2.2	dB

Uplink beam	WS	ES	EH	WH	
Antenna size	1.8	1.8	4.5	4.5	m
Hpa power	7.7	0.9	1.0	1.0	W
Hpa power	8.9	-0.4	0.0	-0.1	dBW
Antenna gain	46.0	46.0	47.7	47.7	dBi
E.I.R.P.	54.9	45.6	47.7	47.6	dBW
Downlink beam	WS	ES	WH	WH	
Beam edge power	11.7	9.8	-4.1	-3.5	dBW
1 MHz lease	13.5	16.0	4.5	4.5	dBW
2 MHz lease	16.5	19.5	7.5	7.5	dBW

MITREX Modem Use

The simplest way to use the MITREX modem on INTELSAT would be to identify a tariffed INTELNET service that provides at least the necessary power and bandwidth. For regular PSK transmissions, the signal bandwidth at the -18 dB points is required to be within the allocated bandwidth. As can be seen in the attached spectrum analyzer plot. Figure 4 shows, the bandwidth of the MITREX modem is 3.5 MHz at the -18 dB point. The 6 dB bandwidth of the MITREX modem output is 2 MHz. It is possible to apply additional filtering to the spread spectrum transmitted signal reducing its -18 dB bandwidth, and thus the nominal tariff. However with the spread spectrum operation at such low levels, the -18 dB bandwidth may not be appropriate for tariffing.

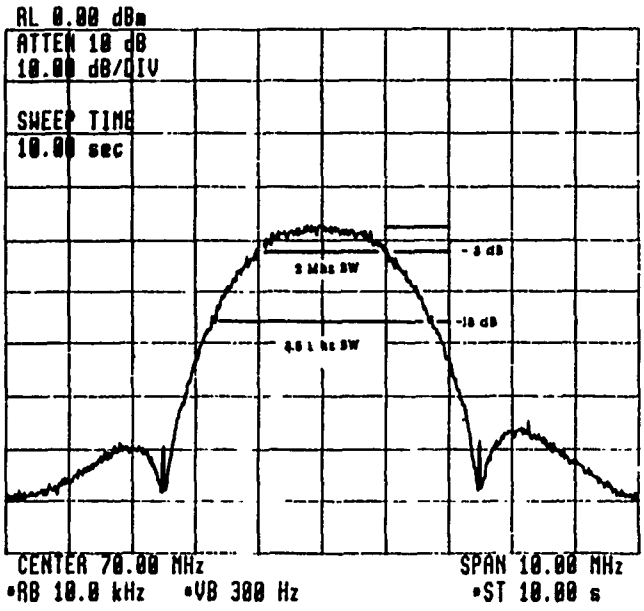


Figure 4. MITREX Modem Output Spectrum

Earth Stations

The earth station requirements to operate a time transfer link are quite modest.

The attached link budgets assume the use of 1.8 m K-band stations as a minimum size. This certainly is not the absolute minimum, but it does represent a useful compromise of physical size, link power requirements, and sidelobe performance. Therefore a station equipped with a solid state amplifier would clearly be suitable. Links involving larger stations will need even less transmitting power.

Any earth station classed as a standard INTELSAT antenna must satisfy the sidelobe gain limit described by the expression:

$$G = 32 - 25 \log \theta,$$

where G is the gain of the sidelobe envelope relative to an isotropic antenna in the direction of the geostationary orbit and is expressed in dBi, and θ is the angle in degrees from the axis of the main lobe.

In addition antennas operating at C-band must use circular polarization with a voltage axial ratio that does not exceed 1.09. However C-band antennas with a diameter of 2.5 m or less are only required to have a voltage axial ratio of 1.3. To operate in the IBS service a K-band antenna must have a minimum G/T ratio of 25 dB/K, qualifying as a standard E1. At C-Band the G/T requirement is 22.7 dB/K and a minimum transmitting gain of 47.7 dBi to qualify as an F1 standard antenna.

Implementation

INTELSAT has always been willing to support innovative uses of satellite technology by granting free use of space segment for tests and demonstrations. A request for free use must be submitted through the national signatory for each station involved. The technical approval process for a test or demonstration has two parts:

(1) Initially, the earth stations involved must be approved. Small stations not having the minimum G/T values (25 dB/K at K-band, 22.7 dB/K at C-Band) for IBS stations, would have to qualify under the standard G specification.

(2) A transmission plan for the proposed experiment will have to be examined to see whether what is proposed will work with the resources requested and, finally, whether the proposed transmissions may cause harm to other users of the space segment.

After approval, the carrier powers are set up in accordance with a test plan issued by INTELSAT and the experiment will then proceed. At the conclusion of the experiment, the participating Signatories are obligated to submit to INTELSAT a test report on the results. This report will be made available to any interested INTELSAT members.

One objective of such an experiment should be an evaluation of the compatibility of such a service with normal INTELSAT operations, with the view to proposing a tariffed technical description of spread spectrum time transfer. This could then be submitted to the INTELSAT Board of Governors for formal approval as a regular international service with the resources allocated and the consequent tariffs appropriate to the unique demands of spread spectrum time transfer.

Alternatively, commercial service could start immediately under the INTELNET service definition and tariffs. The occasional-use option would probably satisfy the requirements for periodic coordination links between various national standard labs.

Appendix 1.

INTELSAT Operations Representatives

Canada
Mr. M. Stephens
Teleglobe Canada
680 Sherbrooke St West
Montreal, Quebec, H3A 2S4
CANADA
Tel. 514-289-7584
Tlx. 9224
Ms. G. Pazos 514-289-7771

Denmark
Mr. Max Albertson
Telecom Denmark
Traffic Division
Telegrade 2
DK-2630 TASTRUP
DENMARK
Tel. 45-2-529111 ext 2320
Tlx. 22999 TELCOM DK

France
Mr. J. Meunier
France Telecom - D.T.R.E.
246 rue de Bercy
75584 Paris, FRANCE
Tel. 1-43426275
Tlx. 670372

Finland
Mr. Rolf Backman
General Directorate of P&T
Radio Department
International Section
Box 526
SF-00101 Helsinki, FINLAND
Tel. 0-7042283
Tlx. 123434

Germany
Mr. A. Binzer
Referat S15
FTZ Darmstadt, GERMANY
Tel. 6151-833459
Tlx. 419201
Mr. Eckstorff 6151-83-2382

Greece
Mr. S. Kontoleon
Hellenic Telecom. Org. (OTE)
International Comm. Dept
15 Stadiou Street
Athens 124. GREECE
Tel. 322-0899
Tlx. 219797

Iceland
Mr. Olafur Tomasson
General Directorate of Posts and Telecommunications
P.O. Box 270
IS-121 Reykjavik, Iceland
Tel. 1-26000
Tlx. 2000 GENTEL IS

Ireland
Mr. James Campoin
Marketing Manager
Telecom Eireann
St. Stephens Green West
Dublin 2, Ireland
Tel. 1 714444
Tlx. 91119
Mr. Louie Garvy Tel. 17-14444 ext 2286

Italy
Dr. Luigi Ruspantini
Telespazio
Via Alberto Bergamini 50
00159 Rome, ITALY
Tel. 498-2355
Tlx. 610654

The Netherlands
Mr. Peter Essers
PTT Telecommunicatie
Directorate for Infrastructure (DIS)
Prinses Beatrixlaan 9
P.O. Box 30000
2500 GA The Hague
The Netherlands
Tel. 70434725
Tlx. 32482 DIS NL

Norway
Mr. Claus Svendsen
Norwegian Telecommunications Admin
PO Box 6701, St. Olavs Plass
N-0130 Oslo 1, NORWAY
Tel. 70434725
Tlx. 71203 Gentel N

Portugal
Mr. Eliseu Crespo
Companhia Portuguesa Radio Marconi
Praca Marques Pombal, 15-4
P.O. Box 2778
1119 Lisbon Codex, Portugal
Tel. 534191
Tlx. 12384

Spain
Mr. J. Lorente
TELEFONICA
Plaza de Espana 4, Pta. 3 - 7th Floor
Madrid 28008, Spain
Tel. 1 241 9380
Tlx. 47793
Mr. A. Martin 1-522-2936

Sweden
Barbro Svensson
Televerket
S-123 86 Farsta
SWEDEN
Tel. 8-713-1568
Tlx. 14970 GENTEL S

Switzerland
Mr. P. Breu
General Directorate PTT
Radio and Television Main Division
Satellite Communications Branch
Speichergasse 6
CH-3030 Berne, Switzerland
Tel. 31-623756
Tlx. 911025
Mr. P. Chablais 31-622533

United Kingdom

Mr. M. Seymour
British Telecom, PLC
Landsec House Room 407
23 New Fetter Lane
London EC4A 1AE, England
Tel. 1-492-3166
Tlx. 883739
Mr. Mike Perry 1-492-2263

United States

Mr. Calvin Harriott
Communications Satellite Corp.
950 L. Enfant Plaza S.W.
Washington D.C. 20024
Tel. 202-863-6427
Tlx. 892688

INTELSAT

Mr. Lester Veenstra
INTELSAT
3400 International Drive, N.W.
Washington, D.C. 20008
Tel. 202-944-7090
Tlx. 64290

Appendix 2.

Occasional-Use Frequencies

Transponder	Configuration	Uplink	Downlink
11/11	WH/WH	5933.7700	3708.7700
41/41	WZ/WZ	5930.5975	3705.5975
41/41	WZ/WZ	5933.7700	3708.7700
51/51	EZ/EZ	5930.5975	3705.5975
51/51	EZ/EZ	5933.7700	3708.7700
42/52	WZ/EZ	6015.5750	3790.5750
52/42	EZ/WZ	6015.5750	3790.5750
43/53	WZ/EZ	6095.5975	3870.5975
53/43	EZ/WZ	6095.5975	3870.5975
13/173	WH/ES	6106.9825	12686.9825
73/13	ES/WH	14176.0425	3876.0425
23/163	EH/WS	6095.5975	11875.5975
63/23	WS/EH	14169.0000	3869.0000
61/171	WS/ES	14058.0000	12563.0000
71/161	ES/WS		
12/22/62/72/162/172	Full Connectivity		
		6017.16	3792.16
		14092.16	11797.16 (WS)
			12597.16 (ES)

43rd Annual Symposium on Frequency Control - 1989

IMPACT OF ATMOSPHERIC NON-RECIPROCITY ON SATELLITE TWO-WAY TIME TRANSFERS

J. Jespersen
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

In the two-way time transfer method no knowledge is necessary concerning the locations of the slave and the master clock stations relative to the satellite nor does one need any knowledge of the atmospheric delay. In this method, signals are exchanged back and forth, via a satellite transponder, between the master and slave stations. Because of the two-way signal exchanges the path delay cancels out in the relevant calculations leaving only the difference in the clock readings at the master and slave stations. (This statement assumes that all cable delays and other associated equipment delays are known and accounted for in the calculations.) However, the cancelling of path delay is critically dependant on the assumption that the signal path delay from master to satellite to slave is equal to the delay from slave to satellite to master--that is, the paths are reciprocal.

In actual practice this is hardly ever the case. Generally speaking the uplink signal frequencies to the satellite from both master and slave to the satellite differ from the downlink frequencies from the satellite to the master and slave stations so that the reciprocity assumption does not hold over the ionospheric portion of the path. Thus one must make calculations to see if the lack of reciprocity is significant in comparison to the degree of timing accuracy required. It is possible, under some circumstances to measure the degree of non-reciprocity directly. This is because the signals relayed by the satellite are generally available at both the slave and master stations simultaneously.

Introduction

The continuing improvement in frequency standards, and consequently clocks, creates a parallel need to perfect the means for comparing these devices at numerous locations throughout the world. In the past, as now, electromagnetic signals are the mainstay for these comparisons.

In the first few decades of this century, clock stability was not better than path stability--the variation in arrival time--of radio signals, so that, for all practical purposes, remotely located clocks could be compared with little or no loss due to signal path delay instability. However, with the introduction of atomic frequency standards in the 1950's, radio signals propagated over long distances through the ionosphere no longer had sufficient path stability to permit unimpaired clock comparisons. To some extent these problems were reduced by averaging out path instabilities--the averaging was usually achieved by integrating a signal for long periods of time--often weeks and even months. Loran-C and VLF navigation signals used in timing were and still are, treated in this way.

Another, less used scheme, entails averaging several signals, referenced to the same clock and broadcast simultaneously from the same location. Thus one could, for example, average WWV time signals on several different carrier frequencies at the same time. However, even in the best case, these schemes do not produce immediate results and often require considerable bookkeeping. Ultimately, path delay variations with periods of a year or more make the averaging approach impractical.

Fortunately, the introduction of atomic frequency standards in the 50's was accompanied by the communications revolution--arriving with the launch of the first artificial earth satellite. With communication satellites one could relay information reliably at high rates over intercontinental distances. The high rates were a consequence of the satellites' broadband communication channels and the reliability was due both to the technical excellence of satellite technology and to the frequencies at which the satellites operated. Whereas normal, long-distance-terrestrial-communication systems operate at frequencies which are low enough to reflect radio signals between the earth and the ionosphere, communication satellites operate at frequencies high enough for their signals to easily penetrate the ionosphere. That is, satellite signals pass through the ionosphere little affected by the ionospheric irregularities that afflict signals propagated around the world by the ionosphere.

The nearly vacuum line-of-sight path stability coupled with the large channel bandwidths was just what the timing community needed to greatly improve clock comparisons over great distances. The latter supported the transmission of very narrow pulses (or more recently their sophisticated cousins--spread spectrum signals) thus insuring that time-of-arrival measurements could be made with great precision and the former insured that path delay variations did not mask the phase variations inherent in the clocks--the phenomena we want to study.

The first satellite clock comparisons were carried out in the early 1960's with the Telstar communication satellite followed by numerous comparisons in the ensuing years. However due to the expense and the complexity, both political and technical, of carrying out these time transfers no routine comparisons were established. It is only today that we are beginning to see the development of routine comparisons over communication satellites as a result of the availability of both more economical ground equipment and modest channel costs. With this greater availability two-way satellite time comparisons are now reaching a mature stage where many factors, once essentially ignored, are now being scrutinized at a level not deemed necessary in the past. In this paper I shall focus on one of these factors--the reciprocity, or non-reciprocity--of atmospheric path delay through the atmosphere.

Background

Most communication satellites are in orbits well above the ionosphere so a signal from earth to satellite and return travels through the bulk of the ionized medium surrounding the earth as well as, of course, the lower neutral atmosphere lying between the surface of the earth and the bottom of the ionosphere, near 60 kilometers.

Several things impact signals in both the neutral and ionized regions. First, there are large scale structures which lead to (1) changes in the propagation velocity and (2) rotation of the plane of polarization of the radio signal. Second, the signal may be partially or completely absorbed by various mechanisms in the intervening media, and finally the signals may be scattered by small scale turbulent structures in the media leading to what is usually termed scintillation. In this paper I shall center on the first effect.

Phase Delay

In the general case, a signal propagating through the atmosphere will be both absorbed and delayed. As an illustration, consider a plane wave represented by the equation

$$\vec{E}(z,t) = \vec{E}_o e^{j(k\hat{n}z - 2\pi\nu t)}. \qquad (1)$$

Here k is the propagation constant $2\pi\nu/c$ in free space, c is the speed of light, ν the signal frequency, and E_o the field amplitude. n is the complex refractive index equal to $n_R + jn_I$. When the imaginary part of the refractive index is negative, the medium is dissipative and the wave decays exponentially. It is usual to define the power absorption coefficient as

$$\alpha = \frac{4\pi\nu}{c} n_I .$$

In the atmosphere the propagation constant is

$$kn_R = \frac{2\pi n\nu}{c} = \frac{2\pi\nu}{V_p} ,$$

where $n = n_R$ and v_p is the phase velocity, c/n_R. As a rule of the thumb the phase velocity is about 0.03 percent less than c in the lower neutral atmosphere

Here we are concerned primarily with how much extra time Δt--more accurately the extra phase path time, as we shall discuss in more detail later--it takes the signal defined by Equation (1) to travel through a medium with refractive index n as compared to the time it takes to travel through free space. In other words we want to evaluate the quantity

$$\Delta t = \frac{1}{c} \int_{path} (n_R - 1) dz.$$

Reciprocity in the Lower Neutral Atmosphere

Since (n -1) in the neutral atmosphere is a very small number it is common to introduce the refractivity N defined by

$$N = 10^6 (n_R - 1).$$

With this definition we may write the excess path length, L, as

$$L = 10^{-6} \int_{path} N(z) dz.$$

Thus the problem of evaluating non-reciprocity in the neutral atmosphere reduces to knowing--experimentally, theoretically or some combination thereof--appropriate expressions for n (or equivalently for N, the refractivity).

An often used empirical expression for N in the neutral atmosphere at radio frequencies is

$$N = 7790 \frac{P_o}{T} - 6480 \frac{P_v}{T} + 377 \times 10^5 \frac{P_v}{T^2} , \qquad (2)$$

which is sometimes referred to at the Smith-Weintraub Equation [1]. Here T is the temperature in Kelvins, P_o the partial pressure of dry air, and P_v the partial pressure of water vapor in pascals. (1 millibar = 100 newtons per square meter = 100 pascals; 1 atmosphere = 1013 millibars).

The first two terms are related to the displacement polarizations (due to the electric field of the penetrating signal) of the gaseous constituents in the air and the third term is due to the permanent dipole moment of water vapor. Equation (2) is accurate to less than 1% at frequencies below 100 GHz.

What concerns us here is the form of Eq. (2.) First we notice that it is dependent on the partial pressures of water vapor and the gaseous constituents of air and the temperature. More important we notice that there is no dependence on the frequency.

In general, ref [2], the up and down frequencies in two-way time transfers are not the same and therefore a frequency dependent term in the Smith-Weintraub would introduce non-reciprocity. That is, the excess path, L, the length in the two directions would not be the same and thus not cancel out. Happily, this is not the case.

Before we leave the subject of reciprocity in the neutral atmosphere, it is useful to consider the situation a little more carefully since, strictly speaking, N does depend, if ever so slightly, on frequency.

At radio frequencies the refractivity due to water vapor in the neutral atmosphere is about 20 times greater than it is in the near infrared or optical regions--certainly a frequency-dependent effect. At first glance this result may seem strange until we consider the underlying mechanisms leading to the Smith-Weintraub Equation.

If we look at the resonant spectrum of the gases making up the neutral atmosphere we find that there are a great many lines at frequencies well above 10 GHz where most satellite two-way time transfers are made today. Figure 1 schematically shows one such line. The figure shows that the tail of the real part of the refractive index approaches 1 as ν approaches ∞ but does not quite reach 1 as ν approaches 0. In other words all the spectral lines above the range of frequencies involved in the time transfer contribute to the refractive index as illustrated schematically in Fig 2. Here we see the tails of many lines adding together to produce the refractive index at the frequencies of interest. Thus, strictly speaking N does have a frequency dependence at radio frequencies.

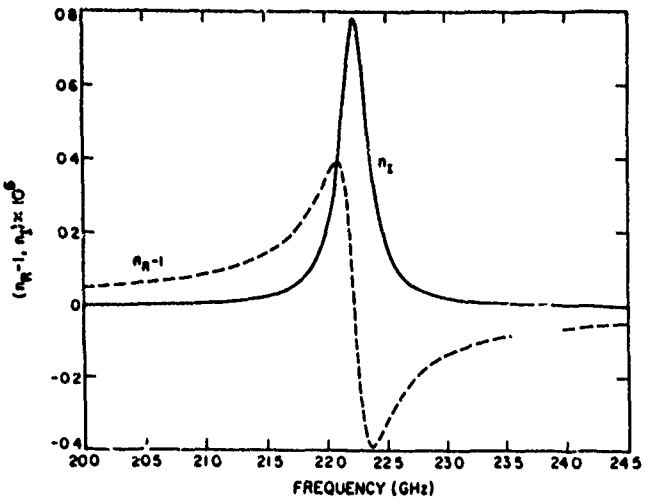


Fig. 1 Refractive index variation of water vapor line with frequency.

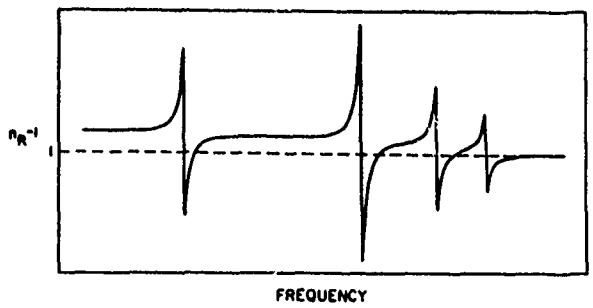


Fig. 2 Low frequency refractive index variation due to summation of tails of higher frequency lines.

How important is this dependence for present-day, two-way time transfers? The answer is that the dependence produces a non-reciprocity that is well below the noise floor, about 0.5 ns, in present two-way exchanges. However, we can imagine circumstances where the effect is not negligible.

Under normal atmosphere conditions a signal propagated at 10 GHz to a satellite directly overhead encounters an extra delay of about 8 ns due to the gaseous part of the atmosphere and about 1 ns more due to water vapor. An optical signal taking the same path undergoes essentially the same 8 ns delay due to the atmospheric gases but only about 0.05 ns due to the water vapor. (This difference is largely due to the fact that there aren't a large number of water vapor line tails adding together to affect the refractive index at optical frequencies.) In other words, assuming that the Smith-Weintraub Equation holds at optical frequencies leads to an error of about 1 nanosecond for propagation in the direction of the zenith. At 5 degrees elevation angle the error increases to about 10 ns.

Thus, two-way time transfers involving both optical and microwave signals could involve non-reciprocities near and above the 0.5 ns noise floors of present-day, two-way exchanges. Although I know of no such recent exchanges or any planned for the future, it is a point to keep in mind.

Absorption in the Troposphere

For all practical purposes atmospheric absorption of electromagnetic signals in the neutral atmosphere can be ignored unless broadcasting signals in or very near an absorption line--an unlikely situation. However, for the sake of completeness we shall consider such a possibility

Absorption itself will, of course, not lead to non-reciprocity. However it may lead to the appearance of non-reciprocity.

Consider a pulse whose initial shape, shown in Fig. 3, propagates through some absorbing medium where the absorption is frequency dependent. In this case it will emerge altered in shape as the figure shows. We will assume that the time reference point on the transmitted pulse, before absorption, is at the half power point also shown in the figure. The question is, "After the pulse has been distorted due to absorption, how does one determine the location of the time reference point on the received pulse?"

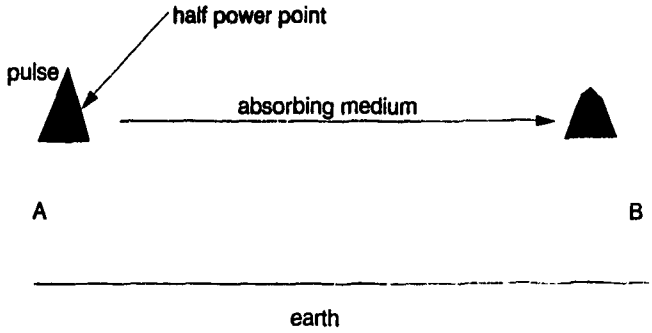


Fig. 3 Pulse distortion due to absorption.

There is no easy answer to this question. However, whatever method is chosen to define the time reference point at one receiver, the same procedure must be used at the other receiver. If different criteria are used then we will have the appearance of non-reciprocity.

In the next section, when we consider radio signals propagated through the ionosphere, I shall pick up this topic again.

Reciprocity in the Ionosphere

As in the neutral atmosphere, reciprocity, or the lack thereof, in the ionosphere depends on the form of the refractive index for an ionized medium imbedded in a magnetic field.

Before giving the appropriate expression for the refractive index in the ionosphere let's consider the important underlying phenomena for a moment.

As an electromagnetic signal propagates through the ionosphere the electrons in the ionosphere "see" the signal's fluctuating electric field and will thus move in some manner under the signal's influence. In addition, the detailed orbits of the electrons are constrained by the magnitude and configuration of the earth's magnetic field.

That this is indeed the case is demonstrated by the Appleton formula

$$\mu^2 = 1 - \frac{X}{1-iZ-[Y_T^2/2(1-X-iZ)] \pm \sqrt{[Y_T^4/4(1-X-iZ)^2 + Y_L^2}}$$

for the square of the refractive index μ in the ionosphere.

The term Z is proportional to the collision frequency between the electrons and the other particles, Y_L is the magnitude of the earth's magnetic field in the direction of propagation of the signal, the longitudinal component, Y_T is the magnitude of the earth's magnetic field transverse to the signal direction, and X is proportional to the ratio of electron density, N_e, to the square of the signal frequency, ν.

One thing we notice immediately, in contrast to the Smith-Weintraub expression, for the radio wave refractive index in the neutral atmosphere, is that the refractive index in the ionosphere is frequency dependent. In fact it varies with the reciprocal of the frequency squared. Thus, any two-way system operating with different up and downlink frequencies will not have reciprocity. Whether this is important or not depends on the desired accuracy of the time comparison. I shall return to this point later.

The other thing we notice in the refractive index formula is that it contains terms involving the longitudinal and transverse components of the earth's magnetic field. Here is a possible source of non-reciprocity. However, looking more closely, we see that all terms involving the magnetic field are even powers of the longitudinal and transverse components so that a reversal in propagation direction does not alter the value of the refractive index. That is, the earth's magnetic field does not introduce any non-reciprocity.

The Group Refractive Index

As we have seen, the refractive index in the ionosphere is frequency dependent. Because of this we must reconsider our notion of signal delay, L, as we defined it earlier.

Any finite duration signal can be decomposed into a packet of constituent signals with different frequencies- the well known Fourier decomposition. When the refractive index varies with frequency--when we have a dispersive medium--each Fourier component travels with a different phase velocity, v_p. In a dispersive medium we are interested in the speed with which the envelope of this group of Fourier components travels--not the speed of each individual component. Also, from a physical point of view, the speed of the envelope, the group velocity, v_g, is the speed with which the signal energy travels.

It is easy to show that the envelope travels with group velocity [3]

$$v_g = 2\pi d\nu/dk.$$

Since the group refractive index is

$$\mu' = c/v_g = (c/2\pi)dk/d\nu = \frac{d}{d\nu}(\nu\mu),$$

we can find the group refractive index for the ionosphere by substituting the Appleton formula in the above expression.

However, before we do this we may simplify the Appleton formula because in the gigahertz frequency range collisions between the electrons and the other atmospheric constituents can be ignored. With this simplification and after some manipulation we obtain for the Appleton group refractive index

$$\mu' = \mu + \frac{2}{2(1-X)-Y_T^2-Y_T^4-4Y_L^2(1-X^2)}\left[1-\mu^2-X^2+\frac{(1-\mu^2)(1-x^2)Y_T^2}{Y_T^4+4Y_L^2(1-X)^2}\right].$$

As before we see that all terms involving the magnetic field are raised to even powers, so 180° changes in the direction of signal propagation do not effect the group refractive index and thus there are no non-reciprocity effects involving magnetic fields.

In summary, then, in the ionosphere a magnetic field does not introduce a reciprocity problem, but frequency does.

As a final remark in this section recall that we calculated the delay, L, in the neutral atmosphere using the appropriate expression for the phase refractive index whereas in the ionosphere we need to use the group refractive index. The reason we use the phase refractive index in the neutral atmosphere and not the group refractive index is that in the neutral atmosphere the phase refractive index is not frequency dependent. In other words, the phase refractive index and the group refractive index are identical in the neutral atmosphere.

Ionospheric Non-Reciprocity at 14.300 and 12.007 GHz

The NIST-USNO two way satellite time transfer system ref [2] operates at 14.300 GHz on the uplink and 12.007 GHz on the downlink. This section presents some results indicating the magnitude of the ionospheric non-reciprocity we can expect at these frequencies.

In the gigahertz frequency range we can further simplify the Appleton formula to obtain

$$\mu' = 1-K_1\frac{I_N}{\nu^2} \pm \frac{K_2 I_N B\cos\theta}{\nu^3} - [K_3 I_N^2 + K_4 I_N B^2(1-\frac{1}{2}\sin^2\theta)]\frac{1}{\nu^4} + \ldots$$

Here I_N is the total integrated electron density along the propagation path, θ is the angle between the signal direction and the earth's magnetic field B, the K's are constants whose values depend on the units employed, and ν is signal frequency. As the equation shows the magnetic field contribution scales as the third and higher order powers of the inverse signal frequency so that we may neglect the magnetic field terms for the analysis that follows.

In these calculations we assume that the total integrated electron content in a vertical column is 1×10^{18} electrons per meter squared. This is a

large value, although near sunspot maximum, not an unheard of value.

At angles of propagation other than the zenith direction the path delay through the neutral atmosphere increases approximately as

$$\left[\csc[(E^2 + 20.3^2)] \right],$$

where E is the elevation angle in degrees.

Using this approximation Fig. 4 shows the residual non-reciprocity for up and downlink frequencies of 14.300 and 12.007 respectively. As the figure shows, at low elevation angles, the non-reciprocity approaches the noise floor of the present two-way exchanges. However, if history is a reliable guide, the noise floor with better techniques will continue to diminish and finally be of the same magnitude or less than the residual non-reciprocity.

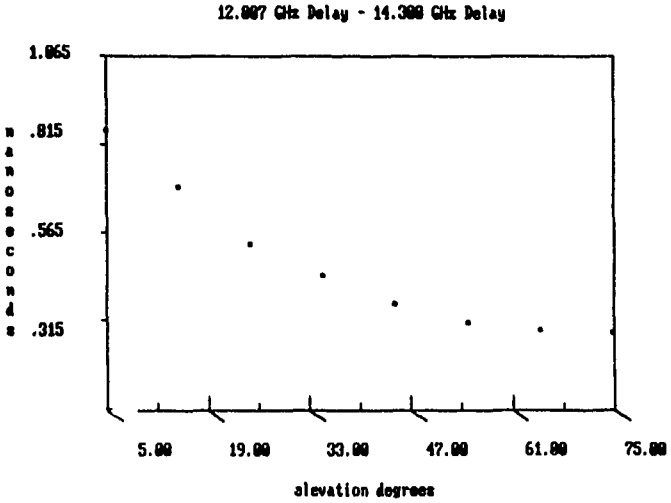


Fig. 4 Path delay non-reciprocity for 12.007-14.300 GHz link corresponding to extreme total integrated electron content.

Of more immediate interest is the fact that many communications satellites operate in the C Band where the uplink frequency is 6 GHz and the downlink frequency 4 GHz. Figure 5 shows the non-reciprocity variation with an elevation angle that we would expect with this pair of frequencies. We have again assumed a total electron content corresponding to daytime maximum sunspot number.

Here we see that the non-reciprocity is well above the measurement noise floor even at high elevation angles. Under more normal circumstances we would expect all the y axis values in this figure to be reduced by about one order of magnitude. But even then we are still above the measurement noise floor at low elevation angles and approach the noise floor as the path direction approaches the zenith.

These results suggest that non-reciprocity should be scrutinized when communications satellites operating at C Band are used in two-way time exchanges. In the next section I suggest a procedure that might be applied to correct for ionospheric delay in circumstances where it cannot be ignored.

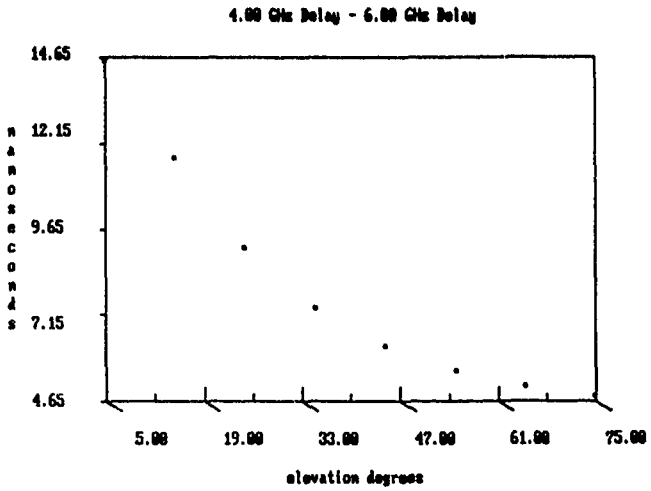


Fig. 5 Path delay non-reciprocity for 4.00-6.00 GHz link corresponding to extreme total integrated electron content.

Measuring the Non-Reciprocity

More often than not in two-way time exchanges the up and downlinks frequencies are not under the control of the participants in the exchange but are determined by available satellites. Thus, there may be ocassions where non-reciprocity effects are not below the noise floor of the measurement--as the discussion of C Band communication satellites in the previous section shows. In a more extreme case one link might be in the VHF range and the other in the microwave region.

Under some circumstances it may be possible to measure the ionospheric path delay if the signals relayed by the satellite can also be received at the sites initiating the signals. The idea explorod here is a variation on the two frequency scheme developed for measuring the ionospheric delay for the Global Positioning Systems (GPS) signals.

Consider the situation shown in Fig. 6. As we have seen the group velocity in the ionosphere is frequency dependent so that the extra delay due to the ionosphere at frequency, ν, is of the form

$$T = \frac{KI_N}{\nu^2}$$

where K is a constant depending on the units and I_N, as before, is the total integrated electron content along the path from A to the satellite.

Consider two signals at ν_1 and ν_2 traveling to the satellite from location A and then transponded back to A at frequencies ν_3 and ν_4. If, on the uplink, two pulses are transmitted (simultaneously, one on the carrier at ν_1 and the other on the carrier at ν_2), from A, they will arrive at the satellite separated in time by an amount,

$$\Delta T_1 = KI_N \left[\frac{1}{\nu_1^2} - \frac{1}{\nu_2^2}\right]$$

as shown schematically in Fig. 6.

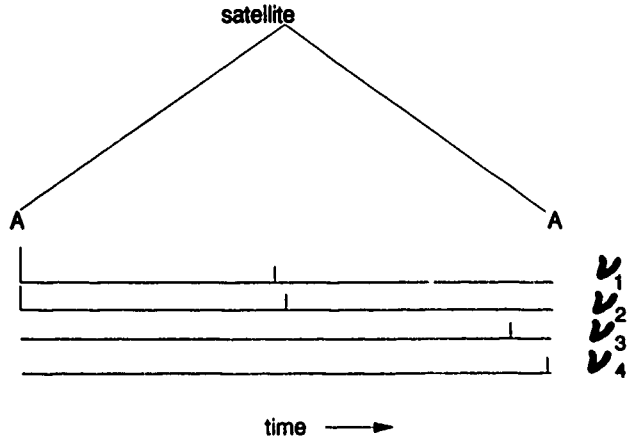


Fig. 6 Schematic representation of pulse delay spreading on transponder up and downlinks to the same location.

These same two pulses transponded back on the downlink signals at frequencies ν_3 and ν_4 will arrive back at A separated by an amount.

$$\Delta T_2 = \Delta T_1 + KI_N \left[\frac{1}{\nu_3^2} - \frac{1}{\nu_4^2}\right]$$

which we can rewrite in the form

$$\Delta T_2 = KI_N \left[\frac{1}{\nu_1^2} + \frac{1}{\nu_3^2} - \frac{1}{\nu_2^2} - \frac{1}{\nu_4^2}\right].$$

Then solving for the total integrated electron content, I_N, we have

$$I_N = \frac{\Delta T_2}{K\left[\frac{1}{\nu_1^2} + \frac{1}{\nu_3^2} - \frac{1}{\nu_2^2} - \frac{1}{\nu_4^2}\right]}.$$

Now, knowing the total electron content we can determine the ionospheric signal path delay at any frequency we desire.

In a practical case the two uplink signals at frequencies ν_1 and ν_2 could be at the upper and lower frequency limits of the uplink transponder and similarly ν_3 and ν_4 could be at the upper and lower frequency limits of the downlink transponder. Of course, whether this scheme would work in practice depends on a number of factors including the transponder characteristics and the satellite transmitted power.

As a practical illustration of the method discussed here consider a C Band communication satellite operating on 4 and 6 GHz. Typically, the transponder bandwidth on these satellites is 30 MHz so we shall assume this figure for the frequency separation between ν_1 and ν_2 on the uplink and ν_3 and ν_4 on the downlink. Again, under daytime, sunspot maximum conditions Fig. 7 shows the total differential delay, ΔT_2, we would expect to find for elevation angles ranging from 5 to 45°. These results suggest that, with care, we can measure the total integrated electron content for C-Band transponders in circumstances where it is important.

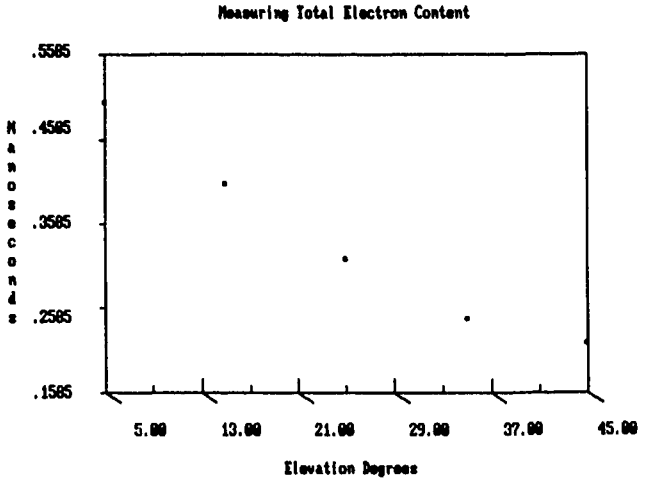


Fig. 7 Variation of pulse delay spreading, ΔT_2, with elevation angle of the satellite. The figure corresponds to extreme total integrated electron content for up and downlinks of 6.00 and 4.00 GHz respectively.

Absorption in the Ionosphere

As stated earlier, absorption in the ionosphere is produced by free electrons colliding with neutral particles. At gigahertz frequencies we have been considering absorption effects may be ignored. However, for the sake of completeness I shall mention one circumstance under which absorption can be a problem.

The problem arises in connection with two-way exchanges involving low frequency signals propagated over long paths by the ionosphere.

At frequencies of 10 kHz and lower, where the wavelength is 30 km and longer, the actual orbits of the electrons as they collide with neutral particles becomes important in computing absorption. Detailed analysis shows that the electron orbits have a smaller radius in one direction of propagation than they do in the reversed direction. The net effect is that for one direction there is more absorption of the wave than there is in the other. The effect is most pronounced for east-west propagation paths where the absorption is greater for signals propagated to the west than it is to the east.

Here again we can have the kind of problem mentioned earlier in the section on absorption in the troposphere. That is when the pulse shape is altered due to absorption we can no longer be sure where the time reference point is. The problem is compounded for low frequency waves reflected to earth by the ionosphere because now the absorption depends on the direction of propagation. Now not only are the pulse shapes altered by absorption but they are altered in different ways depending on the direction of propagation, thus magnifying the problem of identifying the time reference point on the received wave.

Conclusions and Summary

For present two-way satellite time transfer systems operating in the 12.00-14.00 GHz frequency range, propagation non-reciprocity is not ordinarily a problem. Sample calculations show that under typical conditions non-reciprocity effects are about an order of magnitude below the 0.5 ns noise floor of current systems. However, at low elevation angles when the integrated electron density is unusually high, daytime sunspot maximum conditions for example, non-reciprocity begins to approach the two-way noise floor.

For two-way satellite time transfers in the 4.00-6.00 GHz region (C-Band), path non-reciprocity cannot be ignored when the electron content is unusually high at any elevation angle and can be a problem at low elevation angles even under normal conditions.

Under certain circumstances it may possible to correct for ionospheric non-reciprocities using a system analogous to the GPS procedure for correcting for ionospheric delay.

Acknowledgement

I thank D. W. Hanson and W. J. Klepczynski for several comments and suggestions.

References

[1] E.K. Smith, Jr. and S. Weintraub, "The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies," Proc. IRE, 41, 1035-1037 (1953).

[2] D. W. Hanson, "Fundamentals of Two-Way Transfers by Satellite," these proceedings.

[3] Davies, Ionospheric Radio Waves, p. 31, Blaisdell Pub. Co., 1969.

43rd Annual Symposium on Frequency Control - 1989

NIST-USNO TIME COMPARISONS USING TWO-WAY SATELLITE TIME TRANSFERS

D. A. Howe, D. W. Hanson, J. L. Jespersen and M. A. Lombardi
National Institute of Standards and Technology (NIST)
325 Broadway, Boulder, CO 80303

W. J. Klepczynski, P. J. Wheeler, M. Miranian, W. Powell,
J. Jeffries, and A. Myers
United States Naval Observatory (USNO)
Washington, DC 20392-5100

Abstract

NIST and USNO began making two-way sat llite time transfers on August 12, 1987. The time transfers are made using the SBS-3 satellite and take place three times per week. The paper describes the equipment used to make the transfers, the data reduction procedures, and the results obtained.

Introduction

Several two-way time transfer experiments using geostationary satellites have been done over the years [1]. Recent studies of the phase stability of satellite loop tests using present-day equipment have shown that equipment-delay reproducibility over two weeks was 1 ns [2]. Using this same equipment, NIST and USNO have been making routine two-way satellite time transfers since August of 1987.

During the two-way time transfer, NIST and USNO simultaneously transmit and receive two spread-spectrum signals. Both signals are initiated by a 1 pps clock signal that represents UTC(NIST) and UTC(USNO) respectively. The time difference between the local transmitted 1 pps and the received 1 pps from the other station is measured and recorded at each site for a 300 s (5 min) interval. The time differences for each 1 pps epoch are recorded at both sites and stored on a central computer at USNO. These data are then aligned for matching epochs, and the difference between the two values is obtained and then divided by 2. The resulting data show a second-by-second comparison between NIST and USNO.

The mean and standard deviation are also computed for each 300 s measurement run to obtain a single estimate of the time scale difference. Regression analysis shows no discernible slope above the residual white noise, and tests of the residuals show that they follow white noise behavior.

The internal signal delay in each spread-spectrum modem (one at NIST and one at USNO) is measured and recorded before starting each measurement run. These delays are subtracted from the average raw time comparison to obtain an uncalibrated final comparison. Plots of these results are presented in this paper. A full calibration requires subtracting the earth stations' differential time delays, which has not yet been measured, and subtracting the Sagnac effect.

Data reduction was originally done using regression analysis. The regression coefficients were exchanged and the modem delays subtracted to obtain an uncalibrated final comparison. However, this method was sensitive to satellite motion.

Although SBS-3 is geostationary, there is some nearly linear motion which can amount to a few nanoseconds per second during the measurement run. The second-by-second comparison has proved to be a cleaner and simpler method of data reduction than regression analysis.

Equipment

The equipment used at NIST and USNO operates in the Ku-band. The configuration at each location is the same except that NIST has a 6.1 m dish with remote positioning capability and USNO has a 4.5 m dish with a fixed mount. Figure 1 shows a diagram of the principal earth station components. Reference 2 contains a description of the earth station equipment at NIST. The earth station operates at an uplink (transmitting) frequency of 14.307 GHz and downlink (receiving) frequency of 12.007 GHz on SBS-3, transponder 7 lower, which is in geostationary orbit at 95°W.

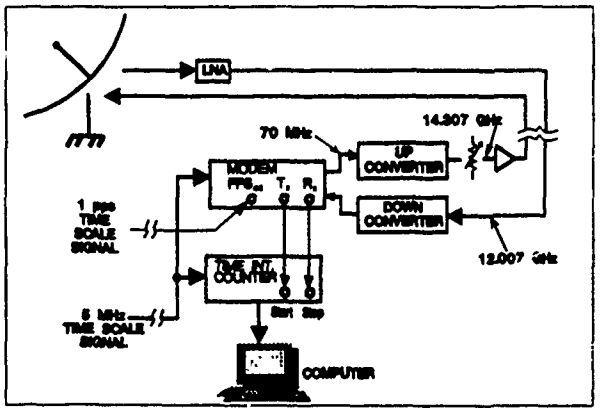


Fig. 1. Earth station equipment.

The 1 pps signal is transmitted from each earth station and received by the other earth station by means of a spread-spectrum modem that operates at 70 MHz [3]. Since orthogonal spread-spectrum code sequences are used, the simultaneous transmissions do not interfere with each other.

Data Acquisition and Analysis

Satellite time on SBS-3 is scheduled for 30 min on Monday, Wednesday, and Friday mornings beginning at 10:30 am eastern time. Raw data are the time differences between a local clock's 1 pps signal and the 1 pps signal received from the other station and demodulated by the spread-spectrum modem (see fig. 2). The once-per-second time interval counter readings at NIST and USNO are respectively

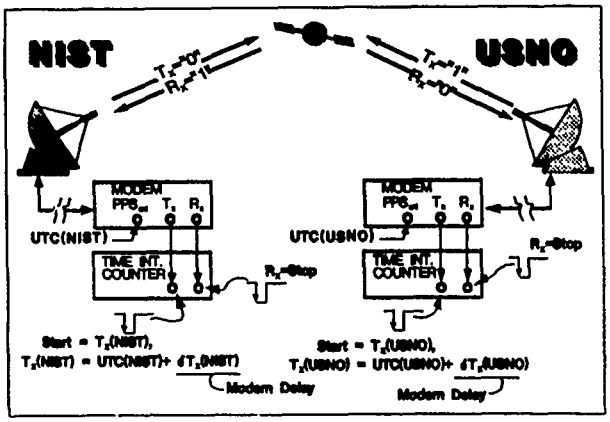


Fig. 2. Each spread-spectrum modem transmits a 1 pps signal(Tx) to the other earth station location via SBS-3. The Tx signal is slightly delayed from the external 1 pps (UTC(NIST) and UTC(USNO)) by δTx.

$$TI(NIST) = Tx(NIST)-[Tx(USNO) + UpC(USNO) + Sat.path(EW) + DnC(NIST)] \quad (1)$$

$$TI(USNO) = Tx(USNO)-[Tx(NIST) + UpC(NIST) + Sat.path(WE) + DnC(USNO)] \quad (2)$$

where "Tx" is the 1 pps signal as sampled on the front panel of the modem, "UpC" and "DnC" are the upconverter and downconverter equipment delays at either NIST or USNO, and "Sat.path" is the signal path up to, through, and down from the satellite going from East to West (EW) or West to East (WE). Sat.path includes a nonreciprocal delay due to the rotating satellite-earth system (the Sagnac effect). Subtracting equations and rearranging, we obtain

$$TI(NIST)-TI(USNO) = 2Tx(NIST) - 2Tx(USNO) + 2R(1) \quad (3)$$

$$\text{where } 2R(1) = [UpC(NIST)-DnC(NIST)] - [UpC(USNO) - DnC(USNO)] - [Sat.path(EW) - Sat.path(WE)]. \quad (4)$$

Thus,

$$Tx(NIST)-Tx(USNO) = \tfrac{1}{2}[TI(NIST)-TI(USNO)] - R(1). \quad (5)$$

The designation "1" for R refers to the NIST-USNO link. There is a combined cable delay and internal modem delay between the local clock and "Tx," the 1 pps signal actually transmitted by the modem. We shall denote this delay as δTx, where

$$\delta Tx(NIST) = UTC(NIST) - Tx(NIST) \quad (6)$$

$$\text{and} \quad \delta Tx(USNO) = UTC(USNO) - Tx(USNO). \quad (7)$$

Combining these equations into Eq. 5 and rearranging, we obtain

$$UTC(NIST)-UTC(USNO) = \tfrac{1}{2}[TI(NIST)-TI(USNO)] - [\delta Tx(USNO)-\delta Tx(NIST)]-R(1). \quad (8)$$

This equation is the basis for the data reduction for the time transfers between NIST and USNO. It contains the raw counter readings, the cable and modem delays (calibration delay), and the nonreciprocal equipment and Sagnac effect R(1) for this specific communications link. Taking one-half the raw counter readings between NIST and USNO relates directly to the difference of the time scales. The calibration delay (difference of the modem delays) remains constant within ±100 ps for a 30 min interval and therefore is essentially a constant during the time transfer. R(1) is unknown but will be determined by using a mobile earth station and by calculating the Sagnac effect.

Several auditional stations will participate in two-way satellite time transfers in the future. Comparisons will be performed between two locations at a time designated as the I^{th} two-way link. With regard to the raw difference data, a generalized expression for an I^{th} link involving locations "A" and "B" is

$$M_k(I) = \frac{TI_k(A)-TI_k(B)}{2}. \quad (9)$$

Although this paper discusses only one link (USNO-NIST), this generalization is useful for future reference. We refer to the raw difference data between NIST and USNO as

$$M_k(1) = \frac{TI_k(NIST) - TI_k(USNO)}{2}, \quad (10)$$

or one-half the k^{th} difference of the time interval counters at NIST and USNO. The designation "1" for M_k refers to the USNO-NIST link. Five minutes worth of data (300 readings) are collected at a time. The time-interval counter readings at both USNO and NIST are stored by PC-type computers. A large computer at USNO is used for data preparation, archiving, and retrieval. Preparation consists of obtaining the computer files by telephone modems, aligning the data so that 1 pps epochs match from each station, and computing each 1 s difference and dividing by 2 (computing $M_k(1)$). The spread-spectrum modem calibration delay is also part of the file and is archived in a separate file at USNO. Both USNO and NIST can retrieve the data from the USNO computer.

Typical 1 s phase stability $\sigma_y(1s)$ is between 3 and 20×10^{-10}, with C/N_0 between 65 and 55 dB-Hz at both USNO and NIST. The range of signal levels encountered is dependent on weather and antenna conditions, shared satellite traffic, and antenna pointing error. C/N_0 (carrier-to-noise density ratio) is a general figure of merit parameter used by the satellite communications industry to describe a link [4].

Figure 3 shows typical plots of the second-to-second difference, $M_k(1)\Big|_{k=1}^{300}$.

Spectral analysis of this kind of data shows that the power spectrum is consistently white over the range of frequencies between 1/2T, or 1.67 mHz, and 1/2t, or 1/2 Hz (the Nyquist frequency), where T is the entire measurement duration (here, 300 s) and t is the minimum sampling interval (here, 1 s). Figure 4 shows a plot of a power spectrum on log-log scale taken from typical data. The method of spectral analysis is a direct, or non-parametric

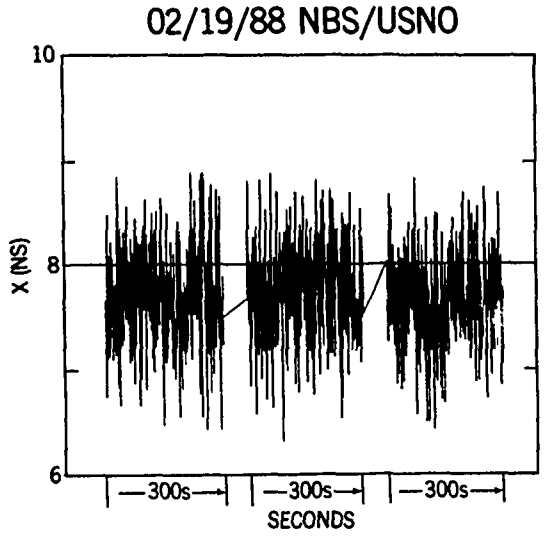


Fig. 3. Example plots of raw time-difference data. Three plots of 300 seconds (5 min) each are shown with a separation between each measurement series of 1 min.

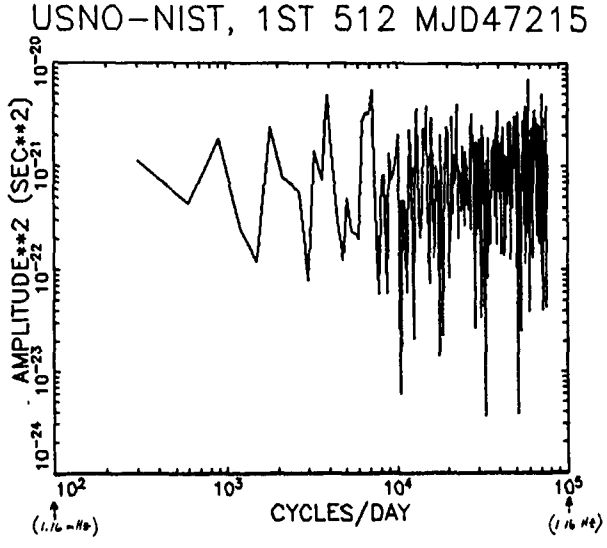


Fig. 4. Typical spectrum of raw time-difference data. This is a direct estimation of the power spectrum using the Fourier transform of the raw data.

estimation in which the Fourier transform of the 300 s time series $M_k(1)$ is computed. For confirming the presence (or lack thereof) of bright lines in the power spectrum, the highest resolution is obtained using an unsmoothed direct method of the entire data set as an appropriate first approximation of the spectrum. This first-level analysis is biased, however, because the Fourier transform of the "window" function (here, a simple rectangle) results in a passband filter which follows a sin(f)/f function around the discrete frequency intervals of the spectrum. This bias is termed "leakage" [5,6]. We are less concerned with bias and more concerned with establishing that the time series $M_k(1)$ is indeed white since the variance can easily be computed classically. Although the unsmoothed spectrum estimate of Fig. 4 is noisy (as noisy as the original data), the distribution of power over frequency f is fairly constant.

The simple mean of $M_k(1)$ is used as the value of the raw time comparison. The level of the spectrum is an estimate (although biased, as explained in the preceding paragraph) of the variance of $M_k(1)$, and its square-root is the standard deviation with respect to the mean. The standard deviation is the measurement precision. For every scheduled transfer, it is simpler to compute the classical standard deviation of the mean to get a measure of precision for that transfer. Typical precisions range from 200 ps (high C/N_0) to 1.5 ns (low C/N_0).

Averaging data for about 100 s exceeds the performance specifications of limiting components, even with transfers having low C/N_0. This is so because the poorest $\sigma_y(\tau)$ is of order $2 \times 10^{-9}\tau^{-1}$. Therefore, the standard deviation of the mean is 115 ps for a 100 s interval. In the best case, the time-interval counter at each location has an inaccuracy of ± 35 ps rms. Furthermore the 1 pps signal from the NIST computer-controlled microstepper has an error of 350 ps rms relative to UTC(NIST), although it is usually better. A test of the round trip delay through the earth station equipment in an in-cabinet test showed reproducibility over two weeks to the level of 1 ns [2]. From this we could expect the "differential" delay to be reproducible to a better level, but it would be presumptuous to say it is reproducible to less than 100 ps without a specific test of this.

Another observation is that regression analysis of $M_k(1)$ shows no discernible linear slope above the measurement precision in 300 s. Daily observations show the rate difference between UTC(USNO) and UTC(NIST) to be several nanoseconds per day (actual data will be given in the next section). This amounts to 30 ps during a 300 s interval. Short-term clock phase fluctuations are expected to be lower than 200 ps/s rms. Therefore, simply computing the mean value of $M_k(1)$ is sufficient for obtaining a time transfer from a second-to-second comparison of 100 s to 300 s given typical two-way transfer precisions.

Results

Second-to-second records of time-interval readings taken simultaneously at NIST and USNO started in mid-February of 1988. A plot of time comparison results since that time is shown in Fig. 5. This plot represents $\bar{M}_k(1)$, or the mean value of the second-to-second differences (300 s per measurement) divided by 2, and corrected by the spread-spectrum modem delays, δT(NIST) and δT(USNO). There is no correction for the differential equipment delay and Sagnac effect (collectively as R(1) in eq. 8) and for cable delays from the UTC signal point of origin (NIST and USNO) to the corresponding spread-spectrum modem. The cable delays are incorporated when a determination of R(1) is completed.

Figure 6 is the two-sample, or Allan, variance $\sigma_y(\tau)$ of the USNO-NIST master clock comparisons plotted in Fig. 5. From 1 d to 10 d, $\sigma_y(\tau)$ behaves as white frequency noise with a level of $1.35 \cdot 10^{-11}$ $\tau^{-\frac{1}{2}}$. The 1 pps estimate of UTC(NIST) is derived from a commercial Cs-standard member of the NIST clock ensemble whose output is fed to a microstepper which is updated every 12 min. Therefore the short

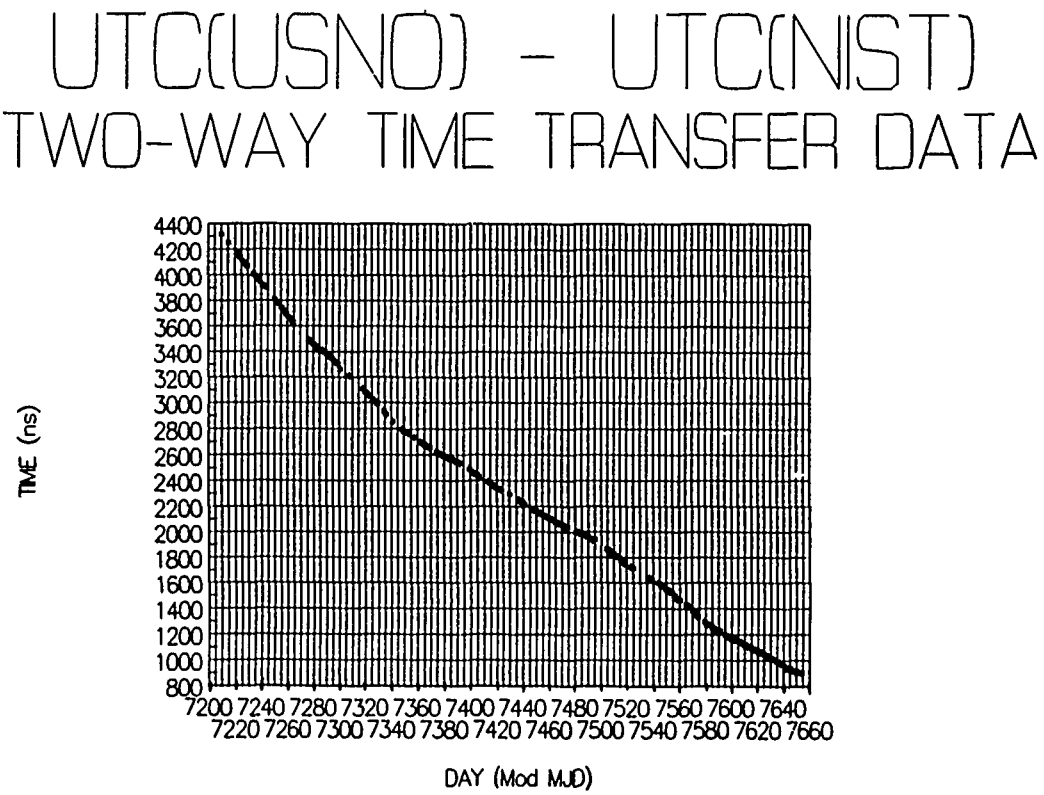


Fig. 5. Over one year of time comparison results taken at scheduled satellite availabilities beginning in mid-February of 1988. Differential signal delays due to earth station equipment and Sagnac effect have not been accounted for.

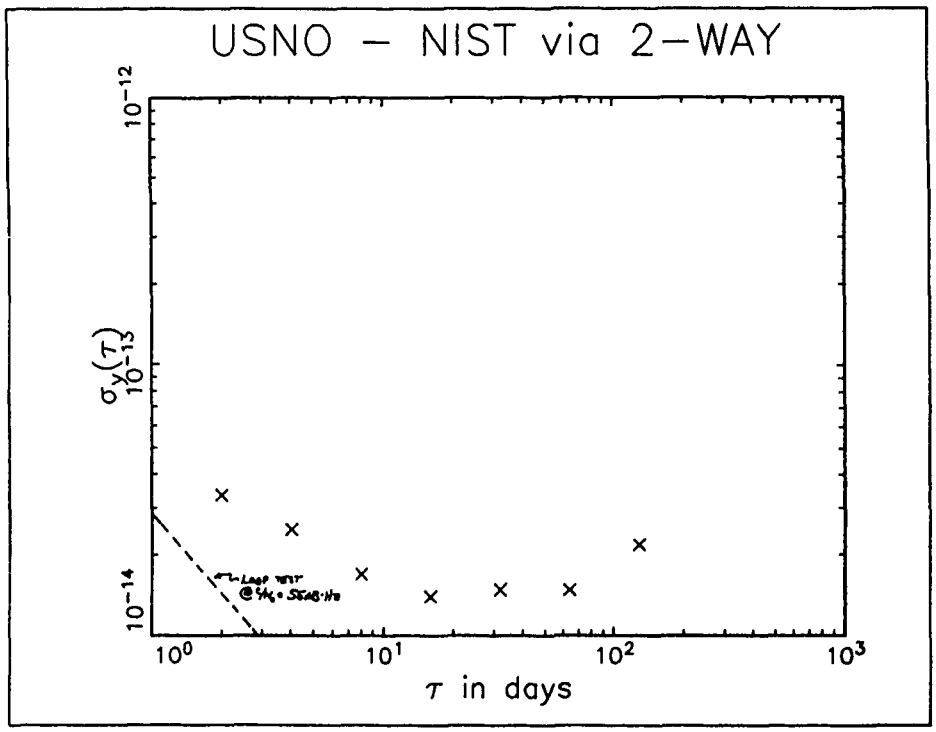


Fig. 6. Plot of $\sigma_y(\tau)$ of data shown in fig. 5. From 1 to 10 d, stability goes as $1.35 \times 10^{-11}\tau^{-\frac{1}{2}}$, which implies white frequency behavior. From 10 to 100 d, stability is 1.4×10^{-14} flicker of frequency. The rightmost point at 2.1×10^{-14} is due to steering corrections applied to UTC(USNO) and UTC(NIST).

term noise of this 1 pps signal is the noise of the commercial Cs standard plus the small steering corrections applied every 12 min. The 1 pps estimate of UTC(USNO) is derived directly from a hydrogen maser which is steered by its frequency synthesizer in time periods of a few days. In-cabinet loop test data of $\sigma_y(\tau)$ of two-way satellite equipment through a satellite simulator is plotted in Fig. 6 [2]. This $\sigma_y(\tau)$ level of $2.0 \cdot 10^{-9}\tau^{-1}$ was the measured result at C/N_0 of 55 dB-Hz; usually the USNO-NIST time transfers have a C/N_0 better than this.

The NIST-USNO time-transfer procedure, although routine, is still in the experimental stage and some equipment was not the same throughout the 1½ years of operation that these data represent. There was every effort to maintain an unperturbed set up and procedure at NIST and USNO, however; there were changes and occasional malfunctions along the way. Since transfers were completed every 2 or 3 d, unknown phase-delay variations of as much as a few nanoseconds are within the full set of data.

In longer term, the level of the flicker of frequency noise extends from roughly 10 to 100 d and shows a level of $1.4 \cdot 10^{-14}$. This agrees with expected results and in-cabinet loop tests show equipment limits to be in the 10^{-15} region, a factor of 10 better. There is therefore no reason to assume any degradation due to the two-way satellite system, and this flicker level is agreeably due to UTC(NIST) and UTC(USNO).

Beyond 100 d, long-term rate corrections of UTC(NIST) and UTC(USNO) account for the increase in $\sigma_y(\tau)$ in Fig. 5 (rightmost point) to $2.1 \cdot 10^{-14}$. Granted there is poor confidence for this point; nevertheless the increase in $\sigma_y(\tau)$ is consistent with noise resulting from these long-term corrections. If a second-order least-squares fit is subtracted from the data, this point drops to $1.6 \cdot 10^{-14}$. Figure 7 shows the time comparison results of Fig. 5 with removal of a first difference of -6.29 ns/d, second difference of 0.003 ns/d^2, and mean of 221 ns.

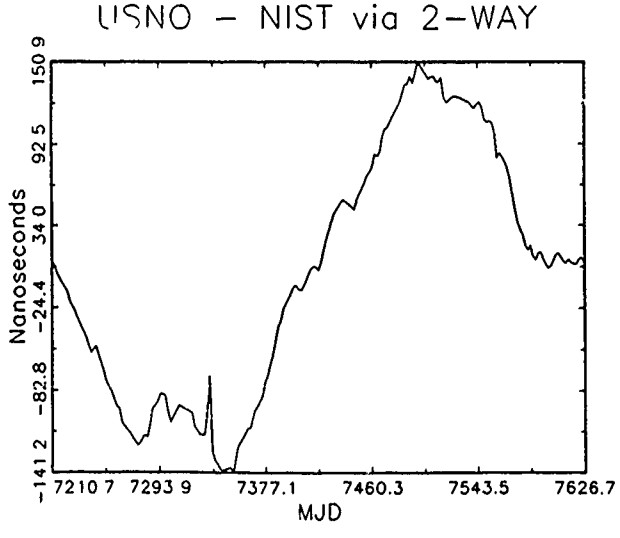


Fig. 7. Time comparison data of fig. 5 with first and second difference and mean removed.

Routinely scheduled two-way time comparisons between NIST and USNO began in August, 1987. The data reduction now involves the exchange of all 300 time-interval counter readings and corresponding times in hours-minutes-seconds. The initial procedure involved regression analysis of the 300 time-interval counter readings taken at each location simultaneously. A fourth-order polynomial was fitted to the readings, and a difference of the coefficients was computed to obtain a mean raw time difference. Curve fitting is less satisfying than taking the simple average of second-to-second differences because uncertainty in the polynomial fit is not readily quantified even by analysis of white-noise residuals. Figure 8 is a plot of the difference between time comparisons using the curve-fit procedure vs. the second-to-second average for a series of days in which both procedures were used. This difference never exceeded ± 1 ns.

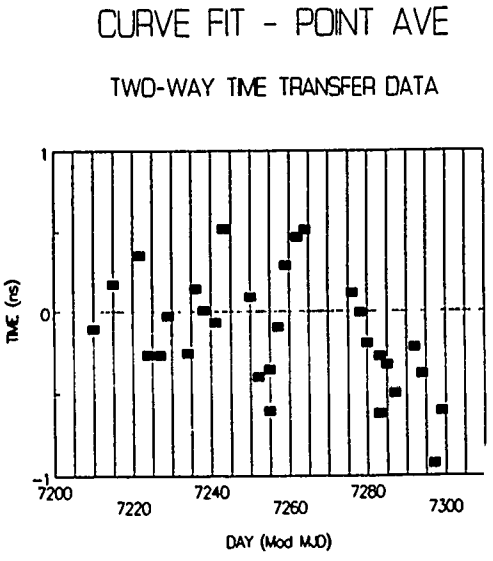


Fig. 8. Difference between reduction of raw data using curve fitting and using the average of second-to-second differences, $\overline{M}_k$.

Conclusion

Time comparison data between UTC(USNO) and UTC(NIST) has been presented along with some analysis of that data. A measurement of the offset time delay due to the earth stations will be performed in the near future so the comparisons here remain uncalibrated. Second-to-second raw difference data is white with no discernible slope for a 300 s measurement, and a simple mean is computed for the value of the time comparison. Long-term $\sigma_y(\tau)$ results are consistent with expected results of a comparison between NIST and USNO, and in-cabinet loop tests of the two-way satellite equipment point to system noise being a factor of 10 below the stability of these results. Therefore the noise contribution due to the two-way time transfer method is assumed to be negligible. Our results show that the measurement time could be reduced from 300 s to 100 s with little or no compromise in precision even with a C/N_0 of as low as 55 dB-Hz.

Acknowledgements

The assistance of M. Weiss and J. Gray are greatly appreciated.

References

[1] See, for example, A. R. Chi and E. Byron, "Two-way time transfer experiment using a synchronous satellite," in Proc. 7th Annual Precise Time and Time Interval Planning Meeting, Dec. 1975; also C. C. Costain et al., "Two-way time transfer via geostationary satellites NRC/NBS, NRC/USNO, and NBS/USNO via Hermes and NRC/LPTF (France) via Symphonie," in Proc. 11th Ann. Precise Time and Time Interval Planning Meeting, Dec. 1979.

[2] "Progress toward one-nanosecond two-way time transfer accuracy using Ku-band geostationary satellites," D. A. Howe, IEEE Trans. on Ultrasonics, Ferroelectrics, and Freq. Control, Vol. UFFC-34, No. 6, Nov. 1987.

[3] Mitrex 25000 documentation, P. Hartl et al., Institute for Navigation, Univ. Stuttgart, Germany, Jan. 1985.

[4] Digital Communications Satellite/Earth Station Engineering, K. Feher, Prentice-Hall Inc., Englewood Cliffs, NJ, 1983.

[5] Spectral Analysis for Physical Applications, Percival and Walden, to be published.

[6] "Pitfalls in digitizing the data," D. A. Howe, D. W. Allan, and J. A. Barnes, in Frequency Stability: Fundamentals and Measurement, edited by Vencesler F. Kroupa, IEEE Press (1984), PC01644, pp. 177-186.

43rd Annual Symposium on Frequency Control - 1989

COMPARISON OF TWO-WAY SATELLITE TIME TRANSFERS

AND

GPS COMMON VIEW TIME TRANSFERS

W. J. Klepczynski, M. Miranian, J. Jeffries,

P. J. Wheeler and W. Powell

U.S. Naval Observatory
34th & Massachusetts Ave., NW
Washington, DC 20392-5100

ABSTRACT

The purpose of this study is to compare the results of remote clock synchronization obtained by two different techniques, i.e., GPS Common View and Two-Way Satellite Time transfers. Values for UTC(USNO)-UTC(NIST) published by the Bureau International des Poids et Mesures (BIPM) and the National Institute of Standards and Technology (NIST), based on GPS Common View data,are compared with values obtained by Two-Way Satellite Time Transfers. The differences between the values published by NIST and those obtained by the Two-Way Satellite technique do not seem to be significant. The differences between the Two-Way Satellite values and those published by the BIPM are probably due to truncation and the way the BIPM smoothes the GPS Common View data which form the basis of their values.

I. INTRODUCTION

During the last 14 years, the Two-Way Satellite Time Transfer Method using geostationary communications satellites has matured from an experimental time transfer system of great promise [1] to an operational system of fulfillment [2]. Both early and current data indicated extremely high precision in the time transfer values obtained by this technique. While the internal precision was great, nothing still could be said of its external precision [3]. This study is an attempt to compare the results of the Two-Way Satellite Technique with that of two different approaches to the GPS Common View Technique to get an estimate of the relative precision of the three techniques. The estimate will be obtained by the Three-Cornered Hat Technique [4], i.e., three different estimates of a quantity will be inter-compared in order to obtain an estimate of the precision of each estimate.

Each of the three processes being inter-compared obtain estimates for the clock differences between two locations. In the case of the Two-Way Satellite Technique discussed here, the clocks at the two locations are physical realizations of the time scales denoted as UTC(USNO,MC) and UTC(NIST,MC). With regard to the NIST GPS Common View data, the values tabulated are UTC(USNO,MC)-UTC(NIST), where the values of UTC(NIST) refer to the smoothed time scale UTC(NIST) and not to its physical realization UTC(NIST,MC). Similarly, the values published by the BIPM using GPS data are the differences UTC(USNO)-UTC(NIST) which are affected by a smoothing process. The procedures used for smoothing are known via private communications [5,6], but the minute details are not readily available Therefore, for a common period of time, when the differences for UTC(USNO)-UTC(NIST) were available from all three sources, an analysis was performed which gives an indication of the external precision of the data and, consequently, a reflection of the techniques and the procedures used in the data reduction procedure.

II. TWO-WAY SATELLITE TIME TRANSFERS

The Two-Way Satellite Time Transfer Method gives an almost immediate estimate of the differences between the two clocks at each terminal in the network with about five minutes of observation after an exchange of data between the stations involved. The internal precision of the technique currently varies between 300 and 700 picoseconds. The reason for these remarkable numbers is that it is not the received signal which is compared against the local 1 PPS but the signal from a crystal which is phase locked to the received signal. In the case of the two clocks used in this experiment, the one at USNO is a Smithsonian Astrophysical Observatory (SAO) Hydrogen Maser which realizes the mathematical Time Scale UTC(USNO). At NIST, the clock is a Cesium Beam Frequency Standard which is steered to UTC(NIST) every twelve minutes.

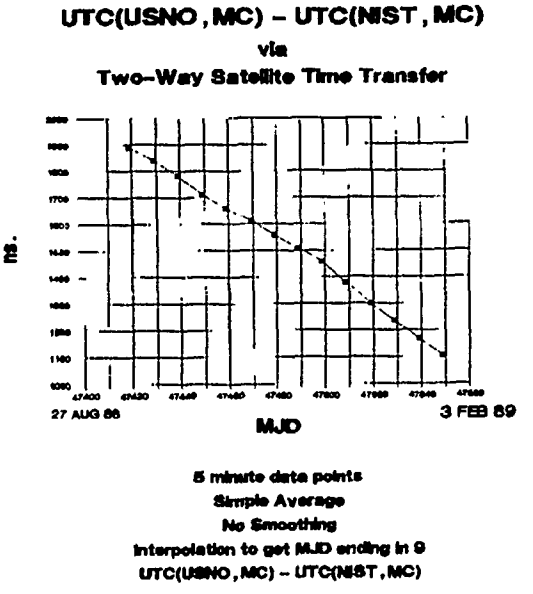


Figure 1.

Figure 1 shows the difference UTC(USNO)-UTC(NIST) as obtained from the Two-Way Time Transfer data. Several things should be pointed out. First, each data point represents only about five minutes of

observing time. Second, because these data will be compared against other data which are listed for MJD's which end in a 9, simple linear interpolation was performed across several points that were on each side of the date in need. Figure 1 then represents the first set of data needed for application of the Three-Cornered Hat Technique.

III. BIPM COMMON VIEW DATA FOR UTC(USNO)-UTC(NIST)

Every month, the BIPM releases their estimates of UTC(BIPM)-UTC(Laboratories) in their Circular T. Values are listed for each MJD which ends in a 9. Normally the publication lists values which are two months in arrears. The values for UTC(BIPM)-UTC(Lab) are derived from GPS Common View data. The data for UTC(USNO)-UTC(NIST) are inferred from the published data. The original data have been smoothed by a weak Vondrak smoothing algorithm '. Furthermore, the data are only released to the nearest 10 nanoseconds. Figure 2 exhibits these data.

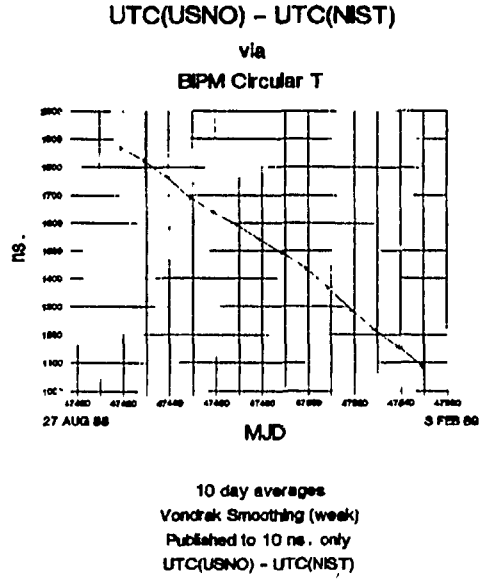


Figure 2.

IV. NIST COMMON VIEW DATA FOR UTC(USNO)-UTC(NIST)

Each month, NIST publishes its time and frequency data. It contains estimates for UTC(NIST)-UTC(Lab). These estimates are also obtained from GPS Common View data. It is probably correct to assume that the GPS data used by NIST are almost identical to the data used by the BIPM. However, it should be pointed out that NIST uses a different optimized smoothing filter to estimate their values of UTC(NIST)-UTC(Lab). Thus, it is difficult to say whether the two data sets are strongly correlated. Figure 3 shows the difference UTC(USNO)-UTC(NIST) as obtained from the NIST publications.

V. THREE-CORNERED HAT ANALYSIS

By inter-comparing the three different data sets, it should be possible to derive an estimate for the external error or a measure of consistency between the three sets by the method known as the Three-Cornered Hat Technique. The procedure

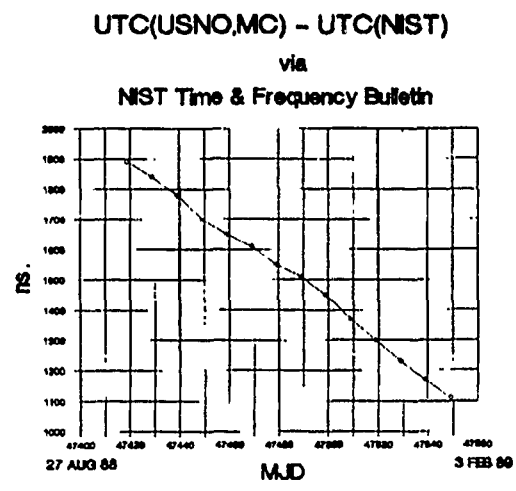


Figure 3.

adopted here is similar to that of finding the stability of a single oscillator through its comparison with two other oscillators '. First, the differences between the estimates for UTC(USNO)-UTC(NIST) from the data in the NIST publications and the data in the BIPM publications are formed. This set will be called (ab), where a refers to the BIPM data and b refers to the NIST data. A linear regression is performed through the differences. Residuals are then formed and an estimate for the variance is made. Similarly, the differences between the BIPM data and the Two-Way data are formed. This set is called (ac), where c refers to the Two-Way data set. The variance for this set is also formed. Finally, the differences of the NIST data and the Two-Way data are formed and estimates for its variance is obtained after a linear regression is fit through the differences. This data set is called (bc). The data so obtained are exhibited in Figure 4. The variance for one of the data sets is obtained as shown in Figure 5, keeping in mind that the variance is the square of the standard deviation. Figure 6 summarizes the resulting standard deviations for each of the three data sets.

VI. DISCUSSION AND CONCLUSIONS

The data in Figure 6 show that the NIST data set has a standard deviation (external precision in this case) of about 1 ns. The Two-Way data set has a standard deviation of about 2 ns. and the BIPM data set has a standard deviation of about 4 ns. While the NIST data set has the smallest standard deviation of the three data sets, it should be noted that the manner in which it is formed employs the strongest smoothing. It is believed that the results reflect this smoothing. The Two-Way data set is not smoothed at all. However, interpolation was done to derive points for the same times as the other data sets. The BIPM data set is derived using a weak smoothing algorithm. However, it is assumed that the large variance of this data set is caused by publication of the data to only 10's of nanoseconds. This is a shortcoming of this

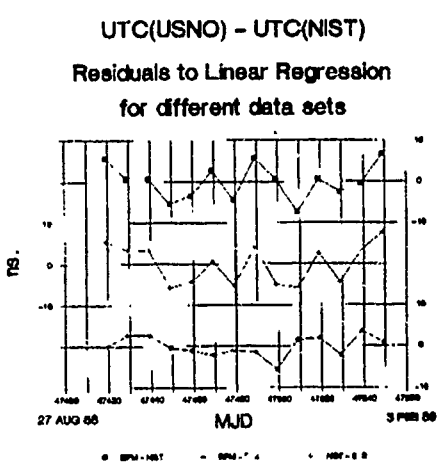


Figure 4.

3 Cornered Hat

$$\sigma_{ab}^2 = \sigma_a^2 + \sigma_b^2$$

$$\sigma_{ac}^2 = \sigma_a^2 + \sigma_c^2$$

$$\sigma_{bc}^2 = \sigma_b^2 + \sigma_c^2$$

$$\sigma_a^2 = 1/2\,(\sigma_{ab}^2 + \sigma_{ac}^2 - \sigma_{bc}^2)$$

σ = Standard Deviation

Figure 5-Equations for deriving the variance of an individual item when three pairs of variances are known.

comparison, but these are the only published data sets which there are to compare.

The fact that the Two-Way data set is almost as smooth as the NIST GPS data set states that method is intrinsically free of a lot of external noise. Therefore, as the technique matures, better knowledge of the processes affecting the results will be gained. As a consequence, it seems that the Two-Way technique will become a very powerful tool in understanding the behavior of remotely separated clocks, as well as one of the most intrinsically precise methods for time transfer.

Results

$\sigma_{BIPM-NIST} = 4.1$ ns.

$\sigma_{BIPM-E.S.} = 4.7$ ns.

$\sigma_{NIST-E.S.} = 2.4$ ns.

$\sigma_{BIPM} = 4$ ns.

$\sigma_{E.S.} = 2$ ns.

$\sigma_{NIST} = 1$ ns.

Figure 6-Results of applying the Three Cornered Hat Technique to determine the Standard Deviation of the BIPM, NIST and Two-Way data sets.

VII. REFERENCES

[1] Chi, A. R. and Byron, E., "Two-way Time Transfer Experiment Using a Synchronous Satellite", Proceedings of 11th Annual PTTI Applications and Planning Meeting, 1975, pp. 357-377.

[2] Klepczynski, W. J., Wheeler, P. J., Powell, W., Jeffries, J., Myers, A., Clarke, R. T., Hanson, W., Jespersen, J. and Howe, D., "Preliminary Comparison between GPS and Two-Way Satellite Time Transfer", Proceedings of 42nd Annual Frequency Control Symposium, 1988, pp. 472-477.

[3] Markowitz, W., "The Photographic Zenith Tube and the Dual-Rate Moon Position Camera", Telescopes, ed. by G.P. Kuiper and B.M. Middlehurst, Univ. of Chicago Press, 1966, pp. 88-114.

[4] Winkler, G. M. R., "Intermediate Term Frequency Measurements with the HP Computing Counter in the USNO Clock Time System", Proceedings of the 4th Annual PTTI Applications and Planning Meeting, 1972, pp. 152-167.

[5] Lewandowski, W., Private Communication, 1989.

[6] Allan, D. W., Private Communication, 1988.

[7] Vondrak, J., "A Contribution to the Problem of Smoothing Observational Data", Astr. Bull. Czech., Vol. 20, No. 6, pp. 349-355, 1969.

[8] Gray, J. E. and Allan, D. W., "A Method for Estimating the Frequency Stability of an Individual Oscillator", Proceedings of 28th Annual Symposium on Frequency Control, 1974, pp. 243-246.

43rd Annual Symposium on Frequency Control - 1989

Applications Of Ultra–Stable Fiber Optic Distribution Systems*

Lori E. Primas,
Ronald T. Logan, Jr.
George F. Lutes

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

Abstract

Fiber optic distribution systems will soon be used as the preferred method for frequency and time distribution. Presently, passively stabilized fiber optic systems can distribute hydrogen maser quality signals, having an Allan deviation of 1 part in 10^{15} for 1000 seconds averaging times, over 30 kilometers with less than 10 percent degradation in the frequency stability. The high distribution stability achievable with fiber optic systems, provides improved performance and permits new experiments to be performed. Specifically, fiber optic systems will be used in applications such as spacecraft navigation, relativity experiments, arrayed antenna experiments and frequency distribution in dynamic environments. This paper will discuss present and future applications of fiber optic frequency distribution systems. A stabilized fiber optic system that can improve distribution stability by more than 100 times will also be described.

Introduction

Diverse applications ranging from gravitational wave detection to coherently arrayed multiple antennas require ultra–stable frequency distribution systems. Other frequency distribution systems use coaxial cable, waveguide and terrestrial microwave and can distribute signals over distances of a few kilometers. Satellites cannot disseminate precise frequency and timing signals because of the resulting low signal to noise ratio (SNR), and can only be used to syntonize frequency references generated by frequency standards and synchronize timing references which are clocked by adjacent frequency standards. Synchronization and syntonization of frequency and timing signals is achieved by averaging the satellite signals over long averaging times. Fiber optic systems are the only systems that can disseminate precise frequency and timing signals to remote users hundreds of kilometers away. Thus, the remote users do not need a frequency reference source at the site. The high cost of developing and maintaining a state–of–the–art frequency standard makes it beneficial to use a fiber optic distribution system. In addition to providing long distance distribution, fiber optic systems provide very stable distribution. Fiber optic distribution systems also enable various scientific experiments that require coherently connected systems separated by tens of kilometers.

The time and frequency systems research group at JPL is involved in the distribution of ultra–stable signals from a hydrogen maser throughout the Deep Space Network (DSN) complex at Goldstone, California (Figure 1). The DSN consists of three complexes located at Goldstone, California, Madrid, Spain and Canberra, Australia. At each complex there are at least four stations, each supported by a parabolic dish antenna with an ultra–sensitive receiving system requiring a precise frequency reference. Currently each complex is supported by a primary hydrogen maser and a backup hydrogen maser [1].

The present capabilities at the Goldstone complex, include distribution of a 100 MHz signal over a 29 km link with a stability less than 1 part in 10^{15} for 1000 seconds [2]. As new, more stable, frequency standards are developed, the distribution systems must improve to achieve the goal of being ten times more stable than the reference signal. Our present goal is to distribute a 100 MHz

*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.

CH2690-6/89/0000-202 $1.00 © 1989 IEEE

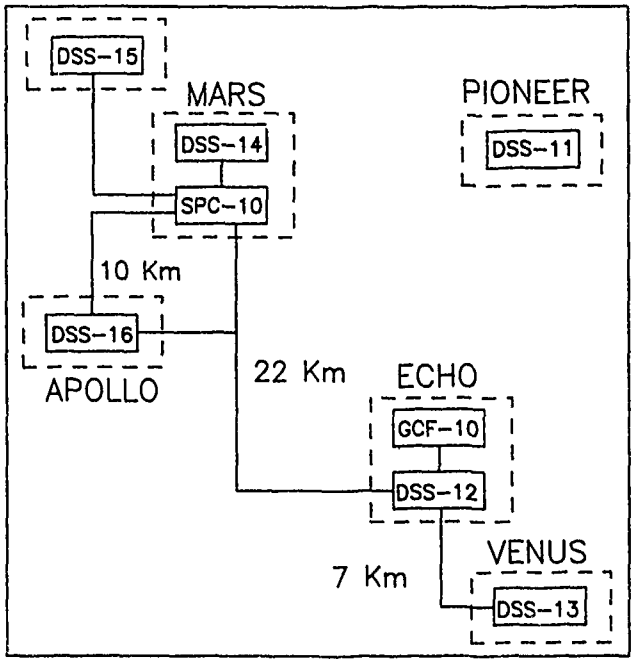


Figure 1: Goldstone Complex Configuration

signal with a stability of 1 part in 10^{17} for 1000 seconds over the 29 km link.

The two ultimate objectives in any distribution system are longer distribution distances and higher distribution stability. With improved lasers, fiber optic systems should have the capability of distributing signals over hundreds of kilometers.

This paper will discuss general considerations important in frequency distribution over fiber optic distribution systems. The present capabilities of the fiber optic links will be presented as well as results from applications currently using fiber optic distribution systems.

For applications requiring distribution stability greater than a part in 10^{15} at 1000 seconds averaging times, a stabilized fiber optic link is the only choice presently available. A stabilized fiber optic distribution system designed at JPL reduces phase variations on the link by 400 times. This system will be discussed.

Fiber Optic Distribution Systems

Fiber optic distribution systems provide high stability because fiber optic cable is a low loss medium and has a low thermal coefficient of delay (TCD). There are two basic limitations in frequency distribution systems. The first is a distance limitation set by the SNR of the received signal. The second limitation is the degradation of the frequency stability due to variations in the group delay of the signal as it is transmitted through the medium.

The SNR is limited by the amount of available output power from the frequency source and the loss in the distribution system. Because fiber optic cable is a low loss medium, a reference signal can be transmitted over long distances and maintain a high SNR. Degradations in the SNR due to the length and loss of coaxial cable is given by,

$$SNR = \frac{10^{-\alpha \cdot L} \cdot P_{in}}{k\,T\,B}, \tag{1}$$

where,

SNR = signal to noise ratio,
α = loss in the medium [dB/Km],
L = length of the medium [Km],
P_{in} = input power [Watts],
k = Boltzmann's constant,
= 1.38054×10^{-23}[J/K],
T = temperature [K] and
B = bandwidth [Hz].

Typically, SNR is given in units of dB, thus

$$SNR[dB] = 10 \times \log(SNR). \tag{2}$$

Assuming a matched load resistance at 300 K in a 1 Hz bandwidth and low loss coaxial cable distribution systems, 1 kW of power would be needed to distribute the signal through 7 km of coaxial cable having a loss of 16.4 dB/km (0.5 dB/100 ft) and maintain a SNR of 120 dB.

Because of the conversion from optical to electrical power, equation (1) does not apply directly to fiber optic cable. For fiber optic distribution systems, a 1 milliwatt optical signal can be distributed in fiber optic cable with a loss of 0.5 dB/km over a distance of 54 km and maintain a SNR of 120 dB. Such parameters as laser optical power, shot noise, dark current noise and thermal noise can degrade the SNR of a fiber optic link [6].

Variations in group delay are caused by physical changes in the transmission medium. A constant rate of change in group delay does not effect the frequency stability, but a change in the rate of group delay variations does degrade the frequency stability as shown in the following equation:

$$\frac{d(\delta f)}{dt} = f_o \cdot \frac{d^2 D}{dt^2} \tag{3}$$

where,

δf = frequency offset [Hz],

f_o = modulation frequency [Hz] and
D = group delay [Seconds].

Frequency stability describes the time behavior of a frequency standard. Frequency stability is typically measured by its Allan deviation, a two sample deviation that describes the frequency stability for different averaging times [3].

Assuming white phase noise, the double sided spectral phase density $S_\Phi(SNR[dB])$ is given by,

$$S_\Phi(SNR[dB]) = 10^{\frac{-SNR[dB]}{10}}. \quad (4)$$

The Allan deviation for white phase noise is [4],

$$\sigma_y(\tau) = \frac{0.276 \cdot \sqrt{S_\Phi(SNR)} \cdot \sqrt{f_h}}{f_o\,\tau} \quad (5)$$

where,

τ = averaging time [seconds],
$S_\Phi(SNR[dB])$ = spectral phase density,
$SNR[dB]$ = signal to noise ratio [dB] and
f_h = measurement bandwidth [Hz],
f_o = modulation frequency [Hz].

Combining equations (4) and (5) we get,

$$\sigma_y(\tau) = \frac{0.276 \cdot \sqrt{10^{\frac{-SNR[dB]}{10}}}}{f_o\,\tau} \quad (6)$$

For a SNR of 120 dB in a 1 Hz bandwidth and a modulation frequency of 100 MHz, the Allan deviation is,

$$\sigma_y(\tau) = \frac{2.76 \times 10^{-15}}{\tau}. \quad (7)$$

The Allan deviation of a signal is degraded due to changes in the group delay. Typically group delay variations are periodic and caused by thermal variations on the fiber optic cable. The Allan deviation of a periodic variation can be describes in terms of the temperature variations on the link or by the phase variations of the distributed signal. The Allan deviation as a function of a sinusoidal temperature variation is given by [5],

$$\sigma_y(\tau) = \frac{L \cdot \alpha \cdot \delta T \cdot n}{c} \cdot \left[\sin^2\left(\frac{\pi\tau}{P_T}\right)\right] \quad (8)$$

where,

τ = averaging time [seconds],
α = thermal coefficient of delay (TCD) [ppm/°C],
L = length of the cable [meters],
δT = peak to peak temperature change [°C],
n = index of refraction of the fiber,
C = speed of light [3×10^8m/sec] and
P_T = period of temperature change [seconds].

For a 4 km length of fiber optic cable having an index of refraction of 1.429, a TCD of 7 ppm/°C and a sinusoidal temperature variation of 10 °C over 24 hours, the Allan deviation is,

$$\sigma_y(\tau) = \frac{1.33 \cdot 10^{-9}}{\tau} \cdot \left[\sin^2\left(\frac{\pi\tau}{86400}\right)\right] \quad (9)$$

The Allan deviation of a signal can be calculated from the measured phase variation of the received signal. The Allan deviation in terms of the peak to peak phase fluctuation is given by,

$$\sigma_y(\tau) = \frac{\Phi_{pp}}{(360 \cdot f_o \cdot \tau)} \cdot \left[\sin^2\left(\frac{\pi\tau}{P_P}\right)\right] \quad (10)$$

where,

Φ_{pp} = peak to peak phase fluctuation [degrees],
f_o = modulation frequency [Hz],
τ = averaging time [seconds] and
P_P = period of phase variations [seconds].

A theoretical estimate of the Allan deviation of a fiber optic link can be determined by looking at the combined effects of the SNR of the received signal and the periodic temperature variations on the link. Figure 2 shows the Allan deviation of a 4 km link having a SNR of 120 dB with a 40 degree phase variation over 24 hours and a 0.13 degree change over 8 hours.

One of the first experiments performed on the 29 km fiber optic link was to measure its Allan deviation. A frequency reference signal from a hydrogen maser was used to modulate the signal on the fiber optic transmitter at 100 MHz. The signal was received at the remote end and then retransmitted back to the source for a total distribution distance of 58 km (Figure 3).

The fiber optic cable is buried 1.5 meters underground to minimize temperature variations. The phase difference between the transmitted and received signals at the source was compared and the Allan deviation of the distributed signals was calculated as in equation (10). At 1000 seconds averaging times, the Allan deviation of the signal is a part in 10^{-15}. The plateau in the Allan deviation at 10000 seconds is due to diurnal temperature variations of the fiber optic cable (Figure 4).

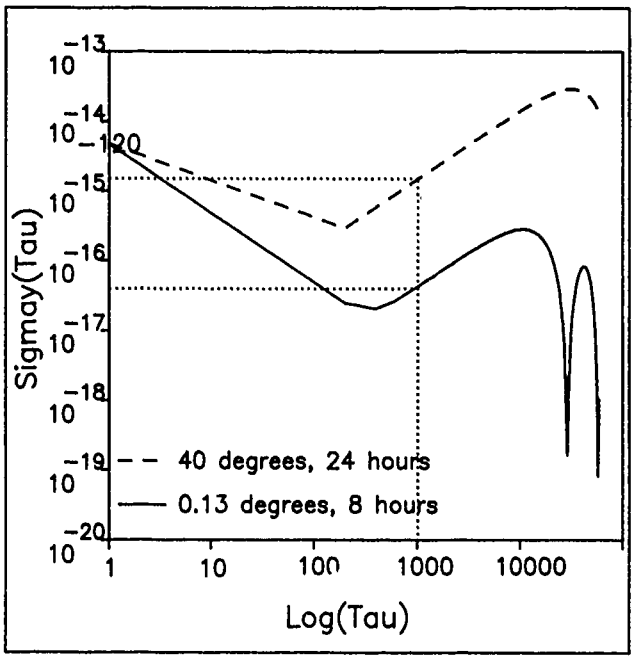


Figure 2: Allan deviation vs. Tau for 4 km link with SNR = 120 dB and phase variations

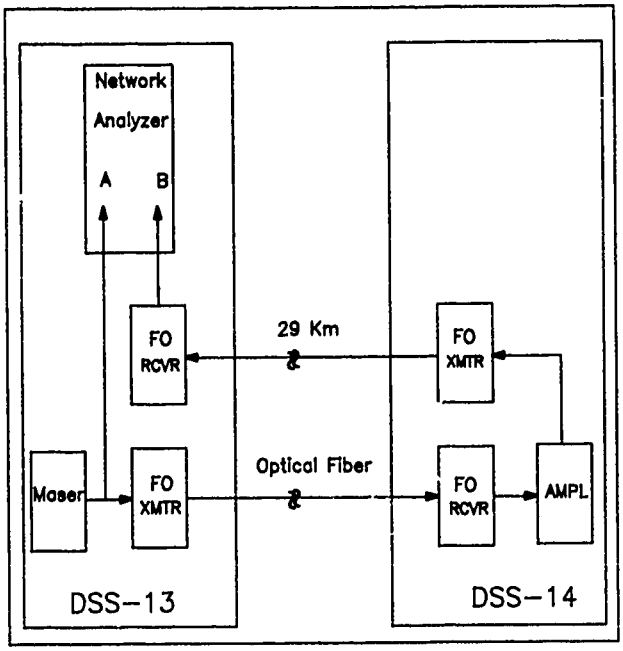


Figure 3: Test configuration for frequency distribution over 58 kilometers of fiber optic cable.

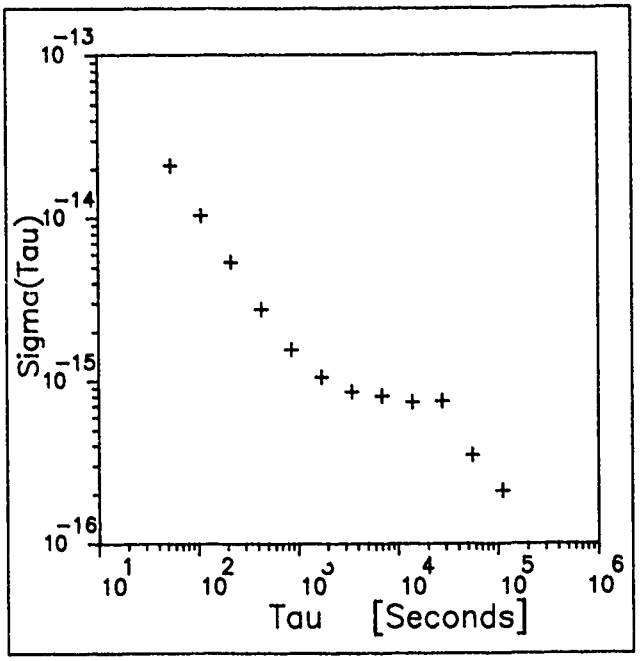


Figure 4: Allan deviation for frequency distribution over 58 kilometers of fiber optic cable.

Fiber Optic Distribution Limitations

Degradations of distributed frequency stability in a fiber optic distribution system can be related to the fiber optic transmitter, the fiber optic receiver or the fiber optic link [6]. Any factors that degrade the SNR or cause thermal degradations of the system components will degrade the distribution stability.

Factors that effect distribution stability at the transmitter include available maximum optical laser output power, modulation depth of the laser, modulation frequency and the relative intensity noise (RIN) of the laser. Typical transmitter parameters are,

$$\begin{aligned} P_o &= \text{optical power} = 1\text{mW}, \\ m &= \text{modulation index} = 0.5, \\ f_o &= \text{modulation frequency} = 100\text{ MHz and} \\ RIN &= \text{relative intensity noise} = -135\text{dB/Hz}. \end{aligned}$$

Factors that effect the distribution stability at the receiver include responstivity of the detector, detector load resistance and the modulation frequency. Typical receiver parameters are,

$$\begin{aligned} R &= \text{responstivity power} = 0.7\text{A/W}, \\ f_o &= \text{modulation frequency} = 100\text{ MHz and} \\ R_L &= \text{detector load resistance} = 50\Omega. \end{aligned}$$

Factors that effect the distribution stability due to the link itself include the length of the link, the modulation frequency, TCD of the fiber optic cable, the attenuation in the cable and splice losses.

Stabilized Fiber Optic Distribution Link

Burying the fiber optic cable underground reduces most of the temperature variations on the cable and thus the fiber optic system is passively very stable. For applications requiring stability greater a part in 10^{-15} at 1000 seconds averaging times, however, active stabilization must be incorporated.

As was stated earlier, the final goal in frequency distribution is to distribute the 100 MHz signal over 29 km with a stability of 1 part in 10^{17} for 1000 seconds. In order to achieve this goal, an active means of stabilization is necessary. Temperature variations on the fiber optic cable are the largest cause of degradation of frequency stability and can be compensated for by adding a corrective amount of phase. A cable delay compensator (CDC) has been built and tested at JPL [7],[8].

The basis of the CDC design is the phase relationship between signals propagating in the fiber optic cable in both directions. A signal passing through the fiber optic cable and its reflection from the remote end experience identical delay in the two directions. The midpoint of the signal propagation time is the far end of the cable, thus the signal undergoes exactly half the round trip delay of this path. If the phases of the transmitted and received signals at the reference end of the cable are conjugate, the phase at the remote end is independent of phase delays in the medium. Thus, the CDC is an electronic device that detects the phases of the transmitted and received signals at the input to the fiber and adds enough phase to maintain conjugation.

The CDC compares the phase of the transmitted and received signals in the reference unit and derives an error voltage from the phase difference (Figure 5). This voltage is then used to control a voltage-controlled oscillator (VCO). The particular design of the CDC requires a 100 MHz reference signal and a 20 MHz auxiliary signal. A previous design used a single 100 MHz reference signal, but required two precisely matched phase detectors and tightly controlled signal levels. By using the 20 MHz auxiliary signal, a single phase detector can be used to measure phase error.

The 100 MHz signal and the 20 MHz signals are multiplied together in mixer M1 to produce 80 MHz and 120 MHz signals. A power splitter (S1) separates the signal

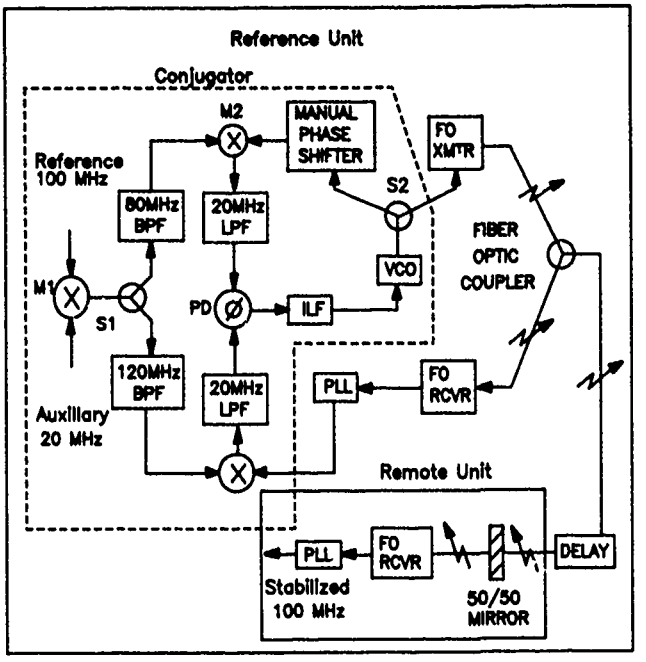


Figure 5: Block diagram of cable delay compensator.

out of the mixer (M1) into two signal paths. Band pass filters in each signal path separate the 80 and 120 MHz signals. The 80 MHz signal and the 100 MHz signal from the VCO are multiplied in mixer (M2) to produce a 20 MHz intermediate (IF) signal. The 20 MHz IF signal contains the instantaneous phase difference between the VCO signal and the 80 MHz signal.

The 120 MHz signal and the 100 MHz signal reflected from the remote unit are multiplied together in mixer (M3) to produce another IF signal. This 20 MHz IF signal contains the instantaneous phase difference between the reflected signal (100 MHz) and the 120 MHz signal.

The phase detector (PD) receives the two 20 MHz IF signals and produces an error voltage that is proportional to the phase difference between them. The error voltage is applied to the VCO control input through the inner loop filter (ILF). Delay changes in the fiber optic cable result in changes in the control voltage. This voltage controls the phase of the VCO relative to the reference 100 MHz signal.

The output of the VCO is divided into two signals in the RF power splitter (S2). One of the signals is received by mixer M2 and the other modulates the optical carrier emitted from the laser transmitter.

The modulated optical signal is transmitted to the remote unit through the optical coupler. The 50/50 mirror at the remote unit reflects half of the optical signal back toward the reference unit while the other half passes through the mirror to the optical receiver. The receiver demodulates the optical signal and amplifies the resulting

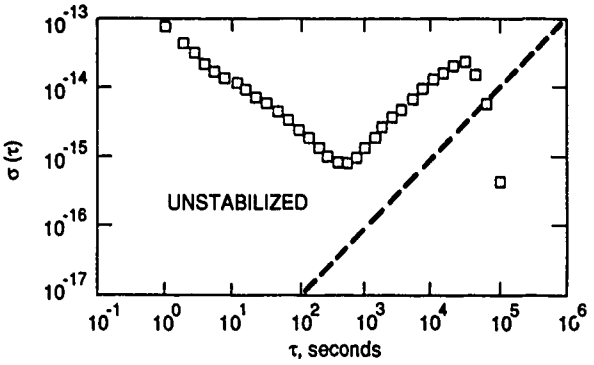


Figure 6: Allan deviation for unstabilized 4 kilometer fiber optic link.

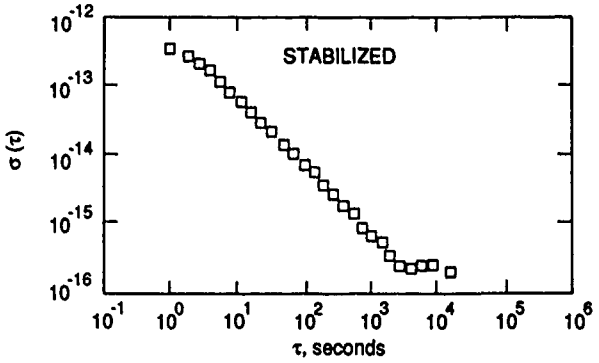


Figure 7: Allan deviation for stabilized 4 kilometer fiber optic link.

100 MHz RF signal. A PLL filters the signal to be used at the remote unit. The reflected optical signal returns to the reference unit where it passes through the optical coupler and is detected by another optical receiver. This signal is also filtered by another PLL and provides a constant amplitude signal into mixer M3. With the signal back at the reference unit, the system loop is closed.

The CDC was tested with a 4 Km link of fiber in an environmental test chamber. The temperature of the fiber optic cable was varied sinusoidally 10 °C over a 24 hour period. The resulting phase variations were reduced by 400 times (Figures 6 and 7). Figure 1 shows the phase variations of a stabilized and unstabilized 4 km link with a 10 °C temperature change over 24 hours.

Applications Of Fiber Optic Distribution Systems

Relativity Experiments

Because of the high stability of fiber optic systems and the fact that light can be transmitted both ways through the fiber optic cable with up to 100 dB isolation, fiber optic cable is the ideal distribution medium for relativity experiments.

Experiments have been done to determine the isotropy of the velocity of light. Previous tests of the isotropy of the velocity of light propagated the signal back and forth in two-way experiments. Two-way experiments are limited in that the isotropy of the average velocity is tested. The effect of anistropy of the speed of light on phase measurements can be shown by considering

$$c = c_o + \delta c(T_R) \tag{11}$$

where,

c = speed light with anistropy [m/sec],
c_o = average speed of light [m/sec],
= 299,792,458 m/sec,
δc = generalized perturbation [m/sec] and
T_R = period of perturbation [seconds].

The phase variation resulting is,

$$\delta\theta(T_R) = \frac{f_o \cdot L}{c_o} \cdot \frac{\delta c(T_R)}{c_o} \tag{12}$$

where,

$\delta\theta(T_R)$ = phase variation [degrees],
f_o = modulation frequency [Hz],
L = baseline of experiment [meters],

According to special relativity, $\delta\theta(T_R)$ should be zero for all T_R. As greater precision is obtained in the measurement of $\delta\theta(T_R)$, more is known about the anistropy of light.

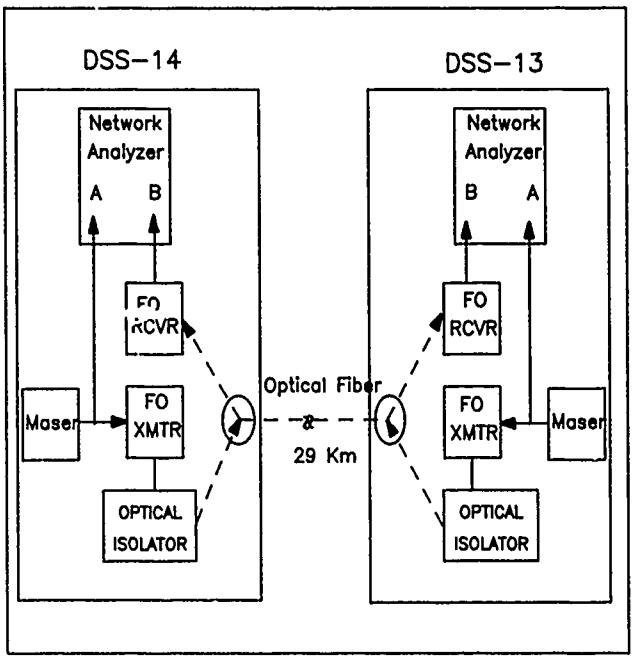


Figure 8: Test configuration for relativity experiment.

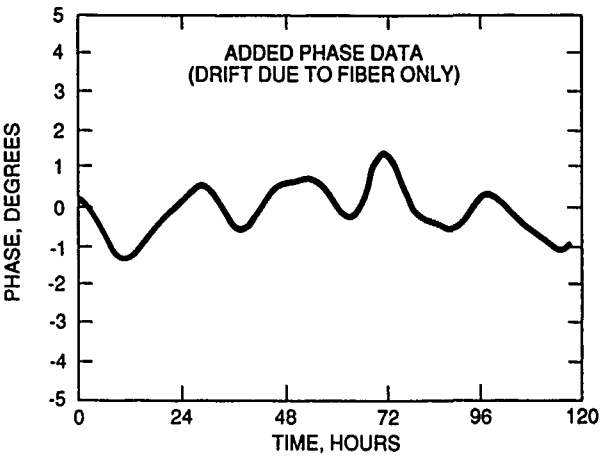


Figure 9: Additive phase results of relativity experiment.

The 29 km fiber optic link at Goldstone has been used to propagate a signal one–way between two ultra–stable frequency sources [9]. Signals (100 MHz) from two hydrogen masers located at each end of the link were transmitted to the opposite end of the link. The 29 km link spans a 10 km distance in the East–West direction. This orientation changes with respect to the stars as the earth rotates and thus provides the test conditions for this experiment. For a frequency of 100 MHz and a baseline of 10 km, equation (12) can be rewritten as,

$$\delta\theta(T_R) = 1.2 \times 10^6 \cdot \frac{\delta c(\tau)}{c_o} \qquad (13)$$

The phase of the transmitted maser signal and the received signal were compared at each end of the link for a period of 4.5 days (Figure 8).

These two sets of data were analyzed to obtain the additive phase and differenced phase. The additive phase represents the drift of the fiber and was determined to be 2 degrees peak to peak at 100 MHz (Figure 9). The differenced data represents the differential maser drift and was determined to be 30 degrees peak to peak (Figure 10). The accuracy of the resolution is thus limited by the stability of the maser.

Distribution in Dynamic Environments

Besides being a low loss medium, fiber optic cable has a very low TCD and low sensitivity to bending. The TCD for standard loose tube fiber is (7 ppm/°C) and less than (0.5 ppm/°C) for new low TCD fiber. Thus, fiber is an excellent medium for transmitting ultra–stable signals in dynamic environments. A specific application in

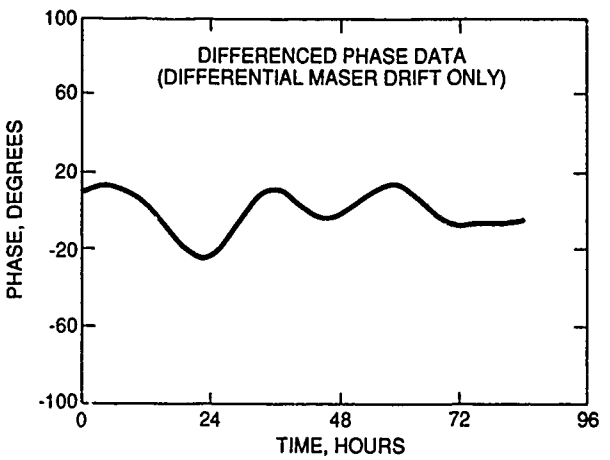


Figure 10: Differenced phase results of relativity experiment.

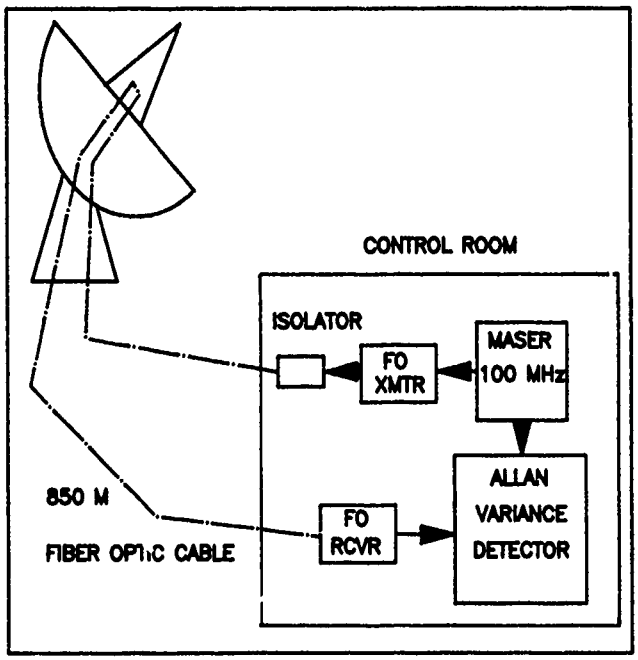


Figure 11: Test configuration for distribution of reference frequencies to cone of the antenna.

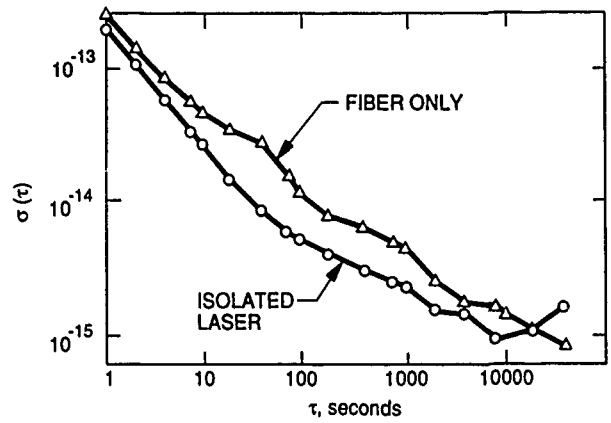


Figure 12: Results of distribution of reference frequencies to cone of the antenna.

the DSN is to distribute a reference signal from a control room and through an antenna wrap- up (Figure 11). Fiber optic cables in the antenna are subject to large temperature variations, bending, and other environmental effects. The stability of an 850 meter long standard loose tube fiber optic cable operating in such an environment was measured and found to be 3 parts in 10^{15} for 1000 seconds in this application. Tests will soon be made using the new low TCD fiber optic cable .

Changes in the polarization and amplitude of reflection from the fiber optic cable back into the laser transmitter cause phase instabilities in the transmitted signal. These changes in reflections occur when the fiber optic cable is flexed. The phase changes due to cable flexure can be eliminated by isolating the laser from reflections. This can be accomplished with an optical isolator. The standard commercial optical isolators have 40 dB isolation, but 60 dB isolation is needed. New isolators having greater than 60 dB isolation have recently been developed and are becoming commercially available.

Results from the measurements made on the stability of the loose tube fiber optic cable routed through the wrap-up of an antenna at the Goldstone complex is shown in Figure 12.

Spacecraft Navigation

Antennas used in the DSN for spacecraft navigation are separated by continental distances. One of the navigation techniques used in the DSN is Very Long Baseline Interferometry (VLBI). In this technique a signal from the spacecraft is received by two widely separated antennas. The time at these antennas is accurately synchronized to a common time standard using GPS satellites . The signal at each antenna is time tagged as it is received and autocorrelated to the time tagged signal received by the other station. The difference in arrival time of the signal between the two stations can be used to pinpoint the location of the spacecraft.

Fiber optic systems are used to keep the local timing at each station very accurately. For instance a frequency reference is transmitted to the antenna front end where it is used to generate tones which fall within the passband of the receiver. These tones are added to the incoming signal and are compared to the station standard at the output of the system to determine the receiver system delays. At the Goldstone DSCC a fiber optic system is used to distribute a frequency reference which is generated by a centralized frequency standard to each station on the complex. Typical distribution distances are 10 to 20 kilometers. This distributed reference frequency is used by the station to maintain its internal time and to syntonize its internal frequency references to the centralized frequency standard. The measurement precision available with VLBI allows ranging, doppler and doppler rate parameters to be measured. An additional measurement supported by VLBI is angular navigation.

The coaxial systems which have been used to distribute reference frequencies within stations are gradually being replaced with more stable fiber optic links.

Antenna Arraying

Fiber optic links can also be used to coherently connect two antennas separated by tens of kilometers. Antenna arraying effectively increases the sensitivity of the receiving system and allows transmitters to be added for greater transmitter power. Unlike VLBI, antennas separated by tens of kilometers can be coherently connected with fiber optic cable. Because of its short baseline, antenna arraying allows real time data processing. By coherently arraying multiple antennas, coherent reception and transmission experiments are enhanced. Some unique experiments such as telemetry arraying, connected element interferometry (CEI), debris radar, gravitational wave detection and planetary ring and atmosphere detection can be performed [11].

Future Directions

As was pointed out earlier, limitations of fiber optic distribution systems are evident in the fiber optic transmitter, the fiber optic receiver and the link itself. To improve the distribution stability, lasers with lower RIN and higher optical output are necessary. Optical isolators are an extremely important part of the distribution system, especially for links in dynamic environments. Receivers with greater responstivity and lower noise could also improve the distribution stability.

Unstabilized fiber optic links provide sufficiently high stability for the frequency standards currently available, but when ultra- high distribution stability is necessary, stabilized links must be used. The CDC previously describe is quite sufficient for current frequency standards, but even greater stability will be necessary for future standards. Work will be done to improve the short term Allan deviation of the link. The performance of the previous CDC seems to be limited by the phase noise characteristics of the VCO used to correct the phase.

Conclusion

Fiber optic distribution systems are the only distribution systems that can disseminate ultra-stable frequencies over hundreds of kilometers. The applications of such a distribution system cover a wide range of scientific research. The present CDC is adequate for hydrogen maser signals and improved methods of active stabilization will further enhance the use of fiber optic distribution systems.

Acknowledgments

The authors wish to thank Phuong Tu, Bill Diener and Al Kirk for their help in making the measurements. Also, the authors would like to thank Lute Maleki, Richard Sydnor and John Dick for their help and support throughout the course of this work.

References

[1] JPL publication 400-333 1/88, "The Deep Space Network."

[2] G. Lutes, "Reference Frequency Distribution Over Optical Fibers: A Progress Report," *Proceedings of the 41st Annual Frequency Control Symposium*, pp. 161-166, May 27-29, 1987.

[3] D. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," *IEEE Transactions on Ultrasonics, Ferroelectronics, and Frequency Control*, pp. 647-654, November 1987, Volume UFFC-34, Number 6.

[4] CCIR Report 580-1, "Characterization Of Frequency And Phase Noise," Volume VII, 1982.

[5] C.A. Greenhall, "Frequency Stability Review," *The Telecommunications and Data Acquisition Progress Report 42-88*, pp. 200-212, October-December 1986.

[6] R. Logan, G. Lutes and L. Maleki, "Impact of Semiconductor Laser Frequency Deviations On Fiber Optic Frequency Distribution Systems," *Proceedings of the 43rd Annual Frequency Control Symposium*, this issue.

[7] L. Primas, G. Lutes, and R. Sydnor, "Stabilized Fiber Optic Frequency Distribution System," *Proceedings of the Twentieth Annual Precise Time and Time Interval Applications and Planning Meeting*, NASA Conference Publication, pp. 23-34, 1988.

[8] L. Primas, G. Lutes, and R. Sydnor, "Stabilized Fiber Optic Frequency Distribution System," *The Telecommunications and Data Acquisition Progress Report 42-97*, 1989.

[9] T. Krisher, L. Maleki, L. Primas, R. Logan, G. Lutes and J. Anderson, "Results Of A New Test Of Relativity," *Proceedings of the Twentieth Annual Precise Time and Time Interval Applications and Planning Meeting*, NASA Conference Publication, pp. 251260, 1988.

[10] C. Edwards, "Angular Navigation Capabilities Of Short Baseline Phase Delay Interferometry," *Proceedings of the AIAA/AAS Astrodynamics Conference*, Aug. 15–17, Minneapolis, MN, PP. 545–548, Paper 88–4287–CP, August 15–17, 1988.

[11] C. Edwards, "Short Baseline Phase Delay Interferometry," *The Telecommunications and Data Acquisition Progress Report 42–97*, pp. 46–56, 1989.

43rd Annual Symposium on Frequency Control - 1989

IMPACT OF SEMICONDUCTOR LASER FREQUENCY DEVIATIONS ON FIBER OPTIC FREQUENCY REFERENCE DISTRIBUTION SYSTEMS

Ronald T. Logan, Jr., George F. Lutes, Lute Maleki
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg. 298
Pasadena, California 91109

Abstract

This paper presents an analysis of the effect of the linewidth of a single–longitudinal–mode semiconductor laser on the frequency stability of a frequency reference transmitted over single mode optical fiber. The interaction of the random laser frequency deviations with the dispersion of the optical fiber is considered to determine theoretically the effect on the Allan deviation of the transmitted frequency reference. It is shown that the magnitude of this effect may determine the limit on the ultimate stability possible for frequency reference transmission on optical fiber, but is not a serious limitation to present system performance.

Introduction

Ultra-stable fiber optic transmission of hydrogen maser reference signals is presently operational at the Goldstone facility of the JPL/NASA Deep Space Network [1]. This capability supports radio science experiments such as Connected Element Interferometry (CEI) by enabling phase coherent arraying of widely separated antennas in real time. Also, distribution of a centralized maser reference throughout the entire complex eliminates the need for a hydrogen maser frequency standard at each Deep Space Station, with substantial cost savings and increased reliability.

A reference signal produced by a hydrogen maser frequency standard is presently distributed over distances up to 29 km with differential fractional frequency stability $\frac{\delta f}{f} \approx 10^{-15}$ for 1000 second averaging times. Although the present fiber optic distribution capability is as stable as the hydrogen maser frequency standard, the ideal distribution system should be an order of magnitude more stable than the distributed signal. With the promise of trapped ion frequency standards [2] and superconducting cavity masers [3] providing more stable frequency references, fiber optic link stability of 10^{-18} at 1000 seconds will be required for stable distribution.

The demanding requirements that a frequency reference distribution system must meet necessitate the examination of all sources of instability at levels far beyond the needs of typical analog and digital fiber optic communication systems. Presently, laser source amplitude noise and thermal variations of the optical fiber have been identified as the limiting factors to distribution system performance. Improved lasers with lower intensity noise and single longitudinal mode operation will be employed in the near future. A thorough examination of fiber optic systems and components has indicated that laser frequency deviations may limit system performance as lower amplitude noise lasers become available. However, a quantitative analysis of the effect of laser frequency noise on a narrow-band frequency distribution system has not previously been performed.

The present analysis theoretically estimates the effects on the amplitude and phase stability of a frequency reference transmitted on single mode optical fiber due to laser frequency fluctuations. An expression for the phase noise spectral density of the modulation signal due to the frequency noise spectral density of the laser is derived and then used to calculate the expected Allan deviation of the transmitted reference signal. The laser induced phase noise is shown to depend on the modulation signal frequency, fiber length, and fiber dispersion, as well as the magnitude of the laser frequency fluctuations. It is shown that the differential frequency stability of a single mode fiber optic link is fundamentally limited by laser frequency noise. Thus, as laser intensity noise is reduced, the laser frequency noise will limit transmission stability.

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

CH2690-6/89/0000-212 $1.00 C 1989 IEEE

Fiber–Laser Interaction

The dispersion of optical fiber causes various optical frequencies to travel with different velocities. Optical carrier frequency deviations couple with the dispersion of the fiber to produce random phase deviations in the envelope of a modulation signal thereby degrading its phase stability. Optical fiber acts as a frequency discriminator to translate random frequency deviations of the laser into random phase deviations of the RF modulation envelope. Although every effort is made to operate the laser at the minimum dispersion point of the fiber, the slope of the fiber index of refraction versus wavelength is typically not zero. As the laser frequency deviates, the signal experiences changes in the fiber index of refraction that cause phase shifts of the modulation envelope.

Under bias current modulation, a semiconductor laser exhibits changes in wavelength, or chirp, that are synchronous with the bias modulation due to the change in refractive index of the laser gain medium [4]. Lasers also exhibit random frequency fluctuations due to the quantum phenomenon of electron–hole recombination in the gain media with an attendant change of refractive index [5]. Temperature excursions of the laser diode also affect the index of refraction and the lasing wavelength, but with time constants from minutes to hours.

In digital transmission systems, laser chirp is the predominant limit on transmission distance [6,7]. These wide–band systems are sensitive to phase deviations of the modulation envelope at all frequencies. However, in such systems, spontaneous emission noise is ignored since chirp is the overwhelming effect [7].

In contrast to digital or wide–band analog transmission systems, frequency distribution systems employ a narrow–band loop filter at the output of the fiber receiver. Therefore, high frequency deviations of the modulation envelope phase are averaged out, leaving only the laser noise within the loop bandwidth. An analysis of the effect of close–to–carrier laser frequency noise on a long distance frequency reference transmission system has not been published (to our knowledge), so the effects of laser frequency noise have not been known. Also, at the levels of signal to noise ratio (SNR) and frequency stability of the frequency distribution systems under consideration, it has been unclear what role laser frequency noise plays in determining the ultimate system stability attainable. The present analysis provides an estimate of the role of intrinsic laser frequency noise in a narrow–band frequency distribution system to determine the level at which system performance might become limited.

Since high–frequency laser chirp can be neglected in a narrow–band system, the present analysis considers only the intrinsic laser frequency noise within the bandwidth of the output filter. As such, the analysis applies to any type of laser system, although semiconductor lasers are typically used. Externally modulated Nd:YAG lasers at 1318 nm may be an attractive alternative to semiconductor lasers for long–haul analog signal transmission. The present analysis applies equally well to these types of lasers by substitution of the appropriate parameters.

Analysis

Intrinsic laser frequency noise has its origins in the discreet random photons spontaneously emitted into the lasing mode that cause random frequency changes of the laser wavelength [8,9,10]. The high frequency character of this noise is well known. It is basically flat within the modulation bandwidth, peaking at the relaxation oscillation resonance of the laser diode cavity, usually in the tens of gigahertz region [8,9]. Within tens of kHz of the carrier, the frequency noise exhibits a $1/f$ character [11]. Close–to–carrier measurements of laser noise are limited to within about 10 kHz, due to the physical difficulty of fabricating frequency discriminators with sufficient resolution at optical frequencies. It is this low frequency FM noise that is of interest for the analysis of narrow–band frequency distribution systems.

We desire an expression for the phase noise spectral density of the modulation signal as a function of the spectral density of laser frequency fluctuations. The resultant phase noise density may then be used to calculate the expected Allan deviation of the reference signal, provided the character of the laser frequency noise is known.

Consider a single longitudinal–mode laser, coupled to a single mode fiber. The laser output is amplitude modulated at RF frequency Ω. The phase delay for the modulation signal envelope along the fiber is given by

$$\phi = \frac{2\pi \cdot n \cdot L \cdot \Omega}{c} \quad \text{[rad]} \tag{1}$$

where n is the fiber index of refraction, L is the fiber length, and c is the speed of light in vacuum. It is assumed that the laser operates in a single longitudinal mode at $\lambda = 1.3\mu$m. Now, consider the effect of a perturbation, such as a change in ambient temperature, on the refractive index of the fiber. This causes a phase change

$$d\phi = \frac{dn \cdot 2\pi \cdot L \cdot \Omega}{c} \quad \text{[rad]} \tag{2}$$

Multiplying the numerator and denominator on the r.h.s by $d\lambda$, gives

$$d\phi = \frac{2\pi \cdot L \cdot \Omega \cdot d\lambda}{c} \cdot \left(\frac{dn}{d\lambda}\right) \quad \text{[rad]} \tag{3}$$

By writing $d\lambda$ in terms of the laser frequency ν, the phase deviations may be expressed in terms of the laser frequency deviations, which have the same (random) time

dependence. Thus

$$d\phi(t) = d\nu(t) \cdot \frac{2\pi \cdot L \cdot \Omega \cdot \lambda^2}{c^2} \cdot \left(-\frac{dn}{d\lambda}\right) \qquad [\text{rad/s}] \quad (4)$$

For the analysis of frequency stability, it is more convenient to look at the last expression in the frequency domain by Fourier transforming as follows:

$$S_\phi(f) = S_\nu(f) \cdot \left[\frac{-2\pi \cdot L \cdot \Omega \cdot \lambda^2}{c^2} \cdot \frac{dn}{d\lambda}\right]^2 \qquad [\text{rad}^2/\text{Hz}] \quad (5)$$

In this expression, $S_\phi(f)$ is the spectral density of the phase fluctuations at an offset frequency f from the RF signal, induced by the spectrum of random frequency deviations, $S_\nu(f)$ of the laser.

The variation of the effective fiber index of refraction with wavelength $\frac{dn}{d\lambda}$ depends on the waveguide parameters and material composition of the fiber. The measured result for typical single mode fiber at 1300 nm is [11]

$$\frac{dn}{d\lambda} = 270.1 \qquad \text{m}^{-1} \quad (6)$$

Inserting this value into Equation (5) and substituting the appropriate constants produces the simple relation

$$S_\phi(f) = S_\nu(f) \cdot L^2 \cdot \Omega^2 \cdot (1.02 \times 10^{-51}) \qquad [\text{rad}^2/\text{Hz}] \quad (7)$$

The above quantity is the mean square phase noise spectral density at an offset f from the modulation signal, Ω. This is the spectrum which would be observed if a perfect oscillator (i.e., an oscillator with no phase noise) modulated the laser, and the output of the photodetector was compared to a second perfect oscillator as in a phase noise measurement system.

Numerical Estimates

The FM noise spectrum of distributed feedback-type (DFB) single mode lasers typically used in fiber optic distribution systems exhibits a power independent $1/f$ character at low frequencies (below about 1 MHz). In the modulation band the FM noise is white and inversely proportional to optical power [10]. The physical mechanism responsible for the $1/f$ behavior is thought to be trapping of carriers due to impurities and interfacial boundaries, but is not fully understood. The white portion of the spectrum is due to spontaneous emission events and is adequately modelled by theory [5,8].

The frequency noise spectrum of typical DFB lasers has been measured experimentally [10]. The above mentioned physical mechanisms may be modelled as

$$S_\nu(f) = \frac{C}{P} + \frac{K}{f} \quad (8)$$

where P is the average output power of the laser and f is the frequency offset from the optical carrier. C and K are empirically determined constants. For the Fujitsu DFB laser diodes measured [10], $C = 1.5 \times 10^4$ [Hz · W], and $K = 5.8 \times 10^{11}$ [Hz^2].

The frequency noise spectrum of a typical DFB laser is illustrated in Figure 1. As laser power is increased, the white portion of the noise spectrum decreases proportional to P^{-1}. A numerical estimate of the additive RF phase noise requires that only the $1/f$ portion of $S_\nu(f)$ be inserted into Equation (7), which gives

$$\begin{aligned} S_\phi(f) &= \frac{5.8 \times 10^{11}}{f} \cdot \Omega^2 L^2 (1.05 \times 10^{-51}) \ [\text{rad}^2/\text{Hz}] \\ &= L^2 \cdot \Omega^2 \cdot \frac{5.9 \times 10^{-40}}{f} \qquad [\text{rad}^2/\text{Hz}] \end{aligned} \quad (9)$$

The $1/f$ laser frequency noise is converted to $1/f$, or "flicker" phase noise in the fiber optic distribution system. This level of $1/f$ phase noise depends on the inherent quantum fluctuations of the laser frequency and represents the ultimate phase noise floor of the system.

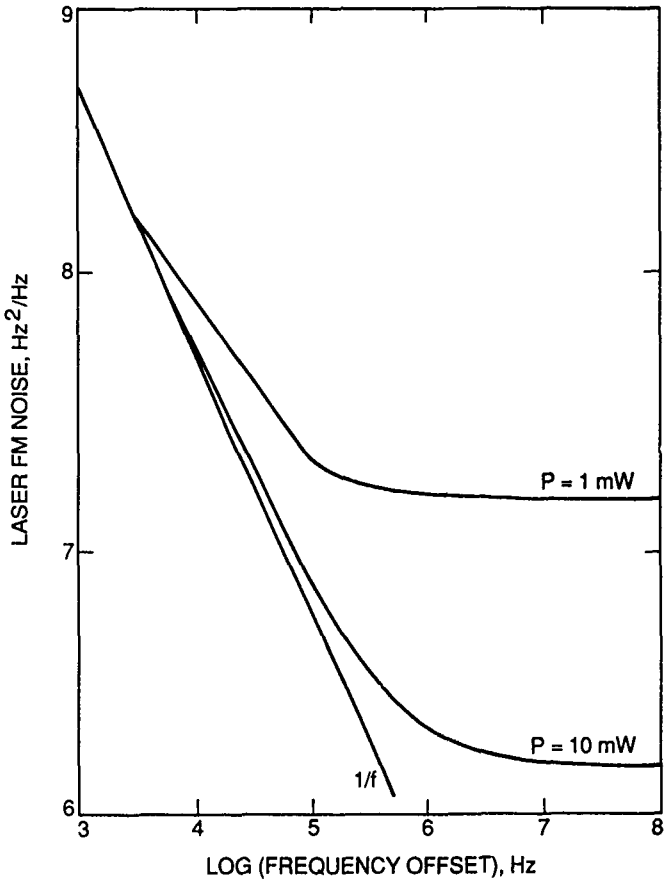


Figure 1. Typical DFB laser frequency noise spectrum.

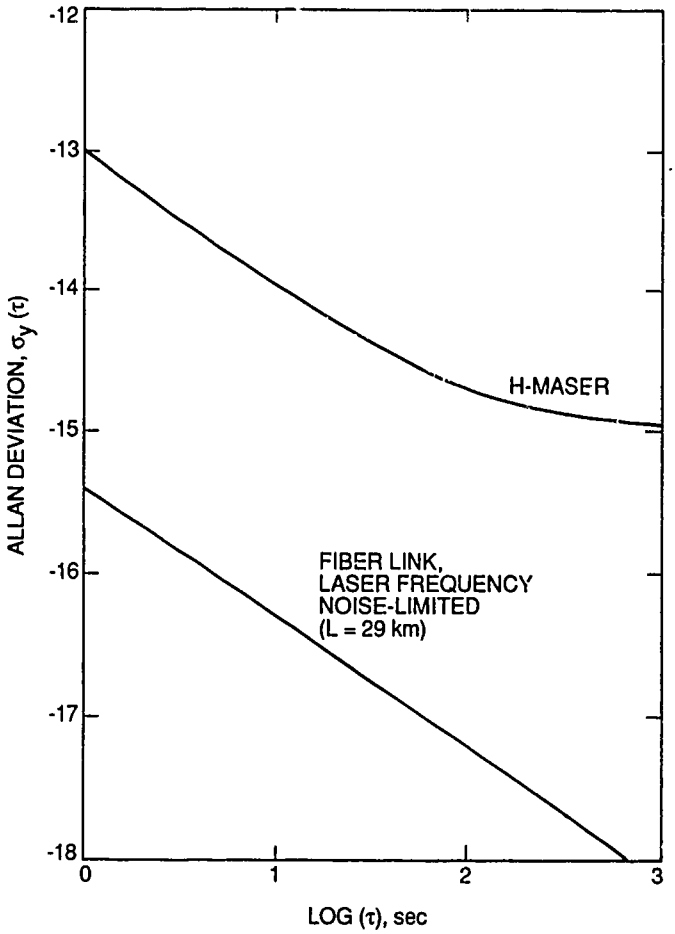


Figure 2. Hydrogen maser and laser FM noise–limited fiber optic link frequency stability comparison.

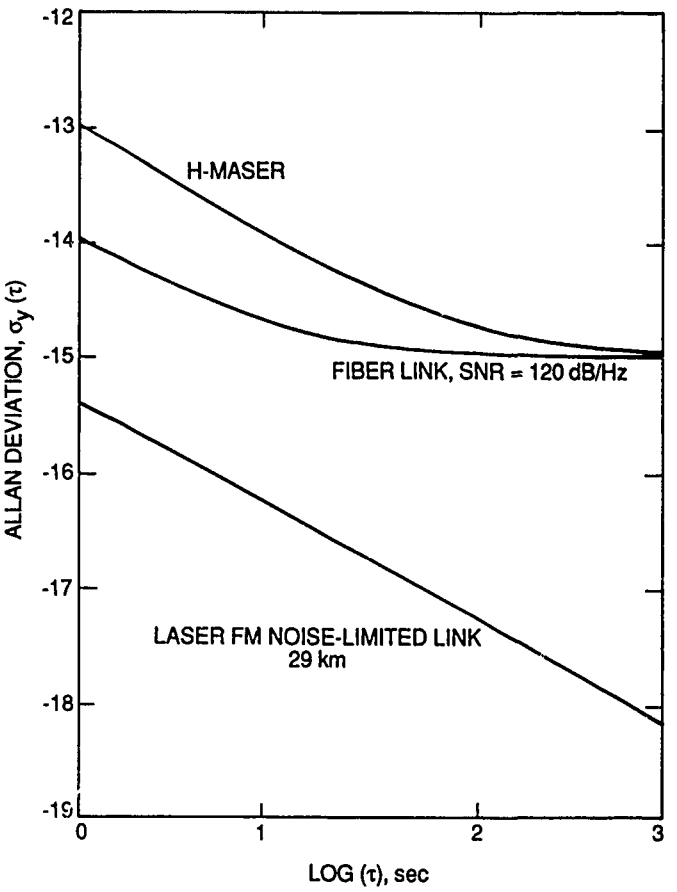


Figure 3. Comparison of Allan deviation: H-maser, actual 29 km link, theoretical FM noise–limited link.

For flicker phase noise, the Allan deviation (square root of the Allan variance) may be calculated from the following relation [12]:

$$\sigma_y(\tau) = \sqrt{\frac{3}{(2\pi)^2\tau^2} \cdot \frac{S_\phi(f) \cdot f}{\Omega^2} \cdot ln(8.88 \cdot f_h \cdot \tau)} \quad (10)$$

where f_h is the frequency cutoff of the phase noise; in this case, one half the bandwidth of the output filter.

Substituting Equation (9) into the last expression, the modulation frequency cancels and the expression for the Allan deviation reduces to

$$\sigma_y(\tau) = \sqrt{\frac{3}{(2\pi)^2 \cdot \tau^2} \cdot L^2 \cdot 6 \times 10^{-40} \cdot ln(8.88 \cdot f_h \cdot \tau)} \quad (11)$$

The laser induced flicker phase noise thus sets the minimum bias level of the $1/\tau$ section of the Allan deviation plot. The fact that the last expression does not depend on the RF modulation frequency, Ω, is significant. This implies that moving to higher or lower modulation frequencies for reference signal distribution will not alter the laser frequency noise "floor" of the Allan deviation.

For purposes of comparison, we consider an actual frequency distribution link in use at the JPL/NASA Goldstone Deep Space Communications Complex. The longest frequency distribution run is 29 km. Assuming that the output filter bandwidth f_h is 10 Hz, the Allan deviation, calculated from Equation (11), is

$$\sigma_y(\tau) \simeq \frac{4.1 \times 10^{-16}}{\tau} \quad (12)$$

The relation between the laser frequency noise–limited Allan deviation of the 29 km link and the Allan deviation of a typical hydrogen maser is plotted in Figure 2. This represents the ultimate frequency stability attainable with such a link, provided all other noise sources are negligible.

Present State of the Art

The ultimate link stability plotted in Figure 2 will only be attained if all other noise sources are negligible. In reality, other noise sources do contribute to the link stability. This is illustrated in Figure 3, where an actual measurement of the 29 km Goldstone link is plotted in addition to the maser and the ultimate stability link curves of Figure 2.

In present-day systems, the Allan deviation $1/\tau$ intercept is set by the SNR of the fiber link, which is determined by the laser intensity noise. The SNR of a typical high performance analog link is 120 dB/Hz. Figure 4 illustrates the 1-second Allan deviation as a function of fiber length. It is immediately apparent from Figure 4 that the laser frequency noise does not limit frequency distribution system performance at this time, since the laser SNR dominates. As lower amplitude-noise lasers become available, the laser frequency noise floor of the Allan deviation may begin to limit frequency reference distribution system performance.

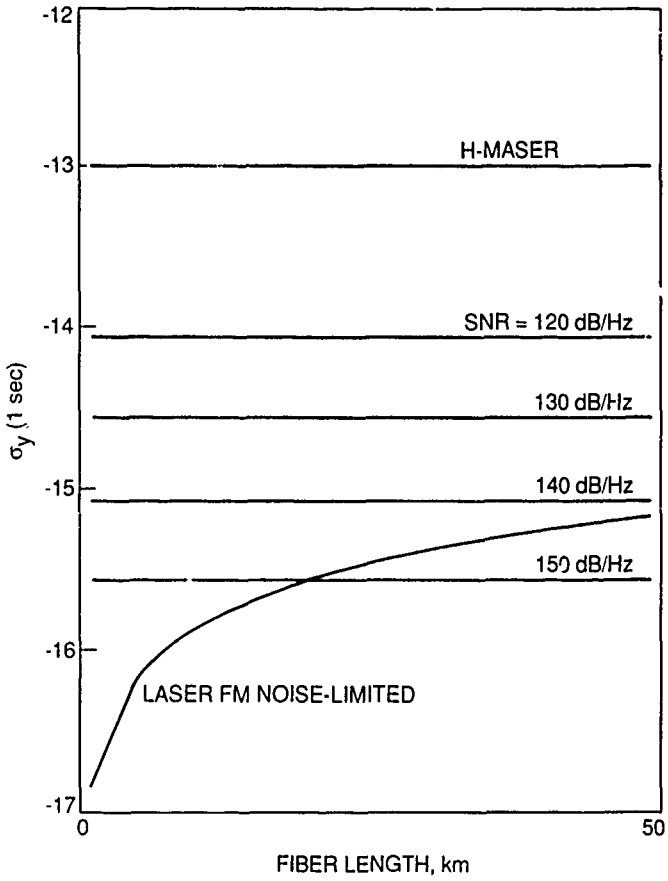


Figure 4. Comparison of frequency stability at one second for maser, FO links of various SNR, and laser FM noise-limited vs. fiber length.

Figure 5 depicts fiber link Allan deviation at 1 second, as a function of link SNR. The laser relative intensity noise (RIN) sets the SNR of the fiber link for short distances [13]. Also shown is the Allan deviation floor due to laser frequency noise for a 29 km link. This plot shows clearly that laser frequency noise limits the frequency stability "floor" of the fiber link to $4 \times 10^{-16}/\tau$ for SNR $\geq$ 145 dB/Hz. Systems with as high as 150 dB/Hz SNR may be achievable with externally modulated high power semiconductor diode-pumped Nd:YAG solid state lasers, or through the use of squeezed light generated directly from semiconductor lasers. As these system improvements are realized, the low frequency $1/f$ FM noise of the laser may begin to limit system performance.

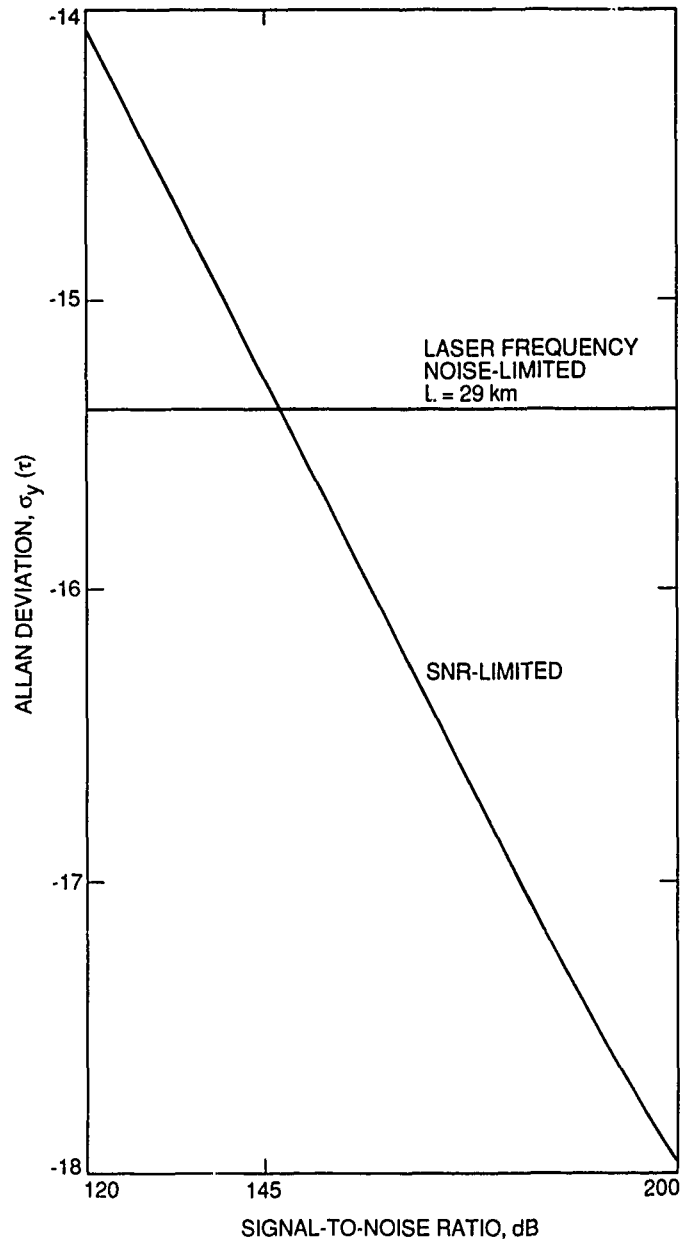


Figure 5. Comparison of fiber link frequency stability at one second vs. SNR with laser FM noise-limited 29-km link.

A final observation: Since the fiber optic transmission system converts laser frequency noise to RF phase noise, it may be the case that a stabilized fiber optic link comprises a very accurate system for measuring the frequency deviations of lasers close to the optical carrier. This approach is under consideration for future research.

Conclusion

At present, the noise floor of fiber optic distribution systems is determined by the laser signal to noise ratio

(SNR) in the RF modulation band. However, lasers with lower relative intensity noise (RIN) or the use of squeezed light, promise increases in link SNR and passive and active temperature stabilization schemes can improve link stability at long averaging times. As these improvements in components and systems are realized, the fundamental limit for frequency stability due to laser frequency noise may be reached.

The present analysis provides the contribution to the phase noise of a transmitted frequency reference due to single–mode laser source frequency deviations. Through interaction with the dispersion of the fiber, the $1/f$ FM noise close to the optical carrier is converted to $1/f$ phase noise close to the RF reference signal. The additive $1/f$ laser–induced phase noise is a function of the fiber dispersion and length, and determines the ultimate Allan deviation floor of the fiber optic distribution system in the 1 to 100 second region.

For the longest fiber optic frequency distribution link in the NASA Deep Space Network (29–km), using data for commercially available DFB lasers, the analysis indicates the link Allan deviation is limited to $4 \times 10^{-16}/\tau$ (for averaging times between 1 second and 100 seconds). This stability limit will be reached at link SNR of 145 db/Hz, which is 25 dB better than the present system.

Further increases in SNR will not yield higher link stability unless laser frequency noise is decreased as well. The laser FM noise stability limit is two orders of magnitude higher stability than the best current frequency standard, which indicates that laser frequency noise will not limit fiber optic frequency distribution capability in the forseeable future.

Acknowledgments

The authors thank G. J. Dick, R. Sydnor, C. Greenhall, and L. Primas for numerous helpful discussions.

References

[1] L.E. Primas, G. Lutes, R. Sydnor, "Fiber Optic Frequency Transfer Link," Proc. of the 42nd Annual Symposium on Frequency Control, 1988.

[2] J. Prestage, G. Janik, G. Dick, L. Maleki, "New Ion Trap for Frequency Standard Applications," Proc. of the 43rd Annual Symposium on Frequency Control, 1989.

[3] G.J. Dick, R.T. Wang, D.M. Strayer, "Operating Parameters for the Superconducting Cavity Maser," Proc. of the 20th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, December 1988.

[4] L. Chinlon, G. Eisenstein, et. al., "Fine structure of frequency chirping and FM sideband generation in single- longitudinal-mode semiconductor lasers under 10–GHz direct intensity modulation," Applied Physics Letters, 46 (1), 1 January 1985.

[5] K. Vahala, A. Yariv, "Semiclassical Theory of Noise in Semiconductor Lasers: Part I, Part II," IEEE Journal of Quantum Electronics, Vol. QE-19, No. 6, June 1983.

[6] J.C. Cartledge, G.S. Burley, "The Effect of Laser Chirping on Lightwave System Performance," IEEE Journal of Lightwave Technology, Vol. 7, No. 3, March 1989.

[7] D.A. Atlas, A.F. Elrefaie, M.B. Romeiser, D.G. Daut, "Chromatic dispersion limitations due to semiconductor laser chirping in conventional and dispersion–shifted single–mode fiber systems," Optics Letters, Vol. 13, No. 11, November 1988.

[8] B. Daino, P. Spano, M. Tamburrini, S. Piazzolla, "Phase Noise and Spectral Lineshape in Semiconductor Lasers," IEEE Journal of Quantum Electronics, Vol. QE-19, No. 3, March 1983.

[9] R. Schimpe, W. Harth, "Theory of FM Noise in Single- Mode Injection Lasers," Electronics Letters, Vol. 19, No. 4, 17 Feb. 1983.

[10] K. Kikuchi, "Effect of $1/f$–type FM noise on semiconductor laser linewidth residual in high–power limit," IEEE Journal of Quantum Electronics, Vol. 25, No. 4, April 1989.

[11] Corning Telecommunications Products Data Sheet, "SMF-21 CPC3 Single Mode Optical Fiber", Corning, New York, February, 1987.

[12] "Characterization of Frequency Stability," NBS Technical Note 394, (1970).

[13] K.Y. Lau, "Signal-to-Noise Calculation for Fiber Optics Links" TDA Progress Report 42-58, Jet Propulsion Laboratory, Pasadena, California, May–June 1980.

43rd Annual Symposium on Frequency Control - 1989

POSITIONING OF GPS ANTENNAS IN TIME-KEEPING LABORATORIES OF NORTH AMERICA

W. LEWANDOWSKI, Bureau International des Poids et Mesures (BIPM), Pavillon de Breteuil, F-92312 Sevres Cedex, France

R. J. DOUGLAS, National Research Council (NRC), Montreal Road Camp, Ottawa, Ontario, Canada KIA0R6

W. J. KLEPCZYNSKI, United States Naval Observatory (USNO), 34th and Massachusetts Ave., N.W., Washington, DC 20392-51000

W. STRANGE, National Geodetic Survey (NGS), Rockwall Building, Rockville, Maryland 20852

J. SUTER, Johns Hopkins University, Applied Physics Laboratory (APL), Johns Hopkins Road, Laurel, Maryland 20707

M. A. WEISS, National Institute of Standards and Technology (NIST), 325 Broadway, Boulder, Colorado 80303

ABSTRACT

One of the problems with the use of the Global Positioning System (GPS) for time transfer is that it is a one-way system. In addition, most of the time-keeping laboratories of the world use only the L1 frequency. However, the use of GPS in the common view approach diminishes the impact of some of the errors such as orbit error, GPS clock error and ionospheric error, in the one-way system. But the common view approach does not cancel the antenna coordinate error.

The Bureau International des Poids et Mesures (International Bureau of Weights and Measures, BIPM) has developed a method of differential positioning using the data of time comparisons themselves. The consistency of the coordinates is within 30 cm for distances up to 1000 km. The agreement with space geodesy positioning for such distances has been verified within involved uncertainties. The method was applied to the European laboratories one year ago. Since then the obtained coordinate corrections have been used for the current BIPM computations of time comparisons in Europe. The consistency of time comparison improved from about 10 ns to about 2 ns

The principles of this technique and the results of its application to the North American time laboratories are presented in this paper. Work on differential positioning by geodetic double-frequency receivers, between U.S. Naval Observatory and Maryland Very Long Baseline Interferometry (VLBI) point, are reported.

1. INTRODUCTION

The Global Positioning System (GPS) now in general use for regional and intercontinental atomic clock comparisons is a one-way system of time transfer. This means that the errors of satellite clock, satellite position, ionospheric delays, and ground antenna coordinates have a direct impact on the accuracy of time transfer between satellites and earth stations.

---

*) Acronym meanings are listed at the end of the text.

However, the simultaneous observations of GPS satellites (called the "GPS common-view technique") cancel or diminish some of these one-way system errors [1]. But the common-view approach does not cancel or diminish the ground-antenna coordinate errors. Thus the adoption of good ground-antenna coordinates appears to be an important factor in the accuracy of time comparisons by GPS satellites.

With the present state of the art of atomic clocks it is desirable that their comparisons be at the level of few nanoseconds of accuracy. In order to achieve such accuracy, the error due to antenna coordinates should not exceed 1 to 2 ns in the global budget of errors of the common-view technique.

To attain this goal the antenna coordinates must fulfill the following requirements:

(a) They must be accurately determined in a common homogeneous geodetic reference system. Uncertainties should be of the order of 30 cm or less.

(b) In order to reduce the residual errors of the common view method, the satellite and antenna coordinates should be expressed in the same geodetic reference system; for example WGS 84 is currently used for the broadcast ephemerides of GPS satellites. But this requirement is less strong than (a): WGS 84 is known with an accuracy of only about two meters with respect to the GPS monitor station network.

To realize the positioning at such a level between antennas located on different continents, the most accurate geodetic technologies, such as Very Long Baseline Interferometry (VLBI) or Satellite Laser Ranging (SLR), are required. The TRANSIT positioning is not sufficiently accurate. Inside the continents or regions a less expensive differential positioning, giving desired uncertainty, can be applied.

We can imagine the following approach for access of time laboratories to accurate coordinates:

(a) realizations of differential positionings between the antennas of the principle timing centers within a given area and

(b) realizations of differential positioning between those principal timing centers of a continent or a region and a VLBI site or SLR site located in the concerned area.

In practice however, the time-keeping laboratories are often using TRANSIT positioning (± 1 m) or a navigation solution provided by GPS time receivers (± several meters) or other types of coordinates not accurate enough.

Recent studies [5,6] have shown that over distances of about 1000 km or less, a main source of biases in the GPS time comparisons is the adoption of wrong geodetic coordinates of the antennas. In BIPM a method of differential positioning over the distances of 1000 km or less has been developed by using data of time comparisons themselves. This method was applied to the European time laboratories one year ago. Since then, the obtained coordinate corrections have been used for the current BIPM computations of time comparisons in Europe, improving their consistency from about 10 ns to about 2 ns.

To examine and eventually establish the homogeneity of antennas' coordinates of North American time-keeping laboratories, several actions have been undertaken:

(a) An analysis of the antennas' coordinates of these laboratories by the BIPM method has been realized; the resulting coordinate corrections are reported here.

(b) A differential positioning by geodetic dual frequency receivers between the U.S. Naval Observatory and, the Maryland Point VLBI site, 20 km away, has been realized by NGS; final reductions of the data should be completed by mid-June 1989.

(c) In order to realize a differential positioning by the BIPM method between NIST in Boulder and the Platteville VLBI site 50 km away, NIST located an atomic cesium clock and GPS time receiver in Platteville in April 1989; the results of this positioning will be available later this year.

We hope these actions will reduce to 1 to 2 ns the impact of errors of antenna coordinates on the accuracy of time comparisons within North America. In addition the links with VLBI sites are a step toward improved accuracy of GPS time links with other continents.

The planned degradation of the GPS system, especially the degradation of the broadcast ephemerides, raises the question of the use, for time comparisons, of post-computed corrections to the broadcast ephemerides. One of the possible ways to resolve this problem is to compute the corrections to the broadcast ephemerides by time comparisons themselves. This approach is facilitated by the use of accurate coordinates of ground stations.

2. DIFFERENTIAL POSITIONING BY BIPM APPROACH

2.1. Biases due to Errors of Antenna Coordinates

For the comparisons of remote clocks by GPS satellites, the time laboratories apply a program of simultaneous trackings, called "common-views," which cancels the role of the satellite clocks and diminishes the errors of satellite position and ionospheric delay.

Let A and B be the positions of the GPS antennas of two laboratories (also designated by A and B) and S_k the position of a satellite they track. The unit vectors of AS_k and BS_k are $\vec{a}_k$ and $\vec{b}_k$ (fig. 1). The true value of the clock comparison at the instant of tracking is designated by

clock reading of A - clock reading of B = U_{ab}.

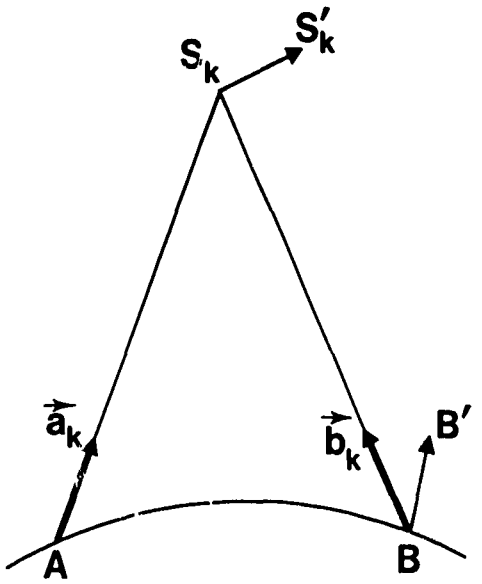


Fig. 1. Errors in the positions of satellite and antenna.

It is assumed throughout section 2 that the relativistic conversion in geocentric coordinate time has been made.

If the broadcast ephemerides correspond to the erroneous position S'_k, the measured time comparison $(U_{ab})_k$ is erroneous, and we define

$$E_k = U_{ab} - (U_{ab})_k . \qquad (1)$$

E_k is given by

$$E_k = - c^{-1} \overrightarrow{S'_k S_k} \bullet (\vec{a}_k - \vec{b}_k), \qquad (2)$$

where c is the velocity of light. For a given $\overrightarrow{S'_k S_k}$, E_k is roughly proportional to the chord |AB|.
If the adopted coordinates for B correspond not to B, but to B', the measured time comparison requires a correction F_k defined as in (1)

$$F_k = - c^{-1} \bullet \overrightarrow{B'B} \bullet \vec{b} , \qquad (3)$$

independent of the location of A.

It is possible by a theoretical approach to estimate the amount of E-type errors over short distances by considering their amount over long distances.

The laboratories of the U.S. Naval Observatory (USNO) and the Paris Observatory (OP) are 6000 km apart. One may compute the differences between the clocks of these laboratories for the same day as given by different common-views (different satellites, or the same satellite at different locations in the sky). The standard deviation of these computations with respect to the mean is of the order of 10 ns (15 ns in 1985, 7 ns in 1987). If we assume that this standard deviation is entirely due to the satellite ephemerides, which would be the worst case for our study, the influence

of the E-type errors theoretically should reduce to 1 ns (one sigma) over distances of 600 km. Averaging over the usual 15 to 20 scheduled common-views leads to a quite negligible contribution.

Let us comment on another aspect of time comparisons by GPS satellites: the repeatability of the same geometry of observations from day to day for periods of several months. The scheduled common-views (SCV) are repeated every 23 h 56 m, so that on account of the sidereal orbits of the GPS satellites, the satellites are observed every sidereal day at nearly the same location on the sky. The common-view schedule is kept without change for about 6 months; then a new schedule is issued.

Repeated time comparisons, each usually averaged over a duration of 13 min, by the same satellite at the same sidereal time typically have standard deviations of 3 ns. But systematic differences reaching 30 to 40 ns appear between the results of measurements using different satellites, or the same satellite at different times. These differences are revealing sources of systematic errors.

As discussed above, over distances up to 1000 km the errors due to the uncertainties of the satellite position theoretically should not exceed 1 ns. Over such distances, the two paths of the signal through the refractive media are similar and we expect that the errors of the differential refraction are small. We assume also that the long series of repeated measurements further decrease the influence of the errors of the satellite ephemerides and differential refraction delays. They also should reduce somewhat the influence of multipath effects. It appears that the observed constant biases persisting during the months over short distances are entirely due to an error of differential antenna coordinates (F - type error).

2.2 Determination of Differential Coordinates

For the same day, we assume that the $(U_{ab})_k$ are all transferred at the same instant (0 h UTC, for instance). Usually the adopted time comparison is

$$\hat{U}_{ab} = \frac{1}{n} \sum_{k=1}^{n} (U_{ab})_k = U_{ab} - \Delta T, \qquad (4)$$

with the unknown ΔT defined by

$$\Delta T = \frac{1}{n} \sum_{k=1}^{n} F_k . \qquad (5)$$

From (4) and (3), we arrive at the equation

$$c^{-1} \bullet \overrightarrow{B'B} \bullet \vec{b} + \Delta T = (U_{ab})_k - \hat{U}_{ab} . \qquad (6)$$

The right side of (6) is obtained every day. On the left side, the unknowns are ΔT and the components DX, DY and DZ of $\overrightarrow{B'B}$ in the usual coordinate system: 0 at the geocenter, OZ toward the pole, OX in the prime meridian, and OY toward East. More precisely, DX, DY and DZ are corrections to add to the adopted coordinates of B in order to express them in the same frame as the coordinates of A (differential positioning).

The $\vec{b}$ in equation (6) are easily expressed in the OXYZ system, using the elevation and azimuth of S, which are included in the standard format for GPS data exchanges. However most of the GPS receivers give elevation and azimuth for instants which do n ot correspond to the center of tracking interval. Corrections are easily derived from these data themselves with the additional knowledge of rough values of the orbit inclination and of the terrestrial longitude of the ascending nodes. These corrections have been made.

We have also taken into account the displacement of stations A and B due to Earth tides, though this effect turns out to be negligible at the level of accuracy we are using here.

The system of equations (6) is solved by the minimum-least-squares method.

2.3. Comparison of BIPM Differential Solution with Space Geodesy Positioning

Four European time laboratories have recently established the local links between space geodesy sites and the center of phase of the GPS time receivers' antennas. These links have made it possible to check the results of the BIPM differential solution.

We give here, for an example, the differential positioning between two French laboratories, the OP and, located 635 km away (chord), OCA at Grasse.

At the OP a Doppler technique was used for absolute positioning by TRANSIT satellites (uncertainty ± 0.5 m), and in OCA laser ranging was used to LAGEOS satellite (uncertainty ± 0.1 m).

In both locations the local links between space geodesy points and GPS antennas have been realized with an uncertainty of ± 0.2 m.

The differences between geodetic coordinates and the coordinates introduced (adopted) into the receivers are following:

OP(Doppler) - OP(adopted)
DX = -0.33 m
DY = 3.19 m (7)
DZ = 0.63 m
uncertainty ± 0.5 m.

OCA(laser)- OCA(adopted)
DX = -3.12 m
DY = 2.64 m (8)
DZ = 1.91 m
uncertainty ± 0.1 m.

If the Doppler and laser positioning are perfectly accurate, the difference between the results in (7) and (8) represents the corrections which should be added to the relative coordinates of OP - OCA.

They are:

DX = -2.79 m
DY = -0.55 m
DZ = 1.28 m
total uncertainty ± 0.5 m.

The determination of these relative coordinate corrections, by the BIPM method realized over the period 18 July 1988 to 14 December 1988, is:

$$DX = (-3.01 \pm 0.18)\ m$$
$$DY = (-0.09 \pm 0.07)\ m$$
$$DZ = (\ 0.52 \pm 0.17)\ m,$$

where the uncertainty stated here is the statistical deviation in the solution.

Given the various uncertainties, the agreement between two determinations is quite satisfactory.

Similar agreements have been found for other pairs of laboratories having the links between GPS antennas and space geodesy sites. The self consistency of the BIPM method appears to be 30 cm. To determine accuracy, this method needs to be compared with a method that is more accurate.

2.4. Application to North American Laboratories

Proceeding with the above method, we have determined the relative coordinates between four time laboratories in North America. The analysis is based on the data of the common-view schedule No. 10, extending from 15 December 1987 to 22 June 1988.

The distances between laboratories are shown in Table I. We notice that the distances between NIST and the other laboratories are more than 2000 km. In this case, the results should be used with caution. They are nevertheless given because the stability of the biases seems to indicate a predominant contribution of coordinate errors.

TABLE I - Distances between the laboratories considered in this study in thousands of kilometers (chords).

	APL	NIST	NRC
NIST	2.41		
NRC	0.70	2.44	
USNO	0.03	0.73	0.73

The results of this differential positioning are given by Table II. Errors of up to 12 m are noted.

TABLE II - Results of the differential coordinates determination between laboratories considered in this study.

Unit: 1 meter.

	DX	σX	DY	σY	DZ	σZ
USNO - APL	1.02	0.11	0.99	0.23	-7.52	0.42
USNO - NRC	9.47	0.10	-2.96	0.24	4.05	0.41
USNO - NIST	-2.12	0.16	5.06	0.35	-3.39	0.61
NIST - APL	2.71	0.16	-3.60	0.40	-5.02	0.69
NIST - NRC	10.95	0.12	-7.84	0.30	6.59	0.51
APL - NRC	8.40	0.12	-4.60	0.24	11.91	0.49

Table III gives the closure by triangles of laboratories. A good consistency of differential solutions including NIST is observed.

TABLE III - Closures of differential coordinates given by Table II by triangles of laboratories.

Unit: 1 meter.

	DX	σX	DY	σY	DZ	σZ
USNO - APL	1.02	0.11	0.99	0.23	-7.52	0.42
USNO - NRC	9.47	0.10	-2.96	0.24	4.05	0.41
APL - NRC	8.40	0.12	-4.60	0.24	11.91	0.49
Closure	-0.05		-0.65		0.32	
USNO - NIST	-2.12	0.16	5.06	0.35	-3.39	0.61
NIST - NRC	10.95	0.12	-7.84	0.30	6.59	0.51
USNO - NRC	9.47	0.10	-2.96	0.24	4.05	0.41
Closure	-0.64		0.18		-0.85	
NIST - APL	2.71	0.16	-3.60	0.40	-5.02	0.69
NIST - NRC	10.95	0.12	-7.84	0.30	6.59	0.51
APL - NRC	8.40	0.12	-4.60	0.24	11.91	0.49
Closure	0.16		-0.36		0.30	
NIST - APL	2.71	0.16	-3.60	0.40	-5.02	0.69
USNO - APL	1.02	0.11	0.99	0.23	-7.52	0.42
USNO - NIST	-2.12	0.16	5.06	0.35	-3.39	0.61
Closure	-0.43		0.47		-0.89	

This analysis is based on a schedule which has not been optimized for the positioning. Only 14 SCV have been used. This explains the quite large uncertainties of differential positioning (up to 0.5 m). Using an optimized schedule (a good geometry and 24 SCV or more) has led to uncertainties below 0.2 m [6].

Special comments are needed for the USNO results. This laboratory is equipped with a receiver which is still using parameters of the WGS 72 coordinate system instead of WGS 84, the system adopted by GPS since January 1987. Thus antenna coordinates enterred in the receiver in the geodetic form ϕ, λ, h, are transformed into cartesian coordinates X, Y, Z using the wrong ellipsoid parameters. In December 1988, the cartesian coordinates X, Y, Z expressed in WGS 84, were directly introduced into the USNO receiver, thus avoiding the use of WGS 72 ellipsoid parameters. Before this date the USNO coordinates should be corrected by:

$$DX = 0.36\ m$$
$$DY = -1.58\ m$$
$$DZ = 1.05\ m$$

These corrections have been applied to the results given by tables II and III.

Another problem arises in the computation of the position of satellites by receivers using WGS 72 parameters. The cartesian coordinates X, Y, Z of the satellite are computed by the receiver software from the broadcast Keplerian elements using WGS 72 values of the universal gravitational parameter and the earth's rotation rate. The estimated error in the satellite position due to the use of wrong parameters is of the order of 2 to 3 m. If the error vector for each satellite has a constant value and direction, it could introduce an error of a similar amount to the differential coordinates computed by BIPM method. This possible error was not removed and could exist in the differential coordinates between

USNO and other laboratories, provided by Tables II and III.

2.5. Use of Coordinate Corrections for Accurate Time Comparisons

Let us consider again a pair of laboratories. We refer to the notation and equations in section 2.1. Instead of using ΔT from the coordinate solution to correct the $\hat{U}_{ab}$, it is better for time transfer to correct each $(U_{ab})_k$ by applying the coordinate corrections to, let us say, B by (3). This is easily done by computation, using the known elevation and azimuth of S_k.

Fig. 2 illustrates the improvement brought by application of coordinate corrections.

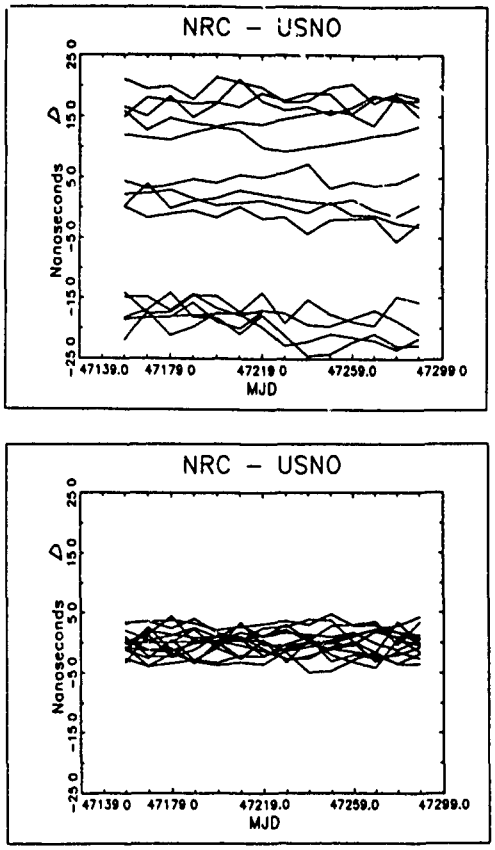


Fig. 2. D is $(Ua_b)_k - \hat{U}_{ab}$ of equation (6), averaged over 10-day intervals. One line is drawn for each k, above with the adopted coordinates, below after coordinate correction.

A bias could remain due to the remaining error of the differential coordinates of B with respect to A. It is maximum for $\overrightarrow{B'B}$ in the direction of $\overrightarrow{BS'_k}$.

A conservative estimate is: $c^{-1} |\overrightarrow{B'B}|/2$, corresponding to $\overrightarrow{B'B}$ vertical and uniform distribution of S_k on the sky B'B is evaluated by

$$\overrightarrow{B'B} = (\sigma_x^2 + \sigma_y^2 + \sigma_z^2)^{\frac{1}{2}},$$

where the σ's are those of the differential positioning. If we take the portion of the error represented by the statistical deviations in Table II, we get a total uncertainty of 2 ns due to positioning for the pair laboratories in that list. This may be optimistic in that we have not fully established the accuracy of this method.

3. LINK OF NORTH AMERICAN LABORATORIES WITH VLBI SITES

The two principal time centers NIST and USNO can be linked to two VLBI primary sites of the BIH Terrestrial System (BTS).

BTS has been established by BIH, with the aim of providing the best possible world-wide reference frame by giving the coordinates of a small number of sites where the most accurate positioning are permanently used (very long baseline interferometry, lunar laser ranging, satellite laser ranging).

The uncertainties of BTS primary sites range from about 5 to 20 cm. The coordinates of sites are revised annually and referred to epoch 1984.0 with a model of tectonic plate motions. This model can be used also to bring the coordinates to date.

The 1987 issue of the BTS, designated BTS (1987), and the formulas for the transformation of BTS coordinates into WGS 84 are given in the BIH Annual Report for 1987 [2].

Since 1 January 1988, the maintenance of the BTS has become one of the tasks of the International Earth Rotation Service (IERS), and the name BTS changed to IERS Terrestrial Reference Frame (ITRF).

3.1. Link Between USNO and Maryland Point VLBI Site

In April 1985 the double-frequency geodetic GPS receivers were used to connect the Maryland Point VLBT Site (BTS site reference: VLBI 7277) to the time transfer antenna at the USNO located 20 km away. The geodetic GPS receivers located a point directly below the time transfer antenna relative to Maryland Point, with the height of the time transfer antenna above this point determined by direct measurement. This connection between the Maryland Point VLBI site and the USNO time transfer antenna is expected to be accurate to 2 to 5 cm. Final reductions of the GPS data should be completed by mid-June 1989.

3.2. Link Between NIST and Platteville VLBI Site

In order to realize the differential positioning by the BIPM approach between the NIST GPS time receiver and the VLBI site in Platteville, 50 km away (BTS site reference: VLBI 7258), NIST installed an atomic cesium clock and a GPS time receiver in Platteville in April 1989.

A specially conceived tracking schedule for the positioning, having 30 tracks has been introduced to both receivers in Boulder and Platteville. The GPS antenna in Platteville is located at the level of the ground without any obstacles around. This location seems to have reduced the noise due to multipath interference to 2 to 3 ns. The antenna in Boulder is shielded by an anti-reflecting plane, reducing noise of multipath to 3 to 5 ns.

The expected consistency of this differential positioning is about 30 cm.

Because the differential positioning by BIPM method requires several months of observations, the results will be available in the Fall, 1989.

4. UNCERTAINTIES OF GPS TIME COMPARISONS

Although it is not the purpose of this paper to discuss in detail other sources of uncertainties in GPS time comparisons [13], we will mention them briefly.

The differential receiver delays can be calibrated with uncertainties of the order of 1 ns [3,10,15,12]. However more recent studies have revealed systematic and random differences between different type of GPS time receivers [4,9].

For short distances, with observations spread in time and direction, we assume that the systematic component of the differential refraction, not corrected by the model, is below 1 ns. Nevertheless this assumption should be verified by measurements of ionospheric delay by dual frequency GPS receivers.

For long distances, a model to compute the ionospheric delay is insufficient, and the measurements of ionosphere by dual-frequency receivers are necessary [7,8].

Over short distances, the errors of satellites' positions are well cancelled by the common-view approach, and contribute less than 1 ns to the uncertainties of time comparisons. The use of post-processed ephemerides can reduce the impact of the ephemerides errors on long distance time comparisons [11].

Multipath propagation is another source of uncertainties. An appropriate location or shielding by anti-reflecting planes of GPS time receivers' antennas can significantly reduce this phenomenon.

5. CONCLUSIONS

(a) The uncertainties of GPS time comparisons due to the antenna positioning can be reduced to the level of 1 ns to 2 ns by the adoption of accurate coordinates (± 0.3 m).

(b) The TRANSIT Doppler positioning (± 1 m) is not sufficient. The navigation solution using the GPS time receiver positioning software (± 6 m) is worse.

(c) Inside a continent or a region (distances between laboratories up to 1000 km) a differential positioning by BIPM approach can be realized from the time comparisons themselves with a consistency of about 0.3 m or less.

(d) The GPS antennas of the principal timing centers of a continent or a region should be linked to VLBI or SLR sites by accurate local geodetic ties (± 0.05 m).

(e) Corrections to the adopted coordinates can be easily performed by software when computing the time differences between sites. We therefore recommend not changing adopted coordinates. All affected laboratories should be informed if any changes are made.

(f) The accurate coordinates of GPS time receivers located on different continents may permit the computations of the corrections to the broadcast ephemerides, assuming that the problems of ionospheric delay errors and multipath have been resolved.

The degradation of GPS system, planned for the near future, makes the improvement of deteriorated ephemerides a matter of urgent interest.

(g) It is important to investigate other sources of uncertainties of GPS time comparisons such as ionospheric delay, satellite position, multipath interference and anomalies of GPS time receivers' software.

6. ACKNOWLEDGEMENTS

The authors wish to express their gratitude to David Allan, Dick Davis, and Judah Levine from NIST for their most helpful advice and their help in the realization of the link between NIST and Platteville. The authors thank the staff of National Geodetic Survey for their work in performing the differential positioning between USNO and the Maryland Point VLBI site.

7. REFERENCES

[1] Allan, D.W. and Weiss, M.A., 1980, "Accurate time and frequency transfer during common-view of a GPS satellite," Proc. 34th Annual Symposium on Frequency Control, Fort Monmouth, NY, pp. 334-346.

[2] BIH Annual Report for 1987, Paris, June 1988.

[3] Buisson, J.A., Oaks, O.J., Lister, M.J., 1985, "R - Cats a new method for calibrating Global Positioning System (GPS) remote sites," Proc. of the 17th PTTI meeting, Washington, D.C., pp. 201-222.

[4] Granveaud, M., Lewandowski, W., Uchrich, R., Tourde, R., 1989, "Comparison of GPS Time Receivers: A User's Point of View," to be published in the Acts of the 3rd European Frequency and Time Forum, Besancon.

[5] Guinot, B. and Lewandowski, W., 1987, "Use of the GPS time transfer at the Bureau International des Poids et Mesures," Proc. 19th PTTI meeting, Redondo Beach, Ca, pp. 3-12.

[6] Guinot, B. and Lewandowski, W., 1988, "Nanosecond time comparisons in Europe using the GPS," Acts, 2nd European Frequency and Time Forum, Neuchatel, pp. 187-193.

[7] Imae, M., Lewandowski, W., Thomas, C., Miki, C., 1988, "A dual frequency GPS receiver measuring ionospheric effects without code demodulation and its application to time comparison," Proc. of the 20th PTTI meeting, Tysons Corner/Vienna, Virginia, pp. 77-86.

[8] Imae, M., Miranian, M., Lewandowski, W., Thomas, C., 1989, "A dual frequency codeless receiver measuring ionospheric effects and its application to time comparisons," to be

published in the Acts of the 3rd European Frequency and Time Forum, Besancon.

[9] Kirchner, D., Ressler, H., Fassl, S., 1989, "Experience with two collocated C/A-code GPS receivers of different type," to be published in the Acts of the 3rd European Frequency and Time Forum, Besancon.

[10] Lewandowski, W., Weiss, M.A. and Davis, D., 1986, "A calibration of GPS equipment at time and frequency standard laboratories in the USA and Europe," Proc. of the 18th PTTI meeting, Washington, D.C., pp. 265-279, [also published in Metrologia, 24, pp. 181-186 (1987)].

[11] Lewandowski, W. and Guinot, B., 1989, "GPS Time Comparisons, Test on the Use of Precise Ephemerides," Technical Note for the 11th Session of Comite Consultatif pour la Definition de la Second, Sevres.

[12] T. Morikawa, et. al., 1989 "Calibration of the Delay Time in the GMS/GPS Time Transfer Receivers Using Portable Reference Receivers," IEEE I & M, 38, No. 2, April 1989, pp. 661-665.

[13] Nard, G., Rabian, J., Gounon, R., 1987, "Utilisation des signaux du GPS en mode differentiel instantane pour les applications temps-frequence de haute precision," Acts, 1st European Frequency and Time Forum, Besanson, pp. 237-243.

[14] Weiss, M.A., 1987, "Apparent Diurnal Effects in the Global Positioning System," Proc. of the 19th PTTI meeting, Redondo Beach, California, pp. 33-48.

[15] Weiss, M.A. and Davis, D., 1988, "A Calibration of GPS Equipment in Japan," Proc. of the 20th PTTI meeting, Tysons Corner/Vienna, Virginia, pp. 101-106.

8. ACRONYMS USED IN THE TEXT

APL Applied Physics Laboratory, Laurel, Maryland, USA.
BIH Bureau International de l'Heure, Paris, France.
BTS BIH Terrestrial System
BIPM Bureau International des Poids et Mesures, Sevres, France.
GPS Global Positioning System
IERS International Earth Rotation System
ITRF IERS Terrestrial Reference Frame
NGS National Geodetic Survey, Rockville, Maryland, USA.
NIST National Institute of Standards and Technology, Boulder, Colorado, USA.
NRC National Research Council, Ottawa, Ontario, Canada.
OCA Observatoire de la Cote d'Azure, Grasse, France.
OP Onservatoire de Paris, Paris, France.
USNO United States Naval Observatory, Washington D.C., USA.
WGS World Geodetic System

43rd Annual Symposium on Frequency Control - 1989

THE MODULAR INTELLIGENT FREQUENCY, TIME AND TIME INTERVAL (MIFTTI) SUBSYSTEM PROGRAM

K.D. Lyon, W.J. Riley, and J.R. Vaccaro
EG&G Frequency Products
Salem, MA 01970

ABSTRACT

The Modular Intelligent Frequency, Time and Time Interval (MIFTTI) Subsystem is a modular time and frequency subsystem that includes built-in intelligence to improve its performance. Modularity is used to achieve maximum configuration flexibility. Intelligence is used to achieve maximum time and frequency performance from the available frequency sources by providing automatic fault detection, synchronization, and syntonization, and by providing compensation for such systematic errors as aging and temperature sensitivity.

This paper describes the continuing research on the MIFTTI Subsystem being sponsored at EG&G by the U.S. Army Electronics Technology and Devices Laboratory (ETDL) under Contract No. DAAL01-87-C-0715.

A primary objective of this program is to study the MIFTTI Subsystem concepts and to define an architecture for it. This is a key task because the choices made will shape all the hardware and software that follow. The overriding goal has been to provide sufficient flexibility to allow the inclusion of all reasonable features while making those specific choices necessary to build real and practical hardware. Hardware is being designed and two MIFTTI Subsystems are being built to test the control system software and evaluate the efficacy and tradeoffs associated with the MIFTTI concepts.

The first year of this 30-month program has concentrated on defining the architecture and packaging concepts of the MIFTTI Subsystem, design of the basic crystal oscillator and rubidium frequency references, and development of the basic control software. The architectural, hardware, and software work so far has resulted in the construction and test of an "Intelligent Rubidium-Crystal Oscillator (RbXO)" that is being used to verify the MIFTTI Subsystem concepts. The XO can be continuously or periodically syntonized to the RFS, and the RFS can be syntonized to an external reference, all under software control. Compensation for XO and RFS aging and temperature sensitivity is included, as is display of XO and RFS monitors. The results are encouraging, and show the potential for significant improvements in the capabilities of military time and frequency subsystems by the incorporation of these concepts.

INTRODUCTION

The Modular Intelligent Frequency, Time and Time Interval (MIFTTI) Subsystem program[1] is a research effort to study the advantages and methods of adding intelligence to a time and frequency subsystem.

A simplified block diagram of the MIFTTI Subsystem is shown in Figure 1. An oscillator drives frequency distribution, time of day, and time interval outputs under the microcomputer control. An external reference may be applied for higher stability and to update the subsystem time and frequency. Power for the subsystem is provided by either external dc or a battery via a dc/dc converter.

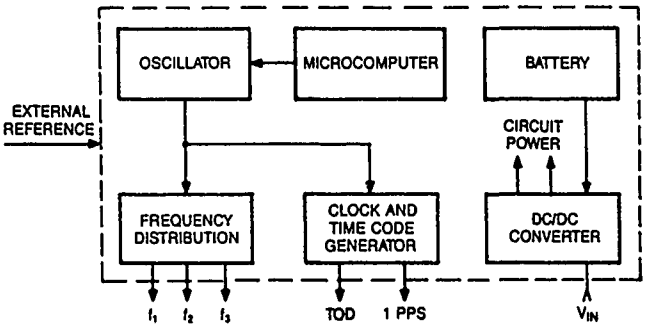


Figure 1. MIFTTI Subsystem Block Diagram.

ARCHITECTURE

The architecture of the MIFTTI Subsystem is described in the following sections of this paper. The MIFTTI architecture is arranged to combine all elements of a time and frequency subsystem under microcomputer control. A detailed block diagram of the complete MIFTTI Subsystem is shown in Figure 2. The elements of the MIFTTI Subsystem are arranged as separate modules that allow for maximum flexibility of its overall configuration. A list of the MIFTTI module types is shown in Table 1. Besides the Chassis, only three of the modules (Power Supply, Microcomputer, and A/D Converter), plus at least one external or internal frequency source, are required to configure a useful system. The frequency source may be a crystal oscillator (XO), a rubidium frequency standard (RFS), an external reference (such as a cesium frequency standard or GPS receiver), or any combination thereof. The system can be enhanced by adding accessory modules for a particular application such as a Battery, a Clock and Time Code Generator (C/TCG), and Distribution Amplifiers.

Table 1. MIFTTI Module Types.

Category	Modules
Essential	Chassis Power Supply Microcomputer A/D Converter
Source	Crystal Oscillator Rubidium Standard External Reference
Accessory	Battery Distribution Amplifier(s) Clock/Time Code Generator Spare(s)

Power/Data Bus. A vital aspect of the MIFTTI Subsystem architecture is the power and data bus arrangement that interconnects and runs the various MIFTTI modules. Means must be provided for powering all anticipated and future modules in a way that is both flexible yet

CH2690-6/89/0000-225 $1.00 C 1989 IEEE

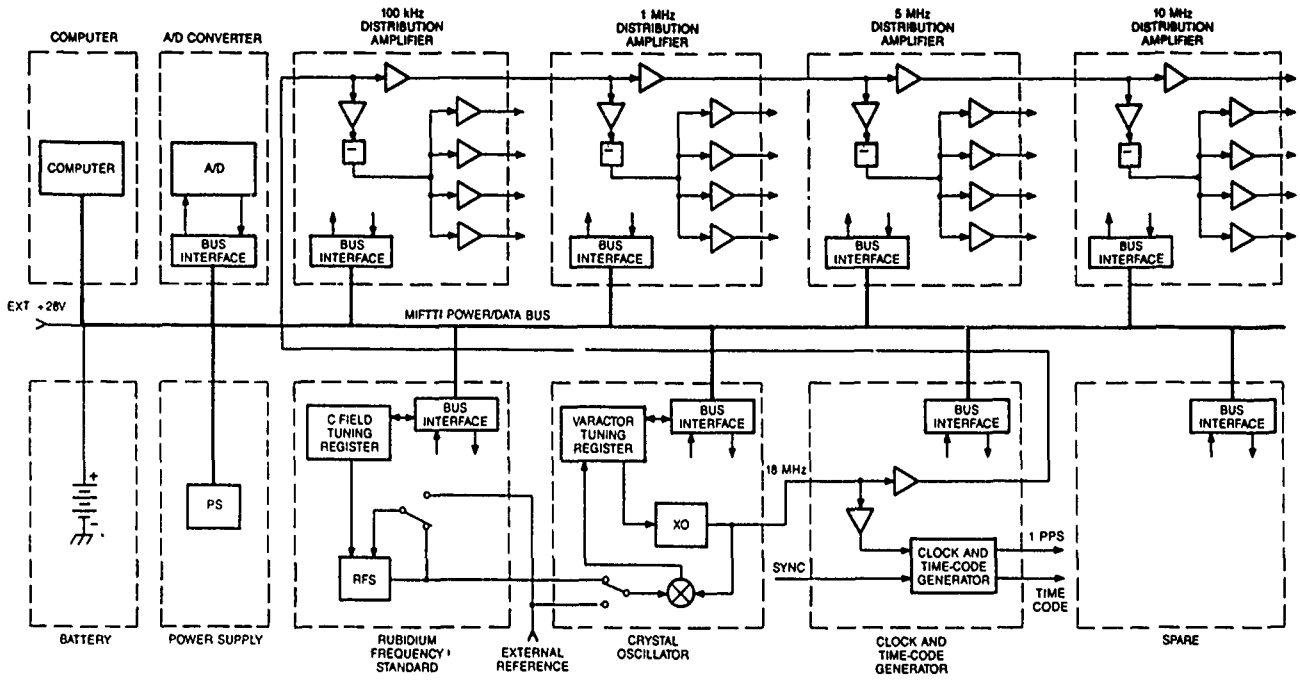


Figure 2. MIFTTI Subsystem Detailed Block Diagram.

efficient. Sufficient choices of voltage and current levels are provided to supply a variety of heater, analog, and digital loads from a single dc input with provisions for battery backup. Means are also provided for the exchange of information between MIFTTI modules. A simple, yet flexible, serial data transmission system was conceived for this purpose.

Power Supply and Bus. The MIFTTI Subsystem is powered by an external +28 V dc supply. Power is supplied to the MIFTTI modules via the MIFTTI power bus, which distributes dc power at +28 V, ±15 V, +12 V, and +5 V. These supply choices are sufficiently flexible for all circuit requirements, and all MIFTTI modules are designed to operate from them. All bus supplies are derived with switching power supply techniques except for the +28 V bus, which is the main power input to the subsystem.

The +28 V bus is intended to supply power to the RFS and any other large loads (up to 170 W total). This line is diode-steered from the +28 V input under normal operation and from the battery module (if included) if the main dc input fails. The RFS module contains a power control switch which can be programmed by the MIFTTI Subsystem to turn the RFS off under battery operation. The external +28 V supply itself can also be battery-backed.

One dc/dc converter supplies +12 V from the internal +28 V bus. This supply is intended primarily for the oven and oscillator power of the crystal oscillator or other relatively high power loads (30 W total). Another dc/dc converter supplies the remaining ±15 V and +5 V MIFTTI buses. The ±15 V supplies are intended primarily for analog circuits (each 1.5 W total) and the +5 V supply is intended primarily for digital circuits (2.5 W total), including those of the MIFTTI data bus interfaces.

Data Bus. The MIFTTI Subsystem uses a unique data bus that emphasizes simplicity and low power. Data transfers within the MIFTTI Subsystem are quite slow by normal computer standards, and the number of addresses required is very low. This dictates a simple serial data transmission arrangement with Data In and Data Out lines and a Clock line, along with three Address and two Write lines that define addresses for up to 16 MIFTTI modules.

A block diagram of the MIFTTI data bus is shown in Figure 3. Data transfers from the microcomputer involve clocking out the desired number of data bits on the Data Out line, which propagate into parallel output shift registers in each module. When the appropriate number of data bits have been sent, they are latched into the addressed registers by a strobe pulse from the address decoder. Data transfers into the microcomputer operate in a similar fashion. A control bit commands the connection of the output of parallel input shift registers in one module to the Data In line, and their data are clocked into the microcomputer.

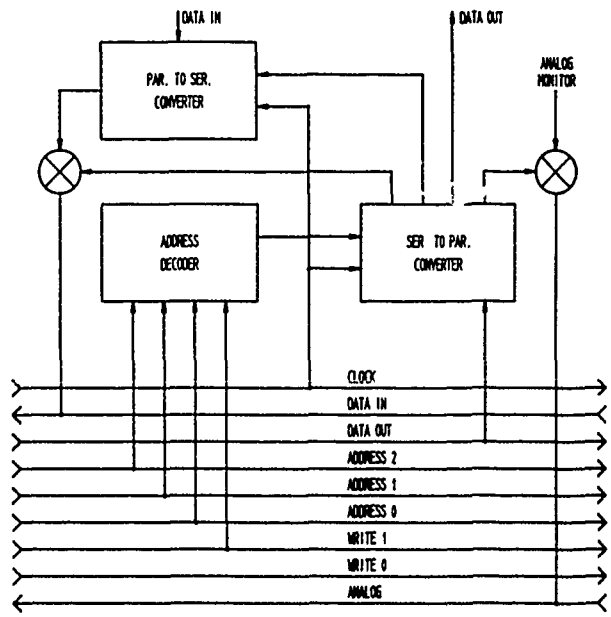


Figure 3. MIFTTI Data Bus.

A special feature of the MIFTTI data bus is an analog line that can be read via an A/D converter by the microcomputer. Analog switches in the various MIFTTI modules can connect analog signals, such as monitors, to this analog line one at a time under control of the microcomputer.

HARDWARE

The hardware under development for the MIFTTI Subsystem is described in the following sections. Common features of the MIFTTI hardware design include (a) packaging into standard modules, (b) interfacing to the MIFTTI power/data bus, and (c) outputting of ID information that uniquely identifies each module.

Crystal Oscillator. Most MIFTTI configurations will contain a Crystal Oscillator (XO) module. The baseline XO for the present hardware is the TMXO[2], but the MIFTTI Subsystem can be adapted to use other comparable voltage-controlled crystal oscillators (VCXOs).

There are two modes of operation for the MIFTTI XO module. In the first operating mode, it can be operated as a computer-controlled crystal oscillator, in which case the control voltage is set by the MIFTTI microcomputer. A local temperature transducer allows the MIFTTI Subsystem to make intelligent frequency compensations to the TMXO due to variations in ambient temperature. In addition, using modeling techniques, the MIFTTI Subsystem can compensate for oscillator aging. This mode of XO operation enhances the effectiveness of the system because of the crystal oscillator's ability to operate without constant access to a superior reference. This can thereby reduce the size, cost, or power consumption required for a certain level of timing stability.

The second mode of XO operation allows the crystal oscillator to be locked to a local or external reference. Although this process is initiated by the MIFTTI Subsystem microcomputer, the computer has little to do with the actual syntonization. XO syntonization is accomplished by means of phase locked loop and digital follower hardware. The microcomputer is able to monitor the process by reading the 12-bit up/down counter via a parallel to serial converter shift register. These data can then be stored along with temperature data in the MIFTTI microcomputer to characterize such parameters as TMXO aging and temperature coefficient.

The Crystal Oscillator board also has logic to monitor the input levels of the local and external references, the varactor voltage of the TMXO, and the ambient temperature.

Rubidium Frequency Standard. The Rubidium Frequency Standard is an optional MIFTTI Subsystem module for applications where the high stability of an atomic frequency standard is required. It consists of a RFS and circuitry to interface it with the MIFTTI Subsystem.

The MIFTTI Rubidium Frequency Standard module uses a modified version of the EG&G Model RFS-10 Rb standard. This RFS has had widespread application in military communications systems. It is also the Rb reference that was used in the original RbXO[3,4], where it proved its stability and reliability while undergoing severe temperature and on-off cycling. The interface circuitry for the Rb Reference Module is primarily to allow tuning of the Rb reference via the MIFTTI data bus. This is accomplished by a tuning register and D/A converter that controls the C-field current of the Rb physics package.

Distribution Amplifiers. Distribution amplifiers at standard frequencies of 0.1, 1, 5, and 10 MHz are planned as optional modules for the MIFTTI Subsystem. Each distribution amplifier module is driven at 10 MHz and has four outputs at a selectable output frequency. Each output level is +7 dBm nominal with harmonic distortion components suppressed by at least 30 dBc. The outputs are isolated from each other by at least 60 dB. An important feature of the distribution amplifier architecture is that the 10 MHz drive is a 5 volt CMOS signal that may be "daisy-chained" to drive any number of modules.

Clock/Time Code Generator. A Clock and Time Code Generator (C/TCG) is planned as an optional module for the MIFTTI Subsystem. Several approaches to the C/TCG are under consideration. These include (a) the use of a commercially available C/TCG board or module, (b) the use of the single-chip CMOS gate array C/TCG used in the U.S. Army SCOTT MILSTAR terminal, or (c) the development of a new general-purpose C/TCG chip.

One attractive possibility is to develop a C/TCG device that could be down-loaded by the MIFTTI host microcomputer with instructions for a specific time code format. This could result in a practical approach to a "universal" device that, augmented with high-speed or specialized interface circuits, could serve in a wide variety of applications.

A/D Converter. The A/D Converter is required in nearly all MIFTTI Subsystem configurations, since it is needed to measure the monitor and other signals on the analog line of the MIFTTI data bus, and it plays an essential role in the syntonization of the RFS to an external reference.

Computer. A single board computer using the Intel™ 80C188 microprocessor has been selected as the MIFTTI Subsystem microcomputer. This board has a MS-DOS™ compatible BIOS that provides low level I/O routines compatible with the IBM™ PC. It has all the necessary memory and port capabilities to support the MIFTTI Subsystem requirements.

Packaging. The primary objectives of the MIFTTI Subsystem packaging are (a) modularity to facilitate reconfiguration both during the study and for various applications, and (b) the use of standardized, commercially available hardware to minimize design and fabrication costs and to ensure the availability of add-on or replacement parts.

An outline drawing of a MIFTTI Subsystem is shown in Figure 4, which illustrates the packaging approach and shows one possible configuration. The chassis is 17.0" wide, 5.2" high, and 16.7" deep. It can be fitted with cover plates for use as a table-top case, or mounted in 19" racks or cabinets. Modules can be installed from the front and back, with guides ensuring proper alignment of the module and back-plane connectors. The connectors comply with DIN 41612 and have both standard and coaxial contacts. The modules are available in various types and widths, provide protection and EMI shielding, and accept 100 x 160 mm Eurocard boards.

SOFTWARE

Software is probably the most important aspect of the MIFTTI Subsystem because it supplies the intelligence that provides the flexibility, automatic features, and performance gains that are the program objectives. It was therefore decided that the software effort should be done in a well-structured, highly professional manner emphasizing flexibility, portability, and maintainability. This led to three fundamental

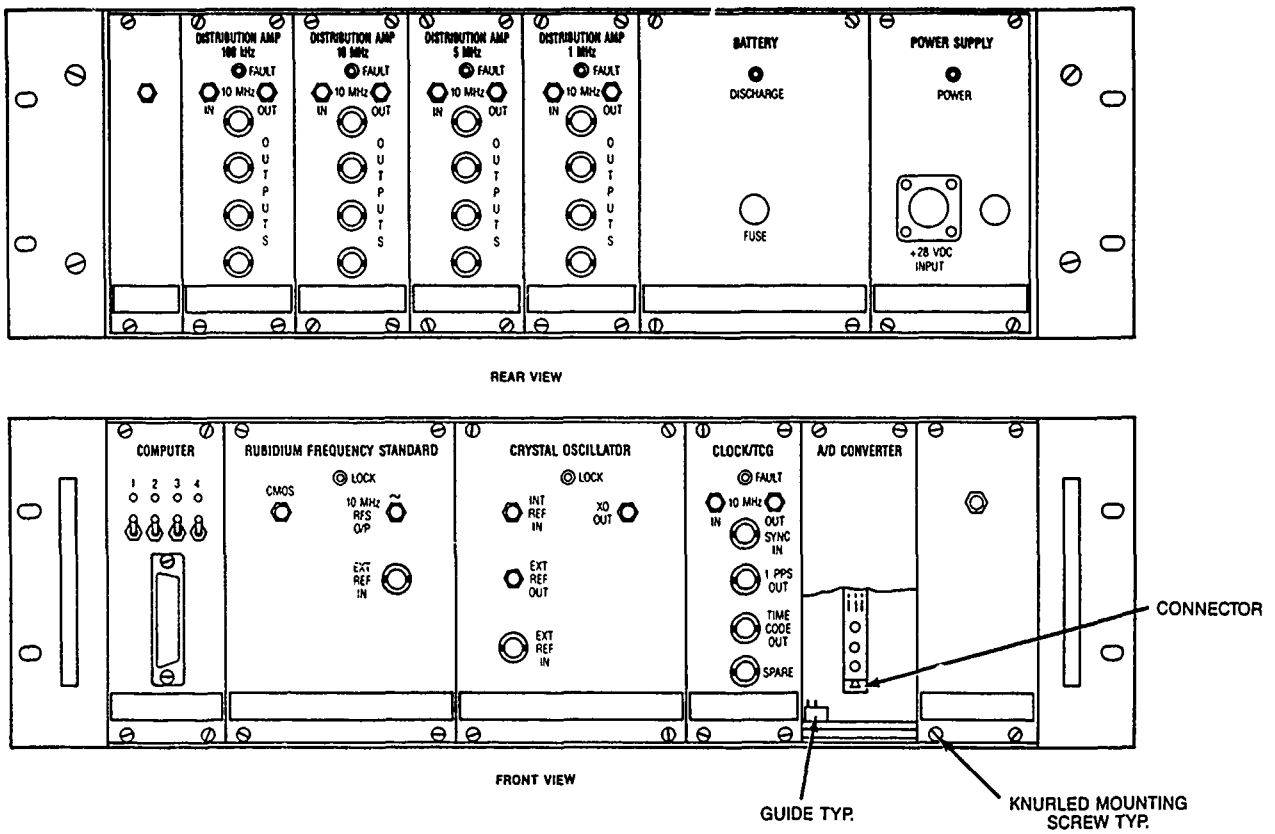


Figure 4. MIFTTI Subsystem Outline Drawing.

choices: (a) the use of a professional programming organization, (b) the use of the C programming language, and (c) the use of the 8086 family of microcomputers and a PC compatible BIOS system.

C is a highly portable programming language well suited for control system applications. C programs are organized into well-structured functions, and the language includes means for hardware bit manipulations. As a compiled language, it results in fast and compact code. PC compatibility allows access to many software tools and hardware options, and allows the MIFTTI Subsystem hardware to be operated by an external PC compatible computer.

These choices have proven to be sound. The selection of a highly professional programming organization has greatly enhanced the quality of the software. The C language is very satisfactory for this application, especially when augmented with a commercially available multi-tasking operating system and other software tools. The major software modules are discussed in more detail below.

Operating System. The Operating System portion of the MIFTTI software provides low level real-time executive functions such as keyboard, event, timer, CRT, printer, port I/O, and EEPROM tasks. Many of these functions are provided by a commercially available multi-tasking operating system and a set of general-purpose software tools.

Application Support. The purpose of the Application Support software is to insulate the details of how things are done from the application software. This approach enhances the maintainability, flexibility, and debugability of the software.

Control Program. The Control Program is the heart of the MIFTTI software, since it is the code that determines most of the subsystem's functionality. This software controls the configuration and operation of the MIFTTI hardware, performing such functions as periodic syntonization of the XO to the RFS, or one-time syntonization of the RFS to the external reference. Flexibility and maintainability are critical for this code since it is likely to undergo frequent change as new features are added.

Modeling. The Modeling block is potentially the most sophisticated portion of the MIFTTI Subsystem software and is the area in which the most research and experimentation will occur. In a sense, all of the current hardware and most of the software are being developed to allow exploration of the advantages of modeling as applied to a time and frequency subsystem. It is primarily through the modeling and subsequent compensation that stability performance improvements will be gained.

Two approaches are being taken regarding the MIFTTI Subsystem modeling software. The first approach (part of this EG&G effort) involves straightforward feedforward compensation of aging and temperature sensitivity for the XO and RFS frequency sources. The second approach (recently completed at Ball Corporation) involves the use of Kalman filter methods to estimate and forecast the time and frequency states of the MIFTTI Subsystem.[5,6] Since both software efforts are using the C programming language and there have been informal meetings held to coordinate these efforts, it is expected that the second approach will eventually be ported into the actual MIFTTI Subsystem hardware.

Certain fundamental principles apply to either approach. Syntonization provides actual frequency correc-

tion of a source (XO or RFS) when a more stable reference is available (RFS or external). This syntonization also provides data for the modeling process. When tagged with the time and temperature, and the syntonized source control voltage is interpreted in terms of frequency, the syntonization performs a measurement for the modeling process. (The microcomputer real time clock provides the time tag, a sensor in the source module provides the temperature data, and prior characterization of the source tuning curve provides the frequency value.)

Simple feedforward (non-adaptive) modeling would use prior characterization of the frequency versus temperature dependence of the XO or RFS to compensate the source for temperature variations. (The modeling software will probably have a learning mode to facilitate the storage of the source temperature characteristic.) Such compensation for temperature (or any other deterministic error such as supply voltage or barometric pressure) would precede any other modeling.

The first modeling approach would then apply an aging compensation factor (based on a single linear drift value or other simple model) between syntonizations. The second approach would use each new syntonization to update the system state vector and make optimal state estimates between syntonizations (which do not have to occur regularly).

Monitoring. The Monitoring portion of the MIFTTI software will monitor the status of the subsystem and make that data available to an external printer, terminal, or PC for logging and analysis. Off-line analysis of monitor data (such as aging corrections) can be done on an external PC using off-the-shelf software packages.

OPERATION

The MIFTTI Subsystem may be operated with several combinations of frequency sources, as shown in Table 2. The frequency sources available determine the possible operating modes of the subsystem. A summary of the MIFTTI Subsystem operating modes is shown in Table 3. Four basic operating modes are defined, depending on what frequency source(s) is included in the particular MIFTTI Subsystem. These MIFTTI Subsystem operating modes are described in more detail below.

Table 2. MIFTTI Source Combinations.

Frequency Sources			Mode Description
Ext.	RFS	XO	
X	X	X	Syntonized RbXO
X	X		Syntonized RFS
X		X	Syntonized XO
X			External
	X	X	RbXO
	X		Rb
		X	XO

Review of RbXO Operating Principles. A RbXO [3,4] is a voltage-controlled crystal oscillator (VCXO) that operates continuously and is occasionally syntonized by a rubidium reference that can remain off most of the time. This concept combines the best features of a rubidium frequency standard (fast warmup and low aging) with those of a crystal oscillator (low power and small size). While not an essential part of a MIFTTI Subsystem, the RbXO is a basic building block for many such

Table 3. MIFTTI Subsystem Operating Modes.

Mode	Sub-Mode	Sources Ext	Rb	XO	Function	Remarks
XO	Manual			X	XO frequency adjustable via user interface	Like ordinary XO plus modeling to correct aging and TC
	Ext. Synt.	X		X	Syntonize XO to external reference	Periodic or continuous syntonization
Rb	Manual		X		RFS frequency adjustable via user interface	Like ordinary RFS plus modeling to correct aging and TC
	Ext. Synt.	X	X		Syntonize RFS to external reference	Periodic syntonization only
RbXO	Synt. by Rb		X	X	Syntonize XO to Rb	Periodic or continuous syntonization
	Ext. Synt.	X	X	X	Syntonize Rb or XO to external reference	When external reference available
Ext.		X			MIFTTI Subsystem driven by external reference	

units. A MIFTTI Subsystem can, however, add significant new features (modularity and intelligence, for example) to a RbXO. The subsystem intelligence can further enhance the power conservation advantages of the RbXO concept.

XO Modes. The simplest useful MIFTTI operating mode is as an ordinary crystal oscillator (except that the XO frequency is controlled by tuning word data rather than a trimmer or external control voltage). At startup, the previously stored XO tuning word is loaded from EEPROM. The XO frequency can then be manually adjusted via the user interface. The operator can command a certain tuning word or a certain frequency, either absolute or incremental. These frequency adjustments will use XO tuning characteristic data previously stored in EEPROM (as a lookup table with linear interpolation).

If an external reference is available, the MIFTTI Subsystem can use it to calibrate (syntonize) the XO. (XO syntonization by the Rb Reference is discussed under the RbXO mode below.) This external XO syntonization can be selected to take place either continuously or periodically. If the external reference controls the XO continuously, then the XO serves as a "flywheel," and would be available to take over if the external reference is lost. (This ability is enhanced greatly by the MIFTTI aging and TC modeling.) The control logic must respond gracefully to the loss of the external reference at any time, causing the XO to operate at the best valid tuning condition.

XO syntonization is accomplished by a digital phase lock loop (PLL), with no direct involvement by the MIFTTI microcomputer. This PLL hardware establishes the correct tuning register word to syntonize the XO to the external or Rb reference. The resulting tuning word can then be read by the microcomputer.

The XO PLL loop is relatively slow so that the phase noise of a high quality crystal oscillator is not degraded by the noise of the reference. The pattern of external XO syntonization can be either (a) one-shot syntonization on command, or (b) continuous syntonization while the external reference is present.

Rb Modes. The operating modes of a MIFTTI Subsystem with only a Rb source are similar to those with only a XO. The RFS frequency is programmable via a C-field tuning register that is initialized from EEPROM at turn on and can be manually adjusted via the user interface. The C-field tuning characteristic is stored in the EEPROM and used to allow absolute or incremental frequency adjustments.

If an external reference is available, the MIFTTI Subsystem can use it to syntonize the RFS. This syntonization uses a successive approximation process that uses the sense of the frequency difference to sequentially adjust the bits of the C-field register to the correct setting. The frequency sense information comes from the RFS servo integrator via the MIFTTI analog monitor bus A/D converter.

RFS syntonization will take place only periodically or in response to a command. Continuous RFS syntonization is not desirable because the successive approximation process disturbs the frequency, and the inherent high stability of the RFS makes frequent syntonization unnecessary. During normal operation, RFS syntonization is most likely (a) when an external reference is temporarily available, or (b) when a significant period of time or change in temperature has taken place. The system can detect the presence of an external reference and it must gracefully restore the previous C-field word if the external reference is lost during a syntonization cycle.

RbXO Modes. The basic RbXO mode also closely resembles conventional RbXO hardware and operates normally without direct MIFTTI microcomputer involvement. Unlike most RbXO configurations, however, the Rb section is a complete RFS (with its own crystal oscillator), not just a Rb Reference. This allows maximum MIFTTI Subsystem flexibility and permits the XO PLL to treat the RFS and external references identically. It also allows the use of the RFS module by itself, and offers a convenient way to use the separate XO module to isolate the system from RFS noise, syntonization disturbance, or failure.

The MIFTTI intelligence does, however, offer many new possibilities for optimum control of RbXO operation. These should certainly include a choice of continuous or periodic operation of the RFS under user control. Great flexibility also exists in the criteria for "syntonization complete" (fixed time, oven status, tuning word stability, etc.) as well as syntonization control (significant temperature change, etc.).

The RbXO can be externally syntonized in either of two ways. The first syntonizes the Rb C-field as described above for the RFS alone. During this process, however, the XO serves as a buffer to isolate the system from the Rb disturbances during its syntonization. This is the preferred method, but it does require a programmable C-field register and a servo integrator monitor. This limits the method to a programmable RFS that can be adjusted to the correct frequency.

An alternative syntonization method is to introduce an offset into the XO tuning word. The XO is first locked to the external reference, and then to the RFS. The difference in the tuning register words is stored and subtracted, and the RFS offset is calculated knowing the XO varactor characteristic. The system then operates at the correct frequency. Subsequent XO aging is corrected for by another RFS syntonization followed by subtraction of its frequency offset. This method can be used with a non-programmable RFS (and even one that is significantly off frequency but stable).

Power Modes. Intelligent power control is another aspect of the MIFTTI Subsystem operation. In particular, the state of the Rb reference is an important programmable parameter during battery operation. A RbXO configuration would probably, but not necessarily, switch to XO-only operation under battery power. An ongoing syntonization cycle might be completed, however, or the Rb reference might turn on infrequently, or only if there were a significant environmental disturbance.

CONCLUSION

Virtually all military communications, navigation, and IFF systems require a time and frequency subsystem. The MIFTTI program is studying the concept of adding intelligence to such subsystems as a means to improve their performance. The results are encouraging, and show the potential for significant improvements in the capabilities of military time and frequency subsystems by the incorporation of these concepts.

ACKNOWLEDGEMENT

The work on the MIFTTI program at EG&G is sponsored by the U.S. Army Electronics Technology and Devices Laboratory (ETDL) under contract No. DAAL01-87-C-0715. The authors wish to acknowledge their stimulating discussions with Messrs. R. Filler and J. Vig of U.S. Army LABCOM concerning the MIFTTI operating concepts.

REFERENCES

1. T.J Lynch, K.D. Lyon, W.J. Riley and J.R. Vaccaro, "Modular Intelligent Frequency, Time and Time Interval (MIFTTI) Subsystem Study," U.S. Army Laboratory Command Research and Development Technical Report SLCET-TR-87-0715-1, Jan. 1989.

2. D. Brown et al, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillator," 38th Annual Frequency Control Symposium, June 1984, pp.380-386.

3. W.J. Riley and J.R. Vaccaro, "Rubidium-Crystal Oscillator (RbXO) Development Program," U.S. Army Laboratory Command Research and Development Technical Report SLCET-TR-84-0410-F, Apr. 1986.

4. W.J. Riley and J.R. Vaccaro, "A Rubidium-Crystal Oscillator (RbXO)," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., Vol. UFFC-34, pp. 612-618, Nov. 1987.

5. S. Stein, "Modular Intelligent Frequency, Time and Time Interval Subsystem Study Program," U.S. Army Laboratory Command Research and Development Technical Report SLCET-TR-87-0717-F, Jan. 1989.

6. R. Filler and S. Stein, "Kalman Filter Analysis for Real Time Applications of Clocks and Oscillator," 42nd Annual Frequency Control Symposium, June 1988, pp.447-452.

KALMAN FILTER ANALYSIS OF PRECISION CLOCKS WITH REAL-TIME PARAMETER ESTIMATION

S.R. Stein*
Ball Communication Systems Division
P.O. Box 1235
Broomfield, CO 80020

Introduction

This paper extends the Kalman analysis of precision clocks presented at the 42nd Annual Symposium on Frequency Control. [1] It addresses the real-time estimation of the noise parameters of the clocks, *i.e.*, the spectral densities of the noises. The work was motivated by the need for updated parameter values for adaptive Kalman filter analysis and clock forecasting and control.

Clock parameters are traditionally calculated by an off-line batch processing technique such as variance analysis, spectrum analysis or maximum likelihood analysis. We have developed a method of analyzing the Kalman filter residuals which provides updated parameter values in near real-time, eliminating the need to interrupt the recursive Kalman computations with periodic parameter analysis. The procedure is similar to estimating the clock noise parameters from the two-sample variance, except traditional variance analysis requires equally spaced data.

Kalman Filter Model for a Precision Oscillator

The states of the clock are formed into an n dimensional state vector $\vec{\mathbf{x}}(t)$. The system evolves from time t to time $t+\delta$ according to

$$\vec{\mathbf{x}}(t+\delta) = \mathbf{\Phi}(\delta)\vec{\mathbf{x}}(t) + \mathbf{\Gamma}\vec{\mathbf{s}}(t+\delta|t) + \mathbf{\Phi}(\delta)\vec{\mathbf{p}}(t), \qquad (1)$$

where the $n \times n$ dimensional state transition matrix $\mathbf{\Phi}(\delta)$ embodies the system model, the n dimensional vector $\vec{\mathbf{s}}(t+\delta|t)$ contains the noise inputs to the system during the interval from t to $t+\delta$, and the n dimensional vector $\vec{\mathbf{p}}(t)$ contains the control inputs made at time t. The state transition matrix is assumed to depend on the length of the interval, but not on the origin. Each element of $\vec{\mathbf{s}}(t+\delta|t)$ is normally distributed with zero mean and is uncorrelated in time. Equation 1 generates a random walk in the elements of the state vector. The observation vector $\vec{\mathbf{z}}(t)$ is described by the measurement equation

$$\vec{\mathbf{z}}(t) = \mathbf{H}(t)\vec{\mathbf{x}}(t) + \vec{\mathbf{v}}(t). \qquad (2)$$

The r observations made at time t are related linearly to the n elements of the state vector by the $r \times n$ dimensional measurement matrix $\mathbf{H}(t)$ and the r dimensional white noise vector $\vec{\mathbf{v}}(t)$. Kalman and Bucy defined a recursive procedure for estimating the next state, which requires the mean squared error of the estimates from the true state to be minimum.

The error in the estimate of the state vector after the measurement at time t is $\hat{\vec{\mathbf{x}}}(t|t) - \vec{\mathbf{x}}(t)$. The error covariance matrix is defined to be

$$\mathbf{P}(t|t) = E\left\{\left[\hat{\vec{\mathbf{x}}}(t|t) - \vec{\mathbf{x}}(t)\right]\left[\hat{\vec{\mathbf{x}}}(t|t) - \vec{\mathbf{x}}(t)\right]^T\right\}. \qquad (3)$$

The diagonal elements of this $n \times n$ matrix are the variances of the estimates of the components of $\vec{\mathbf{x}}(t)$ after the measurement at time t. Next, the error covariance matrix just prior to the measurement at time $t+\delta$ is defined as

$$\mathbf{P}(t+\delta|t) = E\left\{\left[\hat{\vec{\mathbf{x}}}(t+\delta|t) - \vec{\mathbf{x}}(t+\delta)\right]\left[\hat{\vec{\mathbf{x}}}(t+\delta|t) - \vec{\mathbf{x}}(t+\delta)\right]^T\right\}. \qquad (4)$$

Finally, $\mathbf{R}(t)$ is the covariance matrix of the measurement noise and $\mathbf{Q}(t+\delta|t)$ is the covariance matrix of the system noise generated during the interval from t to $t+\delta$

$$\mathbf{R}(t) = E\left[\vec{\mathbf{v}}(t)\vec{\mathbf{v}}(t)^T\right] \qquad (5)$$

$$\mathbf{Q}(t+\delta|t) = E\left[\vec{\mathbf{s}}(t+\delta|t)\vec{\mathbf{s}}(t+\delta|t)^T\right]. \qquad (6)$$

The evolution of the error covariance matrix is determined by the system model

$$\mathbf{P}(t+\delta|t) = \mathbf{\Phi}(\delta)\mathbf{P}(t|t)\mathbf{\Phi}(\delta)^T + \mathbf{\Gamma}\mathbf{Q}(t+\delta|t)\mathbf{\Gamma}^T. \qquad (7)$$

The new estimate of the state vector depends on the previous estimate and the current measurement

$$\begin{aligned}\hat{\vec{\mathbf{x}}}(t+\delta|t+\delta) \;=\; & \mathbf{\Phi}(\delta)\hat{\vec{\mathbf{x}}}(t|t) + \mathbf{\Phi}(\delta)\vec{\mathbf{p}}(t) \\ & + \mathbf{K}(t+\delta)\left[\vec{\mathbf{z}}(t+\delta) - \mathbf{H}(t+\delta)\mathbf{\Phi}(\delta)\hat{\vec{\mathbf{x}}}(t|t)\right] \\ & - \mathbf{K}(t+\delta)\mathbf{H}(t+\delta)\mathbf{\Phi}(\delta)\vec{\mathbf{p}}(t), \end{aligned} \qquad (8)$$

where the gain matrix, $\mathbf{K}(t+\delta)$, determines how heavily the new measurements are weighted. The optimum or Kalman gain, $\mathbf{K}_{opt}$, is determined by minimizing the "square of the length of the error vector," *i.e.*, the sum of the diagonal elements (the trace) of the error covariance matrix.

$$\begin{aligned}\mathbf{K}_{opt}(t+\delta) = \mathbf{P}(t+\delta|t)\mathbf{H}(t+\delta)^T \times \\ \left[\mathbf{H}(t+\delta)\mathbf{P}(t+\delta|t)\mathbf{H}(t+\delta)^T + \mathbf{R}(t+\delta)\right]^{-1}\end{aligned} \qquad (9)$$

Finally, the updated error covariance matrix is

$$\begin{aligned}&\mathbf{P}(t+\delta|t+\delta) = \\ &\left[\mathbf{I} - \mathbf{K}(t+\delta)\mathbf{H}(t+\delta)\right]\mathbf{P}(t+\delta|t)\left[\mathbf{I} - \mathbf{K}(t+\delta)\mathbf{H}(t+\delta)\right]^T \\ &+ \mathbf{K}(t+\delta)\mathbf{R}(t+\delta)\mathbf{K}(t+\delta)^T\end{aligned} \qquad (10)$$

where $\mathbf{I}$ is the identity matrix. If the optimum filter gain is used, Eq. 10 reduces to a simpler form

$$\mathbf{P}(t+\delta|t+\delta) = \left[\mathbf{I} - \mathbf{K}_{opt}(t+\delta)\mathbf{H}(t+\delta)\right]\mathbf{P}(t+\delta|t). \qquad (11)$$

*Work supported by U.S. Army Laboratory Command, Contract No. DAAL01-87-C-0717

CH2690-6/89/0000-232 $1.00 C 1989 IEEE

Equations 7 through 11 define the Kalman filter, and so defined it is an optimal estimator in the minimum squared error sense. Each application of the Kalman recursion yields an estimate of the state of the system which is a function of the elapsed time since the last filter update. Updates may occur at any time. In the absence of observations, the updates are called forecasts. Simultaneous observations may be processed in parallel or in series, in any order, with no change in the final result. The interval between updates, δ is arbitrary and is specifically not assumed to be constant. The state vector is

$$\vec{\mathbf{x}}(t) = \begin{bmatrix} u(t) \\ x(t) \\ y(t) \\ w(t) \end{bmatrix} \tag{12}$$

where $u(t)$ is the output phase, $x(t)$ is a dummy variable, $y(t)$ is the frequency and $w(t)$ is the frequency aging. The corresponding noise vector is

$$\vec{\mathbf{s}}(t+\delta|t) = \begin{bmatrix} \beta'(t+\delta|t) \\ \epsilon'(t+\delta|t) \\ \eta'(t+\delta|t) \\ \alpha'(t+\delta|t) \end{bmatrix}. \tag{13}$$

The state transition and control matrices are

$$\mathbf{\Phi}(\delta) = \begin{bmatrix} 0 & 1 & \delta & \delta^2/2 \\ 0 & 1 & \delta & \delta^2/2 \\ 0 & 0 & 1 & \delta \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{14}$$

and

$$\mathbf{\Gamma}(\delta) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{15}$$

The system covariance matrix is calculated from the definition in Eq. 6. The non-zero elements are

$$\begin{aligned}
\mathbf{Q}(t+\delta|t)_{00} &= S_{\beta'}(t) f_h \\
\mathbf{Q}(t+\delta|t)_{11} &= S_\xi(t)\delta + S_\mu(t)\delta^3/3 + S_\zeta(t)\delta^5/20 \\
\mathbf{Q}(t+\delta|t)_{22} &= S_\mu(t)\delta + S_\zeta(t)\delta^3/3 \\
\mathbf{Q}(t+\delta|t)_{33} &= S_\zeta(t)\delta \\
\mathbf{Q}(t+\delta|t)_{12} &= S_\mu(t)\delta^2/2 + S_\zeta(t)\delta^4/8 \\
\mathbf{Q}(t+\delta|t)_{13} &= S_\zeta(t)\delta^3/6 \\
\mathbf{Q}(t+\delta|t)_{23} &= S_\zeta(t)\delta^2/2 \\
\mathbf{Q}(t+\delta|t)_{ij} &= \mathbf{Q}(t+\delta|t)_{ji}
\end{aligned}$$

where f_h is a rectangular high frequency cutoff. Without this bandwidth limitation, the variance of the white phase additive noise would be infinite. The spectral densities which appear in the system covariance matrix may be written in terms of the standard oscillator noise coefficients. The values and units are shown in Table 1.

The particular case treated here is specified by a series of time observations described by

$$\mathbf{H}(t) = (\ 1\ \ 0\ \ 0\ \ 0\ ) \text{ and } \mathbf{R} = \sigma_{v_\epsilon}^2 \tag{16}$$

where $\sigma_{v_\epsilon}^2$ is the variance of the phase measurement process.

Adaptive Filtering

Estimating the parameters required by the Kalman filter is usually accomplished external to the Kalman filter computation by an independent method. The techniques that are normally applied are Allan variance analysis and maximum likelihood analysis. The problem with using the Allan variance is that it is defined for equally spaced data. In an operational scenario where there are occasional missing data, the gaps may be bridged, but when the data are irregularly spaced, a more powerful approach is required.

Table 1: Relationship between spectral densities and h coefficients

Spectral density	h coefficient	units
$S_{\beta'}(t)$	$(2\pi)^{-2}h_2(t)$	seconds^3
$S_\xi(t)$	$h_0(t)$	seconds
$S_\mu(t)$	$(2\pi)^2 h_{-2}(t)$	seconds^{-1}
$S_\zeta(t)$	$(2\pi)^4 h_{-4}(t)$	seconds^{-3}

The maximum likelihood approach determines the parameter set most likely to have resulted in the observations. Equally spaced observations are not required, but the data are batch processed. Furthermore, each step of the search for the maximum requires a recomputation of the Kalman filter for all the observations, resulting in an extremely time-consuming procedure. Both the memory needs and computation time are incompatible with the use of this approach in real-time or embedded applications.

A variance analysis technique compatible with irregular observations has been developed. The variance of the innovation sequence of the Kalman filter is analyzed to provide estimates of the parameters of the filter. Like the Allan variance analysis, which is performed on the unprocessed measurements, the innovation analysis requires only a limited memory of past data. However, the forecasts provided by the Kalman filter allow the computation to be performed at arbitrary intervals once the algebraic form of the innovation variance has been calculated. Also, the Allan variance is not defined for noise processes more divergent than rw frequency, but the variance of the innovations is well behaved because the state equations account for the appropriate degree of non-stationarity.

Theory

The innovation sequence has been used to provide real-time parameter estimates for Kalman filters with equal sampling intervals. Gelb [2] describes the case of continuous and discrete filters. Mehra [3] developed an approach for discrete Kalman filters with equal sampling intervals, and derived the conditions for estimating all the parameters of the filter:

1. The system must be observable.
2. The system must be time invariant.
3. The number of unknown parameters in $\mathbf{Q}$ must be less than the product of the dimension of the state vector and the dimension of the measurement vector.
4. The filter must be in steady state.

Mehra's technique cannot be used in this case because of the assumed irregular sampling, which prevents the system from ever reaching steady state. However, it is possible to proceed in a similar fashion by calculating the variance of the innovations in terms of the true values of the parameters and the approximate gain and actual covariance of the suboptimal Kalman filter that produced the innovation sequence.

Let $\tilde{\mathbf{x}}(t_-)$ be the the error in the estimate of the state vector before the measurement at time t, and let $\tilde{\mathbf{x}}(t_+)$ be the error after the observation. One can show that, if the estimate is to be unbiased, one must have an estimati equation in the form

$$\tilde{\mathbf{x}}(t_+) = [\mathbf{I} - \mathbf{K}(t)\mathbf{H}(t)]\tilde{\mathbf{x}}(t_-) + \hat{\mathbf{K}}(t)\mathbf{v}(t), \quad (17)$$

where $\hat{\mathbf{K}}(t)$ is the (suboptimal) filter gain obtained from using estimated noise parameters. The innovation vector is the difference between the observation and the prediction

$$\vec{\nu}(t) \equiv \vec{\mathbf{z}}(t) - \mathbf{H}(t)\hat{\tilde{\mathbf{x}}}(t+\tau|t). \quad (18)$$

Substituting Eqs. 17 and the measurement model, yields

$$\vec{\nu}(t) = \vec{\mathbf{v}}(t) - \mathbf{H}(t)\tilde{\mathbf{x}}(t_-). \quad (19)$$

The variance of the innovations may now be calculated. [2]

$$E\left[\vec{\nu}(t+\delta)\vec{\nu}(t+\delta)^T\right] = \mathbf{H}(t+\delta)\mathbf{P}(t+\delta|t)\mathbf{H}(t+\delta)^T + \mathbf{R}(t+\delta) \quad (20)$$

Estimation Procedure

The adaptive modeling procedure is initialized with approximate parameters for the clock from which the first approximate Kalman gain, $\hat{\mathbf{K}}$, is computed. The initial approximations may be obtained from a calibration and two-sample variance evaluation, from the manufacturer's data sheet, or from experience. As the state estimates are computed, the variance of the innovations on the left side of Eq. 20 is estimated. The right-hand side of this equation is written in terms of the actual filter element values (covariance matrix elements) and the theoretical parameters.

$$\begin{aligned} E\left[\vec{\nu}(t+\delta)\vec{\nu}(t+\delta)^T\right] &= \mathbf{P}_{11}(t|t) + 2\delta\mathbf{P}_{12}(t|t) + \delta^2\mathbf{P}_{13}(t|t) \\ &\quad + \delta^3\mathbf{P}_{23}(t|t) + \delta^2\mathbf{P}_{22}(t|t) \\ &\quad + \frac{\delta^4}{4}\mathbf{P}_{33}(t|t) + S_\xi(t)\delta + S_\mu(t)\frac{\delta^3}{3} \\ &\quad + S_\zeta(t)\frac{\delta^5}{20} + \sigma^2_{v_x}(t) \end{aligned} \quad (21)$$

Finally, the equations are inverted to produce improved estimates for the parameters.

It is assumed that the oscillator model contains no hidden noise processes. This means that each noise process included in the model is dominant over some region of Fourier frequency space. The principle of parsimony supports this approach to modeling. If a noise process is not observable with signal-to-noise ratio larger than one, then its inclusion in the model deteriorates the ability to forecast future performance. It is assumed that there are $k \leq 4$ visible noise processes. If the observations were uniformly spaced, it would be possible to calculate the variance of the innovations for k suitable sample times and solve the resulting k simultaneous linear equations. This procedure is not possible when the observations occur at unequal time intervals. In this case it is still possible to use a bootstrap procedure to evaluate the parameters.

For each of the k parameters to be estimated, a Kalman filter is computed using a subset of the observations chosen to maximize the number of predictions in the interval for which that parameter makes the dominant contribution to the innovations. Each innovation is used to compute a single-point (primitive) estimate of the variance of the innovations for the corresponding δ. Substituting the estimated values of the remaining parameters, Eq. 21 is solved for the dominant parameter, and the estimate of that parameter is updated in an exponential filter of the appropriate length. If the minimum sampling interval is too long, it may not be possible to estimate one or more of the parameters. However, there is no deleterious consequence of this situation, since a parameter that cannot be estimated is not contributing appreciably to the prediction errors.

This method of adaptive modeling works over a broader range of situations than Mehra's method. The only criteria are that:

1. The system be observable.
2. The number of unknown parameters in $\mathbf{Q}$ be less than the product of the dimension of the state vector and the dimension of the measurement vector.

Verification

Simulation testing has shown that the this method of real-time parameter estimation combines good data efficiency and high accuracy. First, a simple case was selected for simulation analysis. A clock with white frequency noise and random walk frequency noise was considered and a simulator was used to produce data for analysis. A program was written to compute two Kalman filters, one running as often as possible and the other running at twice the sample interval corresponding to the minimum in the $\sigma_y(\tau)$ curve. The parameter estimation is 'real-time' to the extent that no observations need to be used that are earlier than two days prior to the time the estimate is made. The performance of the algorithm appears superior to two other methods of real-time parameter estimation: calculation of the parameters directly from the estimated autocorrelation function of the second difference of the observations and the real-time ARIMA technique developed Barnes. [4] Both comparison methods require uniformly sampled data. It was found that the direct computation of the parameters from the autocovariance function resulted in higher rms deviations from the true values than the use of feedback. Thus, the feedback method was selected for comparison with the Kalman technique. The ARIMA technique was found to produce biased estimates when the parameter θ was near 0 or 1. Figures 1 through 4 compare the true parameter values to the real-time ARIMA and Kalman estimates. This simulation illustrates the bias of the ARIMA estimate of S_μ, but shows no comparable bias in the Kalman estimate.

Additional verification of the technique has been performed using real clock data with unequal sampling. Allan variance analysis was used to determine initial, approximate parameter values. Subsequently, improved parameter values were calculated for the spectral densities of the white phase noise and random walk frequency noise of the clock. The random walk frequency aging parameter, a level so that the frequency aging rate could change over a period of approximately one year. There was not enough data to improve this estimate. If the Kalman filter were run without any random walk frequency aging, the estimate of the frequency aging rate would converge on a fixed and incorrect value and the resulting time predictions would be grossly in error. Instead, the estimated frequency aging varies slowly during the two years of available data. The estimates of the white pm and random walk fm are shown in Figs. 5 and 6. The white pm level increases dramatically at approximately MJD 45850 and subsequently returns to its original level. The filter tracks with a time constant of approximately 20 days. The change is not an artifact of the

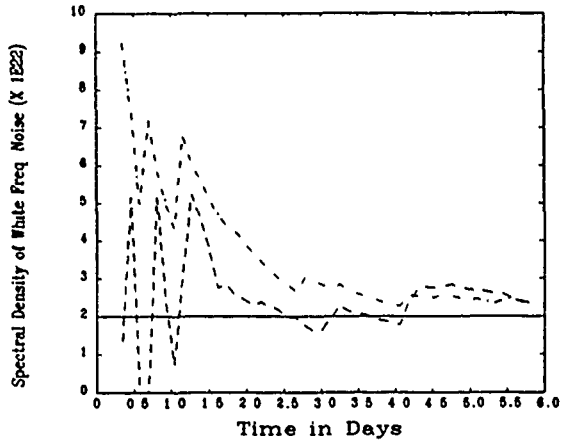


Figure 1: Real-time estimate of the white frequency noise of a simulated clock using ARIMA (dashed line) and Kalman (dotted-dashed line) methods. The true value is 2×10^{-22}.

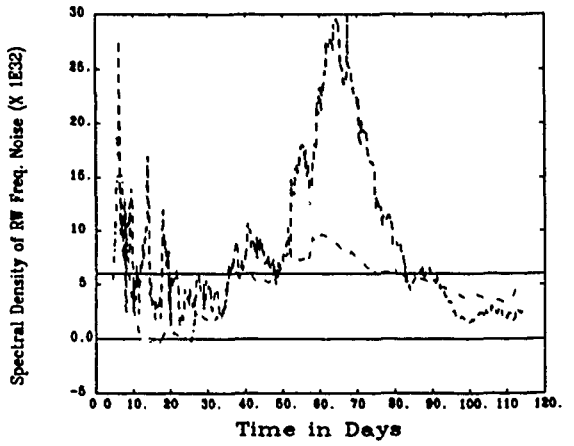


Figure 4: Real-time estimate of the rw frequency noise of a simulated clock using ARIMA (dashed line) and Kalman (dotted-dashed line) methods. The true value is 6×10^{-32}.

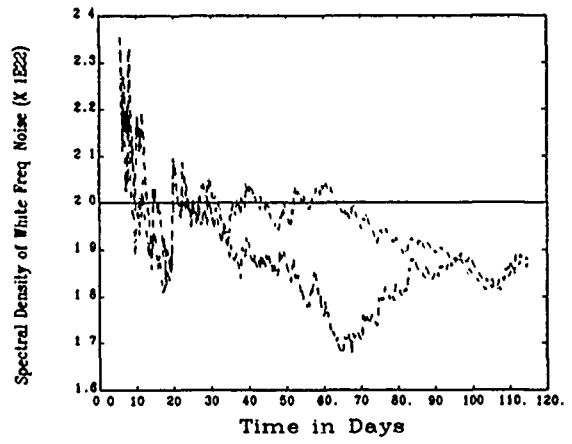


Figure 2: Real-time estimate of the white frequency noise of a simulated clock using ARIMA (dashed line) and Kalman (dotted-dashed line) methods. The true value is 2×10^{-22}.

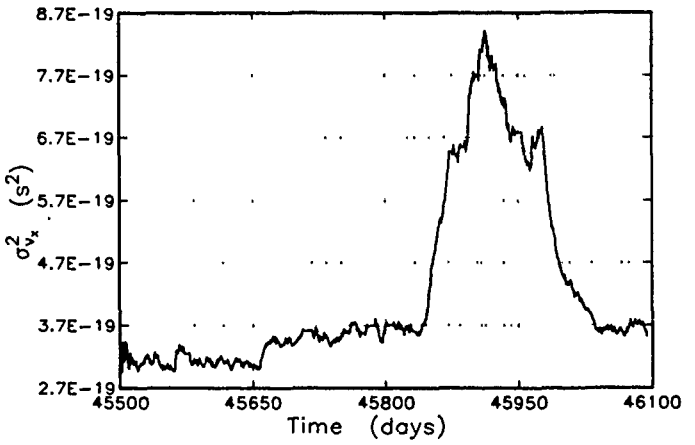


Figure 5: Variance of the white phase noise of clock 7.

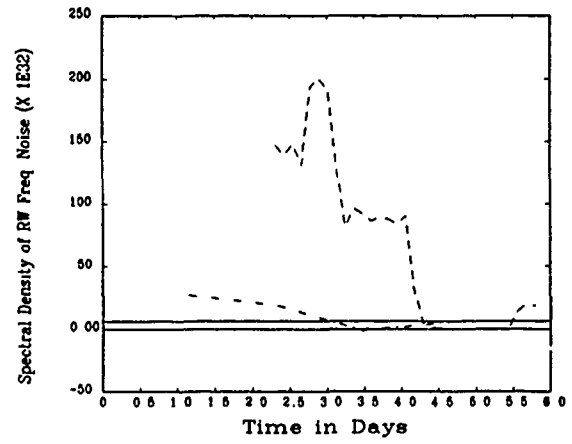


Figure 3: Real-time estimate of the rw frequency noise of a simulated clock using ARIMA (dashed line) and Kalman (dotted-dashed line) methods. The true value is 6×10^{-32}.

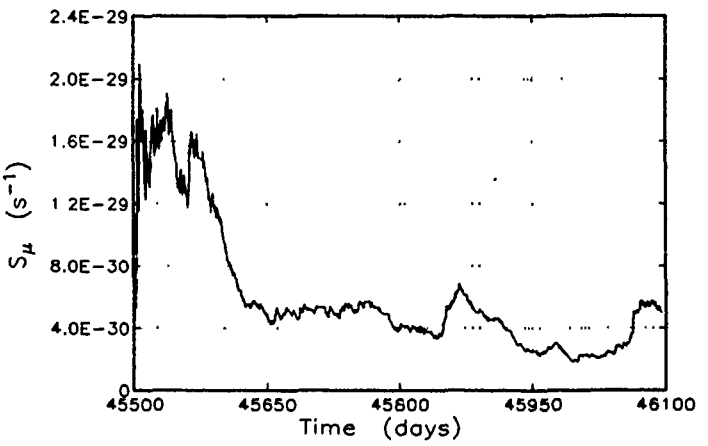


Figure 6: Spectral density of the random walk frequency noise of clock 7.

parameter estimation process. It has been verified using Allan variance analysis and by examining the unfiltered observations. The random walk fm level decreases over the two years of data acquisition. This change is also visibile when one examines the frequency estimates of the clock. Most likely, the random walk fm was elevated for several months after the clock was first turned on and subsequently improved. The software tracks the change automatically and is not significantly affected by the dramatic changes in the white pm.

Conclusions

An "optimum" algorithm has been developed for estimating the time, frequency and frequency aging of clocks and oscillators while simultaneously estimating the noise spectrum. The algorithm is sufficiently general to be used with all types of quartz oscillators and atomic clocks yet the approach is appropriate for use with an embedded controller in automated systems intended for field applications. The technique is adaptive – it calculates and updates estimates for all required clock parameters as observations occur. Data are accepted whenever available and are not required at fixed multiples of a fixed sampling time. Only coarse estimates of the parameters are required to initialize the computation.

References

[1] R. L. Filler and S. R Stein, Kalman Fileter Analysis for Real Time Applications of Clocks and Oscillators, Proc. 42nd Annual Symposium of Frequency Control, 1988, pp. 447-452.

[2] Gelb, A. ed. (1974). "Applied Optimal Estimation." M.I.T. Press,Cambridge

[3] On the Identification of Variances and Adaptive Kalman Filtering, R. K. Mehra, IEEE Trans. Automatic Control, pp. 175-184, April 1970.

[4] J. A. Barnes, An Adaptive Algorithm to Evaluate Clock Performance in Real Time, Proc. of the 20th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 205-218.

43rd Annual Symposium on Frequency Control - 1989

SPACECRAFT GRAVITATIONAL WAVE EXPERIMENTS

J. W. Armstrong
Jet Propulsion Laboratory, MS 238-737
California Institute of Technology,
4800 Oak Grove Dr., Pasadena, CA 91109 USA

Abstract---Doppler tracking of distant spacecraft can be used in broadband (frequencies ~10^{-4}-10^{-1} Hz) searches for low-frequency gravitational radiation. In this experiment, the Earth and spacecraft act as free test masses and the Doppler tracking system continuously measures the relative dimensionless velocity $\Delta v/c = \Delta f/f_0$ as a function of time (here Δf is the Doppler perturbation in a link having center frequency f_0). Gravitational radiation--propagating, polarized gravitational "ripples" produced by nonspherical motions of massive astrophysical objects--is characterized by a dimensionless strain amplitude, $h = \Delta \ell / \ell_0$. A gravitational wave of amplitude h incident on this system causes small perturbations (of order h in the relative Doppler $\Delta f/f_0$) which have a characteristic "three pulse" signature in the Doppler time series. The spacecraft tracking technique is essentially a "one-armed" interferometer; since clock instability is not canceled, precision frequency control of the Doppler system is essential for good sensitivity. In this paper, I give a short review of ideas about astrophysical sources, radiation time scales, and associated gravitational wave amplitudes at the earth. The current and near-future space-based detection approaches and sensitivities are then summarized, along with a discussion of the major noise sources in space-based experiments. Finally, possible improvements in space-based detectors, including approaches that depart significantly from the current prototype systems, are briefly discussed.

I. Sources/Strengths/Timescales

This introductory section gives a theoretical perspective on sources, strengths, and timescales of astrophysical gravitational waves. It is based on the recent exhaustive review by Thorne [1].

Gravitational waves (GW's) are propagating, polarized gravitational fields. They are predicted by all relativistic theories of gravity (with different properties). They manifest themselves by changing the distance between separated test masses and shifting the rates at which separated clocks keep time. The strength of these waves is characterized by the wave's dimensionless strain amplitude, $h = \Delta \ell / \ell$, where $\Delta \ell$ is the change in separation between masses having nominal separation ℓ.

In General Relativity, GW's are propagating solutions of the Einstein field equations, thus tensor strain waves. Like electromagnetic waves, GW's are transverse, have two independent polarization states and propagate at the speed of light. Unlike electromagnetic waves, GW's are extremely weak. This extreme weakness has two consequences. First, GW's are generated at appreciable levels only by *very* massive objects undergoing *very* violent dynamics, i.e. astrophysical sources. Second, in their subsequent propagations GW's interact very weakly with matter, undergoing negligible scattering and absorption.

Why are GW's interesting? There are two reasons: as tests of fundamental physical law (relativistic gravity) and because of their potential for opening a new window for observational astronomy. Thorne [1] thoroughly discusses both cases. I mention here only the potential for gravitational wave astronomy. Since GW's propagate essentially unchanged from their sources, they preserve detailed information about those sources that is unavailable by any other means. (Electromagnetic waves are easily scattered. Even supernova neutrinos are scattered many times before leaving the cores of those explosions. In both cases, information is retained about the last scattering surface, not the interior.) Relativistic bulk motion and strong gravitational fields are central to theoretical views of violent activity in a number of astrophysical objects. GW's from these objects will give the first observations of the *interiors* of those regions. Additionally, Schutz [2] has emphasized that one class of GW source--coalescing binary systems--has the remarkable property that the mathematical combination of (GW amplitude)/(instantaneous GW frequency)/(instantaneous GW frequency derivative) is proportional to 1/(distance to source), *independent of the masses of the objects in the binary system.* This offers a method of accurately measuring distances to very remote objects--a thing which is very difficult to do in astronomy.

From an observational viewpoint, the problem is that the GW's are weak and must be observed in the presence of noise. The waves are far too weak to be generated at detectable levels in the laboratory. For (decently non-spherical) astrophysical systems, however, the situation is more promising. GW's are generated to lowest order by the second time derivative of the "transverse, traceless" mass quadrupole moment of the system (Misner, Thorne, and Wheeler [3]). In order of magnitude (e.g., Tyson and Giffard [4]):

$$h \approx [G M_{eff}/ r c^2] (v/c)^2$$
$$\approx 10^{-17} (M_{eff}/M_O) (v/c)^2 (10 \text{ kpc}/r)$$

where G is the Newtonian gravitational constant, M_{eff} is the part of the system's mass distribution involved in

CH2690-6/89/0000 $1.00 C 1989 IEEE

the time-varying quadrupole moment, M_Q is the solar mass, and v is a characteristic velocity of the source. The scaling of distance to the source, r, in units of 10kpc is motivated because 10kpc is about the distance to the center of the galaxy (thus the nearest of the "reasonable" sources). As will be seen below, current and near-future space-based experiments can only hope to detect very massive systems.

By convention, GW's are classified by the temporal behavior of their waveforms. Bursts are waves that are "on" only for a well-defined time interval. Burst waves might be generated, for example, by the collision of two black holes. Periodic waves, the prototype of which is sinusoidal radiation, are waves having power concentrated in a few harmonics that are sensibly stable over an observing interval. Periodic waves might be generated by black holes in a binary system. Finally, stochastic waves might be generated by density inhomogeneities in the Big Bang itself. Thorne [1] has given an exhaustive review of the possible strengths of GW's at the earth from sources as far away as the Hubble distance (≈ 3 Mpc). In the "low-frequency" band to which spacecraft experiments are sensitive, the estimated strengths of these waves are typically $\sim 10^{-15}$ or less.

II. Detection Methods

Experimental approaches to GW detection vary depending on the time-scale of the radiation: for the theory of the detectors and experimental sensitivities as of 1987, see Thorne [1]. Briefly, at relatively high-frequencies (~100-10,000 Hz) ground based approaches include resonant bars and laser interferometers. Each of these approaches achieves excellent sensitivity. Below about 100 Hz, however, it becomes prohibitively difficult to isolate these detectors from seismic and other acoustic noises. Thus for lower-frequency observations, one or all of the test masses must be moved into the relatively benign space environment. The current prototype space-based detectors us the earth and a distant spacecraft at separated test masses, with a Doppler link driven by a high-quality frequency standard continuously monitoring the relative dimensionless velocity ($\Delta v/c$) between the the two masses. A gravitational wave of strain amplitude h incident on the system causes small perturbations in the tracking record. These perturbations are of order h in $\Delta f/f_0$ (the fractional change in the Doppler frequency of the link) and are replicated three times in the Doppler data (Estabrook and Wahlquist [5]). That is, the "signal part" of the observed Doppler time series is the convolution of the gravitational waveform with the function: $[(\mu-1)/2]\,\delta(t) - \mu\,\delta[t-(1+\mu)L/c) + [(1+\mu)/2]\,\delta(t - 2L/c)$, where μ is the cosine of the angle between the earth-spacecraft vector and the gravity wavevector and L is the earth-spacecraft distance. Two of these three "pulses" are separated by the round-trip light time between the spacecraft and the earth while the third occurs at an intermediate time that depends on the wave's angle of arrival. The sum of the Doppler perturbations of the three pulses is zero. Pulses with duration longer than about the one-way light time produce overlapping responses in the tracking record and the net response then cancels to first order. The tracking system thus has a passband to gravitational excitation: the low-frequency band edge is set by pulse cancellation to be about (1/one-way light time), while thermal and other noises limits the high frequency response to about (1/30 seconds).

II.1 Noise Sources in Low-Frequency Experiments

The main noise sources in a spacecraft gravitational wave experiment are briefly summarized in this section.

At frequencies higher than about 0.033 Hz, thermal noise arising from finite signal-to-noise ratio in the Doppler tracking link dominates. This (white phase) noise corresponds to a power spectrum of fractional frequency $\Delta f/f_0$ going as (Fourier frequency)^2, which dominates at high frequencies. Its level can be reduced by increasing the signal power in the radio link or decreasing the system noise temperature.

At lower frequencies, propagation noise and instrumental instability are important Propagation noise is caused by phase variations imposed on the radio wave as it traverses irregularities in the transmission media between the earth and spacecraft (troposphere, ionosphere, solar wind). These phase variations ("phase scintillation") map to frequency variations in the Doppler time series. Charged particle scintillations (ionosphere and solar wind) are the dominant noise source for the current generation (S-band radio link--2.3 GHz) experiments (Wahlquist et al. [6]; Armstrong, Woo, and Estabrook [7]; Woo and Armstrong [8]). The level of the plasma scintillation varies with sun-earth-spacecraft angle, reaching a broad minimum in the antisolar direction. Expressed as an Allan variance of the fractional frequency deviation at an integration time of 1000 seconds, this component reaches a minimum value of about (3×10^{-15}) (8.4 GHz/radio frequency of link)^2. The plasma scintillation data have a "red" spectrum varying approximately as (frequency)$^{-1}$.

In addition to propagation noise, there is instrumental instability. Here there are a variety of contributors: clock (timekeeping) noise, instability introduced by signal distribution within the ground station, transmitter and receiver instability, mechanical stability of the antenna (e.g. dish sag changing the phase center of the antenna over time as the elevation angle changes), stability of the spacecraft's transponder, etc. Considerable engineering effort within NASA's Deep Space Network (DSN) is being directed toward improving frequency standards and generally minimizing other noise contributions in the ground system. The expected instrumental stability for the Galileo gravity wave experiment (X-band transponder, tracking with NASA's new 34-meter high-efficiency DSN antennas, 1993+) is a few parts in 10^{15} for integration times of 1000 seconds.

Non-gravitational forces (or effects that enter like non-gravitational forces) also act as noise sources. Things like spacecraft buffeting (e.g., leaking thrusters), irregularities in the spacecraft spin rate (for Doppler measurements using circularly polarized signals), microseismic disturbances, etc. can influence sensitivity. It is expected that in the Galileo era the most important of these can be calibrated and removed with engineering telemetry to a level that is less than the propagation and instrumental noise levels.

Finally, systematic errors are always present at some level. These are difficult to quantify and usually influence the sensitivity to different degrees depending on the spectral characteristics of the gravitational wave being sought.

These noises enter through "transfer functions" which connect the noise excitation to the Doppler observable. The transfer function depends on the type of noise. In general the Doppler system responds differently to noise than it does to gravitational radiation; see Estabrook and Wahlquist [5], Wahlquist et al. [6], or Armstrong [9] for a discussion.

II.2 Example: Pioneer 11 Periodic Wave Search

This section gives an example of the sensitivity of current generation detectors to one class of signal: sinusoidal radiation. A more detailed discussion of the data shown here has been given by Armstrong, Estabrook, and Wahlquist [10]. For a discussion of burst and background sources, see e.g. [1], [9], [11], [13], and [14].

The data were taken in 1983 using the Pioneer 11 spacecraft, which was at that time 12.9 AU from the earth at a sun-earth-spacecraft angle of about 160 degrees. Hydrogen masers provided the frequency standard at each of the three 64-meter antennas (California, Spain, Australia) used in this experiment. About 38.8 hours of data having elevation angles of the up- and downlinks above 20 degrees were taken over a three day period. The raw data--transponded electric field--were recorded at the ground stations on magnetic tape and then shipped to JPL. The phase of the received signal was later estimated in the computer using a digital phase lock loop. The time series of fractional frequency fluctuation was computed and corrected for the known orbit of the spacecraft. The resulting time series was searched for sinusoidal signals in the low-frequency band, but with otherwise unknown frequency.

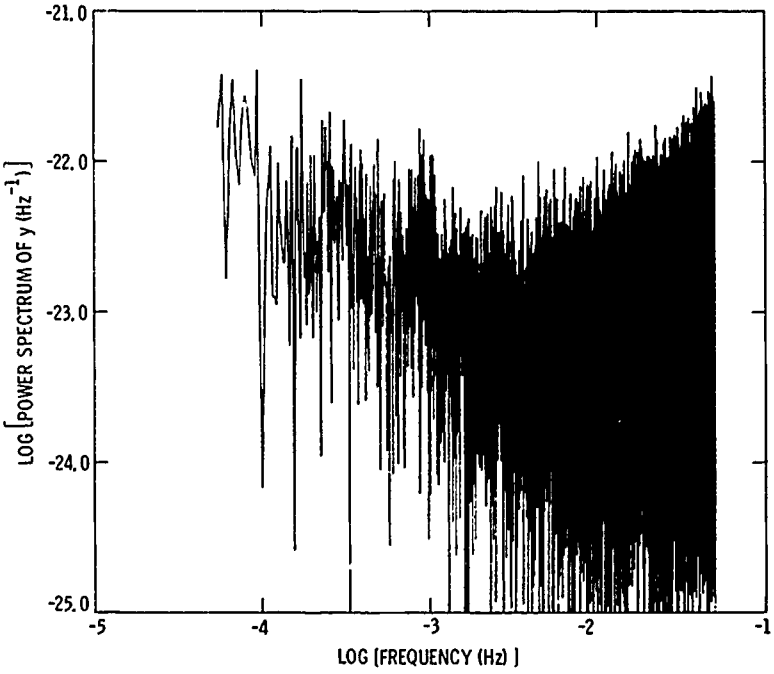


Figure 1. Power spectrum of fractional frequency fluctuations, $S_y(f)$, for the Pioneer 11 data discussed in the text. The main features of the spectrum are a "red" component, dominated by plasma scintillation noise, varying approximately as (frequency)$^{-1}$, and a "blue" component dominated by finite signal-to-noise ratio on the downlink. See text for discussion.

Since the phase of the signal is unknown, it was necessary to form the power spectrum, $S_y(f)$, of the Doppler time series. Figure 1 shows this spectrum for the Pioneer 1983 data formed using standard techniques. This spectrum is two-sided; that is the variance is the integral of the spectrum over both positive and negative frequencies. The main features are "flicker" noise (spectrum roughly proportional to 1/f) at low frequencies due mostly to propagation noise, and thermal noise (spectrum approximately proportional to f^2, rounded at the higher frequencies here by an anti-aliasing digital filter). For these data the interval (0.0005-0.033 Hz) has the lowest noise level; in this interval S_y is roughly constant at 1.6 (+1.6,-0.8) X 10^{-23} Hz^{-1}. Superimposed on this continuum would be any discrete spectral line associated with a gravitational wave signal. The spectrum in Figure 1 is unsmoothed (resolution bandwidth, B, is approximately 4μHz) and statistical estimation error is obvious.

Conversion of S_y to equivalent sinusoidal fractional frequency amplitude at a given fluctuation frequency is done as follows: amplitude = $2[B\ S_y(f)]^{1/2}$. This gives values typically 1.5 X 10^{-14}. However, some spectral lines have associated amplitudes larger than this and one is faced with judging statistical significance. This is determined by the estimation error statistics of the power spectrum. Armstrong, Estabrook, and Wahlquist [10] discuss this point in detail and conclude that the 90% confidence upper limit for a real signal in these data is 6 X 10^{-14}. The important question of how to relate upper limits for fractional frequency fluctuations to upper limits on strain amplitude, given the lack of knowledge of the wave polarization state, is also addressed in [10].

II.3 Experimental Limits in the LF and VLF Bands

All searches to date have yielded only upper limits for the gravitational wave amplitudes at the earth. Table 1 shows current sensitivity and projected near-future sensitivity of searches in the LF and VLF bands.

Table 1. Experimental Limits in the Low-Frequency and Very-Low-Frequency Bands

Burst Waves:

Frequency Band	Strain amplitude	Data Type	Reference
~10^{-3} Hz	< 3 X 10^{-14}	Voyager	Hellings et al. [11]
~10^{-3} Hz	< 6 X 10^{-14}	Viking	Armstrong et al. [7]
~10^{-3} Hz	< 5 X 10^{-14}	Pioneer	Armstrong [9]
10^{-1} to 10^{-4} Hz	~ 3 X 10^{-15}	Galileo	Galileo (1993+; predicted performance)

Sinusoidal Waves:

Frequency	amplitude (1 sigma)	Data Type	Reference
1.04 X 10^{-4} Hz	< 2 X 10^{-14}	Pioneer	Anderson et al. [12]
0.0005-0.033 Hz	< 1.5 X 10^{-14}	Pioneer	Armstrong et al. [10]
≈10^{-1} to 10^{-4} Hz	~ 3 X 10^{-16}	Galileo	Galileo (1993+; predicted performance)

Stochastic Background:

(amplitude expressed as energy density in gravitational waves in bandwidth equal to the center frequency, except for the normal mode data, normalized by the energy density required to close the universe)

Frequency (Hz)	energy density/ closure density	Data Type	reference
~10^{-8}	< 1.4 X 10^{-4}	pulsars	Hellings and Downs [13]
~10^{-6}	< 18	Pioneer	Anderson and Mashoon [14]
~4 X 10^{-6}	< 38	Pioneer	Anderson and Mashoon [14]
~7 X 10^{-5}	< 260	Pioneer	Anderson and Mashoon [14]
~4 X 10^{-4}	< 40	Voyager	Hellings et al. [11]
~3 X 10^{-4}	< 100	sun normal modes	Boughn and Kuhn [15]
10^{-1} to 10^{-4}	~ 0.01	Galileo	Galileo (1993+; predicted performance)

III. Future Experiments

The current generation of spacecraft experiments are limited by plasma noise for bursts, by plasma and thermal noise plus finite duration observations for sine waves, and by clock noise for stochastic backgrounds (at sensitivities $S_y \sim 10^{-23}$ Hz^{-1} for selected frequencies). In addition, all observations to date have been initiated after spacecraft launch, with no opportunity to influence spacecraft design to optimize the experiment. Opportunities for significantly-improved experiments exist in both the near and intermediate terms.

Plasma noise can be minimized with higher frequency radio links or substantially eliminated with dual frequency up- and downlinks. The next generation (Galileo-era) spacecraft gravitational wave experiments will be with X-band uplink. The NASA Deep Space Network is instrumenting a high-stability X-band uplink ground stations capability; the first of the X-band uplink observations (raw sensitivity of a few times 10^{-15} for bursts, $\sim 10^{-16}$ for periodic waves, and anticipated $S_y \sim 10^{-25}$ Hz^{-1} for backgrounds) will be with the Galileo spacecraft during cruise to Jupiter. The planetary observer series (e.g. Mars Observer), and the Mariner Mark II (e.g., CRAF) series will also have X-band uplink.

By the late 1990's, K-band (~32 GHz) uplink or multifrequency link gravitational wave observations can provide very high immunity to plasma noise and very sensitive gravitational wave experiments. (If flight-qualified precision timekeeping becomes practical, then the possibility of onboard Doppler extraction with multiple links offers improved ways to discriminate gravity wave and noise signatures.) To fully exploit this plasma noise immunity would require improved timekeeping on the ground and precision tropospheric monitoring

Finally, it is possible to achieve excellent sensitivity in the low-frequency band by putting a scaled-up version of the ground-based laser interferometer in space. Such a system would be in solar orbit, approximately 30 degrees down-orbit from the earth (to minimize perturbations from the earth-moon system). Although still in the conceptual design stage, it is thought that such a system would have sensitivity to sinusoidal waves at the $\sim 10^{-23}$ level. At this level, one is "guaranteed" detections from known galactic binary stars (in addition to the more exotic sources discused in [1]). Such a system is probably 15-20 years away, but this class of detector represents the pinnacle of sensitivity using foreseeable technology.

Acknowledgments---I thank F. B. Estabrook and H. D. Wahlquist for discussions on all aspects of the spacecraft gravitational wave search problem. I also thank J. D. Anderson, Principal Investigator of the Pioneer Celestial Mechanics Experiment, under whose auspices the Pioneer 11 data shown here were taken, the Pioneer Project Office at NASA/Ames Research Center, and the NASA/JPL Deep Space Network for their cooperation in gathering those data. This work was supported by the NASA UV/Visible/Relativity Office and was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

References

[1] K. S. Thorne in 300 Years of Gravitation, ed. S. W. Hawking and W. Israel, Cambridge: Cambridge University Press, 1987.

[2] B. F. Schutz in Gravitational Radiation, ed. D. Blair, Cambridge: Cambridge University Press, 1989.

[3] C. W. Misner, K. S. Thorne, and J. A. Wheeler, Gravitation. San Fransisco: W. H. Freeman and Company, 1973, chap. 36.

[4] J. A. Tyson and R. P. Giffard, "Gravitational Wave Astronomy", Ann. Rev. Astron. Astrophys., **16**, 521, 1978.

[5] F. B. Estabrook and H. D. Wahlquist. GRG, **6**, 439, 1975

[6] H. D. Wahlquist,J. D. Anderson, F. B. Estabrook, and K. S. Thorne, Atti dei Convengni Lincei, **34**, 335, 1977.

[7] J. W. Armstrong, R. Woo,and F. B. Estabrook, Ap. J., **230**, 570 , 1979.

[8] R. Woo, and J. W. Armstrong, JGR, **84**, 7288, 1979.

[9] J. W. Armstrong in Gravitational Wave Data Analysis, ed. B. F. Schutz,.Kluwer Academic Publishers, 1989.

[10] J. W. Armstrong,F. B. Estabrook, and H. D. Wahlquist, Ap. J.**318** ,536, 1987.

[11] R. W. Hellings,P. S. Callahan,J. D. Anderson, and A. T. Moffet, Phys. Rev. **D23**, 844, 1981.

[12] J. D. Anderson,J. W. Armstrong, F. B. Estabrook,R. W. Hellings,E. K. Lau,, and H. D. Wahlquist, Nature, **308**, 158, 1984.

[13] R. W. Hellings, and G. Downs, Ap. J. Lett. **265**, L139, 1983.

[14] J. D. Anderson, and B. Mashoon, Ap. J., **290**, 445, 1985.

[15] S. P. Boughn, and J. R. Kuhn,. Ap. J., **286**, 387, 1984.

43rd Annual Symposium on Frequency Control - 1989

CHARACTERIZATION METHODS FOR THE SENSITIVITY OF QUARTZ OSCILLATORS TO THE ENVIRONMENT

Jean-Jacques 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

Quartz crystal oscillators are sensitive to external perturbations in their environment. These sensitivities have been intensively studied for many years. But the conditions of measuring them can vary and the results are affected by the way of measuring. The present paper will show the distinction between characterizing a complete oscillator, and characterizing only the quartz resonator by using a passive phase bridge. Advantages and disadvantages of both methods will be discussed. Then measurements of temperature sensitivities, including quasistatic or dynamic thermal conditions will be presented. One important point will be how to measure the real temperature of the device under test (quartz crystal for instance) rather than the temperature of the probe. Then methods for measuring acceleration and pressure sensitivities will be also presented (spurious effects of temperature changes will be considered). Finally different problems will be discussed in connection with the measurement of the sensitivity to magnetic fields, and to electric fields. The first corresponds to a new problem and some emphasize will be given to it.

Introduction

The purpose of this paper is not to describe the sensitivities of oscillators to the different parameters of their environment, such as temperature, pressure, forces, accelerations, fields of different kinds, etc., which have been and are still intensively studied. The paper will rather discuss the methods used for measuring these sensitivities and point out the many pitfalls which can be encountered. For the user it is of upmost importance to know the magnitude of the sensitivities, but for the designer it is also necessary to understand, where the sensitivities come from, and which solutions can be found to reduce them. In the first case a measurement of the signal of the overall oscillator is made. But in the second case it is generally necessary to test separatly different parts of the oscillator in order to localize the causes of the problem.

The first difficulty in a measurement is to measure effectively the property of the element under test, and not the property of the probe. (This is for example the problem of measuring the temperature of a quartz crystal with a temperature probe, which is never completely thermally coupled with the crystal). The second difficulty is to measure the desired quantity without being too much affected by the spurious influence of other parameters on the device under test (for instance measuring acceleration or pressure sensitivities without excessive influence of temperature), or on the measurement system itself. The measurement system is at least to some degree, sensitive to the environment. Finally, the ultimate resolution of the measurement, which is determined by the basic frequency fluctuations of the oscillator under test, is a basic limitation.

Among the different possible characterizations [1], required for quartz oscillators, temperature sensitivity remains the main one, since temperature and its variations are always present, whatever the application domain of the oscillator is. Important efforts have been made in characterizing acceleration and vibration sensitivities and in reducing them. Surprisingly, pressure sensitivity was neglected until recent years, and studies have only just begun. Generally speaking, these effects became more crucial with the development of space and airborne applications of oscillators. Interesting results have also been presented on the influence of humidity [2].

Measuring frequency or phase variations

The effect of the environment directly on the oscillator can be evaluated. This is achieved with the well known methods adapted to frequency or phase measurements. A regular test set is shown on Fig. 1. The oscillator under test is compared to a reference oscillator. The two oscillators can be locked by using a phase lock loop and the error signal at the mixer output is directly analyzed in the frequency or the time domain. Alternatively, the two oscillators are not locked, and the mixer delivers a beat frequency which is directly counted. This is the easiest way to obtain the sensitivity of the oscillator to some given effect.

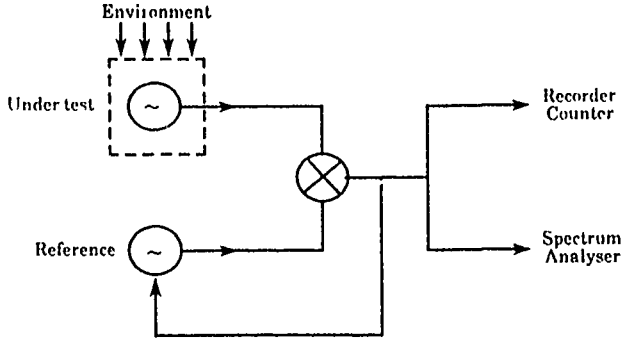


Fig. 1 :
Measurement of the frequency variations of oscillators

But if one needs more information concerning the origin and the physical mechanisms of the sensitivity, it is generally necessary to measure separately different parts of the oscillator ; and in most cases, the quartz crystal resonator itself must be characterized. Several experimental methods are available.

CH2690-6/89/0000-242 $1.00 C 1989 IEEE

The first one can still use an oscillator, but with different components removed. In this case the difficulty is the strong influence of the connections, which induce phase variations in the oscillating loop and therefore frequency variations.

The second method utilizes a passive phase bridge, the resonator (or any other component) being placed in a π transmission network (Fig. 2). Its advantage is the facility with which the tested resonator can be changed. The frequency source can be a synthesizer directly adjustable to the resonator resonance frequency.

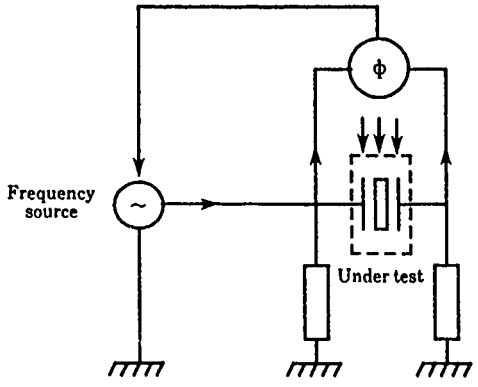


Fig. 2 : Measurement of frequency variations in transmission with a passive phase bridge

Measurement of the phase difference across the resonator directly gives the crystal resonance frequency changes Δf due to the environment, following the relation

$$\Delta\Phi = -2Q\,\Delta f / f_o \qquad (1)$$

where Q is the resonator loaded quality factor, and f_o its mean resonance frequency. The disadvantage of the transmission network method is again the spurious cable effect, which can be important if the crystal is connected to the other elements with long cables.

The third method, which was presented by Stein *et al.* in 1978 [3], uses the concept of a passive reference system. As shown in Fig. 3, a frequency source (which can be a frequency synthesizer) is locked to the resonator under test, which is operated in reflection. The driving signal is phase modulated. As a consequence the reflected signal obtained with a directional coupler is amplitude modulated if the driving signal frequency differs from the resonator resonance frequency. This amplitude modulation is detected and compared to the modulation signal with a lock-in amplifier ; the error voltage locks the frequency of the source to the resonance frequency of the resonator. The advantage of the system comes from the fact that in the comparing the reference signal (incident) and the signal which carries the information on the resonator frequency (reflected) both travel in the same cable (between the resonator and the directional coupler). Thus most of the spurious influence of the environment on the cable will be compensated, and, if necessary, it is possible to use a long cable. This measurement system allows detection of frequency changes on the order of 10^{-12} for a high Q resonator ($Q \geq 2.10^6$), even connected at the end of a several meters long cable. It is therefore particularly adapted for measurements of sensitivity to irradiation (where the electronics cannot be close to the resonator), or to accelerations (without disturbance due to the movement of the cable itself, which is partly subject to the acceleration).

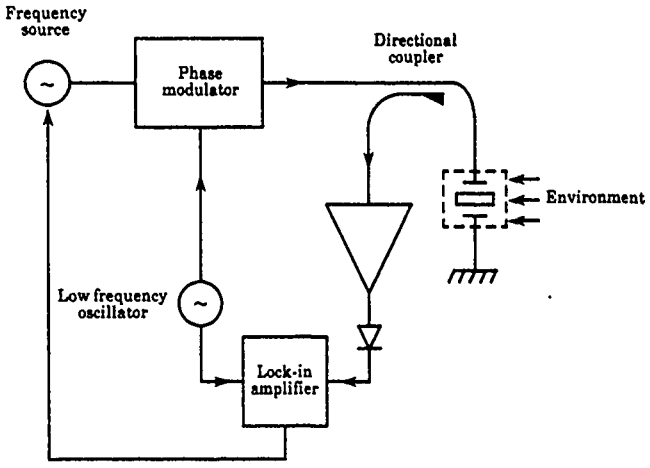


Fig. 3 : Measurement of the frequency variations of a resonator in reflection

Temperature

Temperature remains the most important cause of the frequency instability of oscillators, the resonator being in general at the origin of the sensitivity. For that reason high stability oscillators are ovenized, and at lower cost (and lower performances) oscillators are compensated. In both cases temperature must be measured by means of temperature probes, such as thermistors, thermocouples, or sometimes quartz resonators with appropriate cuts.

The first question is how to measure the right temperature, in other words the temperature of the resonator (or of the oven) rather than the temperature of the probe itself. This is illustrated by Fig. 4, which represents the thermal coupling between a temperature probe (temperature T_p), a quartz resonator (T_Q), an oven (T_o) and the external temperature (T_e). Both oven and external medium are considered in this example as large heat reservoirs with constant temperatures T_o and T_e. The thermal coupling between the crystal and the oven on the one hand and between the crystal and the probe on the other hand is limited at least by the thermal resistor of the crystal mounting structure in its enclosure. Considered as metal ribbons about 8 mm long and with a 0.75 mm^2 cross section, their thermal resistance is evaluated to be ≃ 100 K/W). Additional thermal resistances (resistances of contact) are neglected. The resistance between the probe and the exterior is of the order of 5 000 K/W, corresponding to a cable which is 4 cm long with a thermal conductivity of ≃ 80 W/K. The thermal mass of the crystal is 0.6 J/K for a 5 MHz 5th overtone resonator and 0.3 J/K for a 10 MHz 5th overtone one.

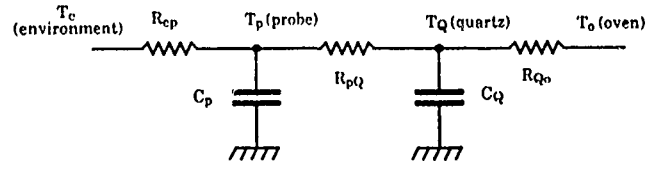


Fig. 4 : Simple diagram of thermal couplings between the quartz resonator, the temperature probe, the oven and the environment

In Fig. 5 is shown the difference of temperature between the probe and the crystal, versus time, for T_e = 20°C and T_0 = 70°C. Different types of temperature probe were used with different thermal masses. This shows that using a temperature probe which is not sufficiently well coupled to the crystal can easily give an error of 1°C in the measurement.

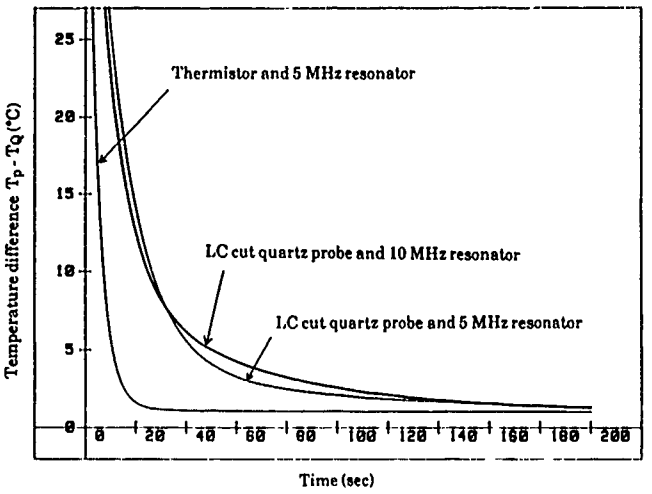


Fig. 5 : Difference of temperature between the probe and the crystal

One possibility for avoiding this problem is to use the crystal resonator itself as temperature probe. This is possible with doubly rotated cut crystals, and in particular with SC-cut crystals, by means of the B mode. The B mode has a quasilinear frequency-temperature characteristics with a sensitivity of 22 ppm/°C. Therefore, if calibrated, the B-mode can be used for measuring the temperature of the crystal with the best confidence. Another application is the possibility of calibrating the temperature stability of the oven by measuring the residual temperature fluctuations of the crystal via the B-mode [4]. Results are shown in Fig. 6.

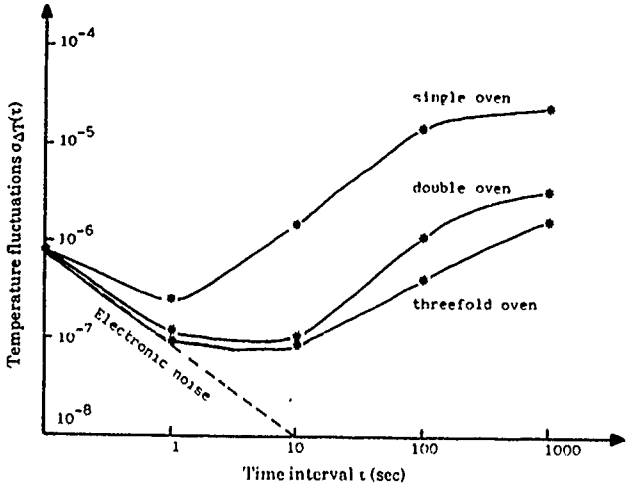


Fig. 6 :
Residual temperature fluctuations of a SC-cut quartz crystal resonator in an oven, by means of the B-mode

The second problem results from the thermal time constants of the devices (temperature probe, oven, crystal, etc.) and the superposition of the dynamic temperature coefficient of the resonator on its static temperature characteristics.

The well known relation [5] between frequency and temperature is [5]

$$(f - f_0)/f_0 = a_0(T - T_0) + b_0(T - T_0)^2 + c_0(T - T_0)^3 + \tilde{a}\frac{dT}{dt} \quad (2)$$

in which f_0 is the resonator frequency at temperature T_0 ; a_0, b_0, c_0 are the static temperature coefficients and $\tilde{a}$ the dynamic one. For AT-cut resonators $\tilde{a} \simeq -10^{-5}$ sec/°C and for SC-cut resonators $\tilde{a} \simeq 10^{-7}$ sec/°C. In view of these values it is clear that the influence of the dynamic term becomes large when the temperature variation rate increases. In Fig. 7 are shown the static and dynamic temperature characteristics of SC-cut resonator measured with temperature variation rate of 2.10^{-3} °C/sec (quasistatic) and 0.1°C/sec (dynamic). The dynamic behavior in that case must be considered as an error in the calibration of the resonator characteristics. For reducing this error below .1 Hz (2.10^{-8} for the fractional frequency of a resonator at 5 MHz), the temperature variation rate cannot exceed 2 m°C/sec for AT-cut crystals and 0.2°C/sec for the best SC-cut crystals (i.e. with $\tilde{a} \simeq 10^{-7}$ sec/°C, but this is not always the case, and the coefficient can be larger). In other words plotting the static characteristics of AT-cut crystals from -40°C up to +80°C in this condition will take at least 16 hours.

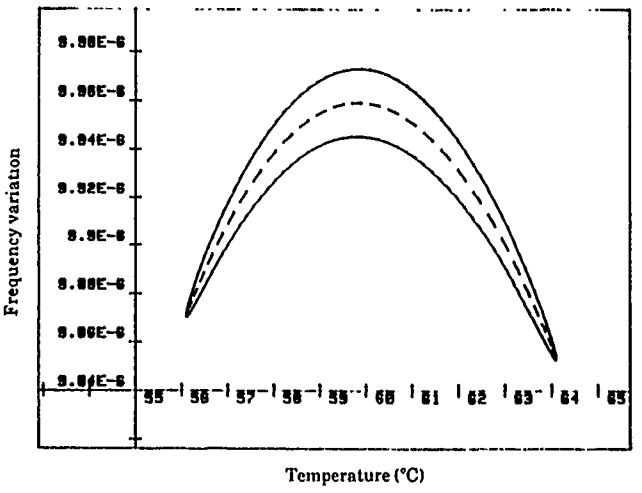
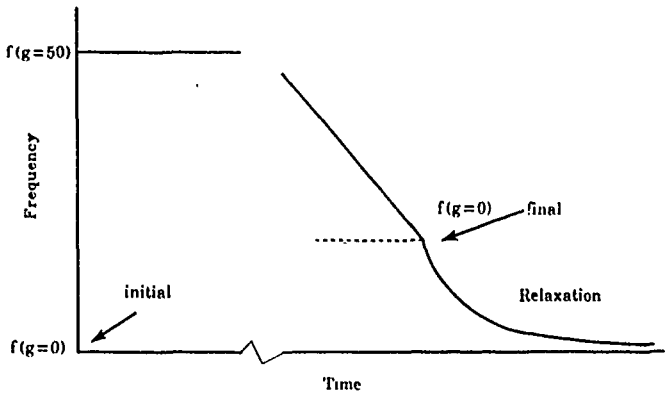


Fig. 7 :Quasistatic and dynamic frequency-temperature characteristics of a 5 MHz SC-cut resonator ($\tilde{a} = 3.10^{-7}$ °C) (- - - quasistatic : 2 m°C/sec ;—dynamic : 50 m°C/sec)

Accelerations

Some of the first measurements of the g-sensitivity of quartz resonators were made by using a centrifuge [6], which applied accelerations over a wide range from 0 to 100 g. The difficulty is the environment of the centrifuge which is not stable in terms of temperature, and also some problems related to the electrical connections with a rotating device. However it is interesting to note a thermal phenomenon that was observed. The oscillator frequency did not come back to its previous value at zero g when the centrifuge was stopped, but a relaxation of several minutes occured (Fig. 8). This relaxation can be attributed to a reequilibration of temperature in the oven perturbed by transfer of air masses, when subjected to the high acceleration field. The phenomenon will be generally neglegible for small accelerations, but it is important to know that it can happen.



Fig; 8 : Thermal relaxation following the application of a 50 g DC acceleration on an ovenized oscillator

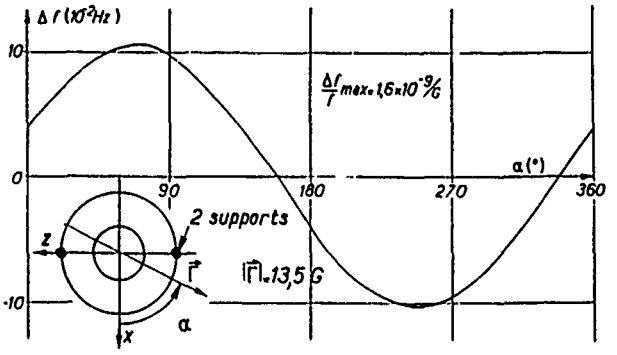


Fig. 10 : Measurement of the g-sensitivity of a quartz crystal resonator by rotation in the earth gravitational field (tip-over test) (Janiaud, 1978)

Measurements of the g-sensitivity at lower accelerations are made easier by using the tip-over method. In Fig. 9 are shown the g-sensitivities of a quartz resonator, the acceleration being applied along three orthogonal axis and in the two opposite directions on the same axis. These results presented by Valdois in 1974 [6] indicate that the g-sensitivity changes sign when the direction of the acceleration is reversed.

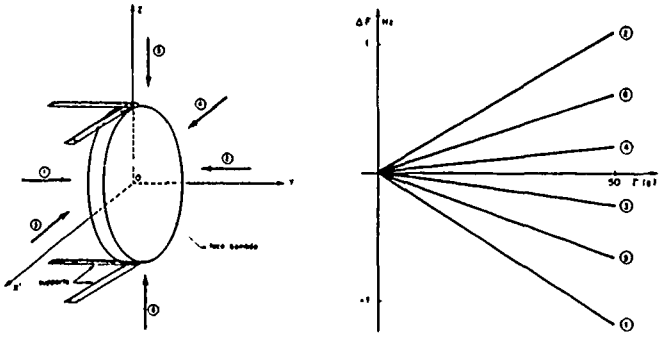


Fig. 9 : g-sensitivity of a quartz resonator as afunction of the acceleration direction

The g-sensitivity can thus be measured by rotating the device around an axis perpendicular to the gravitational field as indicated by Fig. 10. This is the easiest method, which allows one to completely characterize the oscillator or resonator by rotation around three orthogonal axis, following the relation

$$\frac{\Delta f}{f_o} = K_i \cdot \Gamma_i \qquad (3)$$

where Γ_i are the components of the acceleration with respect to the three axis and K_i are the g-sensitivity coefficients along the same three axis.

The resolution in such a measure depends on the accuracy of the angular positioning, which can be for the most sophisticated equipment on the order of a few seconds of arc (equipment used for calibrating high performance accelerometers).

Pressure

The pressure sensitivity problem is encountered in oscillators used for airborne applications. The effect was originally expected to be small since the crystal is mounted in a glass or metal enclosure under vacuum. In practice, variations of the atmospheric pressure induce deformations of the enclosure, which are transmitted by the mounting structure as forces acting on the crystal. The measurement of the pressure sensitivity appears simple, using a closed container and a vacuum pump. But thermal effects can also take place. For example a modification of thermal transfer in the oscillator can occur when the gas is evacuated. This is a drastic effect with ovenized oscillators, because the characteristics of the oven are completely different in air at atmospheric pressure and in vacuum. This is illustrated in Fig. 11, which shows the frequency variations of the resonator itself and of the oscillator in an oven.

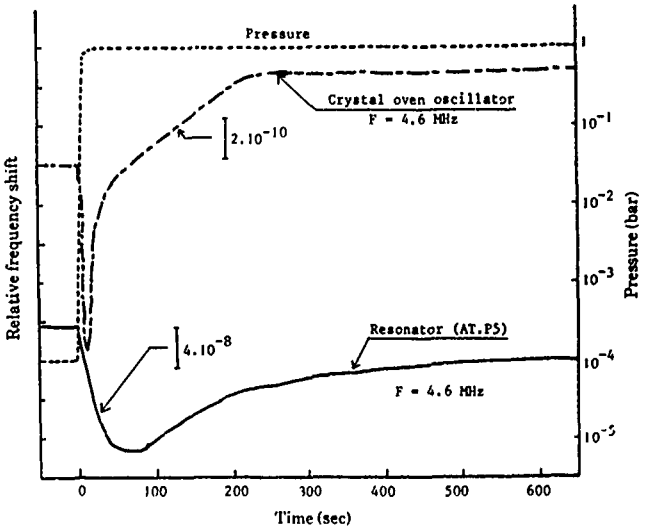


Fig. 11 : Frequency-versus pressure for a quartz oscillator with in addition spurious thermal effects

Magnetic fields

The sensitivity of quartz oscillators to magnetic fields is on the order of 10^{-12}/gauss [7]. This means that the best oscillators can detect the earth's magnetic field, but also that measuring the sensitivity for fields of the order of a gauss requires some care. Of course measurements can be performed in an electromagnet field of several 1000 gauss. In that case the effect is not difficult to detect, but for such high fields saturation occurs (as shown in Fig. 12), and it is not possible to evaluate by extrapolation the sensitivity at low field. It is obvious that the low field effect (shown in Fig. 13) cannot be measured unless the oscillator stability is better than the effect itself.

AC magnetic fields are also a source of problems, since AC, or any transient magnetic field, dissipates energy (proportional to the area of the hysteresis cycle in the corresponding material). This causes spurious thermal effects. In addition eddy currents are generated.

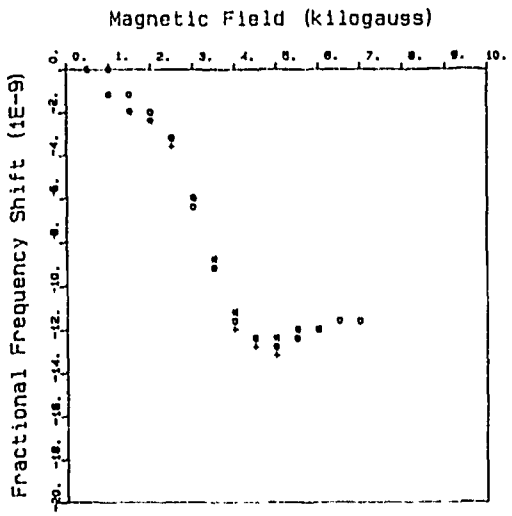


Fig. 12 : Sensitivity of a quartz oscillator to high magnetic field (Brendel, ref. 7)

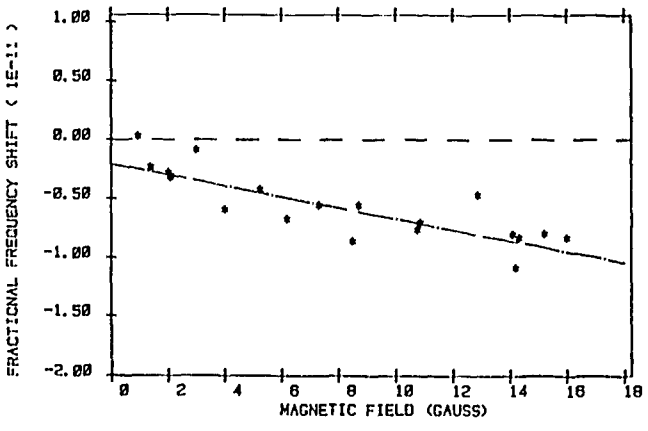
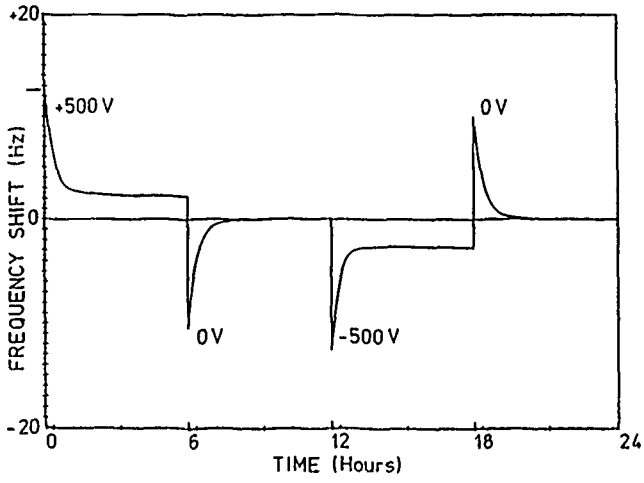


Fig. 13 :
Sensitivity of a quartz oscillator to low magnetic fields (Brendel, ref. 7)

Electric fields

Measuring the sensitivity of a quartz resonator to electric fields is simple, since the effects are of the order of 10^{-9}/V. But the presence of impurities in interstitial positions modifies the initial phenomenon, because the impurities diffuse under the influence of the field and create a depolarization. This behaves as a relaxation (Fig. 14), with a time constant that depends on the nature of the impurity and on the temperature [8]. In fact knowledge of these phenomena is of interest for oscillators when their circuitry is such that polarization is applied on the crystal. The impurity relaxation induces a frequency drift which can appear as long term frequency aging.



Fig; 14 : Sensitivity of a 5 MHz SC-cut crystal to an electric field combined with impurity relaxation

Conclusion

The presentation was limited to some examples of environmental sensitivity measurements and was not intended to exhaustively cover all possible topics. Radiation was not discussed, even though it is an important subject, which brings up the same questions: how to measure radiation sensitivity without being perturbed by temperature ? How to test separatly the different components or parts of the oscillator ? Humidity is another environmental condition which influences oscillator characteristics, as was recently shown for the first time by Walls [2]. Other disturbing parameters in an oscillator's environment are still unknown today and not yet anticipated. But some will certainly contribute to the frequency instabilities of concern in the future.

Acknowledgements

The author would like to thank MM. Djian, Brendel, Marianneau, and El Hassani for their help in preparing the paper, and for providing results and data.

References

[1] H. Hellwig, "Environmental sensitivities of precision frequency sources", 3rd European Freq. and Time Forum, Besançon, 1989.

[2] F.L. Walls, "The influence of pressure and humidity on the medium and long-term frequency stability of quartz oscillators", 42nd Ann. Freq. Cont. Symp., 1988.

[3] S.R. Stein, C.M. Manney, F.L. Walls, J.E. Gray, R. Besson, "A systems approach to high performance oscillators", 32nd Ann. Freq. Cont. Symp., 1978.

[4] G. Marianneau, J.J. Gagnepain, "Digital temperature control for ultrastable quartz oscillators", 34th Ann. Freq. Cont. Symp., 1980.

[5] A. Ballato, J.R. Vig, "Static and dynamic frequency-temperature behavior of singly and doubly rotated oven-controlled quartz resonators", 32nd Ann. Freq. Cont. Symp., 1978.

[6] M. Valdois, J. Besson, J.-J. Gagnepain, "Influence of environment on a quartz resonator", 28th Ann. Freq. Cont. Symp., 1974.

[7] R. Brendel, C. El Hassani, M. Brunet, E. Robert, "Influence of magnetic field on quartz crystal oscillators", This Symposium.

[8] R. Brendel, "Etude des défauts ponctuels dans les résonateurs à quartz par effet électroélastique", 1st European Freq. and Time Forum, Besançon, 1987.

43rd Annual Symposium on Frequency Control - 1989

EXPERIMENTAL PROCEDURE APPLIED TO CHARACTERIZE CRYSTAL OSCILLATOR UNDER FUNCTIONAL ENVIRONMENT

By G. ROBICHON, A. DEBAISIEUX and JP. AUBRY
CEPE 44 avenue de la Glacière 95100 ARGENTEUIL

ABSTRACT

Oscillators performances are very often specified under various kind of environmental stresses.

G sensibility is probably the more popular but others stresses are also specified such as electromagnetic, magnetic fields, barometric....

In order to check the performances of oscillators and resonators under nominal environmental conditions, tests equipments and methods have been specifically designed in various field, these methods being not yet standardized. We will present the methods used by CEPE in some areas :

- g sensibility : dynamic test (spectrum analysis of phase fluctuation under know random vibrations).

- acoustic sensitivity (spectrum analysis of phase fluctuation under known acoustic excitation)

- power supply sensitivity (spectrum analysis of phase fluctuation with ripple on power svpply)

- barometric (pressure) sensitivity

- magnetic field sensitivity

Description of test equipments and procedures of these various measurements will be given. Typical results (including reproducibility) will be given.

INTRODUCTION

For Air borne radar, customers are more and more able to describe the environment of the crystal oscillator. The electrical specification (phase noise, frequency stability output power, etc) has to be met under Functional environment. That one becomes more severe and electrical performances asked are always better. The normalization of these tests follows technical progresses. Few of them are discribed below.

I/ ACCELERATION

I.1/ Static

+/- 100 g is the maximum value which can be tested on the equipement used by CEPE .

This test is performed on a rotating machine (fig. 1) of the ONERA which is a French governmental organization specialised in aeronautic and space applications.

CEPE has to prove that the oscillator fulfill the specification. The first parameter we think about is the frequency deviation.

That one depends on the g sensitivity of the crystal. The difficulty comes from additional effects such as : acoustic excitation and vibration. The phase noise is then affected. The problem is to prove that this degradation is not due to the static acceleration.

FIG 1

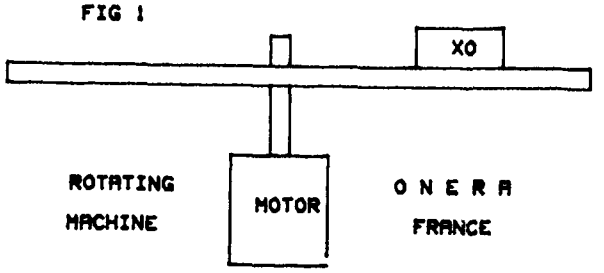


Fig. 1 : Rotating machine

I.2/ Vibration

Phase noise degradation is described by filter's relation :

$$L\ (dBc/Hz) = 20 \log \left(\frac{\sqrt{A^2\ RSM}}{\sqrt{2}} . \gamma . \frac{F}{f} \right)$$

Where :

A^2 RMS is the vibration level in g^2/Hz
γ is the g sensibility value
F is the frequency of the oscillator
f is the frequency of the vibration

CH2690-6/89/0000-248 $1.00 C 1989 IEEE

The max f is from the normalization limited up to 2 KHz.

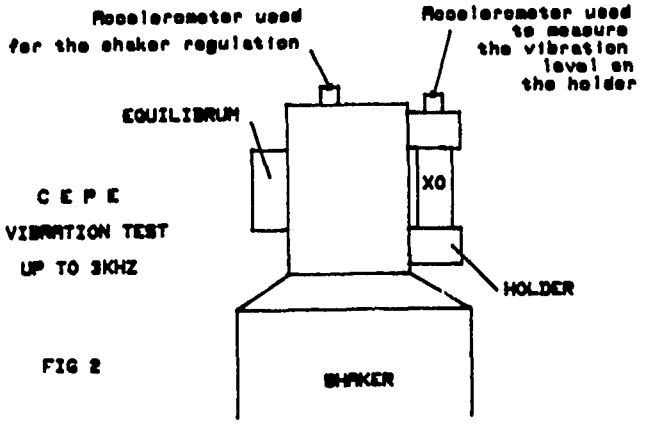


FIG 2

Fig. 2

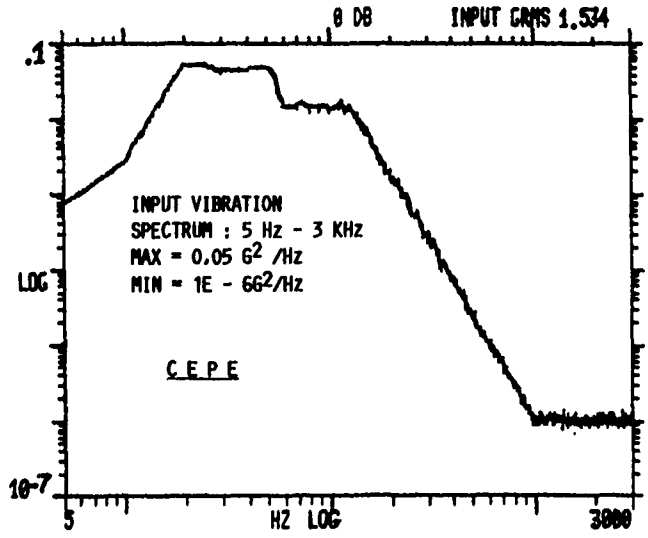


Fig. 3

Some applications require vibration up to 3 KHz and CEPE produce oscillators 100 % tested up to 3 KHz. The test requires a second accelerometer in order to measure the exact vibration level applied on the oscillator. Distorsion of about 6 dB can appears between the first accelorometer used for the phase looked loop and the second one at 3 KHz.

The difficulty to make such an oscillator is to find components tested up to 2 KHz and to use them up to 3 KHz.

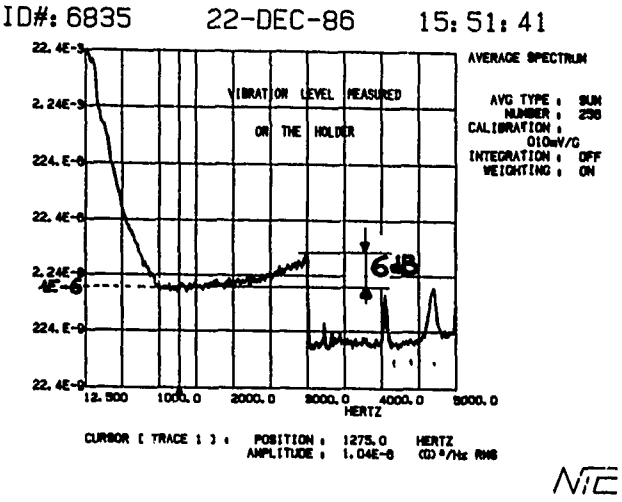


Fig. 4

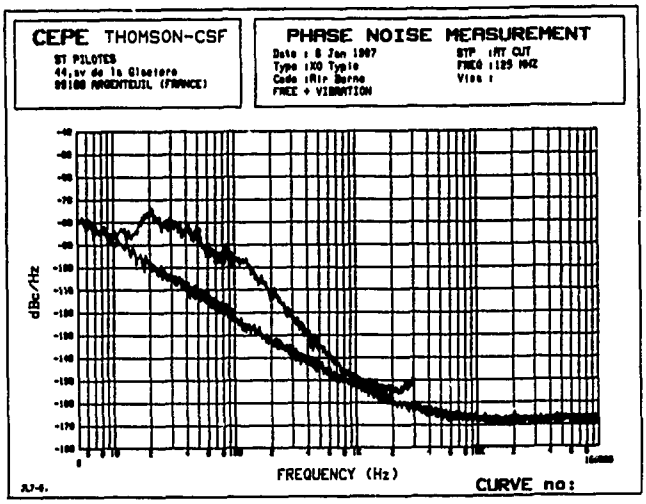


Fig. 5

II/ ACOUSTIC

II.1/ Low level (~ 100 dB)

The equipement used to test crystal oscillator is at CEPE an anechoïc chamber transformed into a reverberent one according to the normalization.
The acoustic level is produced by an electrical loud speaker and controled by a microphone.

Customer gives to CEPE the acoustic spectrum and CEPE proves that electrical specification (phase noise) is met under acoustic excitation. CEPE produces to day crystal oscillator 100 % tested under acoustic environment.

II.2/ ACOUSTIC : High level up to 160 dB (réf. 2.10^{-5} Pascal)

For higher level CEPE performs the acoustic test in BERTIN company. They use a pneumatic loud speaker in a reverberent chamber. Its volume is 12 m3. Sound presure level is measured with a microphone. Note that 160 dB is about the sound presure level obtained at 10 m from the output of the space launcher reactor.

We have at CEPE a study for a VCXO large shift with lithium tantalate crystal able to meet the phase noise specification under a sound pressure level equal to 145 dB. The acoustic spectrum is not yet determinated.

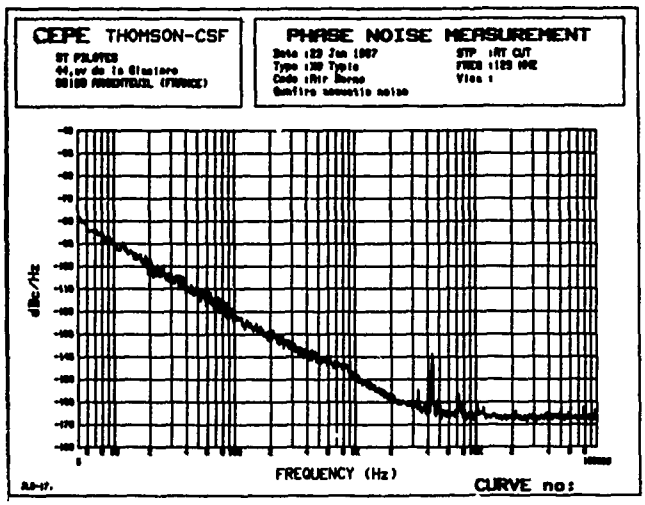
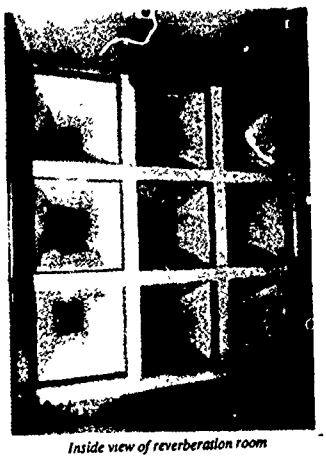


fig. 7



Inside view of reverberation room

Characteristics

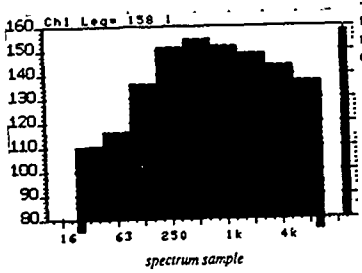
- Metallic reverberation room with non-parallel walls, to achieve a homogeneous diffuse acoustic field
- Useful volume : 12 m3, international dimensions 3 x 2 x 2 m
- Sound source : up to 9 BERTIN fluid-amplification pneumatic sirens, with 9 low-frequency exponential mouths.
- Sound pressure level : 160 dB (reference 2×10^{-5} Pascal) as a global level or in a frequency octave (or third of octave) band.
- Possible spectrum reshaping

Experimental procedures

- Real time measurement and acquisition of the sound spectrum or of the mean spectrum or equivalent throughout the test
- Possible real time acquisition and processing of the test parameters, up to 32 channels (accelerometers, stress gauges, pressure sensors, thermocouples, etc)

BERTIN services

- Supply of test means with skilled personnel for measurement implementation and processing
- On client's request, contribution to instrumentation of the model and analysis of results



spectrum sample

Fig. 6

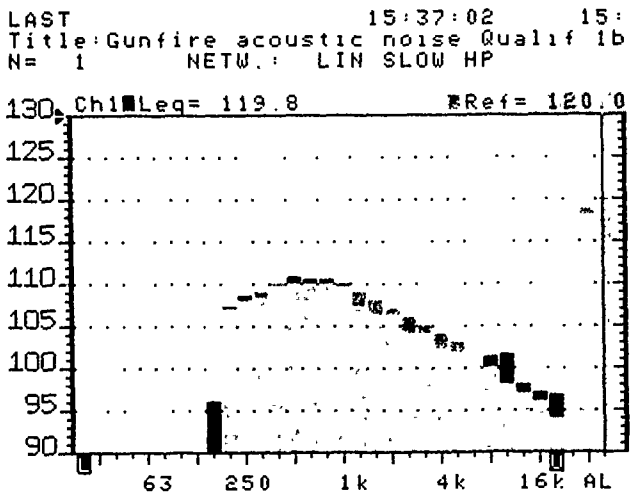


Fig. 8

III/ RIPPLE ON POWER SUPPLY

III.1/ Low frequency

Up to 100 KHz, the test is performed without any difficulties. The pollution is obtained by a power transistor.

There is no difference between the level at the transistor out put and the power supply input of the crystal oscillator.

Up to 1 or 2 MHz, in general, the equipment can be used but a difference level appears between the two points.

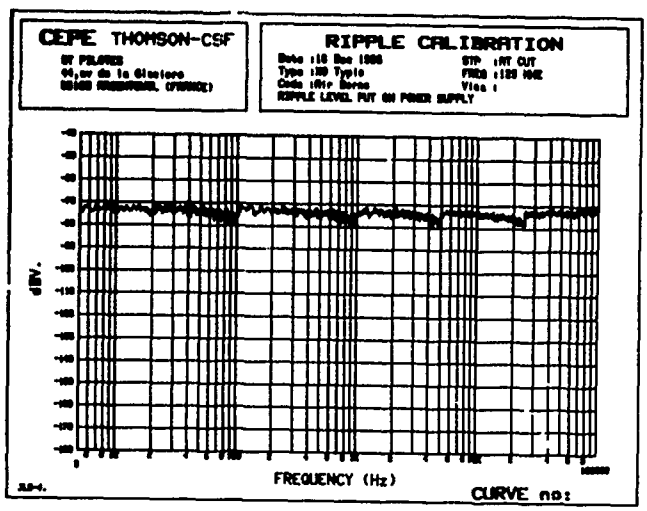


Fig. 9

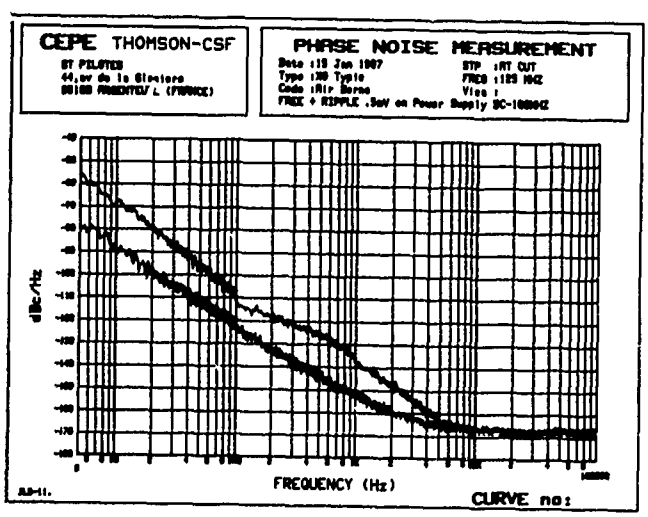


Fig. 10

III.2/ High frequency

The problem of a ripple on power supply between 1 MHz and 1 GHz is difficult if the impedance becomes equal to zero (case of bypass capacitor filter). We propose in this case a calibration of the ripple level when the crystal oscillator is unconnected.

We call this level the potential ripple power. A coaxial cable is used and 50 ohms load is placed as near as possible of the terminal. The output impedance is also equal to 50 ohms.

The phase noise is degraded under ripple on power supply. The performance is improved by increase of internal power supply treatment.

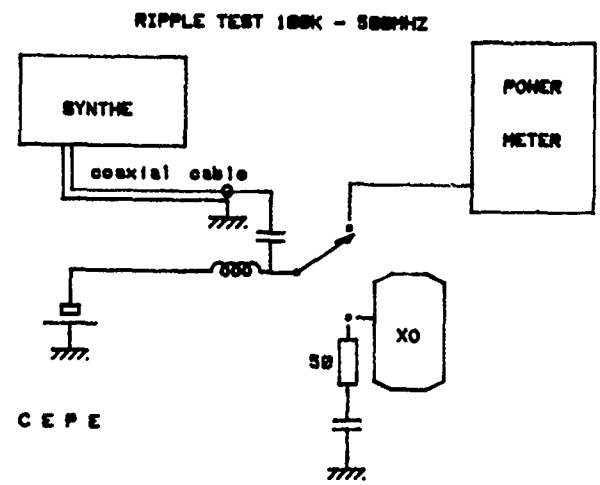


Fig. 11

In order to suppress discussion between supplier and customer we suggest to use a by pass filter whith inductance -capacitor- inductance type.

VI/ MAGNETIC FIELD

IV.1/ Static

A frequency deviation is measured when crystal oscillator is placed into a static field. The test is performed by mean of Helmotz inductance. The frequency deviation is measured when the static magnetic fiel is inversed.

IV.2/ Dynamic

The first method in order to make a characterization of the oscillator under dynamic magnetic field is to put crystal oscillator inside an inductance. The frequency of the excitation is between 20 Hz an 2 KHz. In this case, the magnetic field is alternative only for one direction.

The second method consists on using two perpendicular inductances.

This method allows to generate a turning magnetic field.
Inside the crystal oscillator, it is proved now that the self inductors of the oscillator make the largest part of magnetic sensitivity. Consequently, we think that the turning magnetic field method is the best one.

Under this test, CEPE has to prove that the phase noise of the oscillator fulfils the spurious non-harmonic specification (dBc value).

V/ BAROMETRIC PRESSURE SENSIBILITY

This test is a static experience. It consists on measuring the shift of frequency due to the barometric pressure variation.

Our equipment is able to reach secondary vacuum and to make a simultaneously temperature variation. The typical sensitivity is about 10^{-9}/atmosphere or $\Delta F/F \sim 5.10^{-13}$/Pascal.

We note that, for ovenized crystal oscillators the frequency deviation versus temperature range is modified when pressure is changed.

For the test, the crystal oscillator is rigidely screwed on the floor of the equipement.

CONCLUSION

Technical progresses new measurement method will be followed by the normalization. We hope that CEPE's contribution will improve the clarity of discussions between suppliers and customers.

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] JP AUBRY, G. CARET, G. MAROTEL, BNM Temps-Frequence (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] JP VALENTIN - Thèse de Doctorat es Sciences Physiques Besançon 1983
"Analyse de quelques aspects du comportement thermique des résonateurs à quartz à ondes de volume.

43rd Annual Symposium on Frequency Control - 1989

SPECIFICATION AND MEASUREMENT OF THE FREQUENCY VERSUS TEMPERATURE CHARACTERISTICS OF CRYSTAL OSCILLATORS

RAYMOND L. FILLER, VINCENT J. ROSATI,
STANLEY S. SCHODOWSKI, AND JOHN R. VIG
U.S. Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

SUMMARY

We review several factors to consider when specifying and measuring the frequency vs. temperature (f-T) characteristics of precision quartz crystal oscillators. Topics include static vs. dynamic measurement, thermal time constant, activity dips, condensables, hysteresis, and trim effect.

INTRODUCTION

There is a hierarchy of quartz crystal oscillators (Figure 1), ranging from uncompensated clock oscillators (XO) that have a many parts-per-million (ppm) (f-T) accuracy, to oven-controlled oscillators (OCXO) that can have fractional parts-per-billion (ppb) accuracy. Between these is a wide variety of analog and digital temperature-compensated crystal oscillators (TCXO), and microcomputer-compensated oscillators (MCXO). Generally, the more stable an oscillator is, the more carefully it must be specified, and the more difficult it is to measure. Furthermore, there is an almost direct relationship between accuracy and price, which carries over to the cost of the required test setup.[1] A major component of the price of an oscillator is the cost of testing.

The details of specifying f-T characteristics can be found in MIL-O-55310.[2] This paper reviews some f-T considerations that are not discussed in detail in MIL-O-55310.

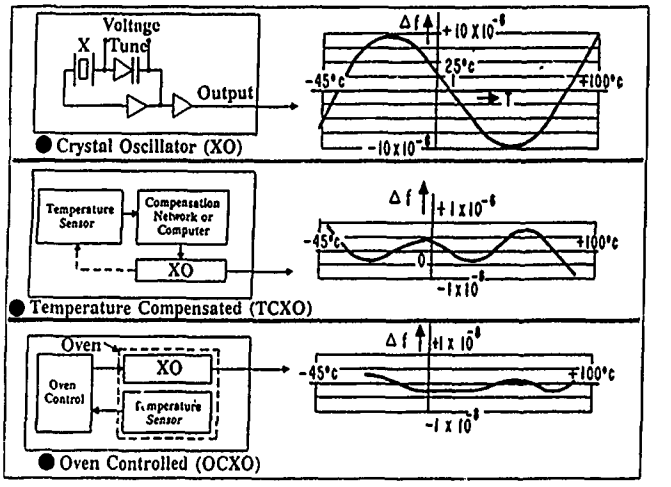


Figure 1. Hierarchy of crystal oscillators.

STABILITY vs. ACCURACY

Two often-misused terms are stability and accuracy. Figure 2 illustrates the difference between the two terms. The definition of frequency-temperature accuracy is the maximum measured deviation of the oscillator frequency from the specified nominal value due to operation over the specified temperature range.[2] The definition of frequency-temperature stability is the maximum measured deviation of the oscillator frequency, with no reference implied, due to operation over the specified temperature range,[2] i.e., ±(max f - min f)/2. Frequency-temperature accuracy considers the combination of f-T stability and calibration-related errors, and should be the preferred method for specifying the f-T characteristics of crystal oscillators.

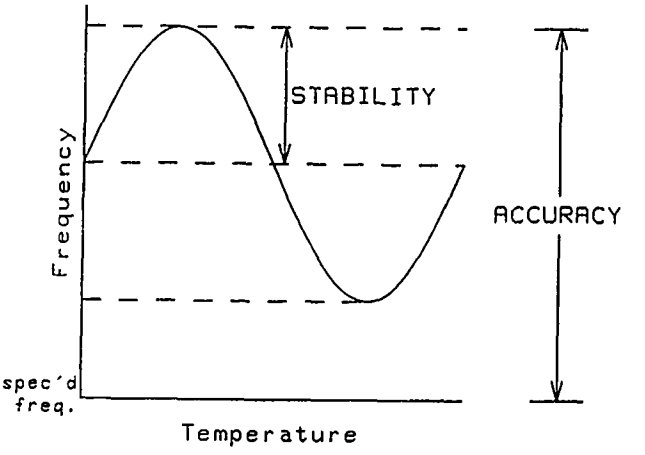


Figure 2. Stability vs. accuracy.

At the outset, designers concern themselves chiefly with stability, assuming that they can always trim the oscillators to the specified nominal frequency. This is often not true, partly because of the trim effect, which will be reviewed later. Systems engineers almost always demand accuracy, although MIL-O-55310 allows for specification of the f-T characteristic in terms of either frequency-temperature stability or frequency-temperature accuracy.

MEASUREMENT OF THE f-T CHARACTERISTIC

Measurement of the f-T characteristics of

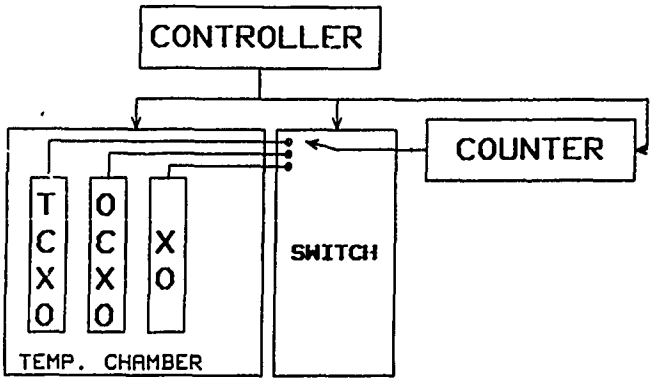


Figure 3. Frequency-temperature measurement setup.

crystal oscillators is normally carried out using a setup similar to the one in Figure 3. The setup is straightforward, but there are several subtleties that must be taken into account to avoid inaccurate characterization.

PITFALLS

The primary cause of the variation of oscillator frequency with temperature is the inherent f-T characteristic of even a perfect quartz resonator. There are, however, other elements that influence oscillator performance. Some are due to imperfect resonators, others to circuit design, and still others to in-process adjustment of the oscillator. Among these influences are: activity dips, condensables, thermal hysteresis, trim effect, and oven offset.

Static vs. dynamic testing

Figure 4 shows a comparison of the response of AT-cut and SC-cut resonators to a large, abrupt change in temperature caused by the turn-on of the oven in an OCXO. The figure demonstrates that sufficient time must be allowed for the resonator to stabilize before accurate frequency can be obtained. Note that the SC-cut shows very little thermal transient effect.

Thermal time constants

Figure 5 shows how the thermal time constant of the oscillator moderates the thermal step. The time constant may add considerable time to the measurement cycle and must be considered.[3] One must be certain to allow enough time for the oscillator to stabilize before making a measurement if the static f-T characteristic is desired.

Activity dips

Activity dips, depicted in Figure 6, are usually caused by interfering modes that can occur within a narrow temperature range and cause severe frequency shifts. In the example shown, the resonator was intended for use in a precision OCXO; note the activity dip near the turnover temperature. A small oven drift could cause a frequency shift of about 0.3 ppm. If this resonator was used in a TCXO, the activity dips could go unnoticed during testing, unless small temperature steps were used during f-T characterization. If testing for high precision, measurements should be made at closely spaced temperatures.

Condensables

Figure 7 is a plot of the f-T curve of a TCXO which shows a phenomenon that, at first glance, appears to be an activity dip. Upon closer examination, the figure shows that the frequency shift occurs over a temperature range wider than is characteristic of activity dips, and it occurs only in one direction of the temperature sweep. The cause of this phenomenon is, probably, condensible vapor in the resonator enclosure. To detect the presence of condensables, one should ramp the temperature in both directions after a rapid cool-down.

Thermal hysteresis

Looking at the low-temperature end of Figure 7, notice that the frequency curves do not

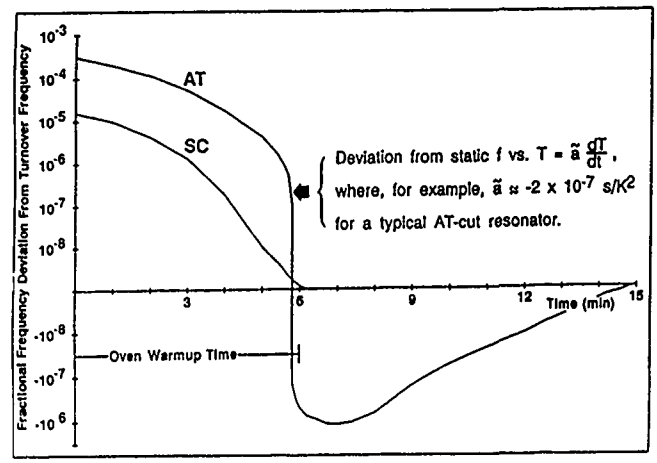


Figure 4. Response of AT- and SC-cut resonators to an abrupt change of temperature.

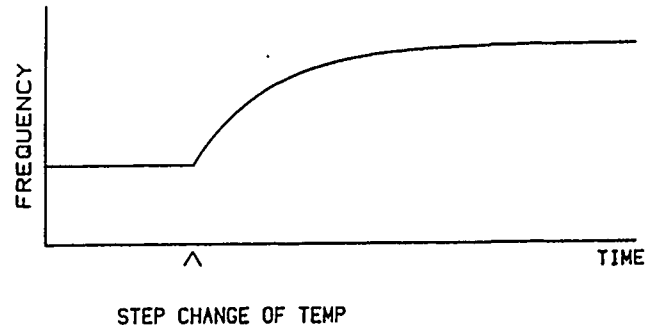


Figure 5. Oscillator thermal time constant.

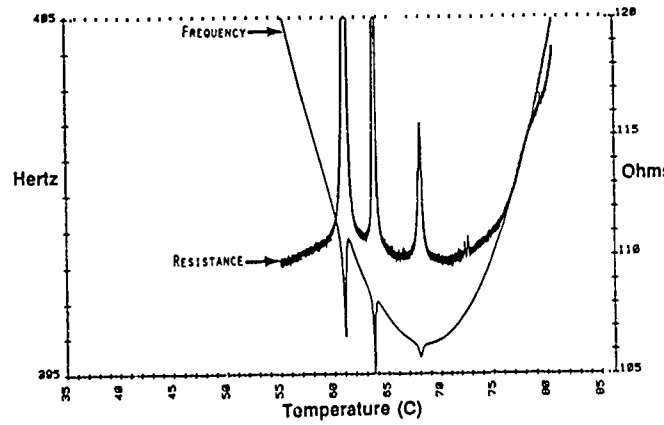


Figure 6. Activity dips.

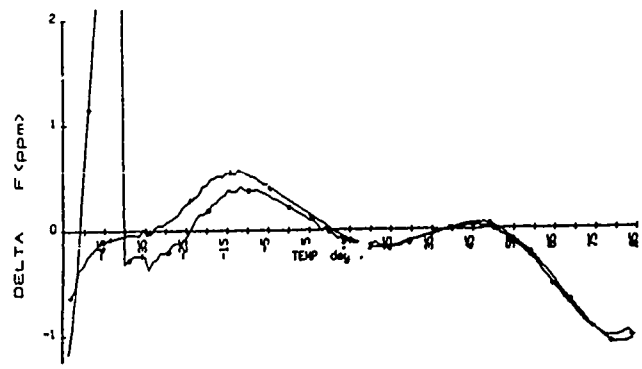


Figure 7. Condensables.

superpose, but are displaced vertically. This is thermal hysteresis, defined as a thermally induced effect in crystal resonators which manifests itself as nonrepeatability of the frequency-temperature characteristic observed over a complete quasi-static temperature cycle.[2] Figure 8 is a plot of a different TCXO; it shows thermal hysteresis present at all temperatures. (Note that thermal hysteresis usually degrades the accuracy of an oscillator.) The top curve in Figure 9 shows yet another example, where there is considerable thermal lag besides thermal hysteresis, which indicates that not enough time was allowed for the gradient between the thermistor and the resonator to stabilize before making a frequency measurement. In the bottom curve, the thermal lag has been eliminated by computer processing of the data, resulting in the "actual hysteresis."[4] To properly characterize an oscillator, it is necessary to measure over a complete up and down temperature sweep, and allow enough time for all components to reach thermal equilibrium.

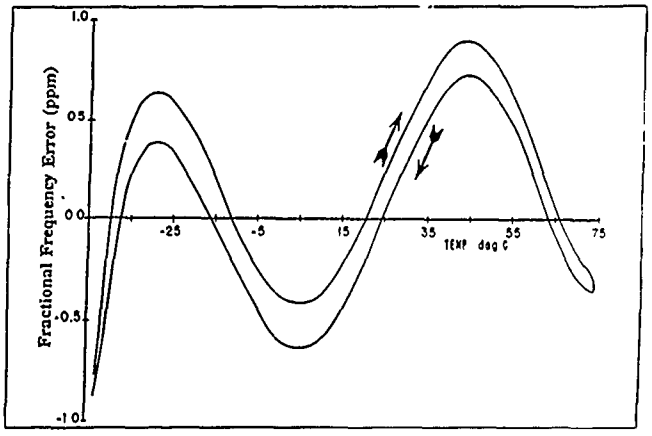


Figure 8. Thermal hysteresis.

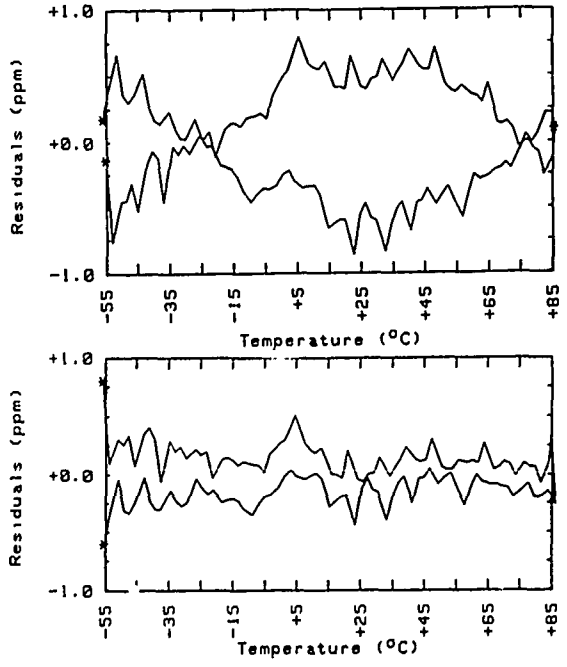


Figure 9. Top: "Apparent hysteresis." Bottom: Actual hysteresis.

Trim effect

In TCXO's, a signal from a temperature sensor (thermistor) is used to generate a correction voltage that is applied to a voltage-variable reactance (varactor) in the crystal network. The reactance variations compensate for the crystal's f vs. T characteristic. Reactance changes also are used to compensate for aging. The frequency vs. reactance relationship is nonlinear, leading to the trim effect, i.e., the degradation of crystal oscillator frequency-temperature stability as a result of frequency adjustment.[2] The trim effect manifests itself as a rotation or distortion, or both, of the initial frequency-temperature characteristic. Figure 10 depicts the nonlinearity of the frequency/load-capacitor relationship, and Figure 11 shows the result of varying the load capacitor. It is important to measure the f-T characteristic at the specified extremes of the frequency adjustment range.

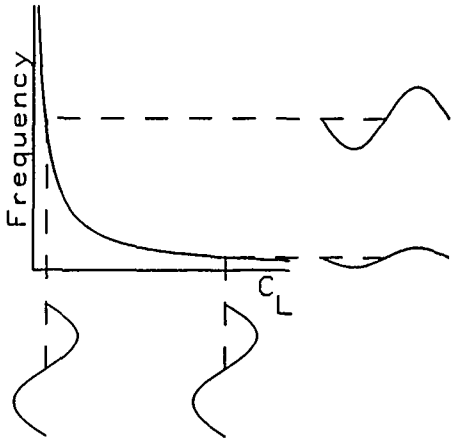


Figure 10. Nonlinear frequency vs. load capacitance

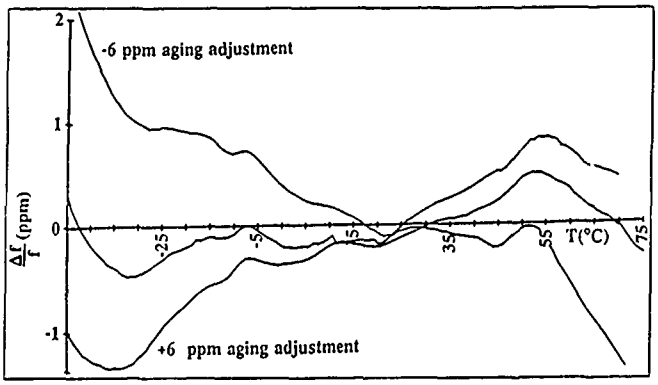


Figure 11. Trim effect.

CONCLUSIONS

To summarize the considerations for specifying and measuring the f-T characteristics of quartz crystal oscillators:

1. Take thermal time constants and thermal transient effects into account.
2. Make measurements at temperature intervals commensurate with the specified performance of the oscillator to search for narrow f-T anomalies.
3. Measure over at least one complete up and down temperature cycle to check for hysteresis and condensables.

4. Measure the f-T characteristic at both extremes of the frequency adjustment range to ensure that the trim effect does not degrade f-T performance to an unacceptable level.

REFERENCES

1. V. J. Rosati, S. S. Schodowski, and R. L. Filler, Temperature Compensated Crystal Oscillator Survey and Test Results," Proceedings of the 37th Annual Symposium on Frequency Control, pp 501-505.

2. " Military Specification, Oscillators, Crystal, General Specification For," MIL-O-55310B, Available from: Naval Publications and Forms Center, 5801 Tabor Avenue, Philadelphia, PA, 19120.

3. S. S. Schodowski, "Temperature Performance Measurement Methods for Temperature Compensated Quartz Oscillators," Research and Development Technical Report ECOM-2896, National Technical Information Service, 1967.

4. R. L. Filler, "Measurement and Analysis of Thermal Hysteresis in Resonators and TCXO's," Proceedings of the 42nd Annual Symposium on Frequency Control, 1988, pp 380-388.

43rd Annual Symposium on Frequency Control - 1989

THE TESTING OF RUBIDIUM FREQUENCY STANDARDS

T. J. Lynch, W. J. Riley, and J. R. Vaccaro
EG&G Frequency Products
Salem, Massachusetts 01970

ABSTRACT

At the present time no standards exist for the testing of atomic frequency standards. The methods used for the qualification and acceptance testing of such devices are devised on a case-by-case basis, with the test procedures defined by the manufacturer or customer. Similarly, there are no standard test methods to aid the user in his evaluation of such devices.

An effort is currently beginning under NIST and IEEE sponsorship to define standard test methods for the environmental sensitivities of atomic frequency standards. That project is expected to eventually result in a document (similar to Section 4.9 of MIL-O-55310B) that will define standard test methods to evaluate, quantify, and report the sensitivity of the frequency of standard frequency generators under environmental influences such as magnetic fields, atmospheric pressure, humidity, shock, vibration, acceleration, temperature, ionizing radiation, and intermittent operation.

This paper describes the test methods used at EG&G for measuring the environmental sensitivities of rubidium frequency standards. It discusses the objectives of such tests, the parameters to be considered and measured, the rationale for the test methods, the setups required to implement the tests, and the way that the test results are reported. The paper also considers Environmental Stress Screening (ESS) of rubidium frequency standards. Examples of these tests are given.

It is hoped that the exchange of this information between the organizations involved will facilitate the adoption of standard methods for such test activities.

INTRODUCTION

The success of a rubidium frequency standard (RFS) in a particular application often depends more on its ability to withstand the operating environment than its performance under benign conditions. This is especially true for a tactical RFS in a military application. It is therefore very important that the proper choices be made in specifying the environmental tests and conditions that convey the operational requirements to the vendor, and allow the product to be correctly designed and qualified. Similarly, it is vital to choose the appropriate acceptance tests that will most economically ensure compliant production hardware.

An example of a Qualification Test Plan for a tactical RFS is given in Reference 1. That test regimen includes just about every kind of environmental test. Some of these tests verify critical performance parameters (such as stability and phase noise under vibration). Others constrain the design and construction; still others guard against unexpected weaknesses. The list of these tests is familiar; it is during the process of incorporating these tests into a detailed test plan that the specific RFS considerations need to be addressed.

ENVIRONMENTAL TESTS

The environmental requirements for an item of military electronic equipment (such as a rubidium frequency standard) can involve a bewildering maze of data sheets, contractor equipment specifications, and a hierarchy of military specifications (some of which are shown in Table 1)[2-11] . It is impossible to generalize about such requirements except to say that the RFS equipment specification (which usually takes precedence) must be carefully written to define the proper requirements. It is unusual for a RFS procurement to involve complete environmental testing because existing test data can often be applied "by similarity." New designs or particularly demanding applications may, however, require a full series of environmental tests. In those cases, a test plan or procedure should be written and approved before testing begins. In addition to the tests themselves, the plan should define the order of the tests and whether more than one unit can be used, and should address such matters as instrument calibration[12] and discrepancy reporting. It is common to include a table of standard operating conditions and a basic functional test that is used before and after each environmental test to verify normal RFS operation. This usually includes measurement of output level and frequency, dc power input, and monitor readings. The test procedure should contain blank data sheet forms.

The most demanding combination of RFS environmental requirements is usually associated with an airborne application which includes high temperature and severe vibration. While it is possible to design a RFS to operate at the +95°C intermittent conditions of MIL-E-5400 Class 2, it should be realized that the stability of a RFS subjected to extreme temperature and vibration can be far worse than under benign conditions. And a RFS design constrained to meet extreme environments can

Table 1. Military Specifications for the Testing of Electronic Equipment.

RFS USAGE	GENERAL SPECIFICATION	TEST REQUIREMENTS	ENVIRONMENTAL TEST METHODS	EMI TEST METHODS
All	MIL-STD-454		MIL-STD-810	MIL-STD-461
Airborne	MIL-E-5400	MIL-T-5422		
Ship	MIL-E-16400			
Ground	MIL-E-4158			
Space	DOD-E-8983	MIL-STD-1540		MIL-STD-1541

CH2690-6/89/0000-257 $1.00 C 1989 IEEE

result in compromised performance under benign conditions. This makes correct environmental specification and test particularly critical.

A list of the most common environmental tests used to evaluate and qualify a rubidium frequency standard is shown in Table 2. Also shown is the most critical RFS parameter associated with each test, whether the test is normally performed as part of the RFS Qualification Test Procedure (QTP) or Acceptance Test Procedure (ATP), and references to common military specifications for these environmental tests [3,4,13,14]. While the specific test conditions (temperature range, vibration levels, etc.) will vary with the particular application, certain general considerations apply to these RFS tests and they will be discussed in the following sections of this paper.

Test Setup. A typical test setup for the environmental testing of a RFS is shown in Figure 1. The RFS under test is installed in an environmental chamber (which may be a temperature, humidity, altitude, or other such chamber, or a shaker, shock machine, or other such apparatus). It is powered from a dc supply and its current and monitors are measured by digital or analog meters, and its recovered signal is observed on an oscilloscope. (Alternatively, a custom "monitor box" may be used, especially if special control or monitor functions are involved.) A stripchart record of the RFS dc supply current is often useful for evaluating performance and diagnosing problems. Besides warmup and steady-state power, the current record shows lamp ignition and oven stability, and can be integrated to determine RbXO [20] syntonization energy. The primary measurements, however, are the RFS rf output level, frequency, and purity. The frequency measuring system may be simply a vector voltmeter or analog frequency difference meter (to measure RFS phase or frequency change), a frequency counter (perhaps the high resolution interpolating reciprocal type), or a very high resolution heterodyne arrangement (perhaps with computer control and data analysis). The frequency reference may be a crystal oscillator (for low noise), another rubidium standard (for stability, portability, and fast warmup) or a cesium standard (for absolute frequency determination). The RFS spectral purity may be observed on a rf spectrum analyzer or on a wave or FFT analyzer (after downconversion). A phase noise

Table 2. RFS Environmental Tests.

ENVIRONMENTAL TEST	CRITICAL PARAMETER	QTP	ATP	MIL-O-55310B PARAGRAPH	MIL-STD-202 METHOD	MIL-STD-810 METHOD
Acceleration/ Orientation	Frequency	0	N/R	4.9.18,.41	212	513
Acoustic Noise	Phase Noise	N/R	N/R	4.9.39		515
Altitude/Pressure	Frequency	0	N/R	4.9.45-.46	105	500
Bench Handling	Function	R	N/R			516
EMI Susceptibility	Frequency	A/R	N/R	4.9.36	MIL-STD-461	
Explosive Atmosphere	Function	N/R	N/R	4.9.42	109	511
Fungus	Function	N/R	N/R	4.9.54		508
Humidity/Moisture	Function	0	N/R	4.9.49	103	507
Immersion	Function	A/R	N/R		104	512
Magnetic Field	Frequency	R	0	4.9.43	ASTM 346-64	
Radiation	Frequency	A/R	N/R	4.9.47		
Rain	Function	0	N/R	4.9.49	106	506
Salt Fog	Function	0	N/R	4.9.50	101	509
Sand and Dust	Function	N/R	N/R		110	510
Shock, Operating	Function	0	N/R	4.9.40	213	516
Temperature, Operating	Frequency	R	R	4.9.10-12		
Temperature, Non Op.	Function	0	N/R	4.9.46		
Temperature/Altitude	Frequency	0	N/R			504
Temperature Cycling	Function	R	N/R		102	
Temperature Shock	Function	0	0	4.9.44	107	503
Vibration, Sine	Frequency	R	0	4.9.18-.38	201	514
Vibration, Random	Phase Noise	R	0	4.9.18-.38	214	514

R=Recommended 0=Optional A/R=As Required N/R=Not Required

measuring system may also be necessary (probably using a double-balanced mixer and phase-locked crystal oscillator). The primary concern for this instrumentation is often its portability to an outside testing lab. In such a case, careful planning is critical and all the test instrumentation should be set up "at home" to confirm the test methodology and to avoid missing or malfunctioning items in the field.

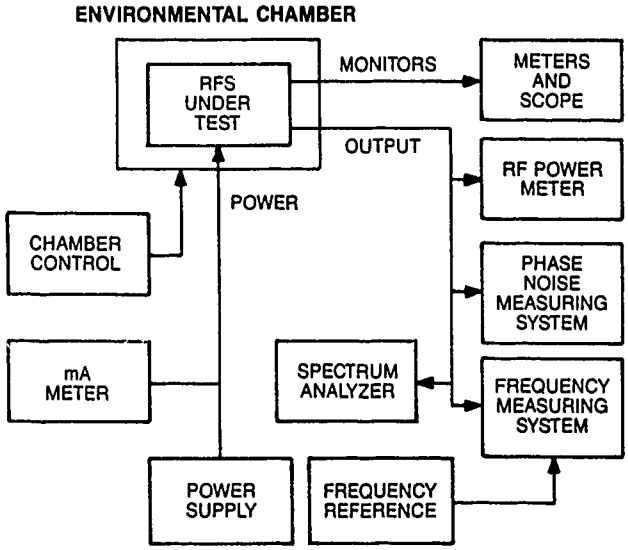


Figure 1. Typical RFS Environmental Test Setup.

Acceleration/Orientation. Unlike a quartz crystal, a RFS does not have any inherent static acceleration sensitivity (either "tip-over effect" or centrifuge). It is possible, however, for a RFS to show some frequency change due to acceleration and reorientation because of thermal effects. A test for RFS orientation sensitivity should therefore allow sufficient time (about 5 minutes) to reestablish thermal equilibrium. It is also possible to have some change in the physical distribution of the rubidium in the lamp under high static g forces that can cause a small frequency change. This effect will occur if the acceleration forces molten rubidium into the body of the lamp, and will be reversed by acceleration in the opposite direction or by quiescent operation. It is possible (but not always easy) to make the precision frequency measurements necessary to see these effects via the slip rings of a centrifuge. Redundant slip rings should be used for both power and RFS output to reduce contact noise.

A separate "2g tip-over" test may be required for the crystal oscillator portion of a RbXO. The test results should be reported as the resultant of the g-sensitivity vector.

Dynamic acceleration can have a profound effect on the stability and purity of a RFS, and those tests are discussed in the Vibration section below.

Acoustic Noise. Acoustic noise is usually not a significant factor affecting RFS operation. The RFS is typically suspended on elastic cords during exposure; frequency and phase noise should be measured.

Altitude/Barometric Pressure. The main considerations for the sensitivity of a RFS to ambient pressure are (a) frequency shift due to absorption cell buffer gas pressure change caused by deflection of the glass cell envelope (about $1x10^{-10}$/atm), (b) frequency shift due to thermal effects involving the physics package ovens and Rb lamp, and (c) thermal effects due to heat transfer of the electronic circuits. Effect (a) is proportional to the pressure change; effects (b) and (c) are small until the pressure is reduced to about 10 Torr (100,000 feet altitude). Oven power will be less, and stabilization factor better in vacuum. The electronic circuits must be designed for adequate heat transfer without air, and temperature profiling may be necessary to ensure reliable operation in vacuum. The barometric sensitivity of a RFS can also be measured as part of a temperature-altitude test. Unlike a cesium frequency standard, a RFS does not use high voltages and can be safely operated throughout the full pressure range from sea level to hard vacuum without any corona discharge hazard.

Bench Handling. A bench handling shock test is perhaps the most severe yet realistic and practical way to check a RFS for ruggedness. The unit should be removed from its enclosure as if it were being serviced. After completing the test, the unit should be inspected for damage and operated to show any resulting frequency offset.

EMI Susceptibility. The most significant form of EMI susceptibility in a RFS is usually due to ripple (CS01) and transients (CS06) on the dc input power. This aspect of RFS design is covered by military and other standards [4,11,15], but specific requirements may also be imposed. Ripple susceptibility is generally worst at the RFS servo modulation rate; large frequency offsets are possible due to interference with the servo that locks the VCXO to the Rb reference. A phase comparator, vector voltmeter, or analog frequency difference meter is very useful for showing RFS EMI susceptibility, as is observation of the RFS recovered signal on an oscilloscope. Ripple susceptibility testing does not actually require a screen room or specialized EMI test instrumentation, and can be investigated with ordinary laboratory equipment. The EMI susceptibility test procedure must include a pass/fail criterion; a frequency change of $1x10^{-10}$ is a reasonable value for a tactical RFS. The radiated susceptibility test setup should pay particular attention to the shielding and grounding of the RFS power/ monitor leads; this is often the primary path for interfering rf into the unit. It is also advisable to use a linear (rather than switching) dc supply to power the RFS during all EMI testing.

The requirements for transient protection [16-18] vary depending on the degree of external power conditioning. Reverse and overvoltage protection is often necessary and must be tested. Applying a pre-charged capacitor across the dc input is one simple way to simulate voltage transients. RFS turn-on (inrush current) and turn-off (voltage spike) transients can be a problem for the host system. A manually tipped mercury relay and a digital storage oscilloscope is an effective means to simulate and observe these transients.

Explosive Atmosphere. An RFS does not use high voltages or mechanical relays and switches, and can be operated in an explosive atmosphere with minimal likelihood of causing an explosion. Since a RFS is not normally sealed, the outer cover should be removed or loosened during the test to facilitate the penetration of the explosive vapor.

Fungus. A fungus test is usually conducted "by analysis" to verify that the design uses only components and materials that are inherently fungus inert.

Humidity/Moisture. A typical humidity test consists of five 48-hour temperature-humidity cycles with RFS operation near the end of each cycle. With an unsealed unit, the test results depend strongly on the adequacy of conformal coating and encapsulating processes, and

these should closely represent the actual production processes for the test results to be valid. RFS moisture sensitivity is most often associated with the high impedance servo amplifier synchronous detector/integrator circuits.

Immersion. The capability to withstand immersion in water is seldom a requirement for a RFS. Immersibility requires a sealed case and connectors, but it otherwise does not affect RFS design or performance.

Intermittent Operation/Retrace. (See Power/Temperature Cycling below.) A test for the frequency retrace of a RFS under intermittent operation should, by definition, return the unit to exactly the same operating conditions so as not to confuse retrace with environmental sensitivity. The main concern is that the retrace error be non-accumulative. The test must define (a) operating conditions (temperature, etc.), (b) off time, (c) restabilization time, and (d) number of retrace cycles. A well-designed RFS should have an excellent retrace characteristic ($pp10^{11}$) that has little dependency on temperature, off time, or restabilization time. The test emphasis should therefore be on performing enough cycles to verify that the retrace error does not accumulate.

Magnetic Field. A RFS (unlike a XO) has an inherent sensitivity to dc magnetic field, and contains magnetic shielding to reduce this sensitivity. The largest magnetic sensitivity is along the physics package optical axis (the direction of the internal bias field), and has a quadratic dependence of frequency on dc magnetic field ($\Delta f/f = 8.38 \times 10^{-8}$ H,Gauss^2). Since the magnetic sensitivity varies with C-field frequency adjustment, the test procedure must specify the RFS frequency setting (minimum, nominal, upper range, or maximum); nominal is usually used.

The test procedure must also define the magnetic field environment. The electrical reversal of a 3 Gauss field from a pair of 4' Helmholtz coils [19] is recommended. This field is easy to produce yet considerably larger than the Earth's field (about 0.5 Gauss), and field reversal tends to eliminate offsets due to the Earth's field (a shielded room is impractical). Other considerations for the test field are (a) uniformity (coils much larger than RFS, away from metallic objects), (b) calibration (measure field at center w/o RFS), and (c) magnitude (affects shielding factor and measurement resolution). Care should be exercised when deenergizing the Helmholtz coils; unplugging them can produce a hazardous high voltage.

Magnetic sensitivity is generally a qualification test, but variations in C-field and/or shield permeability and fit can require 100% testing for critical applications. AC magnetic field sensitivity is generally part of EMI susceptibility testing.

Radiation. The radiation sensitivity of a RFS is essentially that of its electronic circuits; the Rb physics package is inherently quite hard. RFS survivability can be a critical requirement for both transient and total dose radiation environments. Transient radiation testing usually requires a Flash X-Ray (FXR) facility to generate an intense gamma pulse. The RFS is operated during the exposure and may be required to "operate through" or to quickly recover frequency accuracy; in all cases it must not suffer latchup, burnout, or other permanent degradation. The test setup generally includes extensive recording of RFS output level, phase, frequency, light monitor, VCXO control voltage, and dc input current (as well as radiation dosimetry). Transient radiation testing is very desirable to verify analysis of the RFS circuits, and is meaningful even for a single sample having generic parts. The most critical parts for latchup are usually CMOS devices; all circuits may require resistors or other means for current limiting.

Total dose radiation testing requires a different methodology. Analysis (based on piece part test data) is often preferable to a test of the entire RFS because a worst-case error budget can be done. The most critical devices are usually servo amplifier and temperature controller op amps and the C-field voltage reference. RFS testing, if done to confirm the analysis, must be done with hardened parts. Total dose testing is likely to damage the unit as it is exposed to a series of successively higher radiation levels. The test setup is usually quite simple if the unit does not have to be operated during the exposure; before and after frequency (and perhaps monitor) measurements are generally all that is required. Total gamma dose is usually done by placing the unit in a Co^{60} cell. Dose rate and annealing effects should be considered. If the unit is operating during exposure, the RFS photodetector will respond to the gamma irradiation. Neutron exposure is usually done by placing the unit in a reactor. The unit may become radioactive because of neutron activation of the magnetic shields and other high-Z materials. The most critical part for neutron fluence is usually the silicon photodetector, which loses output due to lattice damage.

No standard procedures apply to RFS radiation testing; each case requires careful planning and expert advice.

Rain. The test procedure should specify which sides of the RFS are to be exposed to rain. Typically, waterproof cables are attached throughout the test, and the unit is operated near the end of each exposure.

Salt Fog. Since the RFS is typically not operated during exposure, its connectors should be covered. The unit should not be washed after exposure (as allowed by MIL-STD-810). EMI testing should follow salt fog exposure since salt deposits and corrosion could affect grounding and shielding.

Sand and Dust. The test procedure should specify the RFS face to be exposed. Since the RFS is nonoperating during exposure, connectors should be covered. Penetration of dust should not necessarily be considered a failure.

Shock. The RFS under test should be powered to best show intermittent failures under shock. Test cables should have strain relief and should be supported close to the unit to prevent whipping and damage which could affect RFS measurements. The RFS should be observed for timing error and permanent frequency offset. Rubidium motion in the lamp (as described under Acceleration) may occur.

Temperature, Operating. Temperature sensitivity is often the most significant environmental factor affecting the performance of a RFS. A stability of 3×10^{-10} is typical for a small tactical RFS over a military temperature range, whereas the unit will not have that much frequency aging over several years. Furthermore, there is considerable unit-to-unit variation of this important parameter, which is not necessarily monotonic and which may have regions of high incremental sensitivity. The RFS temperature stability is usually specified as the maximum (peak-to-peak) variation in frequency anywhere within the operating temperature range. RFS operating temperature range is usually specified from a minimum ambient value to a maximum baseplate value. The distinction between ambient and baseplate temperatures is necessary (because of the RFS internal dissipation) unless the unit is attached to a large heatsink or is otherwise well coupled to the

ambient. It is sometimes necessary to specify the thermal resistance of the heat sink (from ambient air to RFS baseplate).

RFS temperature stability should be measured as part of the ATP for every unit. It is best to make continuous frequency measurements on a stripchart recorder as the RFS is stepped or ramped over its full operating temperature range (rather than just measure the RFS frequency at a few discrete temperatures) so that a region of large sensitivity or anomalous behavior is detected. An averaging time should be chosen long enough to average the measurement noise while not masking rapid changes; 100 seconds is usually a good choice. Sufficient dwell time must be allowed, especially at the temperature extremes, and the record should be examined for noise, "glitches," and retrace.

If the RFS uses C-field temperature compensation, it may exhibit disparate frequency-temperature characteristics at different frequency adjustments, similar to the XO "trim effect". (EG&G does not use this method.)

Temperature, Non-Operating. Sufficient soak time must be used to ensure that RFS internal temperatures reach the specified limits.

Temperature Cycling. (See Power/Temperature Cycling below).

Temperature Shock. Temperature shock testing is not as effective with a RFS as with small components because of the slow internal temperature response of the ovens. The severity of the test is indicated more by the internal temperature extremes than the rate of change of ambient temperature, which may be expensive to produce.

Temperature/Altitude. The Temperature/Altitude test has been deleted from the latest revision of MIL-STD-810. It is better to measure the effects of pressure (altitude) and temperature separately.

Vibration. The stability and purity of a RFS are affected by mechanical vibration primarily because of the acceleration sensitivity of the quartz crystal used in the VCXO that produces the output signal. Direct vibrational modulation of the crystal oscillator generally affects the RFS phase noise and spectral purity at vibration frequencies higher than the servo bandwidth without producing a frequency offset. Vibrational modulation of the VCXO at the 2nd harmonic of the servo modulation rate, however, can cause a frequency offset. Low frequency vibrational modulation of the crystal oscillator can cause a frequency offset due to loss of microwave power. These XO effects are reduced by a high modulation rate, a wide servo bandwidth, and a low crystal g-sensitivity.

RFS stability can also be affected by vibrational modulation of the physics package light beam at or near the servo modulation rate. This problem is reduced by rigid physics package construction. Circuit board and wiring microphonics can also affect RFS stability.

RFS vibration testing requires measurement of frequency offset, frequency stability (Allan variance) and phase noise under dynamic conditions while the unit is on a shaker. It should also include before and after measurements to check for a permanent frequency offset. Phase noise is the primary parameter of concern during random vibration, and frequency stability is the most critical factor under sine vibration. Particular attention should be paid to frequency stability while the RFS is vibrated at the servo modulation rate and its harmonics. An analog frequency difference meter is an effective measurement tool; counter measurements are subject to beats due to coherence between the gate time and vibration frequency. The RFS recovered signal should also be observed, as should the lock detector output. Checks should be made for interference from the magnetic field of an electrodynamic shaker (by holding the unit slightly above the top of the shaker). A slip table can help by allowing separation between the shaker and the RFS under test. Ground insulation (rigid) or an isolation transformer in the RFS output can also help to avoid hum and interference; measurements should not be made at 60 Hz harmonics.

POWER/TEMPERATURE CYCLING

RFS units intended for frequent on-off and temperature cycling (in particular, the Rb reference of a RbXO[20]) should be subjected to a Design Verification Test (DVT) to prove their endurance under such cycling. The physics package of a fast-warmup RFS turned on after a cold soak experiences severe thermal stresses; nevertheless, with proper design, a unit can survive such cycling and provide a stability comparable to a unit that operates continuously. A typical DVT would subject 4 RFS units to 20 on-off cycles per day while subjecting them to a -62 to +68°C temperature cycle each day for 180 days. The daily temperature cycles include a cold soak at the low temperature extreme. The internal temperature of the Rb physics package should decay essentially all the way to the air temperature between turn-ons. The 3600 on-off cycles and 180 temperature cycles of the DVT simulate 20 years of normal RFS operation.

The DVT should also include frequency measurement data after each turn-on cycle. A plot of RFS frequency vs. temperature should show a consistent pattern for each daily temperature cycle, similar to the static temperature stability characteristic. A plot of average daily frequency vs. time should show a retrace characteristic trend similar to the normal RFS aging behavior.

A DVT such as this should be performed on any basic Rb physics package design that is intended for an application that involves frequent on-off and temperature cycling. EG&G has also, as part of our RFS development process, subjected a Rb physics package to over 7500 fast warmup turn-on cycles over a period of 25 months without any sign of wearout due to thermal fatigue stress. A cold soak and turn-on is also an effective means of environmental stress screening for production RFS units as discussed below.

ENVIRONMENTAL STRESS SCREENING

The purpose of Environmental Stress Screening (ESS)[21,22] is different from other environmental tests. The latter are intended to verify the adequacy of the RFS design, construction, and performance under the anticipated environments. The purpose of ESS is to reduce the number of service failures by applying temperature and vibration screens to each production unit to precipitate latent manufacturing defects as detectable failures.

To devise a suitable ESS test, vibration and temperature surveys are performed to arrive at tailored levels which will show flaws without damaging a sound RFS unit.[23, 24] A typical tailored vibration screen is random vibration of 5 grms (0.028 g^2/Hz from 80 to 350 Hz, decreasing at 3 dB/octave to 20 Hz and 2000 Hz), applied for 10 minutes in each of three axes with the RFS unit operating and the input current and lock monitor signal recorded. Wideband random vibration is far more effective than sine vibration because the random excites all vibrational modes simultaneously and for the full duration of the exposure. Single axis excitation is usually selected because of shaker availability.

A typical tailored temperature screen consists of 12 temperature cycles with power applied after a cold soak of 1 hour. The cold soak is necessary to adequately stress the physics package ovens and temperature control circuitry in terms of temperature rates and excursions. The cold soak duration is determined by the thermal time constants of the physics package ovens. The effectiveness of temperature cycling in precipitating and detecting failures is greatly increased by operating and monitoring the test unit, and by recording the input current and the lock monitor signal. In particular, a cold soak and turn-on is much more effective than exposing a fast-warmup RFS to an external temperature shock. A temperature rate of 5 C/minute is generally selected (rather than a faster rate which requires a higher performance test chamber) since the internal response of the RFS is not significantly different for the slower rate.

Our experience has shown that a properly chosen ESS test is a cost-effective way to find latent defects and thereby ensure a lower failure rate in the field. An important part of the ESS process is to review the test records to refine both the RFS manufacturing processes and the effectiveness of the ESS test.

CONCLUSION

Environmental performance is critical for many applications of rubidium frequency standards. No standards currently exist for the testing of atomic frequency standards (such as those of Section 4.9 of MIL-O-55310B for crystal oscillators). Standardized environmental test methods would be helpful for both the vendors and users of such devices. Such standardization would ease device specification, eliminate test duplication, simplify test plan preparation, clarify test results, and lead to improved environmental performance for these devices. It is hoped that the exchange of information like this paper will lead to the adoption of such a document.

REFERENCES

1. T. Lynch and W. Riley, "Tactical Rubidium Frequency Standard (TRFS)," RADC-TR-87-166, Final Technical Report, Oct. 1987.
2. MIL-STD-454, " Standard General Requirements For Electronic Equipment."
3. MIL-STD-810, "Environmental Test Methods."
4. MIL-STD-461, "Electromagnetic Interference Characteristics, Requirements For Equipment."
5. MIL-E-5400, "Electronic Equipment, Airborne, General Specification For."
6. MIL-T-5422, "Testing, Environmental, Airborne Electronic And Associated Equipment."
7. MIL-E-16400, "Electronic Equipment, Naval Ship And Shore, General Specification."
8. MIL-E-4158, " Electronic Equipment, Ground, General Requirements For."
9. DOD-E-8983, "Electronic Equipment, Aerospace, Extended Space Environment, General Specification For."
10. MIL-STD-1540, "Test Requirements For Space Vehicles."
11. MIL-STD-1541, "Electromagnetic Compatibility Requirements For Space Systems."
12. MIL-C-45662, "Calibration System Requirements."
13. MIL-O-55310B, "Military Specification, Oscillators, Crystal, General Specification For," 10 May 1988.
14. MIL-STD-202, "Test Methods For Electronic and Electrical Component Parts."
15. MIL-STD-462, "Electromagnetic Interference Characteristics, Measurement Of."
16. MIL-STD-704, "Platform Electrical Power Characteristics."
17. MIL-STD-1275, "Characteristics Of 28 VDC Electrical System In Military Vehicles."
18. ARNIC Report 413, "Guidance For Aircraft And Electrical Power Utilization And Transient Protection," Aeronautical Radio, Inc., Dec. 1976.
19. ASTM 346-64, "Standard Methods Of Test Of Magnetic Shielding," American Society For Testing And Materials.
20. W. Riley and J. Vaccaro, "Rubidium-Crystal Oscillator (RbXO) Development Program," U.S. Army LABCOM R&D Technical Report SLCET-TR-84-0410F, Apr. 1986.
21. MIL-STD-2164(EC), "Environmental Stress Screening Process For Electronic Equipment."
22. NAVMAT P-9492, "Navy Manufacturing Screening Program."
23. A.J. Curtis and R.D. McKain, "A Quantitative Method Of Tailoring Input Spectra For Random Vibration Screens," J. Env. Sci., Sept./Oct. 1987.
24. R.G. Lambert, "Case Histories Of Selection Criteria For Random Vibration Screening," Inst. Env. Sci. Workshop Proc., Sept. 1984.

43rd Annual Symposium on Frequency Control - 1989

TESTING AND SPECIFICATION OF ENVIRONMENTAL SENSITIVITIES IN CESIUM AND PRECISION QUARTZ SIGNAL SOURCES

R.Michael Garvey
Frequency and Time Systems, Inc.
Beverly, MA 01915

ABSTRACT

The stability of a precision timing signal source under the influence of environmental perturbation is a key performance parameter. The testing of this performance is frequently no less important from a system design perspective: the user must define the tests to reflect the actual environment and the designer must optimize the equipment to provide maximum performance margin. Frequently a performance error budget can be a useful tool to define clock behavior under environmental influence.

Standard environmental test methods are generally not suitable for evaluating precision signal source performance. What is needed is a defined framework for environmental test methods which can be tailored as required to meet program test needs.

INTRODUCTION

Environmental testing may, in general, be divided into several types, depending on the maturity of the product design. These types are: development, qualification, and acceptance. Development testing is by its nature informal and somewhat ad hoc, depending upon the design history and the performance desired. Qualification and acceptance testing are formal structured procedures using standardized test methods. Rarely can standard environmental test methods be applied directly without some tailoring to accommodate unique system or product requirements. Military standards form a useful baseline from which to develop these tailored test methods.

The starting point in any product specification is usually a statement of the required performance, a topic which cannot be adequately addressed without considerations of testing. Typically, environmental sensitivities should be addressed one parameter at a time; this ultimately minimizes the design cost as well as the test cost. Occasionally, the required performance is defined as bounded accuracy or stability requirements over a combination of environments. In this case, an error budget is often an effective way to allocate the individual environmental performance requirements. Within the error budget, decisions can be made on how to sum the individual items. The error budget can also usefully handle other (performance) parameters such as design and test margins, specification guardbands, etc.

Specific test techniques have their own considerations. In some cases, standard apparatus and commonly used techniques are available. In other cases it is difficult to isolate and control a single environmental parameter in the test process. A number of pitfalls have to be anticipated.

QUALIFICATION TESTING

Qualification (Qual) testing serves to verify the design; it is usually performed on one of the first articles produced after the design documentation is complete. Qual testing may include extended testing to assess long term performance or device life. It may include exposure to environmental extremes to determine design margin. Examples include higher vibration levels or greater temperature extremes. Qual testing may involve costly or destructive tests to verify performance under exposure or to verify survival and ability to perform after exposure. Safety issues of the design are typically addressed during qual testing.

ACCEPTANCE TESTING

Acceptance testing serves to verify the production process; included are a selection of performance and functional verifications. Specifically what to test is a complex question involving issues of design margin, product complexity and product maturity. The performance verification process seeks to determine, in summary fashion, that all elements of the design are working as intended. In quartz oscillators, critical indicators of performance include, typically, temperature sensitivity, phase noise and aging. In cesium devices temperature sensitivity and Allan variance are critical indicators of performance. The functional verifications determine that, for example, all outputs and instrument controls are functional. Acceptance testing also may include workmanship verification through environmental stress screening. Exposure to random vibration in accordance with MIL-STD-2164[1] is typically used; MIL-STD-2164 also defines a temperature cycling profile for workmanship screening.

PERFORMANCE CHARACTERIZATION

In the case of cesium and quartz instruments, characterization of performance begins with a determination of the frequency offset being produced. Acquisition of this data record has two goals: to provide verification of performance or in the event of a failure, to provide insight into failure mechanisms and to events leading up to the failure.

Automation of data acquisition has many advantages; it has several pitfalls as well, particularly when the observed behavior differs significantly from the norm. These pitfalls must be anticipated and accommodated. Typical areas of concern include acquisition aliasing and windowing, inadequate measurement or recording bandwidth or too infrequent measurement intervals. No general, best measurement system exits, of course; each system must balance cost and complexity against measurement requirements.

CH2690-6/89/0000-263 $1.00 C 1989 IEEE

Data record length is always a troublesome concern. Figure 1 shows a record which demonstrates a pathological behavior that would not have been detected in a truncated measurement. Frequently, historical experience is the best guide to use in determining the proper data record length.

The characterization of cesium performance is usually derived from a measurement of phase at 5 or 10 MHz. For computer data reduction, the RF signal is mixed down and the phase extracted as a time interval measurement. A linear phase detector driving an analog chart recorder can also be a valuable measurement tool particularly where a continuous phase record is desired.

In the characterization of quartz performance, a measurement of frequency against an offset reference is used. For computer data reduction, measurement of beat period is effective. Computer modeling of performance is employed when measuring temperature sensitivity. As shown in Figure 2, the oscillator aging must be modeled and extracted from the measured data in order to obtain the temperature performance. A simple graphic technique can be used when computer aided data reduction is not required. Plotting the high resolution digits of a beat period counter yields a graphic record of quartz device performance.

TEST METHODS

Standardization of test methods has many advantages: the utilization of test apparatus and techniques is enhanced; qualification by similarity and transfer of prior test results is facilitated; and tailoring of standard test methods is usually straightforward when required. In a development effort, standard test methods have the advantage that they can be defined and known to the designer and the user at project inception.

Several military standards provide useful sources of test methods. MIL-STD-810[2] is devoted to environmental test methods. Many of these methods require tailoring in order to be useful for testing of cesium and quartz sensitivities. This tailoring is straightforward, usually involving the extension of dwell times to allow adequate stabilization and measurement resolution.

Other military standards contain test methods: MIL-STD-202[3] is sometimes applied to quartz device testing. MIL-O-55310[4] addresses test methods for quartz oscillators. Certain technical groups have generated standardized test methods for specific applications. An example, for the aeronautical industry, is RTCA/DO-160[5].

ERROR BUDGET

The error budget is a useful tool to define the allocation of various environmental sensitivities to an overall system specification. The budget requires input from both the user and the designer. The user knows the environmental exposure as well as the required performance. The designer knows the intrinsic strengths and weaknesses of the technology and can allocate the budget to produce maximum design margin. The user and the designer must agree on a methodology of summing the individual elements of the budget depending on their random or deterministic nature. Note that environmental factors are sometimes correlated; for example, loss of aircraft pressurization might cause an otherwise controlled temperature environment to degrade significantly. No general methodology for addressing the factors can be given, but they can, if properly addressed, maximize design margin.

Table 1 provides a sample error budget for a cesium instrument. Note that the items summed arithmetically yield 10.7×10^{-12}; when summed in root sum square fashion the items yield 4.2×10^{-12}.

TEST TECHNIQUES

Several general considerations arise in a discussion of specific test techniques. One of these is measurement duration. Transients arising from environmental change must be allowed to decay. Following that, measurement duration must be sufficiently long to average frequency noise to an acceptable level, to allow adequate modeling of quartz aging and to acquire adequate data for the lowest frequency noise components of interest.

Another general consideration addresses transient versus steady state sensitivities. Unless otherwise specified, sensitivities are assumed to be steady state.

The testing of temperature sensitivity can be performed in standard environmental chambers manufactured for this purpose. The measurement duration is driven strongly by thermal time constants which are, for oscillators, approximately 1/2 to 1 hour and for cesium devices, approximately 1 to 2 hours. Air movement is typically implied. Humidity effects, if uncontrolled during temperature testing, can seriously contaminate the measurement. Figure 3 shows an extended measurement of quartz oscillator performance during a temperature test. Note the one day frequency transient for the 20 to 50 °C change which is totally absent for the 50 to 20 °C change. Moisture which condensed in the chamber during the 20 °C period produced a large humidity transient which is absent in the 50 to 20 °C change. The oscillator design was revised to eliminate the humidity sensitivity.

Humidity testing may be performed with standard environmental chambers. Note that humidity effects may have long time constants (days).

Altitude testing may be performed in standard environmental chambers. Temperature control is necessary. The observed time constants are not dissimilar from observed thermal time constants. The magnitude of transient effects may exceed that of static effects.

Vibration testing represents a significant measurement challenge for precision quartz and cesium signal sources primarily due to the large magnetic fields associated with vibration test equipment. In cesium, transient (or permanent) magnetization of the device under test can significantly degrade frequency accuracy. In quartz oscillators, the magnetic fields can modulate oscillator phase or frequency. In general, the large ac fields can couple directly to the measurement apparatus. Great care must be exercised to insure that the sensitivity being measured is vibratory and not magnetic. Hydraulic and mechanical vibration apparatus, which would eliminate this concern, are not generally available nor do they have the capability to perform at or above 1 kHz[6].

Shock and seismic testing are usually performed to demonstrate survival. If electromagnetic apparatus is used, the concerns regarding magnetic effects discussed above are applicable.

Magnetic field testing can be straightforwardly performed in a classical Helmholtz coil. For ac fields, it is sometimes necessary to resonate the coil inductance so that adequate excitation currents can be provided. Note that ambient magnetic fields (e.g. Earth's, building structural steel) can be significant.

Radiation test techniques are perhaps the least standardized. The techniques and levels are usually program-specific and require significant test design input. Dose rate effects in quartz oscillators produce transient as well as longer term effects. Both must be measured. In dose rate testing of cesium instruments, dose rate effects are predominantly transient with time error being the useful parameter of characterization. Total dose effects in cesium have their primary impact on frequency accuracy.

CONCLUSIONS

It is difficult to define generalized test methods. What is needed is a testing framework which can be easily tailored to the specific needs of the product and/or program. In some cases, existing military standards can serve as a starting point.

An error budget is a useful tool to obtain a clear definition of the desired performance in a given environmental scenario. The error budget should be developed early in the system definition phase in order to maximize design and performance margin.

TABLE 1: CESIUM ERROR BUDGET

TEMPERATURE	3.1×10^{-12}
MAGNETIC FIELD	
DC 25 GAUSS EXPOSURE	0.3
AC 2 GAUSS PEAK	
50 HZ	0.3
60 HZ	1.0
400 HZ	1.8
ALTITUDE	0.2
VIBRATION	1.0
HUMIDITY	1.0
SHOCK	1.0
CALIBRATION	1.0

REFERENCES

[1] MIL-STD-2164, Environmental Stress Screening for Electronic Equipment.

[2] MIL-STD-810, Environmental Test Methods and Engineering Guidelines

[3] MIL-STD-202, Test Methods for Electronic and Electrical Component Parts

[4] MIL-O-553310, General Specification for Crystal Oscillator

[5] RTCA/DO-160, Environmental Conditions and Test Procedures for Airborne Equipment; Radio Technical Commission for Aeronautics

[6] W. Tustin and R. Mercado, Random Vibration in Perspective. Santa Barbera, CA: Tustin Institute of Technology, 1984.

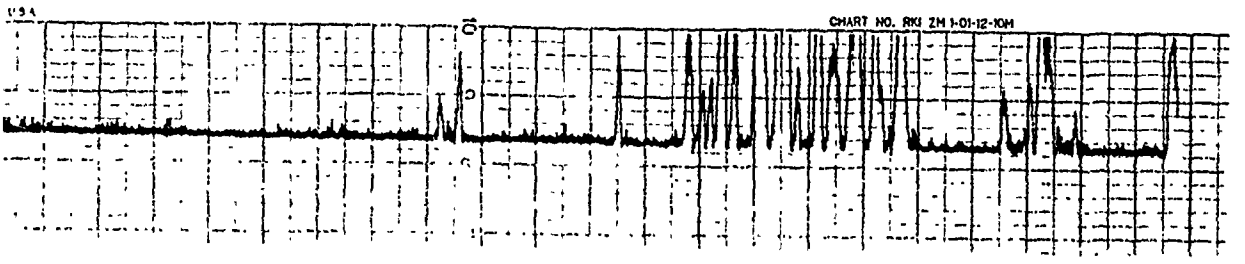


FIGURE 1: NON-PERIODIC NOISE

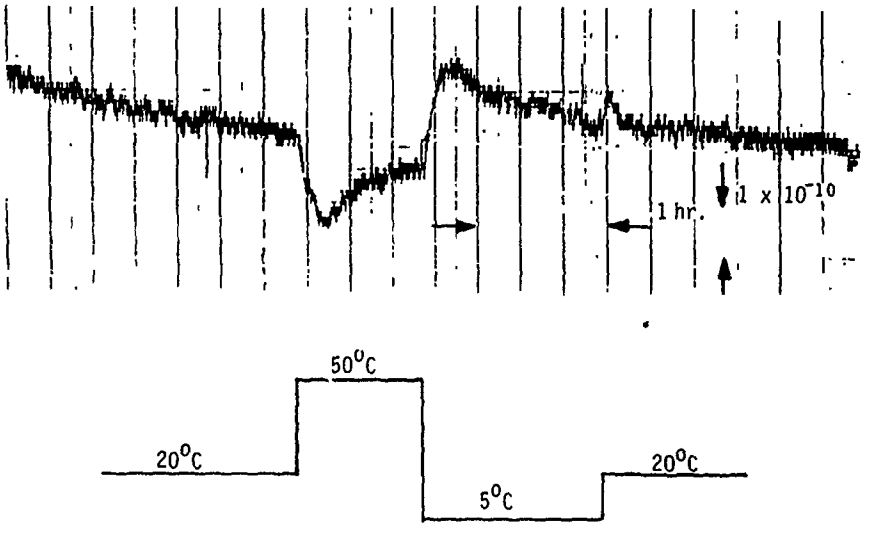


FIGURE 2: OSCILLATOR TEMPERATURE SENSITIVITY

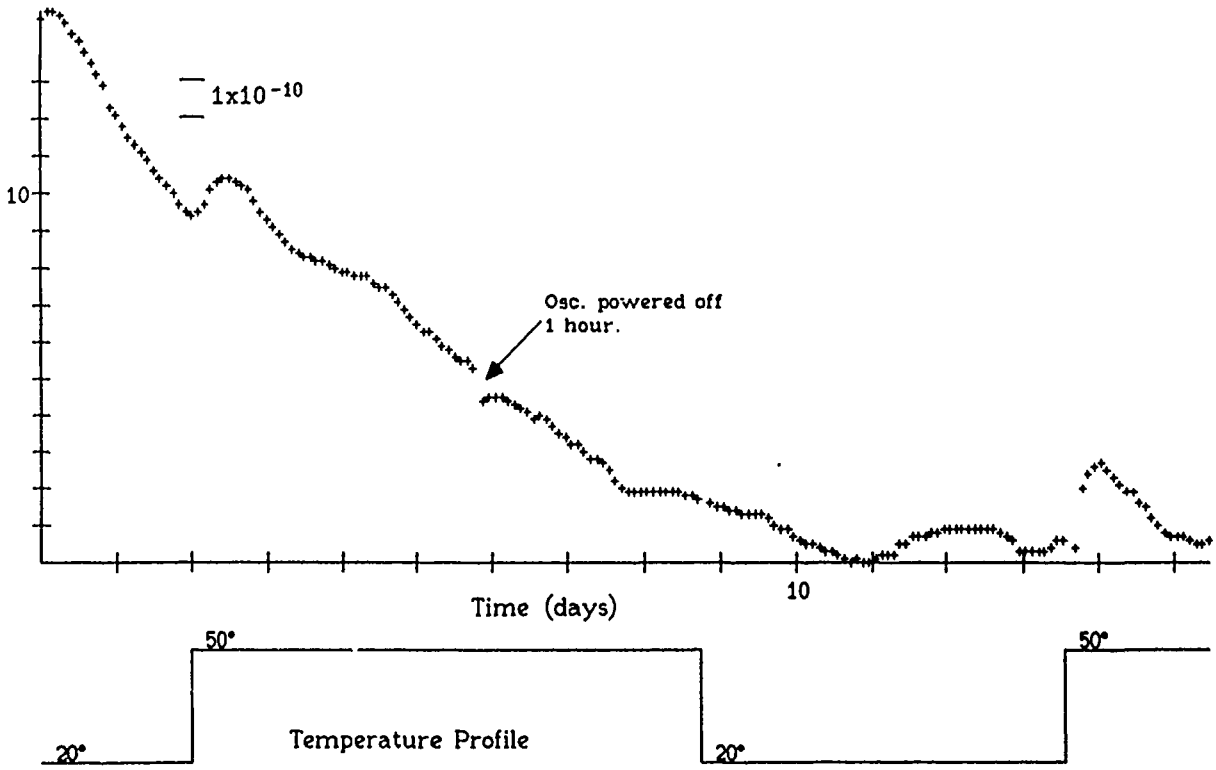


FIGURE 3: QUARTZ OSCILLATOR TEMPERATURE TEST WITH HUMIDITY EFFECTS

INFLUENCE OF MAGNETIC FIELD ON QUARTZ CRYSTAL OSCILLATORS

R. Brendel, C. El Hassani
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

M. Brunet, E. Robert
Centre National d'Etudes Spatiales
18, avenue Edouard Belin - 31055 Toulouse Cedex - France

Abstract

This paper presents recent investigations on magnetic sensitivity of certain on-board quartz crystal oscillators. A part of the study deals with the global magnetic sensitivity of the devices as a function of the magnetic field intensity and direction. Measurements were done in the Frequency and Time department of the CNES and in the laboratory of the Office National d'Etude et de Recherches Aérospatiales (ONERA) on various ovened oscillators used in actual satellites. In addition, in order to determine the origin of the magnetically induced frequency shift, some components have been separated from the oscillator and individually submitted to the magnetic field. The paper presents the experimental setup, measurement procedures and results obtained at the CNRS laboratory in Besançon. Special attention is paid to the quartz crystal resonator itself. It is shown to have a sensitivity on the order of 10^{-12}/gauss. An explanation of the resonator's magnetic sensitivity is given, together with related verification experiments.

Introduction

The increasing accuracy of positioning systems, especially for space applications such as orbitography, requires better and better stability of the on-board oscillators. As a consequence, certain fluctuations of environmental parameters such as temperature, pressure, acceleration ... whose influence on the oscillators has gone unnoticed in the past (being below the required precision) can no longer be neglected.

Interest is now being paid to the influence of magnetic field because it has been observed that starting and stopping a satellite's magnetic stabilizing equipment can induce weak frequency shifts in nearby oscillators. Although the observed fractional frequency variation (a few 10^{-11}/gauss) is acceptable for most applications, it is crippling for future orbitographic missions.

Doris specifications

Doris is a satellite-based radio positioning system whose main purposes are very high precision orbit determination and high accuracy location of ground beacons. Applications of this system mainly concern the monitoring of natural phenomena such as continental drift, sea level measurement, vulcanology or civil engineering construction like dams, tunnels, etc. Locations of beacons or orbitographic measurements are performed by using the Doppler effect. In order to achieve the required positioning accuracy, which has to be less than 10 cm, both the on-board and the ground oscillators must have a short term frequency stability better than 5.10^{-13} over an integration time from 10 s to 100 s [1]. Among the various parameters likely to affect the on-board oscillator stability, attention has been focused on the magnetic environment. In fact, the magnetic induction in satellites may reach intensities of a few 10^{-5} T (a few 1/10 gauss) [2] randomly oriented, due to magnetic stabilizing couplers and earth's magnetic field changes during low altitude flights. The required frequency stability needs an oscillator magnetic sensitivity less than 10^{-13}/gauss [3].

Magnetic sensitivity of an oscillator as a whole

Experiments done by CNES over several years [4,5] on a variety of commercially available oscillators show that their magnetic sensitivity is often much larger than the limiting value previously mentioned.

Figure 1 a, b, c shows the results obtained for an on-board oscillator in a magnetic induction of intensity lying between -5 and +5 gauss and directed along the principal axes of the oscillator as shown on Fig. 2. It can be observed in this case that the fractional frequency change is always negative for all magnetic induction orientations and senses, and that there is a least sensitive direction (here : the Y-axis). Nevertheless in any case the high magnetic sensitivity is crippling (on the order 10^{-11}/gauss).

CH2690-6/89/0000-268 $1.00 C 1989 IEEE

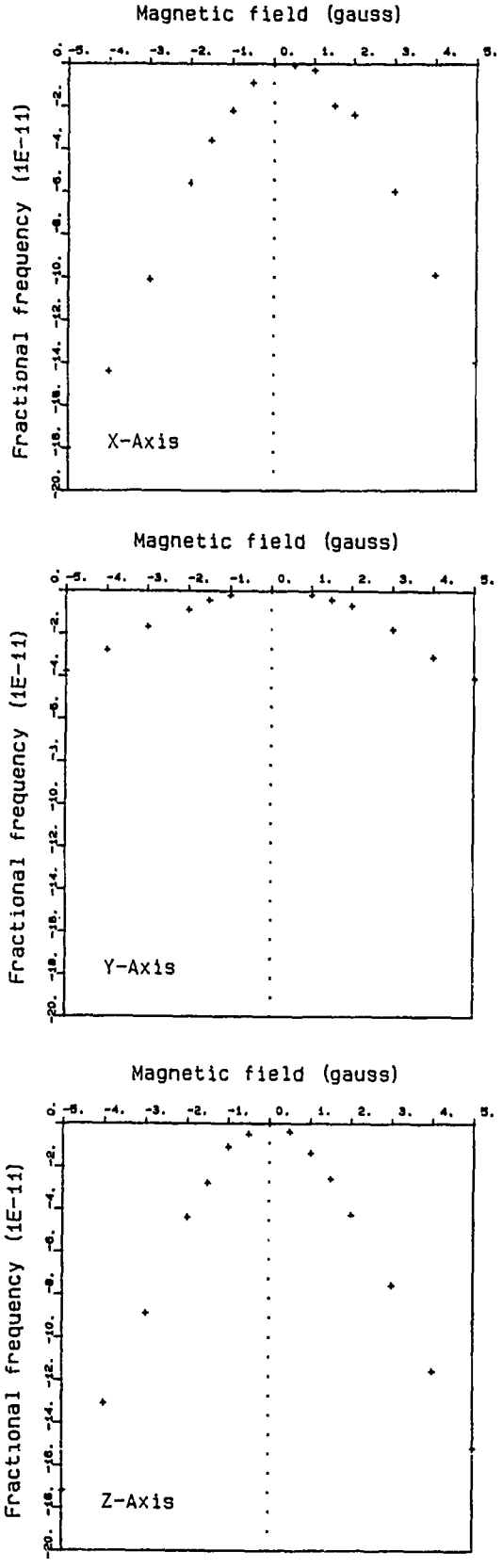


Fig. 1 : Sensitivity of an on-board oscillator in a magnetic induction directed along (a) X-axis, (b) Y-axis, (c) Z-axis (see Fig. 2)

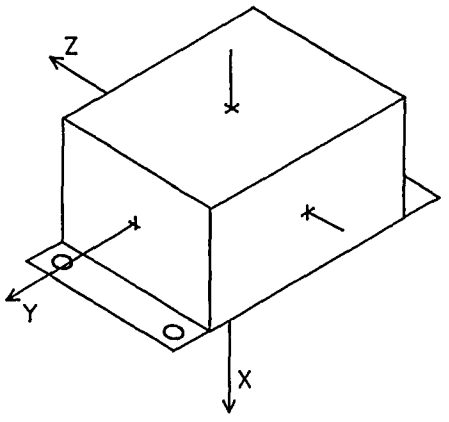


Fig. 2 : Definition of the oscillator axes

Reduction of the magnetic sensitivity

Various attempts have been made to reduce the oscillator magnetic sensitivity, the simplest being to shield the oscillator with a high magnetic permeability material. This solution leading to a reduction of the magnetic induction by about 35 to 40 dB restores the oscillator magnetic sensitivity to an acceptable level [6]. Nevertheless this improvement has other drawbacks : weight and volume increase, thermal behaviour modification, appearance of an undesirable magnetic momentum. Furthermore, the aging characteristics of such shielding materials remain incompletely understood.

Another solution consists of a modification of the oscillator's electronic design [7]. Figures 3 and 4 show the magnetic sensitivity measured before (Fig. 3) and after (Fig. 4) modification.

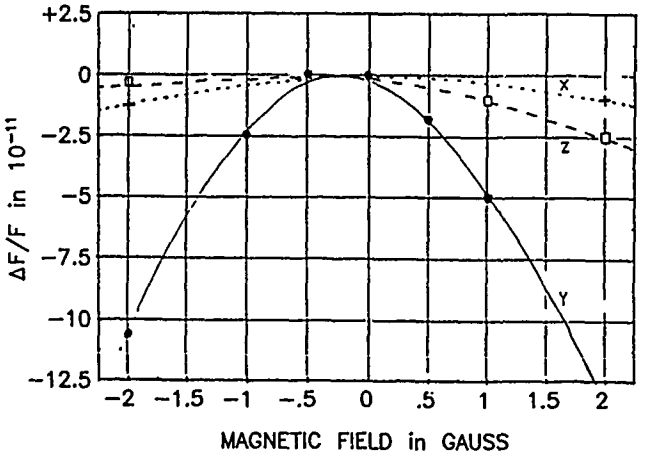


Fig. 3 : Magnetic sensitivity of an oscillator

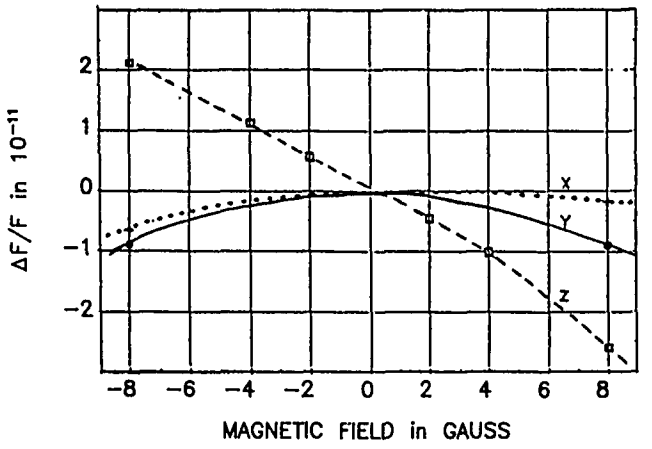


Fig. 4 : Magnetic sensitivity of the same oscillator as in Fig. 3, after electronic design modifications

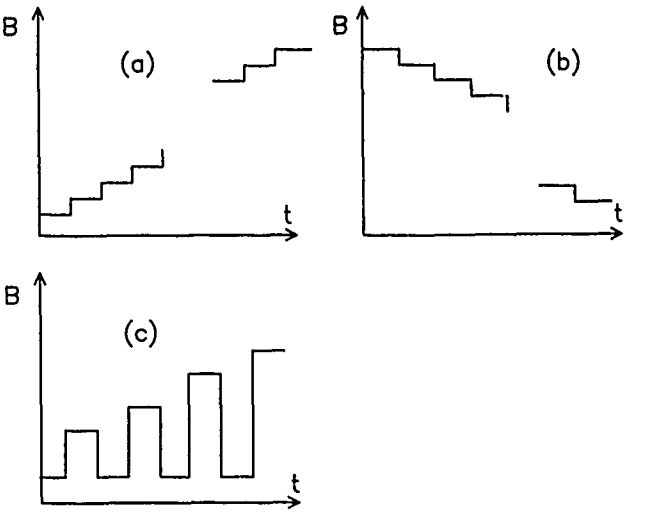


Fig. 6 : Various induction wave shapes used (a) increasing steps, (b) decreasing steps, (c) on-off stages of increasing amplitude

Experimental results obtained with strong magnetic fields

Although the solutions previously presented significantly reduce the magnetic sensitivity, the disavantage or the residual sensitivity observed in every case led us to undertake a more thorough study.

The first experiments were carried out with a 5 MHz ground oscillator used in the SARSAT-COSPAS search and rescue program. The oscillator under test was put in an electromagnet induction field whose intensity could be adjusted from a few gauss to several kgauss (see Fig. 5). The frequency variations of the oscillator were measured by means of the error signal of a locked systhesizer while the magnetic induction B was measured by a gaussmeter whose Hall effect probe was placed close to the oscillator. The two inputs of the plotter were respectively fed by the analog output of the gaussmeter (X-axis) and by the lock-input of the synthesizer (Y-axis).

Various wave shapes of magnetic induction have been used to test the oscillator sensitivity (see Fig. 6a,b,c).

Figure 7 shows that the results obtained depend essentially on the induction strength but not on the way it is applied. Also it can be seen that the fractional frequency shift, which is negative, has a mean slope of a few 10^{-12}/gauss.

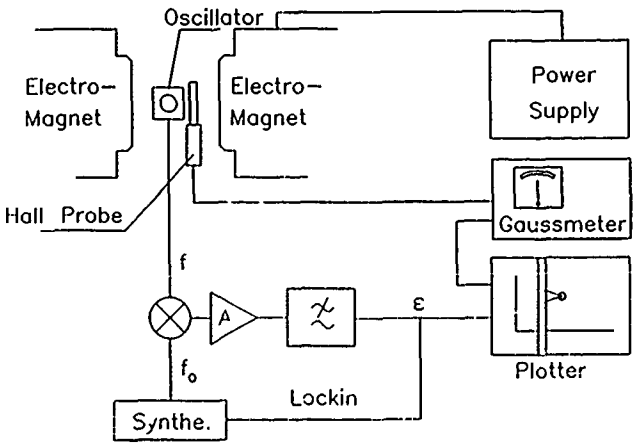
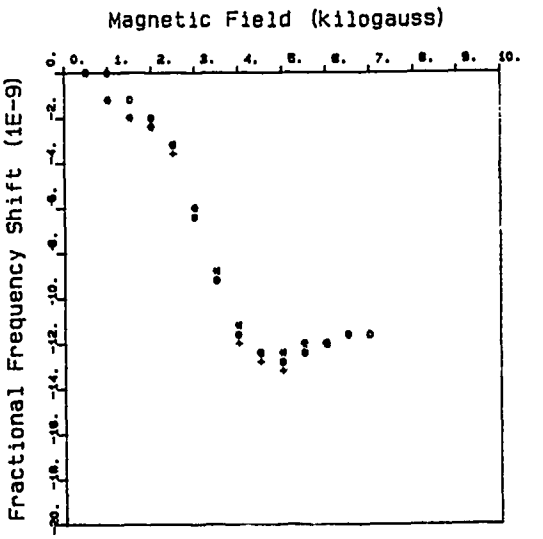


Fig. 5 : Experimental setup for oscillator sensitivity measurements in a strong magnetic field



Fig; 7 : Magnetic sensitivity of a SARSAT-COSPAS ground oscillator exposed to strong magnetic fields of different wave shapes (see Fig. 6)

Resonator magnetic sensitivity

In order to determine what are the most sensitive component(s) of the oscillator it was decided to separately submit each part of the device to the magnetic induction, beginning with the resonator.

The experimental setup used is described in Fig. 8. To limit temperature induced frequency fluctuations, the crystal and the oscillator sustaining circuit were separately ovened, the resonator crystal being maintained at its turn over point.

The main measurement features are identical with those previously described. In addition a microcomputer is used to store and process data. The electromagnet can be driven with a low frequency generator in order to get various time dependent magnetic inductions.

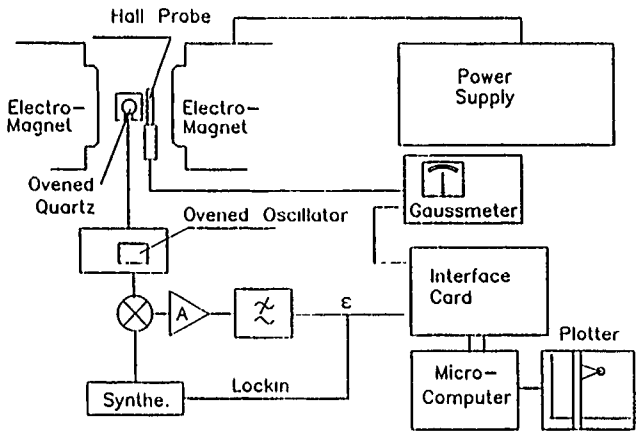


Fig. 8 : Experimental setup for resonator sensitivity measurement in a strong magnetic field

Figure 9 is a typical record obtained, showing that the frequency shift may change in sign as the induction strength increases. Here the component under test is a regular 10 MHz SC-cut silver electroded resonator.

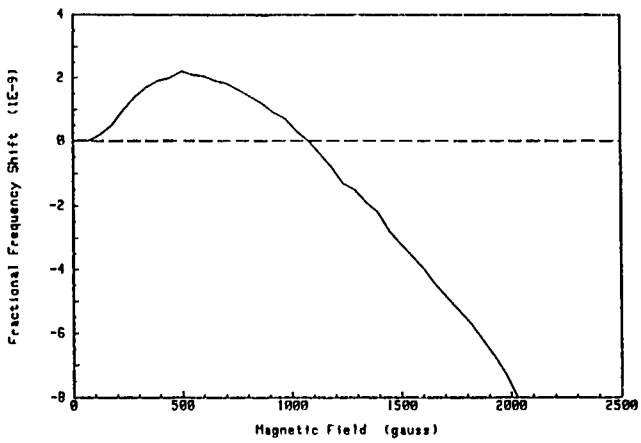


Fig. 9 : Magnetic sensitivity of a standard 10 MHz SC-cut silver electrodes resonator

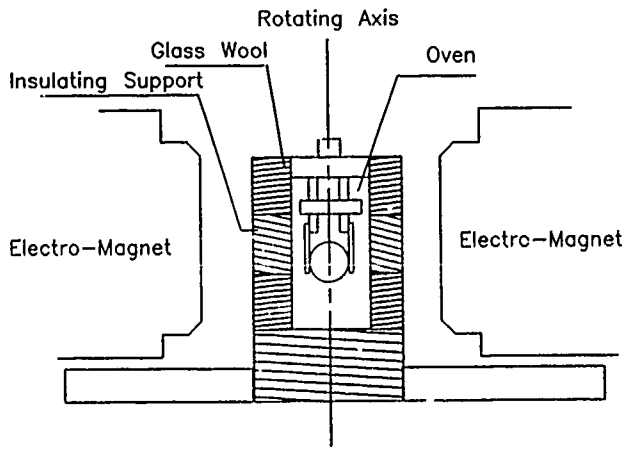
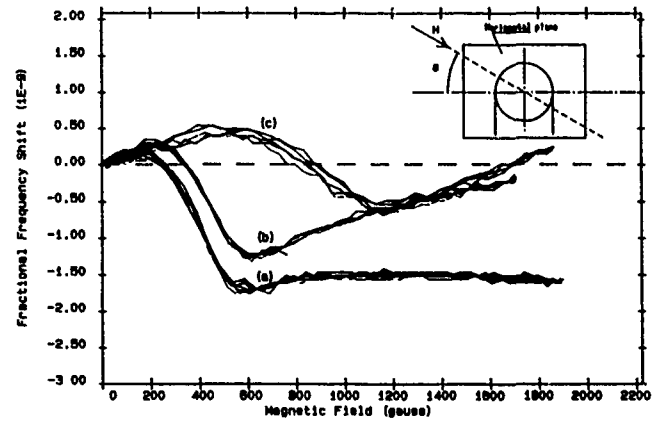
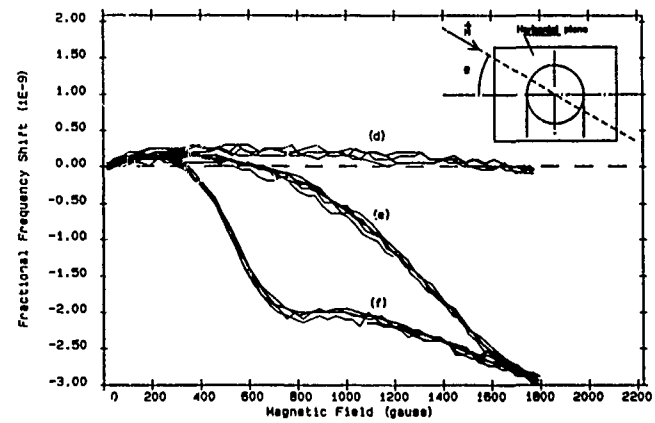
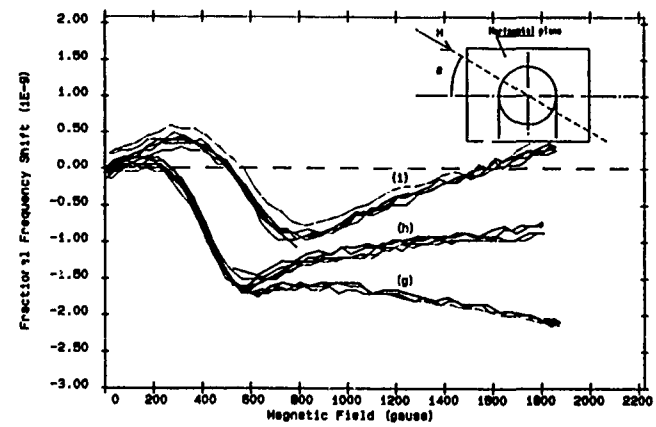


Fig. 10 : Resonator support used to study the magnetic field orientation effect

Other experiments have shown that the shape of the frequency-induction curve strongly depends on the orientation of the crystal with respect to the magnetic field. The orientation effect has been studied by mounting the quartz unit on a rotating support with its axis normal to the magnetic field (see Fig. 10). The results shown in Figs. 11, 12 and 13 were obtained for different orientations of the field with respect to the plane of the crystal.







Figs. 11, 12, 13 : Magnetic sensitivity records obtained with different induction orientations :
(a) $\theta = 0°$ (b) $\theta = 30°$ (c) $\theta = 60°$ (d) $\theta = 90°$ (e) $\theta = 100°$ (f) $\theta = 130°$ (g) $\theta = 160°$ (h) $\theta = 190°$ (i) $\theta = 220°$

Although the amplitude and shape of the frequency-induction curve depends on the orientation of the field, it doesn't depend on its sense, as shown in Fig. 14, the curve obtained is quite symmetrical with respect to the null magnetic induction axis.

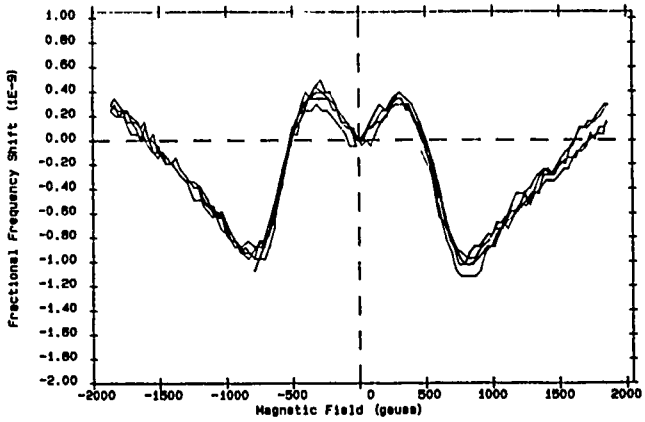


Fig. 14 : The resonator magnetic sensitivity does not depend on sense of the induction vector

Physical interpretation of the resonator magnetic sensitivity

The magnetic sensitivity of quartz crystal resonator has already been the subject of investigations in the past, and various interpretations have been given. They usually eliminate the intrinsic magnetostriction of the quartz material itself because it has a non measurable order of magnitude [8]. Other work has attempted to explain the magnetic sensitivity by the magnetostriction of the electrodes when they are made with a magnetic material such as nickel [9]. But this interpretation cannot be applied in case of the non magnetic silver electrodes used in some of the experiments presented here. Frequency variations could also be explained by an increase of the equivalent series resistance of the resonator, due to the eddy currents produced by the motion of the electrodes in the magnetic field [10], but a calculation gives an effect whose amplitude is two or three order of magnitude smaller than the observed value.

Another explanation of the magnetic sensitivity, mentioned in [8], will be reviewed and developed here. Usually, the crystal is supported by two or more thin metallic strips made of a nickel-based alloy, as shown on Fig. 15. When exposed to a magnetic field, all constitutive magnetic domains initially randomly oriented (Fig. 16a) tend to align along the same direction, corresponding to the easiest magnetization direction (Fig. 16b) [11]. When all domains reach the saturation state, their magnetization vectors $\mathbf{M_S}$ are all oriented along the same direction and are subject to a torque C tending to align them with the magnetic field H as shown on Fig. 17a. When inverting the magnetic field sense, the sense of the magnetization vector $\mathbf{M_S}$ is also inverted so that the resulting torque doesn't change (Fig. 17b).

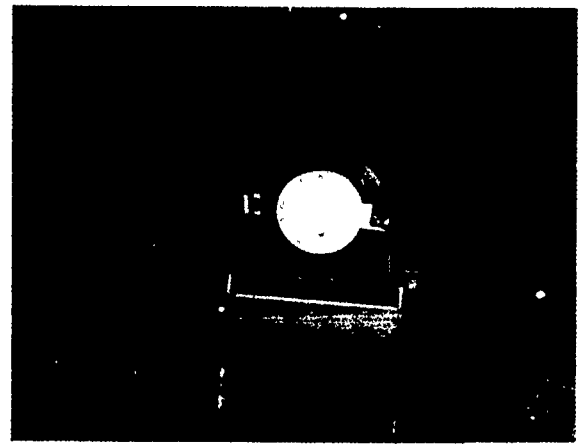


Fig. 15 : Standard quartz crystal supports

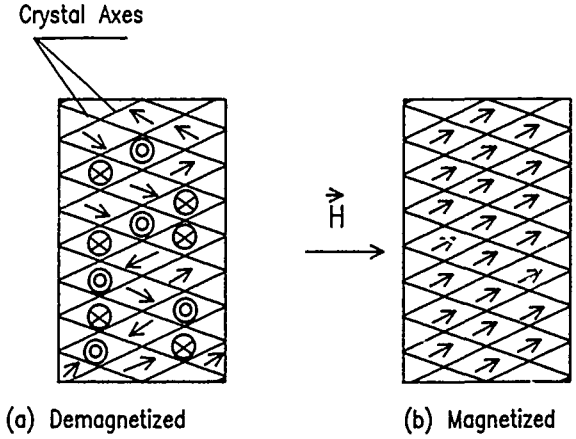


Fig. 16 : Magnetic domains ir a ferromagnetic medium (a) demagnetized state, (b) magnetized state

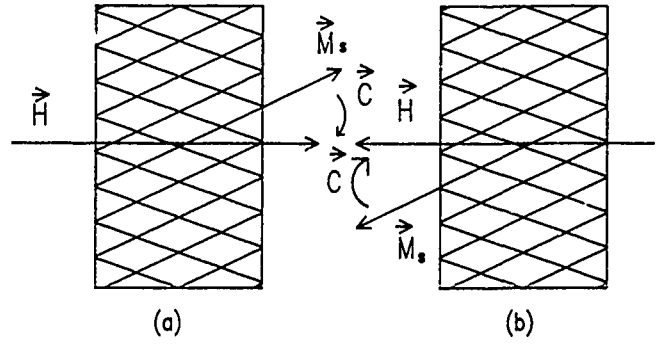


Fig. 17 : The torque C acts in the same sense for (a) direct magnetic excitation, for (b) inverse magnetic excitation

If each strip is twisted at one end, it globally experiences a general bending whose sense does not depend on the magnetic field sense (see Fig. 18). As a result, the crystal fastened to the other end of the strips is subject to a pair of plane torques of opposite signs achieving mechanical equilibrium but inducing inside the crystal a stress field which changes the resonant frequency through nonlinear effects just as for accelerometric sensitivity [12].

The calculation of the mechanical stresses induced by the bending of the support strips is very difficult because of the lack of knowledge concerning the magnetic properties of the strips. In fact, these properties depend not only on the material composition but also on the way it has been mechanically and thermally processed, etc.

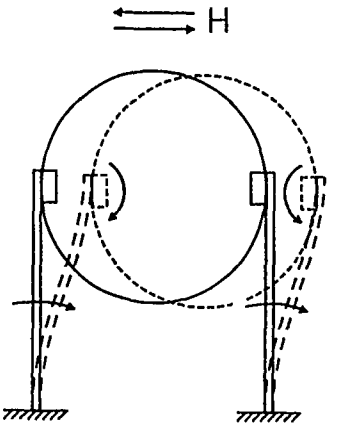


Fig. 18 : Bending of the strip in an applied magnetic excitation field induces plane torques in the crystal

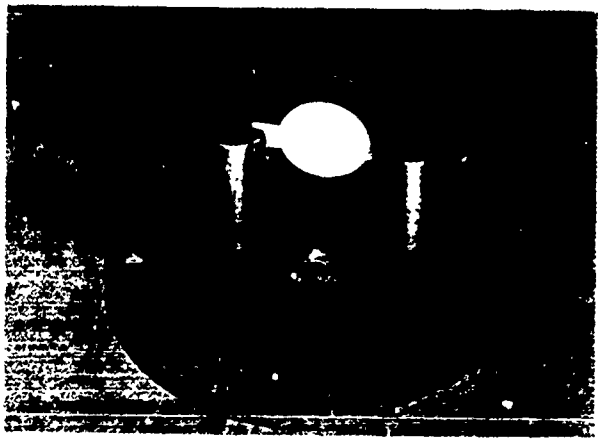
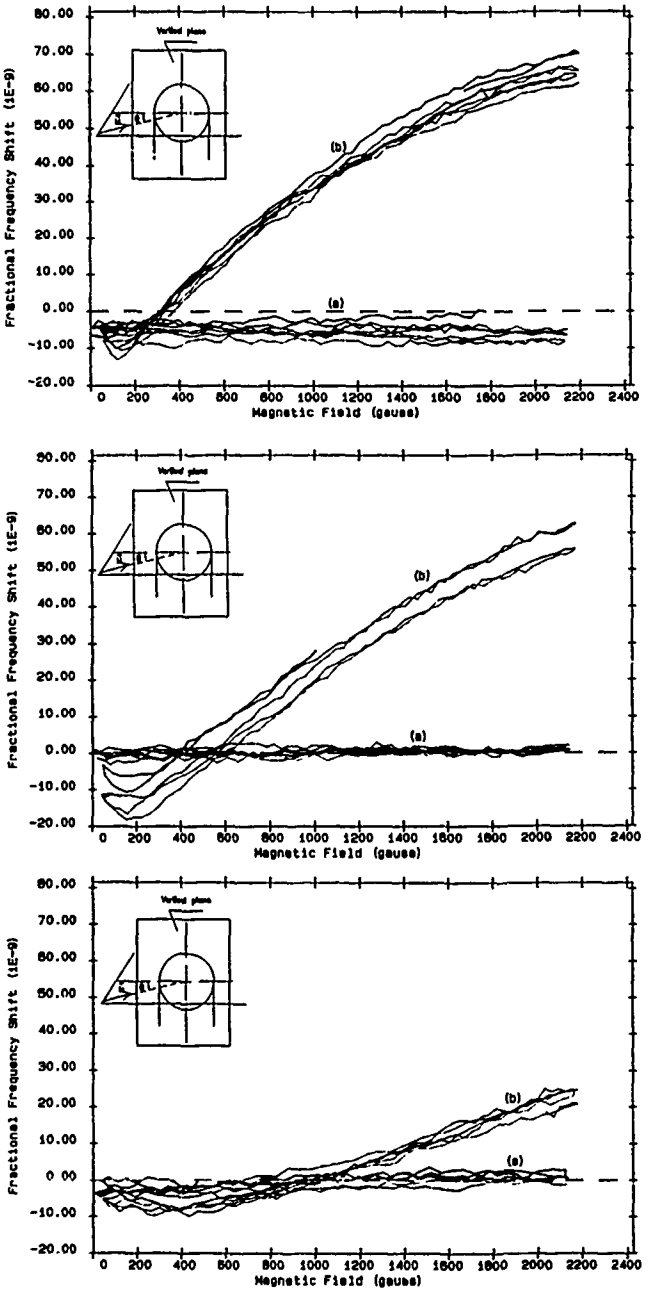


Fig. 19 : Brass supports used to check the magnetic sensitivity of the crystal mounting

Experimental verifications

In order to prove the above mentioned assertion, two experiments have been carried out. In the first one, the magnetic strips of Fig. 15 have been replaced by two thick brass supports (Fig. 19) in such a way that the mounting cannot be bent by the magnetic forces, using the same crystal. Figures 20, 21 and 22 show the records obtained for various directions of the magnetic field both before and after the magnetic strips have been replaced. It can be seen that with the brass supports the resonator no longer exhibits any measurable frequency change with respect to the magnetic field.



Figs. 20, 21, 22 : Magnetic sensitivity records obtained for different induction orientations with brass supports (a) and nickel-alloy supports (b) Fig. 20 : $\theta = 0°$, Fig. 21 : $\theta = 60°$, Fig. 22 : $\theta = 90°$

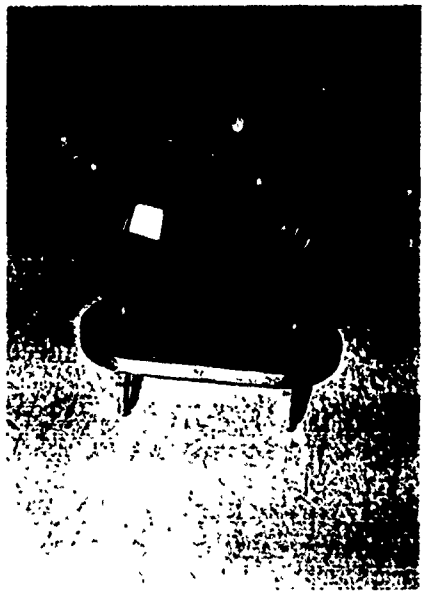


Fig. 23 : A small mirror glued to the end of a strip permits measurement of its motion with a laser beam

The second verification deals with the deformation of the magnetic strip itself. In order to determine the order of magnitude of the bending of the strips, a small mirror of about 1 mm^2 was glued to the free end of one strip (Fig. 23). In the magnetic field the bending of the strip moves the mirror which in turn modifies the reflected path of a laser beam. By using a suitable set of mirrors (see Fig. 24) it is possible to magnify the deformation of the strip, which can be estimated by measuring the displacement of the laser spot on a screen with a magnification factor of about 10 000. A 2 000 gauss magnetic field strength induces a spot displacement of about 5 cm so the end displacement of the strip may be estimated to be of a few microns. Although this experiment doesn't permit measurement of the torques applied to the crystal when mounted, it may be used to get some experimental insights into the magnetic properties of the material used. In particular it has been confirmed that the bending, always in the same sense, doesn't depend on the sense of the magnetic induction.

Conclusion

The magnetic sensitivity of a quartz crystal oscillator may be explained by the stress sensitivity of the resonator, induced by the deformation of crystal supports made from a magnetic material. Nevertheless, it is possible that each of the other components of the oscillator may have its own sensitivity. Experimental investigations are being performed to check this, but they will need a rather sophisticated setup as the effects expected are of very low magnitude. Furthermore, the results presented here have been mainly obtained with high applied magnetic induction fields, and do not give reliable informations on the oscillator behaviour in a weak magnetic field, as in actual practice. Efforts are now being made to observe frequency changes in magnetic field as low as a few gauss when the resonator is far from its sustaining electronics.

Another problem is to find methods for reducing the magnetic sensitivity. Some solutions have already been presented here, but in the case of the resonator one should be thinking of replacing magnetic supports by non-magnetic material having otherwise the same thermal and mechanical properties (aging characteristics, etc.).

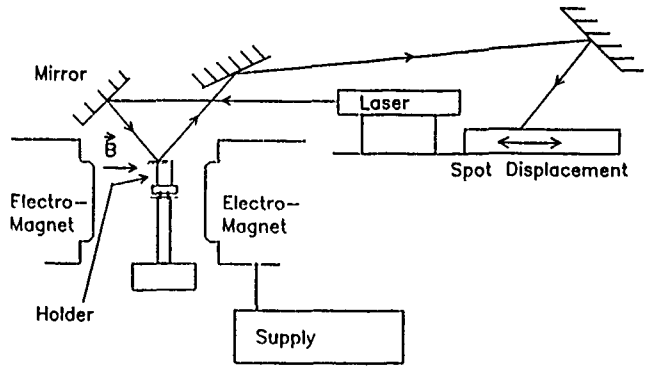


Fig. 24 : Experimental setup for demonstrating bending of the supports in magnetic field

Acknowledgement

This work is supported by CNES and CNRS

References

[1] J.-B. Laporte, "Specifications and performances of ultrastable oscillators in the Doris orbitography and positioning system", Proc. 2nd European Frequency and Time Forum (EFTF), Neuchâtel, 1988

[2] Granger, Roujas, "Implantation de Doris dans Spot 4", 1988

[3] P Sengenes, F. Nouel, "Evaluation des effets de l'instabilité des O.U S sur des mesures Doppler de type Doris", CNES Technical Report 023, 1986

[4] A. Benezet, E. Robert, M. Brunet, "Compte rendu d'essais n° L5160", CNES Technical Report CT/DRT/TIT/TF 87/297, 1987.

[5] E. Robert, M. Brunet, "Susceptibilité magnétique des oscillateurs à quartz spatiaux", CNES Technical Report CT/DRT/TIT/TF 88/354, 1988.

[6] A. Benezet, E. Robert, "Mesure de susceptibilité au champ magnétique et contrôle de l'efficacité du blindage magnétique associé", Technical Report ITS/L5160.

[7] F Deyzac, "Oscillateurs à quartz de bord en environnement sévère Modélisation de comportement", ONERA Technical Report 38/7132PY, 1989.

[8] J -J. Gagnepain, J -G Théobald, "Influence d'un champ magnétique sur la fréquence propre d'un résonateur à quartz", C.R Acad Sc , vol 292, série II, janvier 1981.

[9] L Neel, "Sur la variation, en fonction d'un champ magnétique appliqué, de la fréquence propre d'une lame de quartz, revêtue d'électrodes adhérentes en nickel", C.R. Acad. Sc., vol 292, série II, mai 1981

[10] A. Ballato, T.J. Lukasek, G.J. Iafrate, "Subtle effects in high stability vibrators", Proc 34th Ann. Freq. Cont. Symp , 1980

[11] R M Bozorth, "Ferromagnetism", Van Nostrand, N Y , 1951.

[12] D Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thèse de Docteur-Ingénieur, Université de Besançon, 1978.

43rd Annual Symposium on Frequency Control - 1989

Test and Evaluation Methods at the NRL Clock Test Facility

R.L. BEARD, F. DANZY, E. POWERS AND J.D. WHITE
Naval Research Laboratory
Washington, D.C. 20375-5000

Abstract - The Space Applications Branch of the Naval Research Laboratory has developed a special Clock Test Facility to evaluate atomic clocks developed for space and ground application. The facility itself will be described and the methodologies involved in long-term performance testing, initial qualification testing of spacecraft clocks and post-acceptance testing of candidate spaceflight clocks to be evaluated in orbit. The objective of long-term testing is to build a performance and reliability database on newly developed clocks to support eventual operational system acceptance. Initial qualification testing for developing clocks is used to guide development under contract and to support in-house design of components, subsystems and experimental atomic clocks. Typical testing procedures and methodologies will be discussed. The purpose of testing spaceflight clocks is to perform ground testing beyond current acceptance tests of flight candidate cesium beam frequency standards. The additional data will provide a better understood clock for GPS applications by extending the available baseline data, and could result in pre-flight detection of latent defects. Latent defects as used here, represent weaknesses which could result in a much higher piece part failure rate than that indicated by predictive analysis. The procedures of the post-acceptance testing are designed to maximize the amount of available data on a clock while limiting the total amount of usable life consumed in testing. A special experiment, known as the Benign Environment experiment, will be described. This experiment will be designed to quantify the environmental effects on atomic clocks in space. A question has been raised in the evaluation of atomic clocks in the GPS satellites, in that they appear to be performing better in orbit than they did in ground testing. An approach to answering this question will be described along with the means of parameterizing even small effects ordinarily ignored in ground testing.

I. BACKGROUND

In Dec 1973 the Department of Defense assigned the Air Force as the executive service for the joint-service NAVSTAR Global Positioning System (GPS) program and established the GPS Joint Program Office (JPO). The GPS JPO assumed the responsibility for development of the entire GPS program, with the Navy being assigned specific responsibilities concerning advanced technology, the most notable of which was advanced clock technology. [1] An integrated management approach to proving the system concept was developed. A key element of the overall concept was the improvement, development and deployment of atomic clocks for application in the NAVSTAR satellites and ground segment. Since this beginning of the program, the Naval Research Laboratory (NRL) has had the responsibility of developing these atomic clocks for satellites and ground stations. [2]

The activities involved in the atomic clock development for GPS have a long history. Efforts to develop precision flight frequency standards were initiated in 1964 by

NRL as part of the Timation navigation satellite program. The first two satellites in this program, designated Timation I and II, (launched May 1967 and Sept 1969 respectively) employed quartz-crystal oscillators as frequency standards. These standards were especially developed for spaceflight and stability. [3] While quartz oscillators do not have the required long-term stability for precision navigation, their use on the Timation satellites showed their reliability in a space environment and provided an excellent early demonstration of the use of space-borne clocks for navigation using passive ranging, or as it is called in the GPS sense, pseudo-ranging. It should be noted that each atomic clock employs a quartz oscillator as the output element. Further, the standards in the developmental (Block 1) satellites and current operational GPS satellites (Block 2), are so designed that the crystal oscillator may be operated independently in a backup mode if the atomic clock should fail. Navstars 1 and 2 were operated for a time in this manner. Additional satellites of the Timation series were planned. Timation III was to employ further refined quartz oscillators and two experimental rubidium atomic clocks, and the fourth satellite was to use a cesium AFS. With the formation of the GPS program, the Timation project was merged with an Air Force project and the planned two Timation satellites were re-designated Navigation Technology Satellites (NTS) 1 and 2 respectively.

A key objective of NTS-1 was to demonstrate the performance of space-borne atomic clocks. The satellite was launched in July 1974 and contained two quartz oscillators and two experimental rubidium clocks. [4,5] Even though a stability control problem developed which left the NTS-1 in an unstable condition and precluded a thorough evaluation of the rubidium, sufficient data were collected to show the improved performance of rubidium over quartz. NTS-2, launched in July 1977, was intended to be one of the original four satellites comprising the GPS Phase I constellation. Of the two experimental cesium standards carried on NTS-2, one experienced a power supply failure after a period of satisfactory operation. The other continued to operate long after a satellite failure occurred which prevented collection of data to continue long term performance evaluation.

In the GPS Phase I (Concept Validation) plan, the satellite constellation was to consist of four vehicles, NTS-2 and three Navigation Development Satellites. Rubidium clocks were selected for use on the first Navstar satellites with the intent of adding cesium standards as the progress of the technology permitted. The NTS-2 ranging transmitter failed prior to deployment of the first NDS. Consequently, the Phase I GPS constellation consisted of four NDS vehicles (designated Navstars 1, 2, 3, and 4). Navstars 1, 2 and 3 each carried three rubidium standards. Navstar 4 had, in addition to the three rubidium standards, an engineering development model (EDM) cesium standard. Navstars 1 - 4 were all launched during CY 1978. The rubidium standards on the Navstar satellites provided performance which supported an evaluation of the GPS concept and demonstrated the navigation accuracy requirements. However, single-day prediction errors (one upload) were at times as much as ten meters as compared to a three meter requirement for the operational GPS satellites. By 1979 the experience with atomic clocks in the satellites was such that a dedicated clock development program was initiated between the GPS JPO and NRL.

A. Cesium Development

The first cesium clock for spacecraft was

developed by Frequency and Times Systems, Inc. (FTS) under contract to NRL. This was an interactive development in that many of the tests to qualify the prototypes and engineering model units were performed at NRL. The specification for the unit was developed over the course of the prototype development, and formed the basis for future development. After the prototypes were built and tested in the Laboratory the project proceeded through an engineering development phase to preproduction, before the prime contractor for the operational (Block II) GPS satellites subcontracted for flight units. Since these units were the first to be built for satellites little data was available on the operation of cesium under anything other than laboratory or controlled conditions. To develop a new component for a space system and expect a reasonable projected lifetime would require representative numbers of units to be built and tested to develop a database of performance and operating characteristics.

The dedicated clock development program begun in 1979 called for the development of alternative sources of spacecraft cesium standards. The development with Frequency Electronics Inc. (FEI) and Kernco, Inc., incorporated the experience learned through the inital development. Common specification, qualification and acceptance test plan [6]-[8] were used with the two contractors in a phased development from EDM to PPM. The resulting PPM units are to be launched in the Block II replenishment satellites as the primary clocks for in-orbit evaluation and demonstration.

B. Hydrogen Maser Development

With the development of improved frequency standards for satellites, the need for even better standards in the ground tracking stations became greater The availability in the early 1970's of new hydrogen maser standards from the Smithsonian Astrophysical Observatory offered the possible of these units fulfilling the tracking station need. These units were first developed for Very Long Baseline Interferometry, whose requirements were primarily in the short term, less than one day, and were operated by experienced scientists at observatories. Investigations into improved reliability and smaller size led to a dedicated program to develop hydrogen maser technology for ground and space application [9,10]. This development of a new device represented the same testing problem as the case of cesium, only with an increased requirement for better precision [11].

II. Overall Test Approach

To test and evaluate precision frequency standards for long term applications in a space environment required a specialized testing facility. The facility must be capable of performing evaluation while the clock under test is operating in an environment similar to the anticipated mission environmental profile. The resulting measurements must be very precise and referenced to accurate and precise standards. The test facility developed for this purpose at NRL is divided into four sections, (1) flight candidate hardware testing, (2) prototype and engineering development model evaluation, (3) life testing and (4) Benign Environment evaluation. Figures 1 is a block diagram of the Clock Test Facility, figures 2 through 5 are block diagrams of the individual section configurations. The facility is equipped with a 100 kilowatt emergency power backup system to ensure uninterrupted facility operation. Each section was developed to provide information about candidate clock designs during different phases of development. While the test plans may vary for different articles, procedures for performing the tests are

controlled by the section where testing is being conducted. For instance, section (1) is a limited access area for testing actual flight candidate hardware that will be launched in satellites. This testing requires detailed documentation of all events occurring during unit testing. While section 3 is a life testing section which continuously monitors short term data spans and only records significant events or data for intervals of up to one or two months. All environmental chambers are equipped with fail-safe power shut-down for test articles, should a malfunction occur in a environmental chamber's thermal or pressure control systems. Each clock is interfaced with a telemetry simulator units for controlling normal power on/off cycles, discrete commands and digital tuning. All performance, telemetry and environmental monitoring data is recorded for later analysis.

A. Test Planning

Test plans are tailored to specific test articles and are designed based on, prior knowledge of the article under test, it's stage of development and the performance parameters to be evaluated. The clock development regardless of the technology (cesium and hydrogen maser), typically follows the development pattern derived from MIL-STD 280, from prototype (or advanced development model (ADM)) to EDM to preproduction model (PPM). Prototype, ADM or EDM would have a different test plan than the PPM or flight candidates. PPM and flight candidates are not tested to verify if a unit meets the required specifications but to establish a performance data base in a simulated operational environment, on a design or particular unit prior to flight. For a given design, individual units will exhibit differences in performance and environmental sensitivities. This occurs even when units are built from the same lot of piece parts and materials and fabricated under established controlled procedures.

Typically test programs for developmental units will expose the units to a highly stressful environment. The test programs are designed to precipitate failures to reveal possible design weaknesses or improper selection of piece parts. If no failures occur, a review of stress levels might be in order to determine if the levels should be increased or if modifications in the test sequence is required. Stress levels imposed for these units are not driven by the anticipated mission profiles. Life testing for these types of units is usually not considered because it consumes equipment and provides questionable insight into the final design life. Life testing is generally restricted to PPM units or special subassemblies such as cesium beam tubes. Tubes are a special consideration since the new designs used in the space units have little background and database to establish reliability and operating life. After a baseline performance database of prototype to EDM units has been established, some units are instrumented with sensors to evaluate their thermal profiles. These tests often corrupt the clock stability data, but insight into possible thermal stress levels are obtained which can be a critical parameter in missions with long durations. Environmental testing of subassemblies is often preferred over total system testing since design weaknesses can be masked by good performance in the short term. All test programs include, (when ever possible) a mechanism for providing feedback of test results to the manufacturer of the unit to strengthen their understanding of the final objectives of the development and improve their capability to produce a good unit. Test plans are developed in concert with the manufacturer and end user to ensure the unit will conform to the system of use and the system designers

are fully aware of the units capability.

The testing of prototype, ADM and EDM units is to provide data as follows:

1. Performance

2. Adaptability to ultimate intended use and environment

3. Usefulness of diagnostic monitors

4. Commands and tuning required

5. Configuration, weight and power considerations

6. Formulation of test plans for PPM life testing and flight candidates

7. Development of required test documentation

8. Establish special user operational requirements (storage life, ion pump cycling, stabilization time)

In addition to providing baseline design data the prototypes and EDM models are used to formulate test procedures, data requirements and test systems configurations for long and short term testing. Through this gathering of test data the specifications for PPM units are defined.

B. Flight Testing

A section of the test facility is a limited access area designated for performance testing of PPM or flight candidate units. The intent of this section is to establish a baseline performance of the particular unit to better anticipate on-orbit performance. Experience has shown that clocks of the same design and manufacture will exhibit different on-orbit performance characteristics, and in some cases, these differences have been significant. Tests performed in this section should not be interpreted as a screen or an extension of the acceptance testing performed by a clock manufacturer. Acceptance test program are designed primarily to verify the unit meets the performance specifications, and as a screen to detect workmanship errors and precipitate latent defects in components to detectable patent defects. Latent defects are defects undetected during piece part screening or introduced during assembly. Acceptance testing will provide limited or no insight into the expected long term performance, especially if the unit has a high aging coefficient during early life. A data package is generated as an addendum to the acceptance data which fully documents the total life of the item while at the facility. This data will aid in determining the useful operating life of the unit for future system planning. The data includes any on/off cycles scheduled and unscheduled, and any event such as vacuum chamber anomalies, emergency power switching, telemetry simulator or data acquisition malfunctions, etc. When these flight units are not under powered evaluation they are stored in a controlled thermal vacuum environment. This precaution is taken to prevent possible contamination which might reduce the useful on-orbit operating life by causing accelerated aging in some organic materials or increases in the degradation across thermal interfaces. The specific test goals are to, (a) determine long term stability performance out to a tau of 10 days, (b) observe the unit's operating characteristics under anticipated on-orbit thermal variations, (c) document diagnostic monitors for all conditions to aid in analysis should anomalies occur on-orbit and (d) provide data to evaluate the variances from unit to unit of a particular design and investigate solutions for the possible reduction of these variances.

The cesium standards used on the GPS spacecraft have produced a limited data set on cesium performance. This is because

even in operational deployment, the number of flight clocks that have been manufactured is small. Of this number, only 6 of the Preproduction Model (PPM) series have actually been operated on board a satellite and two production series are just starting in orbit. Only one of the PPM clocks (PPM-14) has undergone extensive ground testing. Clocks designated for flight use are subjected to the required Acceptance Test at the contractor's facility and are then placed in bonded storage until required for spacecraft integration. They are again tested on the spacecraft as part of the overall spacecraft checkout and then turned off until required for use on orbit. None of the current flight production clocks has been subjected to extended ground testing. Knowledge of the behavior of these standards will be limited to acceptance test data and demonstrated, on-orbit performance.

Ground testing of flight frequency standards beyond the current acceptance level testing can provide better clocks for GPS by extending the available database. This will not only result in pre-flight detection of more clock anomalies, but also provide a better basis for understanding the severity of and possible solutions to on-orbit clock problems. This same knowledge would be available to the GPS clock contractors for future product improvements. There is an additional benefit to testing each flight clock to a greater level than currently done in the acceptance testing. GPS uses a Kalman filter algorithm in the Master Control Station software to predict the satellite clocks performance. This method requires accurate input parameters to make reasonable output predictions. The present method is to use the clock specification levels for input. This does not allow the system to take full advantage of a better clock and makes no allowance for differences between clocks bought under the same specification. Iterative improvement by observing operation in the system suffers the same problem as on-orbit clock performance evaluations.

C. Specifications and Procedures

The initial specifications for space cesium standards was developed in 1974 with the initial contracts with FTS. The contract was to build four flight cesium prototypes for the NTS-2 satellite. The key element of this development was to space qualify the cesium beam tube which had been specially tested in a prior cesium beam tube study to determine its capability for surviving the launch environment. NTS-2 was to use the same launch vehicle and stage vehicle system that the Block I NAVSTAR satellites was to use. Therefore, the qualification environment would be directly applicable to both satellite vehicles. The initial procedure for testing were based on the tests developed under the TIMATION program for precise quartz oscillators. The later procedures with the alternative cesium developers incorporated the information gained in the inital PPM program. The development of the hydrogen maser unit for space and ground followed the same development procedure in an even more interactive program to develop the basic technology.

D. Benign Environment Experiment

The testing of precision clocks in a ground test environment is valuable for simulating how the clock will perform in the system. To provide information on how the clocks are actually performing an effort to separate and estimate the clock performance on-orbit is on-going. The results of these efforts has shown a range of performance estimates covering almost and order of magnitude. This performance span could be due to the uncertainty in the distribution of er-

rors in the system, since the clock error is indistinguishable from the satellite position error. However, it is possible that the clocks are performing better than that shown in ground tests due to the satellite environment. As a means of evaluating that possibility, the benign environment experiment was designed to isolate and examine the different parameters of environmental sensitivity. Tests in thermal vacuum chambers ordinarily would control temperature and pressure. Other effects such as electromagnetic fields and vibration would not be controlled or monitored.

The experiment would be designed to separate the different environmental effects and reduce them from ambient levels by one to two orders of magnitude. Eliminating them completely would of course be impossible. With each effect so separated, they would be re-introduced. The effects, if any, on the clock in the test chamber would be identified. This will place even greater demands on the measurement system to measure whatever effects may be detected. The one spacecraft effect not possible to evaluate will be zero gravity. It is hoped that isolation from vibrational influences, including seismic vibrations, will simulate the very low force effects of zero gravity on the clock.

III. Test Facility

The 13 test chambers and other test setups are tied together by a central data acquisition system. This system can collect performance data and telemetry information from up to 48 separate clocks under test and store these data in a 6 month temporary file. The temporary file is duplicated in another long term storage file and transferred to a VAX-750 for further processing and analysis. Integral with the central data acquisition system are two workstations, which enable an operator to view the latest data for performance analysis in a variety of displays or analysis routines.

The test facility primary reference for time/frequency measurements is the Naval Observatory. To facilitate this reference, a number of specially built hydrogen maser standards, which comprise the local reference, are compared to the Master Clock at the Naval Observatory in two ways. A special arrangement with the local channel 5 TV Station enables a comparison of the broadcast carrier at NRL and the Observatory to provide a continuous sub- nanosecond phase comparison to be made. The other technique is by time comparison via the NAVSTAR GPS in the few nanosecond range. These comparison techniques are integrated into the data acquisition system so the comparison data are collected automatically and continuously.

A. Flight Clock Database

GPS has flown and successfully operated cesium clocks in NAVSTAR spacecraft 5,6,8,9,10 and 11. Operational Block II NAVSTARs 14 and 13 are now in orbit for inital test and eventual operational use. Of the six Block I units, one (NAVSTAR 5) was turned off due to spacecraft problems and two others (NAVSTAR 6 and 11) failed due to cesium exhaustion believed due to oven problems. The remaining continue to operate with more or less nominal performance. Such anomalies as have been reported have been remotely diagnosed as well as possible. A good example of the remote diagnosis is NAVSTAR 6 where random frequency jumps occurred during the first year of its operating life. These were on the order of a few parts in 10^{-12} and seemed to be of equal magnitude occurring at random intervals. While several theories have been propounded, no exact diagnosis can be made due to lack of complete data.

The clock designated PPM-14 was built

by FTS at the same time as the initial batch of flight clocks bought by the GPS prime contractor. It is the only clock of the PPM or later series that has been extensively tested on the ground. PPM-14 was built to full flight standards and was, at one time, actually scheduled for flight due to a lack of available standards. Since its return to NRL, it has been tested in thermal vacuum for long term stability and temperature effects. The long term stability data has been consistently within original specification and, just as consistently, worse than its on-orbit siblings as well as can be determined from on-orbit evaluation. The temperature data was done simulating on orbit environment and clearly showed a phase change that should have been detectable by the GPS control segment. PPM-14 has also shown frequency jumps different from those observed in NAVSTAR 6.

B. Example Results

The two workstations incorporated into the data acquistion system for the test facility permit output of the test data during the test performance. The data as is it accumulated, is available on the workstation so the progress of the test or special experiment came be monitored continuously. Figures 6 through 9 are example results of monitoring tests. As the data is accumulated the dual mixer approach to obtaining phase data allows any clock on the system to be compared with any other clock on the system. The software in the workstation allows the experimenter or test personnel to run evaluation routines ranging from a simple plot of the phase comparisions to detailed stability analysis. The data would then also be available for studies into clock performance or modeling.

IV. Conclusions

The data gathered so far from flight clocks on-orbit and PPM-14 have been valuable in understanding and improving the performance and reliability of frequency standards used in the GPS spacecraft. The amount of information available is still limited by the actual number of clocks built and tested. This is partially offset by the extensive experience in the clock community with Hewlett-Packard clocks and a smaller number of FTS commercial and FEI militarized clocks. However, these clocks operate in a different environment than the GPS units. In particular, the thermal environment of space leads to significantly different component operating temperatures and different temperature gradients. The GPS design requirements for reliability are quite stringent in this area but only long term testing of a meaningful number of clocks will verify the design. While this data will eventually come out of GPS' orbital experience, it may come at the cost of premature loss of clocks. Even then it may be quite difficult to sort out the cause of the failure with only the available telemetry data.

The delays in the GPS operational deployment schedule offer an opportunity to significantly expand our knowledge of the GPS flight hardware. Rather than keep these clocks in storage either on a spacecraft or on the shelf, as many as are available could be tested in the space environment (thermal vacuum within the acceptance test range limits). This testing will help find infant mortality type failures prior to launch and will also allow a much broader look at anomalies under controlled conditions. It would also be highly desirable to designate several clocks from the current batch of flight units to be long term test devices. These could be the clocks from spacecraft which will fly the alternate source cesium.

Clocks from the flight group which have been redesignated for long term testing

would be tested over a broader range of conditions. In particular, these units would be tested at constant temperature in thermal vacuum to uncover latent thermal design flaws and also to document long term frequency stability in a stable environment. One or more of these clocks could also be temperature cycled within the acceptance range over a long time to force failures due to design. The resulting data would be given to the manufacturer to allow evolutionary design improvements in future clocks.

REFERENCES

[1] Easton, R., "The Navigation Technology Program", *NAVIGATION: Journal of the Institute of Navigation*, Vol. 25, No. 2, Summer 1978

[2] R.L. Beard, J. Murray, and J.D. White, "GPS Clock Technology and the Navy PTTI Programs at the U.S. Naval Research Laboratory", *Proceedings of the Eighteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, December 2-4, 1986

[3] C.A. Bartholomew, "Quartz Crystal Oscillator Development for TIMATION", NRL Report 7478, 25 October 1972

[4] S. Nichols, J. White and F. Danzy, "Design and Ground Test of the NTS-1 Frequency Standard System", NRL Report 7904, 5 Sept 1975

[5] T. McCaskill and J. Buisson, "NTS-1 (TIMATION III) Quartz and Rubidium Oscillator Frequency Stability Results", NRL Report 7932, 12 Dec 1975

[6] NRL Specification for Advanced Cesium Standard Development dated 1 Jan 1984.

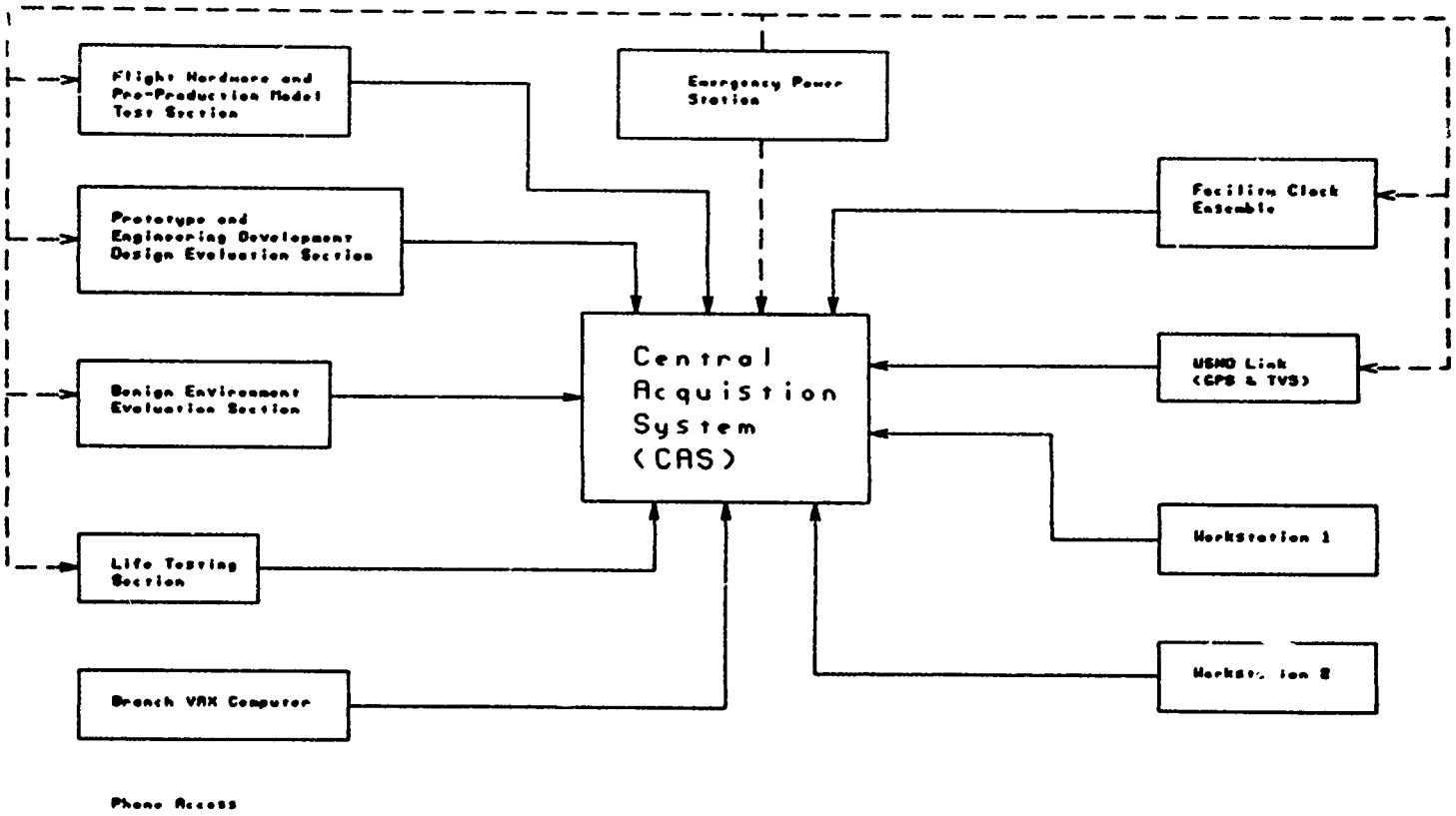
[7] NRL Qualification Test Plan for Advanced Cesium Frequency Standard Development dated 30 Jan 1984.

[8] NRL Acceptance Test Plan for Advanced Cesium Frequency Standard development dated 30 Jan 1984.

[9] R. Easton, "The Hydrogen Maser Program for Navstar GPS", *Proceedings of the Eighth Annual Precise Time and Time Interval Meeting*, December 1976.

[10] R. Moore, et. al., "Design of a Spacecraft Hydrogen Maser," 31st Annual Frequency Control Symposium, Atlantic City, N.J., June 1977.

[11] NRL Specification for Development of Compact Spaceborne Hydrogen Maser, dated 21 Apr 1986.



NRL CLOCK TEST FACILITY

Figure 1

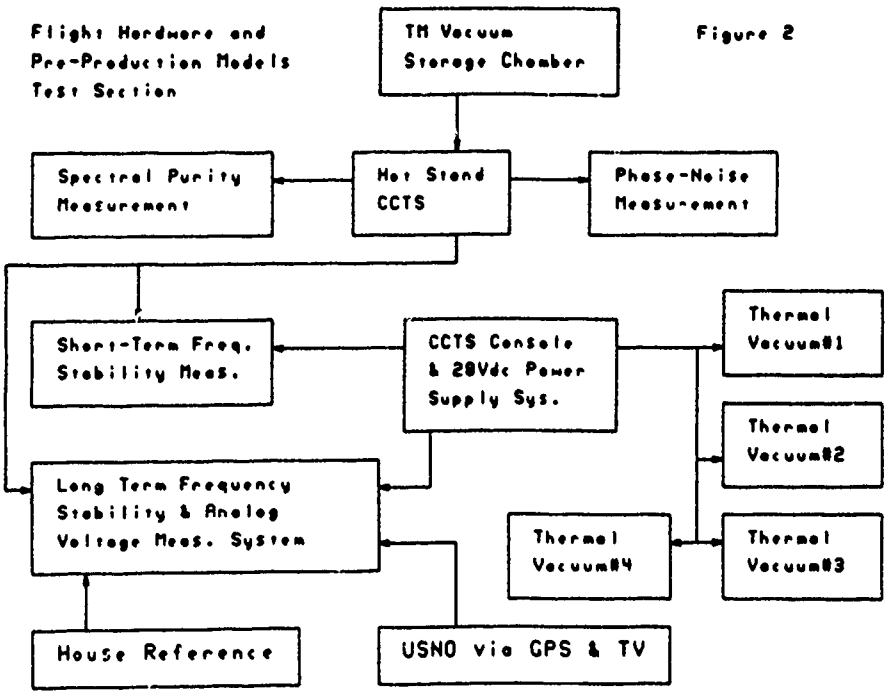


Figure 2

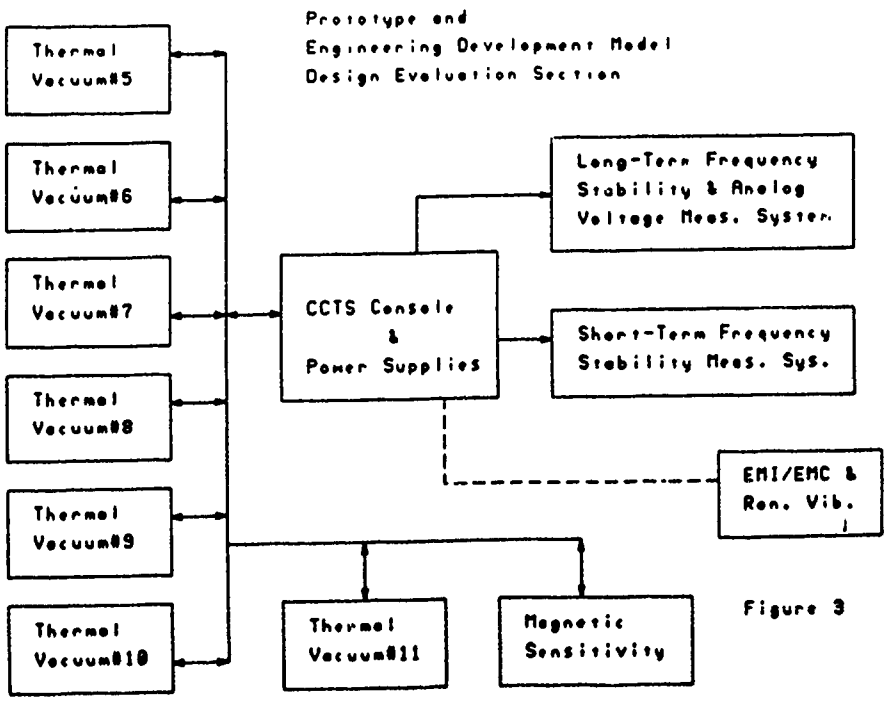
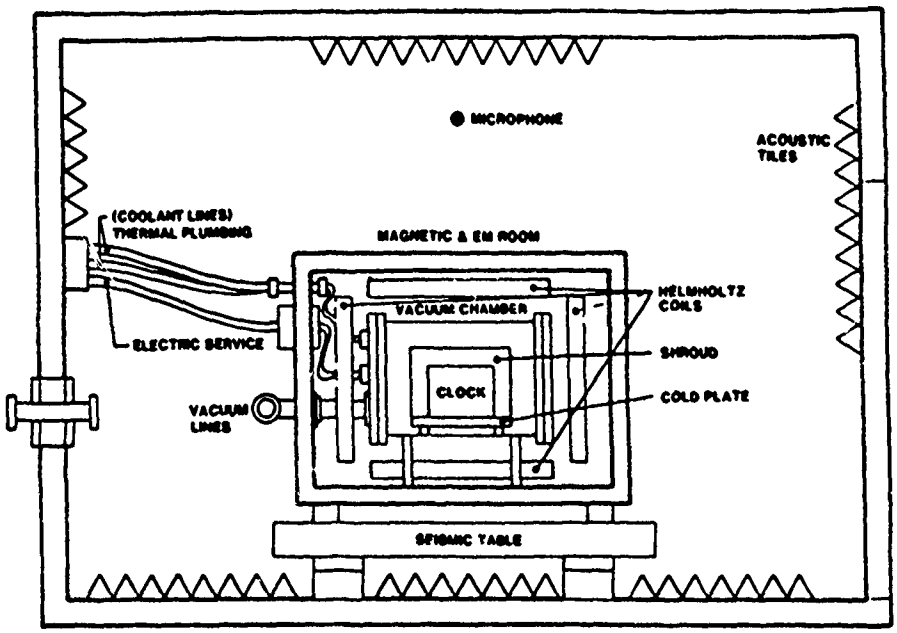


Figure 3

Figure 4



BENIGN CHAMBER

Life Testing Section

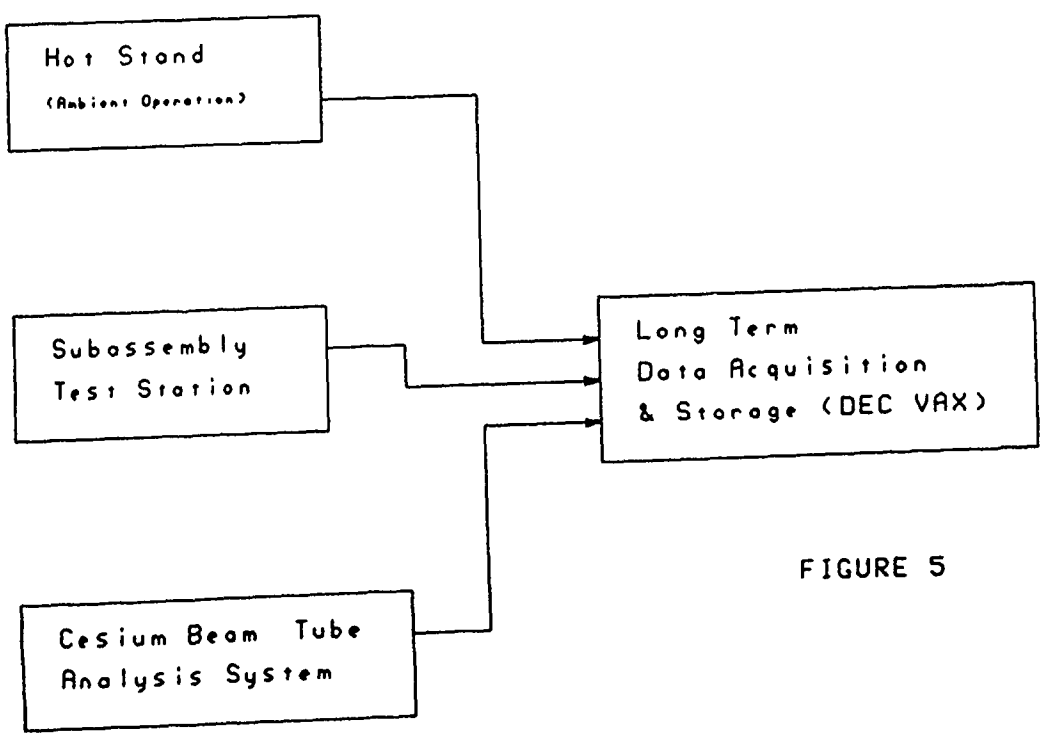


FIGURE 5

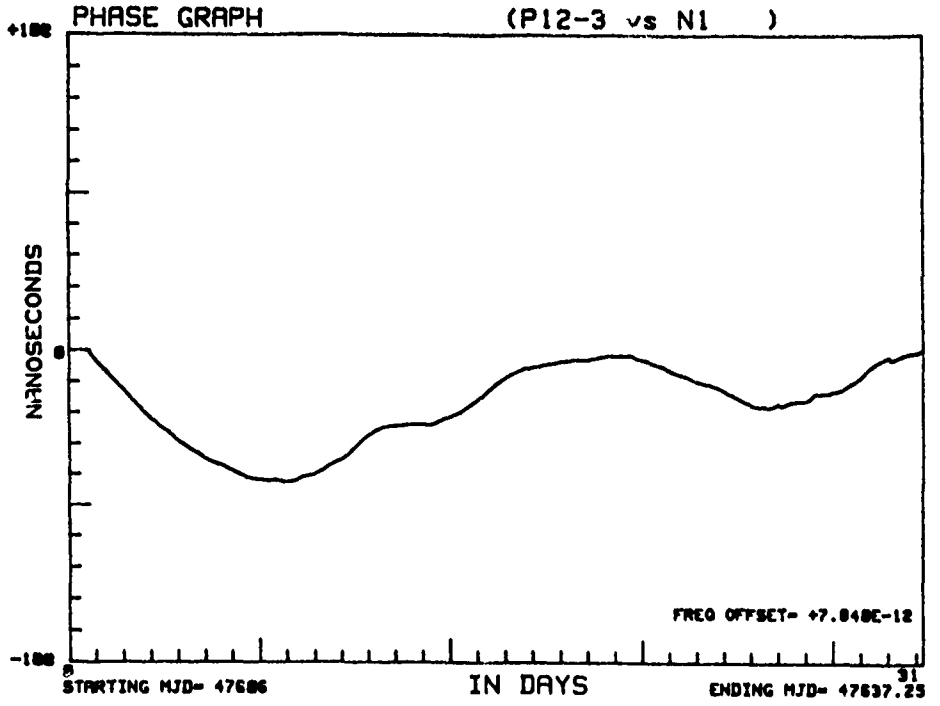


Figure 6

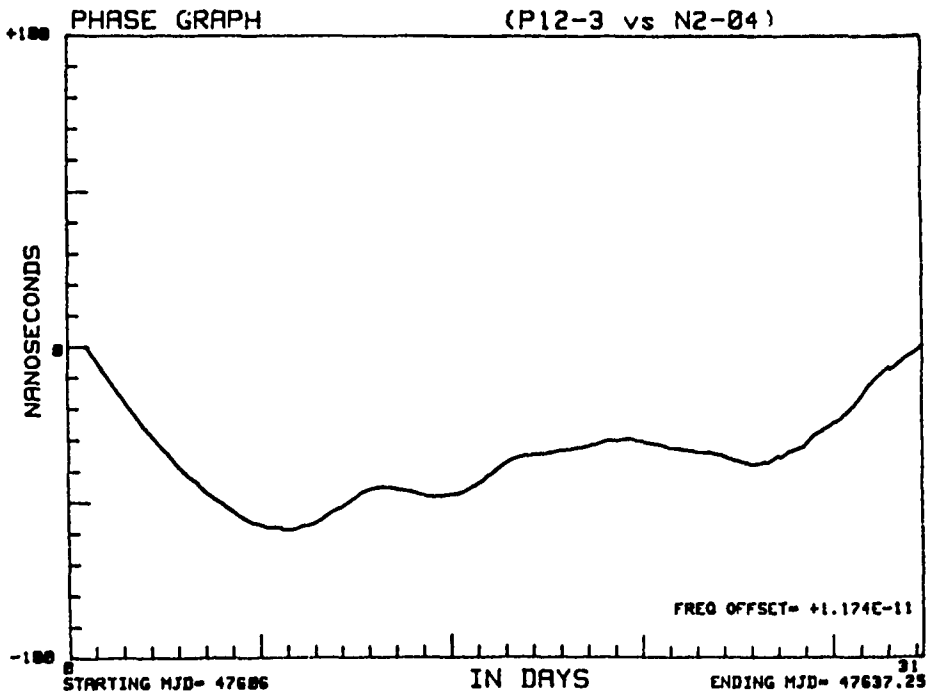


Figure 7

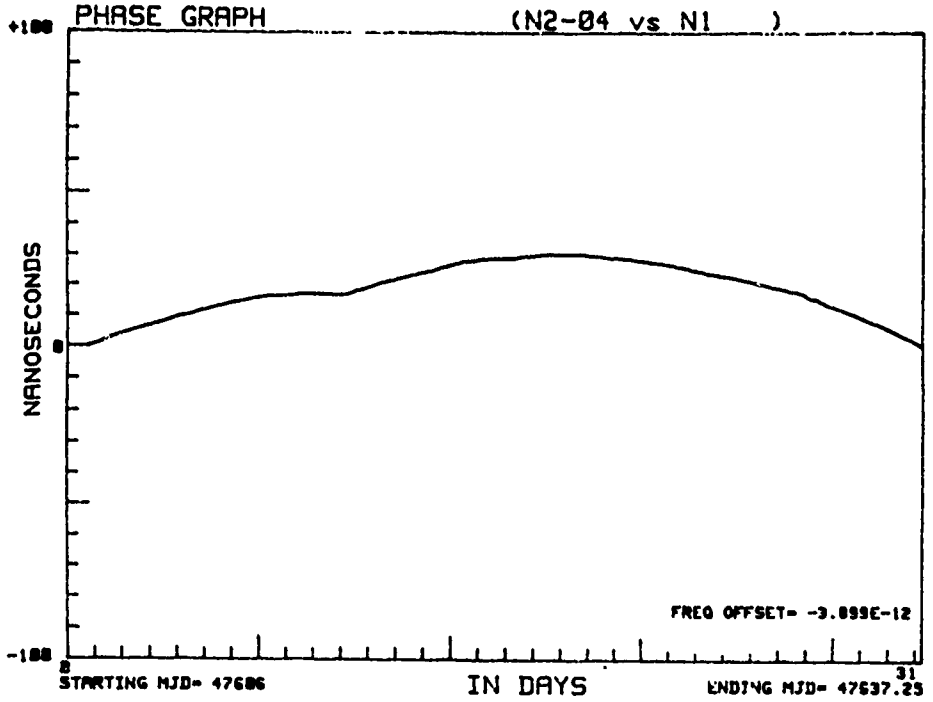


Figure 8

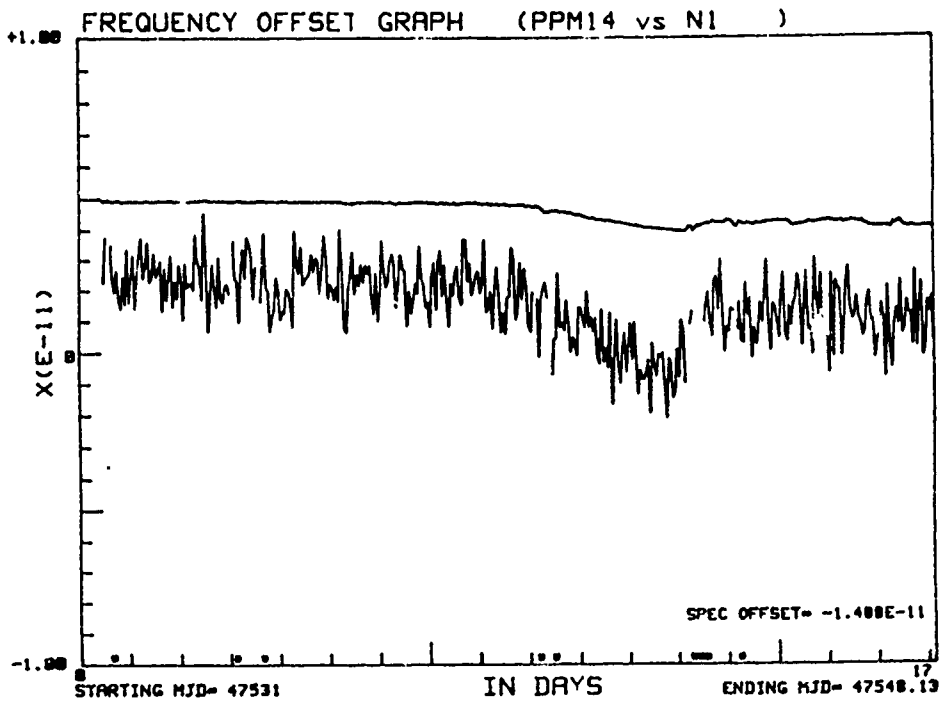


Figure 9

43rd Annual Symposium on Frequency Control - 1989

ENVIRONMENTAL TESTING AT THE JET PROPULSION LABORATORY'S FREQUENCY STANDARDS LABORATORY[1]

RICHARD L. SYDNOR
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

ABSTRACT

The Frequency Standards Laboratory at the Jet Propulsion Laboratory (JPL) is responsible for the research, development and implementation of frequency generation, monitor and control and distribution equipment for the worldwide Deep Space Network (DSN). In order to achieve the required state-of-the-art performance and reliability, an extensive testing capability has been developed. This capability includes special equipment and facilities as well as a standard set of procedures which are described in this paper. The research and development program utilizes these capabilities in the development of advanced frequency standards and distribution equipments that are well characterized for environmental variations. The implementation program tests each piece of equipment to be installed in the DSN for various environmental sensitivities which may affect its field operation. The environmental tests which are performed regularly are:

* temperature,
* barometric pressure,
* humidity,
* magnetic field and
* time.

INTRODUCTION

The Frequency Standards Laboratory (FSL) at the Jet Propulsion Laboratory (JPL) is responsible for the Research and Development and the Implementation of the Frequency and Timing System (FTS) equipment for the NASA/JPL Deep Space Network which operates tracking stations at three stations spaced approximately 120° apart around the world. This network tracks, controls and receives data from the various planetary probes (Mariner, Voyager, Magellan, etc.). The long round trip light times and the desire to extract the most information possible from each launch and fly-by produces extremely stringent requirements on the FTS equipment. Since the equipment is operated in a field environment, complete testing of each component and subsystem is necessary in the research and development phase, before implementation and after every repair. The requirements on the performance of the FTS and the variations in the environment in the stations of the DSN are used to determine the allowable environmental sensitivities of the different parts of the FTS and to select the test regime applicable for each device.

PERFORMANCE REQUIREMENTS

Each mission generates its own performance requirements set for the operation of the entire DSN. From these sets, an error budget is developed for each of the systems in the DSN. This error budget consists of the overall set of specifications which will meet all requirements. The present error budget for the FTS is shown in Table I.

This performance is for the entire FTS, in the field environment, 24 hours per day, 365 days per year. It is apparent that knowledge of the actual environment is essential in order to determine the required environmental sensitivities for the components of the FTS. Typical environmental parameters for the DSN stations are shown in Table 2.

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

CH2690-6/89/0000-289 $1.00 C 1989 IEEE

The actual spectral characteristics (or time domain behavior) for each of these parameters is also important. Typical vibration spectra for the antenna cone environment are shown in Figure 1 for the vibration frequency range of 0 to 1 Hz and 0 to 100 Hz. Similar curves are available for the other antennas as well as the Frequency Standards Room and the Control Room.

By combining the requirements of Table 1 with the vibration environment of Figure 1 and Table 2, the requirements for the vibration sensitivity of an antenna mounted VCXO were derived. In actual usage, these requirements must be modified by the transfer function of any phase-locked loop or filter which affects the performance of the VCXO. In addition, since the performance requirements are near the state of the art, the vibration induced noise must be somewhat lower than the total requirements. We have arbitrarily specified a ten dB margin. The resulting required vibration sensitivity is shown in Figure 2, along with the actual measure sensitivity of the best VCXO available to us.

Test Capability of the Frequency Standards Laboratory at JPL

A set of three environmental test chambers in the Frequency Standards Laboratory (FSL) is shown in Figure 3.

These chambers are 8 ft x 8 ft x 8 ft and have vertically directed air flow to simulate the conditions in the DSN. They are controlled by the left-most rack shown in Figure 4, either manually or by the computer of Figure 5. This computer also performs the data acquisition and analysis.

The interface between the computer and the Zero Crossing Detectors for the measurement of Allan Variance is shown to the right of the computer. The remaining interfaces, patch panels, zero crossing detectors and high isolation distribution amplifiers are shown in Figure 6.

The capability of the chambers and the entire measurement system are shown in Table III and Table IV.

The shake table is shown in Figure 7. This table has the drive system separated from the actual table so as to minimize the magnetic field effects of the drive magnet on the vibration tests. A frequency multiplier from 100 MHz to 8.1 GHz is shown on the table and the interconnecting cables to and from the test equipment. These are special cables to reduce the table motion effects on the cable transmission which can also cause erroneous readings.

Frequency and phase measurements are made using the "picket fence" technique and by use of either identical units, a reference maser, or a high quality VCXO. The capability of these measurements are shown in Table III. One of the reference hydrogen masers in its double-walled magnetic shield as well as the AC Helmholtz coil are shown in Figure 8.

GENERAL TEST PROCEDURES

TEMPERATURE

The general procedure for temperature testing is shown in Figure 9. After stabilization at the nominal room temperature of 25 ^{o}C, the temperature is stepped to a higher temperature near the limit of 35 ^{o}C. The frequency is monitored while stabilization occurs at the higher temperature in order to determine the thermal time constant of the device under test. The remainder of the test consists of varying the temperature between the high limit near 35 ^{o}C and the low limit near 25 ^{o}C several times with a period of at least 5 thermal time constants to enable accurate frequency readings to be obtained. The frequency/temperature profile of Figure 9 is typical for most units. Some devices (usually hydrogen masers) have a double time constant caused by the electronics and the physics units and produce a frequency/temperature profile as shown in Figure 10. The dotted line is the response of the physics unit while the solid curve indicates the extra time constant due to the electronics. It is important that the initial determination of the unit time constant is made over a time long enough to allow both responses to be completed in order to evaluate the overall temperature sensitivity. The values of the time constants are also important since they determine the response to temperature variations of the environment.

MAGNETIC FIELD

The procedure for the magnetic field tests is demonstrated in Figure 11. In this case there is not a problem with time constants, but there are non-linearities and hysteresis. Small changes (± 200 mg) are made in the magnetic field, using an Helmholtz coil. Several cycles are made in order to determine the complete frequency vs magnetic field curve. Then the process is repeated using a larger field (± 600 mg), and lastly using a large field (± 1 g). The three curves thus obtained are the ones shown in Figure 11. Note that the shielding effect is smaller for small fields than for large fields and that, even for small field changes, there is a certain amount of hysteresis.

HUMIDITY

Humidity test are performed over a range of 11% to 90% relative humidity as shown in Figure 12. As in the temperature tests, there is usually a time constant involved in the response of the unit under test, particularly in the case of hydrogen masers. This time constant may be as long as several days, so it is important to be sure that the response has settled before continuing the cycle.

PRESSURE

The barometric pressure tests are normally run between nominal atmospheric pressure (14.7 psi) and ± 12 inches of water (±≈0.5 psi) as shown in Figure 13.

For most devices, the change in frequency due to a pressure change is nearly instantaneous, bu checks should be made with each device to ensure that this is the case.

ALLAN VARIANCE

The measurements of Allan Variance are performed at 100 MHz. Special low noise frequency multipliers are used when measuring devices which have only a 5 MHz or 10 MHz output. One of the reference masers is offset by 1 Hz and the output of the maser is mixed with the output of the device under test to produce a 1 Hz beat tone. This 1 Hz signal is used as the input to a Zero-Crossing-Detector which produces a fast rise, accurately controlled square wave. Using the "picket-fence" technique[2] in a computer system with eight input channels for Allan Variance, the data are collected and analyzed for inspection and archival storage.

POWER SPECTRAL DENSITY OF PHASE

Measurements of the power spectral density of phase are usually made by locking two identical units together in a narrow band phase locked loop, mixing the phase-shifted signals together in a double-balance Schottky diode mixer to produce a zero DC offset, amplifying the resultant signal in a special low-noise amplifier and analyzing the signal with a fast Fourier transform low frequency spectrum analyzer. When measuring single sources, a low noise VCXO is used as the reference. Test procedures have been written to handle nearly any possible configuration and problem that might arise.

FUTURE R&D AND IMPLEMENTATION REQUIREMENTS

While these requirements and test capabilities are for the present needs of the DSN, future requirements are much more stringent. These future requirements are outlined in Table V. The corresponding test capabilities are given in Table VI. Work is in progress, both in the R&D area and in the test area, to meet these future goals. The time scale for these performances and test capabilities lies between 1990 to 1995.

---

2 C. Greenhall, "A Method for Using a Time Interval Counter to Measure Frequency Stability", Proceedings of the 41st Annual Symposium on Frequency Control, 27-29 May, 1987, pp.126-130

JPL

TABLE I

Frequency and Timing System Requirements

Allan Variance	1 s	10^{-12}
	100 s	10^{-14}
	1000 s	2×10^{-15}
	3600 s	2×10^{-15}
	12 h	10^{-14}
	10 d	10^{-13}
Knowledge of vs UTC(NIST)	Frequency	3×10^{-13}
	Time	10 μs
Actual vs UTC(NIST)	Frequency	10^{-12}
	Time	20 μs
L(f) at 5 MHz	1 Hz	−118 dBc/Hz
	10 Hz	−126 dBc/Hz
	100 Hz	−139 dBc/Hz
	1 kHz	−139 dBc/Hz
	10 kHz	−139 dBc/Hz
	100 kHz	−139 dBc/Hz

JPL

TABLE II

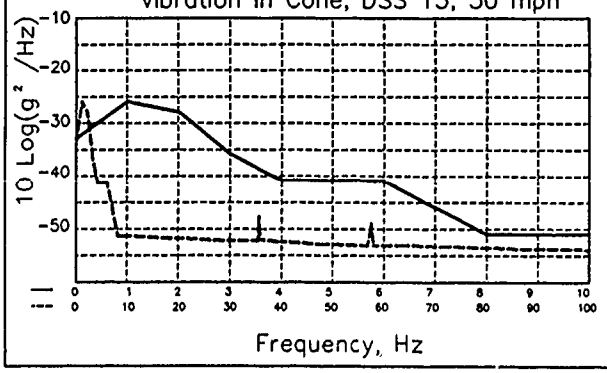
Environmental Parameters in the DSN

Parameter	Frequency Standards Room	Control Room	Antenna Cone
Temperature (°C)	±0.1	±3	±5
Vibration (g/√Hz)	3×10^{-3}	5×10^{-3}	50×10^{-3}
Humidity (% Rel)	10 to 80	10 to 80	10 to 80
Pressure (kPa, pk-pk)	±3	±3	±3
Magnetic Field (T, pk-pk)			
(static)	5×10^{-7}	20×10^{-7}	50×10^{-7}
(dynamic)	10×10^{-7}	200×10^{-7}	200×10^{-7}

JPL

Figure 1

Vibration in Cone, DSS 15, 50 mph



JPL

Figure 2

Required Vibration Sensitivity, 5 MHz VCXO in Antenna Cone

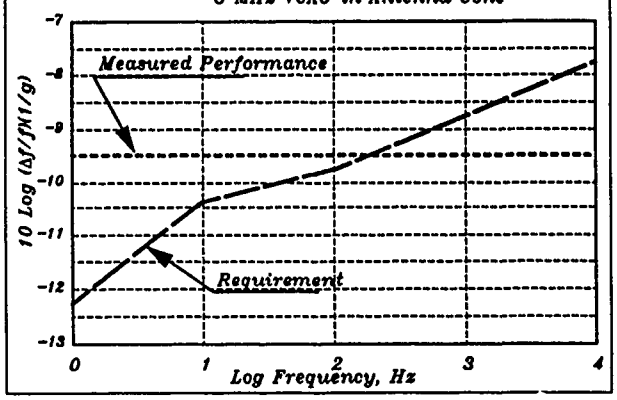


FIGURE 3

TEST CHAMBERS

FIGURE 4

TEST OPERATIONS FACILITY

FIGURE 5

TEST OPERATIONS CONTROL CENTER

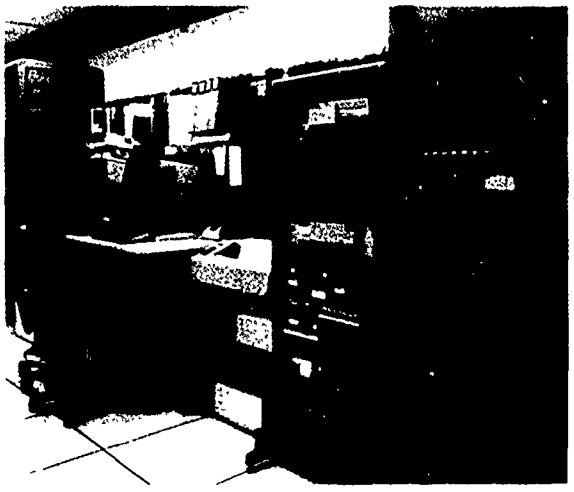
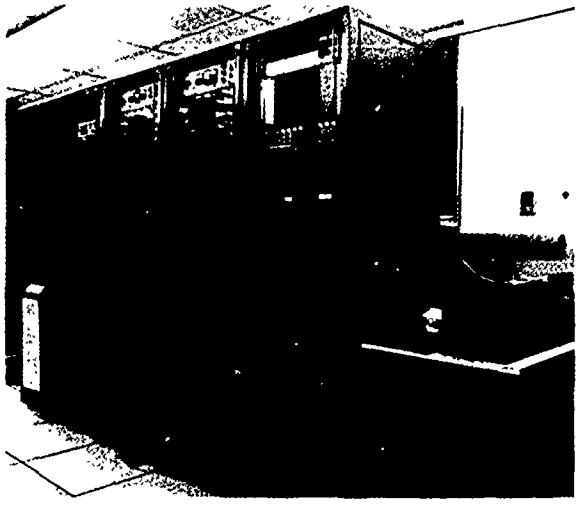


FIGURE 6

INTERFACE ASSEMBLY



JPL

TABLE III

JPL FREQUENCY STANDARDS LABORATORY
ENVIRONMENTAL TEST CAPABILITY

• TEMPERATURE	15 TO 35 °C ± 0.05 °C
• PRESSURE	13.7 to 15.7 PSI ±0.5" H_2O
• HUMIDITY (relative)	11 TO 90%, ± 5%
• MAGNETIC FIELD	
– STATIC	±2 GAUSS
– DYNAMIC (60 Hz)	±1 GAUSS
• VIBRATION	0 TO 1G(RMS), 0.1 TO 200 Hz (limited at low end by excursion limit of 3" pk–pk)

JPL

TABLE IV

JPL FREQUENCY STANDARDS LABORATORY

FREQUENCY AND PHASE MEASUREMENT CAPABILITY

- POWER SPECTRAL DENSITY OF PHASE, $S_\phi(f)$

  −165 dBc/Hz, 0.1 Hz to 100kHz

- ALLAN DEVIATION, $\sigma_y(\tau)$

  $10^{-14}/\tau$, 1 to 10^6 s at 100 MHz

  Floor at 10^{-17}

FIGURE 7

SHAKE TABLE

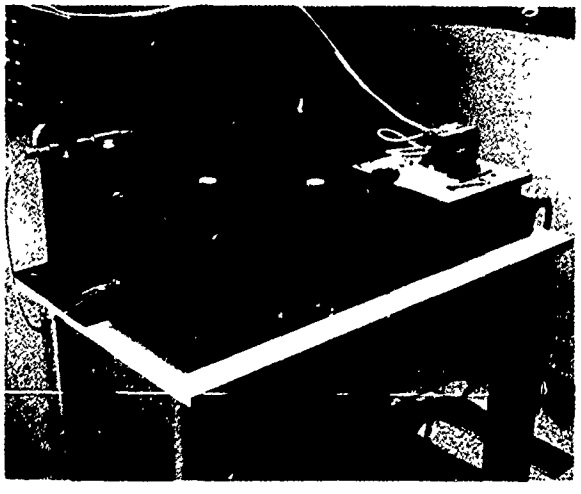


FIGURE 8

REFERENCE MASER & HELMHOLTZ COIL

FIGURE 9

Output Frequency vs Temperature

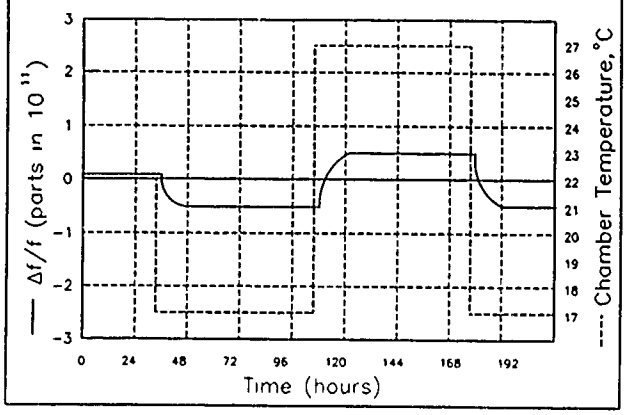


FIGURE 10

Output Frequency vs Temperature

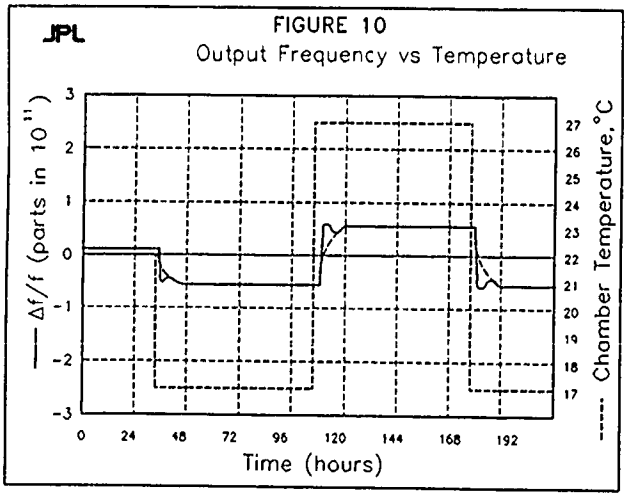


FIGURE 11

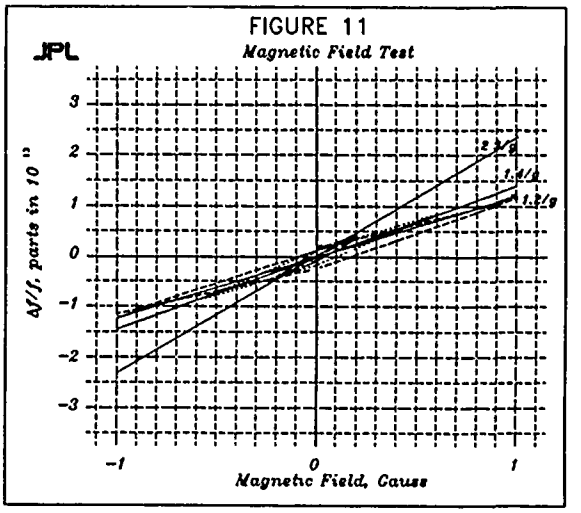


FIGURE 12

HUMIDITY TESTS

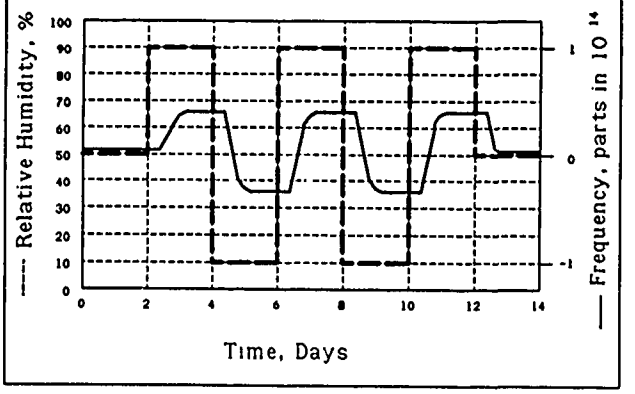


FIGURE 13

PRESSURE TESTS

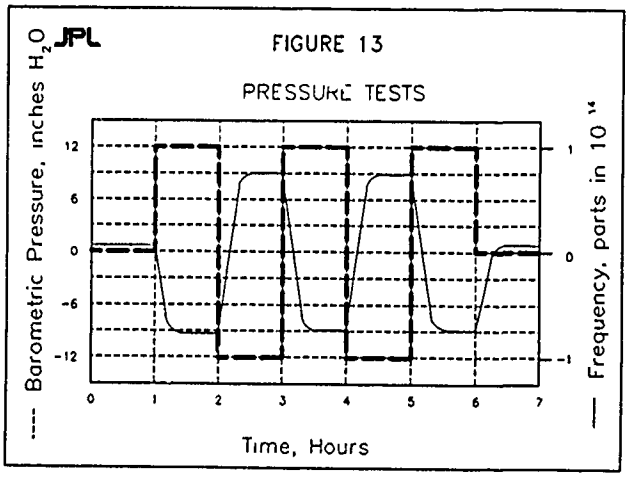


TABLE V

Future R&D and DSN Requirements

Allan Variance	1 s	10^{-13}
	100 s	10^{-15}
	1000 s	2×10^{-17}
	3600 s	10^{-17}
	12 h	10^{-17}
	10 d	10^{-17}
Knowledge of	Frequency	3×10^{-13}
vs UTC(NIST)	Time	10 μs
Synchronization	Time	<3 ns
Syntonization	Frequency	1×10^{-16}
L(f) at 5 MHz	1 Hz	−130 dBc/Hz
	10 Hz	−140 dBc/Hz
	100 Hz	−150 dBc/Hz
	1 kHz	−155 dBc/Hz
	10 kHz	−160 dBc/Hz
	100 kHz	−160 dBc/Hz

TABLE IV

JPL FREQUENCY STANDARDS LABORATORY

FUTURE MEASUREMENT CAPABILITY

- POWER SPECTRAL DENSITY OF PHASE, $S_\phi(f)$

  −170 dBc/Hz, 0.1 Hz to 100kHz

- ALLAN DEVIATION, $\sigma_y(\tau)$

  $10^{-16}/\tau$, 1 to 10^6 s at 100 MHz

  Floor at 10^{-19}

43rd Annual Symposium on Frequency Control - 1989

MEASUREMENT OF THE FREQUENCY RESPONSE OF CESIUM CLOCKS TO TEMPERATURE FLUCTUATIONS

Lee A. Breakiron
U. S. Naval Observatory, Time Service Department, Washington, DC 20392-5100

Abstract

An apparatus is described that is being used to calibrate the sensitivity of several cesium clocks to temperature fluctuations.

Introduction

One of the most critical environmental factors affecting the stability of a cesium-beam atomic clock is the ambient temperature. Analysis of USNO clock data has shown that control of the temperature range to within +1°C is sufficient for the attainment of an Allan deviation of a few parts in 10^{14}/day; cross-correlations with frequency were found only for temperature excursions that were significantly greater (1). Such excursions were not numerous enough to permit determination of accurate temperature coefficients (known to be approximately 2 parts in 10^{14}/°C) because, until recently, we were unable to induce such variations at will.

Equipment

Figure 1 is a schematic of a USNO clock vault with reasonably good temperature and humidity control that contains five cesium clocks. The air conditioner has been fitted with two thermostats set at temperatures differing by several degrees which are alternately selected by a microcomputer controller programmed, for reasons to be seen, to switch between the two thermostats after a period of time varying randomly from one hour to up to several days. The vault temperature and humidity and the clock frequencies are monitored hourly by an external computer.

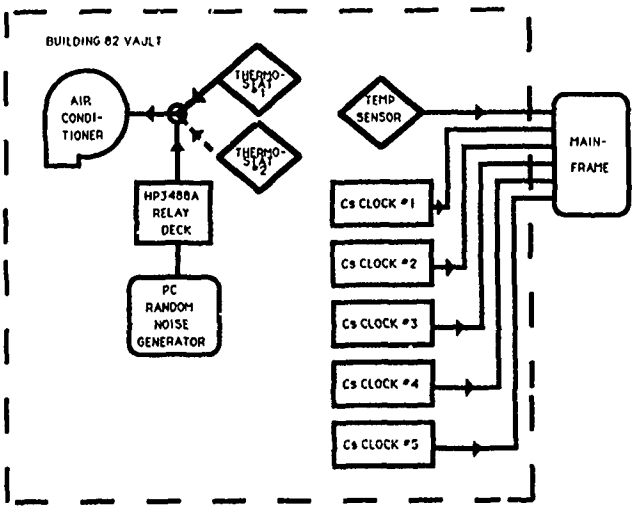


Figure 1.

Intended Analysis

The cross-correlation function between temperature change and frequency change yields the lag time of maximum correlation. If the correlation is sufficiently large, it should be possible to determine each clock's temperature coefficient accurately enough to correct its frequency for most of the temperature effects, though frequency steps due to other causes, such as random walk or human activity, would have to be taken account of.

However, the frequency change $\dot{f}$ may not just be caused by a temperature change $\dot{T}$ at one time τ, however delayed, but by a series of temperature changes over a range of lag times. In order to characterize such a dynamic system, we attempt to model its transfer function. Assuming the simplest form for the relation between the input $\dot{T}$ and the output $\dot{f}$, namely a linear filter:

$$\dot{f}_\tau = v_0 \dot{T}_\tau + v_1 \dot{T}_{\tau-1} + v_2 \dot{T}_{\tau-2} + \ldots$$
$$= (v_0 + v_1 B + v_2 B^2 + \ldots)\, \dot{T}_\tau$$
$$= v(B)\, \dot{T}_\tau$$

where the backshift operator B is defined by $B\,\dot{T}_\tau = \dot{T}_{\tau-1}$, $v(B)$ is the transfer function, and v_0, v_1, v_2, ... comprise the weighting or impulse function of the system (2).

If the series is stationary and produced by a linear system afflicted with white noise, the transfer function can be modelled by:

$$(1 - \delta_1 B - \delta_2 B^2 - \ldots)\, \dot{f}_\tau = (\omega_0 - \omega_1 B - \omega_2 B^2 - \ldots)\, \dot{T}_{\tau-b} + a_\tau$$

where δ_1, δ_2, ... and ω_0, ω_1, ω_2, ... are cross-correlation parameters, b is a delay parameter, and a is white noise. If the series involved is not stationary, it can be replaced with one that is through repeated differencing. Stationarity is indicated by rapid damping of the estimated autocorrelations. Usually only a few parameters need be solved for, and indeed the fewest necessary should be solved for if one is to maximize the accuracy of one's prediction. Such a parsimonious model most appropriate for one's data can be chosen using recursive estimation of the input's autocorrelation function and the cross-correlation function between the input and the output (2).

Considerable simplification of the model-identifying process can be achieved if the input to the system is white noise, for in this case the impulse response function is given by an orthogonal set of equations:

$$v_k = \frac{\rho_{\dot{T}\dot{f}}(k)\, \sigma_{\dot{f}}}{\sigma_{\dot{T}}}$$

for lags k = 0, 1, 2, ...; $\rho_{\dot{T}\dot{f}}(k)$ is the cross-correlation function between $\dot{T}$ and $\dot{f}$; and $\sigma_{\dot{T}}$ and $\sigma_{\dot{f}}$ are the standard errors of $\dot{T}$ and $\dot{f}$ respectively, i.e. the cross-correlation between the input and output of a linear, time-invariant system is proportional to the impulse resonse of the system. This need for white noise is the reason for the random noise generator in our apparatus. v_k can be used to make initial esimates of the cross-correlation and delay parameters, which

can then be improved recursively by nonlinear least squares. If the model is correct, the residuals a will be neither autocorrelated nor cross-correlated with the lagged temperature changes (2).

Initial Results

Preliminary data are given in Figure 2, in which the top curve depicts the temperature and the lower curves show the frequency changes. Ignoring the glitch at MJD 47657 (due to the hydrogen maser against which all the clocks were referred), some significant frequency changes are already evident. We are still optimizing the temperature range (too small in Figure 2) and the maximum time between temperature fluctuations (too long in Figure 2 to lie in the white noise frequency domain of cesium clocks). Planned hardware upgrades will improve our temperature and humidity control. It remains to be seen whether a single probe will suffice to characterize the temperature each clock is experiencing. After enough data has accumulated, the results of an analysis will be published. Considering the importance of humidity effects (3), such an analysis will probably involve solution for humidity coefficients.

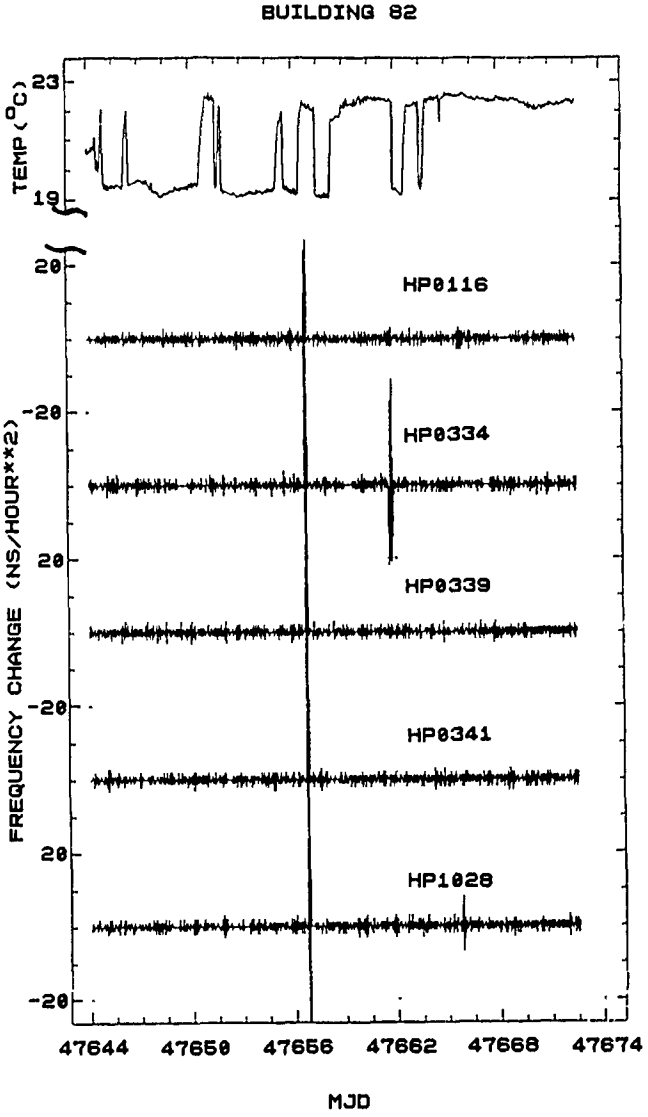


Figure 2.

Acknowledgements

I would like to thank Dr. Gernot M. R. Winkler of USNO for the idea underlying the analysis and William B. Dabney and Len M. McCubbin of USNO for technical assistance.

References

(1) L. A. Breakiron, "The effects of ambient conditions on cesium clock rates," in Proceedings of the 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Redondo Beach, CA, December 1-3, 1987, pp. 175-84.

(2) G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970, ch. 10-11.

(3) L. A. Breakiron, "The effects of data processing and environmental conditions on the accuracy of the USNO timescale," in Proceedings of the 20th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Tysons Corner/Vienna, VA, November 29-December 1, 1988, pp. 221-36.

43rd Annual Symposium on Frequency Control - 1989

ETR CESIUM MAINTENANCE AND REPAIR

Chauncey P. Dunn and James L. Wright
Computer Sciences Raytheon
P. O. Box 4127
Patrick Air Force Base, Florida 32925-0127

Abstract

This paper describes the procedures by which the Air Force Eastern Test Range's (ETR) Cesium Beam Frequency Standards are maintained and repaired. The ETR employs over 75 cesiums, distributed among various instrumentation sites throughout the Range, as primary frequency standards in Precise Time and Time Interval (PTTI) Systems. These cesiums perform critical mission support functions and must be maintained at a high state of readiness. The Air Force regulations authorizing local (ETR) repairs will be briefly examined. Maintenance routines and repair procedures, from fault determination to certification of repair, will be presented. Repair facilities and test equipment will be described. And finally, the authors will express a thought on what is needed from industry and government to facilitate frequency standard procurement and repair.

Introduction

The Eastern Test Range is a specialized complex of facilities available to military and commercial users for the test, checkout, and launch of missiles, rockets, and space vehicles. In addition, the ETR employs major stations for tracking space vehicles at the Cape Canaveral Air Force Station (CCAFS) in Florida, the Jonathan Dickinson Missile Tracking Annex (JDMTA) in Florida, the Antigua Air Station in the British West Indies, the Ascension Air Station on Ascension Island, and the U.S. Naval Ship, the USNS Redstone, which provides tracking coverage for areas otherwise not covered. Additionally, there are a number of other minor instrumentation sites and ships which support and/or are supported by the ETR. (See Figure 1.)

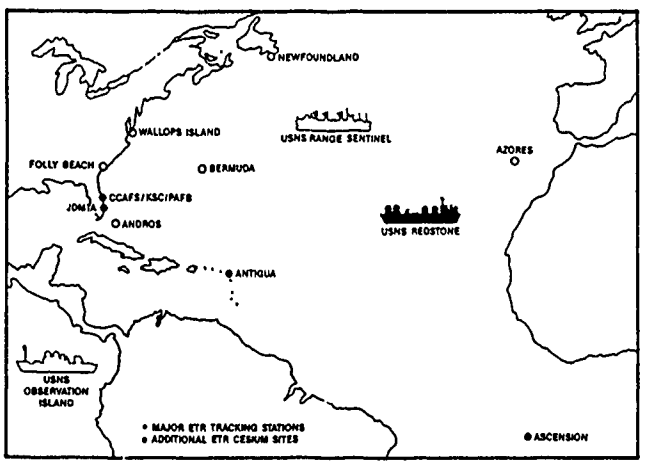


Figure 1. ETR Cesium Instrumentation Sites

There are many PTTI requirements at various locations, with varying accuracies, associated with each ETR operation. Many of the mission PTTI requirements occur during the period just before and after an actual launch (T minus 90 minutes to T plus 30 minutes). The Range also supports Space Transportation System (Shuttle) missions and others for the duration of their missions, which can last several weeks. Due to the many tests supported by the Range and the high accuracy requirements for time and frequency, the ETR PTTI systems must be maintained at a high state of readiness. The most stringent accuracy requirements met by the ETR PTTI systems are synchronization accuracy to the United States Naval Observatory (USNO) Master Clock of 100 nanoseconds and frequency accuracy of 1×10^{-13} (Tau = 1 day) for the major tracking station "Station Clocks".

To meet the Range's PTTI requirements, a hierarchy of Clocks has been established. At each major tracking station, a Station Clock is established and synchronized to both the Range Clock (located with the Station Clock at CCAFS) and the USNO Master Clock. PTTI requirements at outlying sites are supported by Site Clocks synchronized and syntonized to the local Station Clock. The Range Clock is configured with four cesiums and each Station Clock is configured with at least three cesiums. Three more cesiums are used for portable clock applications. The Range also supports a U.S. Navy system, which is configured with over 20 cesiums distributed among nine distant locations. Additionally there are two ships, which are supported by the Range and require cesium maintenance support. In all, with spares, there are over 75 Cesium Beam Frequency Standards supported by the ETR.

Air Force Maintenance Policy

The Air Force maintenance policy for Cesium Beam Frequency Standards is established in Air Force Regulations 66-1 and 74-2, and in Technical Orders 33K-1-100 and 00-20-14. In summary, these regulatory documents establish an Air Force Metrology and Calibration (METCAL) Program, which applies to all Air Force Precision Measurement Equipment (PME) or Test, Diagnostic, and Measurement Equipment (TDME), including cesiums. The Air Force METCAL program is established to ensure that PME used in Air Force programs is accurate, uniform and their accuracies can be traced to standards maintained by the National Institute of Standards and Technology (NIST) or the Department of Defense (i.e. the USNO Master Clock).

The Air Force Measurement Standards Laboratory (AFMSL), located at the Aerospace Guidance and Metrology Center (AGMC), Newark Air Force Station, Ohio, is responsible for maintaining Air Force primary standards, which are traceable to national standards. Precision Measurement Equipment Laboratories (PMELs) are established at selected facilities where a significant number of calibrations are required. The standards at the PMELs are traceable to the national standards, either directly or through the AFMSL.

Normally, all precise time support, maintenance, and calibration for Cesium Beam Frequency Standards is accomplished at either a local PMEL or the AFMSL. However, due to the timekeeping and maintenance capabilities developed in the Cesium Support Center, the ETR obtained a Precise Time and Frequency Self Sufficiency Certification from AFMSL in 1984 and now performs almost all of its maintenance on cesiums. The lone maintenance item not accomplished by the ETR is the replacement of the beam tube assembly which the ETR has been unable to obtain through the national supply system. The ETR now performs all precise time support, cesium maintenance, and calibrations that were formerly accomplished by the AFMSL and the nearby PMEL located at Patrick Air Force Base.

Cesium Support Center

In order to obtain and maintain the Self Sufficiency Certification and to ensure that the cesiums are maintained at a high state of readiness, the ETR has established a Cesium Support Center. This Cesium Support Center is currently distributed between two locations at CCAFS. All records are maintained and repairs per-

CH2690-6/89/0000-298 $1.00 C 1989 IEEE

formed in a timing repair facility. Final calibrations and certifications are performed at the Range Clock where the ETR's best PTTI System exists.

A record file is maintained for each cesium serviced by the ETR. This includes cesiums that have received maintenance attention at the Cesium Support Center, but that are not normally a Range responsibility. In each file are the date of arrival on the ETR, a record of its configuration, records of routine maintenance and repair, a history of the locations where the cesium has been, all other information particular to the cesium. In addition to these records, there are manuals, procedures, and instructions germane to cesium maintenance.

The cesium repair facility contains one large work bench, several cabinets of repair modules and parts, and nearly a full rack of test equipment (see Appendix A for a list of test and support equipment). Although several ETR employees are qualified to operate the repair facility, normally one technician is assigned full time to record keeping and cesium repairs. Generally, ETR employees who are required to repair cesiums have attended the one week Hewlett Packard Cesium Beam Frequency Standard Service Seminar and have extensive on-the-job training. This training, plus well-written manufacturer's service manuals, has proven to be sufficient to perform almost all repairs.

Routine Maintenance

For each equipment type, the ETR develops a Maintenance Procedure to detail periodic and as-required maintenance instructions, which verify pre-service or in-service performance. Typically, these documents provide schedules and instructions for routine maintenance and operational checks derived from the manufacturer's manuals and operators' experience. All routine maintenance for ETR cesiums is performed in accordance with these Maintenance Procedures.

Monthly, each operational or stand-by cesium is checked via front panel monitor meters and by measuring output signal levels. All readings are recorded and initially analyzed by site personnel. Then the readings are forwarded to the Cesium Support Center for final analysis and filing. Additionally, whenever a Portable Clock Team from the Range Clock visits a remote cesium, the same monthly routines are performed by these experienced personnel. At two (and eventually all) Stations Clocks, there is a computer-based PTTI Monitor and Control System which automatically makes at least daily synchronization measurements between all local operational cesiums and other time keeping sources (i.e. GPS, LORAN-C, and Transit receivers). All of this data assists Operations and Cesium Support Center personnel in determining the current and near future health of on-line units. Specifically the Ion Pump and Second Harmonic readings are good indicators of predictable cesium performance.

Repair Maintenance

The Cesium Support Center arranges for replacement units at locations where there are failed or suspect cesiums. Most replacement units are fully configured (high performance beam tube, internal battery, and clock) to meet all the requirements of any Range cesium and thus are loaned to instrumentation sites until their unit is repaired and returned. Most repairs are accomplished using the resources in the Cesium Support Center and take less than one day (not including testing and verification). However, a significant number of failed units require replacement beam tubes. These units are not repaired on the ETR but are returned to the vendor for required beam tube replacements.

Test and Verification

All new and repaired cesiums undergo extensive testing and verification prior to being placed in an operational system. Initially, the cesium is evaluated using the Hewlett Packard model 5390A Frequency Stability Analyzer System. This system produces Sigma Y of Tau graphs for averaging periods of 1×10^{-3} to 1×10^{4} seconds. Long term testing is then conducted comparing the frequency drift of the unit under test with LORAN-C. This testing is conducted for at least one week and, if satisfactory, the cesium is then transported to the Range Clock for final verification. Upon its arrival at the Range Clock, a final degaussing of the beam tube is conducted. Then with the aid of the PTTI Monitor and Control System, the unit under test is compared against the Range Clock and other synchronization sources through automated measurements and phase comparison chart recordings for a second week. If warranted, the C-field is adjusted, and the testing and verification process repeated. Because many of the cesiums must operate to tight tolerances in extremely remote areas, every effort is made to ensure that certified units are as stable and accurate as possible. Typically, the cesiums are returned to operational systems with less than 40 nanosecond per day drift rates.

Problem and Recommendation

In recent years there have been several new manufacturers that have entered the Cesium Beam Frequency Standard market. While most manufacturer's specifications are similar, few units perform equally. Almost each unit requires a unique support program. This problem is not limited to only cesiums, since the government procurement system almost guarantees that multiple procurement activities for similar units will yield a multitude of different manufacturers and model numbers.

A government specification for cesiums should be developed to ensure that all compliant cesiums procured are compatible and maintenance support modules obtained from the supply sources are interchangeable.

Appendix A

Cesium Support Center
Test and Support Equipment

Nomenclature	Manufacturer	Model
Degausser	Hewlett Packard	10638A
Signal Generator	Hewlett Packard	8660C
Spectrum Analyzer	Hewlett Packard	141T
Function Generator	Hewlett Packard	3325A
Audio Oscillator	Hewlett Packard	204CD
Sweep Generator	Hewlett Packard	8620C
RMS Voltmeter	Hewlett Packard	3400A
DC Milliampmeter	Hewlett Packard	428B
RF Power Meter	Hewlett Packard	432A
Vector Voltmeter	Hewlett Packard	8405A
Time Interval Counter	Hewlett Packard	5370A
Counter	Hewlett Packard	5334A
Selective Voltmeter	Hewlett Packard	3581C
Power Meter	Hewlett Packard	431B
Strip Chart Recorder	Hewlett Packard	7132A
Loran-C Receiver	Austron	2100
Oscilloscope	Tektronix	475
Storage Oscilloscope	Tektronix	2430A
Standby Power Supply	Hewlett Packard	5089A
Standby Power Supply	Hewlett Packard	5085A
Terminators and Attenuators	Miscellaneous	
Frequency Stability Analyzer consisting of:	Hewlett Packard	5390A
Instrumentation Controller	Hewlett Packard	9825A
Electronic Counter	Hewlett Packard	5345A
Measurement Storage Unit	Hewlett Packard	5358A
Timing Generator	Hewlett Packard	59308A
Test Tone Generator	Hewlett Packard	10831A
Oscillator	Hewlett Packard	105H
Mixer/IF Amplifier	Hewlett Packard	10830A
Printer/Plotter	Hewlett Packard	9871A

43rd Annual Symposium on Frequency Control - 1989

PRECISION TEMPERATURE TEST STATION for QUARTZ CRYSTALS

Robert Kinsman and Donald Ryback

Motorola, Inc.
1301 E. Algonquin Rd.
Schaumburg, IL 60196

Summary

A new temperature test station has been developed for testing moderate quantities of crystal resonators. The primary function of this station is to accurately measure the frequency versus temperature characteristics of quartz crystals of all types. Additional features include curve fitting routines to characterize the frequency-temperature function and the derivation of the crystal's equivalent circuit parameters as a function of temperature. In this system all electrical tests are performed by a network analyzer which interfaces directly with a computer. All functions of the network analyzer are controllable through the computer interface. This feature allows for such functions as testing parts over a wide range of power levels and for testing parts with different frequencies in the same batch. Crystal measurements are performed with an HP 3577A network analyzer using the general methods outlined in EIA Crystal Measurement Standard 512.

This paper describes the equipment used in this station and provides a general description of the custom software used to operate the system. Estimates of system accuracy and repeatability are provided. Test results for a variety of crystals are shown.

Introduction

A second generation crystal measurement system is now operational in Motorola's Components Research Laboratory. The primary function of this system is to accurately measure the frequency- versus-temperature characteristics of quartz crystals of all types. Additional features include curve-fitting routines to characterize the frequency-temperature function and the derivation of the crystal's equivalent circuit parameters as a function of temperature. In the first generation system developed in 1981-82, measurement of the crystals was performed with a group of instruments including a frequency synthesizer, vector voltmeter, directional coupler, digital voltmeter, and frequency counter which were computer controlled through a multi-programmer interface. In the new system all electrical tests are performed by a network analyzer which interfaces directly with the computer. All functions of the network analyzer are controllable through the computer interface. This feature gives the new test system considerably more flexibility than the old system and allows for such functions as testing parts over a wide range of power levels and testing with a variety of frequencies in the same batch. The capabilities of the new system are listed in Table 1.

System Description

A block diagram of the new system is shown in Figure 1 and the equipment is listed in Table 2. All equipment functions are controlled by the computer and no preliminary adjustments are required except for turning on the power. All test data is processed by the computer and then stored on the hard disk. The processed data may be displayed on the screen or dumped to the plotter. The temperature chamber used in the current system is an analog type unit which has been modified so that the temperature can be selected by a switchable resistance card in the multi-programmer. The chamber may be operated over a range of -65° to +140°C in steps as small as 0.1°. Newer oven controllers are now available which will interface directly with the computer. The oven temperature is monitored by a precision quartz thermometer. The crystal measuring system measures the reflection coefficient of the crystal under test at

CH2690-6/89/0000-300 $1.00 C 1989 IEEE

Frequency range: 100 kHz to 200 MHz
Temperature range: -65° to +140°C
Temperature increment: 0.1°C min.
Temperature measurement accuracy: 0.04°C
Reference frequency accuracy: set by external standard
Measured crystal parameters as a function of temperature:
Resonance frequency of series arm: f_s
Frequency of zero reactance: f_r
Motional parameters: L_1, R_1
Equivalent series resistance at f_r: R_r
Static capacitance: C_o (near resonance)
Crystal drive level: +2 dBm max. (dependent on crystal R_1)
Outputs: All crystal parameters versus temperature. The frequency versus temperature data may be fitted to a curve of maximum order 10. The computed curve, curve coefficients and deviation between the computed curve and actual data can be outputted.
Capacity: 10 crystals
Test time: Typically 20 minutes per temperature test point after reaching initial temperature.

Table 1. System specification

a set of points near its resonance frequency and computes the series resonance frequency and equivalent circuit parameters from this data. The use of the reflection coefficient measurement technique offers the advantages that only a single coaxial line is required between the network analyzer and the crystal, and a simple error correction routine can be used to compensate for the cable and test fixture characteristics. In our system, ten crystals can be tested at one time. This is accomplished by running ten cables into the oven and using an external, 10-position, coaxial switch between the network analyzer and the ten cables. The switch is operated by the computer through the multi-programmer. The system could be expanded by adding more switches. The test program occupies approximately 100 kbytes of memory. For a test run additional memory capacity of 64 bytes per data point is required.

System Operation

Temperature Control

The timing of events during a temperature run is keyed to the operation of the temperature chamber. The desired temperature excursions are entered into the computer prior to starting a test run. In the entry process a pair of temperature end points and the temperature increment is specified for each leg of the total run. Several entries can be made to make up the total temperature profile. For example, a single run might be specified as; end points = +85°,-40°C, increment = 5°C. A multiple step run might be; Step 1: end points = +25°,+85°C, increment = 10°C; Step 2: end points = +85°,-55°C, increment = 5°C; Step 3: end points = -55°,+25°C, increment = 10°C. The soak time after reaching a specified temperature and the temperature window are also specified by the user. A typical soak time would be 10 minutes. For any temperature step greater than 10°, an additional 2 minutes is automatically added to the specified soak time. The temperature window defines the allowable deviation from the specified test temperature. The minimum window size is +/- 0.2°C.

During the test run the computer selects a temperature value and then switches in an appropriate resistance value from the resistance card in the multi-programmer. This resistance is connected to the temperature chamber and makes up one arm of a temperature sensitive bridge. The temperature profile versus resistance for the temperature chamber is stored in the computer. The computer monitors the temperature in the chamber by use of the quartz thermometer. When the temperature rate of change drops below 0.02°C/minute, a soak time period will be initiated if the chamber temperature is within the specified temperature window. If the temperature is

outside this window, a new resistance value will be selected which corresponds to the needed temperature shift. After the desired temperature is achieved and the soak time period has passed, all crystals are measured and the data is stored along with the measurement temperature. After all units are measured, a new temperature value is selected and the cycle is repeated. After all measurements are completed at the final temperature, the chamber returns to room temperature. A flow diagram for the computer software controlling this procedure is shown in Figure 2.

Crystal Measurement

The crystal measurements are performed by a network analyzer which includes an S-parameter test set. The measured data consists of a set of reflection coefficient readings taken in the vicinity of the crystal's resonance frequency. This data is corrected to compensate for the impedance characteristics of the measuring fixture and then the corrected data is used to compute the crystal parameters of interest.

Before starting a test run, values for each crystal's nominal frequency (f_o), approximate Q (Q_e), and desired drive level are entered into the computer. Reflection coefficient error correction data is stored in the computer for each of the ten test ports. If any changes are made in the test fixtures or cabling a calibration run can be performed to obtain new error correction data. After the crystals are installed in the test ports, a search routine is initiated to locate each crystal. An initial frequency sweep equal to 50,000 f_o/Q_e is used and the group delay is measured. A delay peak occurs near the crystal resonance. The sweep is carried out from both directions and two peaks will be found; one above and one below the crystal resonance frequency. These two peaks are then used to define a new, narrower frequency sweep. From this sweep data the crystal parameters can be calculated and the final sweep is adjusted to exactly cover the 3dB bandwidth of the crystal (3dB bandwidth = f_o/Q_{actual}). In the final sweep the drive level is reset and 51 data points are taken. The sweep speed and settling time are automatically adjusted as a function of the sweep bandwidth (the settling time, the time to measure each data point, must be increased for very high Q crystals). These sweep speed and settling time values are then maintained throughout the temperature run.

After the sweep parameters are set for all crystals under test, the temperature chamber is activated and starts moving to the initial test temperature. The crystals are continuously monitored during the temperature excursion and values are calculated for each crystal's series frequency, 3dB bandwidth, series resistance, motional inductance, shunt capacitance, and drive level. Slight adjustments are made in each successive sweep based on the data obtained from the previous sweep. During this period the computer screen displays each crystal's frequency, resistance, and inductance as they are updated. A copy of a typical screen display is shown in Figure 3. When the test temperature is reached and the chamber has stabilized, a final sweep is taken, all of the crystal parameters are calculated, and the data is stored on the hard disk. Eight parameters are saved for each crystal at each test temperature. These parameters are: drive level, f_s, R_1, L_1, C_o, f_r, R_r, and the test temperature. After all parts are measured, the temperature chamber moves on to the next test temperature. At the completion of the run all data is stored and is keyed to the test date.

Crystal Modeling

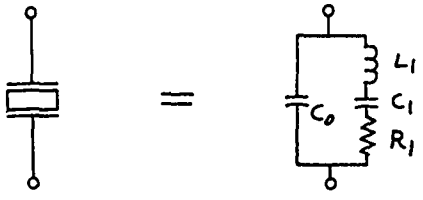
The crystal equivalent circuit model and the method used to derive its parameters are in agreement with the new measurement standard EIA-512, "Standard Methods for Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units, 1kHz to 1GHz." The single-port equivalent circuit as shown here includes the motional parameters L_1, C_1, and R_1 and the shunt capacitance C_o.

The resonance frequency of the series resonator is denoted as f_s where

$$f_s = 1/(2\pi\sqrt{L_1C_1})$$

For oscillator applications the frequency of zero reactance is often of interest. This frequency is denoted as f_r and the equivalent resistance at this frequency is denoted as R_r.

The admittance locus $Y = G + jB$ of the crystal equivalent circuit is shown in Figure 4. The locus describes a circle in the vicinity of the resonance frequency which is tangent to the imaginary (B) axis and may or may not intersect the real (G) axis.



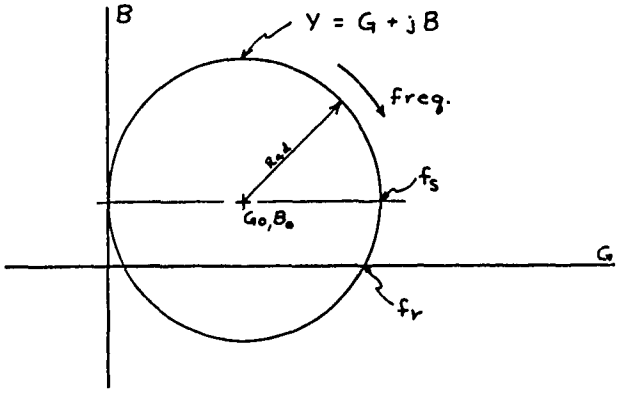


Figure 4. Admittance locus for a crystal resonator

The general equation for the circle is

$$(G - G_o)^2 + (B - B_o)^2 = Rad^2$$

where G_o, B_o are the coordinates of the center of the circle and Rad is the radius. The locations of the series resonance frequency f_s and the frequency of zero reactance f_r are shown.

f_s corresponds to the frequency at which $B = B_o$

f_r corresponds to the lower frequency at which $B = 0$

For the crystal equivalent circuit model

$$R_1 = 1/(2\ Rad)$$

$$C_o = B_o/2\pi f_s$$

$$R_r = 1/(G_o + \sqrt{Rad^2 - B_o^2})$$

The inductance L_1 is derived from the slope of the series resonator arm at frequency f_s.

$$4\pi L_1 = dX(f)/df\Big|_{f = f_s}$$

where $X = -(B - B_o)/[G^2 + (B - B_o)^2]$

The actual computation process is carried out as follows:

The reflection coefficient S_{11} is measured at 51 frequencies equally spaced across the 3dB bandwidth of the crystal (ie. $G \geq G_o$). These values are then error corrected to compensate for the cables and test sockets in the measuring system. The admittance at the 51 points is then calculated from the corrected reflection coefficient values.

The admittance Y is

$$Y = G + jB = (1 - S_{11})/50(1 + S_{11})$$

These data points cover the right one-half of the admittance circle shown in Figure 4. A circle is fitted to this data using a linear-least-squares algorithm. At this point a value for R_1 is calculated from the parameters of the fitted circle. If R_1 is greater than 10kohm, the calculation procedure is stopped and the part is indicated as lost. On the next test of this part the frequency search routine will be initiated. If the crystal is lost at any time during the final two minutes of the temperature soak, it will not be tested at that particular test temperature. If the R_1 value is less than or equal to 10kohm the data is then "smoothed" by shifting all points along a radial vector through the center of the circle (G_o, B_o).

Thus, for a specific data point $Y_1 = G_1 + jB_1$, the angle θ of the vector will be

$$\theta_1 = \tan^{-1}[(G_1 - G_o)/(B_1 - B_o)]$$

The coordinates of the smoothed point $Y_1' = G_1' + jB_1'$ will be

$$G_1' = G_o + (Rad)(\cos\theta_1)$$

$$B_1' = B_o + (Rad)(\sin\theta_1)$$

Crystal parameter values are calculated from the smoothed data as follows:

$$R_1 = 1/(2\ Rad)$$

An approximate value for C_o can be derived from the relationship

$$C_o = B_o/2\pi f_s$$

However, for high Q crystals, the value of C_o is not precise enough to provide an accurate value for C_o. Better results can be obtained by measuring the admittance of the crystal at a frequency well away from its resonance. In our system this measurement is made at $0.95\ f_s$.

To find L_1 and f_s, the circle is shifted so that $B_o = 0$ thus effectively subtracting out C_o. Then the reactance X_s of the series arm is found where

$$X_s = -(B' - B_o)/[(G')^2 + (B' - B_o)^2]$$

This data is fitted to a straight line

$$X_s(f) = m(f - f_c) + a$$

where f_c = center frequency of the sweep.

Then

$$L_1 = m/4\pi$$

$$f_s = f_c - a/m$$

The frequency f_r is defined as the frequency at which the reactance of the crystal is zero. In the admittance diagram of Figure 4, this occurs where the circle Y(f) first intersects the real axis. For higher overtone crystals the circle may not cross the real axis and there is no f_r. Therefore, initially a check is made to see if

$$B_o < Rad$$

If this relationship is not true, no further check is made. If it is true, the reactance X_r is computed where

$$X_r = -B'/[(G')^2 + (B')^2]$$

This data is fitted to a second order curve where

$$X_r(f) = A_0 + A_1(f - f_c) + A_2(f - f_c)^2$$

Then

$$f_r = -A_1/2A_2 + \sqrt{(A_1/2A_2)^2 - A_0/A_2} + f_c$$

The value of R_r is calculated from the Y circle constants where

$$R_r = 1/(G_o + \sqrt{Rad^2 - B_o^2})$$

The crystal drive level (power dissipation) is calculated at frequency f_s. The crystal drive level (pwr) is

$$Pwr = (V_s^2/R_1)\left|Z_x/(50 + Z_x)\right|^2$$

where

V_s is the Thévenin equivalent source voltage (source impedance = 50 ohms)

R_1 is the crystal motional resistance

Z_x is the crystal impedance at f_s.

The source voltage is held constant during a frequency sweep, but is updated for each new sweep to meet the prescribed crystal drive level. The maximum power available from the network analyzer with the S-parameter test set is approximately 2 mW.

Data Processing and Output

At the completion of the test run, all data is stored on the hard disk and is referenced to the test date. The standard output is a plot of frequency versus temperature for each individual crystal. The frequency f_s or f_r can be fitted to a curve of up to the 10th order. AT-type crystals are normally fitted to a 3rd or 4th order curve. The printout consists of the computed response curve plus a residual curve showing the difference between the computed curve and the actual data. All curve constants and an inflection temperature (where applicable) are listed along with the drive level and temperature increment. If requested the actual frequency data is plotted as discrete data points. Also, data can be plotted for R_1 or R_r and L_1 or C_o as a function of temperature. A copy of a typical printout for an AT-cut crystal is shown in Figure 5 and for an SC-cut in Figure 6.

Accuracy

The accuracy of the crystal measuring procedure is enhanced by the use of multiple data points which are fitted to appropriate functions. The improvements resulting from these methods have been extensively investigated by Smith of Bell Laboratories and some of his results will be presented here [1,2]. His conclusions are based on certain assumptions as to the equipment and the devices under test:

1. Equipment accuracy estimates assume that the network analyzer, cables and test fixture are properly calibrated to determine the appropriate error-correction functions.

2. The crystal resonant mode under test is adequately isolated from other modes.

3. Non-linear amplitude effects in the crystal can be neglected across its 3dB bandwidth.

4. Variations due to ambient temperature can be neglected.

The HP3577A network analyzer specification lists a dynamic accuracy of +/- 0.02dB magnitude and +/- 0.2° phase in the test levels most commonly used for crystal measurement. Typical tolerances are purported to be approximately one-half the maximum values. The specified tolerances translate into a maximum area of uncertainty of +/- 0.35% of the magnitude for reflection coefficient measurements. Analysis of these tolerances indicates that impedance measurement accuracy is best for impedance magnitudes near 50 ohms and deteriorates for impedances away from 50 ohms.

Two of the crystals used as examples in Reference 1 were analyzed. In the reference these crystals were measured and modeled and then a 0.5% random error noise was added to the reflection-coefficient and error-correction data. Equivalent circuit parameters were then calculated from the disturbed data. This process was repeated 100 times and the following results were obtained: For a 3.15 MHz

fundamental mode crystal with a Q of 35,000, maximum errors of 0.1 ppm were obtained for f_s, 0.5% for R_1, and 1% for L_1. For this crystal, the maximum error in admittance parameters due to network analyzer maximum tolerances ranges from 0.19 to 0.47%. For a 40.68 MHz third overtone crystal with a Q of 75,000, maximum errors of 0.06 ppm were obtained for f_s, 1.5% for R_1, and 1% for L_1 and the maximum error in admittance parameters due to the network analyzer ranges from 0.34 to 0.37%.

The typical accuracy tolerances for the network analyzer fall well within the 0.5% limits used for the disturbed data. From the parameters derived from the disturbed data, several conclusions may be reached: The worst case frequency accuracy is well within 1×10^{-7} which is more than adequate for characterizing frequency versus temperature. The frequency accuracy is a direct function of crystal Q, but this should not be a limiting factor in evaluating quartz crystals. Worst case errors for R_1 and L_1 are of the order of 1%, which is adequate for most test purposes. Accuracy can be substantially improved by averaging the results of several sequential tests.

Conclusions

A second generation precision crystal measurement system is now operational. The new system takes advantage of the capabilities of the latest generation of network analyzers which are fully digital in operation and interface directly with a computer. Accuracy is enhanced by taking many data points which are "smoothed" and fit to appropriate functions. The network analyzer replaces a combination of several analog and digital instruments to provide a more reliable system which offers considerably more flexibility than the previous system. The new system can test any type of crystal over the range of 100kHz to 200MHz including high overtone types which may not have a zero-reactance frequency. The new system completely characterizes the crystal's equivalent circuit and allows for testing at a wide range of drive levels.

References:

1. W.L. Smith, "An Overview of a Proposed Standard Method for the Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units up to 1GHz", Proc. 6th Quartz Devices Conf., August, 1984, Electronics Industry Association, (1 - 41)

2. W.L. Smith, "Electronic Industries Association Standard 512: Some Further Discussion and Comment", 7th Quartz Devices Conf., August, 1985, Electronics Industry Association, (13 - 51)

Computer: HP Model 216S with 512 kbyte memory and Basic 5.0 software (has complex numbers capability). Also includes an HP 98635A floating point math. card and an HP 98257A one Mbyte RAM board.
Hard disk: HP 9133H (20 Mbyte)
Network Analyzer: HP 3577A with HP 35677A S-parameter test set
Thermometer: HP 2804A quartz thermometer
Multi-programmer: HP 6940B with HP 59500A multi-programmer interface
Temperature chamber: Tenney Jr. (modified for external temperature control)
Coaxial Switch: Matrix 7710-DR-50 (10 position)
Plotter: HP 7550A

Table 2. Equipment list

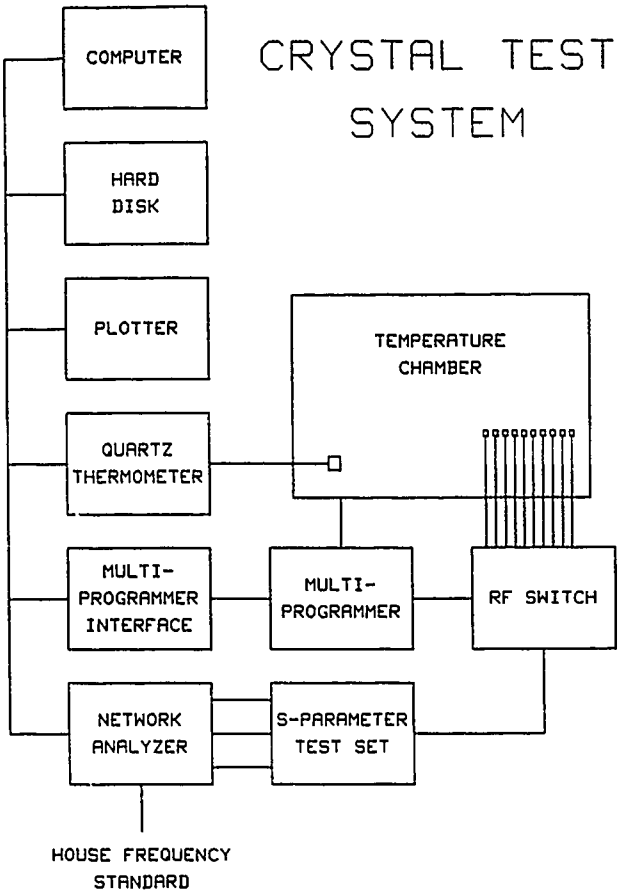


Figure 1. System block diagram

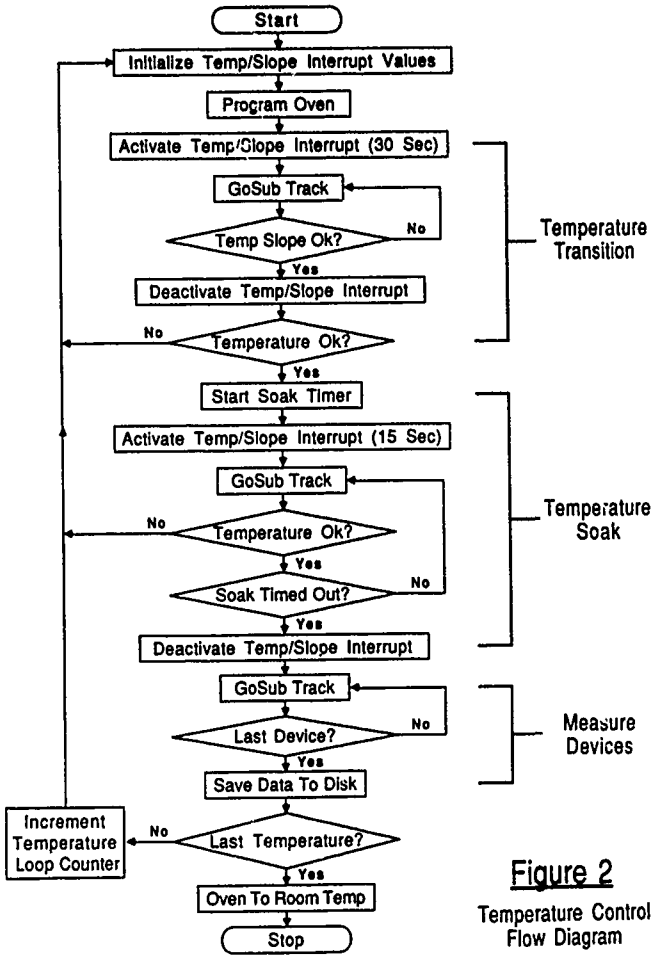


Figure 2
Temperature Control Flow Diagram

```
:::::  Temperature Information  :::::          Start  -  9 Apr 1987  18:56:06
At Temperature #  4 of 121 Points        (Est) Finish - 11 Apr 1987  08:08:00
Desired =  82.000   Slope =  -.034             Elapsed Time     -      1:02:21
Current =  82.050   State = Soaking           Time Until Measure -    0:09:41

Port #      Frequency       Resistance      L1
   1      4 999 981.105        102.5      7.315  H
   2      1 535 606.691        45.95      1.276  H
   3     14 559 882.278        19.40      8.503 mH
   4     16 057 139.900        18.92      10.27 mH
   5     16 041 800.785        13.35      10.69 mH
   6     16 067 839.193        29.09      10.38 mH
   7     16 058 453.304        16.18      10.44 mH
   8     16 049 432.838        34.65      10.58 mH
   9    154 083 987.857        109.8      1.157 mH
  10     43 700 156.719        25.60      10.23 mH
```

```
:::::  Temperature Information  :::::          Start  -  9 Apr 1987  18:56:06
At Temperature #  4 of 121 Points        (Est) Finish - 11 Apr 1987  08:08:0C
Desired =  82.000   Slope =  0.000            Elapsed Time     -       1 12:27
Current =  81.957   State = Measuring

Port #      Frequency       Resistance      L1
   1      4 999 981.108        102.5      7.315  H
   2      1 535 606.669        45.92      1.276  H
   3     14 559 882.090        19.41      8.477 mH
   4     16 057 137.877        18.93      10.40 mH
   5     16 041 799.631        13.37      10.69 mH
   6     16 067 845.343        29.03      10.38 mH
   7     16 058 452.275        16.21      10.46 mH
   8     16 049 431.721        34.68      10.57 mH
   9    154 084 027.642        109.7      1.155 mH
  10     43 700 153.839        25.60      10.22 mH
```

Figure 3. Screen display during test

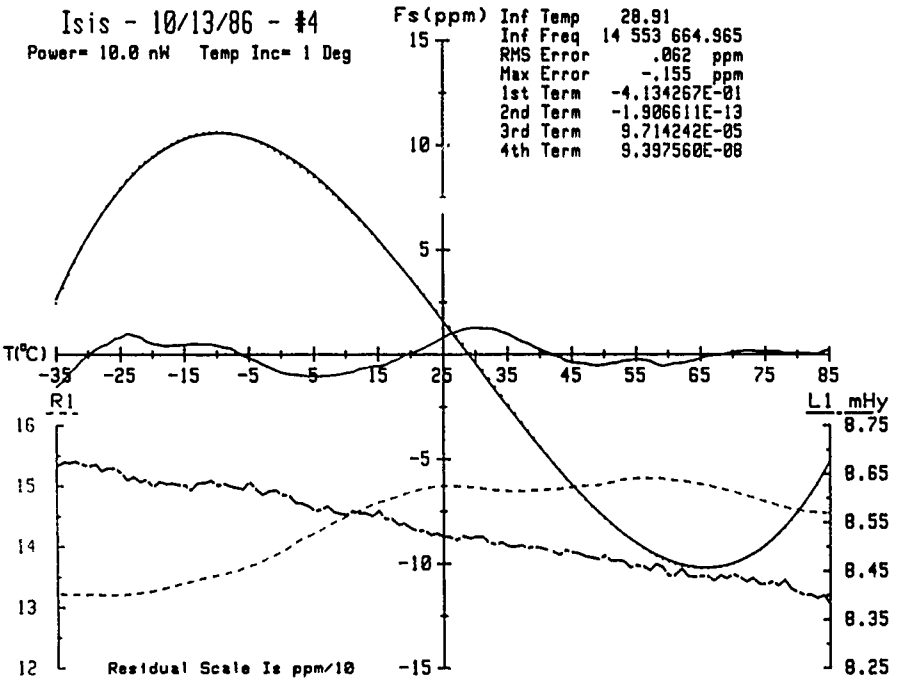


Figure 5. Test results for an AT-cut crystal

Figure 5. Test results for an AT-cut crystal

Figure 5. Test results for an AT-cut crystal

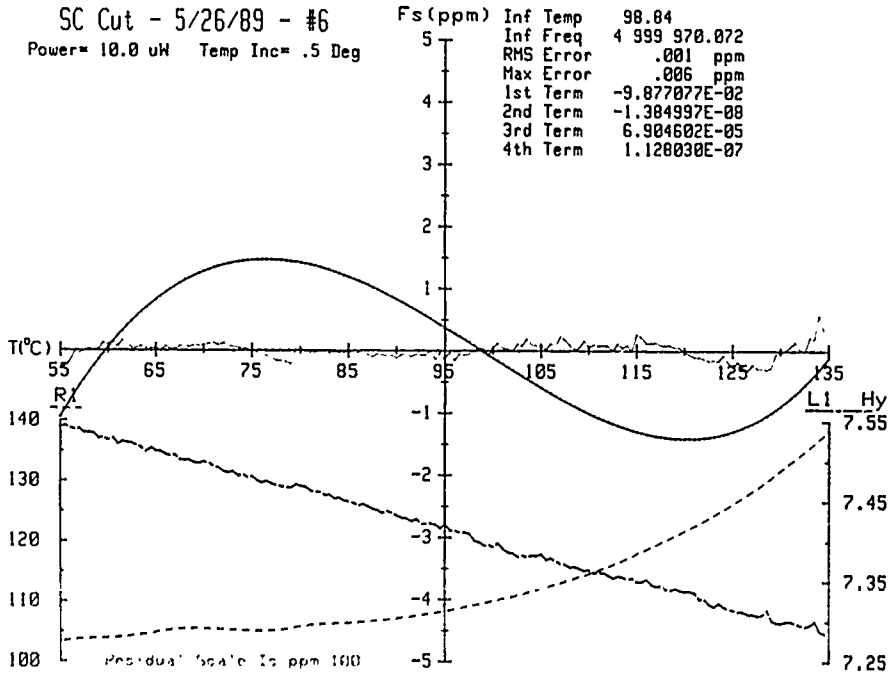


Figure 6. Test results for an SC-cut crystal

43rd Annual Symposium on Frequency Control - 1989

STATISTICAL ANALYSIS OF SSB PHASE NOISE TEST RESULTS

Robert R. Zeigler Jr.
Operations Manager
Piezo Crystal Company

ABSTRACT

Statistical analysis of recent phase noise measurement results for a production lot of 100 MHz to 111.25 MHz oscillators using 3rd overtone SC cut quartz crystals will be reported. The test results are from a lot of 262 units of 10 different frequencies. The results will show the statistical variations of the close-in phase noise measurements of both the crystal and the completed oscillator. Phase noise is reviewed at 10 Hz to 40 MHz offset from the carrier and analyzed in detail at 100 Hz offset where crystal noise is critical to the noise performance of the oscillator. The final results show a significant improvement over previously reported production run data. Measured noise readings as low as –136 dBc/Hz at 100 Hz with limited process capability greater than –140 dBc/Hz is shown.

INTRODUCTION

Low noise quartz crystal oscillators in the 100 MHz region have been the mainstay of UHF and microwave synthesizers by providing a stable internal reference. These sources have either been phase locked to a lower, more stable reference, (i.e., a 5 MHz crystal or atomic low aging system clock) or free running. The 100 MHz region represents the highest frequency range where crystals can be economically produced to achieve a compromise of desired characteristics. High Q, low long term aging, low gravitational sensitivity, and ease of manufacturing are normally the most important crystal parameters specified.

The data presented will show not only the specification limit, but it will be clear as to the statistical chances of meeting the limit in a production environment using commercially available phase noise test equipment. If the system engineer or designer of higher order assemblies is aware of these statistics, each is better able to evaluate his system noise budget when the system is produced in quality. This paper will show plots of the distribution of measurements with calculations for $\bar{x}$ (mean average value) and sigma (standard deviation), 2 sigma, and 3 sigma limits.

Statistical process control (SPC) was introduced by Dr. Deming and utilized with great success by the Japanese. Manufactures in the United States are now becoming more aware of the importance of process control and the value of maintaining each process step to within measurable statistical limits. Piezo Crystal Company has been extensively using SPC for the last two years. Piezo is actively encouraging customers to evaluate our performance using SPC data available to them.

This paper will present statistical data for the following tests:

- Crystal screening test results for SSB phase noise at 100 Hz offset from the carrier for each frequency measured.
- Completed oscillator screening test results for SSB phase noise at 100 Hz offset from the carrier for each frequency measured.
- A comparison between the screening noise data at 100 Hz offset and the completed oscillator noise at 100 Hz offset from the carrier for each frequency measured.
- Data from repeat measurements and experiments to determine the magnitude of various measurement errors.

CRYSTAL DESCRIPTION

The crystal required for this particular oscillator application had to be capable of producing the best combination results for close-in phase noise, high drive capacity, lowest gravitational sensitivity with the best long term stability that can be achieved in the 100 to 111.25 MHz frequency range.

The crystal of choice was an SC cut design manufactured by Piezo Crystal Company with the following design parameters:

- Holder Size: HC–35/U (TO-5) 4 point mount internal construction
- Gravitational Sensitivity: 3 sigma limit of (Γ) = $7 \times 10E^{-10}$, mean average of $4 \times 10E^{-10}$
- Drive Capability: > 5 milliwatts
- Crystal Q: > 100,000/mean average of 115,000
- Long Term Stability: < $5 \times 10E^{-7}$ per year
- Frequency Stability Over Temperature: + and – 8 PPM over the temperature range of 0 to +65 degrees C
- SSB Phase Noise: –120 dBc/Hz maximum at 100 Hz offset

OSCILLATOR DESCRIPTION

The oscillator needed the same design attributes as the crystal. The oscillator was designed to met the following limits

- Output Frequency: 100 to 111.250 MHz range, single frequency at one of 10 discrete frequencies
- Output Power: +3 to +5 dBm into 50 ohm load
- Output Waveform: Sine, Harmonics lower than –30 dBc, Spurious lower than –100 dBc
- SSB Phase Noise

CH2690-6/89/0000-309 $1.00 C 1989 IEEE

Offset Frequency	Max. Level	Mean Average
100 Hz	–120 dBc/Hz	–127 dBc/Hz
1000 Hz	–150 dBc/Hz	–154 dBc/Hz
10,000 Hz	–158 dBc/Hz	–164 dBc/Hz
100 KHz to 60 Mhz	–160 dBc/Hz	–167 dBc/Hz

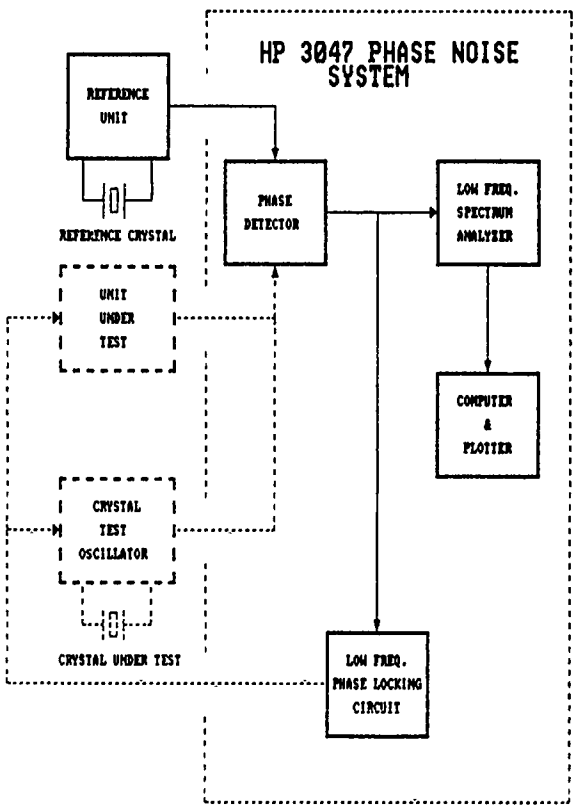
- Package Size: 1.62 x 1.50 x 1.00 inches (LxWxH)
- Absolute Frequency Tolerance: ±10 PPM includes frequency drift over temperature (0 to +65° C) and set tolerance at time of delivery

The oscillator stage is a one transistor Pierce configuration with the crystal running in the parallel resonance mode. The oscillator directly drives a common base one transistor tuned amplifier. The amplifier provides gain, load isolation, and harmonic rejection.

TEST/SETUP DESCRIPTION

Phase noise sidebands in frequency sources are caused by phase fluctuation due to the unintentional frequency modulation of the oscillator. The low offset frequency noise is mainly a measure of the crystal Q, crystal noise and loaded Q of the oscillator stage. The higher offset frequency noise is mainly due to the signal to noise ratio of the individual active stages.

Figure 1: Phase Noise Setup Diagram

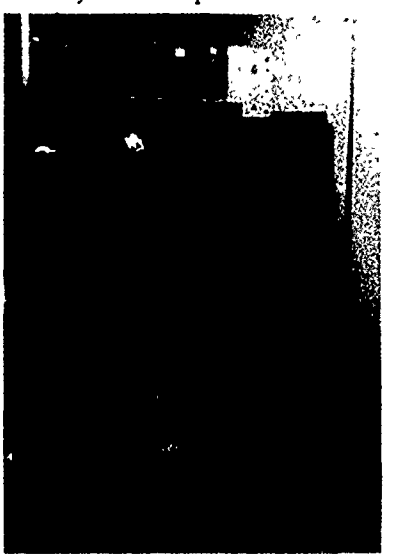


The two oscillator method was used in conjunction with the Hewlett Packard HP 3047 Phase Noise System to measure single side band phase noise. The Reference Unit was a modified version of the Unit Under Test. The modification consisted of the insertion of a Hi Q varactor circuit in the crystal loop of the oscillator stage. This provided control of the center frequency to compensate for normal frequency drift of the Unit Under Test due to room temperature and set tolerance variations of the crystals and oscillators tested. Special care was taken to maintain the loaded Q of the design at same time provide the frequency deviation needed to perform the test.

The system in Figure 1 requires that the noise of the Reference Unit with its crystal, referred to as the reference crystal, be equal to, or better than, the lowest anticipated noise of the Unit Under Test.

Finding a reference crystal for each of the 10 frequencies tested was a time consuming task. Due to the large number of possible frequencies that could be specified in any given frequency range, one does not have available a reference crystal of the exact frequency that represents the lowest noise possible for that design and exact frequency. This task becomes more difficult if the crystal design and process which can affect phase noise is still in the development stage. When measuring noise with the test setup in Figure 1, the indicated noise will not be lower than the noise of the Reference Unit/Crystal or the Unit/Crystal Under Test. If the noise generated by the two sources is the same, then the indicated noise is 3 dB too high. Thus, the test method is inherently conservative. The method we used was to substitute crystals in both oscillators until the lowest total noise performance was achieved. We then used one of the crystals as the reference crystal for that particular frequency.

Figure 2: Photo of Test Setup



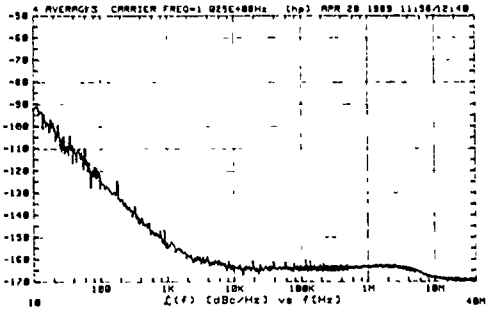
As you will see later, the process capability shows that lower phase noise crystals most likely were available later in the screening process. However, we felt the reference crystals selected were adequate for the test and for production to proceed. All reference crystals selected where a minimum of 12 db better than the specification of –120 dBc/Hz at 100 Hz offset from the carrier.

During the crystal screening test, a Crystal Test Oscillator of the same electrical design, with a crystal socket, was substituted for the Unit Under Test.

The Reference Unit also had a crystal socket for the reference crystals of each frequency tested. When a new crystal or oscillator was tested, the Reference Unit's reference crystal was changed and the oscillator and amplifier stages were retuned to the new frequency.

A typical readout from the system is shown is Figure 3. As noted, the crystals and oscillators with this design show well-behaved phase noise characteristics. The noise shown is total noise measured. No attempt was made in this data, or any data presented, to account for the fact their are two sources contributing to the measured noise.

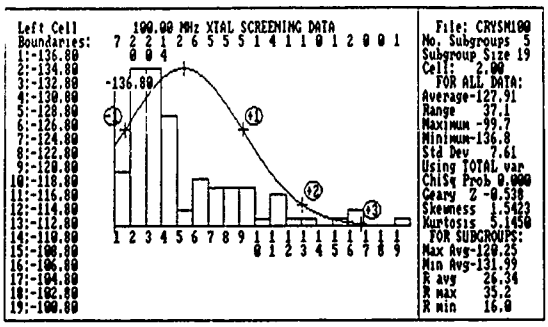
Figure 3: SSB Phase Noise Curve



CRYSTAL SCREENING TEST RESULTS

The crystals were screened for phase noise at 100 Hz offset from the carrier using one Crystal Test Oscillator for all crystals. The Crystal Test Oscillator was retuned for each frequency measured. The Reference Oscillator was also retuned with the reference crystal for that particular frequency.

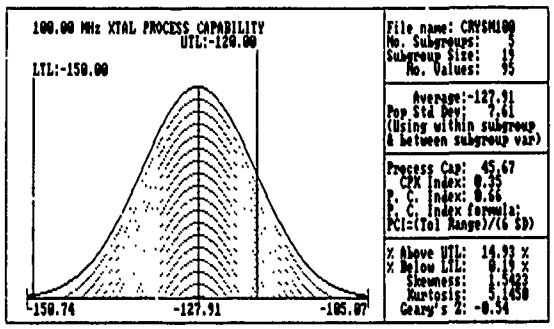
Figure 4: Crystal Screening Measurements



The data shown in Figure 4 are the actual readings for the 100 MHz crystals measured. This represents the typical spread of readings for all 10 frequencies.

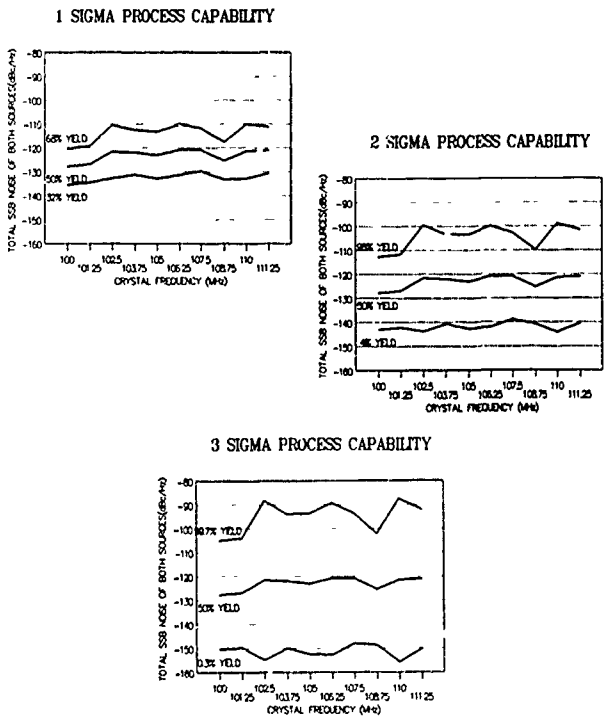
Note that in Figure 4 the peak of the actual data is skewed from the normal curve. The reason for this skewness was the lack of available data below –134.5 dBc/Hz. The measured noise can be no better than the Reference Oscillator with its reference crystal. Statistically a better reference crystal could have been selected. This would have provided readings below –134.5 dBc/Hz and the histogram would not be as skewed.

Figure 5: Crystal Process Capability



The same data at 100 MHz that is shown in Figure 4 is used to generate the curve in Figure 5. This shows the capability of the process used in the manufacturing of this crystal at 100 MHz to meet a phase noise specification of –120 dBc/Hz at 100 Hz offset from the carrier. From this curve we can determine future yields if the same crystal design and process is used.

Figure 6: Crystal Capabilities 100 MHz to 111.25 MHz

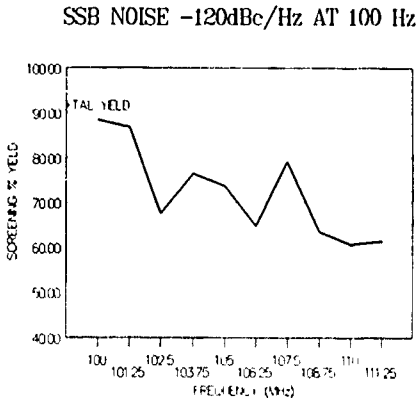


When all frequencies were tested and analyzed, the results in Figure 6 show the process capability of all 10 frequencies measured. From this set of curves, the statistical chance of meeting various crystal noise levels becomes clear. For example, the following yields can be reasonably expected from this crystal design and process in the frequency range of 100 MHz to 111.25 MHz.

Phase Noise Specification	Minimum Yields
–88 dBc/Hz	99.7%
–98 dBc/Hz	96.0%
–110 dBc/Hz	68.0%
–121 dBc/Hz	50.0%
–130 dBc/Hz	32.0%
–139 dBc/Hz	4.0%
–148 dBc/Hz	0.3%

The actual yields for a noise specification of –120 dBc/Hz at 100 Hz offset show a range 88.42% at 100 MHz to 60.53% at 110 Mhz. The average yield for the entire frequency range was 72.64%.

Figure 7: Actual Crystal Yields



Note that in Figures 6 and 7, the yield decreases as the frequency of the crystal increases. This result was expected due to the decrease in Q as the frequency rises. It becomes increasing more difficult to meet the same noise specification (i.e., –120 dBc/Hz at 100 Hz offset) as the crystal frequency is increased.

OSCILLATOR SCREENING TEST RESULTS

The crystals that passed the screening limit of –120 dBc/Hz at 100 Hz offset from the carrier in the first test, were installed in the final oscillator assembly. The completed oscillator assembly was then measured from 10 Hz to 40 MHz offset from the carrier for phase noise. The noise measured at 100 Hz offset was recorded for comparison with the crystal screening results of the first test. The Reference Oscillator was retuned with the reference crystal for that particular frequency.

Figure 8: 100 MHz Phase Noise Measurement Data

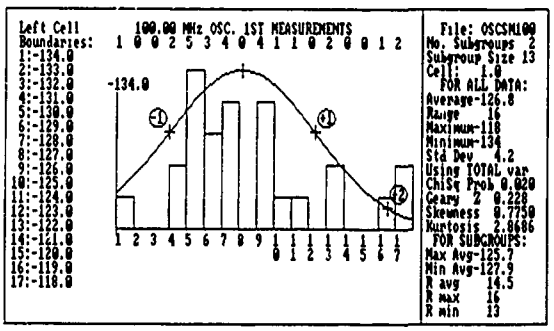
STATARAL.100

CRYSTAL AND OSCILLATOR FREQUENCY 100 MHz

OSCILLATOR SERIAL NO	CRYSTAL SERIAL NO IN OSCILLATOR	CRYSTAL 100Hz NOISE (REFERENCE) dBc/Hz	OSCILLATOR 100Hz NOISE 1ST MEAS'MT dBc/Hz	ERROR 1ST MEAS'MT TO REFERENCE dB	OSCILLATOR 100Hz NOISE 2ND MEAS'MT dBc/Hz	ERROR DIFF. 2ND MEAS,MT TO REFERENCE dB	ERROR DIFF. 1ST MEAS'MT TO 2ND MEAS'MT dB
048	054	132 6	128	4 6			
058	047	120 8	119	1.8			
068	055	132.9	131	1.9			
078	009	119 9	126	6 1			
088	068	129 5	126	5.5			
098	067	131 3	126	5.3			
108	062	133 6	129	4.6			
118	075	119 1	118	1.1			
128	011	125 7	122	3 7			
138	072	132.8	130	2 8			
148	019	128 4	124	4 4			
158	012	125 2	132	6 8			
168	069	134 2	125	9 2			
178	070	133 7	134	0 3			
188	010	122 1	118	4.1			
198	076	132 5	128	4 5	-124	8 5	4
208	080	134	130	4			
218	077	130 2	130	0 2	128	2.2	2
228	097	132	129	3	130	2	1
238	004	125 5	122	3 5			
248	099	129 4	128	1 4	29	0 4	1
258	126	133 8	126	7 8	126	7 8	0
268	106	130 2	128	2 2	-128	2.2	0
278	087	131 5	129	2 5	-129	2.5	0
288	102	-134	130	4	128	6	2
298	095	134 5	131	3 5	-130	4.5	1

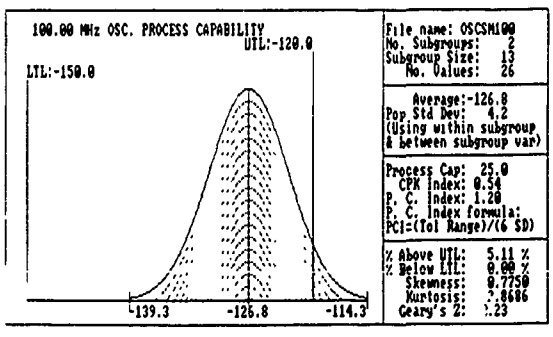
Figure 8 is typical of the data collected at each of the 10 frequencies. Data from all three tests is charted to show the typical spread of readings.

Figure 9: Oscillator Screening Measurements



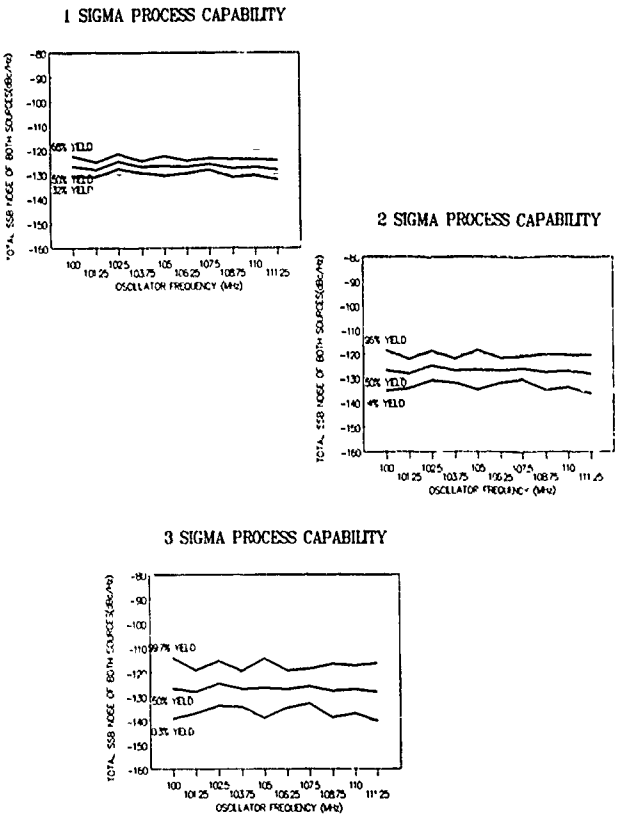
The data in Figure 9 are the actual readings for the 100 MHz final oscillator assembly. Note the skewness still present due to lack of measured data better than the noise of the Reference Oscillator.

Figure 10: Oscillator Process Capacity



The same data at 100 MHz (shown in Figure 9) is used to generate the curve in Figure 10. This shows the capability of this 100 MHz oscillator to meet a phase noise specification of –120 dBc/Hz at 100 Hz offset from the carrier.

Figure 11: Oscillator Process Capability: 100 MHz to 111.25 MHz



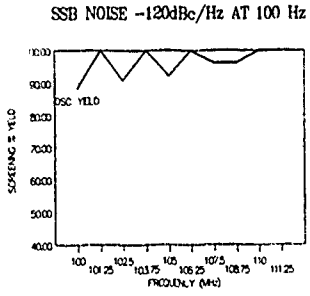
When the 10 frequencies were tested and analyzed, the results in Figure 11 show the total process capability of this oscillator design. From this set of curves the statistical chance of meeting various oscillator noise levels becomes apparent. For example, the following yields can be reasonability expected from this oscillator design and screening process in the frequency range of 100 MHz to 111.25 MHz.

Phase Noise Specification	Minimum Yield
–114 dBc/Hz	99.7%
–118 dBc/Hz	96.0%
–122 dBc/Hz	68.0%
–125 dBc/Hz	50.0%
–128 dBc/Hz	32.0%
–131 dBc/Hz	4.0%
–133 dBc/Hz	0.3%

The actual yields for a noise specification of –120 dBc/Hz at 100 Hz offset show a range of 88.46% at 100 MHz to 100% at 5 different frequencies. The average yield for the entire frequency range was 96.18%.

Note that in Figures 11 and 12 the yield remained relatively constant over the frequency range. This is to be expected due to the fact that the crystals were screened to the same noise specification at each frequency before assembly into the oscillator.

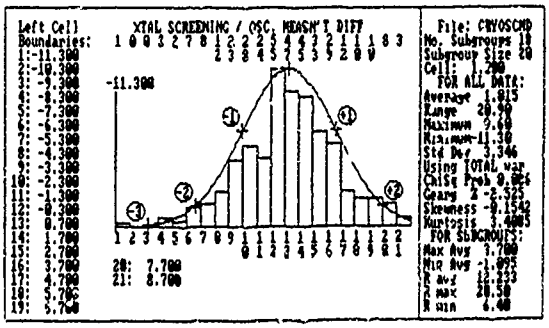
Figure 12: Actual Oscillator Yields



CRYSTAL SCREENING DATA TO OSCILLATOR SCREENING DATA COMPARISON TEST RESULTS

A comparison was made between the test results of the crystal screening and the results from the first oscillator measurement. Figure 13 shows this comparison.

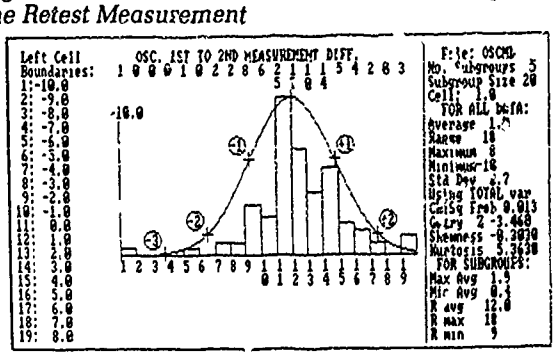
Figure 13: Crystal Screening Data Compared to Oscillator First Test Measurements



REPEAT MEASUREMENT TEST RESULTS

One hundred of the completed oscillator assemblies where retested for the customer on a sample basis. The noise measured at 100 Hz offset was recorded for comparison with the first oscillator measurement in test above. The Reference Oscillator was again retuned with the reference crystal for that particular frequency. Figure 14 shows the results of the retest

Figure 14: First Oscillator Measurement Compared to the Retest Measurement



ERROR ANALYSIS EXPERIMENT TEST RESULTS

The standard deviation (sigma) in both comparisons is approximately 3 dB. Why is there a ±3 sigma spread of 18 dB? An experiment was performed to answer this question. The experiment consisted of three tests.

TEST A: The same crystal was screened 10 different times.

Results: Standard Deviation of 0.66 dB.

TEST B: The same crystal was assembled in 10 different oscillators and noise measured each time.

Results: Standard Deviation of 1.0 dB.

TEST C: The same oscillator assembly was measured 13 times over a 4 day period.

Results: Standard Deviation of 1.0 dB

The statistical sum of the results of all three tests, A, B and C, would explain the majority of the measured Standard Deviation of 3.346 dB shown in Figure 13.

The statistical results of Test C (Tests A and B do not apply) would explain only 1 dB of the 2.7 dB Standard Deviation of the retest of the completed oscillator assembly shown in Figure 14. However another factor also contributed to the error; both the oscillator first reading and the retest reading were visually read from the phase noise curves. The visual error is estimated to be ± 2 dB.

The calibration tolerance of the test equipment has not been considered in either error distributions presented. The absolute calibration error of the system of ± 2 dB must be considered. This error can be measured and added or subtracted from the actual readings. The system error for this production run was less than 0.5 dB.

SUMMARY

Statistical analysis provides a great deal of detail that conventional data listings cannot show. It also gives everyone the opportunity to evaluate the statistical chances of meeting a particular specification in a production environment.

The tests and their results as described have several major areas of concern that must be addressed:

1. The tests are time consuming using expensive equipment.
2. Finding a low noise reference source (crystal and/or oscillator) for each frequency measured is difficult.
3. The oscillator and crystal design must have sufficient margin to allow for the measurement errors of the equipment and those encountered in a production testing environment.

The results of these test shows the statistical possibility of producing quartz crystals and oscillators with SSB phase noise at 100 Hz offset lower than –140 dBc/Hz. At this time however, the yield would be very low and unacceptable for production in quantity. Work is ongoing at Piezo to improve our process yields with regards to phase noise.

43rd Annual Symposium on Frequency Control - 1989

PHASE NOISE FREQUENCY DISTRIBUTIONS OF SC AND AT QUARTZ CRYSTAL RESONATORS

Nicolas K. Bates and Gregory L. Weaver

Piezo Crystal Company
100 K Street
Carlisle, PA 17013

Summary

Phase noise measurements of 1,100 quartz crystal resonators have been collected and analyzed. The focus of the analysis is on the nature and characteristics of the variations, or distribution, of phase noise in groups of similarly produced resonators. Crystal cuts represented in the study include the SC and AT, of various overtones, ranging in frequency from 18.754250 MHz to 111.250 MHz.

The phase noise data displayed as histograms show non-Normal distributions. Many of these histograms display multi-modal characteristics. Analyses has demonstrated that these modal characteristics are statistically significant. The non-Normal nature of the histograms is accentuated by the limitation of the reference crystal which introduces a skewness to the distribution. A technique which aids in the reduction of this problem is described.

Process control theory asserts that assignable causes should exist for multi-modal distributions. The isolation and identification of the various modes observed in the distributions will assist in the control or elimination of any associative causes and narrow the observed screening variances improving resonator yields.

Introduction

A given group of crystals are sent through a unique manufacturing process in order to produce units that meet the specifications of the customer. The quartz and process used is believed to be the main determinants of how the phase noise of the crystals, within groups, will vary and their probability of doing so. Thus, there exists a theoretical probability distribution of phase noise which may be called the parent distribution. The initiative to begin the current work came from a recognition that little is known of the parent distribution, or distributions, of phase noise. However, sample data from manufactured resonators can be collected and analyzed using the tools of descriptive statistics. These include measurements of central tendency, and dispersion. Graphical representations of sample frequency and cumulative frequency distributions prove to be also helpful. Based on these sample results, and certain statistical tests, inferences may be made concerning the nature of the parent distribution.

The present paper begins this process of investigation into the characteristics of the phase noise parent distributions by presenting a great deal of measured crystal phase noise data. The building up of a large database is one of the first steps of characterization, and hence, control and even reduction of unwanted phase noise.

Two measurement systems were used in the course of the study to collect resonator phase noise data. The first is the HP3047A, which uses the carrier suppression method, placing one crystal oscillator in phase quadrature with another identical oscillator. The second is a passive crystal bridge circuit, which uses a balanced bridge and two electrically identical resonators as first described by Walls and Wainwright <1>. All phase noise data were taken at 100 Hz from the carrier.

Corrections Used to Determine Phase Noise Values From Measurement Data

The two measurement systems used in collecting the data measure the combined phase noise of both the reference crystal and the crystal under test. To correct for this, in order to determine the phase noise of the crystal under test, an error is calculated and subtracted from the raw phase noise value. This error is given by the following equation:

$$\text{error(dB)} = 10\log[1+\text{antilog}(L_{ref}-L_{cut}/10)]$$

The inherent advantage of the use of this correction factor is that the crystal data near the reference will shift as much as 3 dB lower on the phase noise level axis while crystals which measured 10 or more dB above the reference will stay in place. This dispersion in the data helps to separate any multi-modal tendencies which may exist. It is particularly important for those distributions which may have modes intersecting with one another. An additional difficulty with either measurement system is that they both use a comparison method to a reference. This causes a "pile-up" of crystals to appear on or within the lowest measurement cells of the distributions. The effect places a skewness and some kurtosis to any distribution modes which appear around the level of the reference crystals.

Skewness is defined by the third moment of the distribution and is related to the symmetry of the frequency distribution about the mean. Kurtosis is the fourth moment and is related to the "peakedness" of the distribution. A significant coefficient of either of these factors in a distribution will make the fit of a regular Gaussian curve less than likely.

Consequently, some correction for this measurement effect was desirable. A straight forward approach was chosen which tended to minimize the "pile-up". This correction was used in the examination of the aggregate of the SC distributions described later in the paper. Simply stated, for the generation of the aggregate distribution, a weighted average figure for the reference based on all the individual distributions was chosen. The average figure was determined by trading the actual calculated average reference and a value which did not create a detrimental loss in accumulated data. Crystals whose phase noise

CH2690-6/89/0000-315 $1.00 C 1989 IEEE

measured within 1-2 dB of this average reference dBc/Hz value were eliminated from the raw data. The effect was to show a marked decrease in the observed skewness in the modes close to the lower portion of the distribution.

Phase Noise Measurement Results

The most familiar way of presenting the frequency distribution of a group of resonators is graphically in the form of a histogram. Here, we present the phase noise data collected in terms of histograms, by groups, and a table of descriptive statistics.

First, a note on the histograms. Technically, phase noise is a continuous random variable. However, measuring it to the nearest integer forces the values of phase noise recorded to become discrete. Therefore, the histograms look as though they are recording a discrete random variable, and not a continuous random variable.

Figure 1 is a histogram of a group of ninety five, fifth overtone, AT cut, 80 MHz crystals. The crystals were measured at 100 Hz from the carrier using the passive crystal bridge circuit.

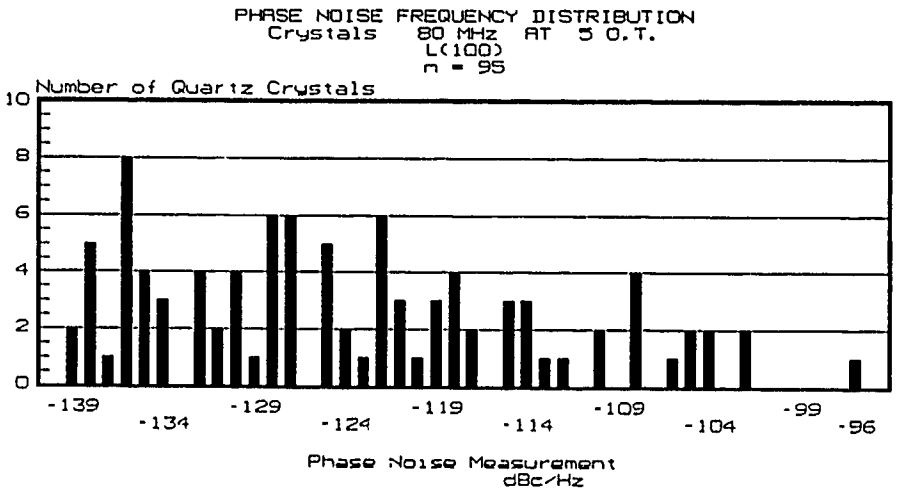


Figure 1

Figure 2 is a histogram of a group of ninety, third overtone, SC cut, 100 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

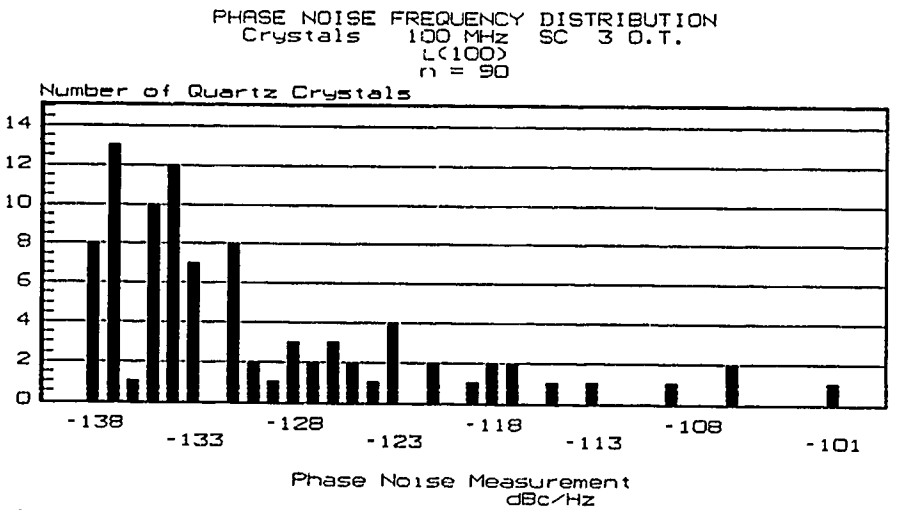


Figure 2

Figure 3 is a histogram of ninety seven, third overtone, SC cut, 101.250 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

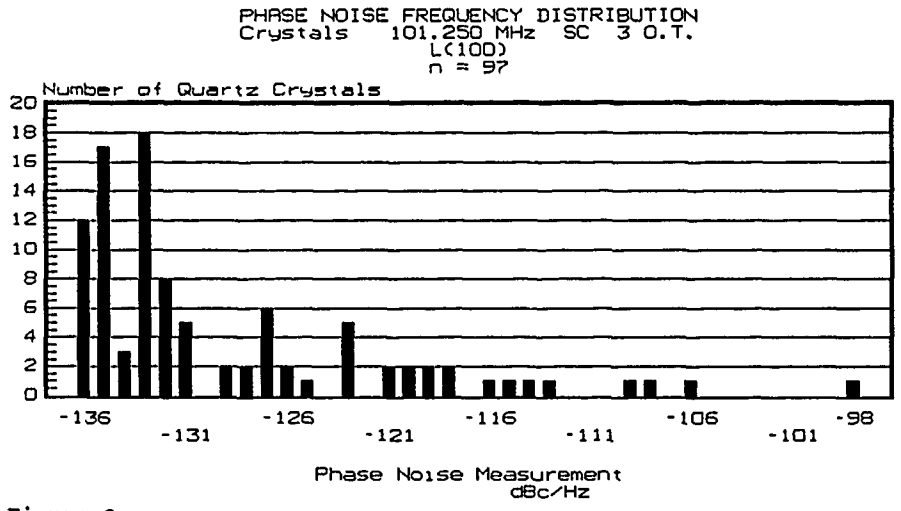


Figure 3

Figure 4 is a histogram of one hundred and eight, third overtone, SC cut, 102.500 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

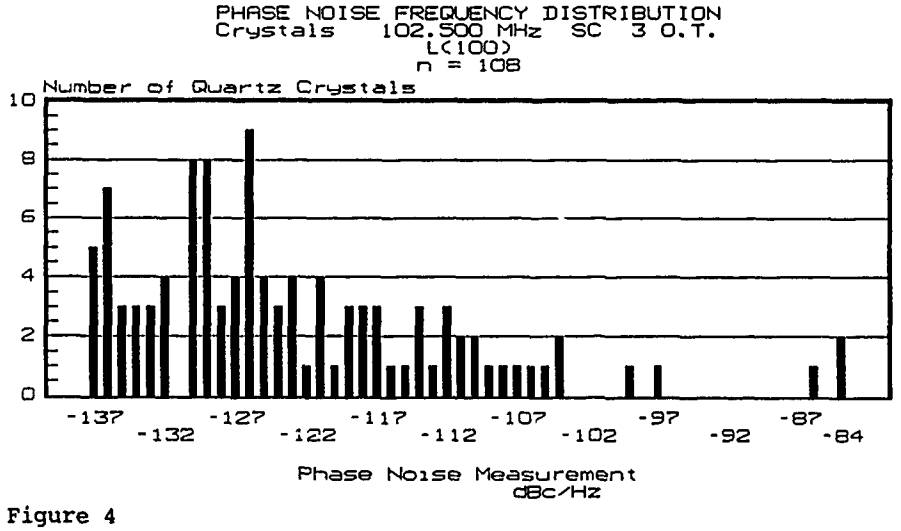


Figure 4

Figure 5 is a histogram of one hundred and thirty four, third overtone, SC cut, 103.750 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

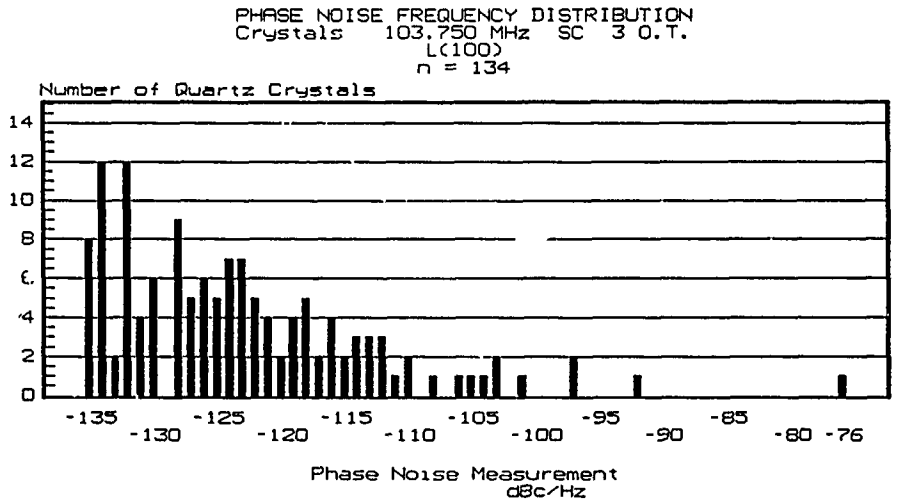


Figure 5

Figure 6 is a histogram of a group of one hundred and forty three, third overtone, SC cut, 105 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

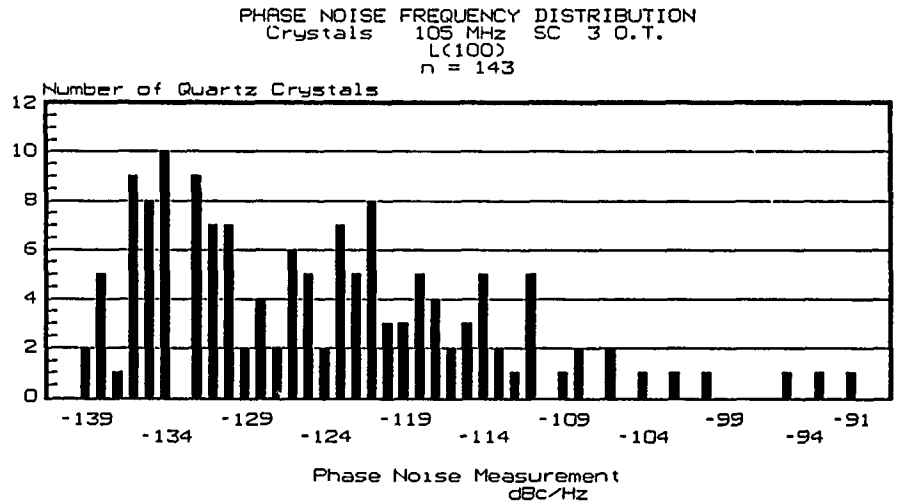


Figure 6

Figure 7 is a histogram of a group of seventy three, third overtone, SC cut, 106.250 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

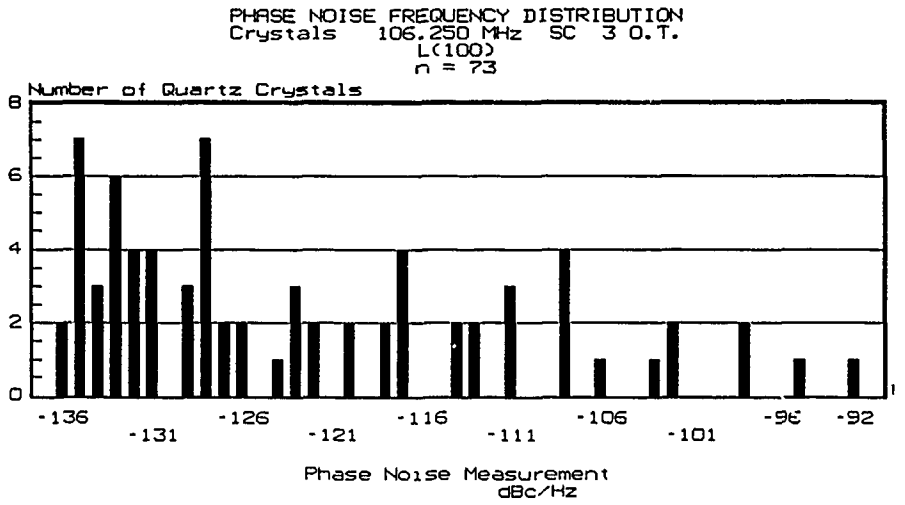


Figure 7

Figure 8 is a histogram of a group of fifty six, third overtone, SC cut, 107.500 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

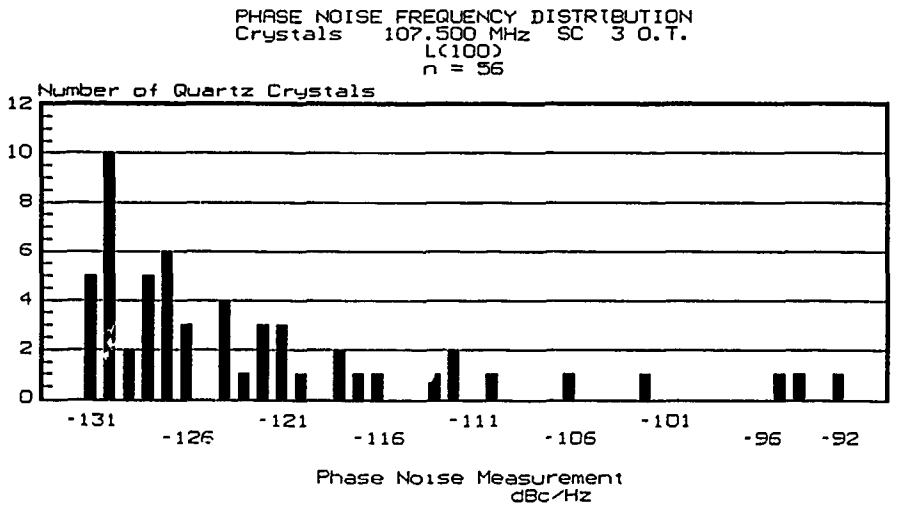


Figure 8

Figure 9 is a histogram of a group of sixty two, third overtone, SC cut, 108.250 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

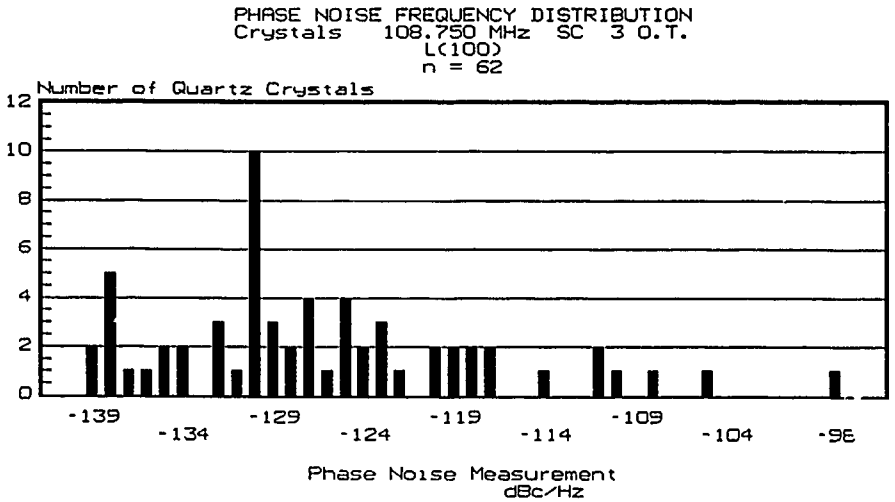


Figure 9

Figure 10 is a histogram of a group of sixty nine, third overtone, SC cut, 110 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

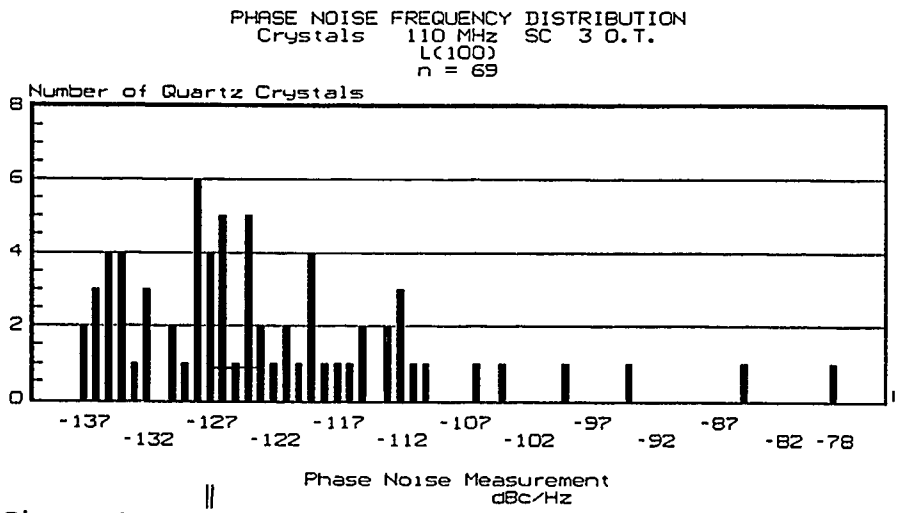


Figure 10

Figure 11 is a histogram of a group of ninety one, third overtone, SC cut, 111.250 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

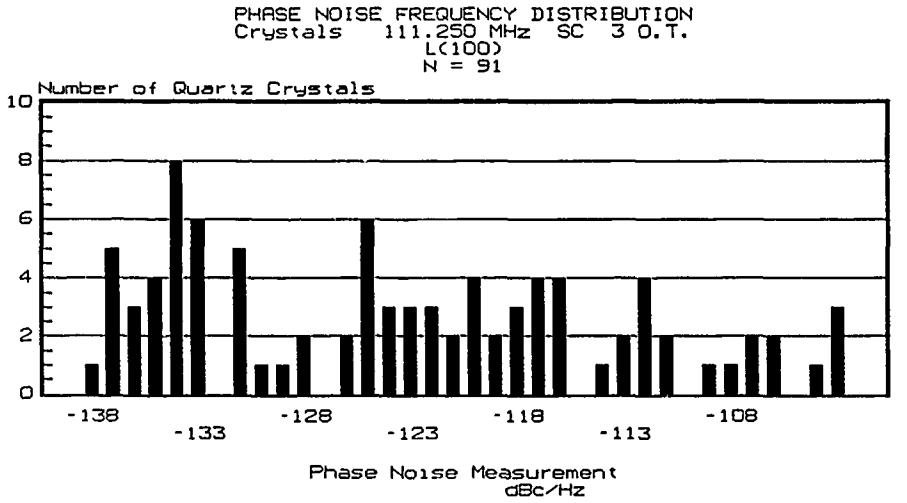


Figure 11

Figure 12 is a histogram of a group of eighty four, fundamental, AT cut, 18.754250 MHz crystals. These were measured at 100 Hz from the carrier using the HP3047A phase noise measurement system.

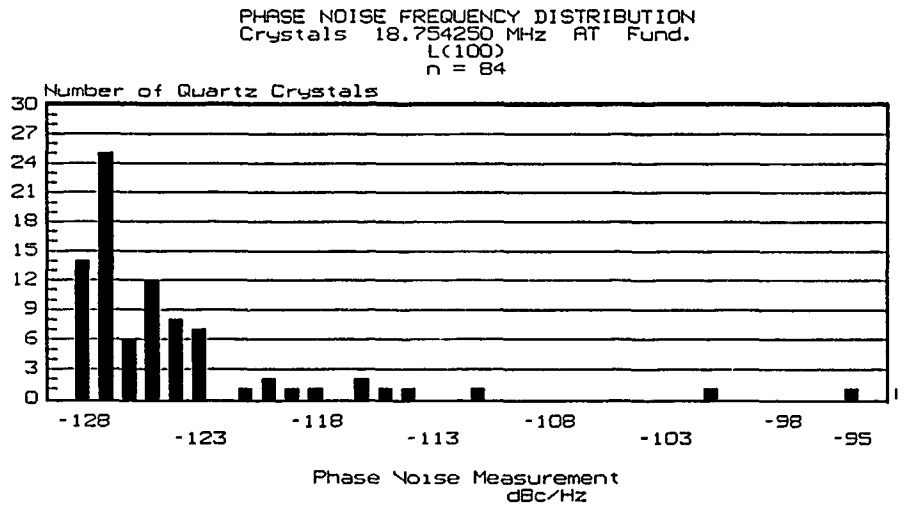


Figure 12

Table 1 is a chart of descriptive statistics for the 12 groups of crystals.

Table I

Xtal Freq. (MHz)	Cut	Ovtn	Number of Xstals	Mean dBc/Hz	Largest Mode dBc/Hz	Range dBc/Hz	Stand. Dev.	Skew-ness	Kurto-sis
80	AT	5	95	-123	-136	44	10.74	-0.62	0.504
100	SC	3	90	-130	-137	38	8.29	2.00	1.521
101.25	SC	3	97	-129	-133	39	7.90	2.62	1.660
102.5	SC	3	108	-122	-126	54	11.59	1.70	1.234
103.75	SC	3	134	-123	-134	60	10.34	2.84	1.383
105	SC	3	143	-124	-134	48	10.41	0.35	0.801
106.25	SC	3	73	-122	-135	45	11.79	-0.36	0.818
107.5	SC	3	56	-122	-130	40	9.78	2.39	1.680
108.75	SC	3	62	-126	-130	42	9.05	0.61	0.868
110	SC	3	69	-122	-128	60	11.98	2.77	1.435
111.25	SC	3	91	-123	-134	36	10.18	-1.00	0.324
18.754	AT	1	84	-124	-127	34	5.47	12.93	3.252

Preliminary Analysis of Results

A couple of general comments can be made concerning the histograms displayed in the previous section. Many clearly show non-Gaussian, multi-modal, characteristics. The mean of the samples do not occur at the same value as the largest mode. Generally, this mode is within the lowest 25% of the dBc/Hz range. There exists a significant probability that some crystal phase noise values will be greater than 10 dBc/Hz from the largest mode, and in some cases greater than 20 dBc/Hz.

The Komogorov-Smirnov test can be used to calculate the probability that the distribution of phase noise of a set of crystals follows a Gaussian distribution. This test is a standard, non-parametric, goodness-of-fit test, which compares the cumulative frequency distribution of the sample, to that of a theoretical distribution, in this case the Gaussian. Table 2 is a chart of the K-S probability that the given sample follows a theoretical Gaussian distribution, whose mean equals the sample mean, and whose variance equals the sample variance.

Table II

Xstal Freq. (MHz)	Cut	Ovtn	Number of Xstals	K-S Prob.
80	AT	5	95	0.035
100	SC	3	90	0.000
101.25	SC	3	97	0.000
102.5	SC	3	108	0.000
103.75	SC	3	134	0.000
105	SC	3	143	0.000
106.25	SC	3	73	0.000
107.5	SC	3	56	0.000
108.75	SC	3	62	0.099
110	SC	3	69	0.011
111.25	SC	3	91	0.002
18.754	AT	1	84	0.000

More sophisticated tests might include skewness and kurtosis variables into the theoretical distributions, as well as more accurate corrections to fix crystal pile up at the value of the reference crystal. Different theoretical distributions may also be used.

Figure 6 is one of the histograms which displays non-Gaussian characteristics. Note that this group of crystals is the largest of a single frequency presented in this paper. The frequency distribution appears to show a distinct mode in the range of -139 to -127 dBc/Hz. It is possible to focus on this single mode and to analyze it. Using a computer generated Gaussian curve the K-S probability is 65.89% that the mode is Gaussian distributed. The probability may be increased by adding a kurtosis factor to the computer generated Gaussian curve.

Because of the large amount of data available for the 100 to 111.25 MHz SC cut, it was desirable to merge or build an aggregate distribution of this data. This merger or aggregate would be statistically large and should provide further confidence in the characteristic shape of the parent distribution. The distribution which was generated contained over 800 data points. This is after the correction for crystal "pile-up" around the various reference crystals was accounted for in regard to the correction technique describred earlier.

Another key element to the development of the aggregate was to account for the linear degradation or "slip" along the phase noise axis as the resonator data progressed across the 100 to 111.25 MHz band. This linear degradation is the effect of the QxF factor or figure of merit for quartz. In simpler terms, the measured Q of the crystal falls as the frequency of resonance increases. Figure 13 illustrates this effect by plotting the observed yield for each SC cut frequency against a constant phase noise level of -120 dBc/Hz. The line drawn through the scatter of points is the best-fit linear regression. Notice that the slope of the regression line somewhat corresponds to the expected 11% change in resonator Q.

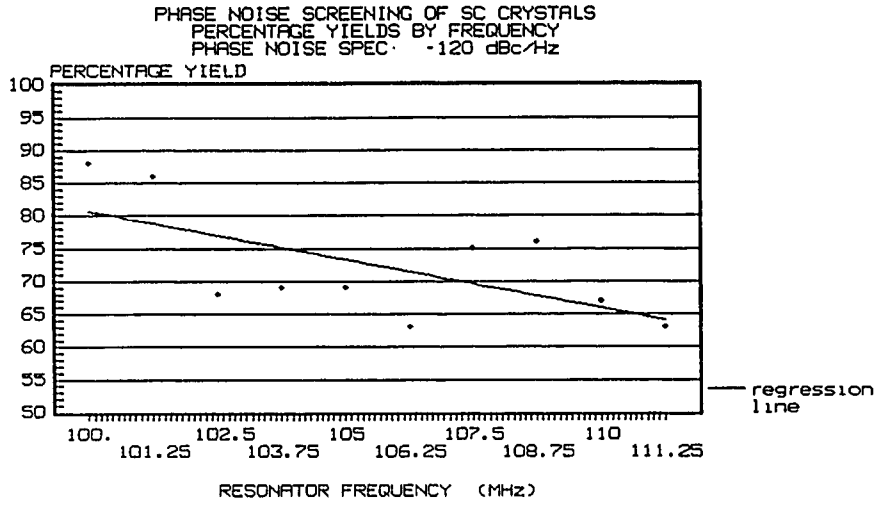


Figure 13

The regression line and the 100 MHz SC cut distribution (Figure 2) were used in order to approximate the amount of slip along the phase noise axis which each distribution would exhibit. This amount was derived by moving the -120 dBc/Hz limit of Figure 13 down the cumulative distribution of the 100 MHz data until the same yield number of about 65% resulted as was observed for the 111.25 MHz distribution (Figure 11). This corresponds to a shift of about 8 dBc. Using this number and the regression line, each of the ten SC cut distributions were then shifted a conforment amount based on their frequency.

Figure 14 is a histogram of the aggregate of all ten SC cut distributions. It is clearly noticable that the distribution is not normal. A K-S test to a best-fit Poisson distribution demonstrated a very low probabilty that the aggregate was Poisson in shape. To the authors' eyes, it seemed likely the shape of the aggregate was the intersection of two fundamentally different distributions. The general shapes of these two appear freely sketched on Figure 14.

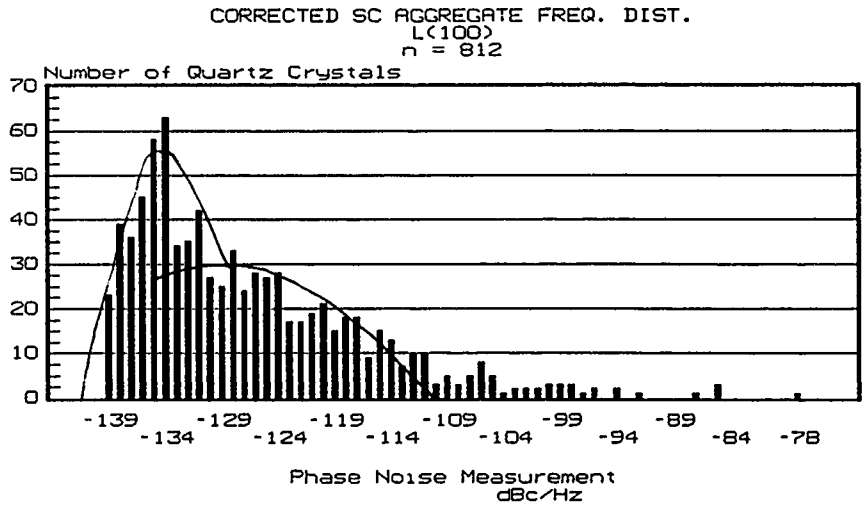


Figure 14

An attempt to seperate these two distributions was performed by subtracting an artificially generated distribution. This artificial distribution could approximate the low dBc/Hz distribution characterised by the large peaked mode, thereby leaving the second for analysis. The nature of the artificial distribution was determined by matching the lower cumulative frequency characteristics of the large peaked mode at the low end of the aggregate distribution. This was done under the assumption that with a great deal of the skewness removed due to "pile-up" the nature of the low end distribution should be somewhat symetrical about its mode.

Figure 15 is a histogram of a randomly generated distribution with the aspect described above. The sample number chosen was 381 points or 47% of the total amount of data because cumulative analysis showed that 23.5% of the aggregate occured below the low peak mode. The mean, variance and kurtosis were then selected to assist in matching the lower cumulative frequency characteristics of the random distribution to the aggregate.

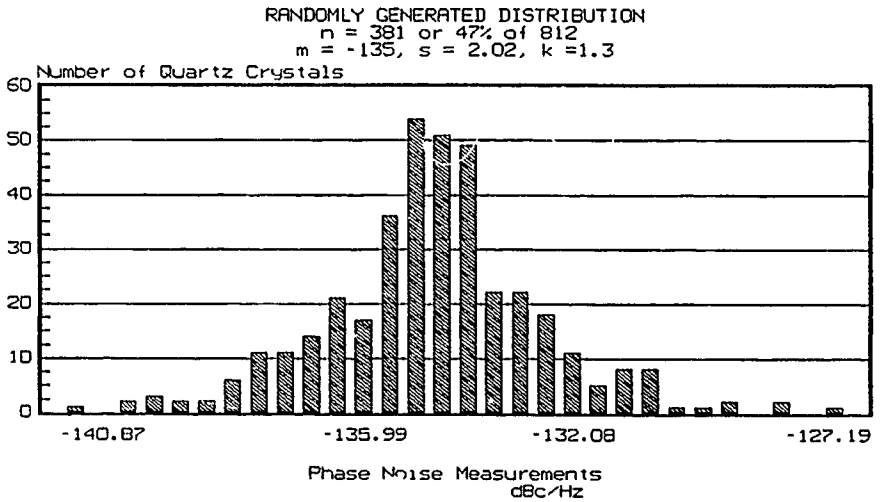


Figure 15

Figure 16 is a histogram of the transformed aggregate after the randomly generated distribution of Figure 15 was subtracted. The nature of the transformed distribution shows a large broad mode centered at approximately -126 dBc/Hz. Cumulative anlysis of the distributions confirmed that the decrease in the frequency of data below the mode of the transformed distribution is real and not an artifact of the subtraction. However, the shape of the transformed distribution is not normal.

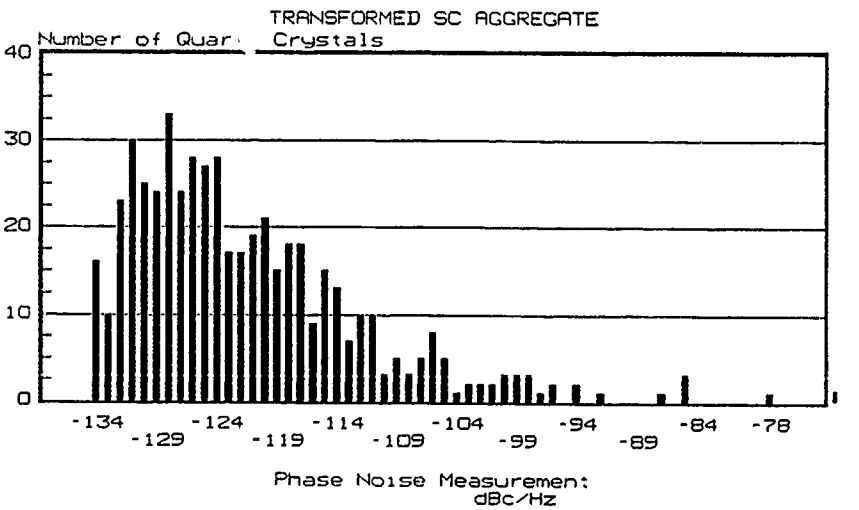


Figure 16

This could be due to several factors. First, the subtraction performed on the aggregate could easily have eliminated some of the lower tail of the transformed distribution (Figure 16). There is no doubt that some intersection between the two distributions speculated in the aggregate (Figure 14) has occured. Second, an inspection of the data of the ten individual files shows that there are large differences in the shapes of these constituent distributions. This will create a "smearing" in the aggregate which will distort any similarities to any standard shaped distributions. Finally, the analysis performed above is by no means exhaustive. More intense study and careful construction of the aggregate will be required before any firm conclusions of the parent population may be drawn.

Conclusions

Phase noise measurements on 1,102 crystal resonators has been presented, in order to investigate the nature of the theoretical parent distributions of phase noise. Histograms of groups ranging from 56 to 143 units show multi-modal characteristics. They also show large ranges that give rise to significant probabilities of having crystal phase noise values greater than 10 and sometimes 20 dBc/Hz from the largest mode. Those which do not seem to show a multi-modal nature have, however, extremely peaked single modes.

The Komogorov-Smirnov test results, as recorded in Table 2, has shown that the frequency distributions of the sample groups of crystals do not follow a Gaussian distribution whose mean equals the sample mean, and whose variance equals the sample variance. This is not to say, however, that as the number of crystals within a group increase to 200, 300, or higher, the distributions will not tend towards the Gaussian. The authors have not presented in this paper any results of the K-S test used to model sample data to other theoretical distributions, which they may legitimately follow.

Should further analysis, and greater numbers of crystals measured confirm the multi-modal nature of the parent phase noise distributions, then an analysis of each mode, as presented for the 105 MHz data, will be an important thing to pursue. Process control theory asserts that assignable causes should exist for multi-modal distributions. Investigations of these causes which lead to their control or elimination will make it possible to attenuate or eliminate the unwanted, high dBc/Hz, phase noise modes.

Acknowledgements

The authors wish to acknowledge the participation and advice of the following individuals: William Hanson, Tim Wickard and Warren Walls. We would also like to thank Mike Driscoll for his contribution of the VHF bridge circuit with which some of the data was taken.

References

1. F. L. Walls, A. E. Wainwright "Measurement of the Short Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications." IEEE Trans. Instrum. Meas., Vol IM-24, no. 1, March 1975

43rd Annual Symposium on Frequency Control - 1989

A NEW" FILTERED ALLAN VARIANCE" AND ITS APPLICATION TO THE IDENTIFICATION OF PHASE AND FREQUENCY NOISE SOURCES

J. Groslambert, 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

F. Vernotte
Ob_ervatoire de Besançon - 41bis avenue de l'Observatoire - 25041 Besançon Cedex - France

F. Walls
National Bureau of Standards - Time and Frequency Division
Boulder, Colorado 80303 - USA

Abstract

In part one different digital noise generators are described. They use

1) a recursive equation known for its properties of successive bifurcations leading to a chaotic behavior.

2) a second method is presented based on the theory of fractals. The noise is generated from a reccurent fractal curve with a given fractal dimension, the slope of the noise spectrum being linked to the fractal dimension.

3) a third method consists in applying the mean theorem to the random number series given by RDN function of a calculator. In this manner a sample series is obtained representing a white noise with a gaussian distribution. With three different numerical filters using the Z transform all the noise types can be generated.

In part two these noise sources are used to generate sample series which simulate frequency samples as given by a counter.

Applying the Allan variance to the samples of six generators (white phase, flicker phase, white frequency, flicker frequency, frequency random walk, filtered flicker frequency) yields the expected theoretical slopes have been observed.

It is well known that both white phase and flicker phase noises give in time domain when using Allan variance the same characteristic in log-log plot slope, which corresponds to τ^{-2}. It will be interesting to separate the two contributions, this is one of the properties of the Modified Allan variance.

In the present work a new method is used. It consists in filtering the f^{-1} and f^{0} phase noise (simulated by the previous noise generators) by means of a digital filter in time domain (using the Z transform), which yields f^{-3} and f^{-2} noises, which thus can be identified by Allan variance.

The combination of the digital filter and the Allan variance corresponds to a new variance, which is described and compared to the well known "Modified Allan Variance".

Introduction

The "Allan Variance" which is used for the characterizations of a signal, gives the possibility to identify in time domain the different types of noises present in a frequency source ; their respective levels in frequency domain can be calculated by Fourier transform, but for white phase noise and flicker phase noise, this transformation is not possible because they exhibit an Allan variance with the same power law τ^{-2} and consequently they cannot be separated.

The "Modified Allan" variance [1] proposed by Allan in 1981 gives the possibility to solve the ambiguity.

For the white phase noise this variance exhibits a τ^{-3} slope which is sligthly different from the τ^{-2} slope obtained for the flicker phase noise : however, on a log-log plot it is difficult to find the respective asymptotes. Therefore, it can be interesting to use a new variance which offers a better separation of the white and flicker phase noises. This is the purpose of the present paper.

The variances are tested on a computer by means of series of well identified samples, which first of all must be generated. If it is simple to achieve a noise generator exhibiting a noise spectrum in $1/f^{\alpha}$ when α is an even integer by using a white noise filtered by a first order filter, it is more difficult to do for the odd values of α, (or when α is not an integer).

In fact all the noise spectra showing an odd value of α can be obtained from the flicker noise corresponding to $\alpha = 1$. Two methods can be used to obtain this flicker noise, either by superposition of weighted Lorentzian spectra or by filtering the white noise with a cascade of first order filters in series with cut-off frequencies distributed as a geometrical series.

These noise generators will be used for comparing the properties of different kinds of variances : Allan variance, Modified Allan variance and Filtered Allan variance.

White noise generators

Recursive equation

The first method which was tested is based on the use of a recursive equation known for its properties of successive bifurcations leading to a chaotic behavior (Fig. 1)

$$x_{n+1} = 4\lambda x_n (1 - x_n)$$

- For $\lambda < .8$	there is only one solution obtained by solving the equation $x_{n+1} = x_n$
- For $.8 < \lambda < .88$	there are two solutions corresponding to $x_{n+2} = x_n$. Then the number of possible solutions increases drastically with λ.
and	
- For $\lambda = 1$	the number of solutions is so large that the behavior is chaotic, and the power spectral density of the corresponding values appears as a white spectrum.

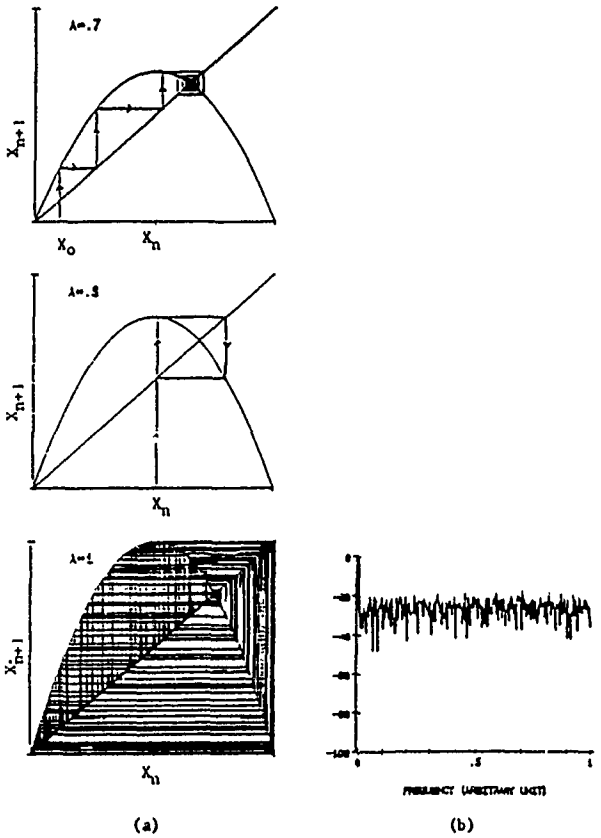


Fig. 1 : Recursive equation $x_{n+1} = 4\lambda X_n (1 - X_n)$ leading to successive bifurcations and chaotic behavior (a) white spectrum corresponding to the chaotic bahavior (b)

This is a very easy method to generate random numbers with white spectrum. But it must be noticed that these distribution functions is not gaussian.

Using fractal

The second method uses the theory of fractals. A deterministic fractal signal, looking like a noisy signal can be easily fabricated by means of a broken curve drawn in a rectangle lattice with m times n elementary rectangles as illustrated by figure 2.

Then by successive iterations of each elementary segment a fractal curve is obtained with a resolution depending on the number of iterations, and with a fractal dimension D which is simply given by

$$D = 2 - \frac{\log n}{\log m}$$

A curve with a given fractal dimension can be easily obtained by simply choosing the number of the elementary rectangles.

The fractal dimension of various curves was plotted as a function of the coefficient α of the measured $1/f^{\alpha}$ power spectral densities (Fig. 3).

Therefore, it is possibie to generate a noise with a given power spectral density by choosing the appropriate fractal dimension, following Fig. 3. But this curve depends on the frequency cut-off of the signal. On the same figure is indicated the limit corresponding to frequency cut-offs and zero and infinity.

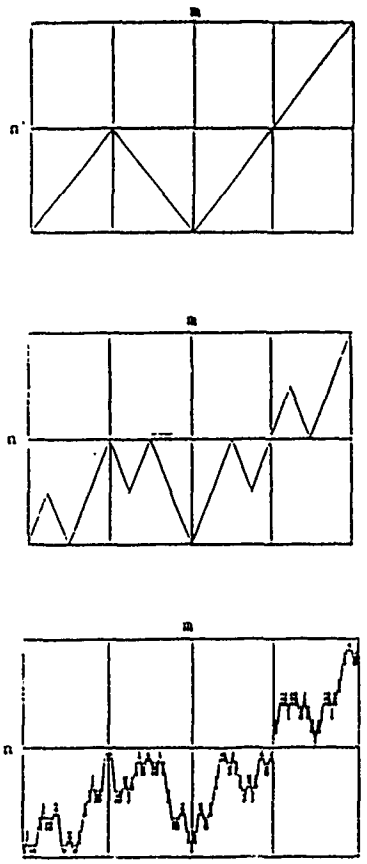


Fig. 2 : Construction of a fractal curve with a given fractal dimension by successive iterations

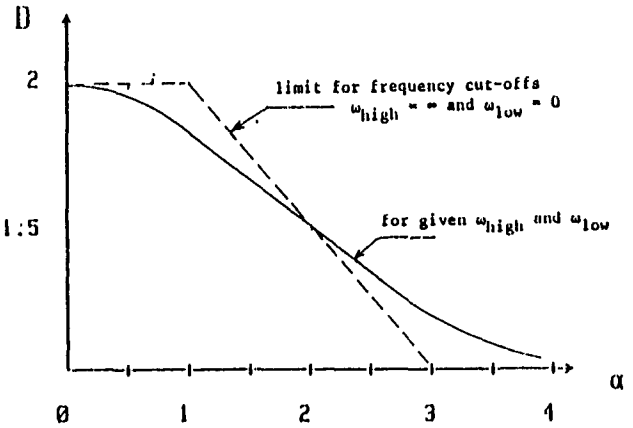


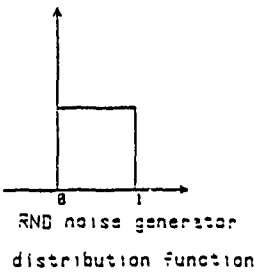
Fig. 3 : Relation between fractal dimension D and power spectral density coefficient α

From a computer random number generator

This other method uses a generator of random number x_i, corresponding to the "RND" functions of a computer, whose values are between 0 and 1 with a white distribution function. This white distribution function can be transformed to a gaussian distribution with a zero mean value

by applying on the x_i samples the mean theorem in order to obtain the $\bar{x}_i$ samples defined by

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{N} (x_i - 0{,}5)$$



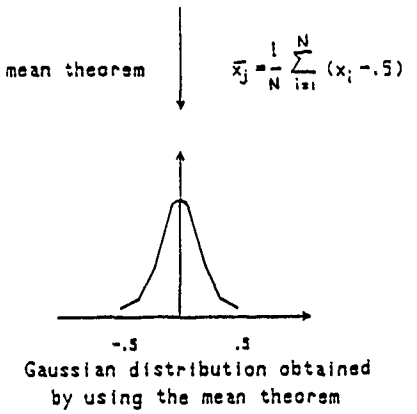


Fig. 4
Distribution function of a RND generator and application of the mean theorem to obtain a gaussian distribution

For $N = 1$ we obtain the original shifted white distributions (Fig. 5) calculated on 8192 successive generated numbers.

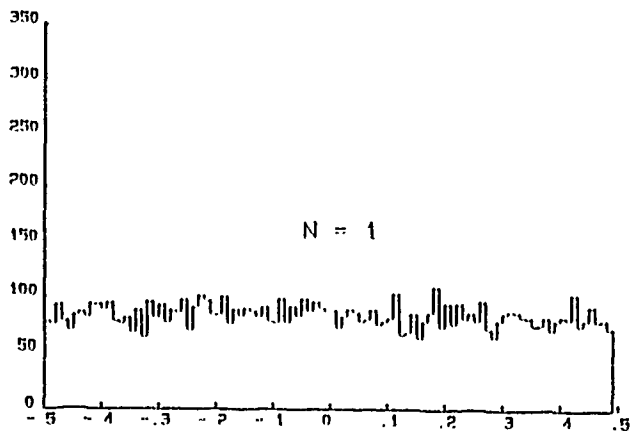


Fig. 5 : Distribution function of the RND number generator after application of the mean theorem for N = 1

Figs. 6 and 7 show the gaussian distribution for $N = 3$ and $N = 8$.

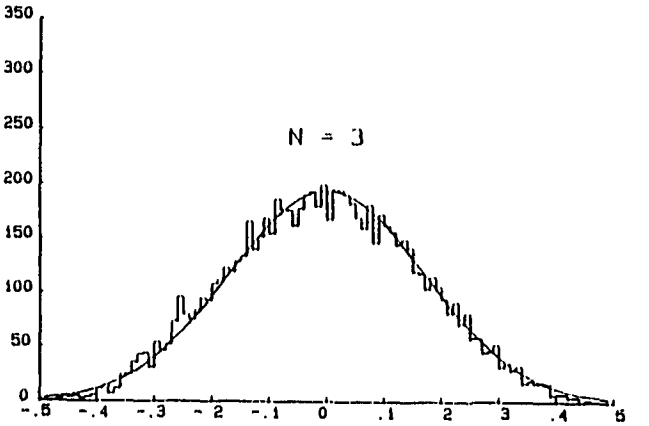


Fig. 6 : Distribution function of the RND number generator after application of the mean theorem for N = 3

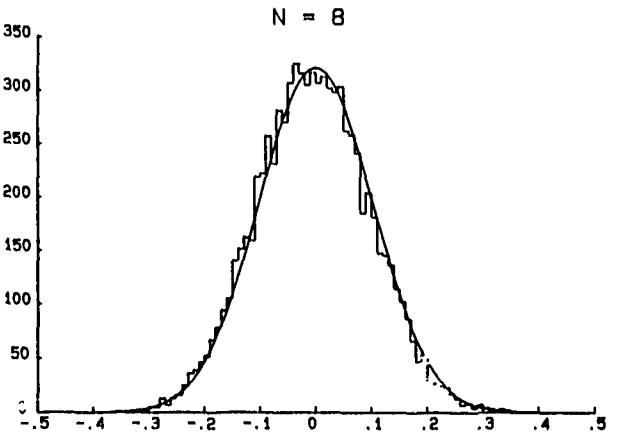


Fig. 7 : Distribution function of the RND number generator after application of the mean theorem for N = 8

This series of samples is considered as a series of time samples integrated over a time interval τ similar to frequency samples furnished by a counter. The corresponding Allan variance $\sigma^2(\tau)$ exhibits a slope of -1 on a log-log plot as expected for white noise (Fig. 8).

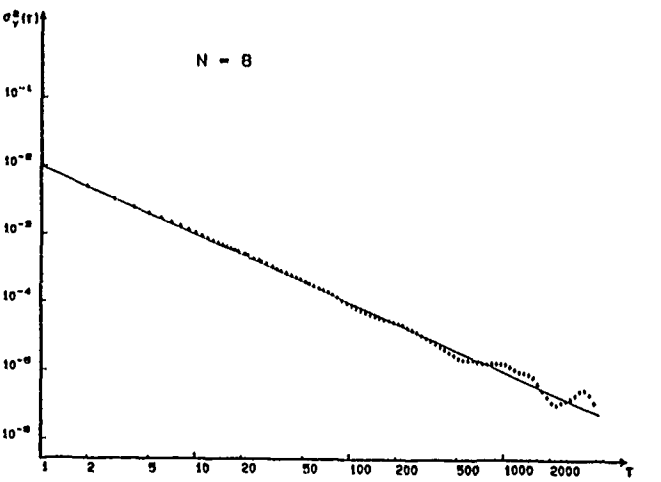


Fig. 8 : Allan variance obtained for a white noise and N = 8

Generation of 1/f$^{\alpha}$ noises

The different noises are obtained by filtering. Only three different types of filters (derivative filter, integrator filter and 1/f filter) are necessary to obtain from white noise all the kinds of noises found in the frequency noise spectrum of oscillators.

$f^0 \times f^2$	→	white phase noise
$f^0 \times f^{-1} \times f^2$	→	flicker phase noise
f^0	→	white frequency noise
$f^0 \times f^{-1}$	→	flicker frequency noise
$f^0 \times f^{-2}$	→	random walk frequency noise.

Two processes can be used, digital filtering or a fast Fourier transform.

Digital filtering

A general transfer function for a filter can be written as

$$\frac{S(p)}{E(p)} = \frac{1 + p k_2}{1 + p k_1} .$$

By applying the complex homographic transform corresponding to the well known Z transform, which links Z to p

$$Z = \frac{1 + p}{1 - p} \quad or \quad p = \frac{Z - 1}{Z + 1} .$$

The output samples $S(x_j)$ are expressed in terms of the output samples $E(x_j)$ by the relation

$$S(x_j) = \frac{1 + k_2}{1 + k_1} E(x_j) + \frac{1 - k_2}{1 + k_1} E(x_{j-1}) - \frac{1 - k_1}{1 + k_1} S(x_{j-1}) .$$

Four types of filters can be obtained :

a) for $k_1 = 0$ and $k_2 \neq 0$, the corresponding transfer function $|H(f)|^2$ is proportional to f^2. This is a derivative filter.

b) for $k_2 = 0$ and $k_1 \neq 1$, the corresponding transfer function $|H(f)|^2$ is proportional to f^{-2}. This is an integrator filter.

c) for $k_2 = a k_1$ with $k_1 < k_2$, the transfer function corresponds to a pseudo-derivative filter.

d) for $k_2 = b k_1$, with $k_1 > k_2$ the transfer function corresponds to a pseudo-integrator filter.

The association of n pseudo-integrator filters with cut-off frequencies distributed by a geometrical series gives an equivalent 1/f transfer function.

For n filters in series

$$|H(f)|^2 = \left| \prod_{i=1}^{n} \left(\frac{1 + a^i f}{1 + a^{i+1} f} \right) \right|^2 \sim \frac{1}{f} .$$

FFT method

In this method the sequence of random numbers, given by the white generator, is first Fast Fourier Transformed, to obtain the components in the frequency domain, then these components are weighted by the corresponding coefficient of the H(f) transfer function of the desired filter and finally an Inverse Fourier Transform is applied to the filtered components in order to obtain a sequence of random number with the desired 1/f$^{\alpha}$ spectrum.

$$x_j(t) \xrightarrow{FFT} E_n(f) \xrightarrow{H(f)} E_{n\,filt}(f) \xrightarrow{IFT} x_{j\,filt}(t) .$$

Then these noise generators are used to test the Allan variance, Modified Allan variance, and the new Filtered Allan variances.

Comparison of variances

For each type of variance a double check is made (Fig. 9). In the first one the variances are obtained in the frequency domain by means of a FFT of the random signal, then by filtering with the transfer function $H_i(f)$ of the corresponding variance and by integrating where :

$H_1(f)$ corresponds to the Allan variance

$H_2(f)$ corresponds to Modified Allan variance

$H_3(f)$ corresponds to the Filtered Allan variance with $(\pi\tau f^2)$ filter

$H_4(f)$ corresponds to the Filtered Allan variance with f^{-2} filter

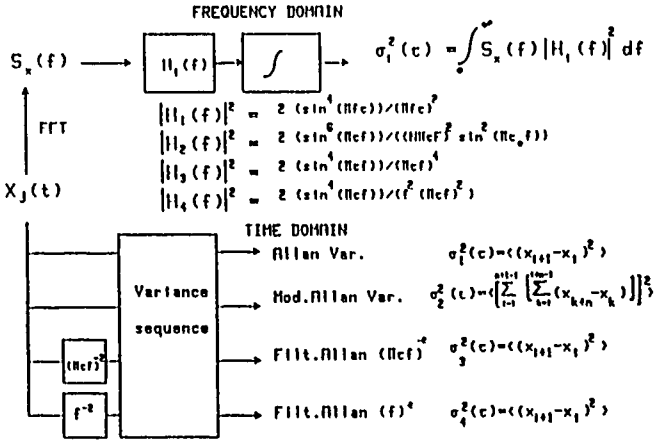


Fig. 9 : Definition of the Filtered Allan variance and schematic diagram of the double check of the four variances in the frequency domain and in the time domain

Simultaneously the second one is the simulation in the time domain performed by calculating the variance with the corresponding algorithm. For the Filtered Allan variance two types of digital filters are applied. The first has a transfer function proportional to $(\pi\tau f)^{-2}$ and the second has a transfer function proportional to f^{-2}. These filters are the digital filters defined in the last paragraph.

An example of this double check for white phase noise is shown on Fig. 10 for a Filtered Allan variance obtained with a (f^{-2}) type filter calculated with 8192 samples.

These curves exhibit a -1 slope as expected. For large values of τ the number of averages decreases and the accuracy of the curve corresponding to the time domain is worse.

To increase the number of averages we can use the Allan variance with overlaping averages. The result of this calculation gives a curve (Fig. 10) which is closer to the curve calculated by transformation from the frequency domain. This means that information on the signal is lost with the Allan variance when the number of averages is too small.

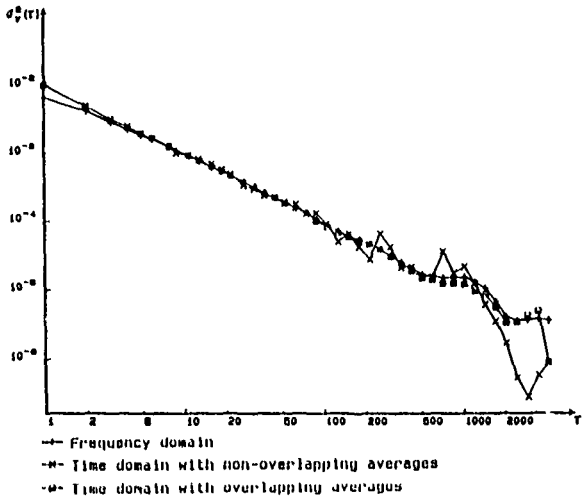


Fig. 10 : Results of the double check for a Filtered Allan Variance obtained with a (f^2) type filter calculated with 8192 samples

These double checks have been made for the four variances in the case of white phase noise, flicker phase noise and white frequency noise. Table I gives the characteristic slopes of the four variances.

$S_y(f)$	Allan variance	Modified Allan variance	Filtered variance	
			$(\pi\tau f)^{-2}$	f^2
2	-2	-3	-3	-1
1	-2	-2	-2	0
0	-1	-1	-1	1

Table I
Characteristic slopes of the different variances

This shows that for white phase noise and flicker phase noise, Allan variance gives the same -2 slope as it well known.

The Modified Allan variance distinguishes these two noises with -3 and -2 slopes. The Filtered Allan variance gives -3 and -2 slopes for filtered transfer function in $(\pi\tau f)^{-2}$ and -1 and 0 for filtered transfer function in f^2.

However if we compare the variances in a log-log plot (Fig. 11) it is clear that the distinction between the two noises (flicker phase noise and white phase noise) is even easier with the Filtered variance since it exhibits an angle of 45° between the asymptot rather than 8° as the Modified variance does.

In Fig. 12 is given the stability curve in the time domain of a 5 MHz quartz oscillator measured with the Filtered (f^2) variance. The value of the transfer function of the digital filter being equal to 1 at 1 Hz, this filter transforms the frequency fluctuations into phase fluctuations and the well known relations between frequency domain and time domain must be shifted by a factor of f^2.

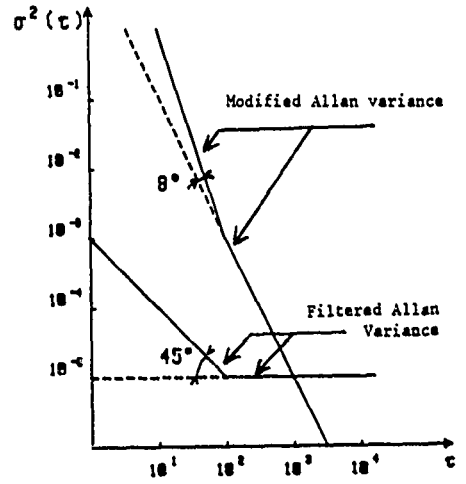


Fig. 11 : Angular difference between the asymptotes of the Modified Allan Variance and the Filtered Allan variance

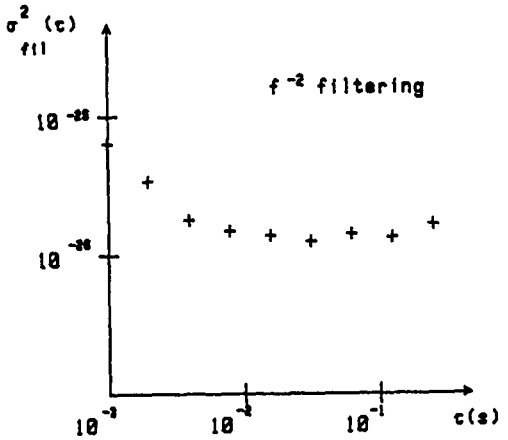


Fig. 12 : Filtered Allan variance of a 5 MHz oscillator showing white phasenoise and flicker phase noise

In other words a white phase noise spectrum will be treated as a white frequency spectrum, a flicker phase spectrum as a flicker frequency spectrum and so on.

For a white phase noise spectrum $S_y(f)$ the Filtered Allan Variance becomes

$$\sigma^2_{filt}(\tau) = h_2/2\tau$$

and for flicker phase noise spectrum $S_y(t)$

$$\sigma^2_{filt}(\tau) = h_1 . 2ln2$$

The h_1 and h_2 values of the signal spectrum can be deduced from the last relations

$$h_1 \simeq 10^{-26}$$

$$h_2 \simeq 10^{-28}$$

References

[1] David W. Allan and James A. Barnes, Proc. 35th Ann. Symp. on Freq. Cont. (1981), p. 470

[2] J.C. Radix, Introduction à la Filtrage Numérique, 1970, Eyrolles.

43rd Annual Symposium on Frequency Control - 1989

EXTENDED VARIANCES AND AUTOREGRESSIVE/MOVING AVERAGE ALGORITHM FOR THE MEASUREMENT AND SYNTHESIS OF OSCILLATOR PHASE NOISE

KIN-WA WAN‡, ENRIC VILAR‡ & JOHN ROBERTS†

‡Portsmouth Polytechnic, Microwave Telecommunications & Signal Processing Research Group, School of Systems Engineering, Anglesea Building, Anglesea Rd., Portsmouth, PO1 3DJ, England.
†Plessey Research Roke Manor Ltd., Department of Mathematics, Roke Manor, Romsey, Hants, SO51 0ZN, England.

Abstract:

This paper is concerned with two related aspects of the recent research work carried out at Portsmouth Polytechnic in the UK regarding numerical and digital techniques for the measurement and synthesis of phase noise in oscillators.

The first aspect concerns the introduction of a statistical quantity termed extended Hadamard variance to be applied in frequency stability analysis by making use of digital frequency counter measurements. The properties, the potential interest and the usefulness of this parameter are briefly demonstrated and applied to the study of a signal from a synthesizer f.m. modulated by white noise followed by the frequency counter and signal processing. It is shown that not only the Hadamard variance but the Allan variance and the extended two-sample variance are such sets of this variance.

The second aspect quite important and closely related to the first one refers to modern requirements for numerical simulation of random noise, emulating phase or frequency fluctuations, and with the following features: (a) variable slope α of the power spectrum $S_y(f) \sim f^{\alpha}$ with: α a positive or negative real number (not necessarily integer), and (b), in parallel with (a), the ability to generate noise with a predetermined amplitude distribution (uniform, Gaussian, etc). The paper describes the autoregressive integrated moving average technique used to generate the various types of random noise including the radio propagation f.m. noise ($S_y(f) \sim f^{-2/3}$) and the Monte Carlo approach to transform from one distribution to another.

1. Introduction

With the development of high-performance, high resolution and programmable electronic counters, cost-efficient FM noise measurements can be carried out for Fourier frequencies not too far away from the carrier. The use of digital frequency counters for the measurement of frequency stability has progressively gained acceptance because counters are relatively insensitive to amplitude variations of a signal after suitable conditioning of the signal. The very high-performance oscillator which is used in an electronic counter using the reciprocal method permits high accuracy of measurement. A reciprocal counter uses a high frequency internal clock and the period of the unknown signal is computed from the counting of the number of clock cycles within a given number of signal periods.

By making use of the counter readings $\bar{\upsilon}$, several algorithms for frequency stability analysis have been developed [1,2,5,8]. A common, fast and simple way to *measure* and *characterize* the phase noise processes (except flicker and white phase noise) present in an oscillator is to carry out the Allan variance analysis [2]. Allan has also suggested a relatively complicated quantity called modified Allan variance [1] to be used in the identification of the flicker and white phase noise processes in the time domain studies of the frequency stability of an oscillator.

Similar to the approach of the Allan variance estimation Baugh [5] proposed to use a statistical quantity termed Hadamard variance, $\langle\sigma_H^2(N,T,\tau)\rangle$, resembling the Hadamard transform [12], for deriving the spectral density of phase noise. One can shows that the associated spectral window is narrower than the Allan one, that is, the equivalent noise bandwidth (*ENB*) of the Hadamard and Allan spectral windows are $1.2337N^{-1}\tau^{-1}$ and $0.476\tau^{-1}$ respectively.

Similar to the concept suggested in [11], Faulkner and Vilar [8] have independently introduced a parameter termed extended Allan variance or two-sample variance for the characterization and the measurement of long-term or close-to-carrier phase noise in oscillator. Extending the concept of the extended Allan variance analysis to the Hadamard variance approach [5], we intend to introduce a new parameter termed extended Hadamard variance [17] used in frequency stability analysis. An attempt has been made in this paper to present the properties, and the potential interest and the usefulness of this extended Hadamard variance.

In parallel with the development of the measurement and characterization techniques for the frequency/phase noise studies [17], we focus on the synthesis of random phase noise processes presented in the oscillators and radio propagation signals [9,16]. A modeling method called autoregressive integrated moving average (**ARIMA**) technique [4,6] employed in the random noise simulations is described. This approach consists in generating $1/f^{\alpha}$ noise process, with $-1 \le \alpha \le 1$. A linear model offering the generation of the radio propagation-like F.M. noise ($\alpha \approx -2/3$) is presented.

Furthermore, one of Monte Carlo methods [15] used in the generation of random numbers with a fairly wide variety of distribution functions is explained. The reason is that the random processes we occasionally deal with are not confined to a normal probability density function, especially to be the case in the relative short-term studies of the random processes.

To summarize; this paper briefly presents numerical/digital techniques applied in the measurement and synthesis of the phase noise presented in oscillating sources or in radio propagation signals. The experimental and computer simulations are also included.

2. Multiple Sample Variance: Extended Hadamard

Based upon the concepts of [8], we propose to apply the extrapolation technique to the analysis provided by the Hadamard variance [14]. This approach combines the *shifting* capability of the extended two-sample spectral window towards the carrier of the signal by increasing the number of the counts *cascaded* and also offers spectral analysis as high as the conventional Hadamard variance does. This new statistical quantity is called extended Hadamard variance [17] and is defined as

$$\begin{aligned}\langle\bar{\sigma}_H^2(N,kT,k\tau)\rangle &= \frac{1}{\upsilon_o^2 k^2}\langle(\bar{\upsilon}_{k,\tau_{i+1}} - \bar{\upsilon}_{k,\tau_{i+2}} + \bar{\upsilon}_{k,\tau_{i+3}} - \\ &\qquad \cdots \quad \cdots + \bar{\upsilon}_{k,\tau_{i+N-1}} - \bar{\upsilon}_{k,\tau_{i+N}})^2\rangle \\ &= \frac{1}{k^2}\langle(\bar{y}_{k,\tau_{i+1}} - \bar{y}_{k,\tau_{i+2}} + \bar{y}_{k,\tau_{i+3}} - \\ &\qquad \cdots \quad \cdots + \bar{y}_{k,\tau_{i+N-1}} - \bar{y}_{k,\tau_{i+N}})^2\rangle \\ &= \langle[y(t) * h_{H_k}(t)]^2\rangle \end{aligned} \tag{1}$$

where N is an even number greater than 2, k is an integer greater than or equal to 2 and υ_o is the carrier frequency set-up in an oscillator. $\bar{\upsilon}_{k,\tau_i}$ represents the additions of k successive frequency counts (not time intervals) measured with the gate time τ, and $\bar{y}_{k,\tau_i}$ represents the addition of the k successive fractional frequency values at the *ith* time interval. In *eq.* (1), * denotes convolution (see [8]) and $h_{H_k}(t)$ called extended Hadamard function is expressed as

$$h_{H_k}(t) = \sum_{i=0}^{N-1}\frac{1}{N\tau}(-1)^i\{\sum_{q=0}^{k-1}[\gamma(t-(ik+q)(\tau+\tau_d)) - \gamma(t-(ik+q)(\tau+\tau_d)-\tau)]\} \tag{2}$$

in the time domain, where τ_d is the dead-time between successive frequency counts. γ is 0 for $t < 0$ and 1 for $t \ge 0$. The corresponding transfer function expressed in the frequency domain, $H_{H_k}(f)$, is given

CH2690-6/89/0000-331 $1.00 C 1989 IEEE

by

$$|H_{H_k}(f)|=\left|\frac{\sin\pi f\tau}{\pi fk\tau}\frac{\sin\pi fk(\tau+\tau_d)}{\sin\pi f(\tau+\tau_d)}\frac{\sin\pi fkN(\tau+\tau_d)}{\cos\pi fk(\tau+\tau_d)}\right| \quad (3)$$

It is noted in here that the extended Hadamard function exhibits the same sidelobe problems (see [17]) as the Hadamard function [14] if one uses this variance for spectral analysis. The limitations can be minimized like for the Hadamard variance analysis by applying the appropriate weighting factors (see [10, 13]).

The equivalent noise bandwidth (ENB) of $|H_{H_k}(f)|$ is found to be

$$ENB_{H_k}=\frac{1}{4Nk\tau r^2}\frac{\sin^2\frac{\pi}{2k}}{\sin^2\frac{\pi}{2kr}} \quad (4)$$

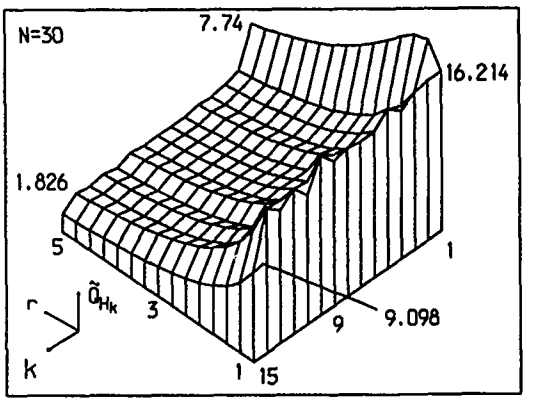
and the maximum value of $|H_{H_k}(f)|$ is located at $f=f_1=\frac{1}{2k(\tau+\tau_d)}$. Following the definition of quality factor Q of a conventional tunned circuit, a parameter called quality factor $\tilde{Q}$ of a function $h(t)$ can be introduced and is defined as $\tilde{Q}=\frac{f_{max}}{ENB}$, instead of $\frac{f_{max}}{BW_{3dB}}$, where f_{max} is the frequency at which the maximum val· e of the transfer function $|H(f)|$ is located, $H(f)$ is the Fourier transform of $h(t)$, and BW_{3dB} is 3 dB bandwidth of $|H(f)|$. Thus, the quality factor $\tilde{Q}_{H_k}$ of the extended Hadamard window can be expressed as

$$\tilde{Q}_{H_k}=\frac{4Nr}{\pi^2}\frac{\sin^2\frac{\pi}{2kr}}{\sin^2\frac{\pi}{2k}} \quad (5)$$

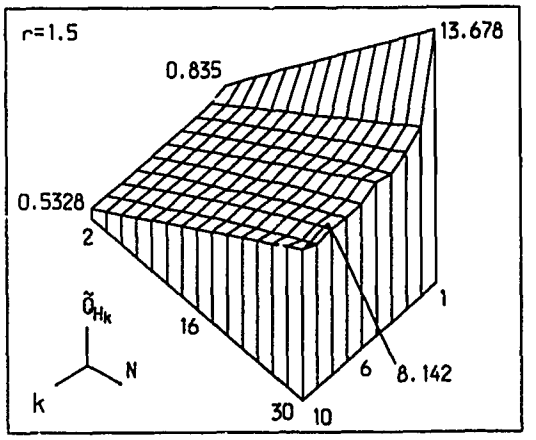
which is shown in *fig.* 1. From *fig.* 1, one can conclude that a good resolution spectral analysis can be achieved using the extended Hadamard variance as the $\tilde{Q}_{H_k}$ of the spectral window is relative high, as long as $N \gg k$ and $1<(\tau+\tau_d)/\tau<3$. Following the concept in [8], the relationship between the extended Hadamard variance and the power spectral density of the fractional frequency deviations $S_y(f)$ can be written as

$$\begin{aligned}\langle\bar{\sigma}^2_H(N,kT,k\tau)\rangle &= \int_0^\infty S_y(f)|H_{H_k}(f)|^2\,df\\ &\approx S_y(f_1)\,ENB_{H_k}|H_{H_k}(f_1)|^2\\ &\approx S_y(f_1)\frac{N}{2k\tau}\end{aligned} \quad (6)$$

Since $f_1=\frac{1}{2k(\tau+\tau_d)}$, one can estimate the power spectral density along the Fourier frequency axis by varying the parameter k. In the other words, one can study the phase noise spectral density $S_\psi(f)$ at the different Fourier frequency by changing the cascading number of the frequency counts as $S_\psi(f)=\frac{\upsilon_o^2}{f^2}S_y(f)$.



(a)



(b)

Figure 1: The quality factor $\tilde{Q}_{H_k}$ of the extended Hadamard function as a function of k, N and r. r is equal to $(\tau+\tau_d)/\tau$.

3. Synthesis of Random Noise Processes

3.1. Autoregressive Integrated Moving Averaging Modeling Technique (ARIMA) [6]

In the process of reviewing the technique called **ARIMA** applied to build a flicker-like random noise model [4], we have found that the **ARIMA** modeling approach can be extended to construct a $1/f^\alpha$ random noise model, with $-1\le\alpha\le 1$.

Suppose we model a noise process $S_y(f)=h_\alpha f^\alpha$ using different single-order low-pass and high-pass two-port networks cascaded in series (see *fig.* 2). The distance between two neighboring poles p_i and p_{i+1} of the two-port networks is assumed to be d in a log-lin scale. That is,

$$p_{i+1}=p_i+d \quad (7)$$

In order to have same positive and negative error for a straight line approximation to the transfer function modeling the $1/f^\alpha$ noise process, the position of the zero z_i is given by

$$z_i=p_i+\frac{\alpha d}{2} \quad (8)$$

Thus, the frequency of the $(i+1)th$ pole, $f_{p,i+1}$ (or $10^{p_{i+1}}$), and the ith zero, $f_{z,i}$ (or 10^{z_i}), can both be found from the frequency of the ith pole $f_{p,i}$ using

$$f_{p,i+1}=f_{p,i}\,10^d \quad \text{and} \quad (9)$$

$$f_{z,i}=f_{p,i}\,10^{\alpha d/2}. \quad (10)$$

(It is noted in here that the Bode plot approach for the design of analog filter, which then permits building a digital filter via matched Z-transformation, has also be used in [7].)

The frequencies of the poles, $f_{p,i}$, or zeros, $f_{z,i}$, of the analog two-port networks and the parameters of the corresponding moving average (**MA(1)**$_i$), θ_i, or auto-regressive (**AR(1)**$_i$), ϕ_i, digital filters are related by [4]

$$\tilde{\theta}_i\approx\frac{1-\pi f_{z,i}}{1+\pi f_{z,i}} \quad \text{and} \quad (11)$$

$$\tilde{\phi}_i\approx\frac{1-\pi f_{p,i}}{1+\pi f_{p,i}}. \quad (12)$$

We can determine the coefficients of the cascaded **AR(1)** or **MA(1)** filters of the desired **ARIMA** model from the frequencies of the poles or zeros for a given value of α, with $0\le\tilde{\phi}_i<1$ and $0\le\tilde{\theta}_i<1$. An example of $\alpha=-2/3$ is shown in table 1.

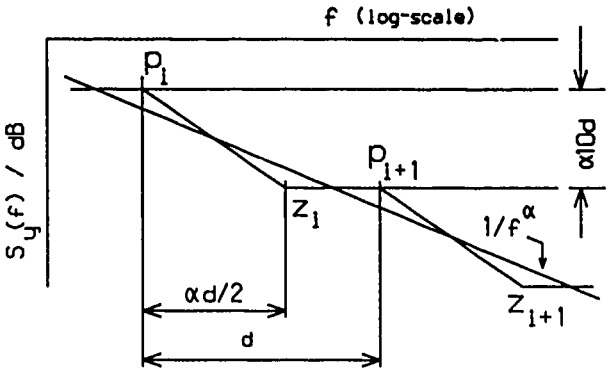


Figure 2:
The asymptotic Bode plot of the transfer function of the analog filter which models a random noise process $h_{\alpha}f^{\alpha}$.

Since the output of the ith MA filter is the input of the ith AR filter (low-pass filtering followed by high-pass filtering), we can consider that the ARIMA(k,0,l=k) linear system can be built by cascading k different ARIMA(1,0,1) models. Then we establish

$$(1-\tilde{\phi}_i B)\, y_{i,n} = (1-\tilde{\theta}_i B)\, a_{i,n} \quad \text{or} \tag{13a}$$

$$y_{i,n} = a_{i,n} - \tilde{\theta}_i\, a_{i,n-1} + \tilde{\phi}_i\, y_{i,n-1} \ , \tag{13b}$$

where $a_{i,n}$ is the input of the ith MA filter, $y_{i,n}$ is the output of the ith AR filter at the nth time interval, and B is a backward shift operator (delay), that is, $B^j a_{i,n} = a_{i,n-j}$. Having calculated $\{\tilde{\phi}_i\ ,\ \tilde{\theta}_i\}$ using the relationships (11) and (12), the concept of *eq.* (13a-b) establishes a system of difference equations representing the desired $1/f^{\alpha}$ noise process. With white noise applied to the input of the designed ARIMA model, one can obtain the desired $1/f^{\alpha}$ noise process.

It is important to note that a smoothing process is used in the simulations, that is, $2N$ random numbers generated by the designed ARIMA model are averaged to generate N-samples which are found to exhibit the best approximation to the desired process. (This *2–to–1* smoothing process has also been applied in generating a flicker-like frequency noise process by means of the Barnes-Jarvis flicker noise generator [3].)

3.2. Distribution Transformation of Random Processes

In the computer simulations of colored noises, a filtering technique is used. White frequency random noises are filtered by means of linear digital filters. The white random noise process being used is not necessary normal and is arbitrary (say uniform). However, the output of the filter exhibits very closely a normal probability density function. The reason is that the digital filtering process

$$v_o(n) = \sum_{m=0}^{n} v_i(m)\, h(m-n) \ , \tag{14}$$

as a discrete convolution, that is, a weighted linear combination (a sum) of random numbers, where v_i is the input (non-Gaussian random) signal to the filter h, and v_o is the output of the filter. According to the Central limit theorem, that sum leads to normal probability density function as long as the impulse response of the filter is longer than the correlation time of the input signal of the filter and the moments of the input process $\{v_i\}$ are finite. One needs also a minimum numbers of terms n, which is usually the case.

In order to extend the studies to non-Gaussian random processes, we use one of the Monte Carlo methods [15] to transform the generated Gaussian process to others (which can be uniform, exponential, Rayleigh, and so on) without altering the power spectral density function of the random process. The fundamental expression connecting random numbers S_i with a given distribution function $f(x)$, to random numbers R_i uniformly distributed over the interval (0,1) is

$$R_i = \int_{-\infty}^{S_i} f(x)\, dx \ . \tag{15}$$

Thus, by taking any arbitrary set of integration limits $\{S_i\}$, we get a set of numbers $\{R_i\}$ which is uniformly distributed in the range (0,1). Now let us define the random variable ξ exhibiting a probability density function $f(x)$, and the distribution of the random variable η being uniform in the interval (0,1). In order to demonstrate the technique of the distribution function transformation, we relate the random variables ξ and η by

$$\eta = \int_{-\infty}^{\xi} f(x)\, dx$$

$$= F(\xi) \quad \text{or} \tag{16a}$$

$$\xi = F^{-1}(\eta) \tag{16b}$$

by following the concept of *eqs.* (15) and (16a-b), and defining $F(\xi)$ as the integral of $f(x)$, where F^{-1} is known as the inverse function of F. ***Clearly F is the cumulative distribution of ξ, the plot of $F(\xi)$ against ξ is a virtual linear curve, and thus the transformation operation is a mild non-linear mapping process.*** One notices that the important aspect in the ***mild non-linear*** transformation approach is the complexity in computing $F(\ )$ and/or $F^{-1}(\ )$. It is also worth noting that η and $(1-\eta)$ exhibit the same uniform distribution in the interval (0,1). Therefore, one may replace $F^{-1}(\eta)$ by $F^{-1}(1-\eta)$ in the simulation of the random process if one wishes (perhaps to simplify the algebraical calculations).

4. Experimental And Computer Simulation Results

The test oscillator consisted of a synthesized carrier F.M. modulated by a white noise degrading the spectral purity and phase in a controlled manner. The frequencies were then monitored by a fast, high-resolution and programmable microcomputer via the IEEE-488 instrumentation bus. The measured frequency counts $v_{(s)}$ can then be processed using *eqs.* (1) and (6) to carry out the Hadamard type of spectral analysis, as well as conventional Fourier analysis. The results are shown in *fig.* 3.

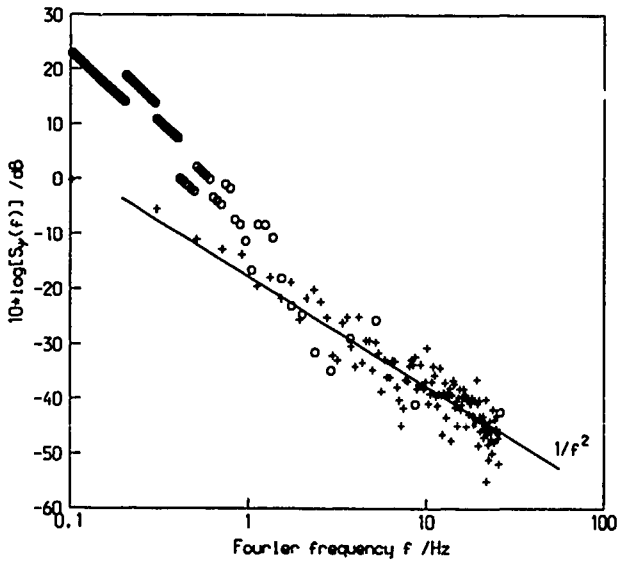


Figure 3: Measured phase noise spectrum. (white F.M.)
+++ using conventional Fourier analysis
ooo using extended Hadamard variance

An attempt to simulate white frequency ($\alpha = 0$), flicker frequency ($\alpha = -1$) and phase ($\alpha = 1$), and radio propagation frequency modulated ($\alpha \approx -2/3$) noise processes exhibiting Gaussian probability density function has been carried out using the **ARIMA** modeling approach. 50 noise sequences of each type of the mentioned processes were used to find the corresponding spectra. Each set contained 1024 samples. The average of the 50 set spectra of each noise processes are shown in *fig.* 4. In addition, the model used to simulate the radio propagation F.M. random noise is shown in table 1.

Applying the **ARIMA** modeling technique and following the concept demonstrated in *eq.* 15, we have also carried out computer simulations of the radio propagation F.M. noise processes ($\alpha \approx -2/3$) with Gaussian and exponential distributions. The selected oscillograms and the results of the corresponding statistical analyses of the Gaussian and exponential-***like*** types of propagation F.M. random noises are shown in figure 5. ***These random processes have zero mean and unity variance in order to compare their spectra.*** The results of the spectral analyses of the simulations are shown in figure 6.

5. Discussion and Conclusion

By comparing the extended Hadamard window presented in this paper with the windows of the Allan [2], extended two-sample [8] and Hadamard variances [5], one concludes that the latter three are only special cases of our extended Hadamard variance. The experimental results shown in *fig.* 3 confirm the possible use of the extended Hadamard variance in frequency stability analysis although the sidelobe structure of $|H_{H_k}(f)|$ can be a disadvantage in the spectral estimation of $S_y(f)$ (see [17]). From *eqs.* (1) and (6), and *fig.* 3, we can conclude that this statistical approach offers a fast, simple and relative accurate way to estimate the spectrum in some frequency ranges where $\tilde{Q}_{H_k}(s)$ are high enough. In the low $\tilde{Q}_{H_k}$ regions of the extended Hadamard analysis, that is, the spectral components shown in *fig.* 3 locating at $f < 1$Hz, a further bias correction will be expected to be applied in the spectral estimation. From the results shown in *fig.* 1, one must notice that $\tilde{Q}_{H_k}$ can be kept at a relative high value as long as $N \gg k$ and $1 < (\tau+\tau_d)/\tau < 3$.

The **ARIMA** modeling technique concerned with the design of analog filters provides a simple way to build the desired $1/f^{\alpha}$-like random noise generator, with $-1 \le \alpha \le +1$. The analog model of the desired noise process is firstly designed by means of the graphical approach called Bode plot. From *eqs.* (11) and (12), the parameters of the **ARIMA** linear system can be easily calculated by the corner frequencies of the corresponding analog filter. The qualities of the simulations have been justified by means of conventional Fourier analysis (see *fig.* 4). Also the success of the synthesis of the random noise process with a theoretical roll-off of -2/3 shows a powerful mean for the studies of radio propagation and frequency standards.

Furthermore, we have considered the statistical properties of the random noise simulations. One of the Monte Carlo methods used to transform the probability density function of a random process to another has been presented. From the results shown in *figs.* 5 and 6, we find that numerical synthesis of random processes with a selected first order probability density and power spectral density function can be achieved by following the concepts demonstrated in section 3. In the other words, the simulation results show that as expected there are no perceptible relationship between the first order probability density function and the corresponding power spectrum density function of a random process. However, since the spectrum analysis is implicated in the second order probability density function, an area concerning with the ***virtual linear mapping process*** and the second order probability density function is exposed for a further investigation.

Table 1: The parameters used to model the radio propagation f.m. noise process.

$d \approx 0.409836$	$\alpha \approx -2/3$	
i	$\tilde{\phi}_i$	$\tilde{\theta}_i$
1	0.993883	0.991631
2	0.984357	0.978637
3	0.960295	0.946014
4	0.901064	0.866925
5	0.764105	0.690403
6	0.488558	0.359999

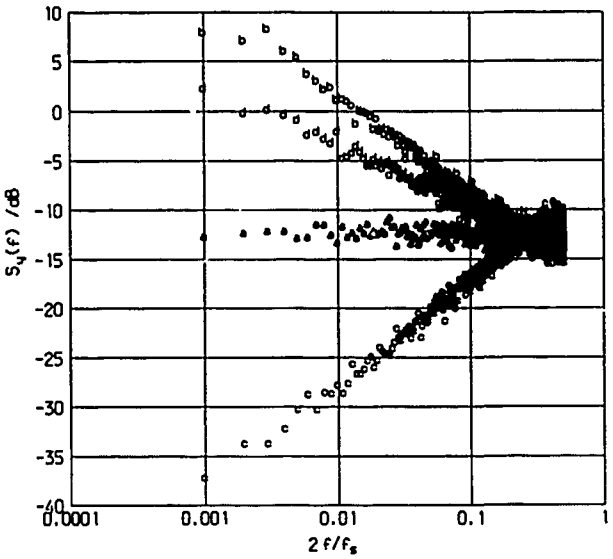


Figure 4: The power spectra of the computer simulated random noises. f_s is the sampling frequency.
aaa: white frequency noise process.
bbb: flicker frequency noise process.
ccc: flicker phase noise process.
ddd: propagation frequency modulated noise process.

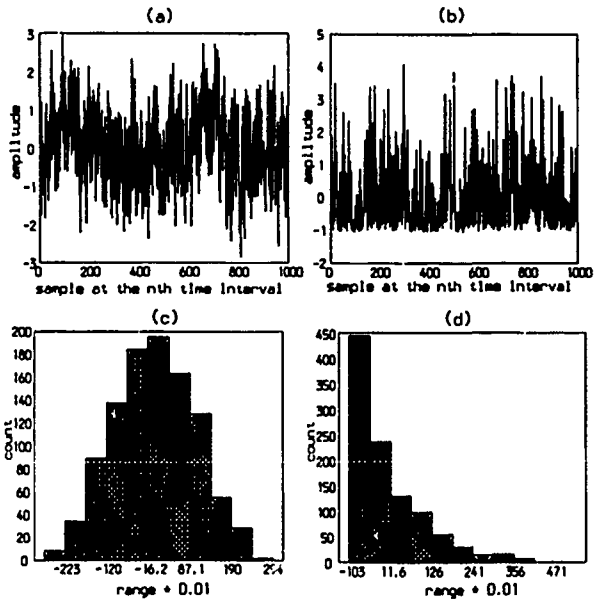


Figure 5: (a) & (b) The oscillograms of the simluated radio propagation f.m. random noises.
(c) The histogram of the simulations shown in (a).
(d) The histogram of the simulations shown in (b).

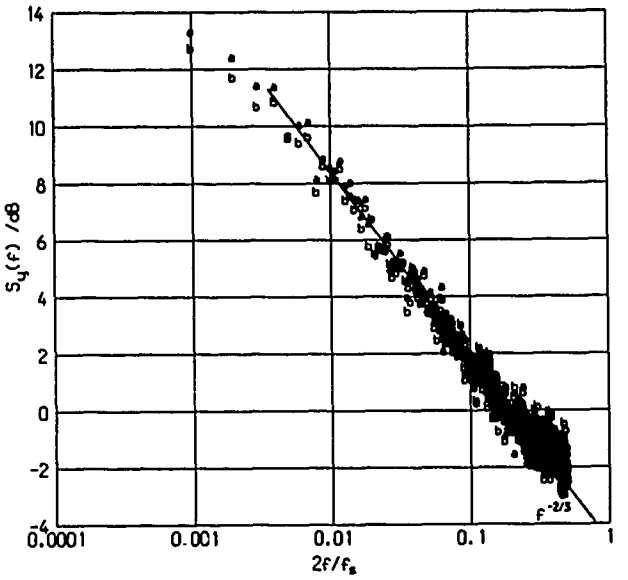


Figure 6: The power spectra of the computer simulated propagation F.M. noises.
f_s is the sampling frequency.
aaa: Gaussian distributed random process.
bbb: Exponentially distributed random process.

Acknowledgment

This project is supported by a research grant from the U.K. National Advisory Body.

References

[1]. Allan, D.W. and Barnes, J.A., "A Modified "Allan Variance" With Increased Oscillator Characterization Ability," in *Proc. 35th Ann. Freq. Control Symposium, USAERADCOM, PP. 470-475*, (May 1981).

[2]. Allan, D.W. and Walls, F.L., "Measurements of Frequency Stability," in *Proceedings of the IEEE, Vol. 74, No. 1, PP. 162-168*, (Jan. 1986).

[3]. Barnes, J.A. and Jarvis, S. Jr., "Efficient Numerical and Analog Modeling of Flicker Noise Processes," in *NBS Technical Noise 604*, (June 1971).

[4]. Barnes, J.A., "Models For The Interpretation of Frequency Stability Measurements," in *NBS Technical Note 683*, (Aug. 1976).

[5]. Baugh, R.A., "Frequency Modulation Analysis With The Hadamard Variance," in *Proc. 25th Ann. Freq. Control Symp., Atlantic City NJ, PP. 222-225*, (April 1971).

[6]. Box, G.E.P. and Jenkins, G.M., "Time Series Analysis: Forecasting and Control," in *Holen-Day, San Francisco, Calif.*, (1970).

[7]. Corsini, G. and Saletti, R., "A $1/f^{\gamma}$ Power Spectrum Noise Sequence Generator," in *IEEE Transaction on Instrumentation and Measurement, Vol. 37, No. 4, PP. 615-619*, (Dec. 1988).

[8]. Faulkner, N.D. and Vilar, E., "Time-Domain Analysis of Frequency Stability Using Nonzero Dead-Time Counter Techniques," in *IEEE Trans. On Instrument. and Measure., Vol. IM-34, No. 2, PP. 144-151*, (June 1985).

[9]. Lawrence, R.S. and Strohbehn, J.W., "A survey of Clear-Air Propagation Effects Relevant to Optical Communications," in *Proceedings of the IEEE, Vol. 58, No. 10, PP. 1523-1546*, (Oct. 1970).

[10]. Lesage, P. and Audoin, C., "Instabilité De Fréquence Des Oscillateurs: Measure Dans Le Domaine Fréquence Par Une Méthode Temporelle," in *L'onde eletrique, Vol. 55, No. 2, PP. 82-89*, (1975).

[11]. Lesage, P., "Characterization of Frequency Stability Bias Due To The Juxtaposition of Time-Interval Measurements," in *IEEE Trans. Inst. & Meas., Vol. IM-32, PP. 204-207*, (March 1983).

[12]. Pratt, W.K., Kane, J., and Andrews, H.C., "Hadamard Transform Image Coding," in *Proceedings of The IEEE, Vol. 57, No. 1, PP. 38-67*, (Jan. 1969).

[13]. Rutman, J., "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," in *Proceedings of The IEEE, Vol. 66, No. 9, PP. 1048-1075*, (Sep. 1978).

[14]. Sauvage, G. and Rutman, J., "Analyse Spectrale Du Bruit De Fréquence Des Oscillateurs Par La Variance De Hadamard," in *Ann. des Téléc., 28, No. 7-8, PP. 304-314*, (1973).

[15]. Shreider, Yu.A., "Method of Statistical Tests (Monte Carlo Methods)," in *Fizmatigiz, Moscow, Russian*, (1962).

[16]. Vilar, E. and Matthews, P.A., "Propagation Noise and Frequency Stability At 36GHz," in *IEE conference: Propagation of Radio Waves at Frequencies above 10GHz, Publication Number 98, PP. 83-89*, (April 1973).

[17]. Wan, K.W., Vilar, E., and Roberts, J.H., "Improved numerical technique for oscillator frequency stability analysis," in *IEE Electronics Letters, Vol. 25, No. 1, PP. 71-72*, (5th Jan. 1989).

43rd Annual Symposium on Frequency Control - 1989

BIASES AND VARIANCES OF SEVERAL FFT SPECTRAL ESTIMATORS AS A FUNCTION OF NOISE TYPE AND NUMBER OF SAMPLES

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

D. B. Percival, Applied Physics Laboratory,
University of Washington, Seattle, WA 98106

W. R. Irelan, Irelan Electronics,
412 Janet, Tahlequah, OK 74464

Abstract

We theoretically and experimentally investigate the biases and the variances of Fast Fourier transform (FFT) spectral estimates with different windows (data tapers) when used to analyze power-law noise types f^0, f^{-2}, f^{-3} and f^{-4}. There is a wide body of literature for white noise but virtually no investigation of biases and variances of spectral estimates for power-law noise spectra commonly seen in oscillators, amplifiers, mixers, etc. Biases (errors) in some cases exceed 30 dB. The experimental techniques introduced here permit one to analyze the performance of virtually any window for any power-law noise. This makes it possible to determine the level of a particular noise type to a specified statistical accuracy for a particular window.

I. Introduction

Fast Fourier transform (FFT) spectrum analyzers are very commonly used to estimate the spectral density of noise. These instruments often have several different windows (data tapers) available for analyzing different types of spectra. For example, in some applications spectral resolution is important; in others, the precise amplitude of a widely resolved line is important; and in still other applications, noise analysis is important. These diverse applications require different types of windows.

We theoretically and experimentally investigate the biases and variances of FFT spectral estimates with different windows when used to analyze a number of common power-law noise types. There is a wide body of literature for white noise but virtually no investigation of these effects for the types of power-law noise spectra commonly seen in oscillators, amplifiers, mixers, etc. Specifically, we present theoretical results for the biases associated with two common windows — the uniform and Hanning windows — when applied to power-law spectra varying as f^0, f^{-2} and f^{-4}. We then introduce experimental techniques for accurately determining the biases of any window and use them to evaluate the biases of three different windows for power-law spectra varying as f^0, f^{-2}, f^{-3} and f^{-4}. As an example we find with f^{-4} noise that the uniform window can have errors ranging from a few dB to over 30 dB, depending on the length of span of the f^{-4} noise.

We have also theoretically investigated the variances of FFT spectral estimates with the uniform and Hanning windows (confidence of the estimates) as a function of the power-law noise type and as a function of the amount of data. We introduce experimental techniques that make it relatively easy to independently determine the variance of the spectral estimate for virtually any window on any FFT spectrum analyzer. The variance that is realized on a particular instrument depends not only on the window but on the specific implementation in both hardware and software. We find that the variance of the spectral density estimates for white noise, f^0, is very similar for three specific windows available on one instrument and almost identical to that obtained by standard statistical analysis. The variances for spectral density estimates of f^{-4} noise are only 4% higher than that of f^0 noise for two of the windows studied. The third window — the uniform window — does not yield usable results for either f^{-3} or f^{-4} noise.

Based on this work it is now possible to determine the minimum number of samples necessary to determine the level of a particular noise type to a specified statistical accuracy as a function of the window. To our knowledge this was previously possible only for white noise — although the traditional results are generally valid for noise that varied as $f^{-\beta}$, where β was equal to or less than 4.

II. Spectrum Analyzer Basics

The spectrum analyzer which was used in the experimental work reported here is fairly typical of a number of such instruments currently available from various manufacturers. The basic measurement process generally consists of taking a string of $N_s = 1024$ digital samples of the input wave form, which we represent here by $X_1, X_2, \ldots, X_{N_s}$. The basic measurement period was 4 ms. This yields a sampling time $\Delta t = 3.90625$ μs. Associated with the FFT of a time series with N_s data points, there are usually $(N_s/2) + 1 = 513$ frequencies

$$f_j = \frac{j}{N_s \Delta t}, \qquad j = 0, 1, \ldots, N_s/2.$$

The fundamental frequency f_1 is 250 Hz, and the Nyquist frequency $f_{N_s/2}$ is 128 kHz. Since the spectrum analyzer uses an anti-aliasing filter which significantly distorts the high frequency portion of the spectrum, the instrument only displays the measured spectrum for the lowest 400 nonzero frequencies, namely, $f_1 = 250$ Hz, $f_2 = 500$ Hz, ..., $f_{400} = 100$ kHz.

The exact details of how the spectrum analyzer estimates the spectrum for $X_1, \ldots, X_{N_s}$ are unfortunately not provided in the documentation supplied by the manufacturer, so the following must be regarded only as a reasonable guess on our part as to its operation (see [1] for a good discussion on the basic ideas behind a spectrum analyzer; two good general references for spectral analysis are [2] and [4]). The sample mean,

$$\bar{X} \equiv \frac{1}{N_s} \sum_{t=1}^{N_s} X_t,$$

is subtracted from each of the samples, and each of these "demeaned" samples is multipled by a window h_t (sometimes called a data taper) to produce

$$X_t^{(h)} = h_t \left(X_t - \bar{X} \right).$$

The spectral estimate,

$$\hat{S}_1(f_j) = \Delta t \left| \sum_{t=1}^{N_s} X_t^{(h)} e^{-i 2\pi f_j t \Delta t} \right|^2, \qquad j = 0, 1, \ldots, N_s/2,$$

is then computed using an FFT algorithm.

The subscript "1" on $\hat{S}_1(f_j)$ indicates that this is the spectral estimate formed from the first block of N_s samples. A similar spectral estimate $\hat{S}_2(f_j)$ is then formed from the second block of contiguous data $X_{N_s+1}, X_{N_s+2}, \ldots, X_{2N_s}$. In all, there are N_b different spectral estimates from N_b contiguous blocks, and the spectrum analyzer averages these together to form

$$\hat{S}(f_j) \equiv \frac{1}{N_b} \sum_{k=1}^{N_b} \hat{S}_k(f_j). \tag{1}$$

It is the statistical properties of $\hat{S}(f_j)$ with which we are concerned in this paper.

Unfortunately some important aspects of the windows are not provided in the documentation for the instrument. One important detail is the manner in which the window is normalized. There are two common normalizations:

$$\sum_{t=1}^{N_s} \left[h_t (X_t - \bar{X})\right]^2 = \frac{1}{N_s} \sum_{t=1}^{N_s} (X_t - \bar{X})^2$$

and

$$\sum_{t=1}^{N_s} h_t^2 = 1. \tag{2}$$

The first of these is common in engineering applications because it ensures that the power in the windowed samples $X_t^{(h)}$ is the same as in the original demeaned samples; the second is equivalent to the first in expectation and is computationaly more convenient, but it can result in small discrepancies in power levels. Either normalization affects only the level of the spectral estimate and not its shape.

There are three windows built into the spectrum analyzer used here. The first is the uniform (rectangular, default) window $h_t^{(U)} = 1/\sqrt{N_s}$. The second is the Hanning data window, for which there are several slightly different definitions in the literature. In lieu of specific details, we assume the following symmetric definition:

$$h_t^{(H)} = C^{(H)} \left(1 - \cos \frac{2\pi(t-0.5)}{N_s}\right), \quad 1 \le t \le \frac{N_s}{2},$$
$$= h_{N_s-t+1}^{(H)}, \quad \frac{N_s}{2} + 1 \le t \le N_s;$$

here $C^{(H)}$ is a constant which forces the normalization in Equation (2). The third window is a proprietary "flattened peak" window, about which little specific information is available (it is evidently designed to accurately measure the heights of peaks in a spectrum).

III. Expected Value and Bias of Spectral Estimates

III.A. Theoretical Analysis

We need to assume a noise model for the X_t's in order to determine the statistical properties of $\hat{S}(f_j)$ in Equation (1). We consider three different models, each of which is represented in terms of a Gaussian white noise process ϵ_t with mean zero and variance σ_ϵ^2. The second-order properties of each model are given by a spectral density function $S(\,)$ defined over the interval $[-1/(2\Delta t), 1/(2\Delta t)]$ in cycles/Δt. The first model is a discrete parameter, white noise process (f^0 noise):

$$X_t = \epsilon_t \quad \text{and} \quad S(f) = \sigma_\epsilon^2 \Delta t.$$

The second model is a discrete-parameter, random-walk process (nominally f^{-2} noise):

$$X_t = \sum_{s=1}^{t} \epsilon_s \quad \text{and} \quad S(f) = \frac{\sigma_\epsilon^2 \Delta t}{4 \sin^2(\pi f \Delta t)}.$$

The third model is a discrete-parameter, random-run process (nominally f^{-4} noise):

$$X_t = \sum_{r=1}^{t} \sum_{s=1}^{r} \epsilon_s \quad \text{and} \quad S(f) = \frac{\sigma_\epsilon^2 \Delta t}{16 \sin^4(\pi f \Delta t)}.$$

Continuous parameter versions of these three models have been used extensively in the literature as models for noise commonly seen in oscillators.

For each of the three models we have derived expressions for $E\{\hat{S}(f)\}$, the expected value of $\hat{S}(f)$. These expressions depend on the window h_t, the number of samples N_s in each block and — in the case of a random-run process — the number of blocks N_b. The details behind these calculations will be reported elsewhere [3]; here we merely summarize our conclusions for the three models in combination with the uniform and Hanning windows and $N_s = 1024$.

First, for a white noise process,

$$E\{\hat{S}(f_j)\} = S(f_j), \qquad j = 1, 2, \ldots, 512,$$

when the uniform window is used. For the Hanning window, the above equality also holds to a very good approximation for $2 \le j \le 511$ and to within 0.8 dB for $j = 1$ and 512 (the latter is of no practical importance since the highest frequency index given by the spectrum analyzer is $j = 400$). These theoretical calculations agree with our experimental data except at f_1 (see Table 1).

Second, for a random-walk process,

$$E\{\hat{S}(f_j)\} = 2S(f_j), \qquad j = 1, 2, \ldots, 512,$$

when the uniform window is used, i.e., the expected value is *twice* what it should be at all frequencies. This theoretical result has been verified by Monte Carlo simulations, but it does *not* agree with our experimental data, which shows no significant level shift in the estimated spectrum. The source of this discrepancy is currently under investigation, but it may be due to either (a) factors in the experimental data which effectively make it band-limited, random-walk noise, i.e., its spectral shape is markedly different from f^{-2} for, say, $0 < f < f_1$ or (b) an incorrect guess on our part as to how the spectral estimate is normalized by the spectrum analyzer. For the Hanning window, we found that

$$E\{\hat{S}(f_j)\} = \begin{cases} 1.08S(f_j) & j = 1; \\ 1.48S(f_j) & j = 2; \\ 1.15S(f_j) & j = 3; \\ 1.07S(f_j) & j = 4; \\ 1.04S(f_j) & j = 5; \\ S(f_j) & 6 \le j \le 511 \text{ to within } 3\%, \end{cases}$$

i.e., $\hat{S}(f_j)$ is essentially an unbiased spectral estimate except for the lowest few frequencies. This theoretical result has been verified by Monte Carlo simulations and also agrees in general with our experimental data.

Third, for a random-run process,

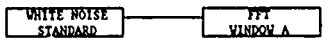
$$E\{\hat{S}(f_j)\} = C_{N_s} f^{-2}, \qquad 1 \le j \le 400,$$

to a good appoximation when the uniform window is used, where C_{N_b} is a constant which depends on the number of blocks N_b and increases as N_b increases. Thus the shape of $E\{\hat{S}(f_j)\}$ follows that of a random-walk process (f^{-2}) rather than that of a random-run process (f^{-4}). This shape has been verified experimentally (see the next subsection), but the dependence of the level on N_b has not. The increase in level of $E\{\hat{S}(f_j)\}$ as N_b increases is due to the fact that the expected value of the sample variance of a block of N_s samples increases with time — by contrast, it is constant with time for the white noise and random-walk cases. For the Hanning window, we found that

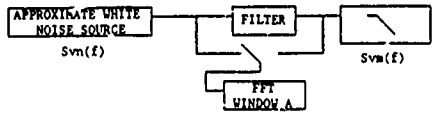
$$E\{\hat{S}(f_j)\} = C'_{N_b} S(f_j), \qquad 4 \le j \le 400,$$

to a good approximation, where again C'_{N_b} is a constant — different from C_{N_b} — which depends on the number of blocks N_b and increases as N_b increases. For frequencies less than f_4 the theoretical results indicate significant (greater than 4%) distortion in the shape, but these do not agree in detail with the experimental values reported in Table 1. For $f_j \ge f_4$ the shape has been verified experimentally, but the dependence of the level on N_b has not. The discrepancy in level between the theoretical and experimental results is yet to be resolved, but it is probably due to a mismatch between the assumed random-run model and the true spectrum for the data (possibly band-limited random-run).

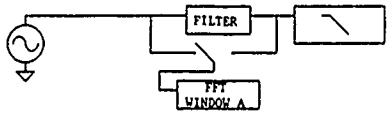
1) Verify Spectral Density Function & Voltage Reference



2) Measure Spectrum Relative to Noise Source Requires Multiple Scans



3) Measure Filter Transfer Function h(f)



4) Calculate Biases

$$B(f) = S_{vm}(f) - h^2(f)\, S_{vn}(f)$$

Figure 1. Outline of measurement procedure for determining the biases in spectral estimators.

III.B. Experimental Determination

The following procedure can be used to experimentally determine the bias in the spectral estimate of any noise spectrum using any window in a particular instrument. The basic concept is to implement a filter that, when applied to white noise, mimics the approximate noise spectra of interest and then measures the level of the white noise and the filter transfer function in a way which has high precision and accuracy as illustrated in Figure 1. First, the level of a known white noise is measured over a convenient range. The higher the frequency span the faster that this is accomplished. Obviously, the chosen range must be one over which the noise source is accurate. To obtain a precision of order 0.2 dB generally requires 1000 samples. This measurement verifies that the spectral density function and the internal reference voltage of the FFT are accurately calibrated and working properly. Virtually all of the windows accurately determine the value of white noise if the first few channels are ignored as explained above. Figure 2 shows the measurement of a noise source, which has been independently determined to have a noise spectral density of 99.8 dBV/Hz by the three windows. (Appendix A shows the circuit diagram for this noise source which has an accuracy of better than 0.2 dB for frequencies from 20 to 20 kHz.)

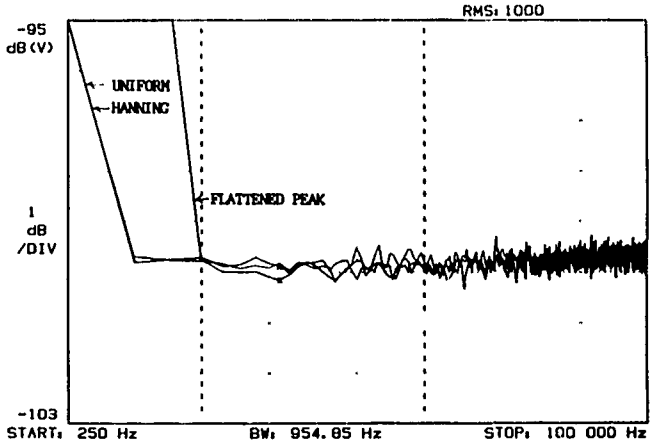
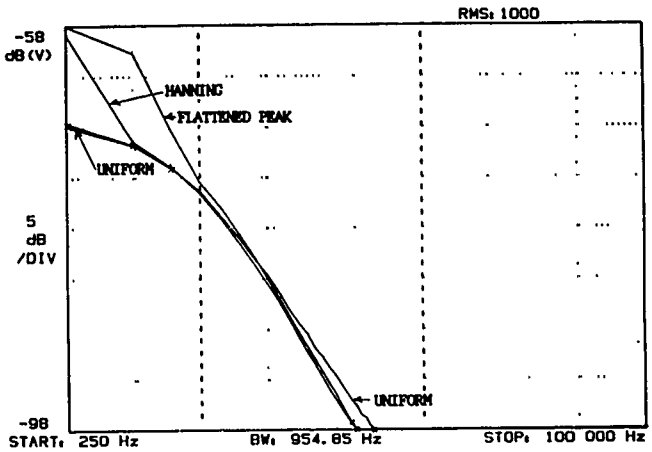
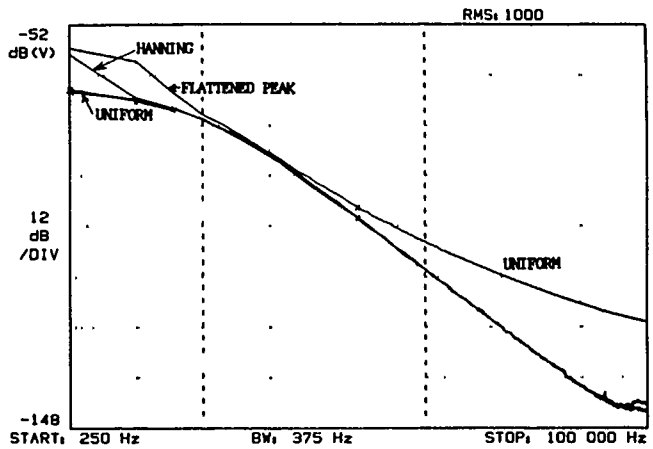


Figure 2. Spectral estimation of a white noise standard using the uniform, Hanning and the proprietary "flattened peak" windows.

Second, an approximately flat spectrum is measured over the frequency range of interest. It is not important if there are small variations in the level that change slowly over the frequency span. Third, the transfer function of the filter is determined for the frequencies of interest using a very narrow spectral source (typically an audio oscillator is sufficient). The very narrow source is accurately measured by the window since there is no problem with either high frequency or low frequency noise biasing the estimate. The use of a window with a flattened peak response is helpful but not necessary if the frequency source is sufficiently stable. This transfer function is then applied to the measured white noise spectrum in step two above. This yields a very accurate value for the "true" spectral density of the white noise source as measured through the filter. This "true value" is then compared to that obtained by the FFT analyzer. The difference between that measured in steps two and three and that measured directly with the FFT is the bias in the spectral estimate for that particular window and noise type. The accuracy of this approach comes from the fact that the calibration has been broken up into steps that can individually be determined with high precision and very small bias. The primary assumption is that the FFT analyzer is linear. Even this assumption can be checked by using precision attenuators. If the known white noise in step one does not extend to the frequencies of interest, then there is an additional assumption that the FFT is flat with frequency. This assumption is nearly always good except perhaps near the last few channels where the effect of the antialiasing filter might cause small inaccuracies.

Figures 3a and 3b show the "true" spectral estimate and the estimates as measured on a particular instrument using the uniform, Hanning, and the instrument's proprietary "flattened peak" windows for noise that varies as f^{-4} over much of the





Figures 3a (top) and 3b (bottom). Difference between the true spectrum (top) which varies approximately as f^{-4} and that estimated by the uniform, Hanning and proprietary "flattened peak" windows (bottom).

range from 1 kHz to 100 kHz. The scan is 0 to 100 kHz, and 1000 samples were taken for all curves. Note the considerable difference between the spectral estimates for channels 1 to 3 for the Hanning and proprietary "flattened peak" windows. These results confirm the theoretical calculations above showing that, for the Hanning window, the first 3 channels should be ignored. For the "flattened peak" window, the first 14 channels should be ignored. For both f^{-3} and f^{-4} noise, the uniform window does not yield usable spectral estimates over *any* portion of the scan. Note in this example that at frequencies above 80 kHz there is a small step in the spectral estimates. This is due to digitizing errors of the signal due to quantization. If the digitizer had more bits, these errors would not occur. This problem of dynamic range is common whenever the spectrum of interest covers many decades. The usual solution is to use filters to divide the spectrum into various frequency range segments which are suitable for the dynamic range of the FFT.

Table 1 summarizes the measured experimental biases in the spectral estimates of a particular instrument with three different windows for power-law noise types varying from f^0 to f^{-4}. This covers most of the random types of noise found in oscillators and signal processing equipment. We do *not* advocate using the biases reported in this table to correct data — they only indicate which channels should not be relied upon for data analysis.

Table 1. Approximate Biases in FFT Spectral Estimates

noise type f^0			
channel #	uniform	Hanning	flattened peak
1	19.6 dB	19.6 dB	20.1 dB
2	small	small	16.7 dB
3	↓	↓	7.2 dB
4			small
5			↓

noise type f^{-4}			
channel #	uniform	Hanning	flattened peak
1	unusable	8.6 dB	10.0 dB
2		0.4 dB	9.1 dB
3		0.4 dB	4.0 dB
4		small	1.2 dB
5		↓	1.1 dB
6			1.1 dB
7			1.0 dB
8			0.8 dB
9			0.6 dB
10			0.6 dB
11			0.5 dB
12			0.4 dB
13			0.4 dB
14			small
15			↓

IV. Variances of Spectral Estimates

IV.A. Theoretical Analysis

We have derived expressions for $var\{\hat{S}(f)\}$ — the variance of $\hat{S}(f)$ — for each of the three models considered in Section III.A. These expressions depend primarily on the number of blocks N_b. Again, the details behind these calculations will be reported elsewhere [3].

First, for a white noise process, the uniform window yields

$$var\{\hat{S}(f_j)\} = S^2(f_j)/N_b, \qquad 1 \le j \le 511,$$

while the Hanning window yields

$$var\{\hat{S}(f_j)\} = \begin{cases} 0.69S^2(f_j)/N_b, & j = 1; \\ S^2(f_j)/N_b, & 2 \le j \le 510; \\ 1.03S^2(f_j)/N_b, & j = 511. \end{cases}$$

These results are consistent with our experimental results and with standard statistical theory.

Second, for a random-walk process, the uniform window yields

$$var\{\hat{S}(f_j)\} = 5S^2(f_j)/N_b, \quad 1 \le j \le 511,$$

while the Hanning window yields

$$var\{\hat{S}(f_j)\} = \begin{cases} 1.30S^2(f_j)/N_b, & j = 1; \\ 2.20S^2(f_j)/N_b, & j = 2; \\ 1.31S^2(f_j)/N_b, & j = 3; \\ 1.15S^2(f_j)/N_b, & j = 4; \\ 1.09S^2(f_j)/N_b, & j = 5; \\ 1.06S^2(f_j)/N_b, & j = 6; \\ 1.04S^2(f_j)/N_b, & j = 7; \\ S^2(f_j)/N_b, & 8 \le j \le 511 \text{ to within 3\%.} \end{cases}$$

Except for the few lowest frequencies, the results for the Hanning window agree with our experimental results and with standard statistical theory; however, the factor of five in the variance for the uniform window disagrees with our experiments and with standard theory (although it has been verified by Monte Carlo techniques). The cause of this discrepancy is under investigation, but we think it is due to the band-limited nature of the experimental data.

Third, for a random-run process, the variance computations are not useful since the variance is dominated by the fact that the expected value of the sample variance for each block of samples increases with time. The agreement which we found between standard statistical theory and our experimental results on the $1/N_b$ rate of decrease of variance is undoubtedly due to the band-limited nature of the experimental data. We will attempt to verify these conclusions in the future using Monte Carlo techniques.

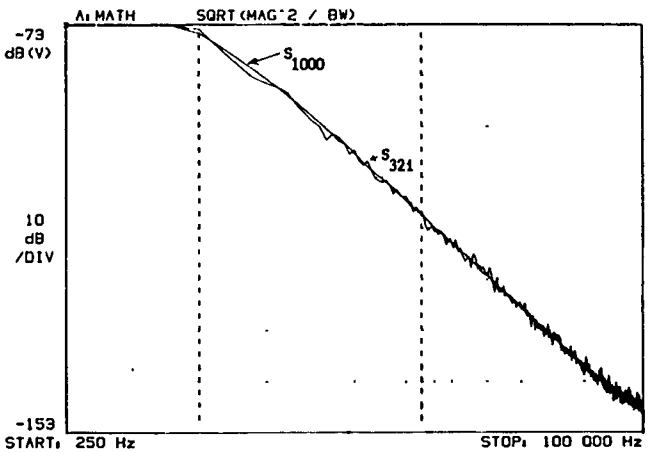


Figure 4. Comparison of the spectral estimate of f^{-4} power-law noise with 1000 samples with that obtained with 32 samples. The text explains how these two curves are used to obtain the fractional RMS confidence of the spectral estimate for 32 samples.

IV.B. Experimental Determination

The following procedure can be used to experimentally determine the variance of the spectral estimates of virtually any type of noise spectrum with any type of window for a particular instrument. Since the spectral density of interest is in general nonwhite, we must determine both the "true value" and a way to normalize the fractional error of the estimate as a function of the number of samples. This can be done by making use of the above theoretical analysis that shows that the variance should decrease as the square root of the number of samples since they are approximately statistically independent (in fact, exactly so in the cases of white and random-walk noise). As an example we have chosen to take $N_b = 1000$ blocks of the various power-law noise types examined in III.B above and compare the value of the spectral estimate with that obtained from $N_b = 32$ blocks (see Figure 4). Since the variance of the 1000 block data is about 32 times smaller than that of the 32 block data, it can serve as an accurate estimate of the "true value." Let $\hat{S}_{1000}(f_j)$ represent this quantity at the j-th channel (frequency). By subtracting the 1000 block data from the 32 block data at the j-th channel, we then have one estimate of the error for the 32 block data; by repeating this procedure over N_c different channels and N_r different replications, we can obtain accurate estimates of the variance for the 32 block data. Let $\hat{S}_{32i}(f_j)$ represent the spectral estimate for the 32 block data at the j-th channel and the i-th replication. To compensate for the variation in the level of the spectral estimates with channel, it is necessary to divide the error at the j-th channel by the "true value" $\hat{S}_{1000}(f_j)$. The mean square fractional error of the 32 block data for the noise type under study is given by

$$\sigma_{32}^2 = \frac{1}{N_r N_c} \sum_{i=1}^{N_r} \sum_{j} \left(\frac{\hat{S}_{32i}(f_j) - \hat{S}_{1000}(f_j)}{\hat{S}_{1000}(f_j)} \right)^2 \approx \frac{var\{\hat{S}_{32i}(f_j)\}}{S^2(f_j)} .$$

It is assumed that all channels with bias — as indicated in Table 1 — have been excluded in the summation over j. It is also important that the changes in the spectral density not exceed the dynamic range of the digitizer because under this condition the quantization errors — in addition to causing biases in the spectral estimates as discussed earlier — can lead to situations where the variance does not improve as N_b increases. These values can be scaled to any number of blocks N_b if care is taken to avoid these quantization errors. Upper and lower approximate 67% confidence limits for $S(f_j)$ — the true spectral density at channel j — using Hanning, uniform and the proprietary "flattened peak" windows for N_b approximately independent blocks are given by

$$\hat{S}(f_j)(1 \pm V(\alpha, N_b))$$

where $\hat{S}(f_j)$ is the spectral estimate given by Equation (1) and $V(\alpha, N_b)$ is the fractional variance given in Table 2 for f^α and $\alpha = 0$, -2, -3 and -4 (these results were obtained by averaging over $N_r N_c = 1200$ channels). The variances obtained are very close to those obtained from standard statistical analysis for white noise, i.e.,

$$\hat{S}(f_j)\left(1 \pm \frac{1}{\sqrt{N_b}}\right) .$$

Table 2. Confidence Intervals for FFT Spectral Estimates

power law noise type	window: uniform	Hanning	flattened peak
f^0	$1.02/\sqrt{N_b}$	$0.98/\sqrt{N_b}$	$0.98/\sqrt{N_b}$
f^{-2}	$1.02/\sqrt{N_b}$	$1.04/\sqrt{N_b}$	$1.04/\sqrt{N_b}$
f^{-3}	unusable	$1.04/\sqrt{N_b}$	$1.04/\sqrt{N_b}$
f^{-4}	unusable	$1.04/\sqrt{N_b}$	$1.04/\sqrt{N_b}$

V. Conclusions

We have introduced experimental techniques to evaluate the statistical properties of FFT spectral estimates for common noise types found in oscillators, amplifiers, mixers and similar

devices, and we have compared these with theoretical calcuations. We have used these techniques to study the biases and variances of FFT spectral estimates using the uniform, Hanning, and a proprietary "flattened peak" window. The theoretical analysis was greatly hampered because the instrument manufacturer does not disclose the exact form of the "flattened peak" window or the normalization procedure for the other windows. Nevertheless, we obtained fair agreement between the theoretical and the experimental analysis. The variances of the spectral estimation were virtually identical to a few percent for f^0 to f^{-4} noise except for the uniform window which is incapable of measuring noise which falls off faster than f^{-2}. There was a very large difference in the biases of the first few channels for the three windows. The Hanning window showed significant biases in the first 3 channels while the proprietary "flattened peak" window showed large biases for f^{-4} noise even up to channel 13. The Hanning window therefore yields useful information over three times wider frequency range than the proprietary "flattened peak" window. In the particular instrument studied, the proprietary "flattened peak" window is the best choice for estimating the height of a narrow band source, while the Hanning window is by far the best choice for spectral analysis of common noise types found in oscillators, amplifiers, mixers, etc. We have also shown that the 67% confidence levels for spectral estimation as a function of the number of contiguous nonoverlapping blocks, N_b, is approximately given by

$$S = S_m \left(1 \pm \frac{0.98}{\sqrt{N_b}}\right)$$

for white noise (f^0) and by

$$S = S_m \left(1 \pm \frac{1.04}{\sqrt{N_b}}\right)$$

for noise types f^{-2} to f^{-4}. This agrees to within 4% of that found by standard statistical analysis for white noise. Using this data one can now determine the number of samples necessary to estimate — to a given level of statistical uncertainty — the spectrum of the various noise types commonly found in oscillators, amplifiers, mixers, etc.

References

[1] D. Babitch and J. Oliverio, "Phase Noise of Various Oscillators at Very Low Fourier Frequencies," in *Proceedings of the 28th Annual Symposium on Frequency Control*, 1974, pp. 150-159.

[2] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, New Jersey: Prentice Hall, 1989, ch. 11, pp. 695-767.

[3] D. B. Percival and F. L. Walls, "Spectral Analysis of Power-Law Processes with a Spectrum Analyzer," manuscript in preparation, 1989.

[4] M. B. Priestley, *Spectral Analysis and Time Series*. London: Academic Press, 1981, ch. 7, pp. 502-612.

Appendix. Precision Noise Source

Figure 5 shows the circuit diagram of a precision noise source whose spectral density can be determined from first principles to ±0.2 dB over the frequency range from 20 Hz to 20 kHz. The spectral density is basically given by the Johnson noise of the 10^5 ohm resistor, $V_n^2 = 4kTR$, where T is in Kelvin, and k is Boltzmann's constant. Corrections due to the input noise voltage and noise current of the amplifier amount to about 0.2 dB for the circuit elements shown. All resistors are precision 1% metal film resistors. The output level can be switched from -100 dBV/Hz to -80 dBV/Hz. By adjusting the noise-gain capacitors one can make the noise spectrum flat to within 0.3 dB out to 200 kHz. There is also provision to measure the input noise voltage of the amplifier by shorting the input to ground or the combined noise voltage and noise current by switching a 220 pF capacitor into the input instead of the 10^5 ohm noise resistor.

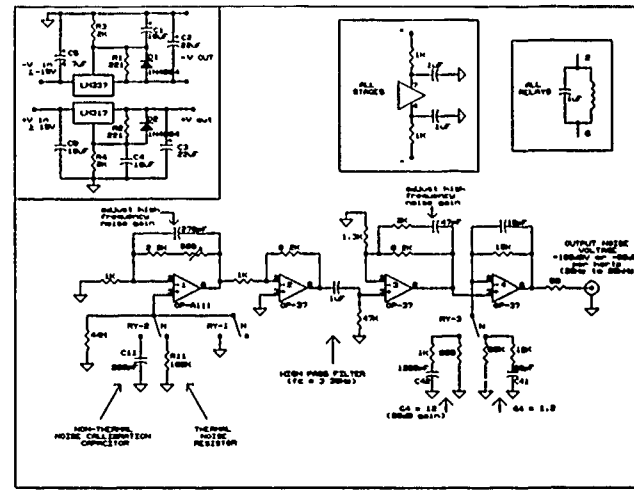


Figure 5. Circuit diagram of a precision noise source.

43rd Annual Symposium on Frequency Control - 1989

PHASE NOISE PERFORMANCE OF ANALOG FREQUENCY DIVIDERS

M. M. DRISCOLL

Westinghouse Electric Corporation
Electronic Systems Group
Baltimore, Maryland 21203

Summary

The phase noise performance obtainable using transistor-transistor logic (TTL) and emitter-coupled logic (ECL), silicon and GaAs-based, digital frequency dividers is available in the literature.

This paper reports on measurement of the spectral performance of two types of analog dividers: a parametric divider using varactor diodes and a regenerative type divider incorporating a double balanced mixer in the oscillator feedback circuit. Both dividers were configured for divide-by-two operation at VHF. The parametric divider typically requires a moderately high input drive level and operates over relatively narrow bandwidth owing to the nature of the circuit constraints for generation of negative resistance at the output frequency = $f_{in/2}$. Prototype dividers were constructed utilizing abrupt junction varactor diodes and were designed without the need for critical tuning, input and output trap circuits normally used. At specified 20 dBm drive, 6 dB conversion loss was obtained. Lowest 1/f noise was obtained with the varactor biased to insure nonconduction over the entire cycle of the RF waveform. Phase noise performance was characterized by £(100 Hz) ~ -148 dB/Hz and a phase noise floor level of -172 dB/Hz. This compares favorably with corresponding values of -148 dB/Hz (and -155 dB/Hz noise floor) obtained using ECL dividers.

Unlike the parametric divider, a regenerative type divider can be operated over a wide range of drive levels and over a much wider bandwidth. This type of divider incorporates a frequency conversion in the oscillator feedback circuit. For divide-by-two operation, the oscillator operates at $f_{out} = f_{in/2}$ and a double balanced mixer externally driven at f_{in} is incorporated in the feedback circuit. The feedback circuit usually contains a low pass filter to prevent unwanted operation at $3f_{in/2}$. Results obtained for prototype divide-by-two circuits using this technique and incorporating low 1/f noise modular amplifiers are characterized by £(100 Hz) ~ -157 dB/Hz and noise floor levels of -174 dB/Hz for dividers operated at +5 dBm 160 MHz input signal drive.

The use of the regenerative type divider appears especially attractive because it provides low 1/f noise and noise floor performance. Because all circuit components except for the filter are ultra-broadband, the nonfilter portion of the circuit could be packaged as a modular type component designed to perform over the entire HF-UHF range via user attachment of specific divider feedback circuit filter.

Introduction

Use of frequency division constitutes a key aspect of synthesized signal generation hardware design. The flexibility offered by programmable, digital logic level dividers has resulted in widespread device use.

Theoretically, carrier signal phase noise sideband level improves as a result of frequency division. In practice, however, actual signal spectral degradation occurs due to the additive noise contribution of the divider. In light of (phase noise level) improvements made with regard to oscillator and frequency multiplier circuitry, divider additive noise often represents the limiting factor with regard to overall signal generator spectral performance and resulting system dynamic range [1-4].

In connection with signal generation hardware employing direct synthesis, critical, low-noise, signal path hardware usually includes fixed-ratio frequency dividers used, for example, in cascaded, iterative mix-and-divide circuitry [3, 4]. This paper describes the results of evaluation of the spectral performance of two types of analog frequency dividers in connection with potential device usefulness in synthesis circuitry. For the purposes of this investigation, divide-by-two operation in the VHF range was evaluated as the simplest form of division.

Parametric Frequency Divider

Figure 1 shows a generalized block diagram for a divide-by-two parametric divider. As shown in the figure, the voltage-variable capacitance (varactor) diode is driven at the input frequency and, as a result of the nonlinear capacitance versus voltage diode characteristic, generates a negative resistance, and thus a signal, at one half the input signal frequency.

Isolation networks (LC tuned circuits) are used to confine input and output signal current flow in the input and output loops of the circuit.

Figure 2 shows a more specific implementation of the circuit. In figure 2, an inductor has been series-connected to the varactor such that the combination exhibits equal magnitude, inductive and capacitive reactance at the input and output frequencies, respectively. This allows the use of simple parallel resonant input and output isolation circuits.

Figure 3 shows the applied RF waveform superimposed on the diode capacitance-voltage characteristic. The diode capacitance is given by:

$$V(C) = C_s \left(\frac{\phi - V_s}{\phi - V}\right)^n \qquad (1)$$

where C_s is the diode capacitance at a specified (Vs) voltage, ϕ is the contact potential, and the exponent $n = 0.5$ for an abrupt junction diode and 0.33 for a graded junction diode. For the abrupt junction diode used, the diode capacitance was specified at 4 volts with $\phi = 0.6$ volts.

For optimum divider performance in terms of negative conductance:

$$V_p = \frac{1}{n+1}\left(V + \phi\right) \qquad (2)$$

with $\phi = 0.6$ volts and $n = 0.5$

CH2690-6/89/0000-342 $1.00 C 1989 IEEE

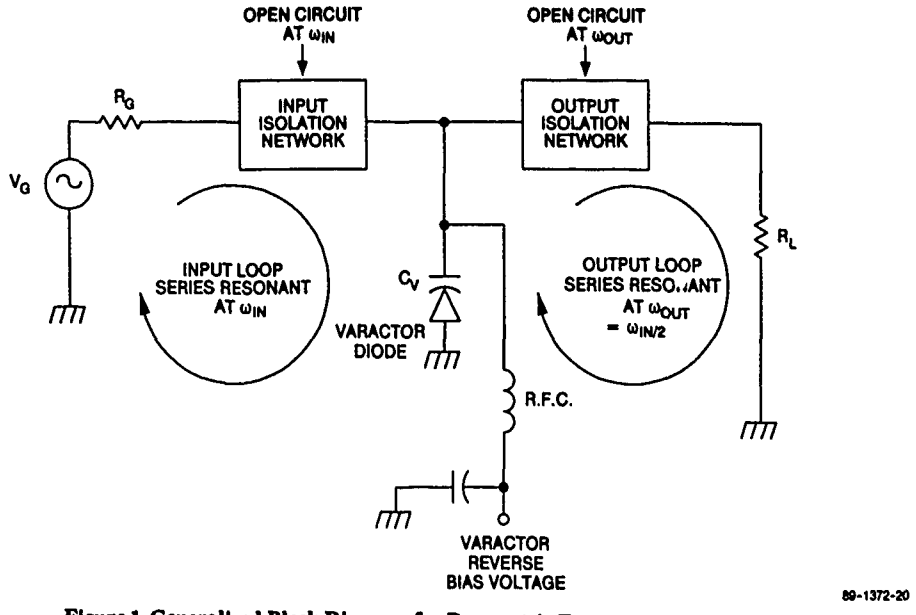


Figure 1. Generalized Block Diagram for Parametric Frequency Divider

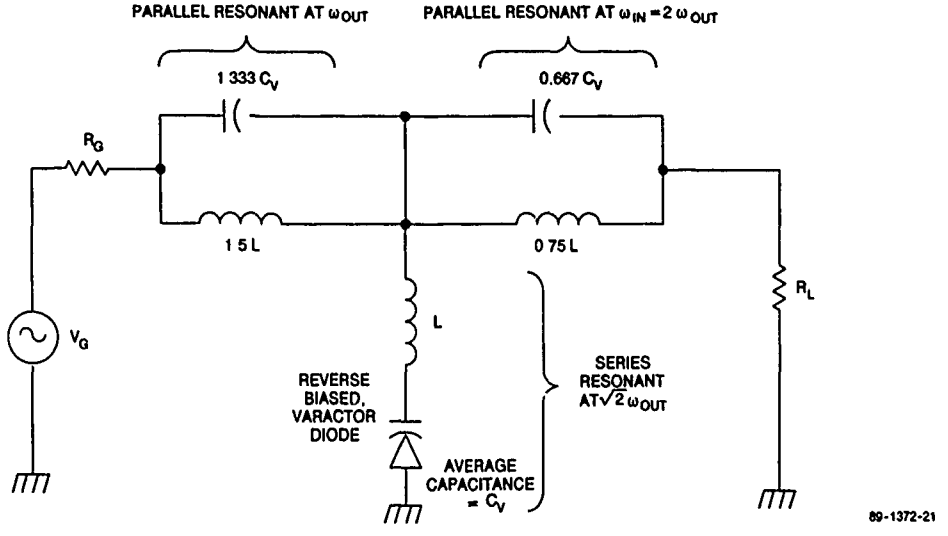


Figure 2. Specific Divide-By-Two Circuit Implementation

$$V_p = \frac{2}{3}\left(V + 0.6\right) \qquad (3)$$

The negative conductance generated at the output frequency, ω_o is:

$$G = \frac{\omega_o}{2}\left(\frac{n}{n+1}\right) Cv \qquad (4)$$

with n = 0.5 this corresponds to a negative resistance equal to:

$$r = \frac{-1}{6\omega_o Cv} \qquad (5)$$

The required input power (to develop V_p across the diode) is:

$$P_{in} = \frac{I_p^2 r_s}{2} = \frac{V_p^2 r_s}{2\omega_{in}^2 Cv^2} \qquad (6)$$

where r_s represents the equivalent input loop series resistance due to losses in the varactor diode and (largely) in the circuit inductors. Equations (1) through (6) were based on parametric divider analysis given in reference [3].

In order to reduce circuit losses and improve divider efficiency, a 40 MHz to 20 MHz divider was designed and fabricated as shown in figure 4. As shown in the figure, the parallel resonant input and output isolation circuits of figure 2 were replaced by a simple series capacitor and inductor. This was done to minimize circuit losses at the expense of slightly reduced input/output isolation. Use of small value reactance shunt inductive input and capacitive output matching, as well as the additional use of an output low pass filter provided more than adequate isolation.

For the circuit of figure 4, 7 volt peak, 40 MHz, input voltage across the (100 pF at 4 volt) varactor provides approximately -20 ohms negative resistance at 20 MHz. Calculated output circuit equivalent series losses total approximately 13 ohms. The two divider were driven at 20

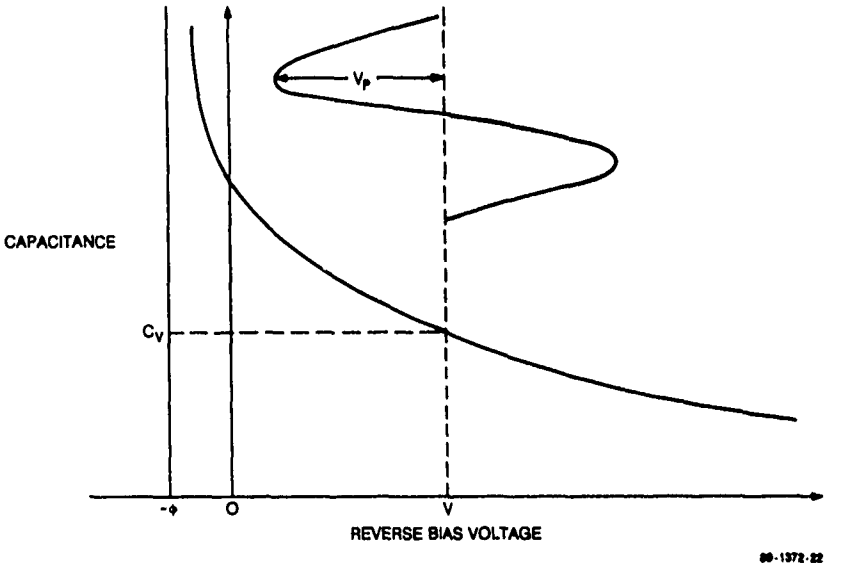


Figure 3. Varactor Diode C-V Characteristic Showing Input Voltage Waveform

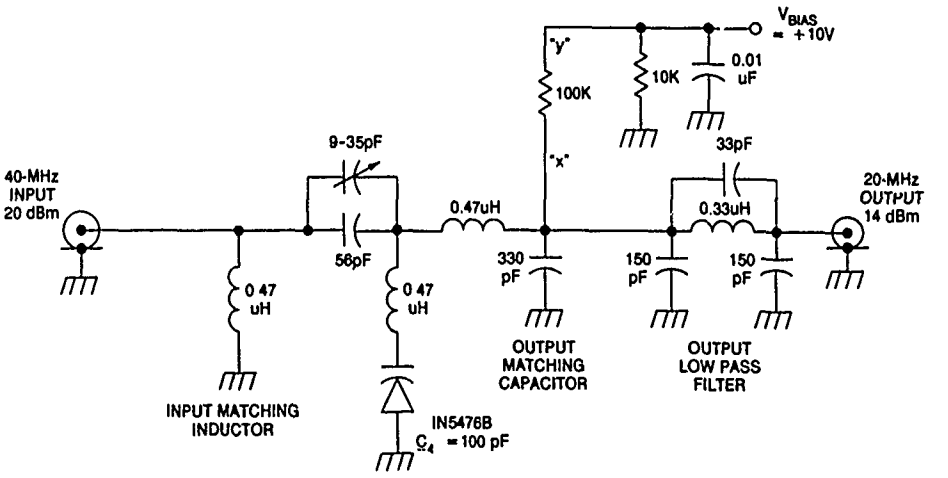


Figure 4. Schematic Diagram for 40-MHz to 20-MHz Divider

dBm input drive and produced 14 dBm output power, consistent with calculated 25 percent efficiency.

Divider phase noise measurements were made using the test setup shown in figure 5. Figure 6 shows the results of measurement of divider phase noise for the pair of figure 4 devices. As shown, the l/f noise (per-device) is characterized by £(100 Hz) = -148 dB/Hz. The per-device noise floor is -172 dB/Hz. The noise floor performance is significantly superior to that measured for digital logic-level dividers. However, requirements for moderately high drive and characteristically small circuit bandwidth limit device usefulness.

As an additional test, the diode bias voltage was decreased to 8 volts and the dividers retuned so as to allow slight diode conduction (I_{xy} = 100 microamps in figure 4). The result shown in figure 7 is a 5 dB increase in l/f noise.

Regenerative Divider

Figure 8 shows a functional block diagram for a regenerative (divide-by-two) frequency divider. As shown in the figure, the circuit operates as a feedback oscillator operating at the output frequency and incorporates a frequency conversion (double-balanced mixer) in the feedback loop.

Loop phase and gain requirements for oscillation are the same as in a conventional feedback oscillator, except that the signal phase shift across the mixer (from point x to y in figure 8) is noncritical in that the oscillating output signal can assume the requisite phase relationship with that of the input signal so that overall $2N\pi$ radians closed loop phase shift occurs. Analysis of regenerative divider signal phase relationships shows that loop amplifier open loop phase noise is reduced by the division factor at the divider output [4].

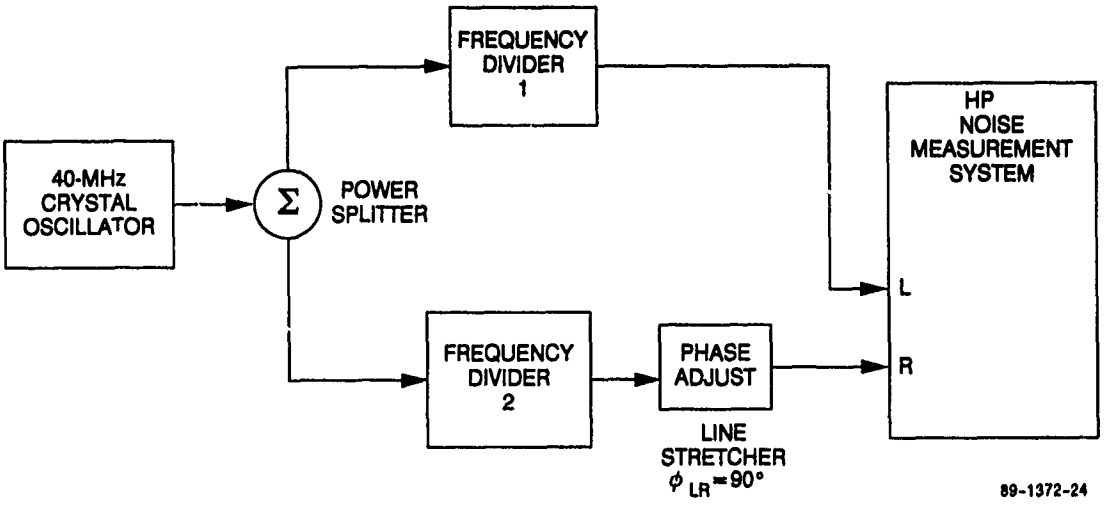


Figure 5. Divider Phase Noise Test Setup

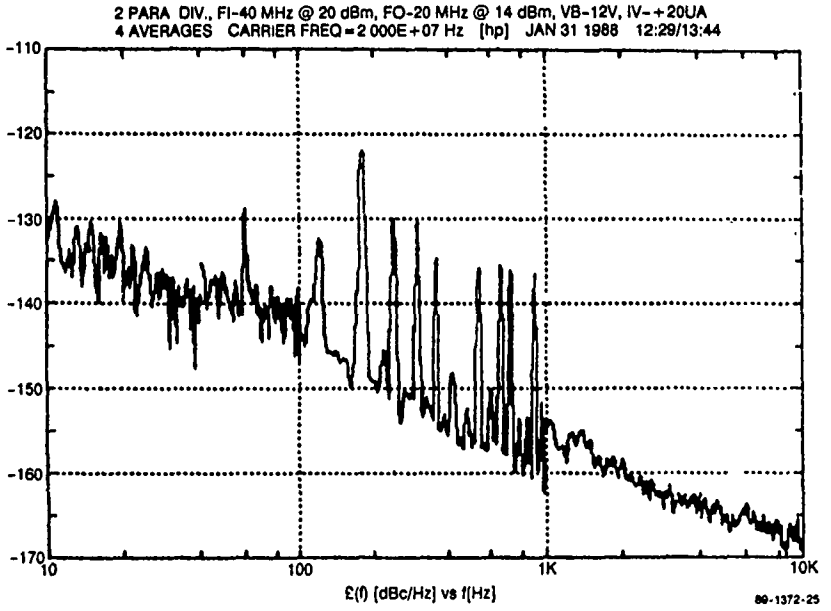


Figure 6. Phase Noise Sideband Performance for a Pair of Parametric Dividers

In addition, if the circuit is designed so that the steady state oscillator feedback signal (mixer) drive level exceeds that of the input signal, the feedback signal level-dependent conversion loss of the mixer provides the required loop AGC function so that the loop amplifier can be operated in its linear operating region.

Figure 9 shows the schematic diagram for prototype regenerative dividers designed for operation at 160 MHz. Input drive level to the mixer was selected to yield steady-state amplifier operation 3 to 4 dB below the device 1 dB compression point.

Figure 10 shows the results of measurement of additive divider phase noise and instrumentation noise made using the figure 5 test setup. As shown in the figure, per-device 1/f phase noise performance is chara terized by £(100Hz) = 157 dB/Hz and is consistent with measured values of open loop 1/f noise for the RF amplifier and double balanced mixer. Per-device noise floor level is on the order of -174 dB/Hz and is primarily a function of -2 dBm amplifier input drive level, 4dB noise figure, and 2:1 (-6 dB) division ratio.

Figure 11 shows the results of measurement of divider input and output signal phase noise plotted as $\sqrt{Sy(f)}$ for the pair of figure 9 dividers driven from two phase-locked, 160-MHz crystal oscillator-multipliers. As shown in the figure, the curves virtually overlap, indicating no special degradation results from divider use.

Figure 12 shows measured divider bandwidth performance. As shown, satisfactory divider operation was achieved over a 100-MHz to 260-MHz input frequency band. No startup problems of any kind were noted. Above 260 MHz, oscillation could not be sustained due to low pass filter stopband loss, and device drive at 80 MHz produced simultaneous 20 and 40 MHz outputs.

Considering the current availability of low 1/f noise ultra-broadband double balanced mixers and modular amplifiers, use of this type of divide-by-two circuit can provide phase noise performance far superior to ECL-level digital dividers over bandwidths in excess of an oc-

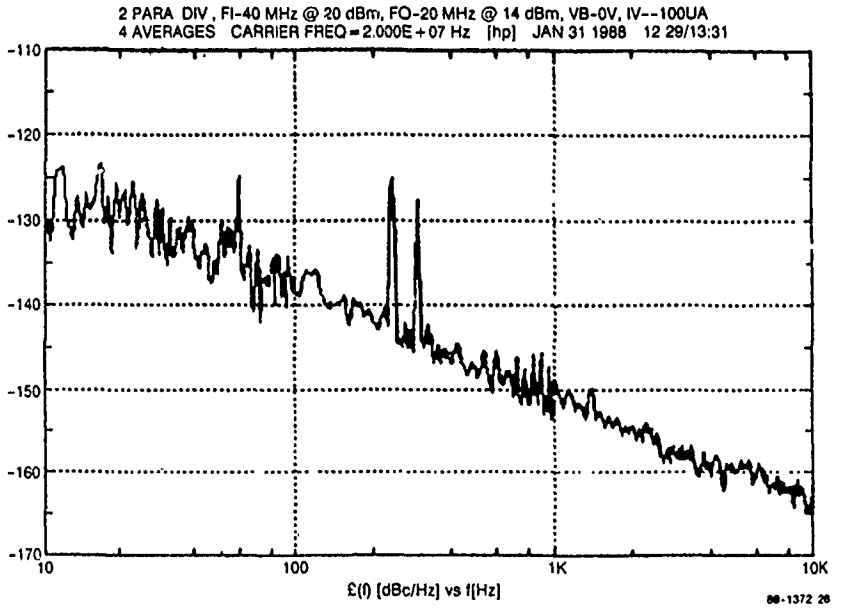


Figure 7. Parametric Divider 1/f Noise Degradation Observed at Onset of Varactor Diode Conduction

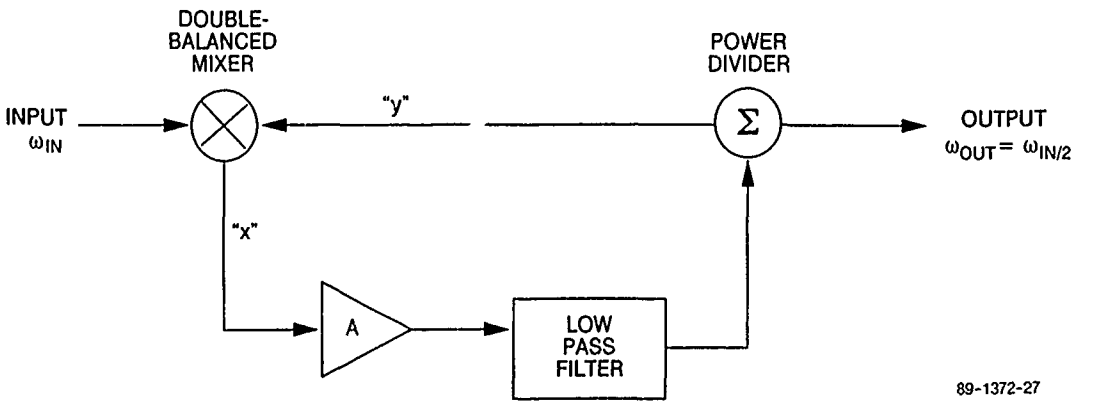


Figure 8. Regenerative Frequency Divider Block Diagram

tave. In addition, the entire divider circuit could easily be packaged as a single modular component with external user connection of the loop low pass filter bandwidth - tailored to the specific application.

Conclusions

Evaluation of the phase noise performance of two alternative types of analog frequency dividers indicates that the regenerative divider is capable of providing much lower phase noise than conventional digital logic level devices. In addition, the regenerative divider can be successfully operated over bandwidths in excess of an octave, and the design lends itself to small (i.e., TO-8) modular package implementation. Operating frequencies are bounded only by the range of the mixer and RF amplifier utilized and, as such, should extend from HF through microwave.

References

[1] F. W. Walls, "Low Noise Frequency Synthesis," Proc. 41st Freq. Cont. Symp., May 1987, pp. 512-518.

[2] B. Flugstad, Hewlett-Packard Co., Private Communication, Spokane Division, 1620 Signal Drive, Spokane, Washington, 99220, 1982.

[3] V. Manassewitsch, *Frequency Synthesizers, Theory and Design*, John Wiley and Sons, New York, 1976.

[4] V. Kroupa, *Frequency Synthesis*, John Wiley and Sons, New York, 1973.

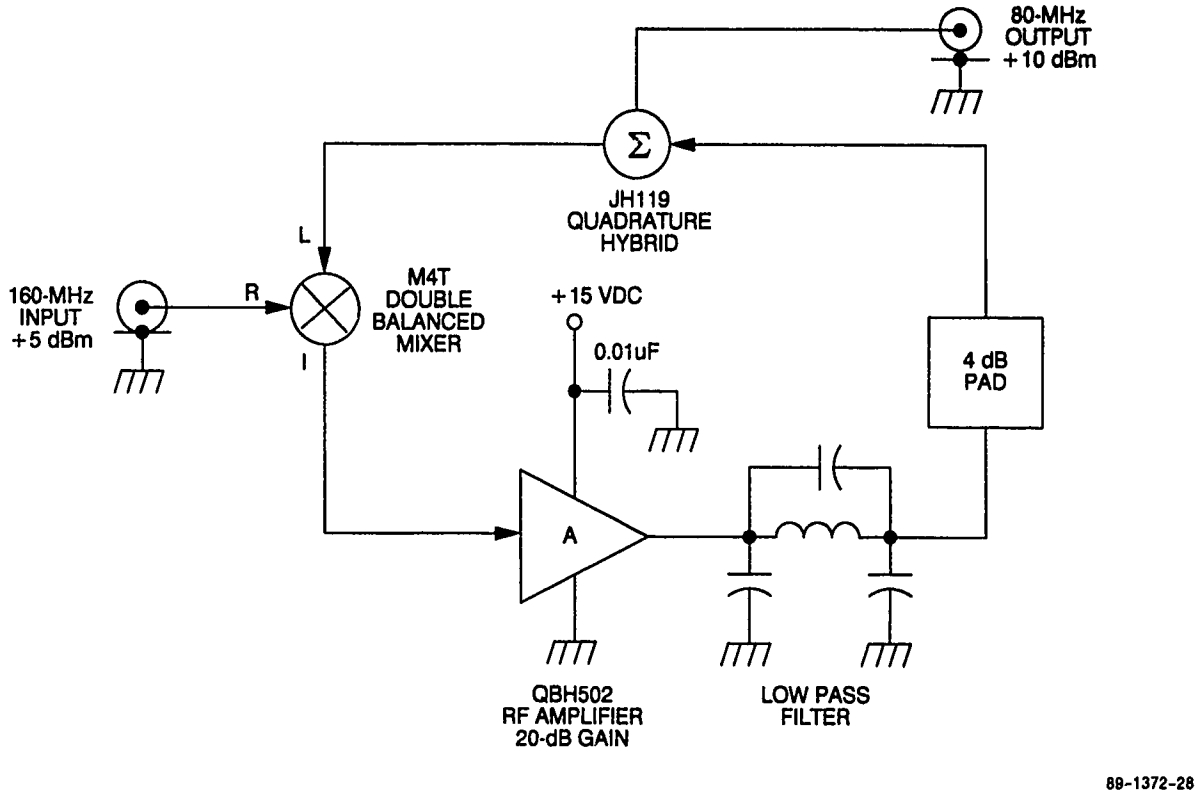


Figure 9. Schematic Diagram for 160-MHz to 80-MHz Regenerative Divider

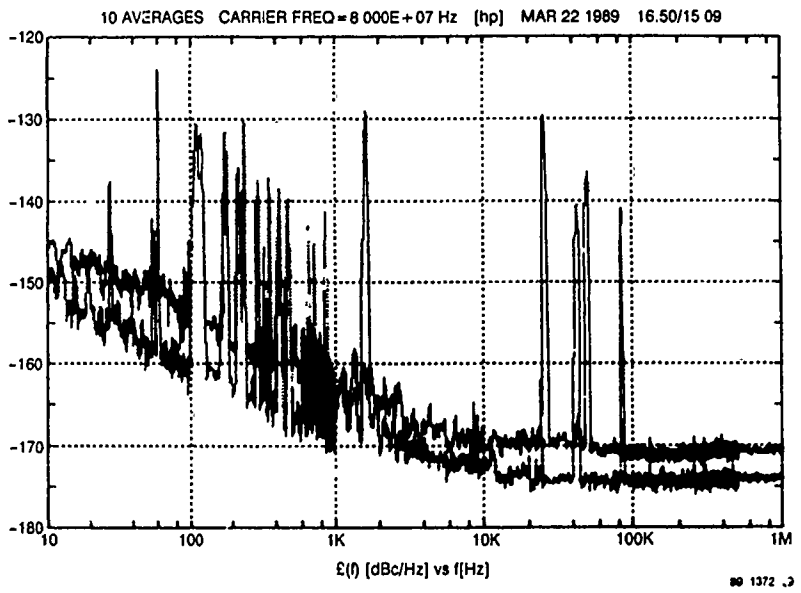


Figure 10. Regenerative Divider Phase Nose Spectra
Top Curve: Measured Performance for Two Devices
Bottom Curve: Measured Instrumentation Noise Level

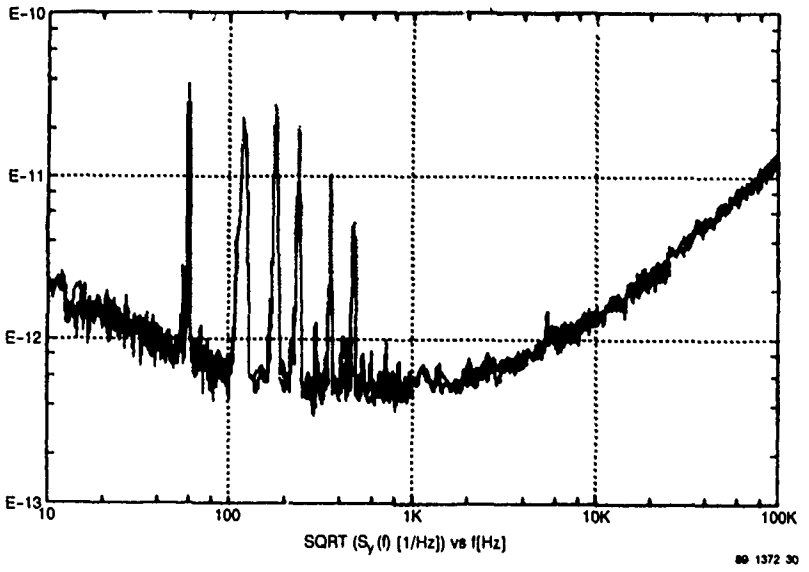


Figure 11. Noise Measurement Results Showing Identical Divider Input and Output Signal Fractional Frequency Spectra

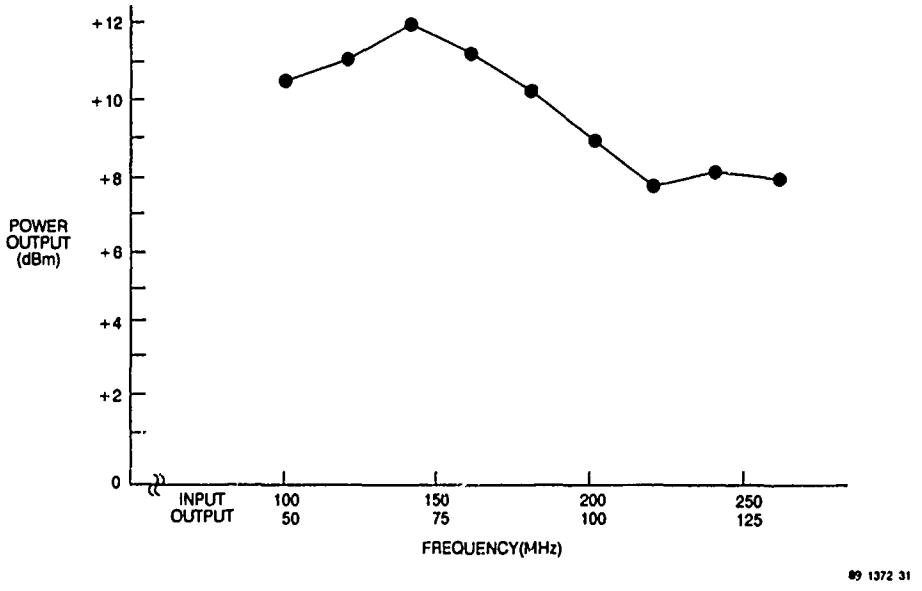


Figure 12. Divider Power Output Versus Frequency

43rd Annual Symposium on Frequency Control - 1989

RESIDUAL PHASE NOISE MEASUREMENTS OF VHF, UHF, AND MICROWAVE COMPONENTS

G. K. Montress, T. E. Parker, and M. J. Loboda
Raytheon Research Division
131 Spring Street
Lexington, Massachusetts 02173

ABSTRACT

The results of residual phase noise measurements on a number of VHF, UHF, and microwave amplifiers, both silicon bipolar junction transistor (BJT) and GaAs field effect transistor (FET) based, electronic phase shifters, frequency dividers and multipliers, etc., which are commonly used in a wide variety of frequency source and synthesizer applications are presented. The measurement technique has also been used to evaluate feedback oscillator components, such as the loop and buffer amplifiers, which can play important roles in determining an oscillator's phase noise spectrum (often in very subtle ways). While some information has previously been published related to component residual phase noise properties, it generally focused on the flicker noise levels of the devices under test, for carrier offset frequencies less than 10 kHz. The work reported herein makes use of an extremely low noise, 500 MHz SAW resonator oscillator (SAWRO) test source for residual phase noise measurements, both close-to- and far-from-the-carrier. Using this SAWRO test source at 500 MHz, we have been able to routinely achieve a measurement system phase noise floor of -184 dBc/Hz, or better, for carrier offset frequencies greater than 10 kHz, and a system flicker phase noise floor of -150 dBc/Hz, or better, at 1 Hz carrier offset. The paper discusses the results of detailed residual phase noise measurements performed on components using this overall system configuration. Several interesting observations related to the residual phase noise properties of moderate to high power RF amplifiers, i.e., amplifiers with 1 dB gain compression points in the range of +20 to +33 dBm, are highlighted.

I. INTRODUCTION

Residual phase noise measurements have been used for some time to characterize devices in the transmission [1] - [4], as well as reflection [5] modes. The generation of low phase noise VHF, UHF, and microwave signals is critically dependent upon the use of components, e.g., amplifiers, electronic phase shifters, frequency dividers and multipliers, mixers, etc., which possess verifiably low residual phase noise levels. While some information has previously been published related to component residual phase noise properties [1] - [4], [6] - [8], it generally focused on the flicker noise levels of the devices under test, for carrier offset frequencies less than 10 kHz. The work reported herein makes use of an extremely low noise, 500 MHz SAWRO test source for residual phase noise measurements both close-to- and far-from-the-carrier. This 500 MHz SAWRO has previously been described in the literature [10] - [12].

The 500 MHz SAWRO test source has been used to achieve a measurement system phase noise floor of -184 dBc/Hz, or better, for carrier offset frequencies greater than 10 kHz, and a system flicker phase noise floor of -150 dBc/Hz, or better, at 1 Hz carrier offset frequency, when used in conjunction with a Hewlett-Packard 3048A Phase Noise Measurement System. The carrier offset frequency region beyond 1 kHz, as well as the flicker noise region, are both of interest when evaluating the residual phase noise properties of components, e.g., amplifiers, for use in extremely low phase noise oscillators. In fact, much of the work concerning amplifier residual phase noise measurements which is reported herein evolved out of our earlier efforts to realize an extremely low phase noise, 500 MHz SAWRO [10] - [12]. The first two prototype low noise SAWROs were subsequently used to develop residual phase noise screening procedures, particularly for amplifiers, in order to insure the expected oscillator phase noise performance of pre-production engineering units.

Detailed residual phase noise measurement results on a wide variety of RF amplifiers, both silicon BJT and GaAs FET based, electronic phase shifters, frequency dividers, and frequency multipliers are discussed. Measurements of this type are invaluable in assessing the potential utility of a particular component in any application which may be sensitive to that component's residual phase noise levels.

II. MEASUREMENT PROCEDURES

The basic residual phase noise measurement set-up shown in Fig. 1 was used [1], [3]. The preferred test source for the measurements is an extremely low noise, 500 MHz SAWRO [10] - [12]. Also shown in Fig. 1, for the purpose of comparison, is a test source configuration based upon the Hewlett-Packard 8662A Frequency Synthesizer. The measurement system's phase noise floor is readily established by replacing the device under test (DUT) with a feedthrough and adjusting the total attenuation in the device test path to maintain the proper RF signal power level to the R-port of the phase detector, in this case +22.5 dBm. Phase quadrature for the phase noise floor measurement is established using a mechanical phase shifter, as indicated in Fig. 1. Figure 2 illustrates the resulting phase noise floor when either the SAWRO based test source or the frequency synthesizer based test source is used. Also shown in Fig. 2 are two solid straight lines which represent an approximate fit to the measured system phase noise floor when using the SAWRO based test source. It is evident that for offset frequencies which are close-to-the-carrier (i.e., less than 1 kHz) there is essentially no discernible difference between the measured system phase noise floors using either test source. However, for carrier offset frequencies greater than 10 kHz there is a noticeable difference between the two test sources, especially in the region between 100 kHz and 10 MHz. This difference between the phase noise floors in the 100 kHz to 10 MHz range is actually attributable to the significantly higher AM noise level of the frequency synthesizer when compared to the SAWRO based test source. This is clearly illustrated in Fig. 3, which shows measured AM noise spectra for both the SAWRO and frequency synthesizer based test sources. The measurements were performed on the test sources exactly as they are shown configured in Fig. 1. A double-balanced mixer was used as the AM detector for these measurements. The two solid straight lines shown in Fig. 3 represent an approximate fit to the AM detector's threshold sensitivity [13], [14]. It is clear that the overall residual phase noise measurement system's AM noise suppression is approximately 30 dB, leading to distortions in the phase noise floor of Fig. 2 when the frequency synthesizer based test source is used. No such distortion in the phase noise floor is evident for the SAWRO based test source. These results suggest that in order to characterize

CH2690-6/89/0000-349 $1.00 C 1989 IEEE

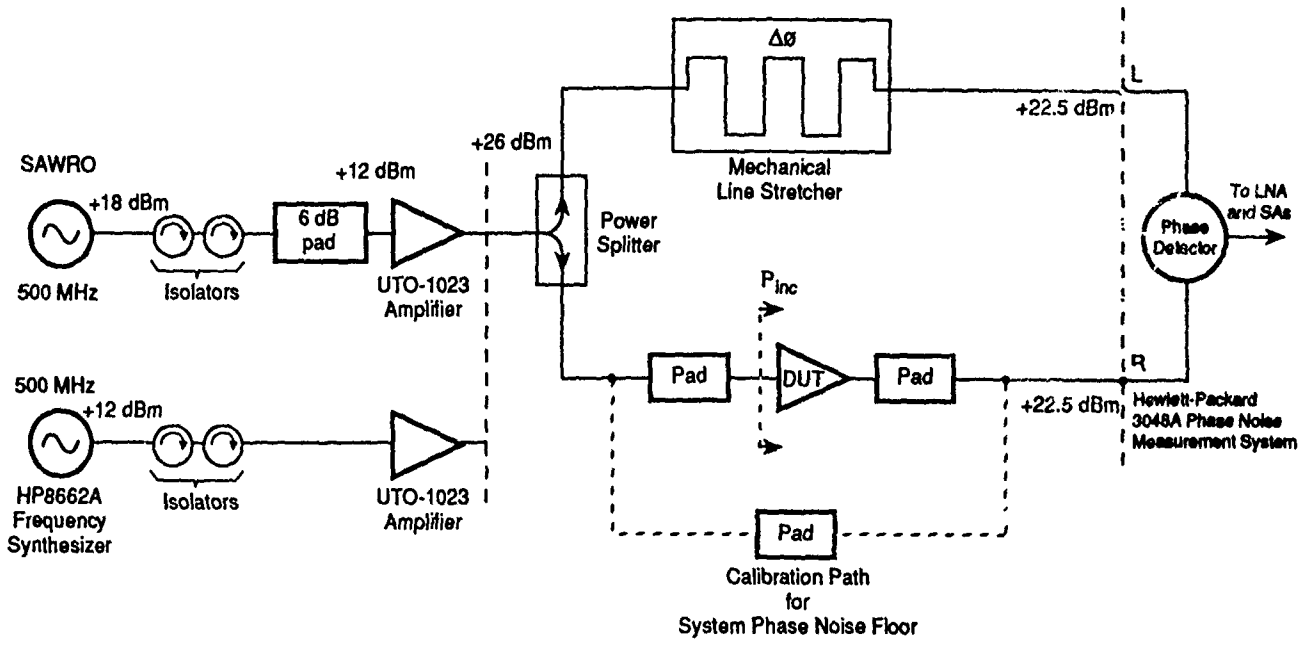


Figure 1. Residual Phase Noise Measurement Set-up at 500 MHz.

PBN-89-1232

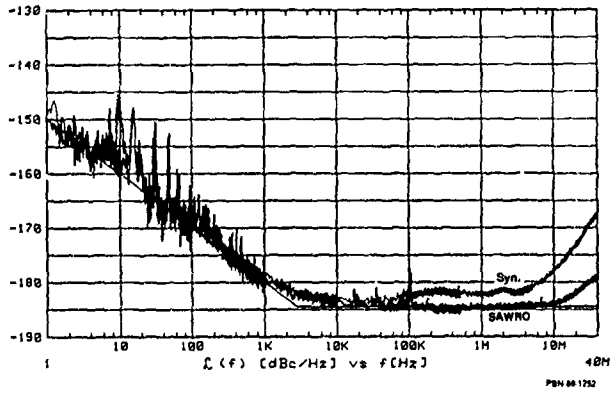


Figure 2. Measurement System Phase Noise Floors at 500 MHz for SAWRO and Frequency Synthesizer Based Test Sources. The Two Solid Straight Lines Represent an Approximate Fit to the Measured Phase Noise Floor When Using the SAWRO Test Source.

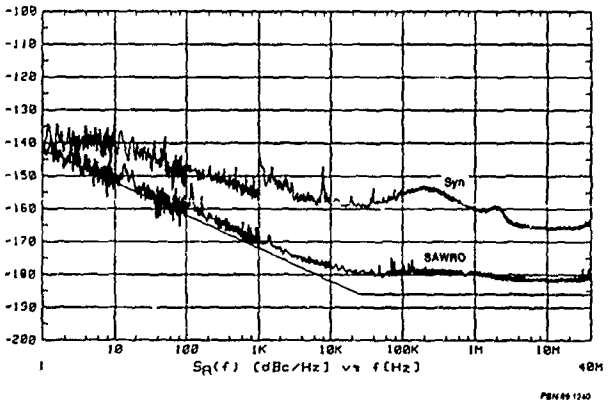


Figure 3. Measured AM Noise at 500 MHz for the SAWRO and Frequency Synthesizer Based Test Sources.

the residual phase noise properties of a DUT for carrier offset frequencies greater than 10 kHz, the test source must possess reasonably low AM noise levels. Certainly, the AM noise of the test source, for carrier offset frequencies beyond 10 kHz, should not be any greater than approximately -160 dBc/Hz. In this case the test source's AM noise will be kept to at least the -185 dBc/Hz level or lower during a residual phase noise measurement, due to the measurement configuration's overall 30 dB suppression of the test source's AM noise.

The upward curvature to the two phase noise floor measurements shown in Fig. 2, for carrier offset frequencies greater than 10 MHz, is attributable to decorrelation of the test source's phase noise due to a slight mismatch in group delay between the two test paths shown in Fig. 1 [15]. A small delay mismatch between the two test paths is unavoidable, since the signals which are input to the phase detector must be maintained in phase quadrature. The phase noise floor measurements in Fig. 2 indicate that the frequency synthesizer based test source's phase noise spectrum is substantially higher than the phase noise spectrum of the SAWRO based test source for a carrier offset frequency of approximately 40 MHz. This is indeed confirmed by the phase noise measurements shown in Fig. 4 for each test source configuration. Only for carrier offset frequencies less than about 400 Hz is the frequency synthesizer based test source's phase noise superior to the SAWRO based test source.

While the phase noise of either test source is not really a limiting factor for residual phase noise measurements on a DUT, at least for reasonably small mismatches in delay between the two test paths, the AM noise properties of the test source are actually much more important than thus far indicated. This is illustrated in Fig. 5, which shows the results of three separate amplifier residual phase noise measurements, specifically: 1) the frequency synthesizer based source used to test a UTO-1023 (Avantek) RF amplifier at ~ 3 dB of gain compression (P_{inc}=+17.0 dBm), 2) the frequency synthesizer based source used to test a UTO-509 (Avantek) RF amplifier at ~ 3 dB of gain compression (P_{inc}=+12.0 dBm), and 3) the SAWRO based source used to test the same UTO-509 (Avantek) RF amplifier under

identical test conditions. It is readily apparent that the DUT is converting the frequency synthesizer based test source's AM noise into PM noise, thereby limiting the measurement configuration's ability to characterize the amplifier's actual residual phase noise. In point of fact, the UTO-1023 is decreasing the effective AM source noise suppression for the measurement to only 2 to 4 dB! Clearly, this fact serves to place a much more stringent requirement upon the test source's AM noise spectrum, if it is to be used for accurate residual phase noise measurements on DUTs with appreciable AM-to-PM conversion factors. Since the UTO-509 only degrades the effective AM source noise suppression for the measurement to approximately 10 dB, the SAWRO based test source may be used to accurately characterize the UTO-509's residual phase noise properties at all carrier offset frequencies. However, note that using the frequency synthesizer based source to test a UTO-509 still presents a problem not only far-from-the-carrier, but also close-to-the-carrier as well. The implications of the results shown in Fig. 5 are that not only must the test source possess very low AM noise levels in order to accurately characterize an amplifier's residual phase noise properties, but an independent residual phase noise measurement with the frequency synthesizer based test source should also be performed. Using a test source with an accurately characterized, relatively high AM noise level provides useful information with regard to the AM-to-PM conversion factor of the DUT, as well as an indication of whether the measured residual flicker phase noise levels close-to-the-carrier are still influenced by the AM noise properties of the test source which is used.

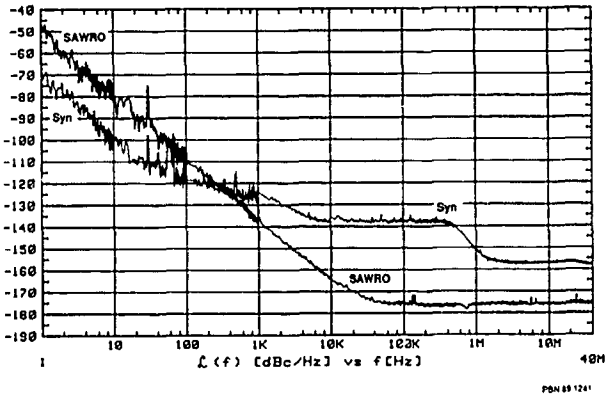


Figure 4. Measured PM Noise at 500 MHz for the SAWRO and Frequency Synthesizer Based Test Sources.

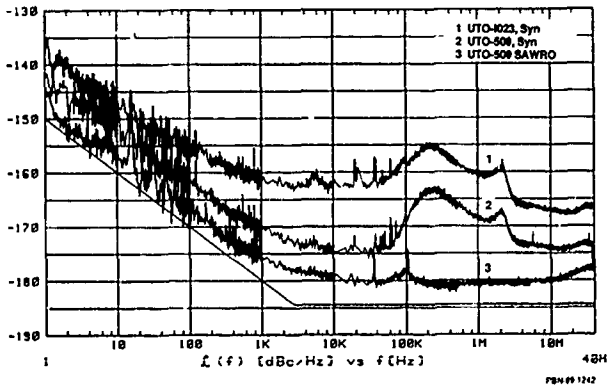


Figure 5. A Comparison of Amplifier Residual Phase Noise Measurements Performed Using the SAWRO and Frequency Synthesizer Based Test Sources. Curve 1: UTO-1023 (Avantek), P_{inc}=+17.0 dBm, with Synthesizer Based Test Source. Curve 2: UTO-509 (Avantek), P_{inc}=+12.0 dBm, with Synthesizer Based Test Source. Curve 3: UTO-509 (Avantek), P_{inc}=+12.0 dBm, with SAWRO Based Test Source.

The frequency synthesizer based test source remains very useful for the residual phase noise characterization of devices with low AM-to-PM conversion factors, such as SAW [15] and BAW [16] resonators and SAW delay lines [17], at least for carrier offset frequencies less than 100 Hz. However, its utility is very limited for the characterization of broadband, low noise components such as moderate to high power RF amplifiers. This caveat will be the case whenever the device under test exhibits an appreciable AM-to-PM conversion factor. While the foregoing observations are based upon using the specific test sources just described, in general the PM ***and*** AM noise properties of the test source used for residual phase noise measurements should be well documented. This will insure that any resulting residual phase noise measurements may be properly interpreted. The following section presents detailed residual phase noise measurements for a variety of VHF, UHF, and microwave components. These results are based upon the basic residual phase noise measurement technique just described, or else slight variants thereof, as when measuring frequency translation type devices such as dividers and multipliers.

III. EXPERIMENTAL RESULTS

A. Silicon Bipolar RF Amplifiers.

Figure 6 illustrates the application of the residual phase noise measurement technique to a moderately high power, TO-8 packaged RF amplifier (~ +23 dBm output power at 3 dB of gain compression), specifically the UTO-509 (Avantek). Generally, for this particular amplifier, we have found that its residual flicker phase noise is usually very close to the measurement system's phase noise floor. Our experience would suggest that this is an exceptionally good result for a UHF amplifier whose power output under 3 dB of gain compression is in excess of +20 dBm. It is interesting to note that several amplifiers of this type (approximately 6 out of more than 50 tested) exhibited an unusual "bump" in the residual phase noise measurement at approximately 80 kHz carrier offset frequency. Figure 6, which shows the residual phase noise of a "normal" UTO-509, also illustrates the residual phase noise characteristic for a different UTO-509, specifically one showing the residual phase noise "bump" at approximately 80 kHz carrier offset frequency. The "bump" was found to be present even for non-gain compressed operation, although the height of the bump did seem to vary somewhat (~ 1 to 2 dB, at most) with the RF input power level. Two feedback loop oscillators, which were assembled using these UTO-509 amplifiers in the loop, exhibited essentially identical "bumps" in their phase noise spectra. While the phase noise "bump"

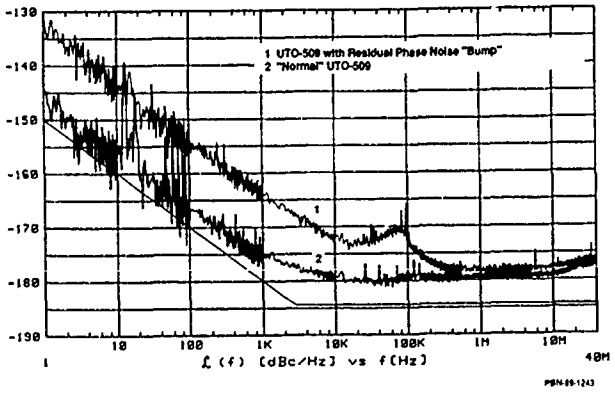


Figure 6. UTO-509 Amplifier Exhibiting a "Bump" in its Residual Phase Noise Characteristic. Curve 1: UTO-509 Amplifier with "Bump". Curve 2: "Normal" UTO-509 Amplifier. P_{inc}=+12.0 dBm (~ 3 dB of Gain Compression) for Both Measurements.

is apparently quite real, at the present time we cannot offer an explanation for its presence in the amplifier's residual phase noise spectrum, although it may be attributable to the "active" DC bias circuit which is used in the amplifier's design. It is, however, a potentially serious problem if this particular amplifier were used in an oscillator application where the phase noise spectrum is to be as low as possible. For example, in a Doppler radar certain targets would be more difficult to detect due to the presence of such a "bump" in the phase noise spectrum of the radar's transmitted signal.

It is interesting to note that neither the amplifier's residual flicker phase noise level nor the possible presence of "bumps" in the residual phase noise spectrum are predictable based upon any other electrical characteristic that we are aware of, including a small-signal noise figure measurement. Furthermore, even very similar amplifier designs from different vendors can differ markedly in their residual phase noise characteristics, even though their other nominal electrical properties are nearly identical. This is illustrated in Fig. 7. Figure 7 simply reproduces the residual phase noise characteristic for the "normal" UTO-509 (Avantek), while also showing the results of measurements performed on a PA3-1 (Watkins-Johnson) and a AC509 (Cougar) amplifier, respectively. In all three cases the incident RF test power at the amplifier's input was +12.0 dBm. These three amplifiers are all remarkably similar in their nominal electrical performance parameters, which are summarized in Table I, along with the results of our own electrical measurements on the devices. The UTO-509's performance shown in Fig. 7 is typical of that observed on more than thirty devices, while the specific PA3-1 and AC509 data shown in Fig. 7 was essentially the same for samples of three and four devices, respectively. It is interesting to note that both the PA3-1 and the AC509 have consistently higher levels of residual flicker phase noise, when compared to the UTO-509. Also, all four AC509s exhibited an anomalous residual phase noise "bump" near 6 kHz carrier offset frequency. Finally, the higher residual white phase noise floor measured for the PA3-1 is consistent with the fact that its noise figure is approximately 2 dB higher than that of the UTO-509. However, the comparatively high phase noise floor of the AC509 cannot be fully explained in a similar way. Separate residual phase noise measurements on each amplifier using the frequency synthesizer based test source, which are shown in Fig. 8, confirm that the AC509 has a considerably higher AM-to-PM conversion factor. This suggests that even with the 500 MHz SAWRO based test source, it simply isn't possible to accurately measure this particular amplifier's residual white phase noise floor properties, at least when only a single device is tested. Testing two amplifiers at a time (one in each test leg shown in Fig. 1) would improve the situation for DUTs with large AM-to-PM conversion factors. This is because the resulting PM in each test path due to the test source's AM noise will be correlated and therefore largely cancel, similar to the measurement configuration's suppression of the test source's PM noise. A disadvantage of this approach is that three DUTs would have to be tested (#1 vs. #2, #1 vs. #3, and #2 vs. #3) in order to extract information about each individual amplifier's residual phase noise properties.

Table I

Summary of Typical RF Amplifier Characteristics at T = 25°C

Amplifier	Small Signal Gain	Noise Figure	Pout @ 1 dB Compression	Pout @ 3 dB Compression	I_{DC} @ V_{DC}	Residual Phase Noise[2] @ 1 kHz
Avantek/UTO-509						
Vendor Typical	14.3 dB	4.5 dB	+22.3 dBm	N.A.	90 mA @ +15V	N.A.
Measured	(14.2 dB)[1]	(4.2 dB)[1]	(+22.2 dBm)[1]	(+23.3 dBm)[1]	(100 mA @ +15V)	(~ -148 dBc/Hz)[1]
Watkins-Johnson/PA3-1						
Vendor Typical	14.5 dB	7.0 dB	+22.5 dBm	N.A.	91 mA @ +15V	N.A.
Measured	(14.2 dB)[1]	(6.8 dB)[1]	(+22.6 dBm)[1]	(+23.6 dBm)[1]	(92 mA @ +15V)	(~ -136 dBc/Hz)[1]
Cougar/AC509						
Vendor Typical	13.5 dB	5.0 dB	+22.0 dBm	N.A.	88 mA @ +15V	N.A.
Measured	(14.0 dB)[1]	(5.3 dB)[1]	(+22.0 dBm)[1]	(+23.1 dBm)[1]	(87 mA @ +15V)	(~ -136 dBc/Hz)[1]

[1] Measured at 500 MHz [2] Measured at 3 dB of Gain Compression N.A. = Not Applicable

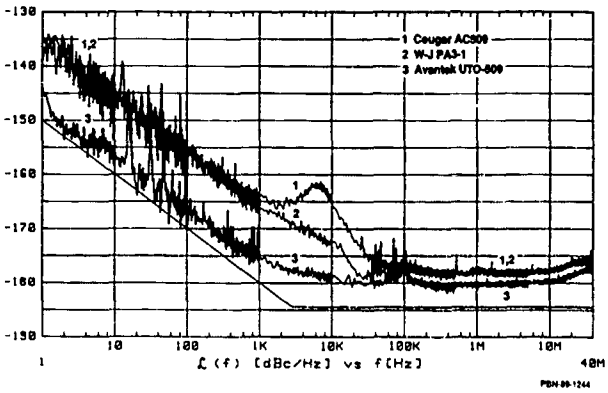


Figure 7. Residual Phase Noise Measured on Three Similar Amplifiers Using SAWRO Based Test Source. Curve 1: Cougar AC509. Curve 2: Watkins-Johnson PA3-1. Curve 3: Avantek UTO-509. P_{inc} = +12.0 dBm (~ 3 dB of Gain Compression) for All Three Measurements.

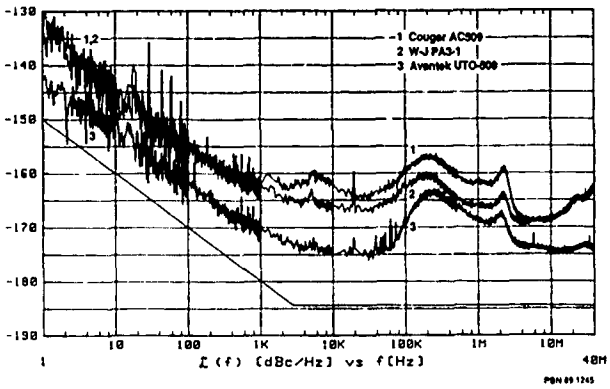


Figure 8. Residual Phase Noise Measured on Three Similar Amplifiers Using Frequency Synthesizer Based Test Source. Curve 1: Cougar AC509. Curve 2: Watkins-Johnson PA3-1. Curve 3: Avantek UTO-509. P_{inc} = +12.0 dBm (~ 3 dB of Gain Compression) for All Three Measurements.

The residual phase noise measurement technique has also revealed that even though an amplifier may have a noise figure (small-signal) which is consistent with the manufacturer's specifications, there may be "anomalous" or excess residual white phase noise present within the device over a relatively wide range of carrier offset frequencies. This is illustrated in Fig. 9 for the UTO-546 (Avantek) TO-8 packaged RF amplifier, for three different incident RF power levels, namely: +4.5 dBm (small-signal, no gain compression), +16.0 dBm (~ 3 dB of gain compression), and +19.5 dBm (~ 5 dB of gain compression). Note the presence of a noise floor "plateau" for carrier offset frequencies between 10 kHz and 1 MHz. The residual phase noise plateau was observed on all seven amplifiers which were tested. The presence of the plateau isn't strongly dependent upon the RF input power to the amplifier. While the exact shape of the residual phase noise plateau is somewhat dependent upon the extent of gain compression in the amplifier, it is still clearly visible even under small-signal operating conditions. Note that the increase in the amplifier's residual white phase noise floor when P_{inc} = +4.5 dBm is due solely to the lower RF input power to the amplifier. The fact that there is essentially no change in the amplifier's residual white phase noise floor for carrier offsets near 10 MHz, when P_{inc} equals either +16.0 or +19.5 dBm, is an indication that once again the amplifier's AM-to-PM conversion factor is coming into play. It is interesting to note that the level of the plateau actually decreases somewhat as the amplifier is driven more strongly into gain compression. This

observation is in distinct contrast to the amplifier's residual flicker phase noise level which increases rather abruptly when it goes much beyond 3 dB of gain compression. Generally, we have found that taking an amplifier beyond 3 dB of gain compression will adversely affect its residual flicker phase noise level. The phase noise plateau was also observed on several feedback loop oscillators which were evaluated with these UTO-546s in the feedback loop [11]. This result suggests that a noise figure measurement with a conventional noise figure meter (small-signal) may not represent the true residual white phase noise floor in a high power amplifier which might be considered for use in a low phase noise oscillator application. Similar behavior has also been noted on a number of amplifier types from several other vendors, including GaAs FET amplifiers, where it is actually quite common to find "excess" residual phase noise for carrier offset frequencies greater than 1 kHz.

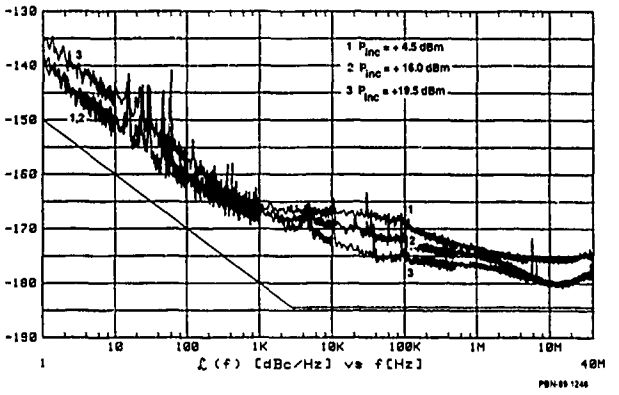


Figure 9. UTO-546 (Avantek) Amplifier Exhibiting a Residual Phase Noise "Plateau". Curve 1: P_{inc}=+4.5 dBm (Small-signal, no Gain Compression). Curve 2: P_{inc}=+16.0 dBm (~ 3 dB of Gain Compression). Curve 3: P_{inc}=+19.5 dBm (~ 5 dB of Gain Compression).

Figure 10 shows residual phase noise measurements for an Avantek UTO-1023 amplifier, using both the SAWRO and frequency synthesizer based test sources. It should be noted here that this particular "UTO-1023" is one of four from a "special" lot which was subjected to screening of the RF transistors' DC βs prior to assembly. In this way several wafers were identified which yielded transistors with particularly low residual flicker phase noise levels [18], [19]. All four "special" UTO-1023s had 1 Hz intercepts for their residual flicker phase noise of -143 dBc/Hz, or better, based upon not only residual phase noise but also oscillator phase noise measurements as well. The AM-to-PM conversion factor for this particular amplifier is such that the effective AM source noise suppression for the residual phase noise measurement is only 2 to 4 dB, at best! Once again, for this particular amplifier even when using the SAWRO based test source, we cannot accurately determine the amplifier's residual white phase noise floor, which is estimated to be approximately -185 dBc/Hz, based upon separate electrical measurements. Furthermore, referring to the measured AM noise of the SAWRO based test source which was shown in Fig. 3, it is questionable as to whether the measured residual flicker phase noise level for the UTO-1023 shown in Fig. 10 is really accurate. Phase noise measurements on a pair of oscillators incorporating UTO-1023s confirm that the 1 Hz intercept for each amplifier's residual flicker phase noise is at least -143 dBc/Hz, or perhaps even better, while the residual white phase noise floor is at least -184 dBc/Hz. These results are in very good agreement with the measured small-signal electrical parameters for the UTO-1023s. The situation for the UTO-1023 once again points out the efficacy of also using a test source with relatively high AM noise to establish the amplifier's AM-to-PM conversion factor, in order to properly interpret the amplifier's residual phase noise measurements. Finally, as mentioned previously, testing two UTO-1023 amplifiers at a time (one in each test path) would quite likely assist in reducing the problem of large AM-to-PM conversion factors for the DUTs.

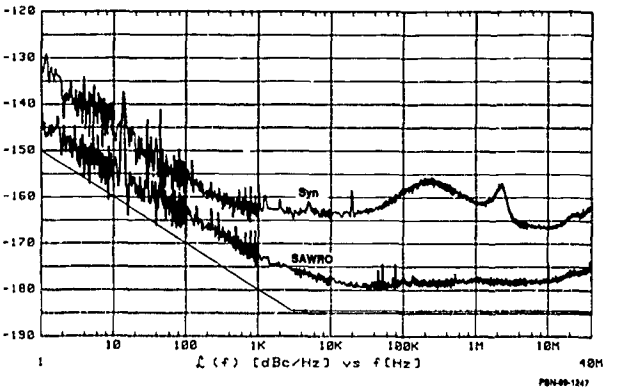


Figure 10. UTO-1023 (Avantek) Amplifier Residual Phase Noise Measurements Using SAWRO and Frequency Synthesizer Based Test Sources. P_{inc}=+17.0 dBm (~ 3 dB of Gain Compression) for Both Measurements.

It should be borne in mind that whether a particular amplifier is used in an oscillator's feedback loop or as a buffer amplifier outside the feedback loop, its residual phase noise properties will directly influence the oscillator's phase noise spectrum. Furthermore, in general, an amplifier can significantly modify the phase noise spectrum of the signal which it is intended to amplify. As oscillator phase noise requirements become progressively more stringent, the residual phase noise properties of the feedback loop and buffer amplifiers become ever more important.

B. GaAs FET RF Amplifiers.

While GaAs FET amplifiers are generally used in higher frequency applications, i.e., frequencies above 2 GHz, there are a number of very high power GaAs FET amplifiers designed for the 500 MHz to 2 GHz frequency range as well. Figure 11 shows the result of a 500 MHz residual phase noise measurement performed on an APG-1021M (Avantek) GaAs FET amplifier, for P_{inc}=+20.5 dBm, which corresponds to approximately 2 dB of gain compression. For the purpose of comparison, the UTO-1023 (Avantek) silicon BJT amplifier's residual phase noise from Fig. 10 is repeated in Fig. 11. Table II contains a summary of the manufacturer's typical electrical performance specifications, along with our measured results on these two amplifiers. As is generally the case when comparing GaAs FET and silicon BJT amplifiers with similar electrical properties, the GaAs FET amplifier will usually have substantially higher residual flicker phase noise levels, typically anywhere from 10 to 30 dB. It should be noted that the predicted residual white phase noise floors, as given by the expression:

$$\{NFkT\}/P_{inc}$$

where NF is the amplifier's noise factor, k is Boltzmann's constant, T is the amplifier's temperature in °K, and P_{inc} is the incident (input) RF power in watts, are -191.9 dBc/Hz for the APG-1021M and -185.5 dBc/Hz for the UTO-1023, respectively. The measured residual white phase noise floor for the APG-1021M is essentially at the system's phase noise floor, indicating that the measurement technique can indeed be used to verify that an amplifier has a residual white phase noise floor substantially better that -186 dBc/Hz. This will be the case when the amplifier's AM-to-PM conversion factor is sufficiently low, as it is for the APG-1021M. The measured effective AM source noise suppression for the residual phase noise

measurement on the APG-1021M was still in excess of 20 dB. As noted previously, truly accurate residual phase noise measurements on a single amplifier such as the UTO-1023 which has a significant AM-to-PM conversion factor, are difficult at best, even when the 500 MHz SAWRO based test source is used.

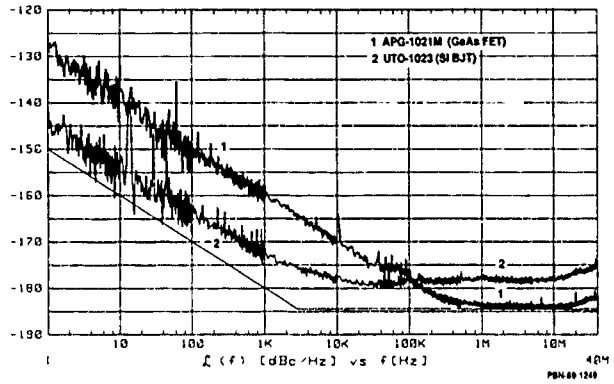


Figure 11. Residual Phase Noise Measurement Comparison for GaAs FET and Si BJT High Power Amplifiers. Curve 1: Avantek APG-1021M, P_{inc}=+20.5 dBm (~ 2 dB of Gain Compression). Curve 2: Avantek UTO-1023, P_{inc}=+17.0 dBm (~ 3 dB of Gain Compression).

The basic SAWRO based test source may also be used at higher frequencies if the appropriate low noise frequency multiplication circuitry is included [11], [12]. Figure 12 shows the typical set-up for a 2 GHz output frequency test source configuration. As mentioned in Section II, the AM and PM noise properties of the test source are very important. Measurements of the 2 GHz test source's AM and PM noise were carried out in order to insure that they don't adversely influence the 2 GHz residual phase noise measurements. The 2 GHz test source configuration was then used to evaluate the residual phase noise properties of two high power GaAs FET amplifiers, namely the Avantek APG-4002M as well as a custom amplifier design. Table III summarizes their electrical properties. Figure 13 shows the measured residual phase noise characteristic for each amplifier. The incident RF power level was +16.0 dBm (~ 1 dB of gain compression) for the APG-4002M amplifier's measurement. For the custom GaAs FET amplifier measurements two incident RF power levels were used, namely +6.5 dBm and +10.0 dBm, which correspond to ~ 1 dB and ~ 4 dB of gain compression, respectively. The effective AM source noise suppression for either amplifier's residual phase noise measurement was still greater than 10 dB. This is sufficient to insure that the test source's AM noise doesn't contribute to the amplifier's measured residual phase noise, at least for the residual phase noise levels shown in Fig. 13. While the residual flicker phase noise of the Avantek amplifier continues to decrease at 10 dB/decade down to the residual white phase noise floor, the custom amplifier exhibits a considerable carrier offset frequency region where the residual phase noise appears to be decreasing at only 5 dB/decade. We have tested a

Table II

Summary of Typical RF Amplifier Characteristics at T = 25°C

Amplifier	Small Signal Gain	Noise Figure	Pout @ 1 dB Compression	Pout @ 3 dB Compression	I_{DC} @ V_{DC}	Residual Phase Noise $\mathcal{L}(f_x 1\ Hz)$
Avantek APG-1021M						
Vendor Typical	>14.0 dB	<3.5 dB	>+33.0 dBm	N.A.	1150 mA @ +15V	N.A.
Measured	(15.5 dB)[1]	(2.3 dB)[1]	(+33.6 dBm)[1]	(+33.8 dBm)[1]	(1125 mA @ +15V)	(~ -130 dBc/Hz)[1,3]
Avantek UTO-1023						
Vendor Typical	13.0 dB	<7.0 dB	+28.0 dBm	N.A.	205 mA @ +15V	N.A.
Measured	(14.5 dB)[1]	(5.9 dB)[1]	(+28.1 dBm)[1]	(+28.9 dBm)[1]	(208 mA @ +15V)	(~ -145 dBc/Hz)[1,2]

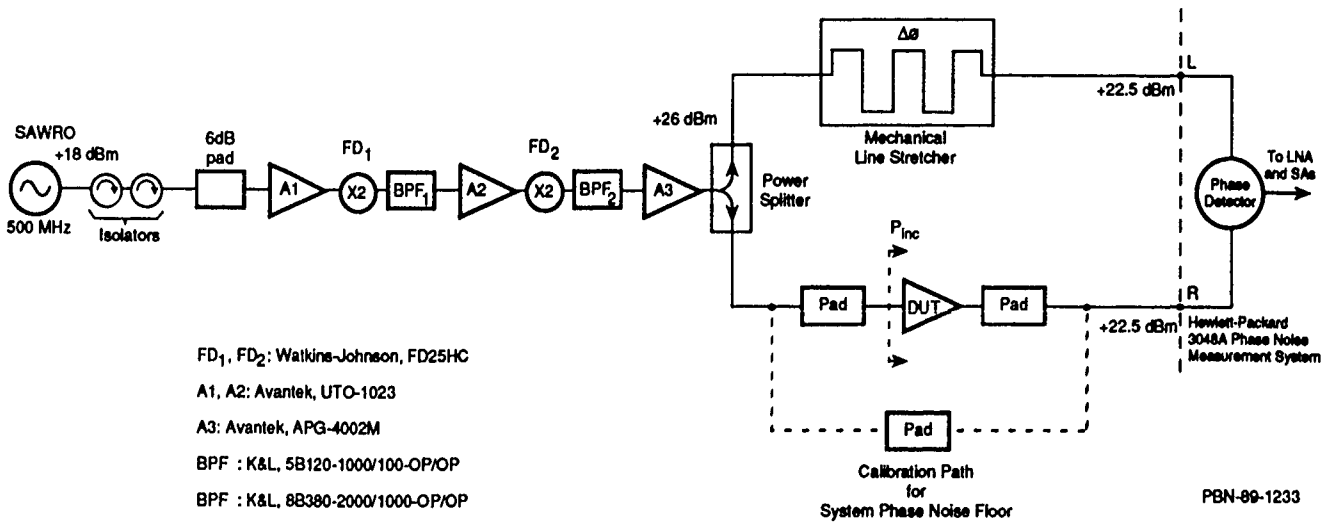
[1] Measured at 500 MHz [2] Measured at 3 dB of Gain Compression [3] Measured at 2 dB of Gain Compression N.A. = Not Applicable

Table III

Summary of Typical RF Amplifier Characteristics at T = 25°C

Amplifier	Small Signal Gain	Noise Figure	Pout @ 1 dB Compression	Pout @ 3 dB Compression	I_{DC} @ V_{DC}	Residual Phase Noise[1] $\mathcal{L}(f_x 1\ Hz)$
Avantek APG-4002M						
Vendor Typical	>15.0 dB	<6.5 dB	>+30.0 dBm	N.A.	1100 mA @ +15V	N.A.
Measured	(17.5 dB)[1]	(4.1 dB)[1]	(+31.3 dBm)[1]	(+33.5 dBm)[1]	(1200 mA @ +15V)	(~ -124 dBc/Hz)[2]
Custom GaAs FET Amplifier						
Vendor Typical	>21.0 dB	<7.0 dB	+28.0 dBm	N.A.	~500 mA @ +15V	N.A.
Measured	(22.4 dB)[1]	(5.4 dB)[1]	(+28.4 dBm)[1]	(+30.1 dBm)[1]	(510 mA @ +15V)	(~ -120 dBc/Hz)[2]

[1] Measured at 2 GHz [2] Measured at 1 dB of Gain Compression N.A. = Not Applicable



FD_1, FD_2: Watkins-Johnson, FD25HC

A1, A2: Avantek, UTO-1023

A3: Avantek, APG-4002M

BPF : K&L, 5B120-1000/100-OP/OP

BPF : K&L, 8B380-2000/1000-OP/OP

Figure 12. Residual Phase Noise Measurement Set-up at 2 GHz.

number of amplifiers, both silicon BJT and GaAs FET, from several vendors with residual phase noise anomalies such as this. While we have not as yet been able to confirm our suspicions, in this specific case the excess residual phase noise for these carrier offset frequencies is perhaps due to the relatively complex voltage regulator and/or temperature compensation circuitry included in this GaAs FET amplifier. Once again, care must be exercised in selecting an amplifier for use in any application that is sensitive to its residual phase noise properties. A perfect case in point is the basic frequency multiplier chain shown in Fig. 12, where an amplifier's excess residual noise, either PM or AM, may seriously degrade the noise spectra of the frequency translated output signal.

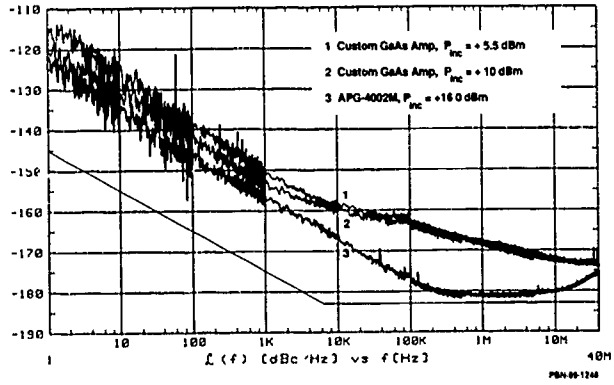


Figure 13. Residual Phase Noise Measurements at 2 GHz for Two Different High Power GaAs FET Amplifiers. Curve 1: Custom GaAs FET Amplifier, P_{inc}=+5.5 dBm (~ 1 dB of Gain Compression). Curve 2: Custom GaAs FET Amplifier, P_{inc}=+10.0 dBm (~ 4 dB of Gain Compression). Curve 3: Avantek APG-4002M, P_{inc}=+16.0 dBm (~ 1 dB of Gain Compression).

C. Electronic Phase Shifters.

The basic residual phase noise measurement technique may also be applied to other linear components as well, e.g., an electronic phase shifter. Since an electronic phase shifter is often used in the feedback loop of a voltage controlled oscillator, its residual phase noise is very important. Figure 14 shows the circuit schematic for a low noise, electronic phase shifter with an extremely linear transmission phase versus voltage tuning characteristic. This basic phase shifter design has been used with considerable success in a number of low noise, 500 MHz SAWROs [11], [12], [20]. The specific component values are as follows:

R_1 = 2 kΩ; TRX, P/N: RP0705BS-2001JN-01,

L_1 = ~ 20 nH; Piconics, P/N: MC5T47PW,

$C_J(V_{MR}$ = 5 v) = 6.4 pF; Frequency Sources, P/N: 15000 Series,

$C_B = C_C$ = 1000 pF; ATC, P/N: 700A-102K-CA-150X,

where V_{MR} is the tuning voltage at the center of the range. The circuit has been fully hybridized onto a 0.240"x0.260"x0.020" alumina substrate and hermetically sealed in a standard four pin, TO-8 package.

Figure 15 shows a typical residual phase noise measurement on an electronic phase shifter. The test was performed with an RF power level of +23.5 dBm incident on the device. This is almost 10 dB higher than the actual RF power level which the phase shifter sees in the SAWRO circuit! The measured residual flicker phase noise level is still within 2 to 3 dB of the measurement system's phase noise floor. Since this result is so close to the measurement system's flicker phase noise floor, we can only infer that the electronic phase shifter's 1 Hz intercept for its residual flicker phase noise is approximately -150 dBc/Hz, or possibly even lower. This is still comfortably below the UTO-509 loop amplifier's typical 1 Hz intercept for its residual flicker phase noise of -147 dBc/Hz. In point of fact, the 500 MHz SAW resonator's typical 1 Hz intercept for its residual flicker phase noise is in the range of -141 to -145 dBc/Hz. While this is still exceptionally good for a SAW resonator, it is nonetheless the most likely source of flicker phase noise in the 500 MHz SAWRO's feedback loop.

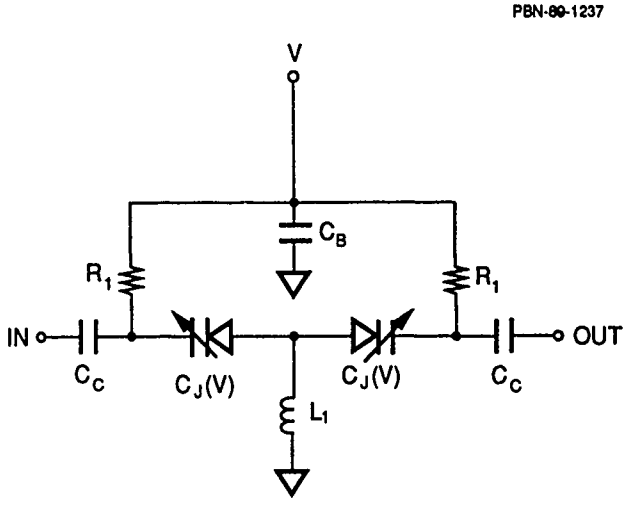


Figure 14. Circuit Schematic of a Low Phase Noise, Electronic Phase Shifter with Very Linear Phase Versus Tuning Voltage Characteristic.

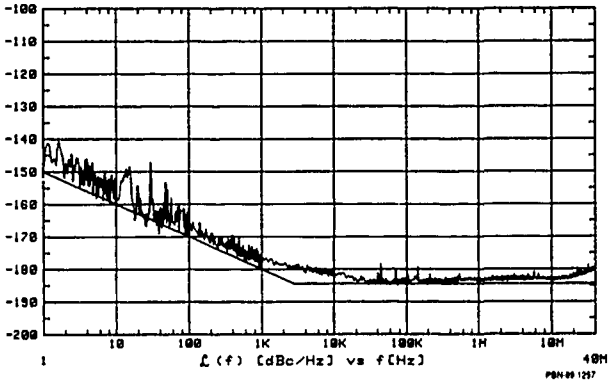


Figure 15. Typical Residual Phase Noise for an Electronic Phase Shifter. P_{inc}=+23.5 dBm.

D. Frequency Dividers and Multipliers.

Frequency division and multiplication are also common system requirements. Ideally, the input source's phase noise after frequency multiplication or division, will be altered by 20*log{N}, where N is the ratio of output to input frequency. In reality, for a variety of reasons, it is difficult to achieve this theoretical limit. Near ideal frequency multiplication to X-Band (X16) has been demonstrated using a simple cascade of four X2 frequency doublers [11], [12], but multiplication with step recovery diodes has generally been found to be worse than theoretical. Digital frequency dividers have generally been shown to

have a residual white phase noise floor (at the output frequency) in the range of -145 to -160 dBc/Hz [2], [6] - [9], at least for carrier offset frequencies greater than 10 kHz. The basic measurement technique reported herein may also be applied to the evaluation of frequency translation devices, as indicated in Fig. 16, which shows a "typical" measurement set-up which might be used to test a ***pair*** of frequency dividers or multipliers. However, each device type to be tested will require some minor adjustments to the basic configuration shown in Fig. 16 in order to accommodate a wide range of input power requirements and conversion losses for the various DUTs. Note the extensive use of isolators, as well as bandpass filters at the outputs of the DUTs. Our experience indicates that this is necessary for accurate phase noise measurements.

Figure 17 illustrates the measured residual phase noise spectrum, at 500 MHz, for a ***pair*** of regenerative frequency dividers (Telemus HB-0102s, +2) tested with +15 dBm input signals at 1 GHz. The measurement system's approximate phase noise floor is shown as a pair of solid lines in the figure. It is interesting to note that the residual phase noise for carrier offset frequencies between 100 kHz and 1 MHz rises quite abruptly and appears to plateau at approximately -132 dBc/Hz for carrier offset frequencies from 1 Hz to 10 kHz. This general behavior was independently confirmed by testing the pair of dividers with two low noise SAWROs in a standard phase locked noise measurement set-up. The level of the device's residual phase noise is some 15 to 40 dB above the measurement system's noise floor in the 1 Hz to 1 kHz carrier offset range and would seriously degrade the divided down phase noise spectrum of an extremely low phase noise microwave source, such as might be derived from the basic 500 MHz SAWRO used herein as a test source. Figure 18 shows the measured AM noise spectrum for one of the signals which is input to the phase detector. The effective AM source noise suppression for the residual phase noise measurement was at least 30 dB. Therefore, the Telemus HB-0102 dividers are a source of residual PM and AM noise at almost equal levels. This fact suggests that perhaps direct additive noise voltage signals within the dividers are the source of both the PM and AM noise observed in our measurements.

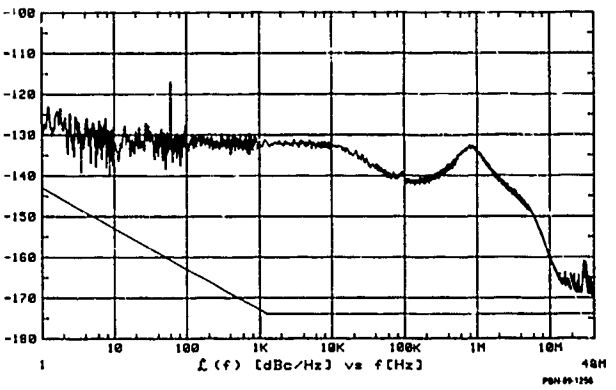


Figure 17. Residual Phase Noise Measurement, at the 500 MHz Outputs, for a ***Pair*** of Telemus HB-0102 Regenerative Frequency Dividers (+2). Input Frequency was 1 GHz, P_{inc}=+15.0 dBm. The Two Solid Straight Lines Represent an Approximate Fit to the Measured System Phase Noise Floor Using the SAWRO Based Test Source.

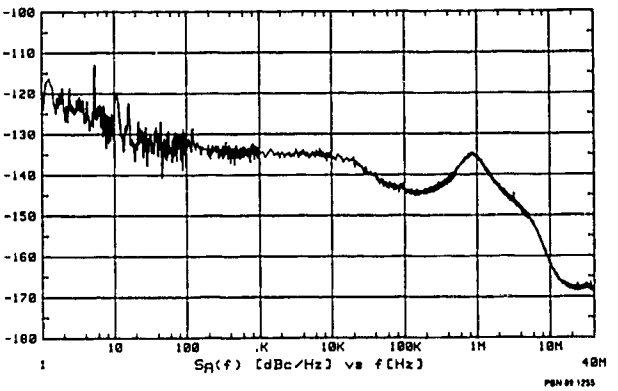


Figure 18. AM Noise at 500 MHz Output for One Telemus HB-0102 Frequency Divider (+2).

A Plessey SP8606A integrated circuit divide-by-two was also characterized using the basic test set-up shown in Fig. 16. The input frequency was 1 GHz, and the incident RF power to each divider was +5 dBm. The measured residual phase noise at 500 MHz for a ***pair*** of SP8606s is shown in Fig. 19. The measurement system's approximate phase noise floor is shown as a pair of solid lines in the figure. This result is in good agreement with other results published in the literature for a variety of integrated circuit dividers [2], [6] - [9]. Generally, these measurements indicate that it is virtually impossible to preserve the phase noise floor of an extremely low noise oscillator signal when it passes through one of these circuits. This is primarily due to the rather low output power of the device after division-by-two.

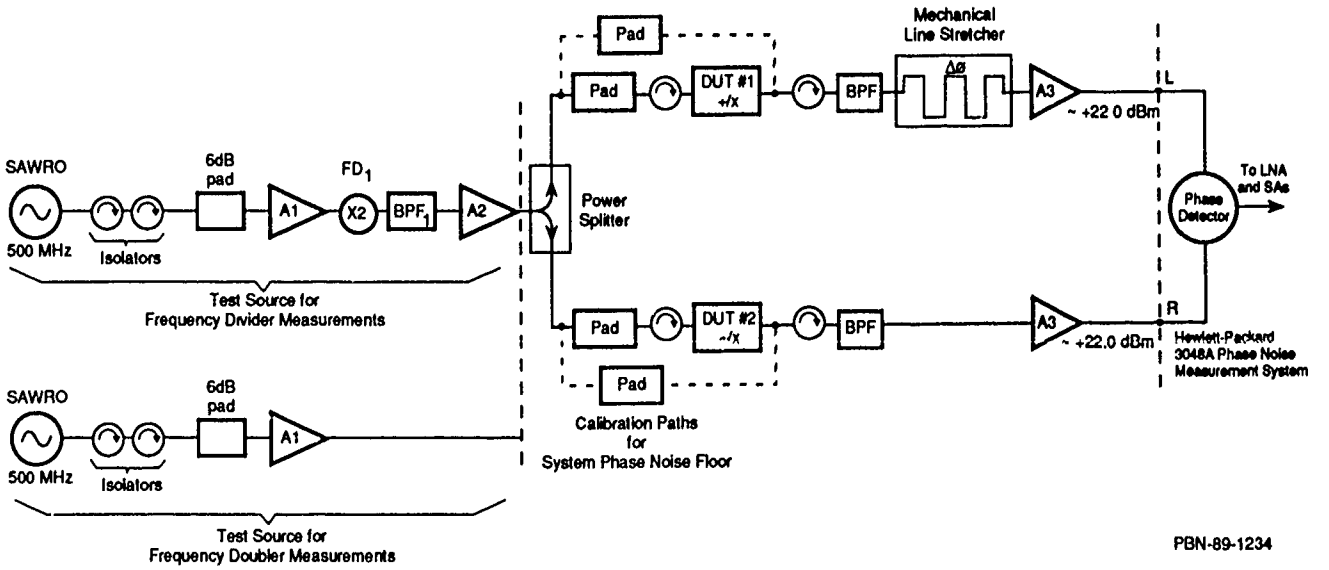


Figure 16. Residual Phase Noise Measurement Set-up for Frequency Dividers and Multipliers.

and the fact that this signal must once again be amplified.

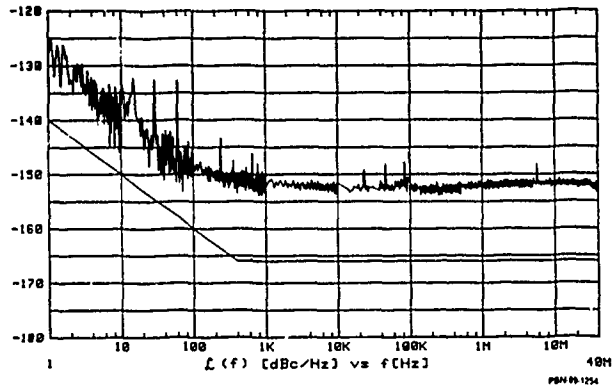


Figure 19. Residual Phase Noise, at the 500 MHz Outputs, for a ***Pair*** of Plessey SP8606A Integrated Circuit Dividers (÷2). Input Frequency was 1 GHz, P_{inc}=+5 dBm. The Two Solid Straight Lines Represent an Approximate Fit to the Measured System Phase Floor Using the SAWRO Based Test Source.

Residual phase noise measurements on frequency doublers are also possible using the measurement test set-up shown in Fig. 16. However, such results tend to be limited in dynamic range due to the rather large conversion losses of the DUTs. Figure 20 shows the result of a residual phase noise measurement on a ***pair*** of Watkins-Johnson FD25HC frequency doublers, for an incident RF power level of +23.5 dBm at 500 MHz. The measurement system's approximate phase noise noise floor is shown as a pair of solid lines in the figure. For all intents and purposes, we are unable to attribute any residual phase noise to the X2 multipliers, given the measurement system's phase noise floor limitation shown in Fig. 20. However, these are the same frequency doublers used in Refs. [11], [12] to achieve

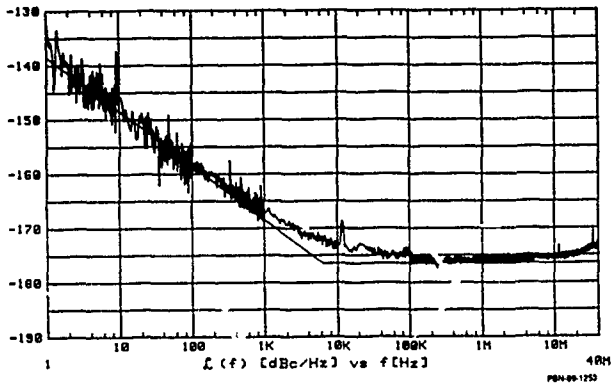


Figure 20. Residual Phase Noise, at the 1 GHz Outputs, for a ***Pair*** of Watkins-Johnson FD25HC Frequency Doublers. Input Frequency was 500 MHz, P_{inc}=+23.5 dBm. The Two Solid Straight Lines Represent an Approximate Fit to the Measured System Phase Noise Floor Using the SAWRO Based Test Source.

essentially ideal frequency multiplication in a pair of X16 multiplier chains with an output frequency of 8 GHz. These previous results suggest that the actual residual phase noise of the FD25HC frequency doubler is substantially below that indicated in Fig. 20, and therefore our measurement simply serves to place an upper limit on its residual phase noise.

Finally, the noise properties for a pair of Hewlett-Packard Step Recovery Diode (SRD) modules (HP33004A) were investigated. Once again the conversion losses of the DUTs preclude a direct residual phase noise measurement. Therefore, a series of two oscillator phase noise measurements were performed at successively higher multiplication ratios. The measurement test set-up is shown in Fig. 21. Note, as indicated in Fig. 21, that two different ***pairs*** of

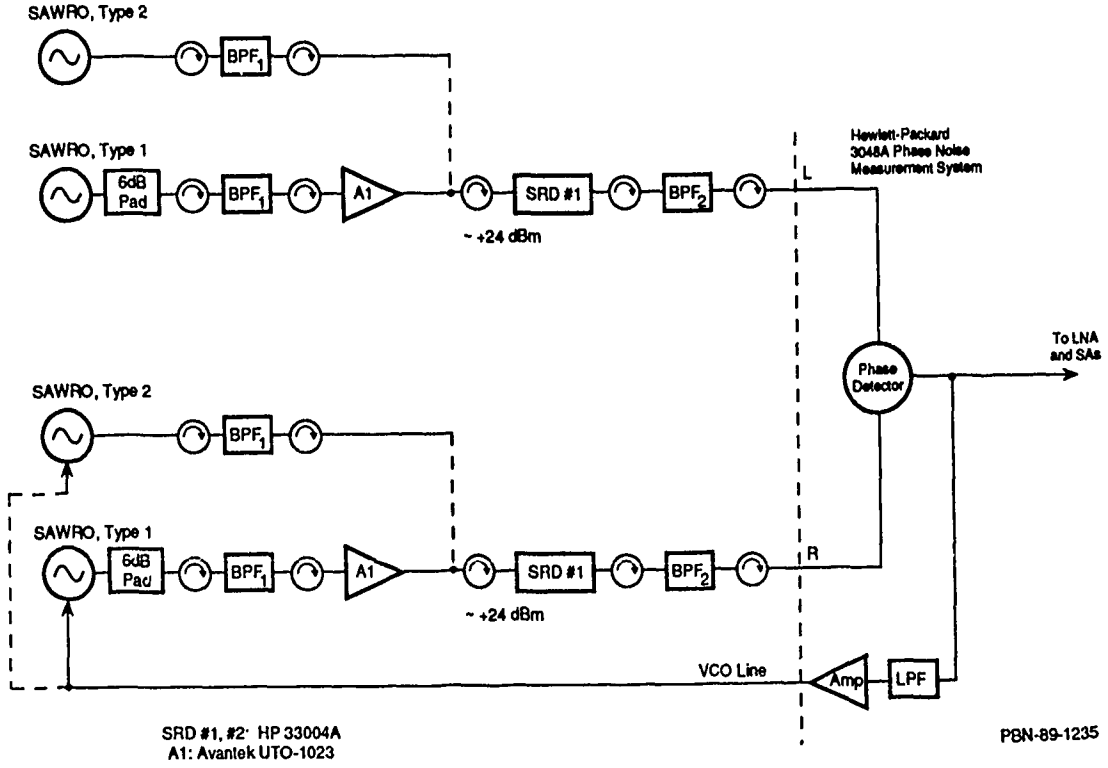


Figure 21. Measurement Test Set-up to Investigate the Noise Properties of a ***Pair*** of Hewlett-Packard 33004A Step Recovery Diode (SRD) Frequency Multipliers.

SAWROs were used for these measurements. The pairs, shown as "SAWROs Type 1" and "SAWROs Type 2" in Fig. 21, were nominally at 500 MHz, however, the second pair (SAWROs Type 2) was almost 7 dB quieter in the white phase noise portion of the spectrum, for carrier offset frequencies greater than approximately 100 kHz. Figure 22 shows data taken at 500 MHz using the first ***pair*** of 500 MHz fundamental frequency oscillators (SAWROs Type 1), as well as data for multiplication ratios of 8 (18.1 dB) and 14 (22.9 dB). It is evident that close to ideal frequency multiplication has been achieved, except for the carrier offset region between 40 kHz and 4 MHz. The overall performance is comparable to the performance previously reported for the frequency doubler based multiplier chains. However, there are indications that the spectral region for carrier offset frequencies between 40 kHz and 4 MHz has been degraded by approximately 4 dB more than theoretical. Thus, the SRDs are only about 4 dB worse than theoretical in their translation of the oscillators' phase noise spectrum. Also indicated in Fig. 22 is a 10 dB/decade line which suggests that for the 4 GHz output frequency (X8) the residual flicker phase noise attributable to the pair of SRDs must be less than -105 dBc/Hz at its 1 Hz intercept. Figure 23 shows similar data taken using the second ***pair*** of SAWROs (SAWROs Type 2). The significantly lower starting white phase noise floor for this pair of oscillators serves to more clearly reveal the carrier offset region over which worse than theoretical phase noise degradation is taking place. Now the SRDs are seen to be as much as 6 dB worse than theoretical over the carrier offset frequency range from approximately 10 kHz to 4 MHz. Also indicated in Fig. 23 is a 10 dB/decade line which suggests that for the 4 GHz output frequency (X8) the residual flicker phase noise attributable to the pair of SRDs must now be less than -110 dBc/Hz at its 1 Hz intercept. While we have seen step recovery diode multipliers from other vendors which would have seriously degraded the overall phase noise spectrum of the mutiplied output signal, the Hewlett-Packard SRD modules are relatively low in their residual phase noise levels, and certainly capable of achieving close to ideal multiplication of extremely low noise input signals, at least when used with care in an appropriate circuit configuration. During the course of these measurements we found that the SRDs' performance was sensitive to the specific circuit environment in which it was embedded. The liberal use of isolators and bandpass filters was once again found to be highly advantageous for these measurements.

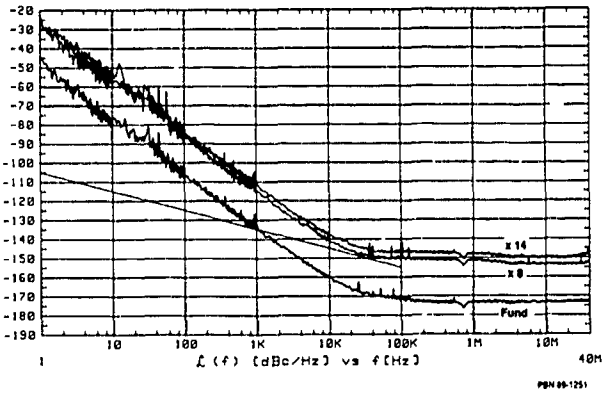


Figure 22. Phase Noise for a ***Pair*** of 500 MHz SAWRO Based Sources Before and After Frequency Multiplication (X8, X14) With SRDs. SAWRO Type 1 Test Sources Were Used. (See Fig. 21.)

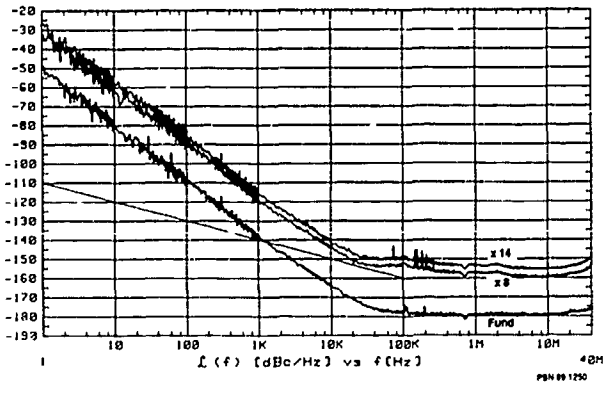


Figure 23. Phase Noise for a ***Pair*** of 500 MHz SAWRO Based Sources Before and After Frequency Multiplication (X8, X14) With SRDs. SAWRO Type 2 Test Sources Were Used. (See Fig. 21.)

IV. SUMMARY

A measurement technique based upon the utilization of an extremely low noise, both AM and PM, SAWRO has been described which permits the accurate evaluation of component residual phase noise properties while maintaining exceptionally good sensitivity for carrier offset frequencies from 1 Hz to 40 MHz. Data for a variety of amplifiers, frequency dividers and multipliers, etc. was included.

REFERENCES

[1] 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.

[2] F. L. Walls and C. M. Felton, "Low Noise Frequency Synthesis," in *Proceedings of the 41st Annual Frequency Control Symposium*, 1987, pp. 512-518.

[3] T. R. Faulkner and R. E. Temple, "Residual Phase Noise and AM Noise Measurements and Techniques," Hewlett-Packard Application Note, H-P Part Number 03048-90011.

[4] G. S. Curtis, "The Relationship Between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques," in *Proceedings of the 41st Annual Frequency Control Symposium*, 1987, pp. 420-428.

[5] T. F. O'Shea, M. R. Lewis, and B. H. Horine, "Low-Noise Voltage Controlled SAW Oscillator for Phase Lock Loop Application," in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 53-59.

[6] D. E. Phillips, "Random Noise in Digital Gates and Dividers," in *Proceedings of the 41st Annual Frequency Control Symposium*, 1987, pp. 507-511.

[7] K. K. Agarwal, W. J. Thompson, and P. M. Asbeck, "Phase-Noise Behavior of Frequency Dividers Implemented with GaAs Heterojunction Bipolar Transistors," *Electronics Letters*, Vol. 21, No. 22, pp. 1005-1006, 24th October 1985.

[8] M. C. Wilson, P. C. Hunt, S. Duncan, and D. J. Bazley, "10.7 GHz Frequency Divider Using Double Layer Silicon Bipolar Process Technology," *Electronics Letters*, Vol. 24, No. 15, pp. 920-922, 21st July 1988.

[9] J. A. Borras, "GaAs Phase Noise Characteristics in High Frequency Synthesizer Prescalers," in *Technical Digest of the 10th Annual GaAs IC Symposium*, 1988, pp. 285-286.

[10] M. J. Loboda, T. E. Parker, J. A. Greer, and G. K. Montress, "Reduction of Close-to-Carrier Phase Noise in Surface Acoustic Wave Resonators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 43-46.

[11] G. K. Montress, T. E. Parker, and M. J. Loboda, "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 47-52.

[12] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low Phase Noise SAW Resonators and Oscillators: Design and Performance," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 6, pp. 657-667, November 1988.

[13] J. G. Ondria, "A Microwave System for Measurements of AM and FM Noise Spectra," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-16, No. 9, pp. 767-781, September 1968.

[14] J. L. Fikart and P. A. Goud, "The Direct-Detection Noise-Measuring System and Its Threshold," *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-21, No. 3, pp. 219-224, August 1972.

[15] 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.

[16] M. M. Driscoll and B. W. Kramer, "Spectral Degradation in VHF Crystal Controlled Oscillators due to Short-Term Instability in the Quartz Resonator," in *Proceedings of the IEEE Ultrasonics Symposium*, 1985, pp. 340-345.

[17] R. L. Baer, "Phase Noise in Surface-Acoustic-Wave Filters and Resonators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 3, pp. 421-425, May 1988

[18] G. D. Alley and H.-C. Wang, "An Ultra-Low Noise Microwave Synthesizer," *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-27, No. 12, pp. 969-974, December 1979.

[19] E. C. Niehenke and P. A. Green, "A Low-Noise L-Band Dielectric Resonator Stabilized Microstrip Oscillator," in *Digest of the IEEE MTT-S International Microwave Symposium*, 1987, Vol. 1, pp. 193-196.

[20] T. E. Parker and G. K. Montress, "Low Noise SAW Resonator Oscillators," to be published in *Proceedings of the 43rd Annual Frequency Control Symposium*, 1989. Paper presented in Session A10 - SAW Devices and Phase Noise.

43rd Annual Symposium on Frequency Control - 1989

MEASUREMENT METHOD OF VHF ELASTIC VIBRATIONS BY OPTICAL FIBER INTERFEROMETRIC SENSING

Yasuo Hirose and Yasuo Tsuzuki

Faculty of Engineering, Yokohama National University
Tokiwadai, Hodogaya-ku, Yokohama, Japan 240

Abstract

This paper presents a measurement method of elastic vibration modes with extremely small vibration amplitudes in the very high frequency (VHF) range. The measurement system is an optical fiber interferometric sensor, which detects optical phase modulations caused by elastic vibrations being tested to give the vibration amplitudes. Extensive use of recent optical fiber technology and a signal processing technique enabled us to achieve detailed measurements of vibration amplitudes as small as 0.1 Å at frequencies as high as nearly 150 MHz.

Introduction

In recent years, electro-mechanical devices using elastic vibrations quartz crystal resonators, piezoelectric ceramic filters and surface-acoustic-wave devices for instance, have been in wide use in electronic equipment. They are indispensable elements for control and selection of frequencies from several kHz to as high as several hundred MHz.

For development and design of such devices, direct measurements of their vibration mode are of particular importance, since their theoretical vibration analysis is in many cases difficult. For this reason, a variety of vibration measurement methods have been developed. Methods which fully utilize the inherent high sensitivity of optical sensing include the holographic method[1] and the laser interferometric method[2],[3] for example. However, these do not appear to have the sufficient capability for extremely small elastic vibrations of VHF frequency devices in most extensive practical use. The vibration amplitudes of these devices are typically below several tens of angstroms. Moreover, high spatial resolution is required to perform accurate mode profile measurements, since the actual VHF devices are quite small.

To overcome these problems, we have developed a measurement method of VHF elastic vibrations by optical fiber interferometric sensing. The method fully utilizes recent optical fiber and micro-optic component technology, and a high S/N ratio signal processing technique at high frequencies. In this method, measurement of any of the in-surface and normal direction vibrations is possible. The minimum detectable amplitude is of the order of 0.1Å for vibrations at frequencies as high as VHF frequencies. A unique feedback compensation scheme is used to stabilize interference signals against phase variations due to external disturbances, and accordingly, to achieve high sensitivity of measurement.

In the following, the measurement method and its performance are described with special emphasis on its applications to VHF elastic vibrations.

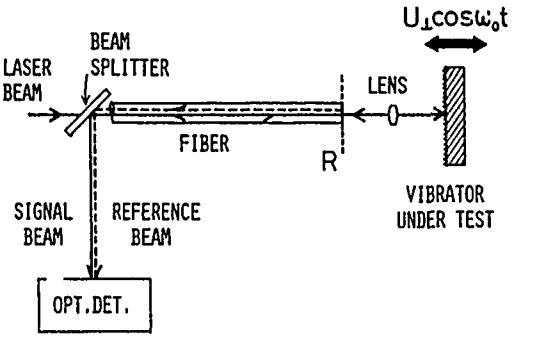
Principle of Measurement Method

In the present method, the amplitude of a vibration is determined by the homodyne-detection and processing of the relative phase modulation which appears between two interfering light beams and which is caused by the vibration to be measured. In the following, the detection principle is outlined.

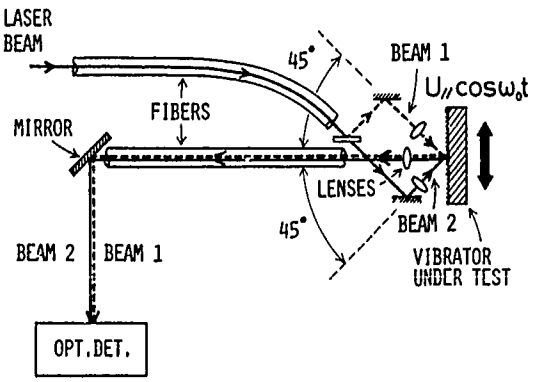
Optical path configurations

Figures 1(a) and 1(b) illustrate optical path configurations for detecting the normal and in-surface direction vibrations, respectively.

In Fig. 1(a), a laser beam transmitted through a fiber is focused onto one point of the vibrator under test. The light reflected back to the fiber is used as the signal beam and guided through the same fiber to an optical detector. The light partially reflected



(a) For the normal direction vibration measurements.



(b) For the in-surface direction vibration measurements.

Fig.1. Optical path configurations.

CH2690-6/89/0000-360 $1.00 C 1989 IEEE

back at the far end R of the fiber is used as the reference beam, as indicated by the dotted line. In this case, the signal beam is phase modulated by the normal direction vibration $U_\perp \cos\omega_0 t$ of the vibrator surface. As a result, the following relative phase modulation appears between the signal and reference beams.

$$\Delta\phi_\perp = (4\pi/\lambda)\, U_\perp \cos\omega_0 t \qquad (1)$$

where λ is the wavelength of the laser light.

In Fig.1(b), a laser beam guided through a fiber is split into two beams. They are then focused onto a point on the diffusely finished surface of the vibrator being tested from upward and downward directions at angles of $\pm 45°$. Scattered lights into the direction normal to the vibrator surface are collected and then guided through another fiber toward an optical detector. These two light beams undergo phase modulations of the same magnitude but of opposite sign by the in-surface direction vibration $U_{//} \cos\omega_0 t$. This is because, at some instant of time, the reflecting point motion due to the in-surface vibration lengthen the optical path of one beam, while this motion shorten that of the other beam. Accordingly, these beams produce the following relative phase modulation between them.

$$\Delta\phi_{//} = (2\sqrt{2}\,\pi/\lambda)\, U_{//} \cos\omega_0 t \qquad (2)$$

Detection principle of vibration amplitudes

Vibration amplitudes are obtained by detecting the relative phase modulations given by eqs. (1) and (2) by some means. In this section, the principle of measuring the in-surface direction vibrations by the homodyne detection is described,[4] since that of the normal direction vibrations is essentially the same.

The interference light intensity incident upon the optical detector is given by

$$I = E_1^2 + E_2^2 + 2E_1E_2 \cos[(2\sqrt{2}\,\pi/\lambda) \times U_{//}\cos\omega_0 t + \phi_\ell \cos\Omega t + \pi/2] \qquad (3)$$

where E_1 and E_2 are the field amplitudes of the two interfering light beams. In eq.(3), the presence of a dc phase difference of $\pi/2$ is essential for the homodyne signal detection, and this phase difference must be maintained at this value during detection. $\phi_\ell \cos\Omega t$ is low frequency (~1kHz) phase modulation which is necessary to monitor light intensity change during measurements. This modulation is also necessary to generate control signals for suppressing phase difference variation from $\pi/2$ by a negative feedback technique, as will be described later.

If we assume that $U_{//} \ll \lambda$ and $\phi_\ell \ll 1$, then expansion of eq.(3) into a Neumann series in terms of Bessel functions and use of approximations $J_0(x) \approx 1$ and $J_1(x) \approx x/2$ for small x give the following two frequency components among others.

$$I_1(\omega_0) = -4\sqrt{2}\,\pi E_1E_2 (U_{//}/\lambda)\cos\omega_0 t \qquad (4)$$

$$I_0(\Omega) = -2E_1E_2\phi_\ell \cos\Omega t \qquad (5)$$

Measuring the maximum amplitudes I_{1P} and I_{0P} of these components, and then taking their ratio to eliminate E_1 and E_2 which directly reflect the intensity change of the interfering light beams, the following expression for $U_{//}$ is obtained.

$$U_{//} = (\sqrt{2}/4\pi)\,\phi_\ell\,(I_{1P}/I_{0P})\,\lambda \qquad (6)$$

The detection principle of the normal direction vibration $U_\perp \cos\omega_0 t$ is quite similar to that of the in-surface direction vibration. The corresponding expression for this vibration amplitude $U_\perp$ is obtained by multiplying the right-hand side of eq.(6) by an inclination factor of $\cos 45° = 1/\sqrt{2}$.

The dc phase difference of $\pi/2$ in eq.(3) is easily disturbed by various factors such as mechanical vibrations of the system or environmental temperature change. To suppress this phase difference variation by a negative feedback control technique, the amplitude of the second harmonics of the low frequency phase modulation component given in eq.(5) can be used. This amplitude is nearly proportional to the phase difference variation from $\pi/2$, and hence, is used as an error signal for the control, as will be described later.

Measurement system

Figure 2 shows a schema of the measurement system. In order to achieve high spatial resolution of measurement, the sensor head is constructed with micro-optic components and phase modulators using piezoelectric-ceramic actuators. In this structure of the sensor head, laser beams can be focused to a spot with a diameter of nearly 10 μm, and hence, the spatial resolution of the same order of magnitude is expected. Single mode fibers approximately 1.5 m long separate this head from the main system consisting of three subsystems, a light source, a signal detection and processing system, and a phase variation compensation system. By the proper selection of fibers, any of the in-surface and normal direction vibrations can be measured on a point by point basis.

The system in Fig.2 is set for measurement of the in-surface direction vibration. The laser beam from a He-Ne laser is guided

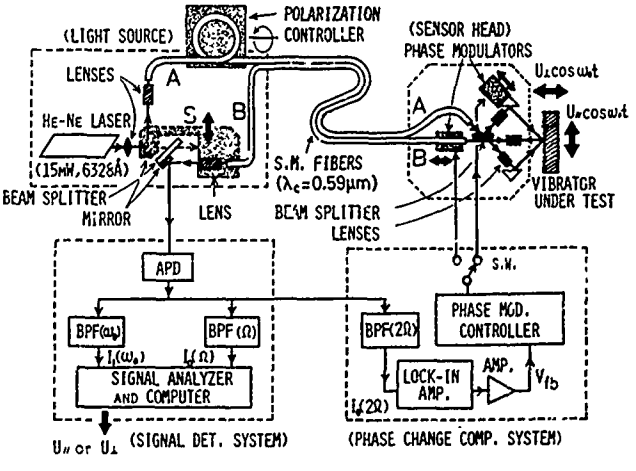


Fig.2. Measurement system.

through fiber A to the sensor head. Scattered lights into the direction normal to the vibrator surface are then collected and guided through fiber B toward an optical detector to achieve the optical path configuration shown in Fig.1(b).

The system can easily be switched to that for measuring the normal direction vibration by simply shifting upward the stage S, on which both the input end of fiber B and a mirror used for directing the laser beam to fiber A are attached. This shift makes the laser beam through a beam splitter be incident only on fiber B to realize the optical path configuration shown in Fig. 1(a).

To achieve highly sensitive detection of vibrations of small amplitude, it is crucial to detect small optical signals with high S/N ratios especially at VHF frequencies. Special care was therefore taken with the signal detection and processing system. For the optical detector, an avalanche photodiode followed by a frequency selective amplifier is used. The amplifier contains a tuning inductor connected in parallel to parasitic or stray capacitances of the photodiode and the amplifier input to prevent the decrease of detector gain at high frequencies. From the output signals of the amplifier, the signal components necessary for detection, $I_1(\omega_0)$ and $I_0(\Omega)$ given in eqs. (4) and (5), are extracted by band pass filters. The signal component $I_1(\omega_0)$, which becomes quite small with decreasing vibration amplitude, is further processed by a high frequency resolution signal analyser with a bandwidth of several Hz to achieve high S/N ratio detection.

In designing the optical system, care was taken to minimize the phase difference variation from $\pi/2$ between the interfering light beams due to external disturbances. Most of the optical paths are made common to these two beams, so that the phase changes are canceled. To further reduce the phase difference variation by a feedback control technique, the amplitude of the second harmonics of the low frequency phase modulation component given in eq.(5) is detected by a lock-in amplifier. The detected signal is then amplified and fed back as a control signal to the phase modulator on the sensor head through the modulator controller, which supplies signals to the modulator to produce the phase difference of $\pi/2$ and the low frequency phase modulation stated earlier.

Measurement System Performance

Figure 3 shows the detection capability of vibration amplitudes of this measurement system which was measured by changing the excitation current of a 53.7 MHz fifth overtone AT-cut quartz crystal resonator depicted in Fig.4. Excellent proportionalities were obtained down to the order of 0.1Å. The detection sensitivities are at least by one order of magnitude higher than those of the previously reported laser interferometric method.[2] The lower sensitivity for the in-surface direction vibration is mainly because of a small amount of light which can be used for detection, since only a small fraction of light is scattered from the 45° to the normal direction of the vibrator surface, as can be seen in Fig.1(b).

Detection of vibration amplitudes at frequencies from several kHz to over 150 MHz was found possible.

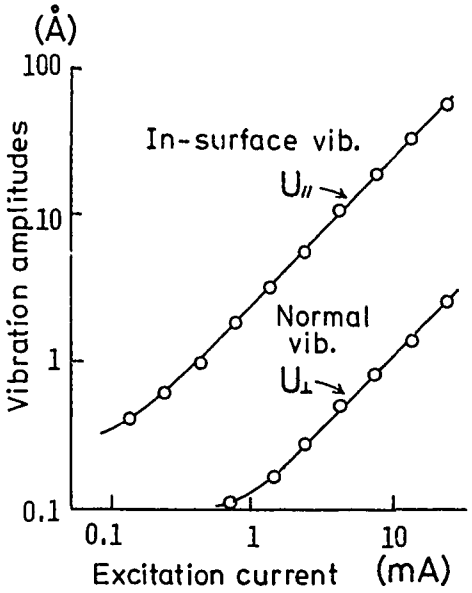


Fig.3. Detection capability of vibration amplitudes.

To demonstrate the validity of the present method, measurements of VHF thickness-shear vibration modes were performed. The vibrator tested is the same AT-cut quartz resonator as that shown in Fig.4. Fig.5(a) shows the frequency spectrum of an excitation current of the resonator near the resonant frequency 53.7 MHz. Fig.5(b) shows a spectrum of an optical signal which was obtained when the in-surface vibrations were measured at the resonator center. In this spectrum, the response peaks correspond to the resonant modes and their magnitudes are proportional to the amplitudes U_x of the vibration in the x direction.

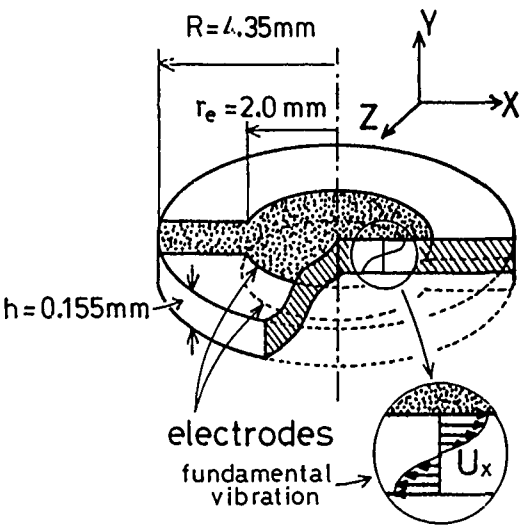
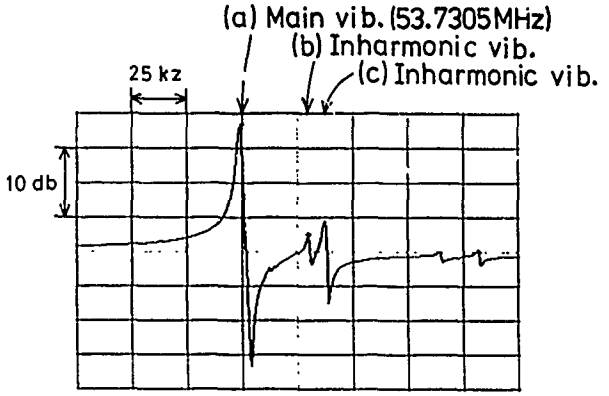
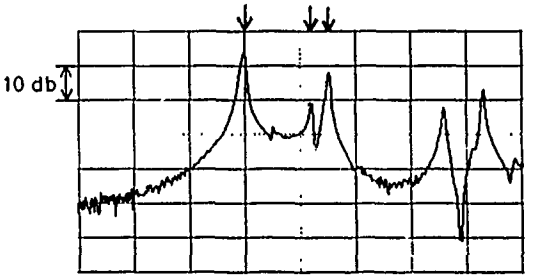


Fig.4. AT-cut quartz resonator.



(a) Excitation current.



(b) Optical signal.

Fig.5. Frequency spectra of excitation current and optical signal near the resonant frequency.

Figure 6 shows an example of vibration amplitude distributions of the main vibration measured along several lines parallel to the x axis of the resonator as shown in the inset of the figure.

Figures 7(a), (b) and (c) show vibration modes of the fifth overtone main and inharmonic vibrations indicated by the arrows in Fig.5, illustrated by loci of equal amplitude and three dimensional profiles. Measurements of these modes were performed by scanning the

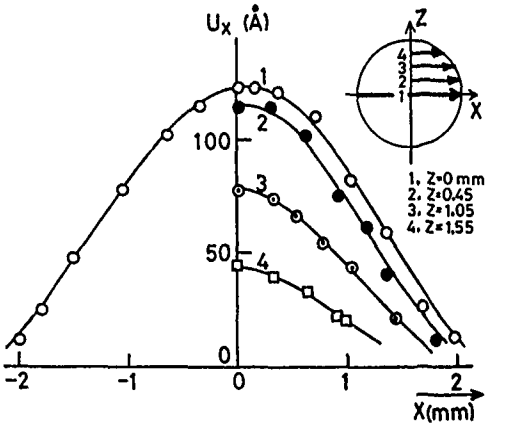
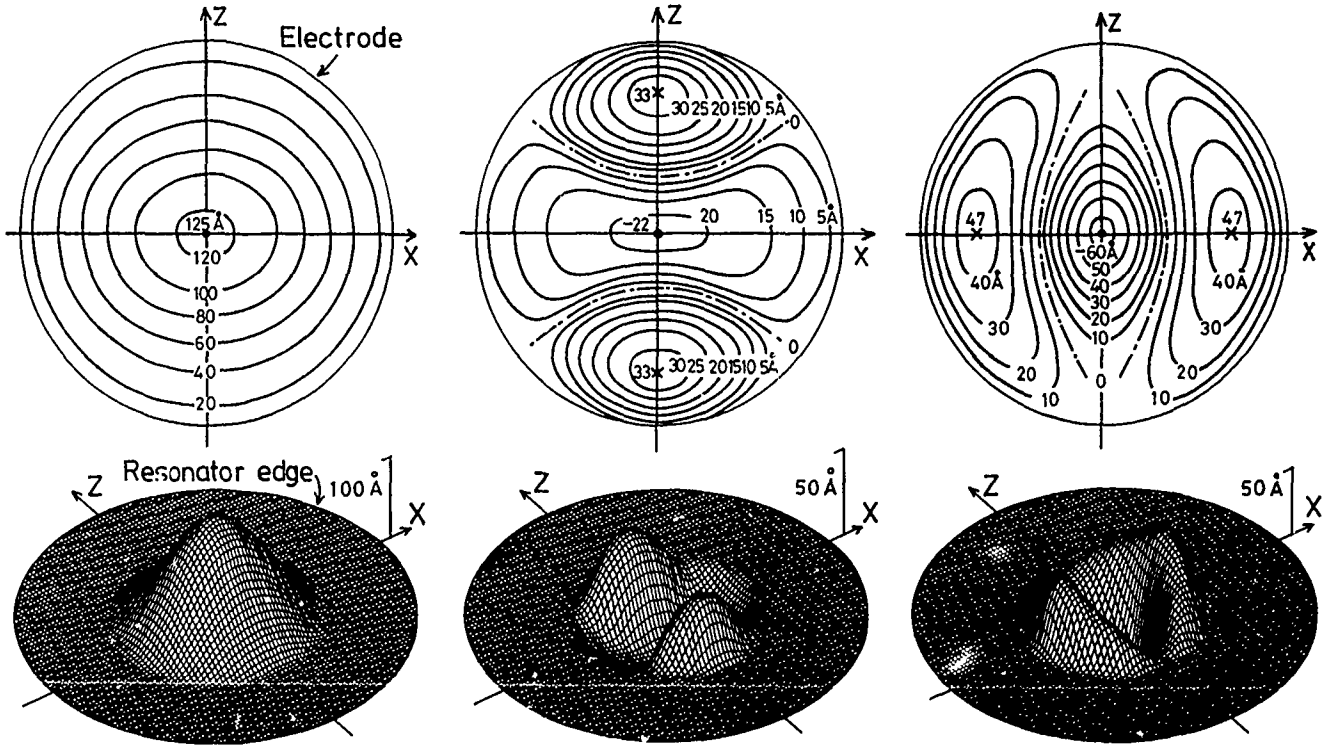


Fig.6. Vibration amplitude distributions of the fifth overtone main vibration.



(a) Main vibration
Res. freq. 53.7305 MHz
Exc. curr. 17.5 mA^{eff}

(b) Inharmonic vibration
Res. freq. 53.7605 MHz
Exc. curr. 3.6 mA^{eff}

(c) Inharmonic vibration
Res. freq. 53.7685 MHz
Exc. curr. 4.4 mA^{eff}
(U_x at the center ~63Å)

Fig.7. Fifth overtone vibration modes of AT-cut quartz crystal resonator.

laser beam along the x and z axes of the resonator, and the symmetry of vibration distributions was first confirmed. Then, in order to obtain the vibration amplitudes at a sufficient number of points to obtain a mode pattern, scanning measurements were made along several lines parallel to the x and z axes only over the first quadrant of the vibrator surface to reduce measuring time.

In the main mode shown in Fig.7(a), all points of the resonator surface are vibrating in phase over the entire surface. In contrast, in the inharmonic modes given in Figs.7(b) and 7(c) with two nodal lines shown by dot-dash lines, the phase of vibration is not the same over the entire resonator surface, and is reversed across the nodal lines. These mode patterns reveal in detail and quantitatively the nature of the vibration modes.

These examples demonstrate the effectiveness of the present method for measuring vibrations at VHF frequencies with extremely small amplitudes, which have not been easy to evaluate by other existing methods.

Conclusions

We have developed a method of measuring vibration modes in the VHF frequency range by optical fibre sensing which makes extensive use of recently developed optical fiber technology and a signal processing technique at high frequencies. The method has been demonstrated to have the capability of performing detailed and quantitative evaluation of vibration modes with extremely small amplitudes.

Acknowledgments

This work was partly supported by a Grant-in-Aid for Special Project Research on "Lightwave Sensing" from the Ministry of Education, Science and Culture, Japan.

References

(1) Y. Tsuzuki, Y. Hirose, and K. Iijima : "Measurement of vibration modes of piezoelectric resonators by means of holography", Proc. 25th Freq. Cont. Symp., p.113 (1971).

(2) K. Iijima, Y. Tsuzuki, Y. Hirose, and M. Akiyama:"Laser interferometric measurement of the vibration displacements of a plano-convex AT-cut quartz crystal resonator", Proc.IEEE, 64, 3, p.386 (1976).

(3) S. Ueha, K. Shiota, T. Okada, and J. Tsujiuchi:"Optical heterodyne measurement of in-plane vibrations", Opt. Commun., 10, 1, p.88 (1974).

(4) Y. Hirose, T. Abe, H. Komai, and Y. Tsuzuki: "Optical fibre vibration sensing system", IEICE Trans., E70, 3, p.195 (1987).

43rd Annual Symposium on Frequency Control - 1989

MASS LOADING MEASUREMENTS OF QUARTZ CRYSTAL PLATES

J. Kosinski*, S. Mallikarjun**, and A. Ballato*

*U.S. Army Electronics Technology and Devices Laboratory
LABCOM, Fort Monmouth, NJ 07703-5000

**Department of Physics, Monmouth College
West Long Branch, NJ 07764

ABSTRACT

Ongoing work aimed at determining the physical properties of cultured quartz material from frequency measurements of bulk wave resonators has highlighted the need for accurate means of measuring the mass loading due to the electrodes. In this paper, various methods of determining electrode mass loading are presented, and measurements of mass loading using several of these techniques are compared. Variations between the values of mass loading obtained by the different techniques are discussed and a model is proposed to explain the various differences.

INTRODUCTION

Measurements of selected quartz material properties are being made as part of a cooperative Army/Air Force/EIA/RPI effort. Preliminary results, based on measurements of TETM resonators, were reported in 1988 [1]. The accuracy of these results was noted as being limited by uncertainties in the values of electrode thicknesses ascribed to individual resonators. In order to resolve some of the uncertainties, a program was undertaken wherein a multiplicity of measurement methods and analysis techniques were used to determine the electrode thicknesses and resultant mass loading on a number of the samples in question. It was hoped that a satisfactory technique could be developed whereby the total electrode thickness (both sides) could be determined from a set of electrical measurements at the resonator terminals.

SIGNIFICANT EFFECTS DUE TO THE ELECTRODES

The effects associated with depositing thin metallic electrodes on the surface of a piezoelectric plate can be separated into three primary categories consisting of 1) inertial effects, 2) energy trapping effects, and 3) stress effects. If the electrodes are made sufficiently thick, wave propagation effects will also be present.

The primary inertial effect due to the electrodes is the frequency lowering commonly used as a means of frequency adjustment. To a very good approximation, the thin electrode films appear as lumped masses of mass per unit area m determined by the film density ρ_e and thickness t_e. While adding mass to the surfaces provides a convenient means of frequency adjustment, it is also a potential source of difficulty as the thermally activated transfer of contamination within a sealed resonator can easily be a significant source of frequency aging. A somewhat smaller inertial effect is an alteration of the resonator frequency-temperature behavior as the metallization thickness is varied [2,3].

Energy trapping effects are another useful result of depositing electrodes on the surface. By creating regions with different cut-off frequencies we are able to confine the vibrational energy of a desired thickness mode to the center of the plate, greatly reducing the coupling to unwanted modes which occurs at the lateral boundaries of the plate. Through careful choice of the amount of mass loading one can achieve suppression of anharmonic modes resulting in a resonator frequency spectrum containing only the desired mode.

The effects of intrinsic stresses in the electrode films and mechanical strains due to thermal expansion mismatch of the electrode-substrate interface are somewhat smaller than the inertial and energy trapping effects, and have not to date been characterized in great detail. There are, however, both theoretical [4] and experimental [5] studies indicating that the effects are sufficiently large as to warrant attention in the manufacture of precision resonators. Of particular interest are data by Filler [6] from which the change in the first order temperature coefficient of frequency can be calculated as a function of increasing mass loading for the case of gold electrodes on AT-cut quartz. After subtracting the inertial contribution noted previously, one finds that interfacial stresses due to the "low-stress" gold electrode contribute $+2\times10^{-7}$ per Kelvin per percent of mass loading to the first order temperature coefficient of frequency. This implies that compressive stresses are being developed in the quartz at the rate of $10^4 N/m^2$ per Kelvin per percent of mass loading, a value in good agreement with that calculable from the thermal expansions of gold and quartz.

ELECTRODE PROPERTIES OF INTEREST

There are a number of properties pertaining to the electrode which have an identifiable impact on resonator performance. These properties include electrical, physical, and geometric quantities.

The primary electrical characteristic of interest is the film resistivity. While it is possible to use resistivity as a tool in analyzing the physical properties of the film [7], we are generally interested only in obtaining a low resistance film so as not to degrade resonator Q.

For thin films, the physical property of most interest is the film density. The density of thin gold films as used in this paper has been reported by different authors as being between 90 percent [8] and 100 percent [9] of the bulk density value. As

the film becomes thicker, the elastic properties of the film become important and the acoustic impedance of the film becomes the physical property of most interest. The combinations of film adhesion and thermal expansion (leading to interfacial strains), and intrinsic stress and surface mobility (leading to metal migration) must also be considered in certain applications.

The geometric quantities of electrode size, shape, thickness, and uniformity directly affect both the static and motional equivalent circuit parameters of a given resonance. For a given cut and blank thickness, electrode size and shape determine both static and motional capacitances, electrode thickness determines the final resonant frequency, and electrode size and uniformity determine unwanted mode content.

PIEZOELECTRIC PLATE RESONATOR

The typical piezoelectric plate resonator consisting of a quartz disk with metallic electrodes top and bottom is shown in Figure 1. The piezoelectric plate of thickness 2h and density ρ_s contains some region A_a known as the active area wherein acoustic waves are driven by voltages applied to electrodes of thickness t_e (each side) and density ρ_e . The region wherein the electrodes overlap to form a parallel plate capacitor structure is denoted as the electroded area A_e.

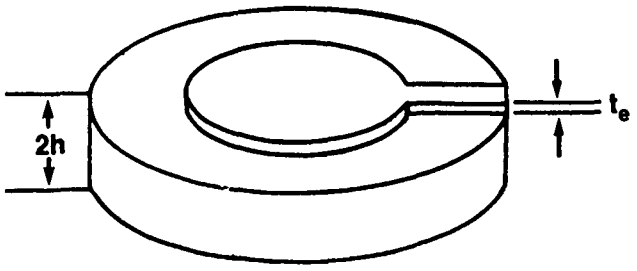


Figure 1. Piezoelectric Plate Resonator.

The inertial interaction of electrodes and substrate is conveniently described by the ratio of electrode mass per unit area to substrate mass per unit area, denoted as mass loading μ . For the case of an infinite plate with no lateral variations, the electroded and active areas are coincident and the mass loading is given by the familiar form

$$\mu = \frac{\rho_e \, t_e}{\rho_s \, h} \qquad (1)$$

wherein mass loading and electrode thickness are directly proportional. The frequency perturbation resulting from a given value of mass loading can be calculated using a number of methods; this will be discussed in a subsequent section.

METHODS OF MEASURING MASS LOADING

Given that the mass loading can be calculated using (1), a means of finding the mass loading using dimensional measurements becomes apparent:

- o assume bulk properties
- o measure the unelectroded critical frequencies
- o calculate 2h
- o measure t_e
- o calculate μ

The assumption of bulk properties for the electrode film is based upon the work of Lewis and Lu [9]. The calculation of blank thickness from the unelectroded critical frequencies is well known. From a practical standpoint, this method reduces to a) being able to measure the critical frequencies of unelectroded plates, and b) being able to measure the electrode film thicknesses.

An alternative method of determining mass loading arises from the ability to calculate the mass loading-induced frequency perturbation:

- o assume bulk properties
- o measure the unelectroded critical frequencies
- o calculate 2h
- o measure the electrode critical frequencies
- o calculate μ
- o calculate t_e

The assumption of bulk properties and the calculation of blank thickness are the same as for the dimensional measurement technique. In this case, however, we need to be able to measure the critical frequencies with and without the electrodes. In addition, the choice of equations used for calculating μ from the frequency perturbation is quite important. The assumption of lateral unboundedness is implicit in the application of (1) in both techniques. If the lateral dimensions are to be included in the treatment, the works of Tiersten [10] and Lee [11] are pertinent.

TABLE I Thickness Measurement Methods [12]

Method	Sensitivity	Remarks
Resistance	-1%	Convenient, empirical thickness-resistance relation required
Microbalance	-1Å/cm^2	Simple, rate and thickness monitor
Stylus	-20Å	Rapid, absolute, thickness and thickness profile
X-ray Emission	-100Å	Simple, relative, thickness and thickness profile
Optical	1%	Relative, rapid, continuous scan, thickness profile
Inter-ferometric	-2Å	Most accurate, absolute, highly reflecting surface or overcoat required in a two-beam method where sensitivity is limited to -50Å

FILM THICKNESS MEASUREMENTS

There are a number of methods available for measuring the thicknesses of metallic electrodes, each with its own combination of accuracy and applicability [12]. The most common techniques are described in Table 1. In this work, we were interested in checking the actual thicknesses of films reported to us as being 1000Å thick. Both interferometric and stylus-type measurements were evaluated, with stylus-type measurements being chosen on the basis of greater practicality in this particular case. Sample data obtained using the stylus method are shown in Figure 2.

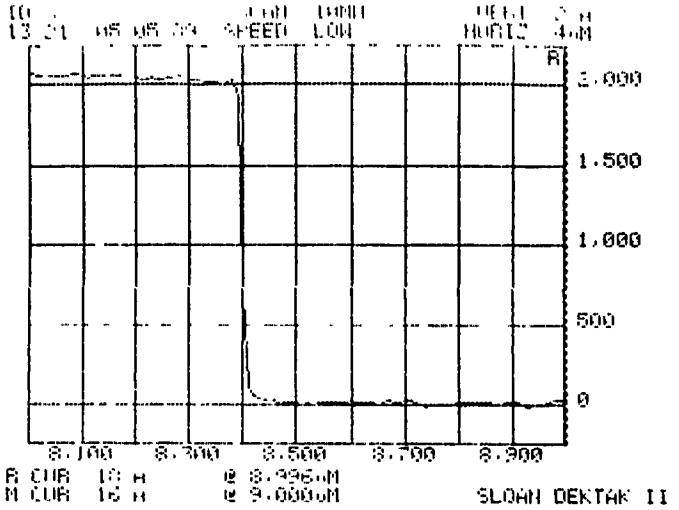


Figure 2. Sample thickness measurement data.

CRITICAL FREQUENCY MEASUREMENTS

The critical frequencies of the unelectroded piezoelectric plates were measured using an adjustable air gap fixture in conjunction with an automatic network analyzer. The resonance frequency was measured as the air gap was varied from zero to one thousand micrometers, and the data were fit by the method of least squares using the well known relationships for frequency pulling with a series load capacitor. Sample data are shown in Figure 3. Both resonance and antiresonance frequencies can be extracted from the data, which may then be converted to values of piezoelectric coupling and motional parameters.

The critical frequencies of the electroded piezoelectric plates were measured using a multiplicity of methods. Unmounted plates were measured using the same air gap fixture and analysis as described above, with the difference between resonator electrode diameter and air gap electrode diameter taken into account. Sample data are shown in Figure 4. The electroded plates were also measured while mounted and sealed in resonator enclosures. The frequency measurements were made using pi-network, impedance analyzer, network analyzer, and automatic microcircuit bridge systems. Good agreement was found among all the measurement methods.

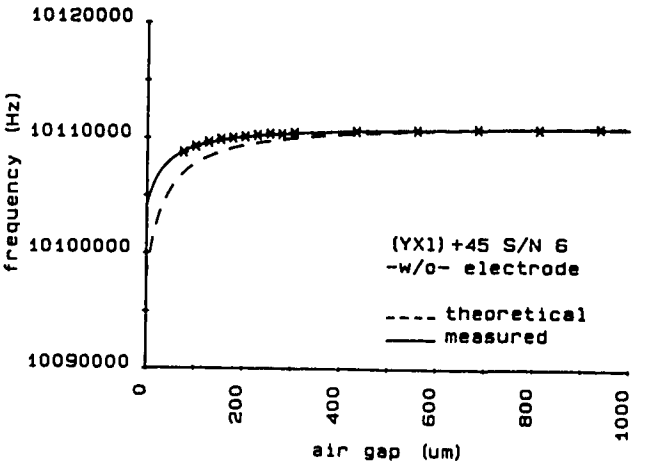


Figure 3. Sample unelectroded plate air-gap measurement data.

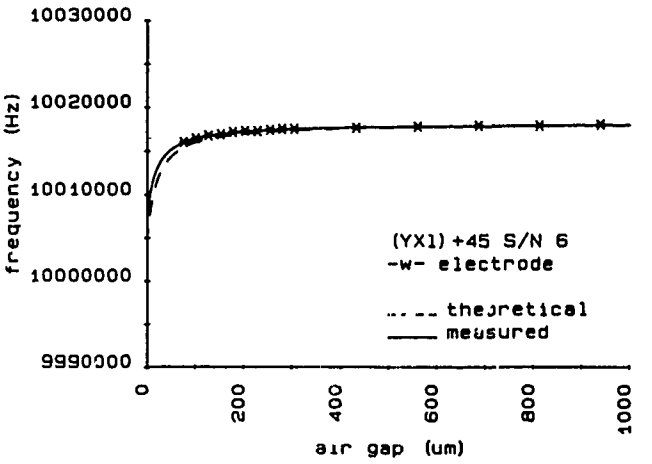


Figure 4. Sample electroded plate air-gap measurement data.

FREQUENCY MEASUREMENT ANALYSIS - MASS LOADING

A number of equation sets exist whereby the frequency perturbation due to the mass loading may be calculated with varying degrees of accuracy, with the greatest accuracy achieved through a three-dimensional equation set with non-linear piezoelectricity. For this discussion, however, we choose to employ the exact transmission line analog of a single thickness mode with no lateral variation [13], and compare these results to those obtained with two other relations commonly used to calculate mass loading.

The first investigator to formalize a relationship describing the mass loading-induced frequency perturbation was Sauerbrey [14]. The relationship, given by

$$\frac{m_e}{m_s} = - \frac{f^{(1)}_{Au} - f^{(1)}_{Ao}}{f^{(1)}_{Ao}} \qquad (2)$$

is considered accurate for values of mass loading up to two percent. An empirically determined constant representing film non-uniformity and lateral variations is sometimes included [15]. One should note that in (1) both m_e and m_s refer to actual masses, not areal densities.

As improvements in experimental techniques allowed thicker films to be evaluated, the relationship from Sauerbrey was modified slightly in order to extend its range of validity [16,17]. This modified relationship, known as the "period-measurement" technique, may be written as

$$\frac{m_e}{m_s} = - \frac{f^{(1)}_{A\mu} - f^{(1)}_{Ao}}{f^{(1)}_{A\mu}} \qquad (3)$$

With this small change, the relationship is considered accurate for values of mass loading up to ten percent for selected materials.

In 1968, Miller and Bolef [18] used acoustic wave analysis to calculate the shift in resonance frequencies which occurs when thin transducers are attached to an ultrasonic crystal resonator. In 1972, Lewis and Lu [19] applied the results of Miller and Bolef to the case of quartz crystal thickness monitors with large mass loading. They showed that, in the case of small losses, the frequency perturbation could be determined from

$$\frac{\rho_f v_f}{\rho_q v_q} \tan \frac{\pi f_c}{f_f} = - \tan \frac{\pi f_c}{f_q} \qquad (4)$$

where v_f and v_q are the shear wave velocities in the film and quartz respectively, and ρ_f and ρ_q are the respective densities. The quantities f_f, f_q, and f_c represent the resonance frequencies of the film, the quartz substrate, and the composite resonator structure.

In 1970, Yamada and Niizeki [20] used the equations of linear piezoelectricity in conjunction with the appropriate boundary conditions to obtain expressions for the admittance of piezoelectric plate resonators. In 1972, Ballato [13] used network synthesis techniques to transform the admittance functions of Yamada and Niizeki into exact transmission line analogs of the piezoelectric plate resonator. In the case of a single mode and small mass loading, the resonance frequencies may be determined from

$$\frac{\tan X}{X} = \frac{1}{k^2 + \mu X^2} \quad , \quad X = \frac{\pi}{2} \frac{f^{(M)}_{R\mu}}{f^{(1)}_{Ao}} \qquad (5)$$

and the antiresonance frequencies may be determined from

$$\frac{\tan X}{X} = \frac{1}{\mu X^2} \quad , \quad X = \frac{\pi}{2} \frac{f^{(M)}_{A\mu}}{f^{(1)}_{Ao}} \qquad (6)$$

Equations for the case of large mass loading are simply developed using the transmission line analog, from which (4) can easily be derived. Applying (6) to the antiresonance frequencies, one can determine the mass loading directly, whereas to use (5) in conjunction with the resonance frequencies, a knowledge of the piezoelectric coupling is required.

Inasmuch as the mass loading is by definition a fixed ratio of areal densities, one should in theory be able to use the antiresonance frequencies of a series of harmonics to determine both mass loading and the unloaded fundamental antiresonance frequency. In practice, however, the situation is complicated by the non-uniform distribution of vibratory motion, so that what we measure as antiresonance is actually the point where the integral of the current density taken over the electroded area vanishes.

FREQUENCY MEASUREMENT ANALYSIS - COUPLING

As an adjunct to determining the mass loading by measuring the critical frequencies, one may also determine the effective value of piezoelectric coupling both with and without mass loading. The simplest relationship for determining the coupling is derived from the Butterworth-Van Dyke equivalent circuit as

$$\frac{f^{(M)}_A - f^{(M)}_R}{f^{(M)}_R} = \frac{4\,k^2}{M^2\,\pi^2} \qquad (7)$$

This relationship can be used both with and without mass loading in which the coupling should be denoted by k_μ or k_o respectively, with similar subscripts on the frequencies.

The transmission line analog resonance frequency equation given in (5) may also be applied either with or without mass loading. In the case of zero mass loading, (5) simplifies to

$$\frac{\tan X}{X} = \frac{1}{k^2_o} \quad , \quad X = \frac{\pi}{2} \frac{f^{(M)}_{Ro}}{f^{(1)}_{Ao}} \qquad (8)$$

from which the coupling is obtained directly from the resonance frequencies, given that the fundamental antiresonance frequency is known. Inasmuch as the piezoelectric coupling for a given mode is a physical constant, one should in theory be able to use the unloaded resonance frequencies of a series of harmonics to determine both the coupling and the unloaded fundamental antiresonance frequency. If the mass loading is known, one can use (5) in conjunction with the mass loaded resonance frequencies to determine the coupling.

TABLE II Sample Data - (YXl)+38° Resonator

Mass Loading (%) and Electrode Thickness (Å)

M	measured $2t_e$	Sauerbrey μ	Sauerbrey $2t_e$	"period" μ	"period" $2t_e$	tanX/X μ	tanX/X $2t_e$
1	5015	1.825	4077	1.859	4153	1.860	4154
3	5015	1.914	4275	1.951	4358	1.964	4386
5	5015	1.921	4290	1.958	4374	1.981	4425
7	5015	1.916	4280	1.954	4364	1.994	4454
9	5015	1.905	4255	1.942	4338	2.003	4474

Piezoelectric Coupling (%)

M	theoretical k	BVD k_μ	BVD k_o	tanX/X k_μ	tanX/X k_o
1	7.77	3.93	2.94	3.89	2.94
3	7.77	3.43	3.18	3.37	5.10
5	7.77	3.84	1.56	4.12	8.13
7	7.77	4.54	4.68	4.86	12.17
9	7.77	6.30	5.88	7.02	16.38

TABLE III Sample Data - (YXl)-49° Resonator

Mass Loading (%) and Electrode Thickness (Å)

	measured	Sauerbrey		"period"		tanX/X	
M	$2t_e$	μ	$2t_e$	μ	$2t_e$	μ	$2t_e$
1	3500	1.305	2325	1.322	2355	1.322	2349
3	3500	1.311	2336	1.329	2367	1.325	2354
5	3500	1.308	2329	1.325	2360	1.318	2344
7	3500	1.316	2344	1.333	2376	1.346	2418

Piezoelectric Coupling (%)

	theoretical	BVD		tanX/X	
M	k	k_μ	k_o	k_μ	k_o
1	5.66	3.45	3.55	3.43	3.55
3	5.66	3.22	3.64	3.20	1.52
5	5.66	3.22	2.92	3.21	--
7	5.66	3.30	4.69	3.31	8.02

EXPERIMENTAL RESULTS

Frequency measurements were made on a number of the rotated Y-cut resonators described last year [1]. In addition, both frequency and film thickness measurements were performed on a number of optically polished (YXl)+38° and (YXl)-49° resonators. Sample data typical of the optically polished resonators are given in Tables II and III. The measured total electrode thickness $2t_e$ is compared to that obtained by combining (1) with (2), (3), and (6). The theoretical piezoelectric coupling is compared to that obtained by applying (7) to the data both with and without mass loading, by applying (5) to the mass loaded resonance frequency data using the values for mass loading obtained using (6), and by applying (8) to the unloaded resonance frequencies. A number of points concerning the mass loading are evident from the data:

1) While the mass loading is by definition a physical constant, the actual effective mass loading varies with harmonic.

2) The harmonic variations observed were consistent for a given angle of cut, but varied between different cut angles.

3) A difference between the bulk and thin film densities cannot by itself explain the mass loading data.

In order to explain the harmonic variations, it is necessary to examine the transmission line analog being used.

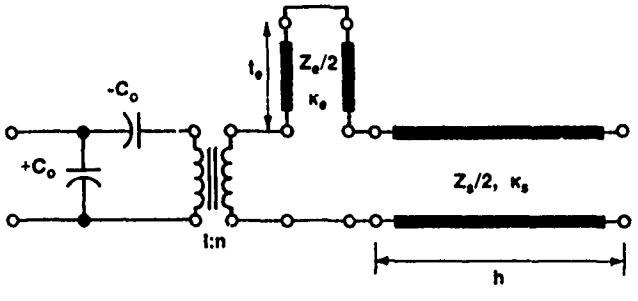


Figure 5. Transmission line analog for a single TETM mode, bisected basis.

FIRST APPROXIMATION TO FINITE ELECTRODES

The transmission line analog for a single piezoelectrically driven mode, in the case of thick electrode films, is shown in Figure 5. Implicit in the calculation of the equivalent circuit elements are the assumptions of lateral unboundedness and a uniform distribution of motion. A simple first approximation to the real world case of lateral boundedness can be made by retaining the assumption of a uniform distribution of motion but allowing for a difference between the electroded area A_e and the active area A_a. The equivalent circuit elements are then calculated as

$$C_o = \frac{\epsilon A_e}{2h} \tag{9}$$

$$N = \frac{e A_a}{2h} \tag{10}$$

$$Z = A \rho v \tag{11}$$

$$k = \frac{\omega}{v} \tag{12}$$

It is also expedient to introduce the concept of the electroded fraction of the active area, here denoted b. The natural upper bound of b = 1 is approached in the work of Lewis and Lu [9]. Since b is always less than or equal to 1, b is conveniently expressed as a percentage. For the common case of equal size electrodes as encountered here, $b = A_e / A_a$.

EFFECTIVE COUPLING AND MASS LOADING

If one employs (9) thru (12) in finding the critical frequencies of the transmission line analog of Figure 5, one obtains equations (5) and (6) with the difference that the piezoelectric coupling k and mass loading μ are now replaced by effective values k_{eff} and μ_{eff}. The effective coupling is given by

$$k^2_{eff} = \frac{k^2}{b} \tag{13}$$

and the effective mass loading is given by

$$\mu_{eff} = \mu \, b \, \frac{\tan X_e}{X_e} \tag{14}$$

The effective mass loading is seen to be the product of the "ideal case" mass loading for thin films given by (1), the electroded fraction of the active area b, and a transcendental term relating to wave propagation in the case of thick films. The term X_e is defined similarly to the X of equations (5) and (6), being equal to one-half pi times the ratio of the critical frequency being measured to the fundamental antiresonance frequency of the total electrode film of thickness $2t_e$. One can readily solve for the electrode thickness required for the "onset" of wave propagation effects by setting the term $(\tan X_e/X_e)$ equal to the sum of unity plus some maximum allowable perturbation level.

TABLE IV Electroded Fraction of Active Area

M	(YXl)+38°	(YXl)-49°
1	82.8%	67.1%
3	87.5%	67.3%
5	88.2%	67.0%
7	88.8%	69.1%
9	89.2%	---

ANALOGY TO BECHMANN'S NUMBER

Table IV lists typical data for the areal ratio b observed as a function of angle of cut and harmonic of operation. For comparison, the harmonic dependence of the energy trapping criterion known as Bechmann's Number is shown in Figures 6 and 7 for the case of AT and BT-cuts respectively [21]. The trend toward better energy trapping with increasing harmonic number for the AT-cut case compares nicely to the increasing electroded fraction of the active area with increasing harmonic number as shown by the data. For the BT-cut, the more complex energy trapping-harmonic relationship is also well mirrored, as is the overall fact that the AT-cut energy traps much better than the BT-cut.

CONCLUSIONS

Measurements have been made of the electrode thicknesses and resultant mass loadings on a number of piezoelectric plate resonators. Significant differences have been found between the theoretical and effective mass loading, including variations in the effective mass loading measured for different harmonics. A simple first approximation to the effects of lateral boundedness which provides qualitative insights as to the cause of the variations has been presented.

ACKNOWLEDGMENTS

The authors wish to thank Mrs. Patricia Fuschetto who fabricated a number of the resonators for this project. We wish also to thank Dr. Erich Hafner of Xotex Corp. who made available to the authors the automatic microcircuit bridge system with which a large number of measurements were made.

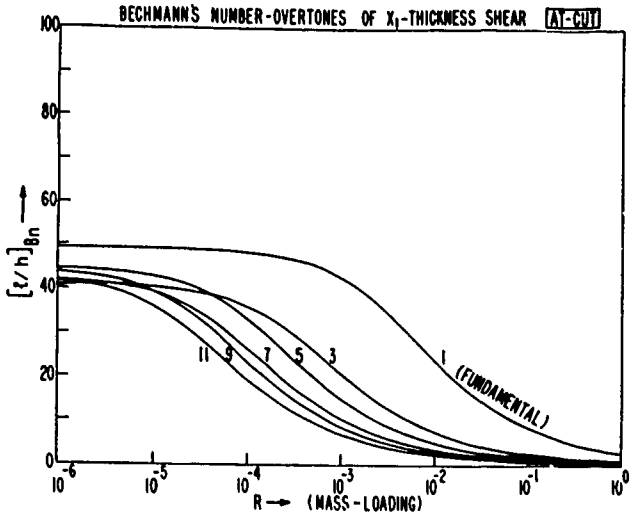


Figure 6. Bechmann's Number for the AT-cut.

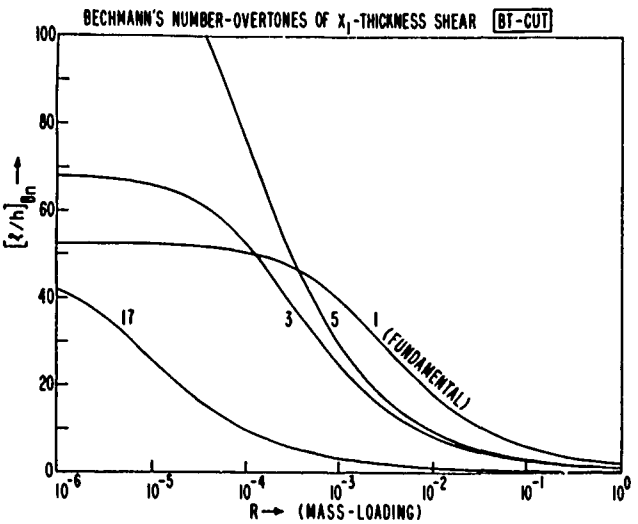


Figure 7. Bechmann's Number for the BT-cut.

REFERENCES

[1] J. Kosinski, A. Ballato, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Temperature Derivatives of the Dynamic Permittivity and Permeability of the Simple Thickness Modes of Quartz Plates," Proc. 42nd Annual Frequency Control Symposium, pp. 53-64, June 1988.

[2] A. Ballato and T. Lukaszek, "Higher Order Temperature Coefficients of Frequency of Mass Loaded Piezoelectric Crystal Plates," Proc. 29th Annual Frequency Control Symposium, pp. 10-25, May 1978.

[3] A. Ballato, T. J. Lukaszek, and G. J. Iafrate, "Subtle Effects in High Stability Vibrators," Proc. 34th Annual Frequency Control Symposium, pp. 431-444, May 1980.

[4] E. P. EerNisse, "Quartz Resonator Frequency Shifts Arising From Electrode Stress," Proc. 29th Annual Frequency Control Symposium, pp. 1-4, May 1975.

[5] J. H. Sherman, "Temperature Coefficient of the Frequency Shift Arising from Electrode Film Stress," IEEE Trans. on Sonics and Ultrasonics, Vol. 30, No. 2, pp. 104-110, March 1983.

[6] R. L. Filler and J. R. Vig, "Fundamental Mode SC-cut Resonators," Proc. 34th Annual Frequency Control Symposium, pp. 187-193, May 1980.

[7] L. Maissel and R. Glang, eds. Handbook of Thin Film Technology. New York:McGraw-Hill, 1970.

[8] E. P. EerNisse, private communication.

[9] O. Lewis and C. Lu, "Relationship of Resonant Frequency of Quartz Crystal to Mass Loading," Proc. 29th Annual Frequency Control Symposium, pp. 5-9, May 1975.

[10] D. S. Stevens and H. F. Tiersten, "An Analysis of Doubly-Rotated Contoured Quartz Crystal Resonators," Proc. 39th Annual Frequency Control Symposium, pp. 436-447, May 1985.

[11] 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," Proc. 40th Annual Frequency Control Symposium, pp. 152-160, May 1986.

[12] K. L. Chopra, Thin Film Phenomena. New York:McGraw-Hill, 1969, pp. 83-136.

[13] A. D. Ballato, "Transmission Line Analogs for Stacked Piezoelectric Crystal Devices," Proc. 26th Annual Frequency Control Symposium, pp. 86-91, June 1972.

[14] G. Z. Sauerbrey, "Verwendung von Schwingungen zur Wägung dünner Schichten und zur Mikrowägung," Zeitschrift der Physik, Vol. 155, p. 206, 1959.

[15] C. D. Stockbridge, "Resonance Frequency Versus Mass Added to Quartz Crystals," in Vacuum Microbalance Techniques, Vol. 5 (K. Behrdt, ed.). New York:Plenum Press, 1966, pp. 193-205.

[16] K. H. Behrndt, "Long-Term Operation of Crystal Oscillators in Thin-Film Deposition," Journal of Vacuum Science and Technology, Vol. 8, No. 5, pp. 622-626, September 1971.

[17] D. R. Denison, "Linearity of a Heavily Loaded Quartz Crystal Microbalance," Journal of Vacuum Science and Technology, Vol. 10, No. 1, pp. 126-129, January 1973.

[18] J. G. Miller and D. I. Bolef, "Sensitivity Enhancement by the Use of Acoustic Resonators in cw Ultrasonic Spectroscopy," Journal of Applied Physics, Vol. 39, No. 10, pp. 4589-4593, September 1968.

[19] C. Lu and O. Lewis, "Investigation of Film-Thickness Determination by Oscillating Quartz Resonators With Large Mass Load," Journal of Applied Physics, Vol. 43, No. 11, pp. 4385-4390, November 1972.

[20] T. Yamada and N. Niizeki, "Admittance of Piezoelectric Plates Under Perpendicular Field Excitation," Proc. of the IEEE, Vol. 58, No. 6, pp. 941-942, June 1970.

[21] H. F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Thickness-Shear," Proc. 28th Annual Frequency Control Symposium, pp. 44-48, May 1974.

43rd Annual Symposium on Frequency Control - 1989

QUARTZ CRYSTAL RESONATOR MODEL MEASUREMENT AND SENSITIVITY ANALYSIS

Karen Park, Donald Malocha, and Madjid Belkerdid

Solid State Devices and Systems Laboratory
Electrical Engineering & Communication Sciences Department
University of Central Florida
Orlando, Florida 32816

Abstract: A quartz crystal resonator can be modelled as a one or two port electrical circuit. Extraction of the equivalent electrical parameters can be accomplished using scattering parameters measured near the resonant frequency of the crystal.

Due to measurement errors inherent in the network analyzer, the extracted parameters will vary from run to run. These measurement inaccuracies lead to greater errors in the electrical parameters as the magnitude of S_{11} (one port model) or S_{12} (two port model) approaches one. This paper will present results of measured and simulated data and a first order sensitivity theory. An expression for the error in the extracted series resistance will be derived. Simulated and actual crystal data will be compared to the theoretical values for the one port model.

Introduction

The importance of quartz crystal resonators stems from their high quality factor, relatively small size, and excellent temperature stability. The use of these devices for frequency control, filter, and timing applications requires an accurate determination of their equivalent electrical parameters.

A quartz crystal resonator near the resonance of an isolated mode of vibration can be represented by the one port model shown in Figure 1 or the two port model shown in Figure 2. [1]

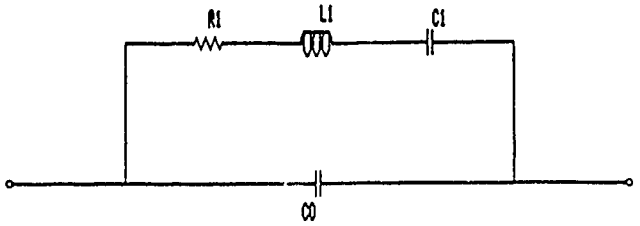


Figure 1. Single Mode One Port Resonator Equivalent Circuit Model.

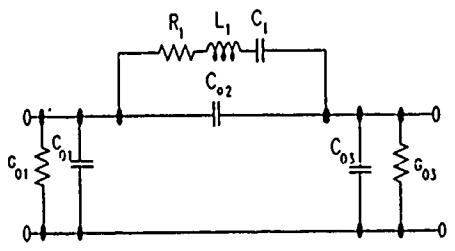


Figure 2. Single Mode Two Port Resonator Equivalent Circuit Model.

** This work was supported, in part, by the Electronic Industries Association, Washington, D.C.

CH2690-6/89/0000-372 $1.00 C 1989 IEEE

There are several means of measuring crystal resonators and extracting the electrical parameters. One in particular, proposed by the Electronics Industries Association (EIA), uses scattering paramters measured around the resonant frequency. [2,3] The extracted parameters depend on the values measured. If measurements could be repeated with exact precision, the extracted parameters would not change from data run to data run. But due to measurement inaccuracies, there are variations in the S-parameters and therefore variations in the electrical parameters. [4] These varitions are assumed to be random and can be due to several sources which include network analyzer noise, frequency inaccuracies, and temperature variations. Tables 1 and 2 show the effects of multiple data acquisition runs on the extracted resonator parameters for two crystals analyzed as one and two port devices.

	R1(OHM)	L1(mH)	C1(nF)	C0(pF)	fs(MHz)
	33.67	204.721	9.661	2.103	3.57872
	34.093	203.321	9.728	1.872	3.57872
	33.408	203.658	9.711	1.823	3.578719
	33.667	204.243	9.684	2.424	3.578719
	33.814	203.825	9.703	2.164	3.578719
	33.472	203.864	9.702	2.101	3.57872
	33.663	202.323	9.776	5.318	3.578718
	33.696	203.765	9.706	3.818	3.578718
	33.642	203.417	9.723	4.193	3.578718
	33.88	202.967	9.744	2.592	3.578718
Average	33.7005	203.6104	9.7138	2.8408	3.578719
Maximum	34.093	204.721	9.776	5.318	3.57872
Minimum	33.408	202.323	9.661	1.823	3.578716
Std. Dev.	.1854261	.6294971	.0301125	1.126042	.0000006
% Std. Dev.	.5502177	.3091675	.3099967	39.63821	.0000181

(a) One Port

	R1(ohms)	L1(mH)	C1(fF)	C02(pF)	C01(pF)	C03(pF)
	37.152	206.116	9.596	2.293	.772	.561
	37.114	206.152	9.594	2.141	.372	.592
	37.1	206.159	9.594	2.101	.351	.595
	37.127	206.074	9.598	2.015	.251	.676
	37.108	206.078	9.597	2.013	.236	.694
	37.104	206.079	9.597	1.925	.168	.691
	37.128	206.106	9.596	1.842	.176	.819
	37.098	206.109	9.596	2.004	.243	.675
	37.103	206.106	9.596	1.808	.086	.721
	37.127	205.914	9.605	1.662	.086	.854
Average	37.1161	206.0893	9.5969	1.9804	.2741	.6878
Maximum	37.152	206.159	9.605	2.293	.772	.854
Minimum	37.152	206.159	9.605	2.293	.772	.854
Std. Dev.	.016257	.06453379	.00294788	.17152154	.18920859	.08953078
% Std. Dev.	.0438004	.03131351	.03071701	8.6609545	69.029036	13.016978

(b) Two Port

Table 1. Extracted Parameters for Device J69 Analyzed as a (a) One Port Device (b) Two Port Device

	R1(OHM)	L1(mH)	C1(fF)	C0(pF)	fs(MHz)
	1696.499	8168.351	3.101	1.043	.9999639
	1701.096	8196.781	3.09	1.048	.9997...
	1709.253	8190.552	3.093	1.06	.999964
	1713.775	8217.336	3.083	.99	.9999635
	1721.039	8153.145	3.107	1.038	.9999638
	1720.972	8179.01	3.097	.996	.9999643
	1720.688	8188.885	3.093	1.059	.9999638
	1725.281	8183.554	3.095	.975	.9999634
	1729.366	8150.726	3.108	.946	.9999643
	1725.963	8162.368	3.104	.914	.9999641
Average	1716.3932	8179.0708	3.0971	1.0069	.99996391
Maximum	1729.366	8217.336	3.108	1.06	.9999643
Minimum	1696.499	8150.726	3.083	.914	.9999634
Std. Dev.	10.428244	19.737245	.00752928	.04822333	.00000028
% Std. Dev.	.60756731	.24131403	.2431073	4.7892874	.00002844

(a) One Port

R1(OHM)	L1(mH)	C1(fF)	C02(pF)	C01(pF)	C03(pF)
1655.017	9030.96	2.805	.91	.065	-13.985
1649.075	9024.523	2.807	.987	.071	-14.188
1652.011	9025.034	2.807	.971	-.407	-13.871
1647.969	9026.511	2.806	.999	-.462	-14.002
Average	1651.018	9026.757	2.80625	.96675	-.18325	-14.0115
Maximum	1655.017	9030.96	2.807	.999	.071	-13.871
Minimum	1647.969	9024.523	2.805	.91	-.462	-14.188
Std. Dev.	2.7408803	2.5340289	.00082916	.03423722	.25201029	.11367168
% Std. Dev.	.16601154	.02807242	.02954677	3.5414765	-137.5227	-.8112741

(b) Two Port

Table 2. Extracted Parameters for Device XX1 Analyzed as a (a) One Port Device (b) Two Port Device

As seen in Tables 1 and 2, the extracted parameters vary from run to run. The variations in C_0 are due to the extration technique and were unimportant to this study. Normally C_0 is measured off resonance to much greater accuracy. The variations in the motional arm parameters, R_1, C_1, and L_1, can be attributed to the inaccuracies of the measurement system. The data taken was not averaged; including the calibration data. Therefore the statistical variations in the extracted parameters also include the variations in calibration data. Normally data averaging should be implemented to improve the accuracy of the extraction technique, but this provides a base line for data and represents a "worst case" approach.

Extraction Technique

For the one port model, the admittance can be written as

$$Y = jB_0 + \frac{R_1 - jX_1}{R_1^2 + X_1^2} \tag{1}$$

where $B_0 = \omega C_0$. Also, the normalized impedance is given by

$$Y = \frac{1 - S_{11}}{1 + S_{11}} \tag{2}$$

Therefore, using values of S_{11} measured near resonance, the model parameters can be extracted using curve fitting methods.

Defining the motional arm admittance as Y_c, the admittance matrix for the two port model can be written as

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{21} \\ Y_{12} & Y_{22} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \tag{3}$$

where

$$Y_{11} = G_{01} + j\omega C_{01} + Y_c + j\omega C_{02} \tag{4}$$

$$Y_{12} = Y_{21} = -Y_c - j\omega C_{02} \tag{5}$$

$$Y_{22} = G_{03} + j\omega C_{03} + Y_c + j\omega C_{02} \tag{6}$$

Here the motional arm terms and the shunt capacitance, C_{02}, are isolated in the Y_{12} term. The parasitic elements can be eliminated from Y_{12} using simple manipulations. Therefore the measured values of S_{12} can be used to extract the equivalent electrical parameters using the same technique implemented for the one port model. This paper will examine the effects of measurment errors in S_{11} on the extraction of R_1 for the one port model and the results will be extended to the two port model.

Sensitivity Analysis

One Port

To derive an expression for the error in R_1, evaluate equation (1) at $\omega \approx \omega_s$ (i.e, $X_1 \approx 0$). This yields

$$Y = jB_0 + \frac{1}{R_1} \tag{7}$$

letting $\beta = R_1 B_0$,

$$Y = \frac{j\beta + 1}{R_1} \tag{8}$$

Therefore,

$$R_1 = \frac{j\beta + 1}{Y} \tag{9}$$

Take the derivative of R_1 with respect to S_{11} using the chain rule, which yields

$$\frac{\delta R_1}{\delta S_{11}} = \frac{\delta R_1}{\delta Y} \frac{\delta Y}{\delta S_{11}} \tag{10}$$

From equation (9)

$$\frac{\delta R_1}{\delta Y} = -\frac{j\beta + 1}{Y^2} \tag{11}$$

From equation (2)

$$\frac{\delta Y}{\delta S_{11}} = -\frac{2}{(1 + S_{11})^2} \tag{12}$$

Substituting equations (11) and (12) into (10) gives

$$\frac{\delta R_1}{\delta S_{11}} = \frac{2(j\beta + 1)}{Y^2 (1 + S_{11})^2} \tag{13}$$

Multiplying both sides by δS_{11} and dividing by R_1 yields

$$\frac{\delta R_1}{R_1} = \frac{2(j\beta + 1)}{Y^2(1 + S_{11})^2} \frac{\delta S_{11}}{R_1} \tag{14}$$

Substituting equation (9) for R_1 on the right hand side and using equation (2) for Y yields

$$\frac{\delta R_1}{R_1} = \frac{2}{1 - S_{11}^2} \delta S_{11} \tag{15}$$

Equation (15) is the fractional change in R_1 as a function of S_{11} and the change in S_{11}. In order to develop a theoretical model for the measured error, equation (15) will be extended assuming statistical variations in the extracted values of R_1. Begin with the ith extracted value of R_1 and S_{11} which can be expressed as

$$R_{1i} = \bar{R}_1 + \delta R_1 \tag{16}$$

and

$$S_{11i} = \bar{S}_{11} + \delta S_{11} \tag{17}$$

or

$$\delta R_1 = R_{1i} - \bar{R}_1 \tag{18}$$

$$\delta S_{11} = S_{11i} - \bar{S}_{11} \tag{19}$$

where $\bar{R}_1$ and $\bar{S}_{11}$ are the average values of the resistance and scattering parameter, respectively. Substituting (18) and (19) back into equation (15) yields

$$\frac{R_{1i} - \bar{R}_1}{R_{1i}} = \frac{2}{1 - S_{11i}^2} (S_{11i} - \bar{S}_{11}) \tag{20}$$

Squaring both sides and taking the variance gives

$$\frac{\sigma_{R_1}^2}{\bar{R}_1^2} = \left(\frac{2}{1 - \bar{S}_{11}^2}\right)^2 \sigma_{S_{11}}^2 \tag{21}$$

where σ^2 is the variance. Taking the square root of both sides yields

$$\frac{\sigma_{R_1}}{\bar{R}_1} = \frac{2}{1 - \bar{S}_{11}^2} \sigma_{S_{11}} \tag{22}$$

where σ is the standard deviation.

Equation (22) represents the fractional standard deviation of R_1 with respect to the one port scattering parameter and the standard deviation of the measured value of S_{11}.

These results indicate that the degree of accuracy of the extracted model parameters are a function of S_{11} and the accuracy degrades as S_{11} approaches unity.

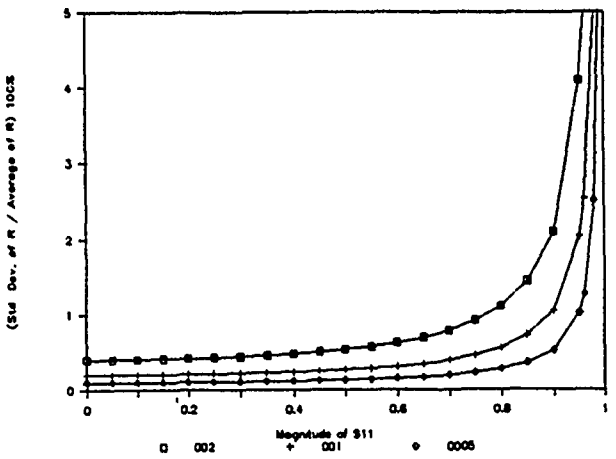


Figure 3. Standard Deviation of R_1 Divided by the Average of R_1 vs the Magnitude of S_{11} for $\sigma_{S_{11}}$ = .0005, .001, and .002.

Simulated Data

To confirm that equation (22) predicts the inaccuracy in the extracted value of R_1, sets of S_{11} values were simulated and given some random error. A program, which used input values of R_1, L_1, C_1, and C_0, generated the admittance of the one port model over a specified frequency range around resonant frequency. These admittance values where then converted to S_{11} values. Next a random number generator was used to simulate the error of the network analyzer. Error was added to each S_{11} value. The random number generator used had a guassian distribution with zero mean and a standard deviation input by the user. The distribution is is plotted in Figure 4, superimposed on a guassian with a standard deviation of 0.01.

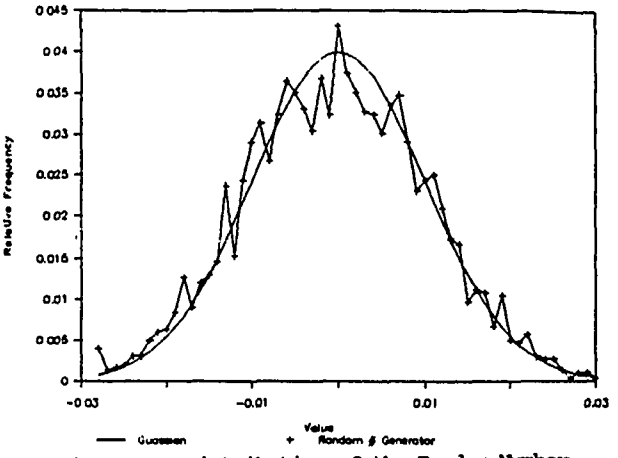


Figure 4. Distribution of the Random Number Generator with σ = .01.

Ten files of S_{11} values were created having a range of S_{11} at resonance from .00125 to .95122. The inputs and the value of S_{11} at resonance are shown in Table 3. Each set of S_{11} values was randomized twenty five times. The electrical parameters were extracted and a spread sheet program was used to calculate the average value and standard deviation of R_1 for each set of S_{11} values.

Figures 5, 6, and 7 show the simulated data and the theoretical data for three different standard deviations. There is close agreement between the simulation and the theory in each case.

R_1 (ohms)	L_1 (mh)	C_1 (pf)	C_0 (pf)	S_{11}*
50	50	.0025	5	.00125
75	145	.0032	5	.20015
100	300	.0037	5	.33338
125	450	.0035	5	.42857
150	850	.0041	5	.50001
200	1000	.0031	5	.60000
300	1500	.0021	5	.71429
500	2600	.0013	5	.81819
1000	3600	.00043	5	.90477
2000	8250	.00025	5	.95122

* at resonance

Table 3. Inputs to Simulation Program and S_{11} at Resonance

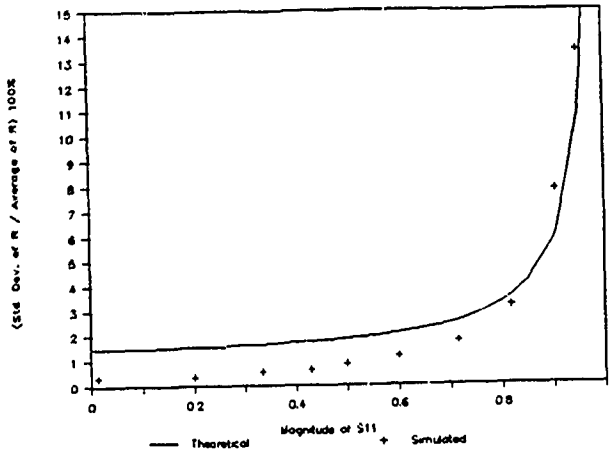


Figure 5. Comparison of Simulated Data to Theoretical Data for $\sigma_{S_{11}} = .002$

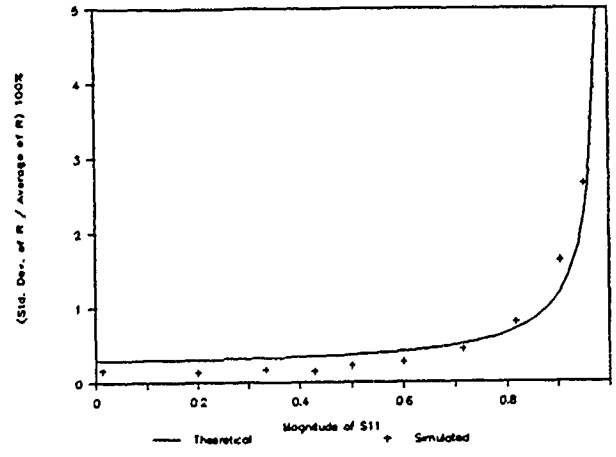


Figure 6. Comparison of Simulated Data to Theoretical Data for $\sigma_{S_{11}} = .001$

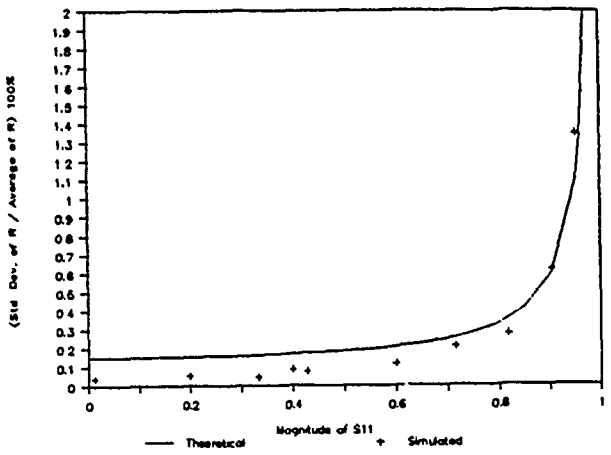


Figure 7. Comparison of Simulated Data to Theoretical Data for $\sigma_{S_{11}} = .0005$

Crystal Data

Data taken from measured crystals was also compared to equation (22). As examples, three resonators were measured twenty five times each and the electrical parameters were extracted. Measurement of the crystal was done with an HP 3577 network analyzer controlled remotely with a PC using a data aquisition program. Each crystal was measured at sixteen frequencies in the 6 dB bandwidth around center frequency using a resolution bandwidth of 10 Hz. Only S_{11} was measured since the one port model error is being analyzed. Calibration data was taken for a short circuit, an open circuit, and a 50 ohm load. [6] No data averaging was implemented on any of the calibration sample points or the measured sample points. All of the data was stored on the PC hard disk in directories corresponding to the device name and the resolution bandwidth.

The measurements were taken with the network analyzer uncalibrated. A three term S-parameter error correction model was implemented. Once the S_{11} file was corrected, the equivalent electrical parameters were extracted. The program reads in the corrected S_{11} values and outputs the electrical parameters, along with some statistical data, to a user specified file.

A spread sheet program was used to calculate standard deviations and average values of R_1 and S_{11}. The crystals, the average value of S_{11} and the standard deviation of S_{11} at resonance, and the fractional standard deviation of R_1 calculated from both equation (23) and the measured data are shown in Table 4.

Device Name	$\bar{S}_{11}$*	$\sigma_{S_{11}}$	$(\sigma_{R_1}/\bar{R}_1)$ x 100 Measured	Theory
J69	.157	.00096	.4218	.2000
CC4	.182	.00044	.0415	.0904
XX1	.947	.00013	.4763	.2438

* at resonance

Table 4. Crystal Data

For all three cyrstals, there is a strong correlation between the measured and the theoretical fractional standard deviation of R_1.

Two Port Model

The extraction technique for the two port model is similar that for the one port model except that a modified S_{12} is used rather that S_{11}. Therefore, the derivation of the error equation for R_1 follows that for the one port model, with S_{12} as the scattering parameter of interest. This yields an equation for the fractional standard deviation in R_1 as

$$\frac{\sigma_{R_1}}{\bar{R}_1} = \frac{2}{1 - \bar{S}_{12}^2}\,\sigma_{S_{12}} \qquad (23)$$

Conclusions

For the one port model, the derived relationship for the fractional standard deviation of R_1 as a function of S_{11} shows that as the magnitude of S_{11} approaches one, the inaccuracies in the extracted parameter also increases. One consequence of this is that crystals with a high or low series resistance at resonance (which gives a magnitude of S_{11} closer to one) will have larger errors in the extracted parameters. In these cases, multiple data and extraction runs should be taken and the results averaged and processed.

The comparison of the derived expression for the fractional standard deviation of R_1 to actual crystal data shows strong correlation and the comparison of the derived equation to simulated data also shows close agreement. Only twenty five data runs were used for the calculation of the statistical data and this number should be increased to give better accuracy.

The fractional standard deviation of R_1 as a function of S_{12} for the two port model follows that for the one port model and the same conclusions can be drawn.

This study has shown the results of a first order sensitivity model and further work is warranted to extend these results for the broad range of crystal resonators and measurement techniques.

Acknowledgements

The authors would like to thank Mike Fletcher who wrote the data aquisition program and Huat Ng who wrote the error correction and parameter extraction programs.

References

[1] E. Hafner, "Piezoelectric Crystal Unit Definitions and Methods of Measurement," in Proceedings of the IEEE, Vol. 57, No. 2, February 1969, pp. 179-201.

[2] D. C. Malocha, H. Ng, and M. Fletcher, "Quartz Resonator Model Measurement and Sensitivity Study," 1988 EIA Quartz Crystal Conference Proceedings, Vol 2, pp. 31-39.

[3] K. Anderson, "Measurements of High Frequency Crystals Using the EIA-512 Measurement Standard," in 1988 EIA Quartz Devices Conference Proceedings, Vol. 2, August 1988, pp. 40-43.

[4] W. Smith, "Electronic Industries Association Standard 512: Some Further Discussion and Comments." correspondence.

[5] W. Smith, "Final Report EIA-P11 Round Robin Crystal Measurement Experiment," in EIA Quartz Devices Conference Proceedings, August 1988, pp. 51-90.

[6] D. Rytting, "An Analysis of Vector Measurement Enhancement Techniques," HP Application Note.

43rd Annual Symposium on Frequency Control - 1989

VARIATIONAL METHOD FOR MODELING STATIC AND DYNAMIC STRESSES IN A RESONATOR DISC WITH MOUNTS

Errol P. EerNisse
Larry D. Clayton
Milton H. Watts

Quartztronics, Inc.
1020 Atherton Drive
Building C
Salt Lake City, UT 84123

SUMMARY

A number of important quartz resonator performance characteristics depend on the stress patterns set up in the resonator disc as it reacts against its mounting structure during static or dynamic acceleration. A desk-top computer program has been developed that can handle most practical stress problems including full anisotropy such as encountered in the SC-cut. The model is a hybrid combination of variational calculus and lumped elastic models. The continuum mechanics problem of the stress pattern in the resonator disc (plano-plano) is treated variationally with a three-dimensional polynomial trial function set for each of the three displacements. Body forces such as acceleration force are included. Boundary conditions are lumped. Forces and moments can be applied anywhere on the surface of the disc. The mounting structure is handled as a lumped model spring for each mount. Solutions can be obtained for static acceleration and for modal analysis of the combined structure.

The stress pattern due to acceleration in a 10 MHz, 3rd overtone, SC-cut with four mounts will be shown and compared to the stress patterns for an isotropic solution of the same structure. The contour plots of the various stress tensor components show how the disc and mounts interact to support the disc and what stress patterns exist in the active region of resonant vibration.

INTRODUCTION

Stress patterns occur in a thickness shear mode quartz resonator as it reacts against its mounting structure during static or dynamic acceleration. The resonant frequency shifts caused by this stress create frequency shifts or phase noise in crystal oscillator outputs.[1,2] Acoustic noise excites the low-frequency resonant modes of the crystal/mount combined structure, which causes stress patterns that lead to noise in crystal oscillator outputs.[3] Shock resistance depends on the stresses that support the crystal. Unfortunately, elastic continuum problems are too complicated for analytical solution, particularly if the anisotropy of the quartz elastic properties is included.

One method of solution is Finite Element Analysis (FEA). This technique subdivides the structure into large numbers of smaller units (called elements) that are solved approximately and coupled together to analyze the complete structure. Practical quartz crystals have large diameter to thickness ratios, which necessitates a large number of elements for proper FEA results. A mainframe computer becomes necessary, particularly if one is interested in several computational runs for characterizing the effects of crystal size, mount design, crystal shape variations, mount location variations, etc.

The present work establishes calculational procedures implemented in a desk top computer program which provide adequate solutions for most problems of interest in reasonable computation times. It uses variational calculus to approximate the displacements in the quartz crystal. Body forces are included to represent acceleration. The mounting structures are included as point boundary conditions represented by spring matrices relating deflections and rotations to the forces and moments from the mount reactions.

VARIATIONAL FORMULATION

The stationary expression which is often used for calculation of resonant modes and admittance matrices of piezoelectric bodies[4] can be used for static and quasistatic problems by setting the frequency to zero[5]. Alternately, if resonant frequency is of interest for the low frequency modes of the crystal/mount structure, frequency can be left in the expression. The stationary expression L is reproduced here in the form to be used:

$$L = 1/2\int(u_{i,j}\, c_{ijkl}\, u_{k,l} - \rho\omega^2 u_i u_i)dV - \int u_i f_i dV - \int F_i u_i dS \qquad (1)$$

Here V is the crystal volume, u_i is displacement, c_{ijkl} is the elastic stiffness tensor, ρ is mass density, ω is radial frequency in rad/sec, f_i is body force, and F_i is an externally applied surface traction acting over surface area S. A comma denotes differentiation, piezoelectric effects are excluded, and i,j,k,l, run 1 to 3. All surface area other than S is stress-free, which is the "natural boundary" condition of the variational problem[4]. The coordinate system is the plate axes system as described in Fig. 1.

The displacement u_i is expanded with trial functions in the form of a power series for the present work. A "trial" vector TR(NT,4, is used as a bookkeeping tool. If NT terms are used in the power series expansion,

CH2690-6/89/0000-377 $1.00 C 1989 IEEE

$$u_i(x_1,x_2,x_3) = \sum_{N=1}^{NT} A(N)\, x_1^{TR(N,1)} x_2^{TR(N,2)} x_3^{TR(N,3)} UV_{TR(N,4)}. \quad (2)$$

Here, A(N) is the magnitude of the Nth term and UV is the unit vector. The four numbers TR(N,i) for i=1 to 4 characterize the Nth trial function. The TR(N,1)-TR(N,3) run 0 to the maximum power of x_i, and TR(N,4) runs 1 to 3 for the three axes directions.

The series in Eq. 2 can be substituted into the c_{ijkl} volume integral of Eq. 1. The resulting integrations over V are simple integration of powers of x_i. The present work used a plano-plano disc, for which the integrals are available in standard tables.

The series in Eq. 2 can also be substituted into the body force volume integral. For acceleration a_i, $f_i = \rho a_i$. Again, the volume integrals are over powers of x_i.

If resonant frequencies are of interest, such as the lowest few frequencies needed for shock, vibration, and acoustic noise studies, the series in Eq. 2 can be substituted into the $\rho\omega^2$ volume integral. Again, the integrations are of powers of x_i.

The boundary condition te· in Eq. 1 is broken into two parts to account for point forces and localized moments. There will be NTM locations for a point force, NTM being the number of mounts plus the number of any externally applied forces (such as the diametric force pair used in earlier work[5]). At the Pth location (x_{1P},x_{2P},x_{3P}) on the surface, with $\delta(x_1,x_2,x_3)$ being the delta function.

$$F_i = F_{iP}\ \delta(x_1-x_{1P})\ \delta(x_2-x_{2P})\ \delta(x_3-x_{3P}). \quad (3)$$

The surface integrand in Eq. 1 becomes

$$u_i\ \delta(x_{1P},x_{2P},x_{3P})\ F_{iP} = u_{iP}\ F_{iP} \quad (4)$$

Use of Eq. 2 for u_i provides the complete expression contributing to L. The expression in Eq. 4 is summed for all NTM locations.

Moments are handled as shown in Fig. 2. In Fig. 2 a positive moment acts at point (x_{1P},x_{2P},x_{3P}) about x_1 due to the point forces F_{1P} acting through delta functions at $\pm$ d/2 spacing. The moment M_{1P} about x_1 is

$$M_{1P} = 2dF_{1P} \quad (5)$$

The point forces are given by

$$F_2 = -F_{1P}\ \delta[x_{1P},x_{2P},x_3-(x_{3P}+d/2)] + F_{1P}\ \delta[x_{1P},x_{2P},x_3-(x_{3P}-d/2)] \quad (6)$$

$$F_3 = F_{1P}\ \delta[x_{1P},x_2-(x_{2P}+d/2),x_{3P}] - F_{1P}\ \delta[x_{1P},x_2-(x_{2P}-d/2),x_{3P}]. \quad (7)$$

Integration of the surface term in Eq. 1 gives the contribution to L of

$$-[u_3(x_{1P},x_{2P}+d/2,x_{3P})+u_3(x_{1P},x_{2P}-d/2,x_{3P})] *M_{1P}/2d + [u_2(x_{1P},x_{2P},x_{3P}+d/2,)-u_2(x_{1P},x_{2P},x_{3P}-d/2)] *M_{1P}/2d, \quad (8)$$

or, in the limit of small d,

$$-M_{1P}[u_{3,2}(x_{1P},x_{2P},x_{3P})-u_{2,3}(x_{1P},x_{2P},x_{3P})]/2. \quad (9)$$

By the definition of rotation R_i:

$$\begin{aligned} R_1 &= [u_{3,2}-u_{2,3}]/2 \\ R_2 &= [u_{1,3}-u_{3,1}]/2 \\ R_3 &= [u_{2,1}-u_{1,2}]/2 \end{aligned} \quad (10)$$

we have the contribution at the Pth location:

$$-R_1(x_{1P},x_{2P},x_{3P})\ M_{1P} = -R_{1P}\ M_{1P} \quad (11)$$

Use of Eqs. 10 and 2 provides the necessary boundary condition term for any moment acting at a point.

The complete expression for L is differentiated with respect to A(N), which leads to a linear algebra problem for obtaining the A[N] coefficient.

The effect of mounts is added after the differentation of L with respect to A(N) by considering that the forces and moments come from massless springs. The mount at the Pth location is represented by a 6x6 matrix composed of 4 each 3x3 matrices.

$$\begin{bmatrix} F_{1P} \\ F_{2P} \\ F_{3P} \\ M_{1P} \\ M_{2P} \\ M_{3P} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} 3x3 \\ KMFF \end{bmatrix} & \begin{bmatrix} 3x3 \\ KMFT \end{bmatrix} \\ \begin{bmatrix} 3x3 \\ KMTF \end{bmatrix} & \begin{bmatrix} 3x3 \\ KMTT \end{bmatrix} \end{bmatrix} \begin{bmatrix} u_{1P} \\ u_{2P} \\ u_{3P} \\ R_{1P} \\ R_{2P} \\ R_{3P} \end{bmatrix} \quad (12)$$

The 3x3 matrices in Eq. 12 relate the appropriate variables. As an example, Fig. 3 shows a typical mount and its matrix.

Note in Eq. 12 that u_{iP} and M_{iP} appear. We break up Eq. 12:

$$\begin{bmatrix} F_{1P} \\ F_{2P} \\ F_{3P} \end{bmatrix} = \begin{bmatrix} KMFF \end{bmatrix} \begin{bmatrix} u_{1P} \\ u_{2P} \\ u_{3P} \end{bmatrix} + \begin{bmatrix} KMFT \end{bmatrix} \begin{bmatrix} R_{1P} \\ R_{2P} \\ R_{3P} \end{bmatrix} \quad (13)$$

and

$$\begin{bmatrix} M_{1P} \\ M_{2P} \\ M_{3P} \end{bmatrix} = \begin{bmatrix} KMTF \end{bmatrix} \begin{bmatrix} u_{1P} \\ u_{2P} \\ u_{3P} \end{bmatrix} + \begin{bmatrix} KMTT \end{bmatrix} \begin{bmatrix} R_{1P} \\ R_{2P} \\ R_{3P} \end{bmatrix} \quad (14)$$

These Eqs. 13 and 14 can be substituted into the linear algebra problem formed by the differentation of L by A(N). The net result is a group of terms that appear in conjunction with the volume integral terms of Eq. 1 to add the stiffness of the mounts to the stiffness of the quartz crystal. In this way, the shape obtained for the distortions caused by acceleration account for not only

the weight of the quartz crystal reacting against the spring constants of the mounts but also for the complex bending rigidity of the quartz crystal and how it reacts against the bending rigidity of the mounts. Also, the complex interaction of displacement and rotation come into play through the KMFT and KMTF matrices.

The algebra involved in deriving the complete expressions for the linear algebra problem is somewhat tedious, but the "Trial Vector" of Eq. 2 helps considerably.

The stress T_{ij} can be obtained from

$$T_{ij} = c_{ijkl}\, u_{k,l} \qquad (15)$$

by substituting Eq. 2 for $u_{k,l}$. The differentiation of the series in Eq. 2 is subject to adequate convergence behavior of the series. The series we have obtained in our work have had adequate convergence.

The perturbation formula developed by Tiersten[6] calculates the frequency shift with a perturbation integral that integrates the stress times the vibrational mode displacements. This integration enhances convergence when Eq. 15 is used in the integrand of the perturbation integral.

COMPUTER PROGRAM FORMULATION

The computer program is written in HPBASIC v.4.0. and runs on the HP9816 (Model 216). This platform was chosen because of the ease of developing relatively efficient code. Compilers are available if increased computational speed is required.

The program, called SEASHELL, is divided into three main sections: a command file editor, a command file processor, and a post processor. The program is menu driven, including complete error checking. Subroutines are automatically loaded from disk as needed.

The command file contains all of the information necessary to describe the problem. The editor allows the user to select from among the various options. These options include disc radius, thickness and crystallographic orientation using the IEEE (YXwl) ϕ, θ notation. A choice of materials other than quartz is also offered including an isotropic approximation of quartz. This provides a useful comparison to the full anisotropic results.

The trial vector is created in the editor. Arbitrary powers are allowed in each direction. The only limitation is the amount of memory required to solve the problem. (The array indexing routines used currently limit the trial function order to 256 terms.)

The 6x6 matrix for the mounts to be used must be computed by the user before the program is run. This is typically done either with an FEA program such as the commercially available SAP90 [7] on a PC, or analytically with conventional beam and plate bending formulas. Computation of the 6x6 matrix need only be calculated for one orientation. The program allows insertion of the mount at any location and rotates the 6x6 matrix relative to the plate axes x_i.

Additional information needed to complete a command file is the gravity or acceleration load and any explicit applied forces or moments. Finally, choice of static or resonant frequency solution is made. If resonant frequency is requested, the maximum number of frequencies must be defined.

The Command File Processor performs all the necessary calculational steps. If static solution is requested, the solution of the linear algebra problem is obtained by Gaussian Elimination. If resonant frequencies are requested, Subspace Iteration is used.

The static solution is stored on a disc file with all the pertinent batch information and the solution vector for u_i. If resonant frequencies were requested, a file for each resonant frequency contains the eigenvector for that resonant frequency.

The Post Processor calculates a variety of outputs from the solution vector. Displacements and stresses are available for either static or resonant frequency solutions. The method of presentation can be either numeric printout or plots. Plots may be viewed from any direction. Contour plots of displacement or stress are available and have been found to be useful in visualizing what happens in the active region of vibration of a thickness shear mode resonator.

RESULTS

Calculations were performed for a plano-plano 10 MHz, 3rd overtone, SC-cut resonator, diameter 1 cm, thickness 0.051 cm. Four supports as shown in Fig. 3 were located symmetrically around the perimeter. Figure 4 shows the configuration examined along with the axes convention x,y,z, corresponding to x_1, x_2, x_3. SEASHELL handles multiple loads such as gravity in all three directions in one calculation. Results will be presented here only for x gravity. Calculations were carried out with both isotropic elastic properties and the full SC-cut anisotropy for comparison purposes. As it turns out, some major conclusions can be drawn from the isotropic results. In addition, some calculations were performed with SAP90[7] using both isotropic and full SC-cut anisotropy for comparison.

The trial function choice was made based on prior knowledge of the shape under acceleration. Along x_2 (or y) the deflection is constant for rotations and lateral displacements, and linear for bending. Higher order than 2 is necessary to properly satisfy the $T_{yy}=T_{yz}=T_{yx}=0$ boundary condition. However, memory space of the computer limited our trial function size to first-order only. For this reason, T_{yy}, T_{yz}, and T_{yx} are not accurate in our calculations. However, the weight of the crysta is carried almost entirely by T_{xx}, T_{zz}, and T_{xz}, so the major stress-related effects are covered.

The "S" shape bending expected suggested that at least third-order in x_1 and x_3 were necessary. This proved out in initial calculations. In order to provide a thorough fit of the x_1 and x_3 dependences of T_{xx}, T_{xz}, and T_{zz}, a sixth order polynomial was selected. The resulting trial function is sixth order for the x_1 and x_3 dependencies of u_1, u_2 and u_3, and first order for the x_2 dependence of u_1 and u_3, the 0th order for the x_2 dependence of u_2. The total number of trial terms is 245. On an HP9816, the SEASHELL program solves the complete static acceleration problem in 3 hours. With a math coprocessor and compiled basic, the run time would be approximately 45 minutes. For comparison, a 700 to 900 degree of freedom problem with the SAP90 finite element run takes 5 hours 30 minutes with compiled FORTRAN on an ATT 6300 PC with a math coprocessor.

Figure 5 shows a side view of the isotropic SEASHELL displacement results looking down the z axis for x gravity. Note that because of the angle bracket configuration of Fig. 3, the crystal rotates about z as well as deflects along x. Also note the "S" shape bending. For comparison, Fig. 6 shows a similar view of SAP90 finite element isotropic results. The shapes are the same in displacement, rotation, and bending between Fig. 5 and Fig. 6.

The stress T_{xx} (in dyne/cm^2) for the isotropic displacement result of Fig. 6 is shown in a SAP90 contour plot in Fig. 7. The contour plot is limited to the central region of the crystal, radius 0.2 cm, and the central plane. The view is down the y axis. Note that the zero contour passes through the disc center. The corresponding SEASHELL result is shown as a similar contour plot in Fig. 8, again, limited to the central plane for radius of 0.2 cm and viewed down the y axis. Again, the zero contour passes through the crystal center. In both Figs. 7 and 8 there is a considerable slope to T_{xx} along the x axis. If the mode (active region of vibration) is centered along the x axis, the average stress in the mode region is zero and no frequency shift will occur due to T_{xx}[8]. If the mode is off-center along the x axis, the average stress is not zero and a frequency shift will occur due to T_{xx}[8]. This is one main conclusion gleaned from the isotropic results.

Note in Fig. 7 and 8 that the slope is reversed. Whether the slope is positive or negative depends on the fraction of the crystal weight supported by the two x mounts aligned with the acceleration as compared to the fraction of crystal weight supported by the side mounts on the z axis. The slope of Fig. 8 correlates with a larger fraction carried by the x mounts than the case of Fig. 7.

The calculations leading to Figs. 7 and 8 were made as identical as possible between SEASHELL and SAP90. Unfortunately, the load sharing between the mounts in-line with acceleration and on the sides is very sensitive to details of the mount attachment and mount spring stiffness, particularly bending rigidity. The SAP90 results could be reversed in slope by simply unclamping a few nodes in the mount end to simulate an imperfect spot weld. Apparently changing from a distributed model for the mounts in SAP90 to the point representation in SEASHELL causes different results in T_{xx}. Minor variations in mount location and mount stiffness caused large changes in the T_{xx} in the central plane of the crystal. The conclusion reached from these isotropic calculations is that small mount variations that will occur in production make sizeable changes in the T_{xx} stress patterns. The T_{xx} variable is basically out of control in present-day production of crystal units.

Figures 9 and 10 show isotropic results for x acceleration from SAP90 and SEASHELL, respectively, for T_{xz} in the central plane. Again, the 0.2 cm radius center region viewed from the y axis is shown. Note that the zero contour passes through the center of the crystal for both calculations. Note also that the slope for T_{xz} is along the z axis, orthogonal to the contour lines of T_{xx} seen in Figs. 7 and 10. In this case, a perfectly-centered mode has no frequency shift effect for T_{xz}, but a mode shifted off-center along the z axis has a non-zero net stress and has a T_{xz}-related frequency shift. Thus, if the mode is off-center in any direction, acceleration sensitivity will be affected. Figure 11 shows the SEASHELL results for T_{zz} in the central plane for the x acceleration case. The contour lines show a slope in the x direction similar to Figs. 7 and 8. A mode shifted along the x direction will have a T_{zz}-related frequency shift.

Incorporation of the full anisotropy of a doubly-rotated cut of quartz such as the SC-cut is possible in SEASHELL and SAP90. Figure 12 shows the SEASHELL anisotropic T_{zz} result for the situation of Fig. 11. Note the zero contour line is shifted off-center. Figures 13 and 14 show the anisotropic T_{xx} results for the situation of Figs. 7 and 8. Note that the zero contour line is not at the center of the crystal. A centered mode has a net T_{xx} stress average in Figs. 13 and 14, and a net T_{zz} stress average in Fig. 12. There will be T_{xx}- and T_{zz}- related acceleration effects on resonant frequency. It should be noted here that reversal of the acceleration direction only changes the sign of the contour line values, the off-center shift stays the same direction.

Figure 15 shows the anisotropic SAP90 T_{xz} results for the situation of Fig. 9. Here the zero contour line has shifted slightly off the center of the crystal.

DISCUSSION

It is clear from the isotropic results that several gradients for the T_{xx}, T_{zz}, and T_{xz} stresses occur along the two-in-plane axes x and z for x acceleration. In addition, for z-directed accelerations, the contour patterns are 90° from those shown here for x-direction acceleration. If the thickness-shear mode is off-center from the crystal center in any direction, the stress gradients cause a non-zero net average stress and a frequency shift[8].

The anisotropic results point out that besides the gradients in stresses T_{xx}, T_{zz} and T_{xz}, the zero contour line falls off the center of the crystal for all three stresses. Thus, there will be frequency shifts even for a centered mode.

It is concluded that useful information can be obtained relatively easily with SEASHELL. The mounts can be moved or rotated with only one minute of setup time. The resonator crystallographic orientation diameter and thickness can also be changed in short time. This ease of set-up allows the user to do parameter studies with a minimum of effort. In contrast, finite element programs require that the user move numerous mesh points to simulate mount changes, or crystal diameter and thickness changes. In defense of finite element analysis, given enough computer space, a more thorough treatment of T_{yy}, T_{yx}, and T_{yz} is possible.

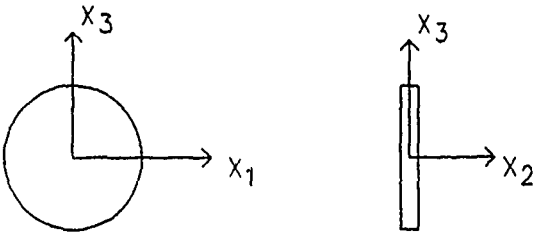


PLATE AXES SYSTEM

Figure 1. The plate axes system used in the theoretical development.

REFERENCES

[1] R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," Proc. 41st Annual Symposium on Frequency Control, 1987, p. 398.

[2] R.L. Filler, et. al, "Further Studies On The Acceleration Sensitivity of Quartz Resonators," Proc. 37th Annual Symposium on Frequency Control, 1983, p. 265.

[3] R.D. Weglein, et. al, "Microphonics In VHF Quartz Resonators and Oscillators," Proc. 1984 Ultrasonics Symposium, 1984, p. 378 (Available from IEEE).

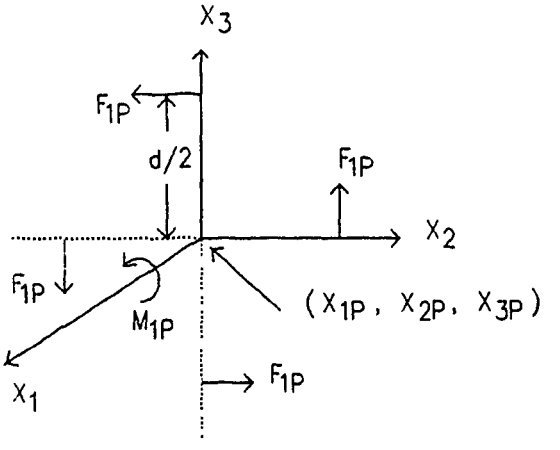
[4] R. Holland and E.P. EerNisse, "Design of Resonant Piezoelectric Devices," Research Monograph No. 56, The M.I.T. Press, Cambridge, 1969.

[5] E.P. EerNisse, et. al, "Variational Calculation of Force-Frequency Constants," IEEE Trans. Sonics and Ultrasonics SA-25, 1978, p. 132.

[6] H.F. Tiersten, "Perturbation Theory For Linear Electroelastic Equations For Small Field Superposed On A Bias," J. Acoust. Soc. Am. 64, 1978, p. 832.

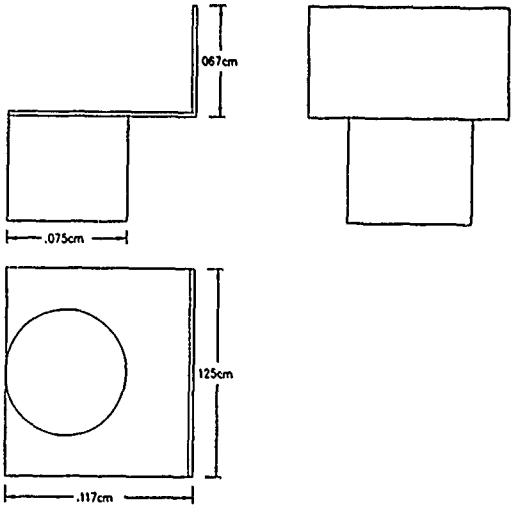
[7] Computers and Structures, Inc., 1918 University Ave., Berkeley, CA 94704 USA.

[8] E.P. EerNisse, et. al, "Acceleration-Induced Frequency Shifts In Quartz Resonators," these proceedings.



MOMENT REPRESENTED BY FORCE COUPLES

Figure 2. A moment about x_1 acting at point (x_{1P}, x_{2P}, x_{3P}) and its representation by a pair of force couples for the present theoretical development.



2.6E8	9.4E8	0	0	0	1.0E7
9.4E7	9.8E7	0	0	0	1.9E6
0	0	6.6E8	-4.2E7	5.0E6	0
0	0	-4.2E7	2.9E6	-4.1E5	0
0	0	5.0E6	-4.1E5	6.0E5	0
1.0E7	1.9E6	0	0	0	5.4E5

MATERIAL: NICKEL
E=2.1E12 dyne/cm^2
THICKNESS = 0.0076cm
ν = 0.3

Figure 3. The metal strip mount used for the calculations. The 6x6 spring matrix for the mount is shown in units of dyn/cm and dyn cm/rad. It was obtained from a SAP90 calculation.

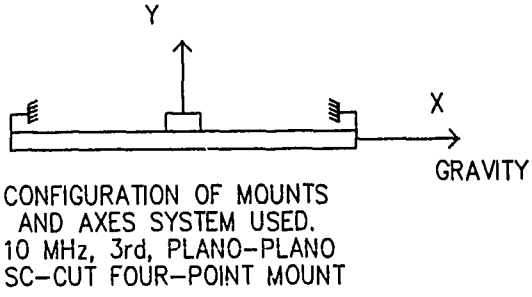


Figure 4. The mount configuration used in the calculations. The x,y, and z axes (equivalent to x_1, x_2, and x_3) are shown. Gravity acts along + x.

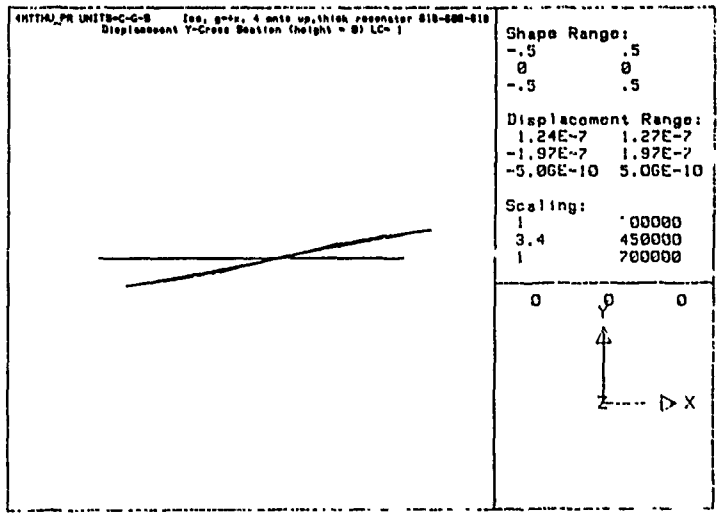


Figure 5. The SEASHELL calculated displacement for the case of Fig. 4, i.e., a 10 MHz, 3rd, plano-plano, SC-cut with gravity acting along + x. Isotropic elastic properties were used and the mounts of Figs. 3.

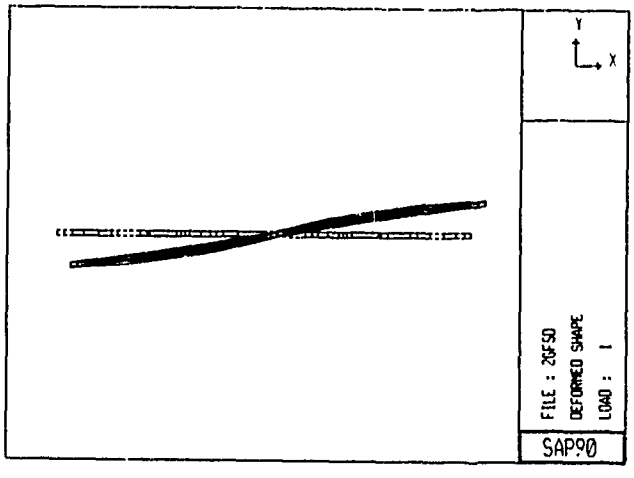


Figure 6. The SAP90 calculated displacement for the Fig. 4 configuration, for comparison to the Fig. 5 results.

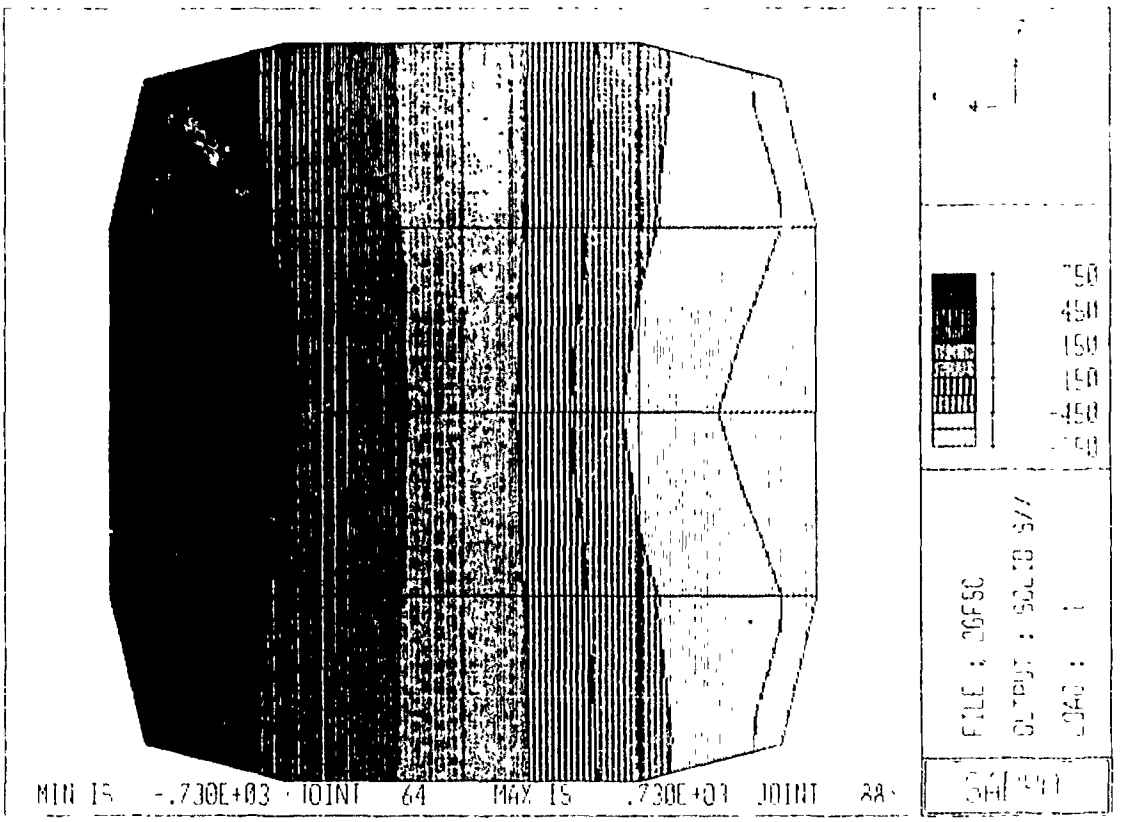


Figure 7. A SAP90 contour plot of T_{xx} for the case of Fig. 4, i.e., a 10 MHz, 3rd, plano-plano, SC-cut with gravity acting along the + x. Isotropic elastic properties were used. The view is down the y axis and covers the central plane of the crystal for a radius of 0.2 cm.

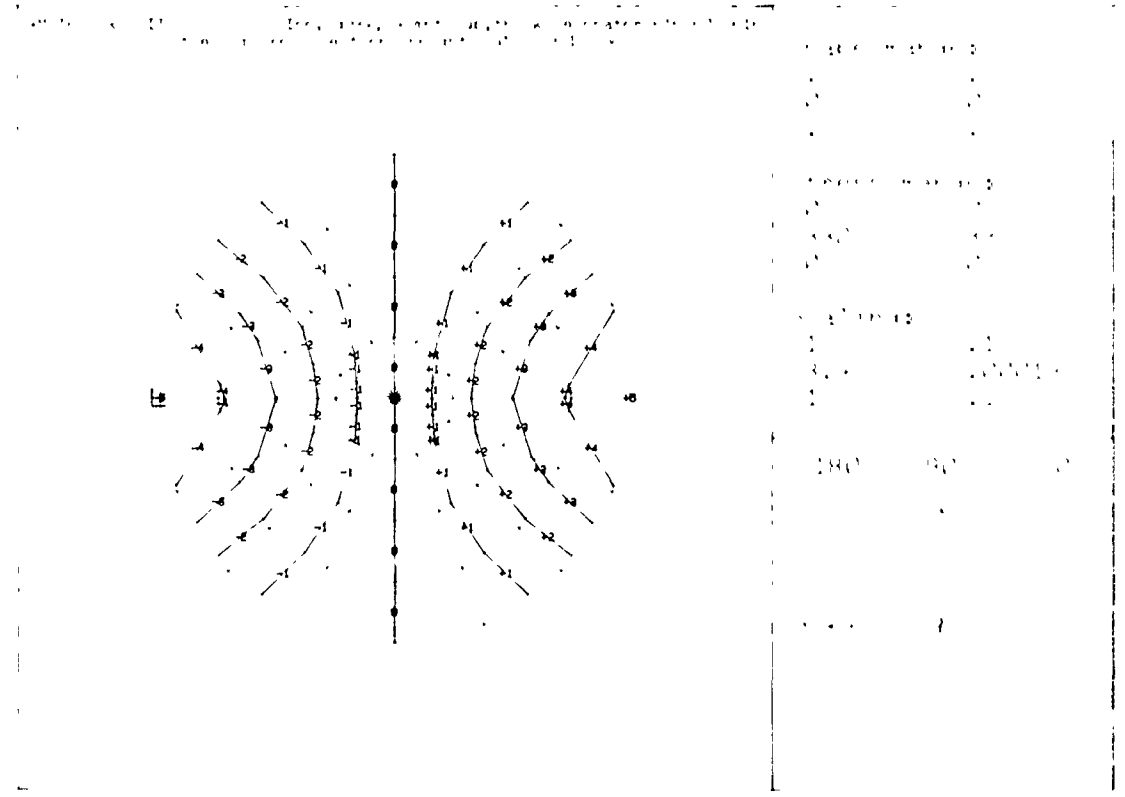


Figure 8. The SEASHELL T_{xx} results in dyne/cm^2 for the case of Fig. 4 using isotropic elastic properties. Again, the contour plot is for the central plane of the crystal for a radius of 0.2 cm.

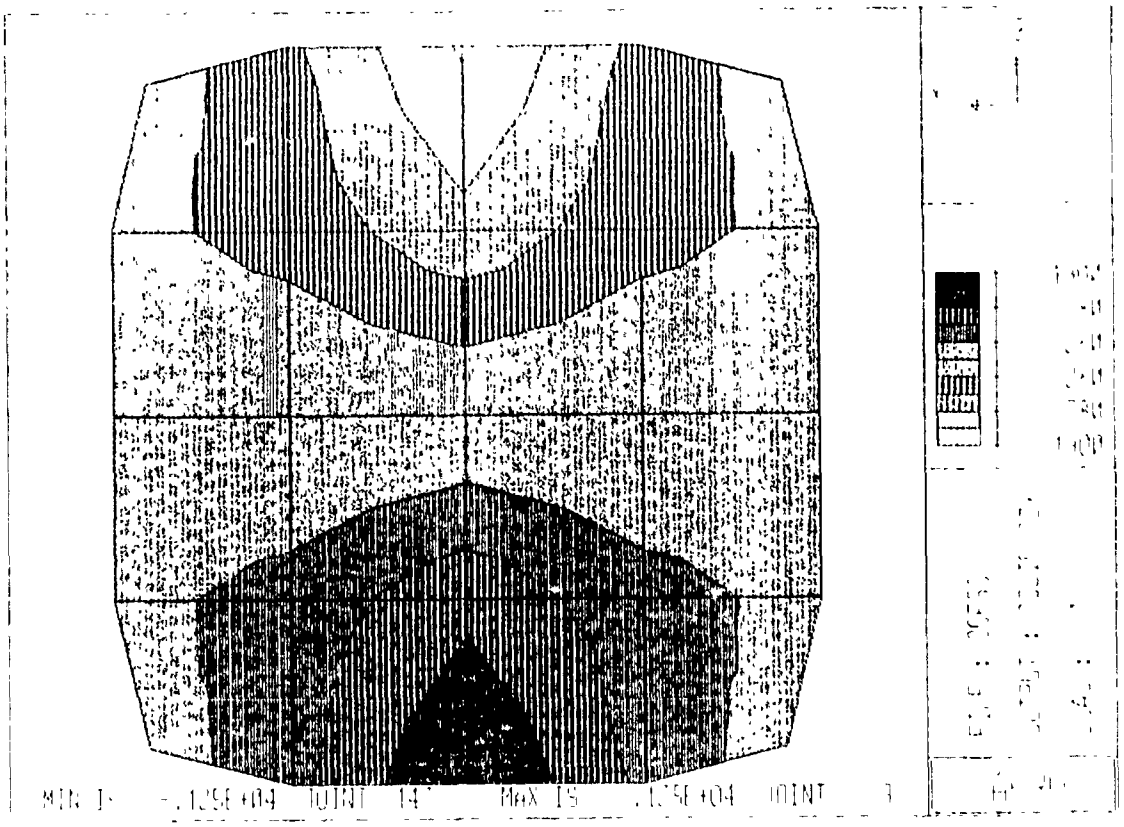


Figure 9. The SAP90 results for T_{xz} in dyne/cm^2 for the case of Fig. 4 using isotropic elastic properties.

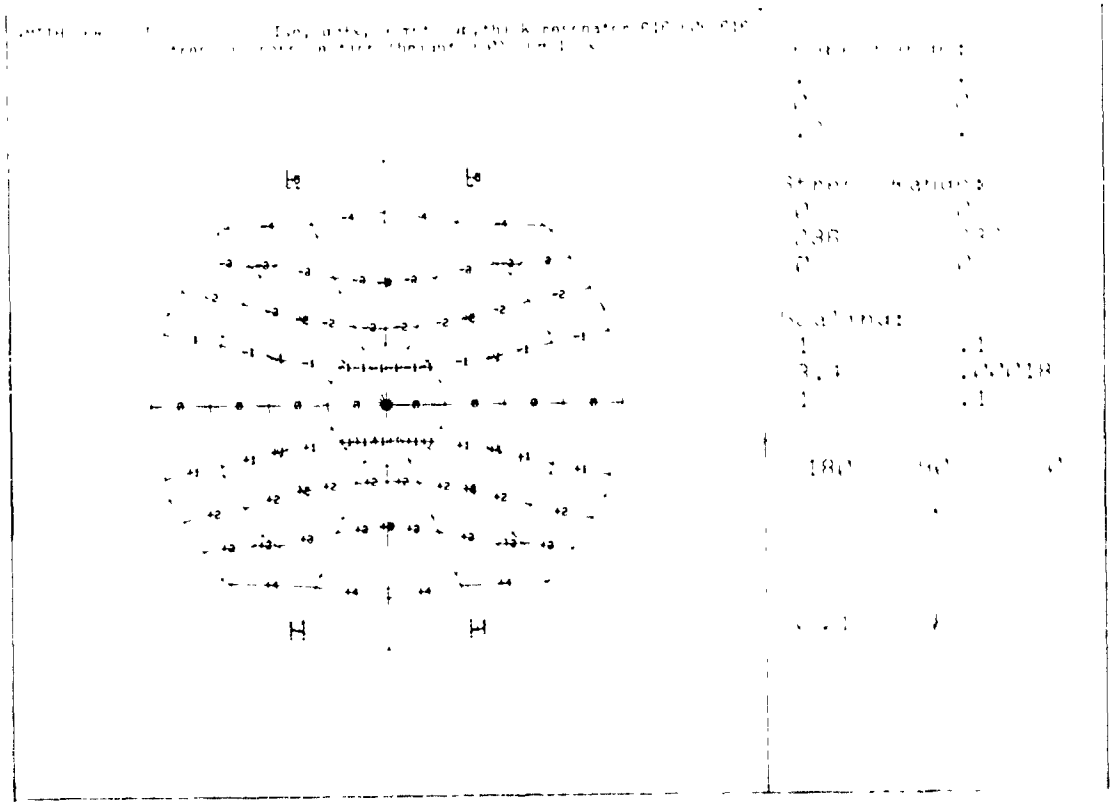


Figure 10. The SEASHELL results for T_{xz} in dyne/cm^2 for the case of Fig. 4 using isotropic elastic properties.

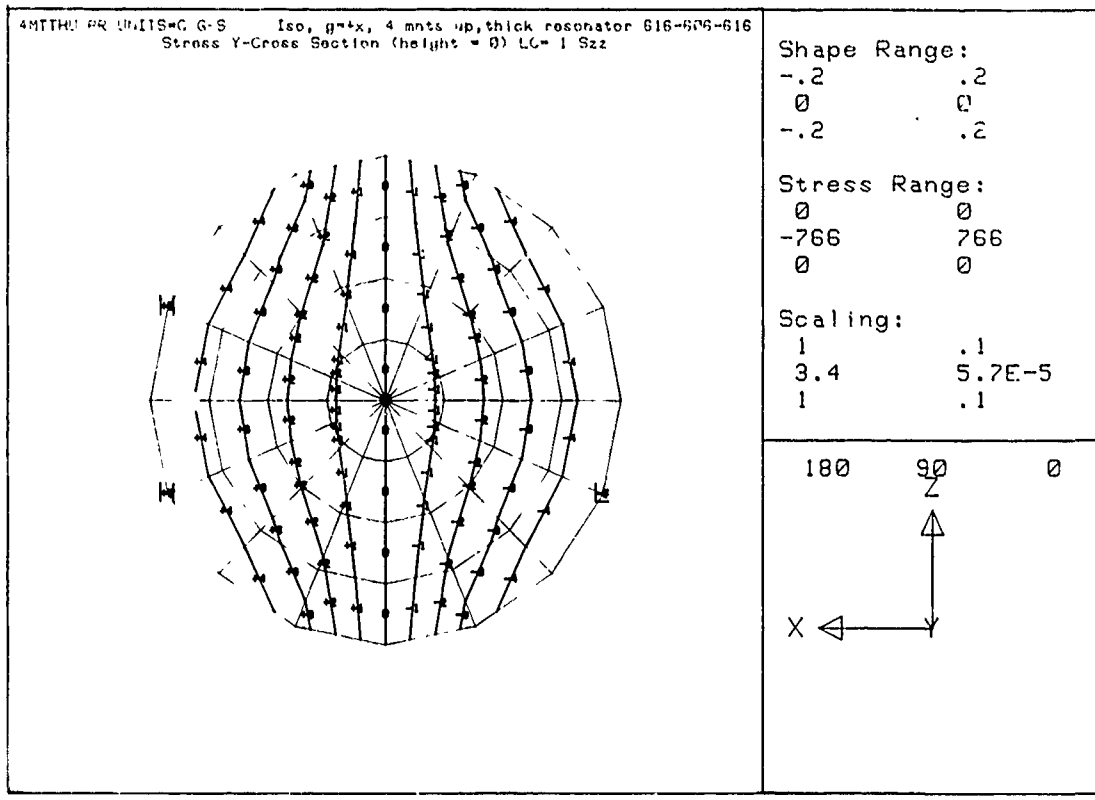


Figure 11. The SEASHELL results for T_{zz} in dyne/cm^2 for the case of Fig. 4 using isotropic elastic properties.

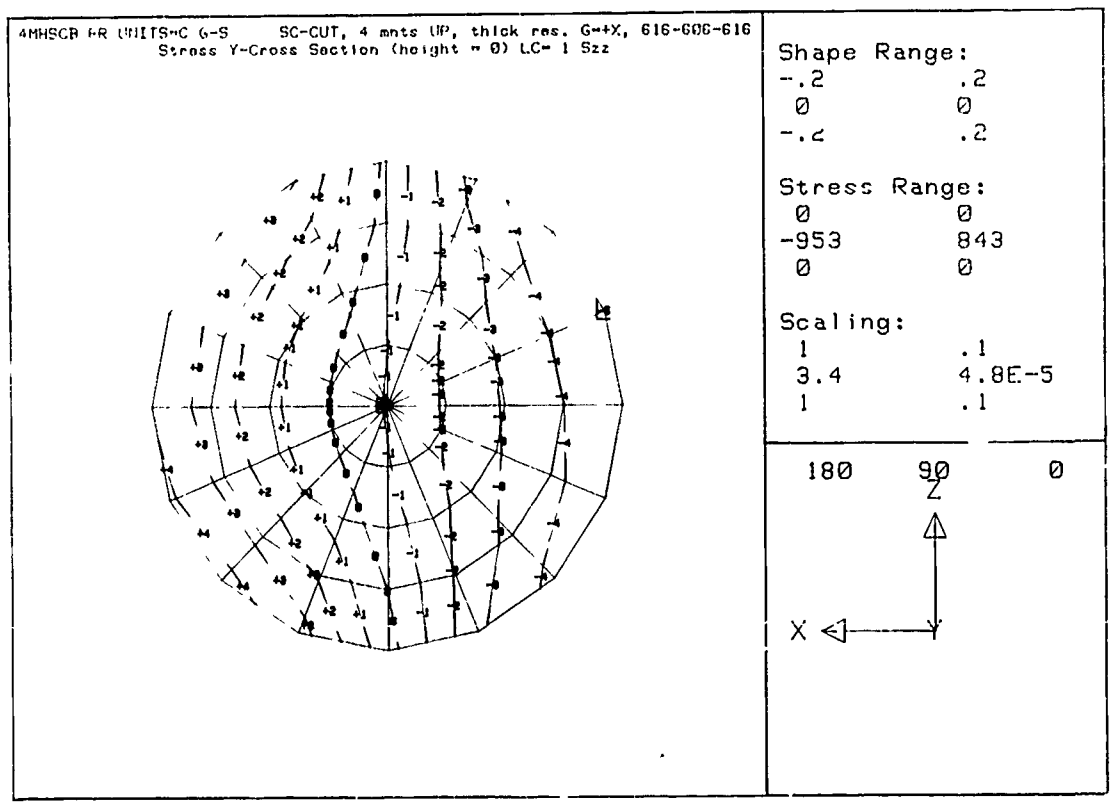


Figure 12. The SEASHELL results for T_{zz} in dyne/cm^2 for the case of Fig. 4 using the full anisotropy of the elastic properties.

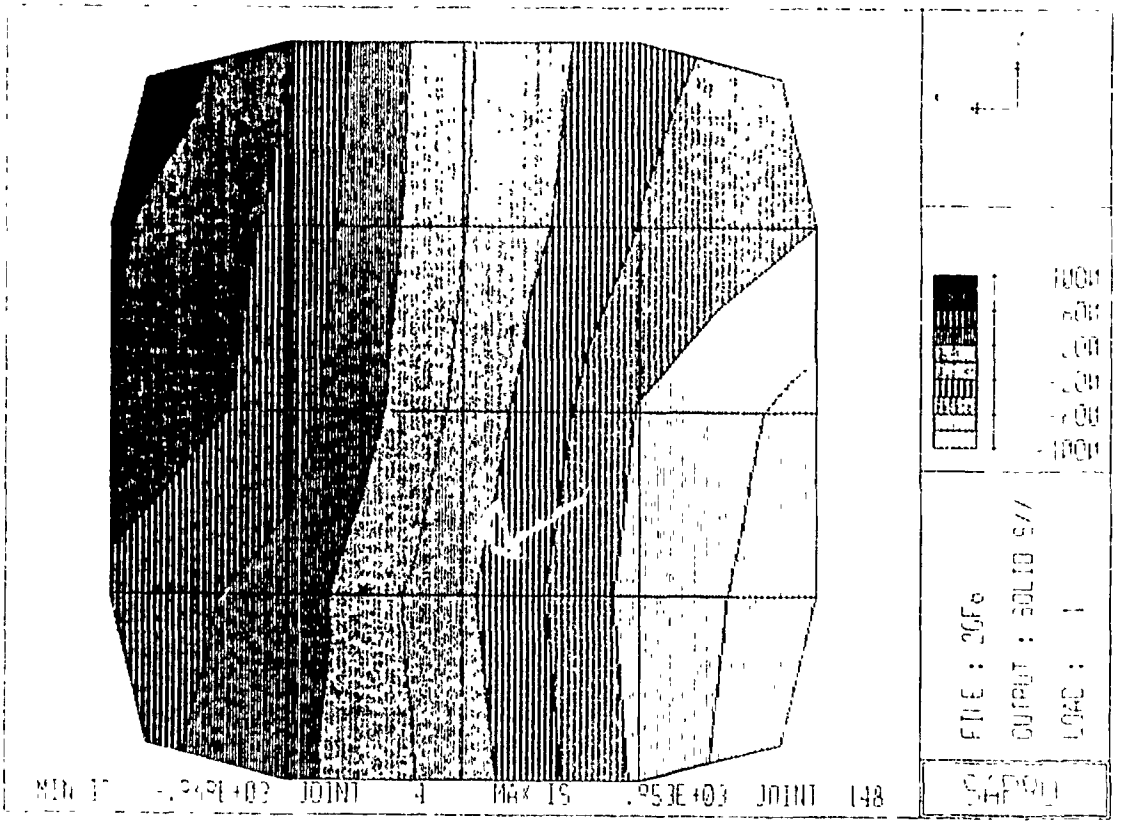


Figure 13. The SAP90 results for T_{xx} in dyne/cm^2 for the case of Fig. 4 using the full anisotropy of the elastic properties.

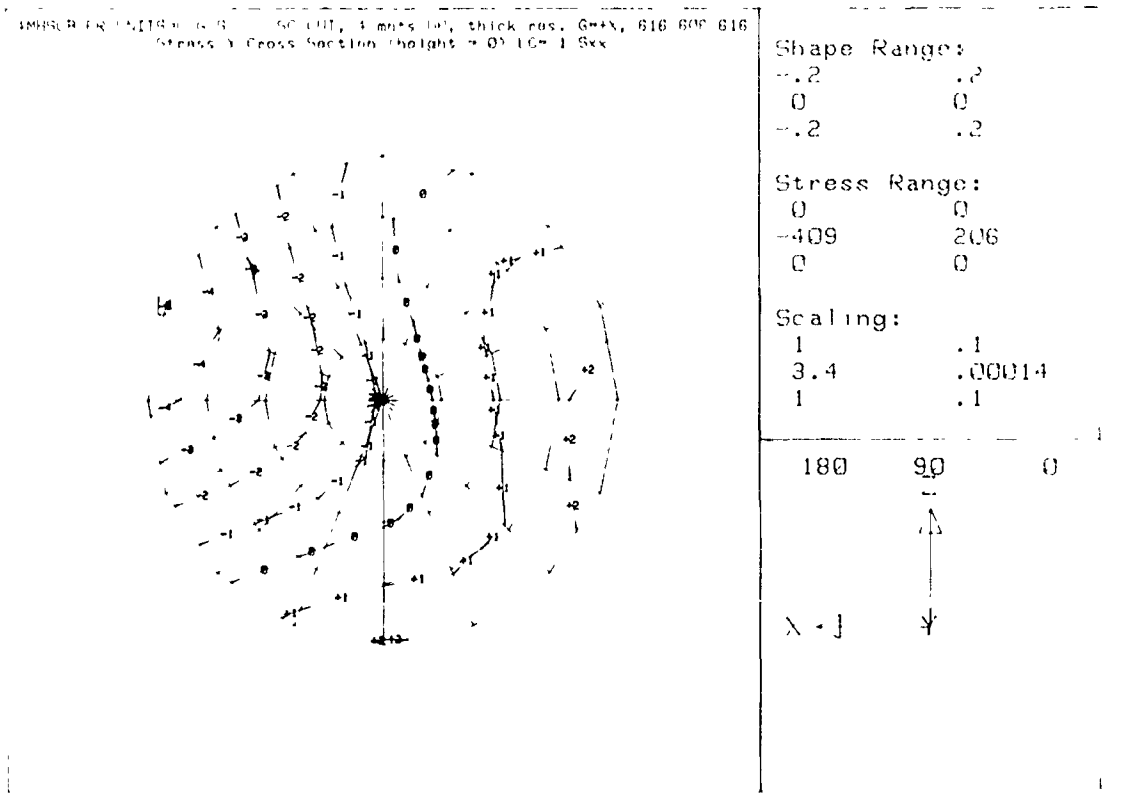


Figure 14. The SEASHELL results for T_{xx} in dyne/cm^2 for the case of Fig. 4 using the full anisotropy of the elastic properties.

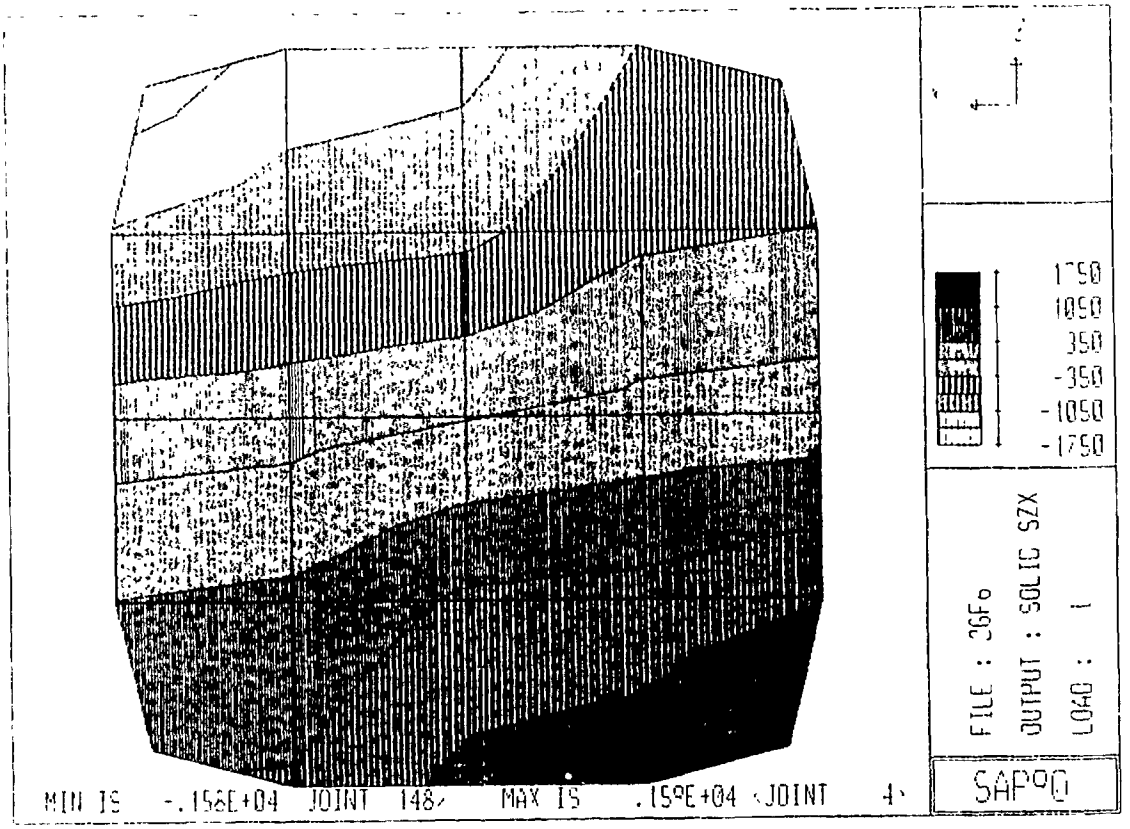


Figure 15. The SAP90 results for T_{xz} in dyne/cm^2 for the case of Fig. 4 using the full anisotropy of the elastic properties.

43rd Annual Symposium on Frequency Control - 1989

ACCELERATION-INDUCED FREQUENCY SHIFTS IN QUARTZ RESONATORS

Errol P. EerNisse
Roger W. Ward
O. Lew Wood

Quartztronics, Inc.
1020 Atherton Drive
Building C
Salt Lake City, UT 84123

SUMMARY

Extensive experimental and theoretical results are presented to support the hypothesis that the shape and location of the active region of vibration of a thickness shear mode quartz resonator is the dominant factor in determining the acceleration sensitivity of the resonator. The shape and location of the mode in a "real world" resonator vary sufficiently from unit to unit due to material and processing variations that the effect swamps out all other considerations. It is shown theoretically and experimentally that the mode shape and/or location can be "trimmed" by judicious addition or subtraction of mass to produce resonators with improved acceleration sensitivity.

INTRODUCTION

Acceleration induces a shift in the frequency of a quartz crystal resonator. Although the effect is small by most standards, e.g. 5×10^{-10}/g fractional frequency shift, it can cause close-in sidebands in a crystal oscillator output during vibration that degrade the performance of modern communications, navigation, and radar systems[1]. The acceleration-induced frequency shift is characterized in the literature by a "gamma" vector, the three orthogonal components of the gamma vector being the fractional frequency shift observed for unit acceleration in the three orthogonal directions[1]. There have been extensive theoretical and experimental investigations on the subject, which are summarized by Filler[1]. Of particular note is the experimental work by Filler, et. al.[2], wherein it is stated, "No significant systematic variation of acceleration sensitivity has been found as a function of any of these parameters except overtone for the same resonator." The parameters were quartz blank geometry, angle of cut, mounting orientation, temperature, drive current, and overtone. Figure 1 shows experimental results from Watts, et.al.[3], for a 10 MHz, 3rd overtone, two-point mounted design. Note the large scatter in results, which is characteristic of experimental measurements on real-world resonators.

Extensive past theoretical work has proceeded in the direction of refinement of analytical techniques[1], culminating recently in some suggested configurations that promise low acceleration sensitivity in one direction [4,5], the direction normal to the blank. It is clear that the theoretical developments are becoming powerful, particularly now that finite element analysis

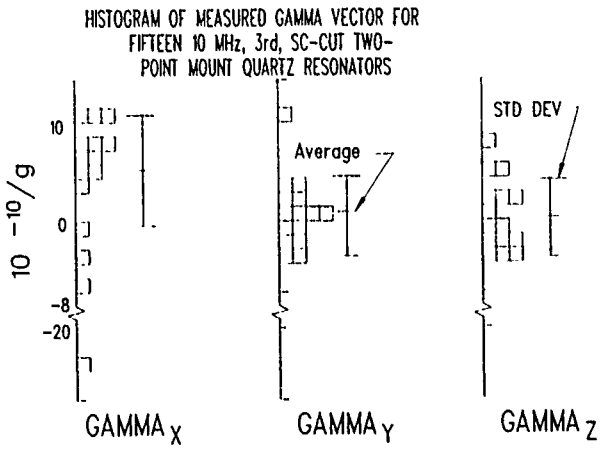


Figure 1. Histogram of gamma measured on each axis of 15 resonators (10 MHz, 3rd, SC-cut, 2-point mount). (From reference 3).

on desk top computers is available. However, it is the hypothesis herein that the theoretical efforts have not properly included one major variable and, therefore, have not treated the source of the large experimental scatter seen by the experimentalists. It is a further hypothesis herein that the source of the large scatter in acceleration sensitivity in production crystal units must be brought under control experimentally and theoretically before any practical designs of low-acceleration sensitivity quartz crystal resonators can be manufactured for commercial and military use.

Experimental and theoretical evidence will be presented to prove that the major variable yet to be properly treated is the shape and location of the mode (active region of vibration) in a "real world" resonator. Once this fact is established, the reader will realize that this problem in crystal fabrication technology will never go away completely. In fact, as technology improves in the future, allowing sensitivities to drop into the range of 5×10^{-11}/g, the precision needed for controlling the actual shape and location of the mode will become even more demanding.

It is proposed herein that, because of material imperfections (such as dislocations), and fabrication imperfections (thickness variations, contouring imperfections, electrode mass density and thickness variations), the demands on mode shape control will be impractical to accomplish during manufacturing without some ability to "trim" the mode shape and location

CH2690-6/89/0000-388 $1.00 C 1989 IEEE

for each individual resonator. This holds particularly true for any forthcoming "low-sensitivity" designs because of the inherent impact of the actual mode shape and location on the acceleration sensitivity.

The favored method of "trimming" the acceleration sensitivity[6] is to move the mode by slightly altering the mass distribution of a resonator after first measuring its acceleration sensitivity. The mass distribution is altered by adding or subtracting mass on selected areas of the active region of vibration. Modifying the electrode material is the most straight forward trimming technique for one to consider, but quartz could be removed by reactive etching, etc. for lateral-field-excited designs, for example.

Finally, the theoretical results suggest that the ultimate performance of a quartz resonator undergoing acceleration may be obtained with the mode location slightly off-center of the blank[6,7].

EXPERIMENTAL EVIDENCE

A combination of experimental results with supporting theoretical background will be presented. The experimental results are convincing by themselves and will be presented first. Our hypothesis deals with deviations in the mode shape from the ideal. Normally, an ideal mode in a contoured blank is considered to have its vibration energy density of Gaussian shape due to energy trapping, centered in the resonator blank, and somewhat elliptical when viewed from the blank thickness direction because of the anisotropy of quartz. However, past experimental measurements of the vibration intensity (energy density) by X rays has shown that there are usually minor (sometimes major) deviations in the smooth elliptical shape predicted by theory. As an example, we reproduce here the 5 MHz, 5th overtone, AT-cut plano-convex results of Zarka, et.al.[8], in Figs. 2a and 2b. In the figures, the dark area in the center of the crystal blank is the main vibrational area. As discussed by Zarka, et.al.[8], the lateral confinement of the vibration energy is degraded by the presence of dislocations, as exhibited by the many dark lines in Fig. 2a. Natural quartz with growth bands but no dislocations produces the results in Fig. 2b. Note in Fig. 2b that the dark area in the center of the blank is not a clean ellipse. Zarka, et.al., suggest that the dislocations guide vibration energy away from the center region. They also suggest that surface defects cause similar effects.

As will be seen in the Theoretical Section herein, a small deviation in mode shape and location causes large effects in the acceleration sensitivity. It has been known for years that the use of a quartz resonator as a thin film thickness monitor is accurate only if the complete mode is covered by the film. Partial mode coverage gives different sensitivities from resonator to resonator. This effect can be used here. The frequency shift observed for a 150A film of Pt deposited in a small pie-shaped area was used as a probe to explore mode

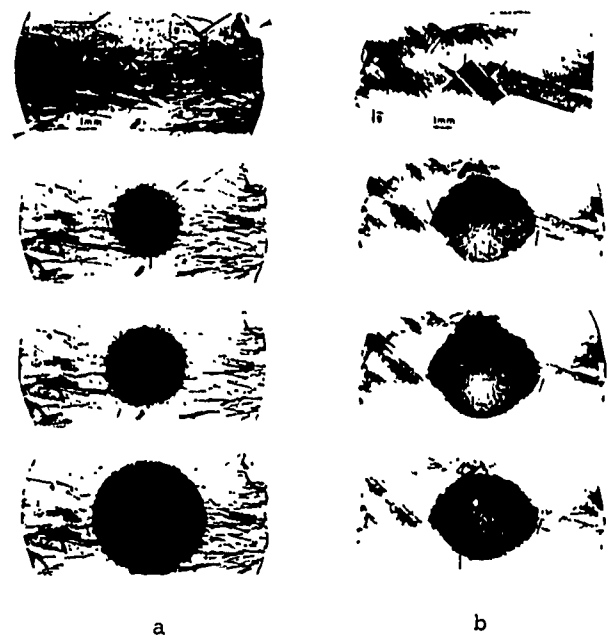


Figure 2. Stroboscopic Xray Topography of AT-cut, 5MHz, 5th, plano-convex resonators. a) Cultured quartz showing how dislocations guide vibration energy away from the center. b) Natural quartz with growth bands showing irregular mode shapes. (From reference 8).

USE OF PIE-SHAPED Pt DEPOSITIONS TO PROBE UNIFORMITY OF ACTIVE REGION

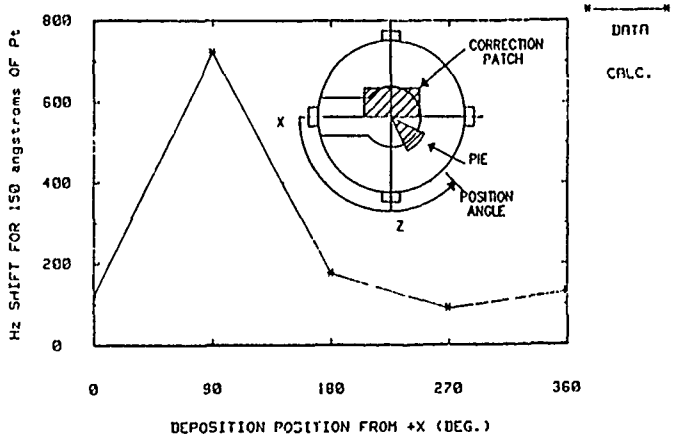


Figure 3. Use of a pie-shaped deposition of 150A of Pt to probe the actual distribution of vibration energy in a 10 MHz, 3rd, SC-cut, four-point mount, plano-convex resonator. The large, rectangular patch of Pt was added to the side with the lowest vibration energy to improve the uniformity of the mode. The dashed line is 1/6 of the frequency shift estimated for a full coverage Pt film.

uniformity in a 10 MHz, 3rd, SC-cut, four-point mount resonator. The results are shown in Fig. 3. Frequency shift is proportional to the fractional amount of the total vibrational energy covered by the pie-shaped deposition. The pie shape used was a 60° wedge, so one would expect a frequency shift 1/6 of that obtained by a full-coverage layer of the same thickness. Note the large variation in the frequency shift about the dotted line, which is 1/6 of the uniform coverage estimate, observed as the pie-segment is deposited at different locations around the center of the blank. Large area depositions of Pt were subsequently placed on the -z side (directions as defined in Fig. 3), which exhibited the lowest frequency shift (lowest vibration energy density), in order to move the mode in the -z direction. The acceleration sensitivity was measured before and after the depositions of the Pt. Figure 4 shows that all three of the components of the gamma vector improved after the mode was forced to be more "uniform" by the deposition of Pt on the originally less-active -z side of the mode. Thus, there is strong evidence that mode shape and location have a critical effect on the acceleration sensitivity.

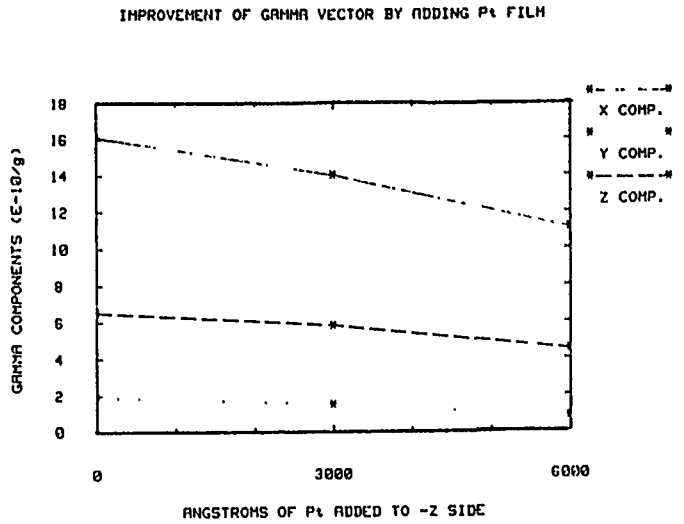


Figure 4. The gamma vector components of the resonator of Fig. 3 before and after two depositions of the large rectangular patch of Pt.

Watts, et.al.[3] improved the gamma vector of four 10 MHz, 3rd, SC-cut two-point mounted resonators by a "proprietary" process. That process was the judicious use of Pt depositions on selected areas of the blank. Figures 5a and 5b show how the three gamma vector components changed for those resonators as patches of mass were added in locations off-center in two orthogonal directions X and Z on the surface of the resonator. The directions are noted in Fig. 6, as is the location of the X' crystallographic direction.

Several of the resonators of the group of 15 units were used to obtain Figs. 5a and 5b. The results were reversible, i.e., if the patch of Pt was deposited in the -X direction, the gamma vector components changed opposite to the change observed for the +X patch. The change in the gamma vector components were similar for all samples examined, which verifies that the stress patterns in the quartz were similar from device to device and that the Pt depositions moved the mode similarly from device to device. Considerable scatter occurred, however, as evidenced in Figs. 5a and 5b. The lines in Figs. 5a and 5b are estimates of how each of the gamma components changed.

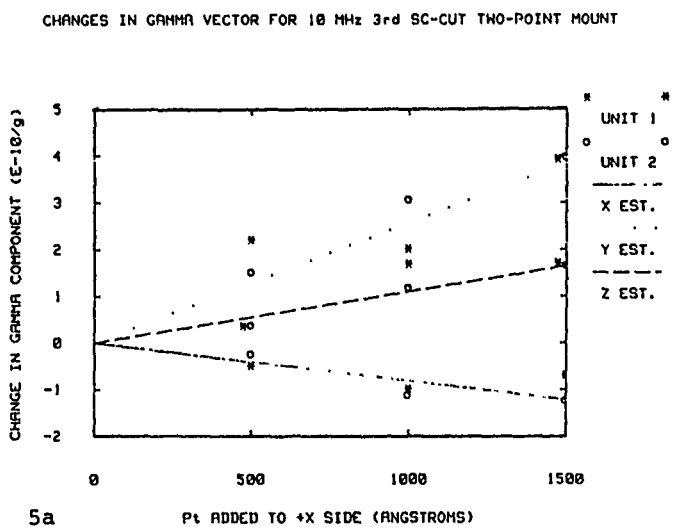
Four of the units were selected for attempts to improve the gamma vector. It is no coincidence that all four were successfully improved. The fact that they were dissimilar in gamma vector in the first place was because the modes were not uniform. Our adding of patches of Pt simply improved the irregular mode shapes to the more desirable shape and location envisioned by the theorticians (and the engineers who designed the original plano-convex crystal shape).

Figure 7 shows the progressive improvement of one device as successive patches of Pt were added to the +X side. The X and Z notations in Fig. 6 indicate the convention for the locations wnere patches were added. The thicknesses of Pt seen in Fig. 7 were thicker than usually used in resonator fabrication. These large thicknesses were needed to overcome both the flaws in manufacturing and the contour in order to change the mode shape. The future application of this technique will be best applied to resonators with less curvature[2,9].

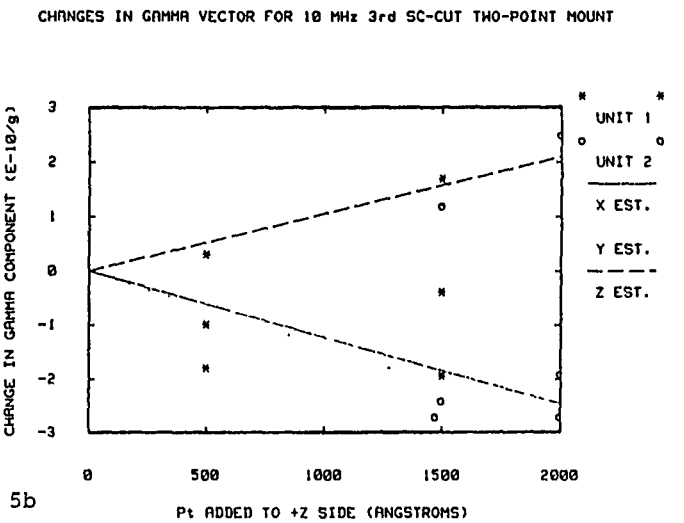
Figure 8 shows the final gamma vector of all four resonators compared with the original spread of the gamma vector. Patches of Pt were added to a variety of locations (not the same from unit to unit). The fact that such a remarkable improvement in gamma vector can be accomplished with such a simple procedure, after the original manufacture and mounting of the resonator, supports the hypothesis of mode shape and location being the most critical variable.

It should be noted that the two-point mount used on the devices of Figs. 5-8 is not symmetric in that one header post is longer than the other. The final gamma vector of Fig. 8 is actually smaller than predicted by theory if the mode is assumed to be at the geometric center of the blank. This occurs because enough Pt has been deposited to pull the mode off-center slightly to the most favorable location within the stress pattern set up in the resonator during acceleration. More on this can be found in the Theory Section.

It must also be mentioned that it was not possible to obtain improvements as dramatic as seen in Fig. 8 for 10 MHz, 3rd, SC-cut four-point mounts obtained from two different manufacturers. The modes were so distorted that the Pt layers became too thick and peeled off. In order to use this "trimming" technique to effectively reduce the acceleration sensitivity of mass-produced



5a



5b

Figure 5a and b. Changes in the measured components of the gamma vector for 10 MHz, 3rd, SC-cut, two-point mount resonators as a Pt film is added to the surface in directions and locations shown in Fig. 6. a) Patch added to +X location; b) Patch added to +Z location.

crystal units, they will have to be made with a low contour number[2,9], careful fabrication practices, and quality material selection; otherwise, the mode cannot be moved selectively to accomplish optimum results by addition of reasonable amounts of mass.

In case the reader is not yet convinced that mode shape and location can be used to improve the gamma vector of resonators, additional data is presented here. A number of 50 MHz, 5th, SC-cut four-point mount resonators were studied. These units had low contour, and were thinner, so the amount of mass to be added was smaller. Figure 9 shows the definitions of X and Z, and the X' crystallographic direction. Figure 10 shows a collective result of the gamma vector before and after addition of the Pt patches for 10 units. The X and Y gamma components

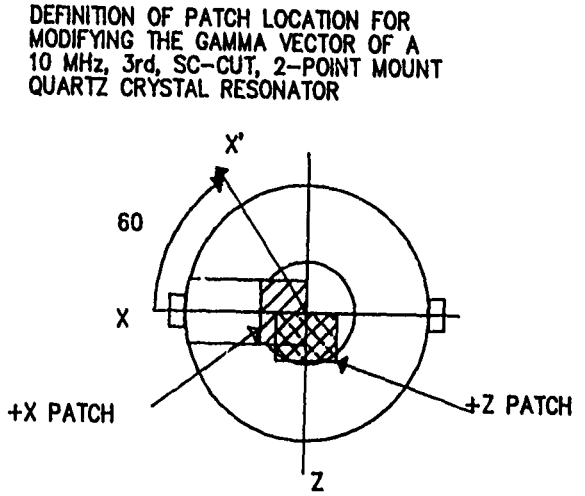


Figure 6. Top view of a 10 MHz, 3rd, SC-cut, two-point mount resonator showing the two locations for the deposition of small patches of Pt. The experimental axis X and Z and the crystallographic X' axis are shown.

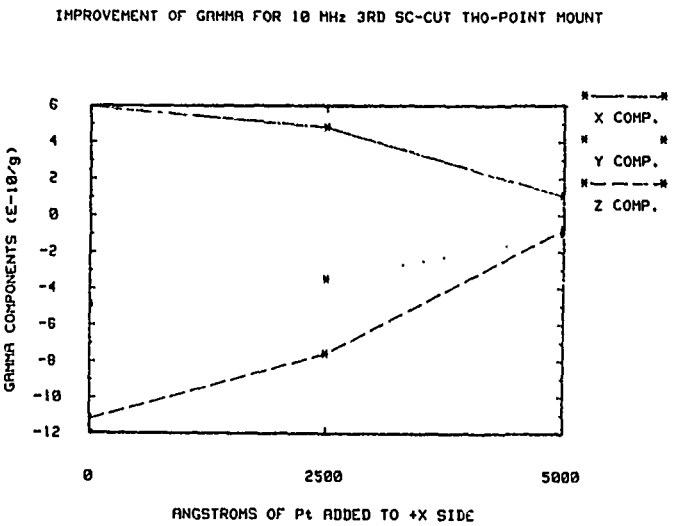


Figure 7. Progressive improvement of a 10 MHz, 3rd, SC-cut, two-point mount resonator gamma vector as sequential patches of Pt are deposited on the surface in the +X direction (see Fig. 6).

responded well to the mode movement. The fact that the results are more closely grouped after processing shows first that the addition of the Pt patches has shaped the mode to the optimum shape and location for this design and second, that the future study of effects such as mounting location, blank contour, etc., will be much easier when the methods presented here are used to create a uniform mode.

Further improvement in the results of Fig. 10 can be obtained only after a careful theoretical and experimental study of the effects of crystallographic orientation of the mounts and different mounting structures.

IMPROVEMENT OF GAMMA VECTOR OF FOUR 10 MHz, 3rd, SC-CUT, TWO-POINT MOUNT QUARTZ RESONATORS BY DEPOSITING Pt PATCHES ON SELECTED AREAS

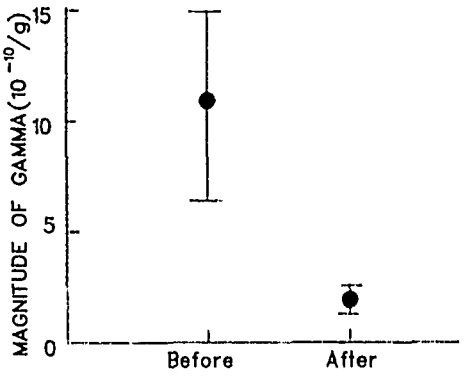


Figure 8. Comparison of the magnitude of the gamma vector of four resonators from the group of Fig. 1 before and after depositing selected patches of Pt.

DEFINITION OF PATCH LOCATION FOR MODIFYING THE GAMMA VECTOR OF 50 MHz, 3rd, SC-CUT, 4-POINT MOUNT QUARTZ CRYSTAL RESONATOR

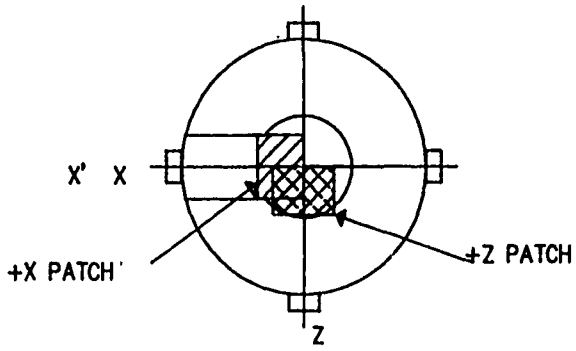


Figure 9. Top view of a 50 MHz, 5th, SC-cut, two-point mount resonator showing the convention for the deposition of small patches of Pt. The experimental axes X and Z and the crystallographic X' axis are shown.

IMPROVEMENT OF GAMMA FOR 50 MHz 5th SC-CUT FOUR-POINT MOUNT

0 = BEFORE 1 = AFTER

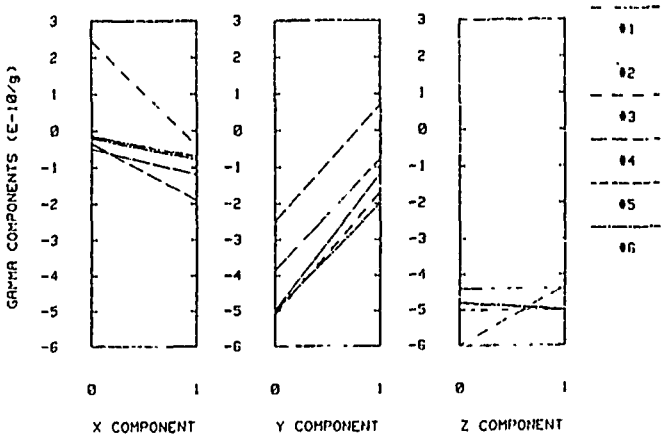


Figure 10. Comparison of the gamma vector components of a group of six 50 MHz, 5th, SC-cut, four-point mount resonators before and after depositing selected patches of Pt.

THEORETICAL BACKGROUND

The mode shape plays a critical role in determining acceleration sensitivity. In order to focus attention on concepts instead of details, the theoretical description presented here will leave out tensor notation. Any calculated results that are presented are derived from computer programs based on rigorous formulations[10,11]. The perturbation theory of Tiersten[10] has one term that contributes the majority of the frequency shift effect. It is represented here conceptually as

$$df/f = C_3 \int G^2(X,Y,Z)\ E(X,Y,Z)dV \qquad (1a)$$

where df/f is fractional frequency shift, G is the normalized strain distribution associated with the mode of vibration (G^2 is proportional to the strain energy density), E is the strain distribution associated with the stress caused by acceleration, C_3 is the third-order elastic constant tensor, V is the volume of the resonator, and dV=dXdYdZ. The coordinate system is the same as used in Figs. 6 and 9, with Y being the normal to the resonator blank.

In examining Eq. 1a, we see that this frequency shift term is simply the integration of the third-order elastic constant times the mode energy density times the strain pattern set up by the acceleration.

Equation 1a can be rewritten to calculate the same result in terms of stresses instead of strains, again conceptually:

$$df/f = K \int G^2(X,Y,Z)\ T(X,Y,Z)dV \qquad (1b)$$

Here, K is the frequency shift coefficient for the stress T.

Consider the general case where the stress from acceleration has components that are even symmetry with X and also some that are odd symmetry with X. The stress T (X,Y,Z) is divided into the even component and odd component:

$$T(X,Y,Z) = T_e(X,Y,Z) + X\ T_o(Y,Z). \qquad (2)$$

where T_o is a stress slope.

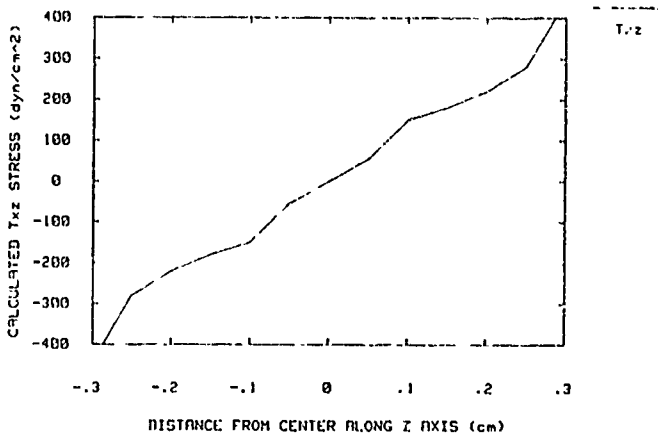


Figure 11. The calculated in-plane stress T_{xz} vs. Z for a 10 MHz, 3rd, SC-cut, four-point mount accelerated along the X direction.

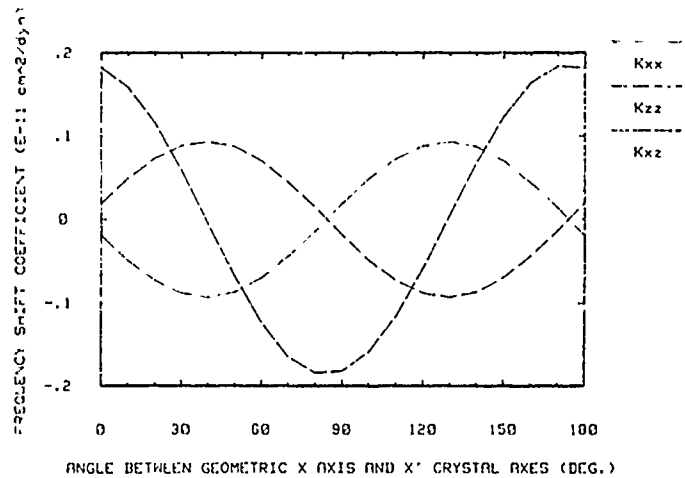


Figure 12. The frequency shift coefficient for the SC-cut for uniaxial T_{xx}, T_{zz}, and T_{xz} vs. angle between the geometric axes and the crystallographic X' axis.

Then,

$$df/f = K\int G^2(X,Y,Z)\ T_e(X,Y,Z)dV + K\int G^2(X,Y,Z)\ X\ T_o(Y,Z)dV. \qquad (3)$$

As we shall see shortly, T_o can be a large stress slope.

The second term in Eq. 3 is nothing more than the usual equation for locating the center of moment $\underline{X}$ of G^2 along the X axis. Thus, in any stress distribution with an odd symmetry along the X axis, the location of the center of the mode $\underline{X}$ along X is a key determining factor for what the gamma vector component value is. Consider Fig. 2b. Where is the center of the mode? It clearly is not at the exact center of the crystal blank.

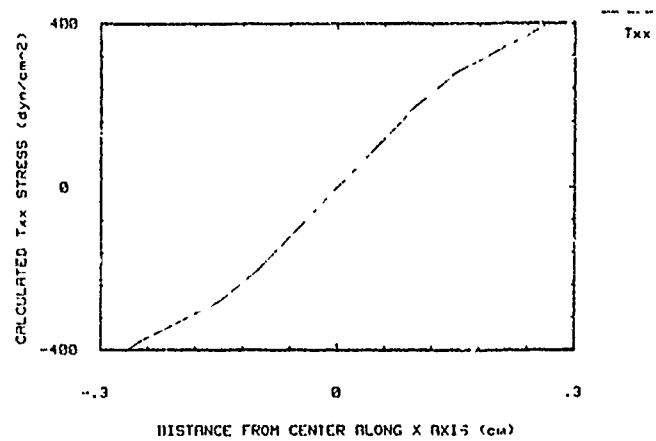


Figure 13. The calculated in-plane stress T_{xx} vs. X for a 10 MHz, 3rd, SC-cut, four-point mount accelerated along the X direction.

Similar conceptual formulation for even and odd symmetry stress along the Z axis (and the Y axis) leads to the same conclusion. Where is the $\underline{X}$, $\underline{Y}$ and $\underline{Z}$ center (net average location) of the mode?

In order to appreciate that this effect is important, we need two more pieces of information: the odd symmetry stress slope magnitude T_o and K. First the stress slope magnitude. Figure 11 shows the stress T_{xz} averaged over the blank thickness as calculated with finite element analysis (isotropic material properties) vs. Z for a 10 MHz, 3rd, SC-cut four-point mount resonator with typical metal strip mounts when it is subjected to an X directed acceleration of 1g. The slope is T_o and is 1600 dyn/cm^{3}g.

Figure 12 shows results of a rigorous formulation for the SC-cut frequency shift coefficients K_{xx}, K_{zz}, and K_{xz} for uniaxial stresses T_{xx}, T_{zz}, and T_{xz} vs. orientation angle between the geometric axis and the X' crystallographic axis per the IEEE standard.

If the resonator is oriented in the worst possible direction, the coefficient K_{xz} is 1.9×10^{-12} cm^2/dyn from Fig. 12. We thus can estimate that this term contributes to df/f as $1600 \times 1.9\times10^{-12}$, or

$$df/f = 3.0\times10^{-9}\ \underline{X}/g, \qquad (4)$$

where $\underline{X}$ is the distance in cm from the geometric center of the blank to the center of the mode. Equation 4 tells us that for every mm that the mode net average center is displaced from the geometric center along the Z axis, we have 3.0×10^{-10}/g of acceleration sensitivity! Alternately, if one desires a sensitivity of 5×10^{-11}/g, the mode, in order to satisfy T_{xz} conditions (T_{xz} is only one of six stress components), must be uniform and centered to 0.16 mm! As it turns out, use of

the X' crystallographic axis along the X geometric axis is the worst choice, yet this orientation is common in the industry.

If the reader thinks one can alleviate the problem by changing the orientation of the mounts to force K_{xz} to zero for T_{xz} stress, consider Fig. 13. Figure 13 shows the stress T_{xx} vs. X for the same resonator again accelerated along the X_3 axis. Here the stress gradient is 2000 dyn/cm^{3}g, and K_{xx} is $1x10^{-12}$ cm^2/dyn from Fig. 13. The result for the worst case of crystallographic orientation is

$$df/f = 2.1x10^{-9} \underline{X}/g. \qquad (5)$$

This means that, for every mm the mode center is displaced along the X axis, a frequency shift effect of $2.1x10^{-10}$/g will occur. Alternately, a $5x10^{-11}$/g specification requires the mode location to be repeatably controlled in production by 0.24 mm.

Since the mode can be off-center in either the X or Z direction, or any direction with a combination of X and Z component, and since the zero crossings for the three frequency shift coefficients in Fig. 12 are not coincident, no one choice for mount orientation can be made for zero gamma. In addition, we have to deal with Z acceleration and Y acceleration, so the number of variables becomes burdensome.

An accompanying paper[11] describes one method for determining the stress patterns in a resonator subjected to acceleration. That work shows the variety of stress patterns that occur. Other analysis techniques, such as finite element, are becoming available. Since the mounting structure will never be perfectly symmetric, the detailed analysis of stress patterns can not treat all possible situations in a manufactured lot of resonators, but can only be used as a guide for resonator design.

The mode shapes in Fig. 2b, and the vibration energy leaking out along dislocations in Fig. 2a, suggest that the effective mode center can be most anywhere near, but not likely at, the geometric center of the blank. Besides, even if the mode were perfect, the strain pattern is never perfectly centered because of perturbations in mount location and mount stiffness.

The real world case in a given resonator is much more complicated than what has been presented here. The conceptual theory presented here is only meant to add credibility to the hypothesis that mode shape and location are critical factors in determining the acceleration sensitivity of manufactured resonators. Even though the two in-plane accelerations along X and Z are used for the theoretical discussion, Figs. 4, 5a, 5b, 7, 8, and 10 show that mode location affects the normal acceleration along Y.

Another important feature appears in Eq. 1b. The even symmetry term is independent of mode center location. Therefore, there may be situations where the mode center can be moved slightly off-center so that the odd symmetry term of Eq. 1b can partially compensate for the even symmetry term in Eq. 1b. Thus, the optimum location for the mode center is not necessarily the geometric center[6,7].

CONCLUSIONS

The present work supports the hypothesis that mode shape and location are not uniform in as-manufactured resonators. Furthermore, it has been shown that mode shape and location influence the acceleration sensitivity of resonators to such an extent that most other considerations are swamped out. It is concluded that the only way a manufacturable, low-acceleration sensitivity resonator can be designed is to allow for "trimming" of the mode shape after mounting. This "trimming" is most easily accomplished by addition of electrode metallization on selected areas of the active region of vibration after the gamma vector has first been measured[6,7]. Iterations of the "trimming" may be necessary to produce resonators that are "on-frequency" while at the same time having small gamma. Finally, the optimum location for the mode may be slightly off-center of the geometric center of the blank[7].

The fact that a few resonators have been observed to have gamma vectors in the 10^{-11}/g range suggests that those resonators have an optimum mode shape and location. The task is to take the manufacturing process from a random one where only few resonators per thousand have 10^{-11}/g gamma vectors to a controllable one where a high percentage have acceleration vectors of 10^{-11}/g. This can be accomplished by using the above mentioned theoretical techniques to optimize the resonator blank/mount design, and the use of trimming to make each resonator more "perfect".

ACKNOWLEDGEMENTS

The authors wish to thank Mr. Kevin Kelly for making the measurements described herein.

REFERENCES

[1] R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," Proc. 41st Annual Symposium on Frequency Control, 1987, p. 398.

[2] R.L. Filler, et. al, "Further Studies On The Acceleration Sensitivity of Quartz Resonators," Proc. 37th Annual Symposium on Frequency Control, 1983, p. 265.

[3] M. H. Watts, et.al, "Techniques For Measuring The Acceleration Sensitivity of SC-Cut Quartz Resonators," Proc. 42nd Annual Symposium on Frequency Control, 1988, p. 442.

[4] H.F. Tiersten and D.V. Shick, "An Analysis Of The Normal Acceleration Sensitivity Of Contoured Quartz Resonators Rigidly Supported Along the Edges," Proc. 42nd Annual Symposium on Frequency Control, 1988, p. 230.

[5] P.C.Y. Lee, Private Communication.

[6] E.P. EerNisse and R.W. Ward, "Crystal Resonator With Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent Pending.

[7] E. P. EerNisse, et.al., "Improved Crystal Resonator With Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent Pending.

[8] A. Zarka, et.al., "Studies of Quartz Resonators By Stroboscopic Topography," Proc. 41st Symposium on Frequency Control, 1987, p. 236.

[9] J.R. Vig, et.al., "Acceleration Sensitivity Reduction Method," U.S. Patent No. 4,451,755, May 29, 1984.

[10] H.F. Tiersten, "Perturation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 1978, p. 832.

[11] E.P. EerNisse, et.al., "Variational Method For Modeling Static and Dynamic Stresses In A Resonator Disc With Mounts," Proc. 43rd Symposium on Frequency Control, 1989, to be published (this Symposium).

43rd Annual Symposium on Frequency Control - 1989

AN ANALYSIS OF THE IN-PLANE ACCELERATION SENSITIVITY OF ST-CUT QUARTZ SURFACE WAVE RESONATORS WITH INTERIOR RECTANGULAR SUPPORTS

H.F. Tiersten and D.V. Shick
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Abstract

An analysis of the in-plane acceleration sensitivity of ST-cut quartz surface wave resonators with the substrate extending beyond interior rectangular supports is performed. In the treatment the variational principle with all natural conditions is extended in such a way as to permit interior surfaces of discontinuity with spring supports, at which the approximating solution functions need not satisfy the constraint-type conditions. The resulting extensional biasing state is employed in the existing perturbation integral along with the proper continuous representation of the acoustic surface wave mode shape to calculate the in-plane acceleration sensitivity for this support configuration. Results are presented as a function of planar aspect ratio for a number of overhang configurations

1. Introduction

It has been shown[1] that vanishingly small normal acceleration sensitivities of ST-cut quartz surface wave resonators can be obtained over a reasonably wide range of planar aspect ratio when the substrate extends beyond rectangular supports, as it does in the configuration employed at Raytheon[2-4]. It has also been shown that although the in-plane acceleration sensitivity for the same configuration does not have a zero-crossing, it can be kept lower than a few parts in 10^{11} per g by appropriate dimensioning. It was also shown[5] that the small amount of flexure induced in the active substrate produces a negligible increase in the in-plane acceleration sensitivity. Accordingly, we ignore this small amount of flexure here.

In this work an analysis of the in-plane acceleration sensitivity of ST-cut quartz surface wave resonators with the substrate extending beyond interior rectangular supports is performed. This is a revised version of the frit support configuration used by Raytheon and is treated in this work to see if the in-plane acceleration sensitivity can be reduced by using this revised configuration. In order to use our new variational approximation procedure[5,6] in the determination of the biasing state, an appropriate extension of an existing variational principle[7,8] is required. Consequently, the existing variational principle with all natural conditions is extended in such a way as to permit interior surfaces of discontinuity with spring supports, at which the approximating solution functions need not satisfy the constraint-type conditions. The extensional biasing deformation field is determined by means of the new variational approximation procedure using the extended version of the variational principle. The resulting extensional biasing state is employed in the existing perturbation integral[9] along with the proper continuous representation of the acoustic surface wave mode shape[10] for arrays of reflecting grooves to calculate the in-plane acceleration sensitivity for this revised support configuration. The calculations reveal that this revised support configuration results in only a slight reduction in the in-plane acceleration sensitivity, but they also indicate how to appropriately select the distances between the supports so as to minimize the in-plane acceleration sensitivity.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the first perturbation of the eigenvalue obtained from the perturbation analysis[9] mentioned in the Introduction may be written in the form

$$\Delta_\mu = H_\mu/2\omega_\mu, \quad \omega = \omega_\mu - \Delta_\mu, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively, and

$$H_\mu = -\int_V \hat{K}^n_{L\gamma} g^\mu_{\gamma,L}\, dV, \tag{2.2}$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g^μ_γ denotes the normalized mechanical displacement vector, and $\hat{K}^n_{L\gamma}$ denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g^μ_γ, and is given by

$$\hat{K}^n_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M}, \tag{2.3}$$

where

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM}\delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN}E^1_{KN} + c_{2L\gamma KM}w_{\alpha,K} + c_{2LKM\alpha}w_{\gamma,K}, \tag{2.4}$$

and

$$T^1_{LM} = c_{2LMKN}E^1_{KN}, \quad E^1_{KN} = \frac{1}{2}(w_{K,N} + w_{N,K}). \tag{2.5}$$

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 $c_{3L\gamma M\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g^μ_γ and $\hat{f}^\mu$ is defined by

$$g^\mu_\gamma = \frac{u^\mu_\gamma}{N_\mu}, \quad \hat{f}^\mu = \frac{\tilde{\varphi}^\mu}{N_\mu}, \quad N^2_\mu = \int_V \rho u^\mu_\gamma u^\mu_\gamma\, dV, \tag{2.7}$$

where u^μ_γ and $\tilde{\varphi}^\mu$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

CH2690-6/89/0000-396 $1.00 C 1989 IEEE

$$\hat{K}^{\ell}_{L\gamma} = c_{2L\gamma M\alpha} u_{\alpha,M} + e_{ML\gamma}\tilde{\varphi}_{,M},$$

$$\tilde{\mathcal{D}}^{\ell}_{L} = e_{LM\gamma} u_{\gamma,M} - \epsilon_{LM}\tilde{\varphi}_{,M}, \qquad (2.8)$$

$$\hat{K}^{\ell}_{L\gamma,L} = \rho\ddot{u}_{\gamma}, \quad \tilde{\mathcal{D}}^{\ell}_{L,L} = 0, \qquad (2.9)$$

subject to the appropriate boundary conditions. The quantities $\hat{K}^{\ell}_{L\gamma}$, $\tilde{\mathcal{D}}^{\ell}_{L}$, $e_{ML\gamma}$ and ϵ_{LM} denote the linear stress tensor, the electric displacement vector and the piezoelectric and dielectric constants, respectively, 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.9, as is the remainder of the notation in this section. The fact that the capital Latin and lower case Greek indices refer to the reference and intermediate positions of material points, respectively, is not important here, and in this work they may be used interchangeably. We employ Cartesian tensor notation, the summation convention for repeated tensor indices, the convention that a comma followed by an index denotes partial differentiation with respect to a reference coordinate, and the dot notation for differentiation with respect to time.

The substitution of (2.3) in (2.2) yields

$$H_{\mu} = -\int_{V} \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} g^{\mu}_{\gamma,L}\, dV. \qquad (2.10)$$

Since g^{μ}_{α} denotes the normalized mode shape in the surface wave resonator and from (2.4) $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_{μ} can be evaluated when the surface wave mode shape and biasing state are known.

3. Formulation of Biasing Problem

A plan view and cross-section of the substrate extending beyond the interior rectangular supports is shown in Figure 1 along with the coordinate system. Since, as noted in the Introduction, the small amount of flexure in the active substrate caused by the offset between the center-plane of the plate and the location of the resisting force at the base of the frit results in a negligible increase in the in-plane acceleration sensitivity[5], it is ignored in the treatment. Nevertheless, the in-plane resistance of the frit must of course, be included, although it is to be treated as acting at the center-plane of the plate. Since the plate is thin, the equations of anisotropic plane stress apply and, as already noted,we take the resisting forces due to the frits to lie in the center-plane of the plate rather than below the plate as shown in the cross-section in Figure 1. For anisotropic plane stress all K_{2L} vanish and only K_{AB} exist, where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2.

Now we note that in the plan view of Figure 1 the outside edges, which are denoted by solid lines, are traction-free and along these edges we have the conditions

$$N_A K_{AB} = 0. \qquad (3.1)$$

On the other hand the interior dotted lines refer to the frits, along which we have support. As a consequence of the geometry of the supporting frit, we now take the displacement w_s in the tangential direction of the supporting frit to vanish, and along these interior lines we have

$$w_s = 0. \qquad (3.2)$$

However, the condition in the direction normal to the supporting frit is another matter. Clearly, in the normal direction the frit exerts a spring type resistance to the displacement w_n of the center-plane of the plate, which we must include in the treatment.

A cross-section of the frit and adjacent portions of the plate are shown in Figure 2 along with the associated geometry and stresses in the plate and the frit. On account of the stubbiness of the cross-section of a frit, we model its resistance to the normal displacement w_n of the center-plane of the plate as a simple variable shear strip, which yields

$$[K_{(n)(n)}] = k w_n, \qquad (3.3)$$

where[11]

$$k = 2\mu t_f / 3h^2, \qquad (3.4)$$

in which μ is the appropriate shear modulus, i.e., either c_{44} or c_{66} for the ST cut depending on the direction in which the frit runs and we have ignored the deformation in the frit because $h_f \ll h$.

Thus we have shown that the boundary conditions are given by (3.1) along the outside edges of the thin plate and by (3.2) and (3.3) for the tangential and normal conditions, respectively, along the interior supporting frits. Along with the edge conditions we have the equations of anisotropic plane stress, which may be written in the form

$$K_{AB,A} - \rho a_B = 0, \qquad (3.5)$$

where a_B denotes the acceleration in the plane of the plate and the anisotropic constitutive equations for plane stress may be written in the form

$$K_{AB} = \gamma_{ABCD} E_{CD}, \quad E_{CD} = \tfrac{1}{2}(w_{C,D} + w_{D,C}), \qquad (3.6)$$

where γ_{ABCD} are Voigt's anisotropic plate elastic constants, which are given by

$$\gamma_{RS} = c_{RS} - c_{RW} c^{-1}_{WV} c_{VS}; \quad R,S = 1,3,5; \quad W,V = 2,4,6. \qquad (3.7)$$

From the relaxation of the stresses K_{2L}, we have

$$E_W = -c^{-1}_{WV} c_{VS} E_S, \qquad (3.8)$$

which enables us to obtain all the strains E_{KL} when a solution is known as well as the rotation Ω_{13} from the relation

$$\Omega_{13} = \tfrac{1}{2}(w_{3,1} - w_{1,3}). \qquad (3.9)$$

However, the rotations Ω_{2A} cannot be obtained from a solution because of the relaxation of the K_{2L}. Nevertheless, the flexural type plate rotations Ω_{2A} are taken to vanish in the case of uncoupled anisotropic plane stress, which is the kinematic uncoupling assumption for anisotropic plane stress. Accordingly, we have

$$\Omega_{2A} = 0, \qquad (3.10)$$

and now with (3.6) - (3.10) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \tag{3.11}$$

we have the required three-dimensional displacement gradients for the biasing deformation when a solution has been found.

Since it is impossible to solve the problem that has been posed exactly, an approximation procedure must be employed. As noted in the Introduction, in order to use our very accurate, computationally efficient and convenient variational approximation procedure, an extension of the variational principle with all natural conditions is required. This extension is provided in the next section.

4. Unconstrained Variational Principle with Interior Spring Supports

In this section we present the unconstrained variational principle for the static extension of thin plates when spring supports are present in the interior of the plate. This principle is required for the determination of the extensional biasing state for the problem posed in Section 3 using the approximation procedure we employ. Before proceeding with the variational principle we consider it advisable for clarity to discuss the plan view of the plate shown in Figure 3. The outside edge of the plate consists of a curve with unit outward normal N_A and the entire closed circuit is denoted by C. Some portions of the outside 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 $\bar{w}_A$ prescribed and are denoted C^C and the bar denotes that a quantity is prescribed. The portions of the edge which are subject to prescribed loadings $\bar{t}_A$ are denoted C^N. Thus C^N and C^C denote the portions of the edge on which natural- and constraint-type conditions, respectively, are prescribed and we note that C^N and C^C can refer to different portions of the surface for different terms in the edge conditions depending on the actual condition at a point. The interior curve labeled C^d denotes the edge along which the spring support acts. Since the edge C^d represents a spring, there are no prescribed conditions along the edge which stores internal energy quadratic in the displacement w_A of the edge. However, since the continuity of w_A across C^d would be a constraint condition and we want the variational principle to have all natural conditions, i.e., to be unconstrained, we express the energy of the spring in terms of the mean displacement w_A, where

$$w_A^d = \frac{1}{2}\,(w_A^{(2)} + w_A^{(1)}) \tag{4.1}$$

and we note that if w_A is continuous across C^d, we have

$$w_A^{(2)} = w_A^{(1)} = w_A^d. \tag{4.2}$$

Thus, the in-plane spring energy takes the general form

$$U^d = \frac{1}{2}\,k_{AB} w_A^d w_B^d = \frac{1}{8}\,k_{AB}(w_A^{(2)} + w_A^{(1)})(w_B^{(2)} + w_B^{(1)}), \tag{4.3}$$

where for the case of interest k_{AB} may be obtained from Section 3. 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[7,8] in the variational principle. Accordingly, the variational principle for the static extension of thin plates with an in-plane interior spring support may be written in the form

$$\delta \sum_{m=1}^{2} \Big[\int_{S^{(m)}} \left(\bar{f}_A^{(m)} w_A^{(m)} - U^{(m)} \right) dS + \int_{C^{(m)N}} \bar{t}_A^{(m)} w_A^{(m)}\, ds + \int_{C^{(m)C}} \lambda_A^{(m)} \left(w_A^{(m)} - \bar{w}_A^{(m)} \right) ds \Big] + \delta \Big[\int_{C^{(d)}} -U^{(d)}\, ds + \int_{C^{(d)}} \lambda_A^{(d)} \left(w_A^{(2)} - w_A^{(1)} \right) ds \Big] = 0, \tag{4.4}$$

where $f_A^{(m)}$ denotes an applied body force per unit area of the mth region, $S^{(m)}$ denotes the area of the mth region of the plate

$$U^{(m)} = \frac{1}{2}\,\gamma_{ABCD}^{(m)} E_{AB}^{(m)} E_{CD}^{(m)}, \tag{4.5}$$

and $\lambda_A^{(m)}$ and $\lambda_A^{(d)}$ are Lagrange undetermined multipliers.

Taking the variations in (4.4), employing (4.5), (4.3) and (3.6) and using the surface divergence theorem[12], we obtain

$$\sum_{m=1}^{2} \Big[\int_{S^{(m)}} [(K_{AB,A}^{(m)} + \bar{f}_B^{(m)})\,\delta w_B^{(m)}]\, dS + \int_{C^{(m)N}} [(\bar{t}_B^{(m)} - N_A^{(m)} K_{AB}^{(m)})\,\delta w_B^{(m)}]\, ds + \int_{C^{(m)C}} [(\lambda_B^{(m)} - N_A^{(m)} K_{AB}^{(m)})\,\delta w_B^{(m)} + \delta\lambda_B^{(m)} (w_B^{(m)} - \bar{w}_B^{(m)})]\, ds \Big] + \int_{C^{(d)}} \Big[-(\lambda_B^{(d)} + N_A^{(d)} K_{AB}^{(1)})\,\delta w_B^{(1)} + (\lambda_B^{(d)} + N_A^{(d)} K_{AB}^{(2)})\,\delta w_B^{(2)} + \delta\lambda_B^{(d)} (w_B^{(2)} - w_B^{(1)}) - \frac{1}{4} k_{AB} (w_A^{(1)} + w_A^{(2)})(\delta w_B^{(1)} + \delta w_B^{(2)}) \Big]\, ds = 0, \tag{4.6}$$

where $N_A^{(m)}$ and $N_A^{(d)}$ denote the unit normals outwardly directed to the outside edge of the mth region and directed from region 1 to region 2 across $C^{(d)}$, respectively. Since on account of the use of Lagrange multipliers with the constraint conditions and (4.3) all variations in (4.6) are arbitrary, we obtain the differential equations (3.5) for the mth region[11] with $\bar{f}_B = -\rho a_B$ the natural edge conditions on $C^{(m)N}$

$$N_A K_{AB} = \bar{t}_B^{(m)}, \tag{4.7}$$

the natural form of the constraint-type edge conditions on $C^{(m)C}$ and $C^{(d)}$

$$w_B^{(m)} = \bar{w}_B^{(m)}, \quad w_B^{(2)} = w_B^{(1)}, \tag{4.8}$$

the Lagrange multipliers[11]

$$\lambda_B^{(m)} = N_A^{(m)} K_{AB}^{(m)} \text{ on } C^{(m)C}, \tag{4.9}$$

$$\lambda_B^{(d)} = -\frac{1}{2}\,N_A^{(d)} (K_{AB}^{(1)} + K_{AB}^{(2)}), \tag{4.10}$$

and with the use of $(4.8)_2$ the natural conditions[11]

$$N_A^{(d)}(K_{AB}^{(2)} - K_{AB}^{(1)}) - k_{BA}w_A^{(1)} = 0, \qquad (4.11)$$

across $C^{(d)}$. Thus it is clear that the variational principle in (4.4) with unconstrained variations yields the differential equations (3.5), the natural edge conditions (4.7) and (4.11) on $C^{(m)N}$ and $C^{(d)}$, respectively, and the constraint-type edge conditions $(4.8)_1$ and $(4.8)_2$ on $C^{(m)C}$ and $C^{(d)}$, respectively. Hence, the variational principle (4.4) yields the required differential equations and boundary conditions governing the problem. In addition the Lagrange multipliers have been expressed in terms of the field variables in (4.9) and (4.10). Consequently, the variations $\delta\lambda_B^{(m)}$ and $\delta\lambda_B^{(d)}$ may be obtained from (4.9) and (4.10) and substituted[11] in (4.6), which may then be used for obtaining an approximate solution for the extensional biasing state when spring supports are present in the interior of a plate without any a priori conditions on the approximating functions.

5. Extensional Biasing Deformation

The problem of interest was formulated in Section 3 where it was noted that the problem could not be solved exactly and an approximation procedure was required. Since the variational equation required for the approximation procedure we employ is given in (4.6), we present the approximate solution in this section. We now have recourse to the plan view of the plate shown in Figure 4, in which the five distinct regions required for the solution presented in this section are denoted L, R, T, C and B, respectively. The L and R regions are the entire regions to the left and right of the left and right vertical dotted lines, respectively, and T, C and B refer to the top, center and bottom portions of the central region between the two vertical dotted lines.

Substituting from (3.6) into (3.5), we obtain

$$\gamma_{ABCD}w_{C,DA} - \rho a_B = 0, \qquad (5.1)$$

which constitute the differential equations in each of the five regions. From Eqs.(3.1), (3.2) and (3.3) the edge conditions for the problem take the form

$$K_{11} = 0,\ K_{13} = 0 \quad \text{at } X_1 = \pm a_\ell,\ |X_3| < b_\ell, \qquad (5.2)$$

$$K_{31} = 0,\ K_{33} = 0 \quad \text{at } X_3 = \pm b_\ell,\ |X_1| < a_\ell, \qquad (5.3)$$

$$w_3 = 0,\ [\underset{\sim}{K}_{11}] = \pm k_1 w_1 \quad \text{at } X_1 = \pm a,\ |X_3| < b, \qquad (5.4)$$

$$w_1 = 0,\ [\underset{\sim}{K}_{33}] = \pm k_3 w_3 \quad \text{at } X_3 = \pm b,\ |X_1| < a. \qquad (5.5)$$

In obtaining the solution to this problem we first transform the inhomogeneities from the differential equations (5.1) in each region into the boundary conditions (5.2) - (5.5) in such a way that certain of the conditions and certain portions of other conditions in (5.2) - (5.5) remain homogeneous. The particular choice of the transformation is strongly influenced by the regions defined in Figure 4. To this end we write

$$\tilde{w}_A = \hat{\tilde{w}}_A + A_{(A)}\tilde{f}_A(X_3), \qquad (5.6)$$

where $\sim$ denotes the regions L, T, C, B and R defined in Figure 4. The $\sim$ notation is used here because the detailed equations are too cumbersome to write[11]. The coefficients A_1 and A_3 are selected as

$$A_1 = \frac{\rho a_1}{2\gamma_{55}}, \quad A_3 = \frac{\rho a_3}{2\gamma_{33}}, \qquad (5.7)$$

and substituting into (5.1) for ST-cut quartz, for which $\gamma_{15} = \gamma_{35} = 0$, we obtain the homogeneous equations

$$\gamma_{11}\hat{\tilde{w}}_{1,11} + (\gamma_{13} + \gamma_{55})\hat{\tilde{w}}_{3,13} + \gamma_{55}\hat{\tilde{w}}_{1,33} = 0,$$
$$\gamma_{55}\hat{\tilde{w}}_{3,11} + (\gamma_{13} + \gamma_{55})\hat{\tilde{w}}_{1,13} + \gamma_{33}\hat{\tilde{w}}_{3,33} = 0, \qquad (5.8)$$

in each of the five regions with numerous edge and continuity conditions on the $\hat{\tilde{w}}_A$.

As a solution of (5.8) for each of the five regions take

$$\hat{\tilde{w}}_A = \sum_m \tilde{\beta}_{Am} e^{i(\tilde{\nu}_m X_1 + \tilde{\eta}_m X_3)}, \qquad (5.9)$$

where $\tilde{\nu}_m = m\pi/2\sigma a$ for regions T, C and B and $\tilde{\nu}_m = m\pi/2\sigma(a_\ell - a)$ for regions L and R and m takes all positive and negative integers considered in the sum, and σ is a parameter which must be an irrational number so that neither sin nor cos vanish at the appropriate edges $X_1 = \pm a$ or $X_1 = \pm a_\ell$ and can be chosen to improve convergence. It has been found that $\sigma = \pi$ works well. The substitution of (5.9) into (5.8) yields

$$(\gamma_{11}\tilde{\nu}_m^2 + \gamma_{55}\tilde{\eta}_m^2)\tilde{\beta}_{1m} + (\gamma_{13} + \gamma_{55})\tilde{\nu}_m\tilde{\eta}_m\tilde{\beta}_{3m} = 0,$$
$$(\gamma_{13} + \gamma_{55})\tilde{\nu}_m\tilde{\eta}_m\tilde{\beta}_{1m} + (\gamma_{55}\tilde{\nu}_m^2 + \gamma_{33}\tilde{\eta}_m^2)\tilde{\beta}_{3m} = 0, \qquad (5.10)$$

for each m. Equations (5.10) constitute a system of two linear homogeneous algebraic equations in $\tilde{\beta}_{1m}$ and $\tilde{\beta}_{3m}$, which yields nontrivial solutions when the determinant of the coefficients vanishes. By scaling this is independent of m. The resulting algebraic equation yields four roots for each $\tilde{\nu}_m$ and corresponding sets of amplitude ratios $\tilde{\beta}_1^{(p)}:\tilde{\beta}_3^{(p)}$, which are independent of m. This holds for all regions. Hence, as a solution of the boundary value problem we take[11]

$$\tilde{w}_A = \sum_m \sum_{p=1}^{4} \tilde{D}_p^{(m)}\tilde{\beta}_A^{(p)} e^{i(\tilde{\eta}_m^{(p)}X_3 + \tilde{\nu}_m X_1)}, \qquad (5.11)$$

where the $\tilde{D}_p^{(m)}$ are amplitude coefficients still to be determined. By satisfying the 11 homogeneous edge conditions for each m in the T, C, B region, we obtain[11]

$$D_p^{(m)} = g_p^{(m)}D_1^{(m)},\ D_p^{T(m)} = g_p^{T(m)}D_1^{(m)},$$
$$D_p^{B(m)} = g_p^{B(m)}D_1^{(m)}, \qquad (5.12)$$

which reduces the number of unknown coefficients in the central region from 12 to 1 for each m. Similarly, by satisfying the 3 homogeneous edge conditions in the L and R regions, we obtain[11]

$$D_p^{R(m)} = \tilde{g}_p^{(m)}D_1^{R(m)},\ D_p^{L(m)} = \tilde{g}_p^{(m)}D_1^{R(m)}, \qquad (5.13)$$

and we note that the known $\tilde{g}_p^{(m)}$ depend on whether a_1 or a_3 is zero[11]. The substitution of (5.12) and (5.13) into (5.11) yields the reduced form

$$\hat{w}_A^C = \sum_m \Big(\sum_{p=1}^4 \beta_A^{(p)} g_p^{(m)} e^{i\eta_m^{(p)} X_3} \Big) D_1^{(m)} e^{i\nu_m X_1},$$

$$\hat{w}_A^T = \sum_m \Big(\sum_{p=1}^4 \beta_A^{(p)} g_p^{T(m)} e^{i\eta_m^{(p)} X_3} \Big) D_1^{(m)} e^{i\nu_m X_1},$$

$$\hat{w}_A^B = \sum_m \Big(\sum_{p=1}^4 \beta_A^{(p)} g_p^{B(m)} e^{i\eta_m^{(p)} X_3} \Big) D_1^{(m)} e^{i\nu_m X_1},$$

$$\hat{w}_A^R = \sum_m \Big(\sum_{p=1}^4 \beta_A^{(p)} \tilde{g}_p^{(m)} e^{i\tilde{\eta}_m^{(p)} X_3} \Big) D_1^{R(m)} e^{i\tilde{\nu}_m X_1},$$

$$\hat{w}_A^L = \sum_m \Big(\sum_{p=1}^4 \beta_A^{(p)} \tilde{g}_p^{(m)} e^{i\tilde{\eta}_m^{(p)} X_3} \Big) D_1^{L(m)} e^{i\tilde{\nu}_m X_1}, \qquad (5.14)$$

for the approximating solution functions.

The solution functions in (5.14) with (5.6) satisfy the differential equations (5.1) and many of the boundary conditions in (5.2) - (5.5), but they do not satisfy all of the boundary conditions. In addition, they do not satisfy certain continuity conditions[11] not contained in (5.2) - (5.5). We satisfy the presently unsatisfied boundary and continuity conditions approximately by means of a variational approximation procedure, which has been used in earlier work[5,6]. Since the solution (5.14) with (5.6) satisfies the differential equations (5.1) exactly and all exterior edge conditions are traction free and the displacements $\bar{w}_3$ vanish at $X_1 = \pm a$, all that remains of the variational principle given in (4.6) with (4.9) and (4.10) is

$$\sum_{m=1}^{2} \Big[- \int_{C^{(m)}N} N_A^{(m)} K_{AB}^{(m)} \delta w_B^{(m)} \, ds + \oint_{C^{(m)}C} N_A^{(m)} w_B^{(m)} \delta K_{AB}^{(m)} \, ds \Big] + \oint_{C^{(d)}} \frac{1}{2} \Big(N_A^{(d)} (K_{AB}^{(2)} - K_{AB}^{(1)}) - \frac{1}{2} k_{AB} (w_A^{(1)} + w_A^{(2)}) \Big) (\delta w_B^{(1)} + \delta w_B^{(2)}) + N_A^{(d)} (w_B^{(1)} - w_B^{(2)}) (\delta K_{AB}^{(1)} + \delta K_{AB}^{(2)}) \Big] ds = 0, \qquad (5.15)$$

where the constraint-type condition in (5.15) arises because in the application at hand there is a tangential constraint condition along both sides of $C^{(d)}$ by virtue of (3.2). Substituting from (5.6) and (5.14) into (5.15), employing (3.6) and (5.2) - (5.5) and performing the integrations, we obtain[11]

$$\sum_m \sum_n a_{mn} \delta D^{(m)} D^{(n)} + \sum_m \sum_n \bar{a}_{mn}^R \delta \bar{D}^{R(m)} \bar{D}^{R(n)} + \sum_m \sum_n \bar{a}_{mn}^L \delta \bar{D}^{L(m)} \bar{D}^{L(n)} + \sum_m \sum_n a_{mn}^R \delta D^{(m)} \bar{D}^{R(n)} + \sum_m \sum_n a_{mn}^L \delta D^{(m)} \bar{D}^{L(n)} + \sum_m \sum_n a_{nm}^R \delta \bar{D}^{R(m)} D^{(n)} + \sum_m \sum_n a_{nm}^L \delta \bar{D}^{L(m)} D^{(n)} + \sum_m b_m \delta D^{(m)} + \sum_m \bar{b}_m^R \delta \bar{D}^{R(m)} + \sum_m \bar{b}_m^L \delta \bar{D}^{L(m)} = 0, \qquad (5.16)$$

where we have taken the liberty of dropping the subscript 1 on the D.

Since the variations in (5.16) are arbitrary, we obtain

$$\sum_{n=-N}^{N} (a_{mn} D^{(n)} + a_{mn}^R \bar{D}^{R(n)} + a_{mn}^L \bar{D}^{L(n)}) = - b_m,$$

$$\sum_{n=-N}^{N} (a_{nm}^R D^{(n)} + \bar{a}_{mn}^R \bar{D}^{R(n)} = - \bar{b}_m^R,$$

$$\sum_{n=-N}^{N} (a_{nm}^L D^{(n)} + \bar{a}_{mn}^L \bar{D}^{L(n)}) = - \bar{b}_m^L, \quad m = \pm 1, \pm 2, \ldots \pm N \qquad (5.17)$$

which constitute 6N inhomogeneous linear algebraic equations in 6N unknowns $D^{(m)}$, $\bar{D}^{L(m)}$ and $\bar{D}^{R(m)}$. Convergence is determined by increasing the number N and comparing the solution for N with that for N - 1. In this case good convergence was obtained for N = 16.

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[10]. The straight-crested surface wave displacement field may be written in the known form[13,14]

$$u_j = \alpha_j(X_2) e^{i\xi(X_1 - Vt)}, \quad \alpha_j = \sum_{m=1}^{4} C^{(m)} A_j^{(m)} e^{i\beta_m \xi X_2}. \qquad (6.1)$$

A plan view of the resonator showing the reflecting arrays of grooves, the coordinate system and the associated planar geometry is shown in Figure 1, 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[10] that the variable-crested resonant surface wave mode shape with variable amplitude along the transmission path is very accurately approximated by

$$u_j = \cos \frac{\pi X_3}{2w} \, \mathrm{Re} \Big[\alpha_j(X_2) \hat{C}^R(X_1) e^{i\xi(X_1 - s)} + \alpha_j^*(X_2) \hat{C}^L(X_1) e^{-i\xi(X_1 - s)} \Big] e^{-i\omega t}, \qquad (6.2)$$

where the variations along the transmission path are given by

$$\hat{C}^R(X_1) = \Big[r_1 e^{-\alpha_1 N} e^{-\beta_2 X_1} - r_2 e^{-\alpha_2 N} e^{-\beta_1 X_1} \Big] \bar{C}/d,$$

$$\hat{C}^L(X_1) = r_1 r_2 \Big[e^{-\alpha_1 N} e^{-\beta_2 X_1} - e^{-\alpha_2 N} e^{-\beta_1 X_1} \Big] \bar{C}/d. \qquad (6.3)$$

In (6.3) $\bar{C}$ denotes the amplitude of the input wave, N denotes the number of grooves,

$$d = r_1 e^{-\alpha_1 N} - r_2 e^{-\alpha_2 N}, \quad \beta_{1,2} = \alpha_{1,2} N/(\ell - s), \qquad (6.4)$$

and $\alpha_{1,2}$ and $r_{1,2}$ are given in Eqs.(32) and (34) of Ref.10. The standing surface wave mode, i.e., the part multiplying cos ωt, along the transmission path is used in the perturbation integral (2.10).

7. Acceleration Sensitivity

From Sec.6 we now know g_γ^μ and from Sec.5 we know $\hat{c}_{L\gamma M\alpha}$ for in-plane acceleration for this configuration. Hence, we can now evaluate H_μ in Eq.(2.10). Such calculations have been performed using the known values of the second order[15] and third order[16] elastic constants of quartz. For acceleration in the propagation direction the calculated in-plane sensitivity is plotted in Figure 5 as a function of aspect ratio a/b for values of b of 2.5 and 5.0 mm. For acceleration normal to the propagation direction the calculated in-plane sensitivity for the same configurations is plotted in Figure 6. The curves in Figures 5 and 6 for b = 5.0 mm are for the resonator and support geometry treated in Ref.5 and a comparison with Figures 4 and 6, respectively, of that work reveals that they are virtually indistinguishable. This shows that the somewhat crude assumptions made in that work are adequate. The curves in Figures 5 and 6 for b = 2.5 mm are for the resonator treated in Ref.5 but for the b dimension used in Ref.4. The curves show that the larger b results in a significantly smaller acceleration sensitivity. Figures 7 and 8 are similar curves, respectively, but are for the resonator treated in Ref.4 and encompasses the support geometry used in Ref.4, which is indicated by the dots shown in the figures. It can be seen from the curves that the acceleration sensitivities are considerably higher in Figures 7 and 8 than in Figures 5 and 6, respectively. There are essentially two reasons for this fact. One is that the aperture is smaller for Figures 5 and 6 than for Figures 7 and 8, and the other is that the planar Substrate dimensions are relatively larger with respect to the aperture for Figures 5 and 6 than for Figures 7 and 8.

The curves shown in Figures 9 and 10 are for the resonator used in Ref.4 with fixed overhang dimensions and different fixed values of b. Each figure shows that the acceleration sensitivity decreases as b increases, but that the effect begins to saturate as the value of b is increased beyond a certain point. The value of b used in Ref.4 is the smallest shown and yields the highest acceleration sensitivity. The figures show that the use of an appropriately larger value of b would result in an appreciable reduction in the in-plane acceleration sensitivity. Figure 11 shows the resultant in-plane acceleration sensitivity obtained from Figures 9 and 10 and reveals the same information in a more useful form. The curves in Figures 12 and 13 show the influence of the surface wave aperture on the in-plane sensitivity for acceleration normal to the propagation direction for two different values of b and are as expected from the foregoing discussion.

The results of this investigation along with that of Ref.1 indicate how the support and substrate dimensions for a given surface wave resonator should be determined for low acceleration sensitivity. First, the frit location should be set for (near) minimum in-plane acceleration sensitivity and then the overhang dimensions should be determined for zero normal acceleration sensitivity.

Acknowledgement

This work was supported in part by the Army Research Office under Contract No. DAAL03-88-K-0065.

References

1. H.F. Tiersten and D.V. Shick, "An Analysis of the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Supports," Proceedings of the 2nd European Frequency and Time Forum, EFTF88, 753 (1988).
2. G.K. Montress, T.E. Parker and J. Callerame, "A Miniature Hybrid Circuit SAW Oscillator Using an All Quartz Packaged Resonator," 1985 Ultrasonics Symposium Proceedings, IEEE Cat. No. 85CH2209-5, Institute of Electrical and Electronics Engineers, New York, 277 (1985).
3. J.A. Greer, T.E. Parker, M. Rothschild and D.J. Ehrlich, "A Novel Technique for Trimming the Frequency of a Sealed Surface Acoustic Wave Resonator," Proceedings of the 41st 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. 87CH2427-3, 351 (1987).
4. J.A. Greer and T.E. Parker, "Improved Vibration Sensitivity of the All Quartz Package Surface Acoustic Wave Resonator," Proceedings of the 42nd 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. 88CH2588-2, 239 (1988).
5. D.V. Shick and H.F. Tiersten, "An Analysis of the In-plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with the Substrate Extending Beyond the Supports," Proceedings of the 42nd 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. 88CH2588-2, 220 (1988).
6. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along Rectangular Edges," J. Appl. Phys., 64, 4334 (1988).
7. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., 9, 1445 (1968).
8. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4.
9. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
10. H.F. Tiersten, J.T. Song and D.V. Shick, "A Continuous Representation of the Acoustic Surface Wave Mode Shape in Arrays of Reflecting Grooves," J. Appl. Phys., 62, 1154 (1987).
11. For more detail see H.F. Tiersten and D.V. Shick, "On the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
12. L. Brand, Vector and Tensor Analysis (Wiley, New York, 1967), p.222.
13. 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).
14. B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided by Thin Films," J. Appl. Phys., 44, 4830 (1973). Sec.II.
15. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

16. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third-Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

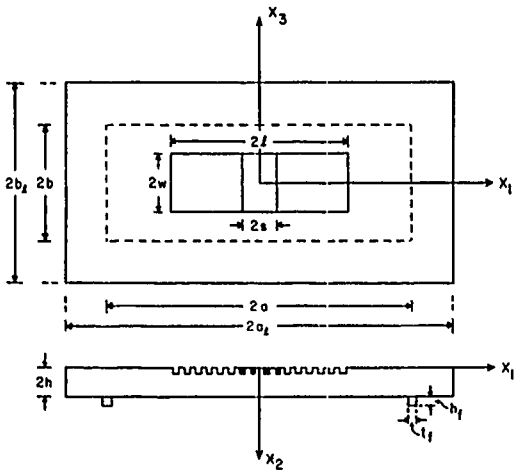


Figure 1 Plan View and Cross-Section of Rectangular ST-Cut Quartz Plate with the Substrate Extending Beyond Interior Rectangular Supports

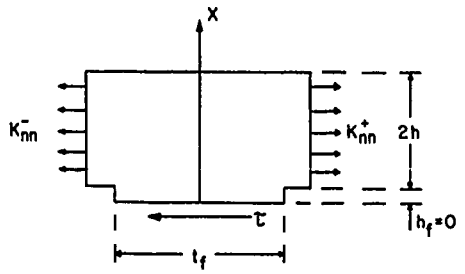


Figure 2 Cross-Section Showing Resisting Shearing Stresses at the Base of a Frit and Extensional Stresses in the Plate

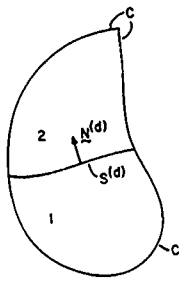


Figure 3 Plan View of a Typical Plate with an Arbitrary Curved Edge and an Interior Spring Support

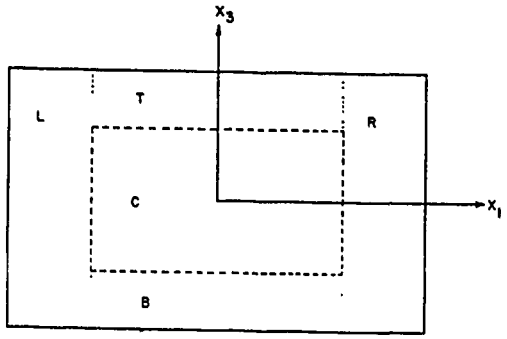


Figure 4 Plan View of the Plate Denoting Regions for the Analysis of the Biasing State

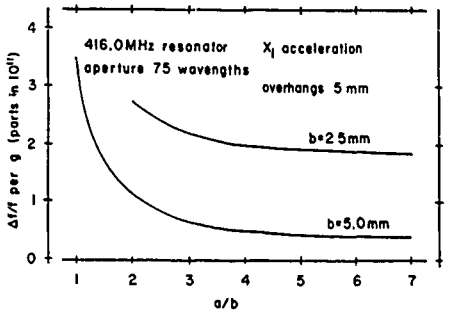


Figure 5 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio a/b for Acceleration in the Propagation Direction for Resonator of Ref.5

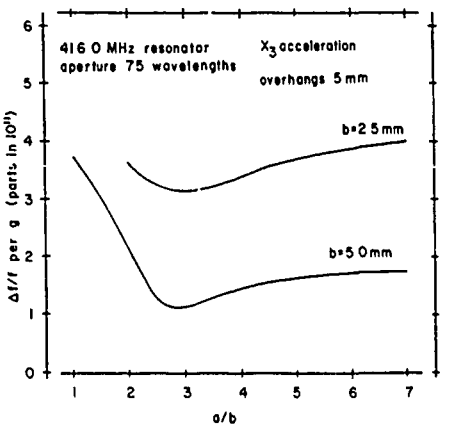


Figure 6 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio for Acceleration Normal to the Propagation for Resonator of Ref.5

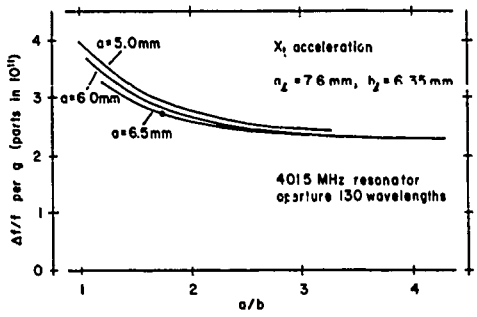


Figure 7 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio for Acceleration in the Propagation Direction for Resonator of Ref.4, Raytheon Resonator and Substrate, Varying Support Dimensions

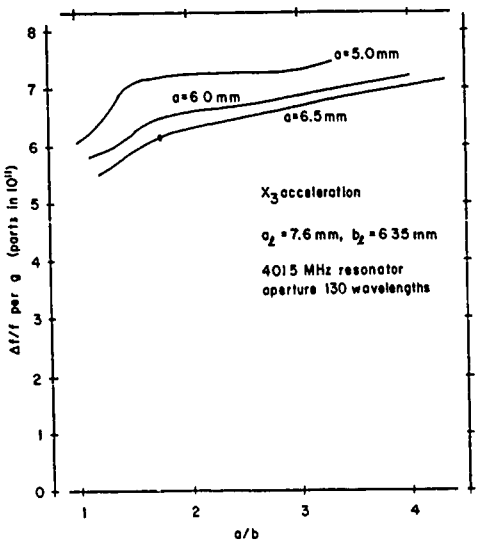


Figure 8 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio for Acceleration Normal to the Propagation Direction for Resonator of Ref.4, Raytheon Resonator and Substrate, Varying Support Dimensions

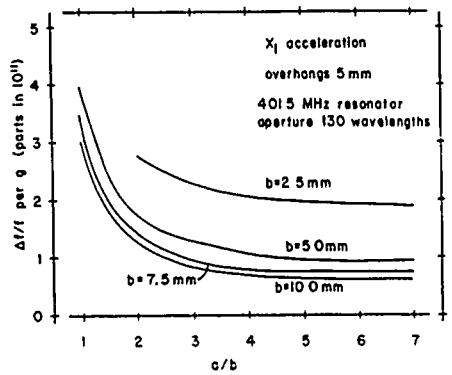


Figure 9 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio a/b for Acceleration in the Propagation Direction for Resonator of Ref.4 and Different Support Dimensions

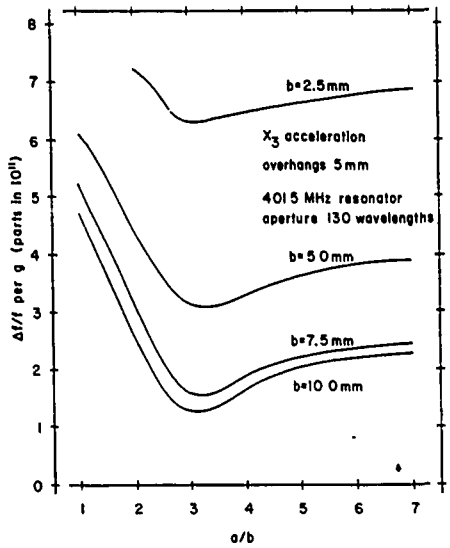


Figure 10 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio a/b for Acceleration in the Propagation Direction for Resonator of Ref.4 and Different Support Dimensions

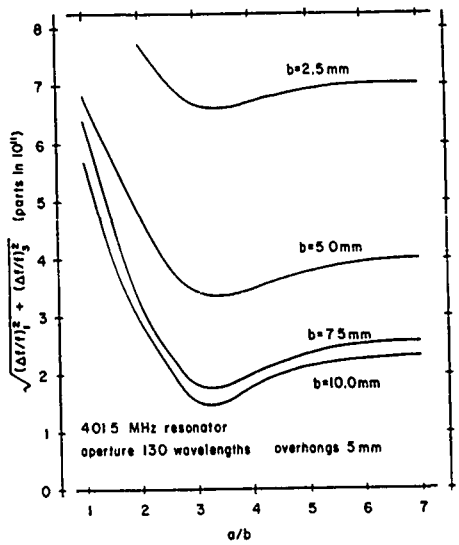


Figure 11 Resultant In-Plane Acceleration Sensitivity Versus Planar Aspect Ratio Obtained from Figures 9 and 10

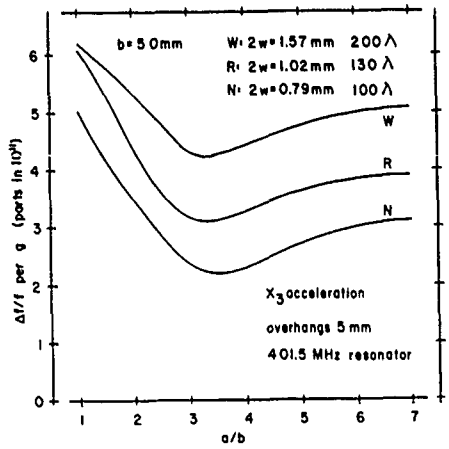


Figure 12 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio for Acceleration in the Propagation Direction for Different Resonator Apertures

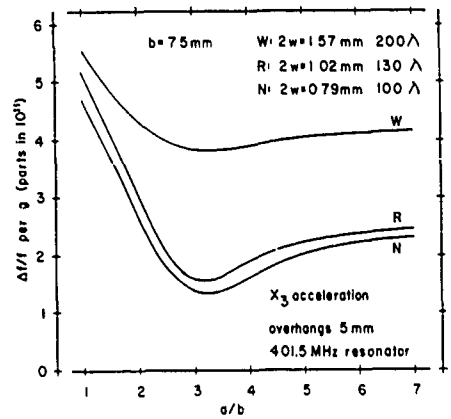


Figure 13 Calculated In-Plane Sensitivity Versus Planar Aspect Ratio for Acceleration Normal to the Propagation Direction for Different Resonator Apertures

43rd Annual Symposium on Frequency Control - 1989

AN ANALYSIS OF THE IN-PLANE ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS RIGIDLY SUPPORTED ALONG THE EDGES

D.V. Shick and H.F. Tiersten
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Abstract

It has recently been shown that the normal acceleration sensitivity of contoured quartz resonators rigidly supported along rectangular edges vanishes for certain values of the planar aspect ratio for all modes considered in both the AT and SC cuts. Consequently, an analysis of the in-plane acceleration sensitivity of contoured AT- and SC-cut quartz crystal resonators with the same support configuration is performed. The extensional biasing deformation field is determined by means of the new variational approximation procedure using the variational principle in which all conditions appear as natural conditions. The resulting extensional biasing state is employed in the existing perturbation equation along with the mode shapes of the contoured resonators to calculate the in-plane acceleration sensitivities. Results are presented as a function of the planar aspect ratio and orientation of rectangle.

1. Introduction

Since the mounting clips commonly used to support contoured resonators do not yield consistent acceleration sensitivities[1], an alternative support system has been examined[2] with which we feel it should be easier to obtain consistent results. The alternative configuration consists of contoured resonators rigidly supported along rectangular edges. The idealized case of rigid supports is treated because it simplifies the analysis while still providing the most essential information and the rectangular configuration enables the influence of the anisotropy to be investigated simply by varying the aspect ratio and orientation of the rectangle.

Since the normal acceleration sensitivity of this configuration has been shown[2] to vanish for certain values of the planar aspect ratio for all modes considered in both the AT and SC cuts, in this work an analysis of the in-plane acceleration sensitivity of contoured resonators with this support configuration is performed. The biasing deformation fields are calculated by means of the new variational approximation procedure, which has been used in earlier work[3,4]. The calculated biasing deformation fields are employed in an existing perturbation equation[5] along with the equivalent trapped energy mode shapes[6] for the contoured resonator to calculate the in-plane acceleration sensitivities. It is shown that each component of the in-plane acceleration sensitivity vanishes for certain planar aspect ratios for a range of orientations of the rectangle for both the AT and SC cuts. In particular, for a given aspect ratio an orientation of an AT cut is found for which the resultant in-plane acceleration sensitivity vanishes. The investigation further reveals that the range of acceleration sensitivities calculated for the SC cut is almost an order of magnitude larger than that for the AT cut.

2. Preliminary Considerations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency[5] may be written in the form given in Eq.(2.1) of the previous paper (Ref.7), which we reproduce here for completeness

$$\Delta_\mu = H_\mu/2\omega_\mu, \quad \omega = \omega_\mu - \Delta_\mu, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively. It has further been shown that H_μ may be written in the form given in Eq.(2.10) of Ref.7, which we also reproduce here for completeness

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} g^\mu_{\alpha,M}\, dV, \tag{2.2}$$

where $\hat{c}_{L\gamma M\alpha}$ is defined in Eq.(2.4) of Ref.7 and all other quantities are defined in Sec.2 of Ref.7. Since g^μ_α denotes the normalized mode shape in the contoured resonator and from Eq.(2.4) of Ref.7 $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_μ can be evaluated when the mode shape in the resonator and biasing state are known.

Since the plate is thin, for in-plane acceleration the equations of anisotropic plane stress apply. For anisotropic plane stress the stresses K_{2L} vanish and only K_{AB} exist, where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2. The equations for anisotropic plane stress may be written in the form given in Eq.(3.5) of Ref.7, which are reproduced here for completeness,

$$K_{AB,A} - \rho a_B = 0, \tag{2.3}$$

where a_B denotes the acceleration in the plane of the plate and the anisotropic constitutive equations for plane stress may be written in the form

$$K_{AB} = \gamma_{ABCD} E_{CD}, \quad E_{CD} = \frac{1}{2}(w_{C,D} + w_{D,C}), \tag{2.4}$$

and γ_{ABCD} are Voigt's anisotropic plate elastic constants, which are defined in Eq.(3.7) of Ref.7. Since it is shown in Sec.3 of Ref.7 [between Eqs.(3.8) and (3.10)] that when a solution to a plane stress problem has been found all three-dimensional strains E_{KL} and rotations Ω_{LK} can be obtained, when such a solution is known we may write

$$w_{K,L} = E_{KL} + \Omega_{LK}, \tag{2.5}$$

for the three-dimensional displacement gradients of the biasing deformation field, which are required for the evaluation of $\hat{c}_{L\gamma M\alpha}$ of Ref.7.

3. Extension of Rectangular Quartz Plates Rigidly Supported Along the Edges

A plan view and cross-section of the plate is shown in Fig.1 along with the coordinate system. Since the plate has all four edges fixed, we have the boundary conditions

CH2690-6/89/0000-405 $1.00 C 1989 IEEE

$$w_1 = w_3 = 0 \text{ at } X_1 = \pm a, \ |X_3| \le b,$$
$$w_1 = w_3 = 0 \text{ at } X_3 = \pm b, \ |X_1| \le a. \qquad (3.1)$$

In obtaining the solution to this problem we first consider acceleration in the X_1-direction only so that $a_3 = 0$ and then transform the inhomogeneity from the differential equations, which consist of (2.3) with (2.4), into the boundary conditions (3.1) by writing

$$w_1 = A_1(X_3^2 - b^2) + \hat{w}_1, \quad w_3 = A_3(X_1 - a)(X_3 - b) + \hat{w}_3, \qquad (3.2)$$

which when substituted into (2.3) with (2.4) yields

$$\gamma_{11}\hat{w}_{1,11} + 2\gamma_{15}\hat{w}_{1,31} + \gamma_{55}\hat{w}_{1,33} + \gamma_{35}\hat{w}_{3,33} + (\gamma_{13} + \gamma_{55})\hat{w}_{3,31} + \gamma_{55}\hat{w}_{3,11} = 0,$$
$$\gamma_{33}\hat{w}_{3,33} + 2\gamma_{35}\hat{w}_{3,31} + \gamma_{55}\hat{w}_{3,11} + \gamma_{15}\hat{w}_{1,11} + (\gamma_{13} + \gamma_{55})\hat{w}_{1,31} + \gamma_{35}\hat{w}_{1,33} = 0, \qquad (3.3)$$

since A_1 and A_3 have been selected as

$$A_1 = \rho a_1/(\gamma_{55} - \gamma_{13}), \quad A_3 = -A_1. \qquad (3.4)$$

The further substitution of (3.2) into (3.1) yields the edge conditions

$$\hat{w}_1 + A_1(X_3^2 - b^2) = 0, \ \hat{w}_3 = 0 \text{ at } X_1 = a, \qquad (3.5)$$
$$\hat{w}_1 + A_1(X_3^2 - b^2) = 0, \ \hat{w}_3 - 2aA_3(X_3 - b) = 0 \text{ at } X_1 = -a, \qquad (3.6)$$
$$\hat{w}_1 = \hat{w}_3 = 0 \text{ at } X_3 = b, \qquad (3.7)$$
$$\hat{w}_1 = 0, \ \hat{w}_3 - 2bA_3(X_1 - a) = 0 \text{ at } X_3 = -b. \qquad (3.8)$$

As a solution of (3.3) consider

$$\hat{w}_A = \sum_m \beta_{Am} e^{i\eta_m X_3} e^{i\nu_m X_1}, \qquad (3.9)$$

where $\nu_m = m\pi/2\sigma a$ and m takes all + and - integers considered in the sum, and σ is a parameter which can be chosen to improve convergence. In this work it was found that $\sigma = \pi$ worked well. The substitution of (3.9) into (3.3) yields

$$[\gamma_{11}\nu_m^2 + 2\gamma_{15}\nu_m\eta_m + \gamma_{55}\eta_m^2]\beta_{1m} + [\gamma_{35}\eta_m^2 + (\gamma_{13} + \gamma_{55})\eta_m\nu_m + \gamma_{15}\nu_m^2]\beta_{3m} = 0,$$
$$[\gamma_{15}\nu_m^2 + (\gamma_{13} + \gamma_{55})\eta_m\nu_m + \gamma_{35}\eta_m^2]\beta_{1m} + [\gamma_{33}\eta_m^2 + 2\gamma_{35}\eta_m\nu_m + \gamma_{55}\nu_m^2]\beta_{3m} = 0, \qquad (3.10)$$

for each m. Equations (3.10) constitute a system of two linear homogeneous algebraic equations in β_{1m} and β_{3m}, which yields nontrivial solutions when the determinant of the coefficients vanishes. If we define $\eta_m = h\nu_m$, then the vanishing of the determinant yields

$$[\gamma_{33}\gamma_{55} - \gamma_{35}\gamma_{35}]h^4 + [2\gamma_{33}\gamma_{15} - 2\gamma_{13}\gamma_{35}]h^3 + [\gamma_{11}\gamma_{33} - \gamma_{13}^2 + 2\gamma_{15} - \gamma_{35} - 2\gamma_{13}\gamma_{55}]h^2$$
$$2[\gamma_{11}\gamma_{35} - \gamma_{15}\gamma_{12}]h + [\gamma_{11}\gamma_{55} - \gamma_{15}^2] = 0, \qquad (3.11)$$

which has four roots $h^{(p)}$ $(p = 1,2,3,4)$ and corresponding sets of amplitude ratios $\beta_{1m}^{(p)} : \beta_{3m}^{(p)}$, which are independent of m. Hence as a solution of the boundary value problem we take

$$\hat{w}_A = \sum_m \sum_{p=1}^{4} D_p^m \beta_A^{(p)} e^{i(\eta_m^p X_3 + \nu_m X_1)}, \qquad (3.12)$$

where the D_p^m are amplitude coefficients still to be determined and $\eta_m^p = h^{(p)}\nu_m$. By requiring (3.7) and $(3.8)_1$ to be satisfied for each m, we can reduce the number of unknown coefficients D_p^m from 4 to 1 for each m. Accordingly, we obtain

$$D_p^{(m)} = g_p^{(m)} D_1^{(m)}, \qquad (3.13)$$

where the amplitude ratios $g_p^{(m)}$ are lengthy expressions, which are too cumbersome to present here[8]. The substitution of (3.13) into (3.12) yields the reduced form

$$\hat{w}_A = \sum_m \left(\sum_{p=1}^{4} \beta_A^{(p)} g_p^{(m)} e^{i\eta_m^p X_3} \right) D_1^m e^{i\nu_m X_1}, \qquad (3.14)$$

for the solution function.

The solution functions in (3.14) satisfy the differential equations and boundary conditions (3.7) and $(3.8)_1$ but they do not satisfy the boundary conditions $(3.8)_2$, (3.5) and (3.6). We satisfy these latter conditions by means of a variational approximation procedure. Since the solution functions satisfy the differential equations exactly, all that remains of the appropriate variational principle in which all conditions are unconstrained[9] is

$$\int_S N_A(w_B - \bar{w}_B)\, \delta K_{AB}\, dS = 0, \qquad (3.15)$$

where N_A is the outwardly directed unit normal to the edges and $\bar{w}_B$ is the prescribed edge displacement (here $\bar{w}_B = 0$). Substituting from (3.2) into (3.15) and using the fact that the edge conditions (3.7) and $(3.8)_1$ are satisfied exactly, we obtain

$$-\int_{-a}^{a} \left[\left(\hat{w}_3 + 2bA_1(X_1 - a) \right) \delta\hat{K}_{33} \right]_{X_3 = -b} dX_1$$
$$+ \int_{-b}^{b} \left[\left(\hat{w}_1 + A_1(X_3^2 - b^2) \right) \delta\hat{K}_{11} + \hat{w}_3 \delta\hat{K}_{13} \right]_{X_1 = a} dX_3$$
$$+ \int_{-b}^{b} \left[\left(\hat{w}_1 + A_1(X_3^2 - b^2) \right) \delta\hat{K}_{11} + \left(\hat{w}_3 + 2aA_1(X_3 - b) \right) \delta\hat{K}_{13} \right]_{X_1 = -a} dX_3 = 0, \qquad (3.16)$$

since the variation of prescribed quantities vanishes. Substituting from (3.14) into (3.16) and performing the integrations, we obtain

$$\sum_m \sum_n a_{(mn)} D_1^m \delta D_1^n + \sum_n b_{(n)} \delta D_1^n = 0 , \qquad (3.17)$$

where the expressions for the $a_{(mn)}$ and $b_{(n)}$ are too lengthy to present here[8]. Since the variations δD_1^n are arbitrary, we have

$$\sum_m a_{(mn)} D_1^m = - b_n , \qquad (3.18)$$

which constitute N inhomogeneous linear algebraic equations in the N unknowns $D_1^{(m)}$. Convergence is determined by increasing the number N and comparing the solution for N with that for N - 1.

For acceleration in the X_3-direction only, $a_1 = 0$ in (2.3) and we use the transformation

$$w_1 = B_1(X_1 - a)(X_3 - b) + \hat{w}_1 ,$$
$$w_3 = B_3(X_3^2 - b^2) + \hat{w}_3 , \qquad (3.19)$$

in place of (3.2). Then Eqs.(3.3) hold if we select B_3 and B_1 as

$$B_3 = \frac{a_3 \gamma_{15}}{2\gamma_{33}\gamma_{15} - \gamma_{35}(\gamma_{13} + \gamma_{55})} , \quad B_1 = - \frac{\gamma_{35}}{\gamma_{15}} B_3 , \qquad (3.20)$$

in which $\gamma_{35}/\gamma_{15} = 0/0$ is to be taken as 1 for the AT-cut and in place of (3.5) - (3.8) we have

$$\hat{w}_1 = 0, \ \hat{w}_3 + B_3(X_3^2 - b^2) = 0 \quad \text{at } X_1 = +a \qquad (3.21)$$

$$\hat{w}_1 - 2aB_1(X_3 - b) = 0, \ \hat{w}_3 + B_3(X_3^2 - b^2) = 0 \quad \text{at } X_1 = -a \qquad (3.22)$$

$$\hat{w}_1 = \hat{w}_3 = 0 \quad \text{at } X_3 = +b \qquad (3.23)$$

$$\hat{w}_1 - 2bB_1(X_1 - a) = 0, \ \hat{w}_3 = 0 \quad \text{at } X_3 = -b . \qquad (3.24)$$

Furthermore, Eqs.(3.9) - (3.14) hold in this case also provided that Eqs.(3.23) and $(3.24)_2$ are satisfied in place of (3.7) and $(3.8)_1$ in obtaining the $g_p^{(m)}$, which as already noted are too cumbersome to present here[8]. In addition, for this case Eq.(3.15) yields

$$- \int_{-a}^{a} \left[\left(\hat{w}_1 - 2bB_1(X_1 - a) \right) \delta \hat{K}_{31} \right]_{X_3 = -b} dX_1$$
$$+ \int_{-b}^{b} \left[\hat{w}_1 \delta \hat{K}_{11} + \left(\hat{w}_3 + B_3(X_3^2 - b^2) \right) \delta \hat{K}_{13} \right]_{X_1 = a} dX_3$$
$$+ \int_{-b}^{b} \left[\left(\hat{w}_1 - 2aB_1(X_3 - b) \right) \delta \hat{K}_{11} + \left(\hat{w}_3 + B_3(X_3^2 - b^2) \right) \delta \hat{K}_{13} \right]_{X_1 = -a} dX_3 = 0 , \qquad (3.25)$$

in place of Eq.(3.16), and from which we obtain a system of equations having the same form as (3.18).

4. The Equivalent Trapped Energy Mode Shape

A schematic diagram of the contoured crystal resonator along with the associated coordinate system is shown in Fig.2. It has been shown that the eigensolution for the dominant displacement of the harmonic modes in a contoured quartz resonator can be written in the form[10,8]

$$u_{1n} = \sin \frac{n\pi X_2}{2h} \, u_{noo} e^{i\omega_{noo} t} , \quad n = 1,3,5, \ldots , \qquad (4.1)$$

where

$$u_{noo} = e^{-\alpha_n \frac{\hat{X}_1^2}{2} - \frac{\hat{X}_3^2}{2}} , \qquad (4.2)$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{8Rh_o^3 M_n'} , \quad \beta_n^2 = \frac{n^2\pi^2 c^{(1)}}{8Rh_o^3 P_n'} , \qquad (4.3)$$

and M_n' and P_n' are involved lengthy expressions that are defined in Sec.II of Ref.10 along with $\hat{X}_1$ and $\hat{X}_3$, which are orthogonal directions in the plane of the plate for which the scalar differential equation for the nth anharmonic family of modes does not contain mixed derivatives and R denotes the radius of the spherical contour. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2\hat{c}^{(1)}}{4h_c^2\rho} \left[1 + \frac{1}{n\pi} \left(\frac{2h_o}{R} \right)^{1/2} \left(\sqrt{\frac{M_n'}{\hat{c}^{(1)}}} + \sqrt{\frac{P_n'}{\hat{c}^{(1)}}} \right) \right] , \qquad (4.4)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.10.

In addition to the thickness-eigendisplacement u_{1n} given in (4.1) there are thickness-eigendisplacements u_{2n} and u_{3n} which are an order of magnitude smaller than u_{1n} but are included in this work in order to evaluate their influence. However, since the u_2 and u_3 displacement components accompanying the larger u_1 component are known only for the flat plate, we fit the Gaussian mode shape given in (4.1) and (4.2) for the contoured resonator to a trapped energy mode[11] in a flat plate in accordance with the diagram shown in Fig.3. We first note that the location of the edge of the electrode is irrevelant for the determination of the equivalent trapped energy mode in a contoured resonator because the trapping is due essentially to the contouring. The equivalent trapped energy mode is fitted to the Gaussian as in Sec.V of Ref.6, i.e., by matching the Gaussian at the center of the plate and requiring the volumes under the Gaussian mode shape and the equivalent trapped energy mode to be separately the same under the inner rectangular region defined by the lines of inflection of the Gaussian mode shape in the two directions and the outer regions. This is a reasonable procedure for our purposes because such a function can match the Gaussian quite well.

Since the equivalent trapped energy mode is determined by matching to the <u>known</u> Gaussian mode shape in accordance with the foregoing procedure[6], the dispersion relations are not needed and only the continuity of the mechanical displacement u_1 need be imposed at the line of inflection in each direction.

The solution functions for the trapped energy eigenmodes have been shown[12] to be of the form

$$\bar{u}_1 = \bar{B} \sin \frac{n\pi X_2}{2h} \cos \hat{\bar{\xi}}\hat{X}_1 \cos \hat{\bar{\nu}}_3\hat{X}_3,$$

$$u_1^{SR} = B^{SR} \sin \frac{n\pi X_2}{2h} e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)} \cos \hat{\bar{\nu}}\hat{X}_3,$$

$$u_1^{TM} = B^{TM} \sin \frac{n\pi X_2}{2h} \cos \hat{\bar{\xi}}\hat{X}_1 e^{-\hat{\nu}^{TM}(\hat{X}_3 - a_3)},$$

$$u_1^{TR} = B^{TR} \sin \frac{n\pi X_2}{2h} e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)} e^{-\hat{\nu}^{TM}(\hat{X}_3 - a_3)}, \quad (4.5)$$

in the upper right quadrant of Fig.3, which is all that is needed because of the symmetry of the mode. Following the matching procedure outlined[8] in Sec.V of Ref.6 we obtain

$$\bar{B} = 1, \; B^{SR} = \cos \hat{\bar{\xi}}a_1, \; B^{TM} = \cos \hat{\bar{\nu}}a_3,$$

$$B^{TR} = \cos \hat{\bar{\xi}}a_1 \cos \hat{\bar{\nu}}a_3, \quad (4.6)$$

along with the transcendental equations for the trigonometric wavenumbers $\hat{\bar{\xi}}$ and $\hat{\bar{\nu}}$

$$\frac{\sin \hat{\bar{\xi}}a_1}{\hat{\bar{\xi}}a_1} = \sqrt{\frac{\pi}{2\alpha_n a_1^2}} \operatorname{erf}\left(\sqrt{\frac{\alpha_n a_1^2}{2}}\right),$$

$$\frac{\sin \hat{\bar{\nu}}a_3}{\hat{\bar{\nu}}a_3} = \sqrt{\frac{\pi}{2\beta_n a_3^2}} \operatorname{erf}\left(\sqrt{\frac{\beta_n a_3^2}{2}}\right), \quad (4.7)$$

for the fundamental and harmonic overtones in the interior regions and the decay numbers $\hat{\xi}^{SR}$ and $\hat{\nu}^{TM}$

$$\hat{\xi}^{SR} = \frac{\cos \hat{\bar{\xi}}a_1}{(\pi/2\alpha_n)^{1/2} \operatorname{erfc}[(\alpha_n/2)^{1/2} a_1]},$$

$$\hat{\nu}^{TM} = \frac{\cos \hat{\bar{\nu}}a_3}{(\pi/2\beta_n)^{1/2} \operatorname{erfc}[(\beta_n/2)^{1/2} a_3]}, \quad (4.8)$$

for the SR and TM regions, respectively.

As noted earlier, in addition to the u_{1n} displacement fields there are accompanying u_{2n} and u_{3n} displacement fields, which are an order of magnitude smaller than u_{1n}. Furthermore, there is a correction to u_{1n} which is of the same order as u_{2n} and must be included. From Eqs.(65) of Ref.10 we find that to the order of interest the u_n^{SR} displacement field for the SR region may be written

$$u_1^{SR} = A_+^{(1)SR}\left[1 + \frac{(c_{16}\xi^{SR} - c_{56}i\bar{\nu})}{\bar{c}^{(1)}} X_2\right] \times \sin \frac{n\pi X_2}{2h} e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)} e^{i\hat{\bar{\nu}}\hat{X}_3},$$

$$u_2^{SR} = \left[\frac{(r_2\xi^{SR} - r_4 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR} \cos \frac{n\pi}{2h} X_2 + iC_+^{(2)SR} \cos \kappa_2 \frac{n\pi}{2h} X_2\right] e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)} e^{i\hat{\bar{\nu}}\hat{X}_3},$$

$$u_3^{SR} = \left[\frac{(r_5\xi^{SR} - r_3 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR} \cos \frac{n\pi}{2h} X_2 + iE_+^{(3)SR} \cos \kappa_3 \frac{n\pi}{2h} X_2\right] e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)} e^{i\hat{\bar{\nu}}\hat{X}_3}, \quad (4.9)$$

where from (4.9) and Eqs.(97) of Ref.10 we have[8]

$$\xi^{SR} = R_{11}\hat{\xi}^{SR} - R_{31}i\hat{\bar{\nu}}, \quad i\bar{\nu} = R_{33}i\hat{\bar{\nu}} - R_{13}\hat{\xi}^{SR}, \quad (4.10)$$

where R_{ij} is defined in Eqs.(98) of Ref.10 and from Eqs.(72) of Ref.10 we know $C_+^{(2)SR}$ and $E_+^{(3)SR}$ in terms of $A_+^{(1)SR}$ and to save space we do not write the relations or the corresponding displacement fields for the other regions. We note that $A_+^{(1)SR}$ here corresponds to B^{SR} in Eqs.(4.5) and (4.6). It should be carefully noted that the constants appearing in Eqs.(4.9) and Sec.II of Ref.10 are not the usual elastic constants appearing in the previous sections of this work. This is a result of the fact that they are referred to both the natural SC-cut plate axes and the orthogonal eigenvector triad of the pure thickness solution, as shown in Ref.10, where they are defined in Eq.(23).

5. Acceleration Sensitivity

From Sec.4 we now know g_γ^μ and from Sec.3 we know $\hat{c}_{LYM\alpha}$ for in-plane acceleration, i.e., extension with rigid rectangular supports. Hence, we can now evaluate H_μ in Eq.(2.2). Such calculations have been performed using the known values of the second order[13] and third order[14] elastic constants of quartz. The calculated in-plane acceleration sensitivities for the three lowest harmonic modes of the SC cut are plotted in Fig.4 as a function of the planar aspect ratio a/b for a fixed rectangular area of 200 mm^2 for two orientations of the rectangle with respect to the conventional axes of the SC-cut plate. It can be seen from the figure that for the rectangle oriented along the conventional axes there is a zero-crossing for an acceleration in the X_1-direction but not in the X_3-direction. It can also be seen that for a rectangle oriented at 30° with respect to the conventional axes there are no zero-crossings and that the sensitivities are considerably larger than for $\beta = 0$. Similar curves for the AT cut are plotted in Fig.5, which shows that the range of acceleration sensitivities is considerably smaller for the AT cut than for the SC cut and that both orientations of the AT cut have zero-crossings for acceleration in the X_1-direction. At this point it should be noted that the acceleration sensitivity curves for the fundamental and the third and fifth harmonics are indistinguishable in all the figures and that the fixed planar area of 100 mm^2 is adhered to throughout. Figure 6 shows the in-plane sensitivities for the SC cut for a fixed a/b ratio of 2 as a function of the orientation of the rectangle. The figure shows that there are zero-crossings for acceleration in both the X_1- and X_3-directions, but they do not occur at the same orientation. It also shows that the range of calculated sensitivities for

the SC cut is quite large. Figure 7 shows the resultant in-plane acceleration sensitivity obtained from Fig.6. Figure 8 shows the same type of information as Fig.6 for an SC cut, but for an a/b ratio of 1.5 instead of 2. The figure shows that things have changed a bit with the A_1 curve having a smaller range of variation and the A_3 curve having a larger range of variation than in Fig.6. Figure 9 shows the resultant in-plane acceleration sensitivity obtained from Fig.8. Figure 10 shows the same type of information as Figs.6 and 8 for an SC cut with an a/b ratio of 1. The figure shows that the A_1 and A_3 curves have become identical but displaced by 90°, as expected for a square plate. Figure 11 shows the resultant in-plane acceleration sensitivity obtained from Fig.10. Figure 12 shows the in-plane acceleration sensitivities for the AT cut for three distinct a/b ratios as a function of the orientation β. The figure shows that there are zero-crossings for acceleration in both the X_1- and X_3-directions and that the range of acceleration sensitivities is much smaller for the AT cut than for the SC cut. Furthermore, it can be seen from the figure that for a/b = 2.9 there is a zero-crossing for both in-plane acceleration components at the same angle of β = 84°. This is a result of potential practical importance. Figure 13 shows the resultant in-plane acceleration sensitivities obtained for the a/b ratios considered in Fig.12. The figure clearly shows the above-mentioned zero in-plane acceleration sensitivity at β = 84° for a/b = 2.0

Acknowledgements

This work was supported in part by the Army Research Office under Contract No. DAAL03-88-K-0065.

References

1. R.L. Filler, J.A. Kosinski and J.R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," Proceedings of the 37th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No. 83CH1957-0, 265 (1983).
2. H.F. Tiersten and D.V. Shick, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," 1988 Ultrasonics Symposium Proceedings, IEEE Cat. No. 88CH2578-3, Institute of Electrical and Electronics Engineers, New York, 357 (1988).
3. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," J. Appl. Phys., 64, 4334 (1988).
4. D.V. Shick, Y.S. Zhou and H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Quartz Surface-Wave Resonators Rigidly Supported Along the Edges," J. Appl. Phys., 65, 35 (1988).
5. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
6. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).
7. H.F. Tiersten and D.V. Shick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," Proceedings of the 43rd Annual Symposium on Frequency Control (these proceedings).
8. For more detail see D.V. Shick and H.F. Tiersten, "On the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
9. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4, Eq.(6.44).
10. D.S. Stevens and H.F. Tiersten, "An Analysis of Doubly-Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., 79, 1811 (1986).
11. H.F. Tiersten, "Analysis of Trapped Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness-Twist," J. Acoust. Soc. Am., 59, 879 (1976).
12. The amplitude factors B, B^{SR}, B^{TM} and B^{TR} given in Eqs. (4.6) and related in Eqs.(5.13) of Ref.6 were inadvertently omitted in Eqs.(103) and (104) of Ref.10.
13. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).
14. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

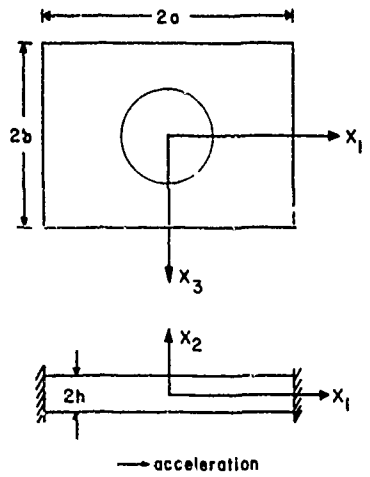


Figure 1 Plan View and Cross-Section of Rectangular Quartz Plate Rigidly Supported Along the Edges

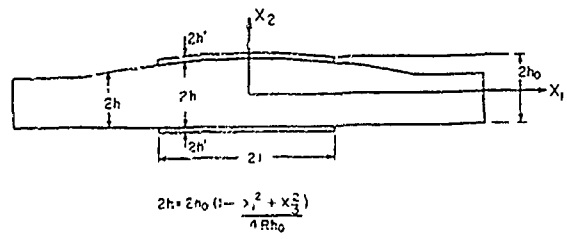


Figure 2 Cross-Section of the Plano-Convex Resonator

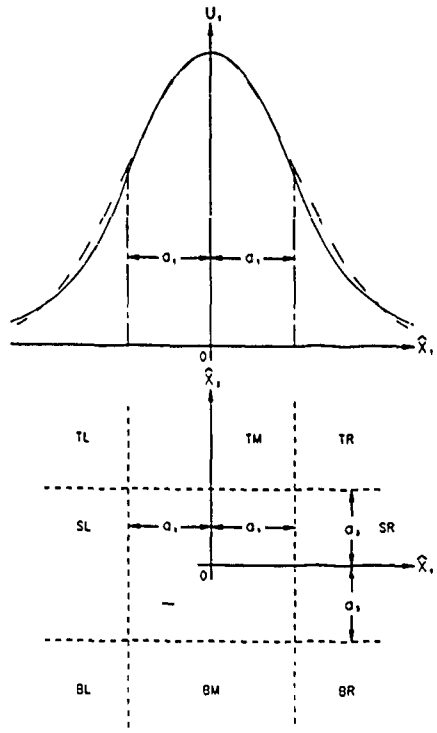


Figure 3 Plan View and Cross-Section for Obtaining the Equivalent Trapped Energy Mode Shape from the Gaussian Mode Shape on the Contoured Resonator

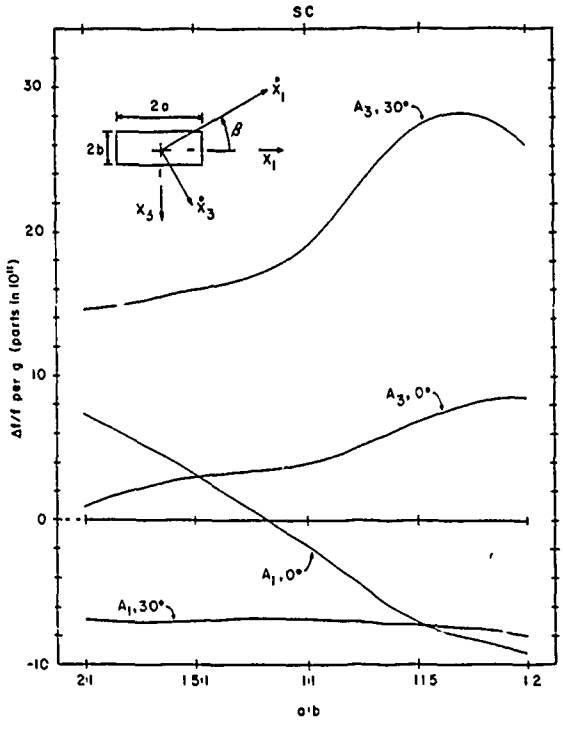


Figure 4 Calculated In-Plane Acceleration Sensitivity Versus Planar Aspect Ratio a/b for the Lowest Three Harmonics of the Contoured SC-Cut Plate for Two Orientations of the Rectangular Support System

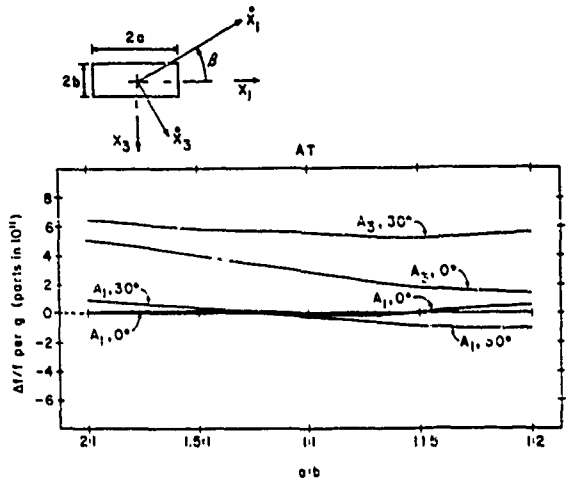


Figure 5 Calculated In-Plane Acceleration Sensitivity Versus Planar Aspect Ratio a/b for the Lowest Three Harmonics of the Contoured AT-Cut Plate for Two Orientations of the Rectangular Support System

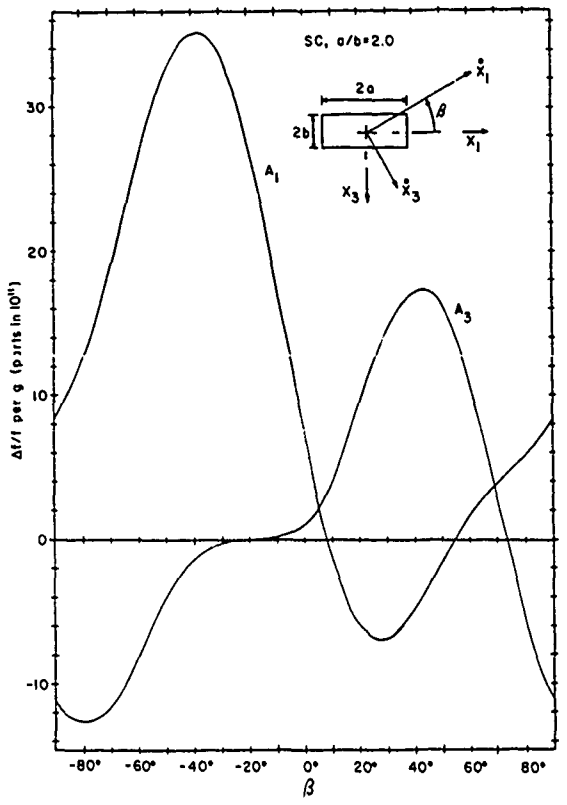


Figure 6 Calculated In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a Fixed a/b = 2.0

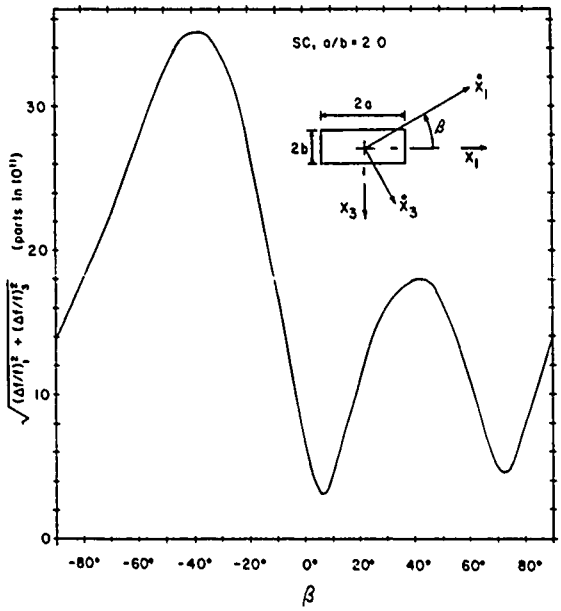


Figure 7 Resultant In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a/b = 2.0

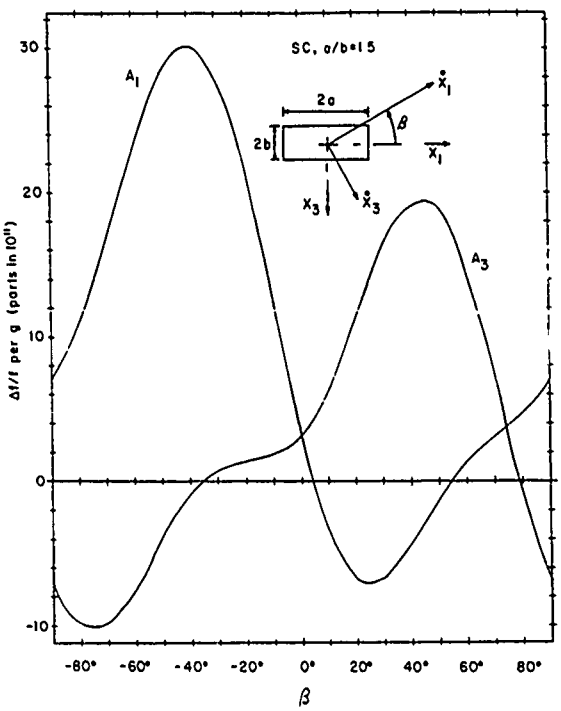


Figure 8 Calculated In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a/b = 1.5

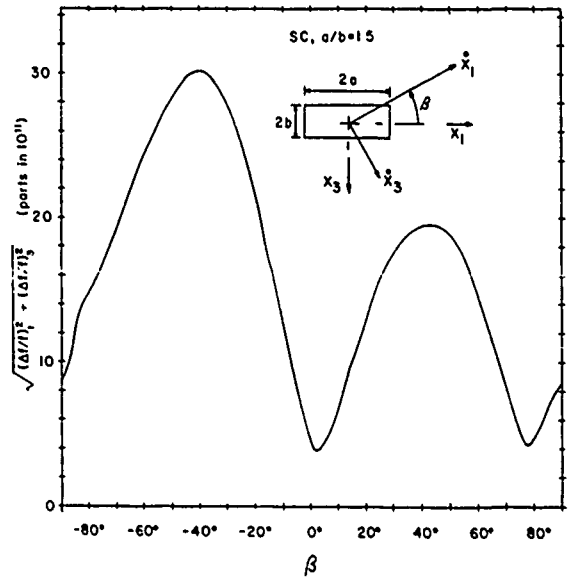


Figure 9 Resultant In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a/b = 1.5

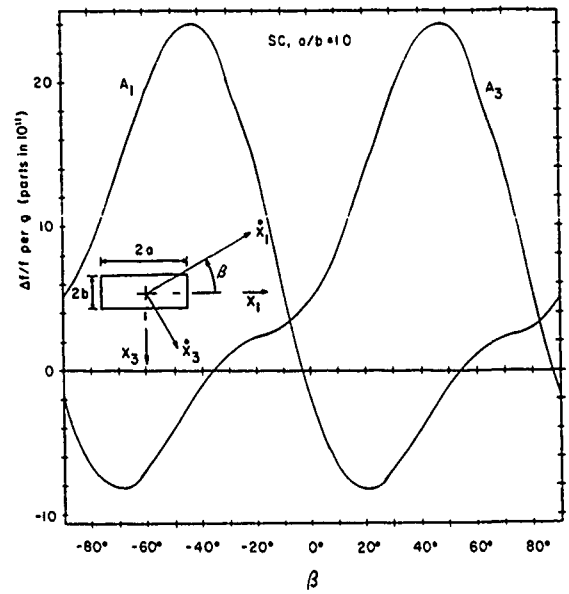


Figure 10 Calculated In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a/b = 1.0

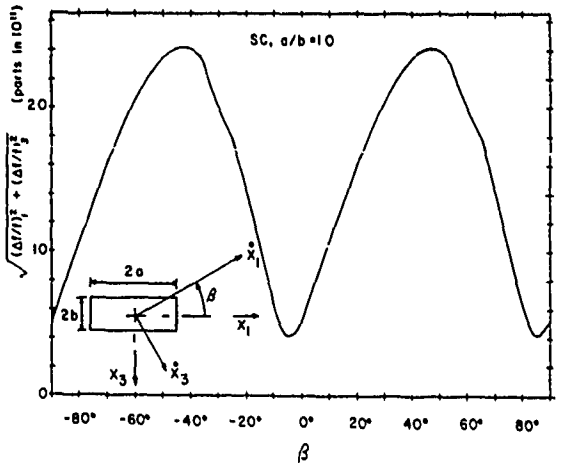


Figure 11 Resultant In-Plane Sensitivity Versus Orientation of the Rectangle for the Contoured SC-Cut Plate for a/b = 1.0

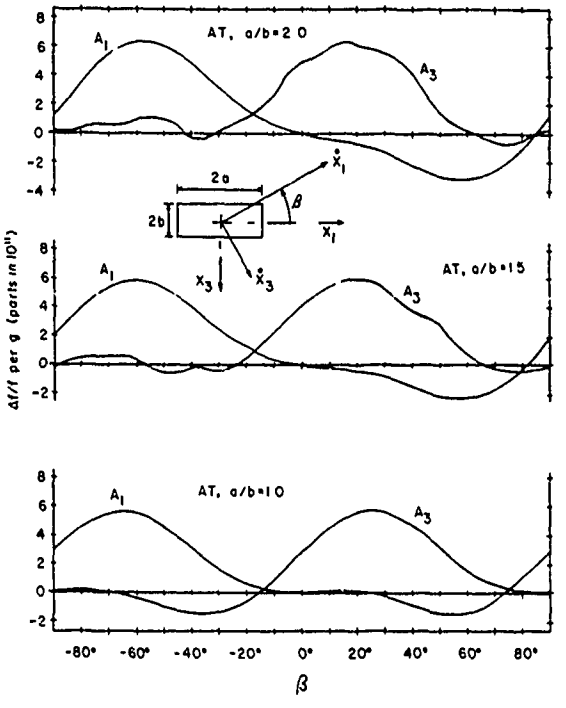


Figure 12 Calculated In-Plane Sensitivity Versus Orientation of the Rectangle for Contoured AT-Cut Plates for Three a/b Ratios

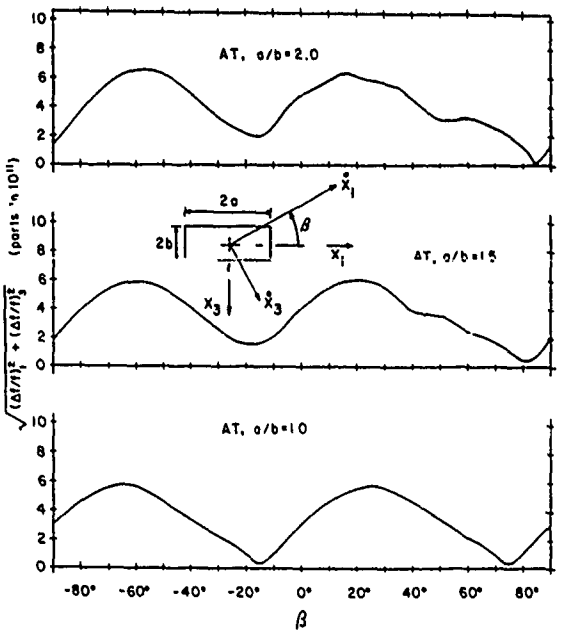


Figure 13 Resultant In-Plane Sensitivities Versus Orientation of the Rectangles for Contoured AT-Cut Plates for Three a/b Ratios

43rd Annual Symposium on Frequency Control - 1989

Parametric Study of Acceleration Insensitivity of Doubly Rotated Circular Quartz Resonators

P.C.Y. Lee and M.S.H. Tang*
Department of Civil Engineering & Operations Research
Princeton University
Princeton, N.J. 08544

Summary

Initial displacement gradients and strains which were caused by the steady acceleration of an arbitrary direction acting on an crystal disk are calculated from the finite element solution of Mindlin's two-dimensional, first-order equations of equilibrium of crystal plates.[1] These initial displacement gradients and strains are taken into account as known functions in the frequency equation of incremental thickness vibrations[2] which is, in turn, solved for frequency changes by a perturbation method.

To search for potentially acceleration-insensitive, circular resonators, the acceration Γ is calculated as a function of (1) the number of resonator supports N, (2) the orientation angle of the support configuration α, (3) the fast or slow thickness-shear mode of vibrations (the b- or c-mode), (4) the orientation angles of doubly rotated cut of quartz ϕ, θ. In the present study, the number of supports N is limited to 4 or 6, and the angles of cut ϕ and θ are limited to those along the upper and lower loci of zeros of the first-order temperature coefficient of frequency of slow thickness-shear mode c, along the lower locus of fast thickness-shear mode b[3], and in the AK-cut region for thickness-shear mode c[4], for their desirable frequency-temperature behavior.

1. Introduction

In two previous papers[5,6], effects of support numbers, support orientation, support structures, and plate thickness on the acceleration sensitivity of SC- and AT-cuts of quartz disks were studied. The background of related previous investigations and the definition of the acceleration sensitivity Γ were given in Ref. 5.

The purpose of the present study is to search for potentially acceleration-insensitive circular quartz resonators by computing the combined effect from the various resonator parameters, listed in the Summary, to reduce or minimize the acceleration sensitivity.

$$\Gamma = \Gamma(N, M, \alpha, \phi, \theta) \qquad (1)$$

where

N = the number of supports,

M = b- or c-mode,

α = orientation angle of support configuration,

ϕ, θ = orientation angles of doubly rotated cut.

2. Acceleration Sensitivity for c-mode with 4-pt. Mount

(a) Along the Upper locus[3]:

$$\Gamma = \Gamma(\phi, \theta, \alpha, N = 4, M = c\text{-}mode)$$

where

$$\theta = 32.25° - (11/180)\,\phi.$$

Γ is computed for various values of α and for $\theta = 0° \sim 30°$. The result is plotted in Fig. 1. It may be seen that on the curve $\alpha = 0$, Γ has a minimum value $0.27 \times 10^{-10}/g$ at $\phi = 0$, i.e., the value for an AT-cut with 4-pt. mount and $\alpha = 0$ as predicted previously[6]. It is interesting to see that for $\alpha = 7\,18°$ and $\phi = 3° \sim 7°$, $\Gamma \le 0.28 \times 10^{-10}/g$. At $\phi = 5°$, $\Gamma \sim 10^{-11}/g$ and it is not sensitive to the incremental change of ϕ.

(b) Along the lower locus[7]:

$$\Gamma = \Gamma(\phi, \theta, \alpha, N = 4, M = c\text{-}mode)$$

where

$$\theta = 0.072\,\phi - 35.58°$$

Γ is computed for various values of α and for $\phi = 15° \sim 30°$. The values of Γ are always greater than or equal to $0.5 \times 10^{10}/g$, therefore they are not shown here.

3. Acceleration Sensitivity for b-mode with 4-pt. Mount

The values of ϕ and θ for doubly rotated cuts along the lower locus of b-mode are chosen according to those given in Ref. 3.

The Γ vs ϕ curves for $\phi = 0° \sim 12°$ are shown in Fig. 2 for various values of α. It may be seen that for $\alpha = 0$, $\phi = 0° \sim 6°$, $\Gamma \le 10^{-11}/g$. At $\phi = 3°$, $\Gamma \sim 0.2 \times 10^{-11}/g$ and it is not sensitive to the incremental changes of ϕ or α.

4. Acceleration Sensitivity for c-mode with 6-pt. Mount

(a) Along the upper locus:

Computations of Γ are similar to those described in Section 2(a), except now N = 6.

The Γ vs ϕ curves are given in Fig. 3 from which we see that for $\alpha = 0$, Γ has a minimum value of $0.45 \times 10^{-10}/g$ at $\phi = 22°$, i.e, the value of an SC-cut with 6-pt. mount and $\alpha = 0$ as predicted previously.[5,6]. It is interesting to note that for $\alpha = 7.18°$ and $\phi = 6° \sim 14°$, $\Gamma \le 0.45 \times 10^{-10}/g$. At $\phi = 9°$, $\Gamma = 0.25 \times 10^{-10}/g$ which is comparable to acceleration sensitivity of an AT-cut with 4-pt. mount and $\alpha = 0$.

(b) Along the lower locus:

Computations are similar to those described in Section 2(b), except now N=6.

The Γ vs ϕ curves are shown in Fig. 4 which reveals that for $\alpha = -7.18°$ and $\phi = 27 \sim 30°$, $\Gamma \le 0.45 \times 10^{-10}/g$. It is comparable or slightly better than that for the SC-cut with 6-pt. mount and $\alpha = 0$.

5. Acceleration Sensitivity for b-mode with 6-pt. Mount

Computations of Γ are similar to those described in Section 3 except no N = 6.

Results are given in Fig. 5 which shows that $0.15 \times 10^{-10}/g \le 0.9 \times 10^{-10}/g$ for $-30° \le \alpha \le 30°$ and $\phi = 0° \sim 12°$.

CH2690-6/89/0000-413 $1.00 C 1989 IEEE

6. Acceleration Sensitivity for c-mode with 4-pt. Mount in the AK-cut Region

Acceleration sensitivity are computed for cuts in the AK-cut[4] region, i.e., for $\phi = 30° \sim 46°$, $\theta = 20° \sim 32°$, and for support orientation angles $\alpha = 0°, 30°$, and $60°$.

Values of Γ in $(10^{-10}/g$ for $\alpha = 0$ is given in Fig. 6 from which we see that except in a narrow strip the values of Γ are greater than $1.5 \times 10^{-10}/g$ in the whole region. Predicted values of Γ for $\alpha = 30°$ and $60°$ are, in general greater than $2 \times 10^{-10}/g$ in the AK-cut region.

7. Acceleration Sensitivity for c-mode with 6-pt. Mount in the AK-cut Region

Computations of Γ for resonators with 6-pt. mount are made for $\alpha = 0, 14.5°, 30°$, and $-14.5°$. Predicted Γ in the AK-cut region for $\alpha = 30°$ and $-14.5°$ are shown in Figs. 7 and 8, respectively. It can be seen from Figs. 7 and 8 that there exists regions in which Γ attains values between $0.5 \times 10^{-10}/g$ to $1.0 \times 10^{-10}/g$.

8. Conclusion

By controlling the number of supports, mount orientation, angles of doubly rotated cut, and modes of thickness-shear vibration, computational results show that values of acceleration sensitivity can be further reduced and they are less sensitive to the incremental changes of ϕ and α as compared with those of the AT-cut with 4-pt. mount and the SC-cut with 6-pt. mount.

These acceleration-insensitive cuts or AI-cuts are summarized as follows:

Four-point mount

(a) c-mode, along the upper locus

$\phi = 0,\ \alpha = 0,\ \Gamma = 0.28 \times 10^{-10}/g$ (AT)

$\phi = 5°,\ \alpha = 7°,\ \Gamma = 10^{-11}/g$

where $\theta = 32.23° - (11/180)\,\phi$.

(b) b-mode, along the lower locus

$\phi = 3°,\ \alpha = 0,\ \Gamma = 0.2 \times 10^{-11}/g$.

Six-point mount

(a) c-mode, along the upper locus

$\phi = 22°,\ \alpha = 0,\ \Gamma = 0.45 \times 10^{-10}/g$ (SC)

$\phi = 9°,\ \alpha = 7°,\ \Gamma = 0.25 \times 10^{-10}/g$

where $\theta = 32.25° - (11/180)\,\phi$.

(b) c-mode, along the lower locus

$\phi = 28°,\ \alpha = -7°,\ \Gamma \le 0.45 \times 10^{-10}/g$

where $\theta = 0.72\,\phi - 35.58°$.

Acknowledgement

This work was supported by the U.S. Army Research Office, Contract No. DAAL 03-87-K-0125.

* Present address: Motorola, Inc. Franklin Park, IL.

References

1. P.C.Y. Lee and M.S.H. Tang, **Proce. 40th Ann. Freq. Cont. Symp**, pp. 152-160, 1986.
2. P.C.Y. Lee, X. Guo, and M.S.H. Tang, **J. Appl. Phys.**, 63(6), pp. 1850-1856, 1988.
3. A. Ballato, "Doubly rotated thickness mode plate vibrations", in **Physical Acoustics: Principles and Methods**, (W.P. Mason and R.N. Thurston, eds.), Vol. 13, Chap. 5, Academic Press, New York, pp. 115-181, 1977.
4. A. Kahan, **Proc. 36th Ann. Freq. Cont. Symp.**, p. 180, 1982.
5. P.C.Y. Lee and M.S.H. Tang, **Proc. 41st Ann. Freq. Cont. Symp.**, pp. 277-281, 1987.
6. P.C.Y. Lee and M.S.H. Tang, **Proc. 42nd Ann. Freq. Cont. Symp.** , pp. 14-18, 1988.
7. A. Ballato, the equation for the lower locus of the c-mode is suggested by Dr. Ballato through private communication.

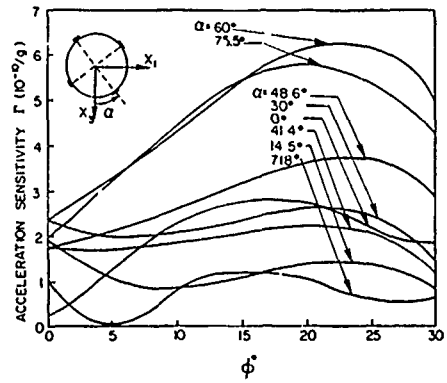


Fig. 1 Acceleration sensitivity Γ vs ϕ for 4-point mount and for cuts along the upper locus of c-mode, where $\theta = 35.25° - (11/180)\,\phi$.

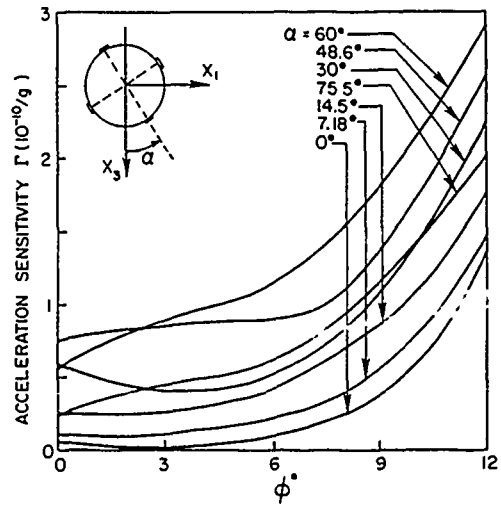


Fig. 2 Γ vs ϕ for 4-point mount and for cuts along the lower locus of b-mode.

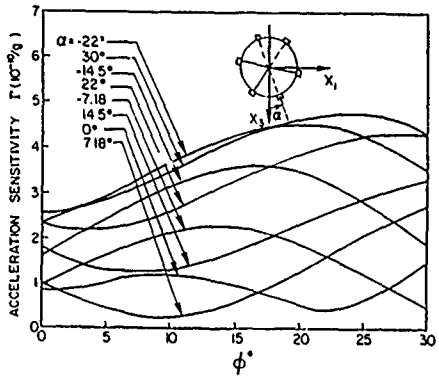


Fig. 3 Γ vs ϕ for 6-point mount and for cuts along the upper-locus of c-mode.

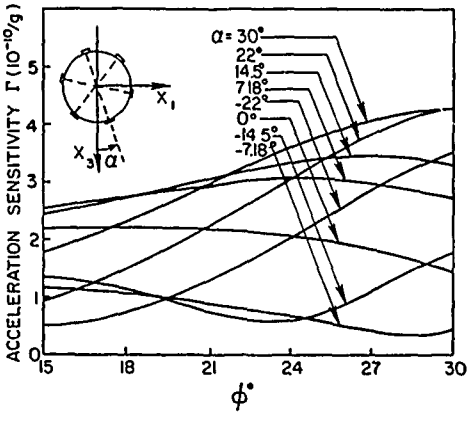


Fig. 4 Γ vs ϕ for 6-point mount and for cuts along the lower locus of c-mode, where $\theta = 0.072\,\phi - 35.58°$.

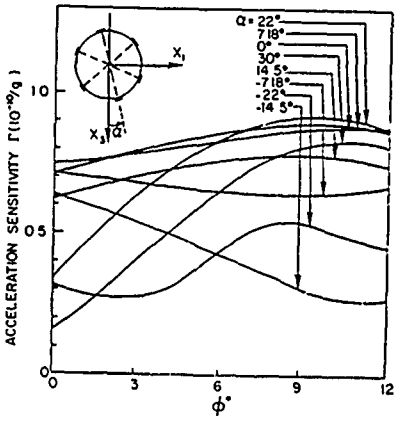


Fig. 5 Γ vs ϕ for 6-point mount and for cuts along the lower locus of b-mode.

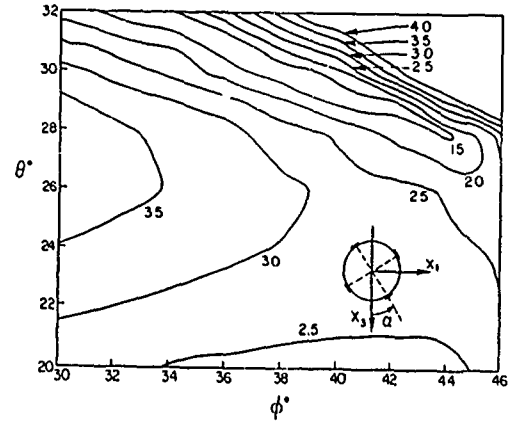


Fig. 6 Γ (in $10^{-10}/g$) for 4-point mount and $\alpha = 0$ in the AK-cut region.

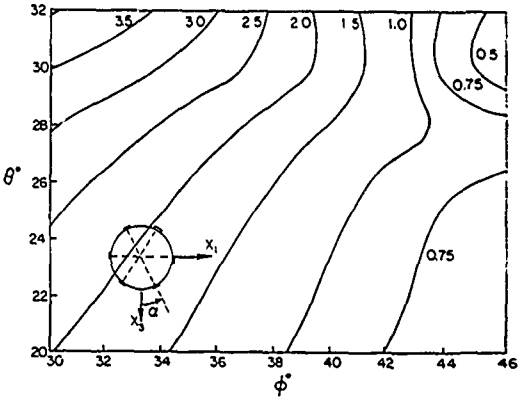


Fig. 7 Γ (in $10^{-10}/g$) for 6-point mount and $\alpha = 30°$ in the AK-cut region.

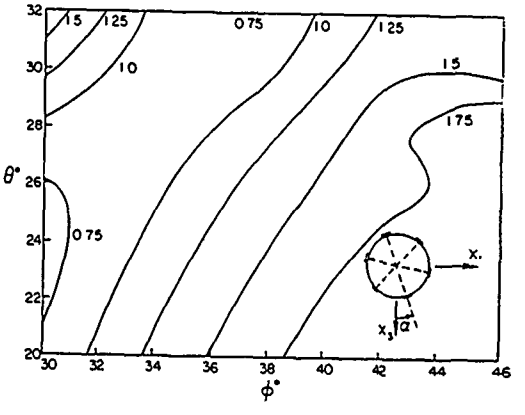


Fig. 8 Γ (in $10^{-10}/g$) for 6-point mount and $\alpha = 14.5°$ in the AK-cut region.

Effect of Electrodes on the Acceleration Sensitivity of Crystal Resonators

P.C.Y. Lee and X. Guo
Department of Civil Engineering and Operations Research
Princeton University
Princeton, N.J. 08544

Summary

When an electroded, crystal plate is subjected to a steady acceleration and a steady, uniform temperature change, stresses are induced by the body forces exerted on the plate and electrodes and by the differences in thermal expansions between the plate and electrodes.

Based on Mindlin's first-order equations of plated, crystal plates[1], a set of six coupled equations is derived in which the effect of mass of electrodes, the extensional elastic stiffnesses of the electrodes, and the thermal expansions of the plate and electrodes are taken into account. The principle of virtual displacements corresponding to these equations is obtained.

Strains and displacement gradients in an electroded crystal disk subject to a steady acceleration are obtained from the finite element solutions of the presently derived equations. By inserting these strains and displacement gradients in the frequency equation of the fundamental thickness vibrations of the disk and solving the frequency equation by a perturbation method, acceleration sensitivities of crystal resonators with various electrode arrangements are obtained.

It is founded that for small plate backs, say $R \leq 0.5\%$, the percent increase of acceleration sensitivity of an electroded plate as compared with that of an unelectroded plate is essentially proportinal to the ratio of the total mass of attached electrodes to the total mass of the plate.

1. Introduction

In our previous papers[2,3,4], effect of support numbers, orientation of the support configuration, support structures, plate thickness, and plate orientation on the acceleration sensitivity of circular quartz resonators were studied.

In the present investigation, in addition to the effect from the above-mentioned parameters, the mechanical effect of attached electrodes, i.e., the mass density, electric stiffnesses, and thermal expansion of electrodes, on the frequency changes is considered.

2 Equations of Plated Plate Under Steady Accelerations and Uniform Temperature Changes

Consider a crystal plate with thickness 2b and mass density ρ. It is plated with an upper plating (or electrode) of thickness $2b'$ and density ρ' and a lower plating of thickness $2b''$ and density ρ as shown in Fig. 1.

The governing equations of plated plate with thermal expansion are obtained as an extension of Mindlin's work of 1963[1].

(a) Governing equations for the plate

Stress equations of equilibrium

$$T_{i,i}^{(o)} + F_j^{(o)} + \rho B_j^{(o)},$$
$$T_{i,i}^{(1)} - T_{2j}^{(o)} + F_j^{(1)} \rho B_j^{(1)} = o, \qquad (1)$$

where $i,j = 1,2,3$, $T_{ij}^{(n)}$, $n = o,1$ are components of the nth-order stress, $F_j^{(n)}$ are components of the nth-order face traction, and $B_j^{(n)}$ are components of the nth-order body force,

$$F_j^{(o)} = T_{2j}(b) - T_{2j}(-b),$$
$$F_j^{(1)} = b\,[T_{2j}(b) + T_{2j}(-b)],$$
$$B_j^{(o)} = \int_{-b}^{b} B_j\, dx_2,$$
$$B_j^{(1)} = \int_{-b}^{b} x_2 B_j\, dx_2. \qquad (2)$$

We note that B_j is the body force per unit mass.

Stress-strain-temperature relations

$$T_i^{(o)} = 2b\, K_{(ij)} K_{(kl)} C_{ijkl} (E_{kl}^{(o)} - \alpha_{kl}\,\theta),$$
$$T_i^{(1)} = \frac{2b^3}{3} \hat{C}_{ijkl} E_{kl}^{(1)}, \qquad (3)$$

where $e_{kl}^{(n)}$ are components of the nth-order strain,

$$K_{(ij)} = \pi/\sqrt{12} \text{ for } i,j = 21,22,23$$
$$= 0 \quad \textit{otherwise}$$
$$\alpha_{kl} = \alpha_{kl}^{(1)} + \alpha_{kl}^{(2)}\theta + \alpha_{kl}^{(3)}\theta^2,$$
$$\hat{C}_{ijkl} = C_{ijkl} - C_{ij22} C_{22kl}/C_{2222}. \qquad (4)$$

In (4), $\alpha_{kl}^{(n)}$ are components of the nth-order thermal expansion coefficient[5], T is the temperature of the plate, T_o is a reference temperature, and $\theta = T - T_o$ is the temperature increase.

Strain-displacement Relations

$$E_{ij}^{(o)} = \frac{1}{2}(U_{j,i}^{(o)} + U_{i,j}^{(o)} + \delta_{2i} U_j^{(1)} + \delta_{2j} U_i^{(1)},$$
$$E_{ij}^{(1)} = \frac{1}{2}(U_{j,i}^{(1)} + U_{i,j}^{(1)}), \qquad ij \neq 22$$
$$E_{22}^{(1)} = -C_{22kl} E_{kl}^{(1)}/C_{222}, \qquad kl \neq 22 \qquad (5)$$

where $U_j^{(n)}$ are components of the nth-order displacement. Equations $(5)_3$ is obtained by setting $T_{22}^{(1)} = o$ to allow free expansion of $E_{22}^{(1)}$, and $T_{22}^{(1)} = o$ has been accommodated in $(3)_2$

(b) Governing equations for the upper plating

When the thickness of the plating is much less than that of the plate, i.e., $b'/b \ll 1$, only the "in-plane" components of

CH2690-6/89/0000-416 $1.00 C 1989 IEEE

stress in the plating are retained, since the rest are insignificantly small. Hence we have

$$T^{(\rho)\prime}_{ab,a} + F_b^{(o)\prime} + \rho' B_a^{(o)\prime} = o \ , \quad a,b = 1,3$$
$$F_2^{(o)\prime} + \rho' B_2^{(o)\prime} = o \ , \tag{6}$$

$$T_{ab}^{(\rho)} = 2b' \bar{C}_{abcd} \left(E_{cd}^{(\rho)\prime} - \alpha'_{cd}\,\theta\right), \tag{7}$$

$$E_{cd}^{(\rho)\prime} = \frac{1}{2}\left(U_{d,c}^{(o)\prime} + U_{d,c}^{(o)\prime}\right), \tag{8}$$

$$F_j^{(o)\prime} = T_{2j}^{(\rho)\prime}(b') - T_{2j}^{(\rho)\prime}(-b'),$$
$$B_j^{(o)\prime} = \int_{-b'}^{b'} B_j \, dx_2 , \tag{9}$$

$$\bar{C}'_{abcd} = \bar{C}'_{abcd} - \bar{C}'_{ab12}\,\bar{C}'_{12cd}/\bar{C}'_{1212} ,$$
$$\bar{C}'_{abcd} = \hat{C}'_{abcd} - \hat{C}'_{ab23}\,\hat{C}'_{23cd}/\hat{C}'_{2323} ,$$
$$\hat{C}'_{abcd} = C'_{abcd} - C'_{ab22}\,C'_{22cd}/C'_{2222} . \tag{10}$$

In (7), the plane-stress conditions $T_{21}^{(\rho)\prime} = T_{22}^{(\rho)\prime} = T_{23}^{(\rho)\prime} = o$ have been taken into account through the relations among the elastic stiffness (10).

(c) Governing equations for the lower plating

Equations for the lower plating are exactly the same as those given in (6) to (10), except replacing $(\,)'$ by $(\,)''$.

(d) Continuity of face tractions

We require that face tractions at the interface $x_2 = \pm b$ be continuous,

$$T_{2j}(b) = T'_{2j}(-b'),$$
$$T_{2j}(-b) = T''_{2j}(b''). \tag{11}$$

By substituting definitions of $F_j^{(o)}, F_j^{(1)}, F_j^{(o)}$, and $F_j^{(o)''}$ into (11) and defining

$$\mathrm{F}_j^{(o)} = \mathrm{T}'_{2j}(b') - \mathrm{T}''_{2j}(-b''),$$
$$\mathrm{F}_j^{(o)} = b\,[\,\mathrm{T}'_{2j}(b') + \mathrm{T}''_{2j}(-b'')\,], \tag{12}$$

as the face tractions of plated plate, we obtain

$$F_j^{(o)} = -(F_j^{(o)\prime} + F_j^{(o)''}) + \mathrm{F}_j^{(o)},$$
$$F_j^{(1)} = b\,(-F_j^{(o)\prime} + F_j^{(o)''}) + \mathrm{F}_j^{(1)}. \tag{13}$$

By further substituting equations of equilibrium of the plate and platings, i.e. (1) and (5), into (12), we have

$$\tau_{ij,i}^{(\rho)} + \mathrm{F}_j^{(o)} + \rho\,(\mathrm{B}_j^{(o)} + \mathrm{B}_j^{S}) = o \ ,$$
$$\tau_{ij,i}^{(1)} - \tau_{2j}^{(\rho)} + \mathrm{F}_j^{(1)} + \rho\,(\mathrm{B}_j^{(1)} + b\,\mathrm{B}_j^{D}) = o \ , \tag{14}$$

where

$$\tau_{ij}^{(o)} = T_{ij}^{(o)} + \delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,(T_{ab}^{(\rho)\prime} + T_{ab}^{(\rho)''}),$$
$$\tau_{ij}^{(1)} = T_{ij}^{(1)} + \delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,b\,(T_{ab}^{(\rho)\prime} - T_{ab}^{(\rho)''}),$$
$$B_j^{S} = \frac{\rho'}{\rho} B_j^{(o)\prime} + \frac{\rho''}{\rho} B_j^{(o)''},$$
$$B_j^{D} = \frac{\rho'}{\rho} B_j^{(o)\prime} - \frac{\rho''}{\rho} B_j^{(o)''},$$
$$\delta_{ia}^{\rho} = \delta_{ia} \quad \text{when } i = c$$
$$= o \quad \text{when } i = 2 . \tag{15}$$

Equations (14) are the stress equations of equilibrium of plated plate.

(e) Continuity of displacements

We require that displacements at the interfaces $x_2 = \pm b$ be continuous,

$$U_i^{(o)\prime} = U_i^{(o)} + b\ U_i^{(1)},$$
$$U_i^{(o)''} = U_i^{(o)} - b\ U_i^{(1)}. \tag{16}$$

Insertion (16) into (18), the strain-displacement relations for platings, leads to

$$E_{ab}^{(\rho)\prime} = E_{ab}^{(\rho)} + b\ E_{ab}^{(1)},$$
$$E_{ab}^{(\rho)''} = E_{ab}^{(\rho)} - b\ E_{ab}^{(1)}. \tag{17}$$

By substituting (3) and (7) into (15), and $(15)_2$ and eliminating $E_{ab}^{(\rho)\prime}$ and $E_{ab}^{(\rho)''}$ by (17), we have

$$\tau_{ij}^{(o)} = 2b\ K_{(ij)}\,K_{(kl)}\,C_{ijkl}\,E_{kl}^{(\rho)}$$
$$+ 2b\ \delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\bar{C}^{S}_{abcd}\,E_{cd}^{(\rho)} + 2b\ \delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\bar{C}^{D}_{abcd}\,E_{cd}^{(1)}$$
$$- 2b\ (K_{(ij)}\,K_{(kl)}C_{ijkl}\,\alpha_{kl} + \delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\lambda^{S}_{ab})\,\theta ,$$

$$\tau_{ij}^{(1)} = \frac{2b^3}{3}\,\hat{C}_{ijkl}\,E_{kl}^{(1)}$$
$$+ 2b^2\,\delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\bar{C}^{D}_{abcd}\,E_{cd}^{(\rho)} + 2b^3\,\delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\bar{C}^{S}_{abcd}\,E_{cd}^{(1)}$$
$$- 2b^2\,\delta_{ia}^{\rho}\,\delta_{jb}^{\rho}\,\lambda^{D}_{ab}\,\theta , \tag{18}$$

where

$$\bar{C}^{S}_{abcd} = \frac{b'}{b}\,\bar{C}'_{abcd} + \frac{b''}{b}\,\bar{C}''_{abcd} ,$$
$$\bar{C}^{D}_{abcd} = \frac{b'}{b}\,\bar{C}'_{abcd} - \frac{b''}{b}\,\bar{C}''_{abcd} ,$$
$$\lambda^{S}_{ab} = \frac{b'}{b}\,\bar{C}'_{abcd}\,\alpha'_{cd} + \frac{b''}{b}\,\bar{C}''_{abcd}\,\alpha''_{cd} ,$$
$$\lambda^{D}_{ab} = \frac{b'}{b}\,\bar{C}'_{abcd}\,\alpha'_{cd} - \frac{b''}{b}\,\bar{C}''_{abcd}\,\alpha''_{cd} . \tag{19}$$

Equations (18) are the constitutive equations for plated crystal plates with thermal expansions.

3. Principle of Virtual Displacements

Let $\bar{U}_j^{(o)}$ and $\bar{U}_j^{(1)}$ be components of the zero- and first-order virtual displacements, and

$$\bar{E}_{ij}^{(o)} = \frac{1}{2}\left(\bar{U}_{j,i}^{(\rho)} + \bar{U}_{i,j}^{(\rho)} + \delta_{2i}\,\bar{U}_j^{(1)} + \delta_{2j}\,\bar{U}_i^{(1)}\right),$$
$$\bar{E}_{ij}^{(1)} = \frac{1}{2}\left(\bar{U}_{j,i}^{(1)} + \bar{U}_{i,j}^{(1)}\right), \tag{20}$$

be components of the zero- and first-order virtual strains. By multiplying (14), by $\bar{U}_j^{(o)}$ and $(14)_2$ by $\bar{U}_j^{(1)}$ and integrating the sum of these products over the plate area A, we have

$$\int_A [\tau_{ij,i}^{(\rho)} + \mathrm{F}_j^{(o)} + \rho\,(\mathrm{B}_j^{(o)} + \mathrm{B}_j^{S})]\,\bar{U}_j^{(o)}\,dA$$
$$+ \int_A [\tau_{ij,i}^{(1)} - \tau_{2j}^{(\rho)} + \mathrm{F}_j^{(1)} + \rho\,(\mathrm{B}_j^{(1)} + b\mathrm{B}_j^{D})]\,\bar{U}_j^{(1)}\,dA = 0 . \tag{21}$$

by the use of the two dimensional divergence theorem and (20), (21) can be written as follows.

$$\int_A (\tau_{ij}^{(\rho)}\,\bar{E}_{ij}^{(o)} + \tau_{ij}^{(1)}\,\bar{E}_{ij}^{(1)})\,dA$$

$$= \int_c (P_j^{(o)} \bar{U}_j^{(o)} + P_j^{(1)} \bar{U}_j^{(1)})\, ds + \int_A (F_j^{(o)} \bar{U}_j^{(o)} + F_j^{(1)} \bar{U}_j^{(1)})\, dA$$

$$+ \int_A [\rho\,(B_j^{(o)} + B_j^{\phi})\, \bar{U}_j^{(o)} + \rho\,(B_j^{(1)} + bB_j^{p})\, \bar{U}_j^{(1)}]\, dA\,, \qquad (22)$$

where

$$P_j^{(o)} = n_a\, \tau_{aj}^{(o)}\,, \quad P_j^{(1)} = n_a\, \tau_{aj}^{(1)}\,, \qquad (23)$$

$P_j^{(n)}$, $n = 0,1$, are components of the nth-order edge tractions of plated plate, c is the intersection of the cylindrical boundary of the plate and the middle plane of the plate, s is the coordinate measuring along the curve c, and n_a a = 1,3, are components of the unit normal to c.

Equation (22) is called the principle of virtual displacement of plated plate with thermal expansion. Equation (22), in conjunction with (18), is employed in the finite element method for calculating strains and displacement gradients in the plate.

We note that (14), and (18) reduce to Mindlin's frist-order equations for crystal plates without electrodes and no thermal stresses if $b' = b'' = o$ and $\theta = o$, and they reduce to Mindlin's equations of plated crystal plates[1] if $\theta = o$, excepting that six modes instead of five are included in the present set of equations.

4. Effect of Electrodes on the Acceleration Sensitivity

In case there is no thermal stresses, i.e., $\theta = o$, the functional dependence of acceleration sensitivity on electrode parameters can be expressed by

$$\Gamma = \Gamma\left(\frac{b'}{b}, \frac{b''}{b}, \frac{\rho'}{\rho}, \frac{\rho''}{\rho}, \frac{r'}{r}, \frac{r''}{r}\right), \qquad (24)$$

where r, r', and r'' denote radii of the plate, upper electrode, and lower electrode respectively as shown in Fig. 1.

We consider an SC-cut quartz disk with four-point mount and $\alpha = o$. Acceleration sensitivity Γ is computed for gold electrodes with the following four cases:

(1) No electrodes

$r' = r'' = o$, $r = 7mm$, $2b = 1.1469mm$

$\Gamma_o = 2.583 \times 10^{-10}/g$

(2) Symmetric (w.r.t. $x_1 x_3$ plane) and full electrodes

$r' = r'' = r = 7mm$

$b' = 3870\mathring{A}$, $b'/b = b''/b = 0.34/1000$

$R = (\rho' b' + \rho'' b'')/\rho b = 0.5\%$

$\Gamma = 2.597 \times 10^{-10}/g$

$\delta\Gamma/\Gamma_o = (\Gamma - \Gamma_o/\Gamma_o) = 0.54\%$

(3) Symmetric and partial electrodes

$r'/r = r''/r = 0.5$

$b' = b'' = 3870\,\mathring{A}$

$\Gamma = 2.589 \times 10^{-10}/g$

$\delta\Gamma/\Gamma_o = 0.27\%$

(4) Asymmetric and partial electrodes

$r'/r = 0.3$, $r''/r = 0.5$

$b' = b'' = 3870\,\mathring{A}$

$\Gamma = 2.584 \times 10^{-10}/g$

$\delta\Gamma/\Gamma_o = 0.12\%$

It can be seen that (a) the plate without attached electrodes has the lowest value of Γ, (b) the value of Γ can be reduced to certain extent by electrode arrangement as that given in the case (4) compared with those in cases (2) and (3), (c) for plate back $R \leq 0.5\%$, $\delta\Gamma/\Gamma_o$ is essentially proportional to the ratio of the total mass of electrodes to the total mass of the plate.

Acknowledgement

This work was supported by the U.S. Army Research Office, Contract No. DAAL 03-87-K-0125.

References

1. R.D. Mindlin, "High Frequency Vibrations of Plated Crystal Plates", *Progress in Applied Mechanics*, Prager, Anniversary Volume, pp. 73-84, MacMillan, 1963.
2. P.C.Y. Lee and M.S.H. Tang, *Proc. 41st Ann. Freq. Cont. Symp.* pp. 277-281. 1987.
3. P.C.Y. Lee and M.S.H. Tang, *Proc. 42nd Ann. Freq. Cont. Symp.* pp. 14-18, 1988.
4. P.C.Y. Lee and M.S.H. Tang, *Proc. 43rd Ann. Freq. Cont. Symp.*, 1989.
5. R. Cechmann, A.D. Ballato, and T.J. Lukaszek, *Proc. IRE* Vol. 50, p. 1812, 1962.

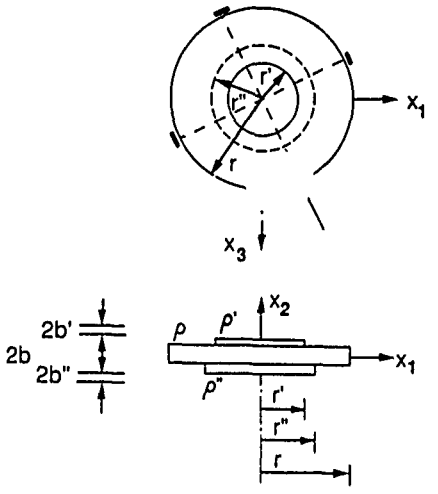


Figure Caption

Fig. 1 Partially plated circular plate with four-point mount.

43rd Annual Symposium on Frequency Control - 1989

QUARTZ CRYSTAL RESONATOR G-SENSITIVITY MEASUREMENT METHODS AND RECENT RESULTS

M. M. Driscoll

Westinghouse Electric Corporation
Electronic Systems Croup
Baltimore, Maryland 21203

Summary

In 1986, the author described the design of a low-noise crystal oscillator in which a (lumped element approximation) quarter wavelength transmission line impedance inverter was used to connect the quartz crystal resonator to the oscillator sustaining stage [1]. If the line is realized using actual coaxial cable, use of the oscillator circuit provides an excellent vehicle for measurement of quartz resonator acceleration sensitivity. Because the resonator can be physically separated and coaxial cable connected to the sustaining stage, independent and accurate determination of resonator and sustaining stage vibration sensitivities can be made by mounting either portion of the circuit on the shake table.

Also, effects of connecting coaxial cable vibration can be determined by measurement of differences in (1) the vibration-induced FM sideband level, and (2) the accelerometer output/FM sideband phase relationship obtained by inverting the position of device on the shake table.

In addition to oscillator-type measurements, it is also possible to utilize a quarter wave, impedance-inverting transmission line for single cable connection to the resonator for acceleration sensitivity measurements using a synthesized signal generator and a passive phase bridge. The passive bridge measurement results clearly demonstrate that, using this technique, valid results can only be obtained for vibration rates less than the half-bandwidth of the resonator circuit transmission response.

Test results were obtained for 40 and 50 MHz, fifth overtone AT-cut and third overtone SC-cut crystals. Acceleration sensitivity (gamma vector) values for the SC-cut resonators were typically four times smaller (5×10^{-10} per g) than for the AT-cut units. However, unit-to-unit gamma vector magnitude and direction uniformity was much better for the AT-cut devices.

Oscillator signal frequency modulation levels resulting from vibration of the 50-ohm modular amplifier TO-8 sustaining stage were significantly lower than the 5×10^{-11} per g value due to cable vibration.

Introduction

The use of quartz crystal resonators as oscillator frequency control elements provides a means for generation of signals exhibiting exceptionally good short-term frequency stability. The electrical properties of the crystal resonator that play an important role in this regard are:

- High Q-f product
- Low 1/f noise
- Moderate drive level capability
- Near-zero temperature coefficient (at turnover)
- Low-acceleration sensitivity

Although the (resonant frequency) sensitivity of the crystal to acceleration is quite low in comparison to other types of resonators, excessive signal spectral degradation can result from crystal oscillator operation in an environment subject to relatively mild levels of mechanical vibration. For example, figure 1 shows the measured signal spectra of a pair of ultralow noise Westinghouse crystal oscillators utilizing fifth overtone, 40 MHz, AT-cut crystals exhibiting acceleration sensitivity (gamma) values on the order of 2×10^{-9} per g. As shown in the figure, per-device output signal spectra in the range 1 to 1000 Hz is characterized by fractional frequency fluctuation levels, $\sqrt{Sy(f)} \sim 1$ to 15 parts in

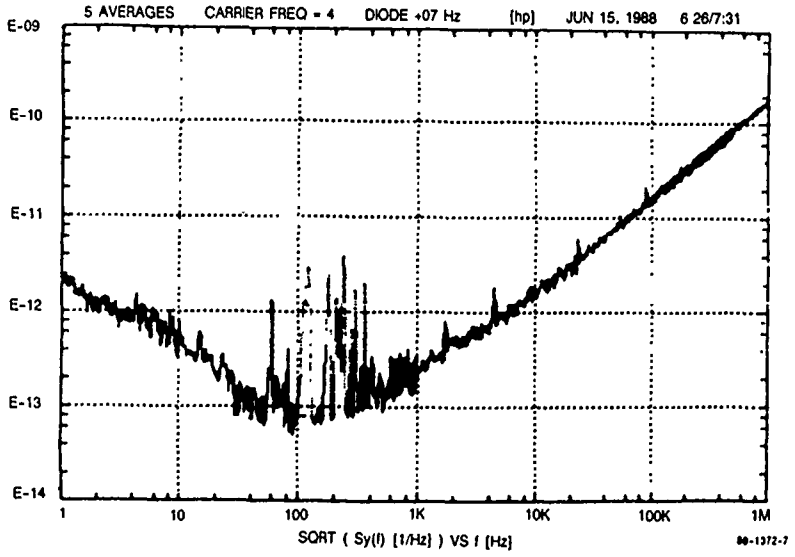


Figure 1. Spectral Density of the Fractional Frequency Fluctuations (Plotted as $\sqrt{Sy(f)}$ for Two Low-Noise Crystal Controlled Oscillators

CH2690-6/89/0000-419 $1.00 C 1989 IEEE

10^{13} [2]. Vibration levels on the order of 0.0001 g per $\sqrt{Hz}$ would therefore be sufficient to degrade oscillator at-rest spectral performance.

The acceleration sensitivity of quartz crystal oscillators is primarily associated with changes in the crystal resonant frequency resulting from stresses induced in the crystal blank reacting against the blank mounting structure during vibration. A great deal of analytical and experimental work has been undertaken with regard to reducing both crystal oscillator and crystal resonator acceleration or g-sensitivity [3-6].

Progress has been somewhat hampered, however, due to differences between theoretical predictions and measured results. In general, large unit-to-unit variations are usually measured with regard to resonator acceleration sensitivity vector magnitude and angle (direction). Moreover, the variability observed in measured data is usually significantly larger than that predicted as a result of various resonator manufacturing tolerances such as blank mount positioning accuracy. [7, 8].

An important aspect of crystal g-sensitivity measurement techniques involves requirements for measurement accuracy and repeatability and the associated ability to identify and quantify signal frequency/phase modulation resulting from vibration of the nonresonator portion of the measurement apparatus such as oscillator sustaining stage circuitry and/or RF coaxial connecting cables.

Oscillator Circuit

Figure 2 shows a schematic diagram for a crystal oscillator circuit employing a broadband TO-8 modular amplifier sustaining stage with crystal resonator connection to the sustaining stage input via a quarter-wave impedance inverter [1]. All or a portion of the quarter-wavelength line of figure 2 can be realized using actual coaxial cable. In this regard, the circuit provides an excellent vehicle for performing crystal resonator g-sensitivity measurements with the resonator alone mounted on the shake table and connected to the (at-rest) sustaining stage via the quarter-wavelength cable. Effects of sustaining stage circuit vibration are eliminated, and the measurement problem is reduced to one of quantifying individual signal modulation levels induced by resonator and by cable vibration.

Measurement Setup

Figure 3 shows a block diagram of the crystal g-sensitivity measurement setup. As shown in the figure, the apparatus is quite conventional except that (1) the resonator alone is mounted on the shake table, and (2) a phase comparison is made between the recovered, vibration-induced, oscillator FM sideband and the accelerometer output. The procedure consists of measurement of the vibration-induced oscillator sideband level $\mathcal{L}(f)$ or fractional frequency modulation $\sqrt{Sy(f)}$ resulting from sinusoidal vibration in a given direction along each of three orthogonal crystal mounting block axes.

Both the FM sideband amplitude and its phase relationship to the accelerometer output is recorded. Modulation rates in excess of the oscillator/VCXO phase lock loop bandwidth are utilized in order to eliminate the added necessity of accounting for test-to-test variations in closed loop bandwidth and associated phase response. The crystal mounting block (figure 4) is then rotated 180 degrees (turned upside down) with the coaxial cable connector loosened and retightened for identical cable positioning, and the measurements are repeated.

Measurement of the recovered sideband amplitude and phase change resulting from crystal orientation reversal allows computation of both the individual frequency modulation levels, resulting from vibration of the crystal and the crystal-connecting coaxial cable, and the relative direction of the crystal vibration sensitivity (gamma) vector.

Figure 5 shows a vector diagram indicating how actual measured results relate to individual crystal and cable vibration sensitivities. As shown in the figure, the cable vibration-induced sensitivity vector direction is not necessarily parallel to that of the crystal. However, in most cases (including the one depicted in figure 5), only slight accuracy degradation results from the simplified calculation

$$|\Gamma_X| = \frac{|\Gamma_{MU}| + |\Gamma_{MD}|}{2} \qquad (1)$$

This measurement procedure (in terms of its ability to determine relative gamma vector direction and isolate cable contributions) is similar to that described in reference [8], except that auxiliary vibration frequency signal drive to the crystal is not required and the procedure can be used for evaluation of AT-cut resonators as well as SC-cut devices.

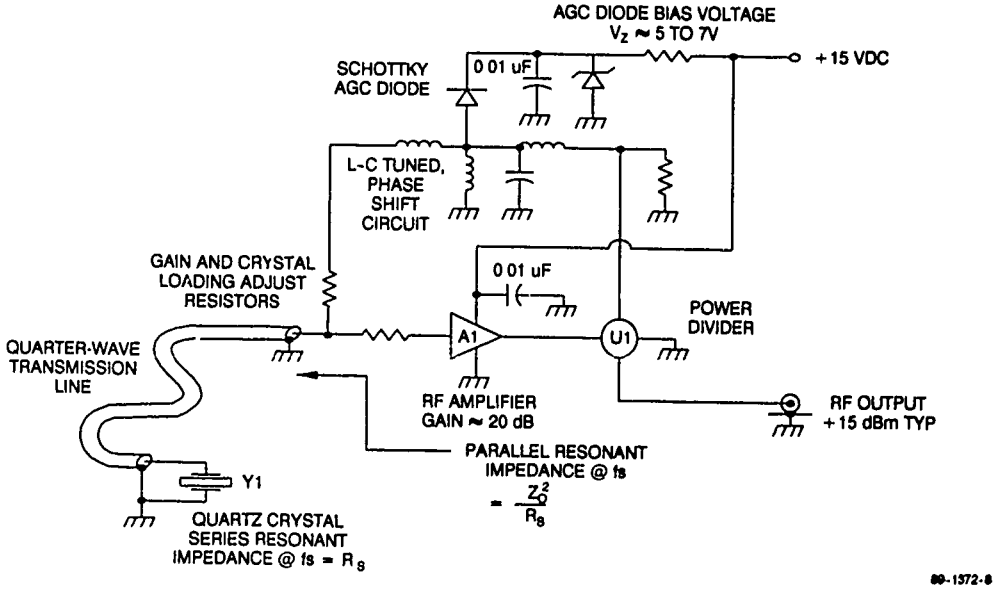


Figure 2. Basic Oscillator Circuit Employing Quarter-Wave Transformation of Crystal Impedance

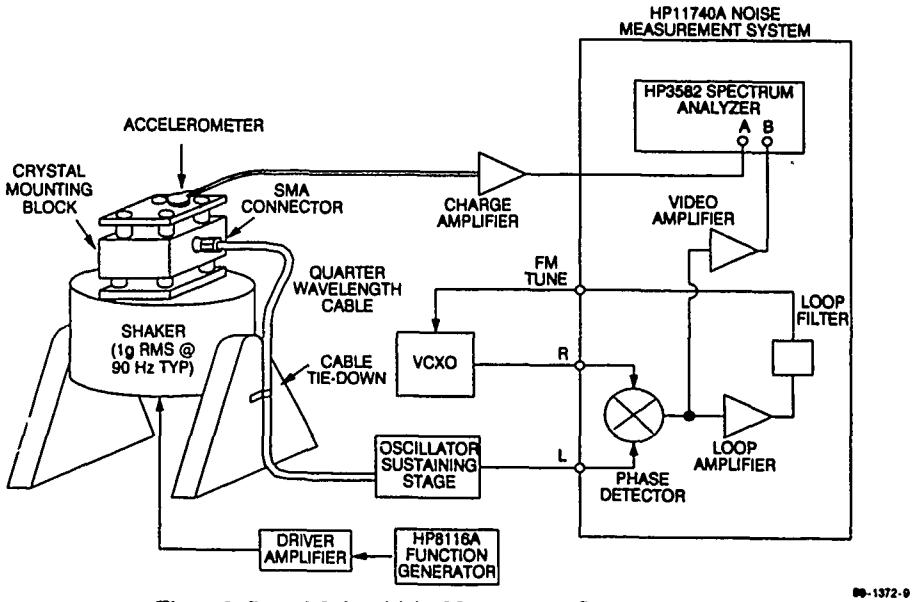


Figure 3. Crystal G-Sensitivity Measurement Setup

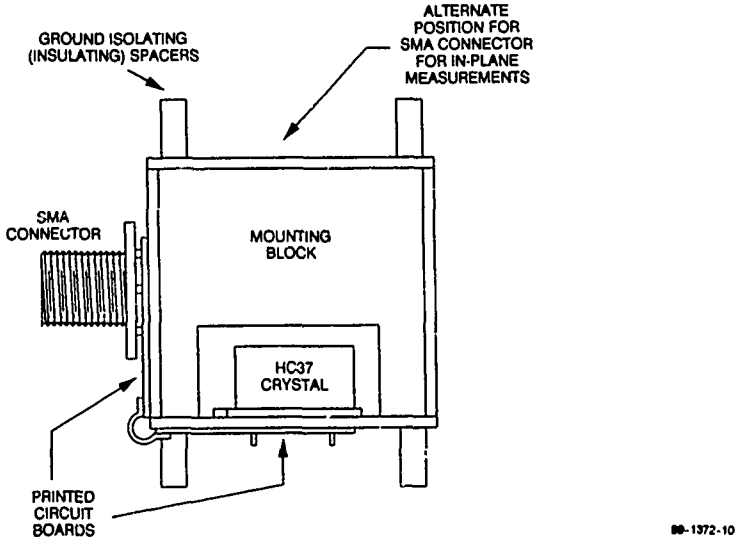


Figure 4. Crystal Mounting Block

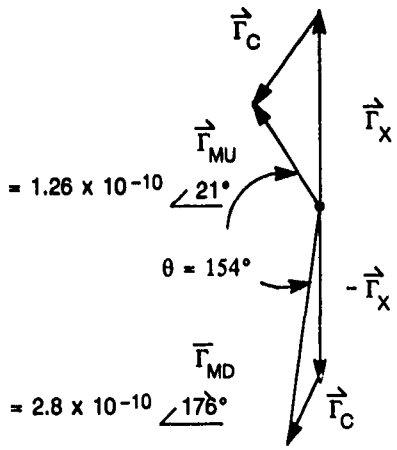
Measurement Results

Table 1 summarizes the results of g-sensitivity measurements for a group of 40-MHz, fifth overtone AT-cut crystals. The directions of the three orthogonal measurement axes with respect to the HC-37 crystal enclosure and crystal blank are depicted in figure 6. Measurements were made at 1-g rms levels with 90-Hz sinusoidal drive. Figure 7 shows a sample output plot of the vibration-induced oscillator 90 Hz, FM sideband level depicted, for convenience, as a fractional frequency fluctuation level. The relative phase angle between the recovered FM sideband and the accelerometer output is measured by operating the phase measurement system (figure 3) in the real-time mode and depressing the phase "XFR FCTN" key on the HP 3582 with the analyzer marker at the 90-Hz vibration frequency.

Tables 2 and 3 show additional measurement results for third overtone, 40 MHz and 50 MHz, SC-cut resonators. Comparison of the SC-cut and AT-cut data shows that AT-cut crystal (vibration sensitivity) gamma vector amplitude is typically three times larger than for the SC-cut crystals, but AT-cut gamma vector direction is much more uniform, as noted by the smaller net vector amplitude variation and by the fact that the vector direction is confined to a single spherical quadrant.

Sustaining Stage Vibration

As previously mentioned, an advantage associated with the use of the oscillator circuit (figure 2), is that either the crystal resonator or the sustaining stage can be separately mounted on the shake table. Figure 8 shows the results of vibration tests conducted with the oscillator sustaining stage, rather than the crystal resonator, mounted on the shake table. As shown in the figure, reversal of the sustaining stage position on the shaker does not result in a corresponding change in the sideband/accelerometer phase reading. This indicates that the induced FM sideband is primarily due to vibration of the sustaining stage-to-crystal cable rather than the sustaining stage itself. From the figure 8 data, cable vibration at 90 Hz is characterized by a sensitivity factor of approximately 5×10^{-11} per g, and the corresponding sustaining stage sensitivity is on the order of 5×10^{-12} per g. Using hard-line cable, cable sensitivity was reduced to approximately 1.5×10^{-11} per g.



$\vec{\Gamma}_X$ = 1.99 x 10^{-10}

$\vec{\Gamma}_C$ = 0.87 x 10^{-10}

$\vec{\Gamma}_{MU}$ = Measured net crystal and cable vibration sensitivity, "up" direction

$\vec{\Gamma}_{MD}$ = Measured net crystal and cable vibration sensitivity, "down" direction

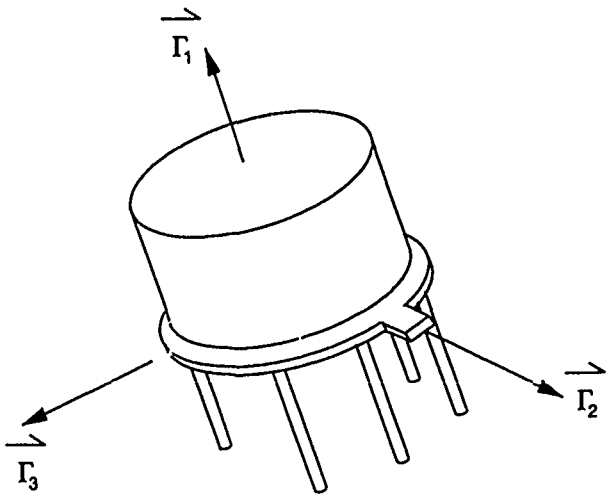
$\vec{\Gamma}_X$ = Crystal vibration sensitivity

$\vec{\Gamma}_C$ = Cable vibration sensitivity

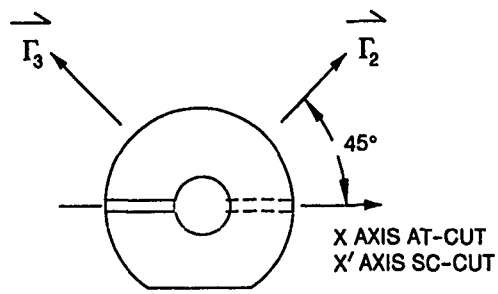
Figure 5. Vector Diagram Showing Typical Out-of-Phase Crystal and Cable Vibration Sensitivities

Table 1. G-Sensitivity Measurement Results for 40 MHz, Fifth Overtone, AT-Cut Crystals (Parts in 10^{10})

Crystal No.	Γ_1	Γ_2	Γ_3	$\lvert\Gamma_{Net}\rvert$
25	-20.4	+2.0	-6.0	21.4
26	-24.0	+9.1	-14.9	29.7
27	-21.9	+4.1	-6.0	23.1
28	-22.3	+7.0	-0.1	23.4
29	-18.8	+1.2	-2.4	19.0



(a) HC37 ENCLOSURE



(b) CRYSTAL BLANK

Figure 6. G-Sensitivity Measurement Axes Γ_1, Γ_2, Γ_3

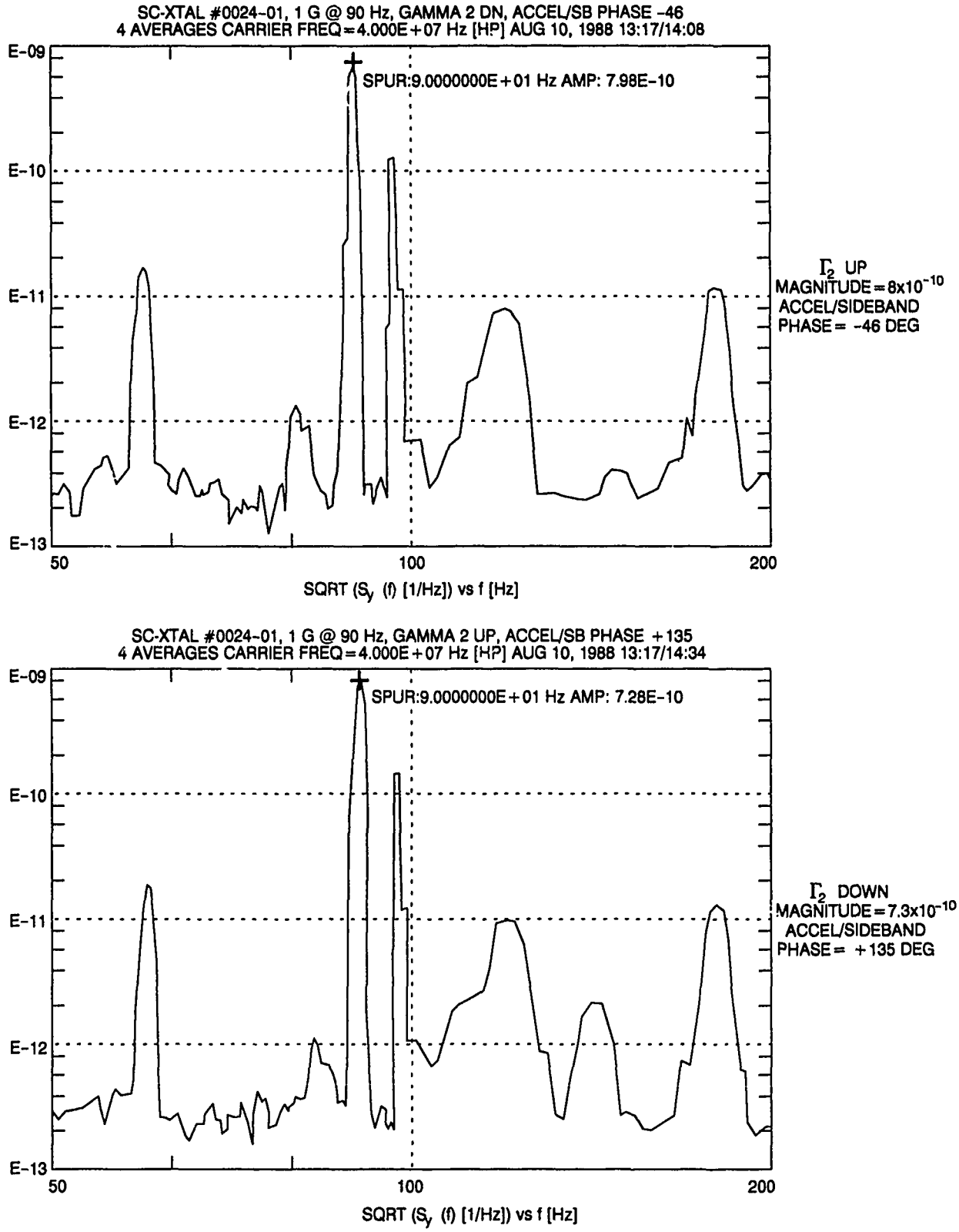


Figure 7. Typical Plotted Measurement Results

Table 2. G-Sensitivity Measurement Results for 50 MHz, Third Overtone, SC-Cut Crystals (Parts in 10^{10})

Crystal No.	Γ_1	Γ_2	Γ_3	$\lvert\Gamma_{Net}\rvert$
71	-4.7	+0.6	+5.3	7.1
02	-5.2	-7.4	-3.4	9.7
73	-1.5	-6.6	-2.0	7.0
74	+3.7	-4.6	-2.3	6.3
09	-11.4	+6.3	+6.3	14.5
76	+2.4	-4.4	-0.5	5.0

Table 3. G-Sensitivity Measurement Results for 40 MHz, Third Overtone, SC-Cut Crystals (Parts in 10^{10})

Crystal No.	Γ_1	Γ_2	Γ_3	$\lvert\Gamma_{Net}\rvert$
09	-1.4	-5.6	+0.3	5.8
11	-6.8	+1.6	+9.5	11.8
13	-4.4	-7.8	+1.8	9.1
15	+0.3	-5.7	+3.3	6.6
19	-3.3	-6.5	-11.5	13.6
20	-3.0	-5.6	-3.7	7.3
01	+0.6	-7.6	+0.9	7.7
05	-0.8	-6.1	-3.2	7.0
10	+2.4	+2.4	-3.1	4.6
08	+4.8	+1.2	+5.2	7.2
02	+0.5	-3.8	+1.4	4.0

Passive Crystal Measurements

In addition to active (in-oscillator) measurements the test setup of figure 3 was slightly modified to perform passive crystal g-sensitivity measurements with the (quarter-wavelength line-connected) crystal installed in a phase bridge. Figure 9 shows the test setup. Using the phase bridge, vibration-induced resonant frequency fluctuations in the crystal result in carrier signal phase modulation, and accurate measurement of the signal transmission path group delay is required to convert the measured phase modulation sideband to corresponding crystal frequency modulation (i.e., vibration sensitivity).

It is very important to note, however, that, when performing passive vibration tests, the level of the vibration-induced PM sideband is modified by the selectivity of the crystal circuit transmission response. Therefore, vibration frequencies substantially less than the resonator (loaded) half-bandwidth must be utilized.

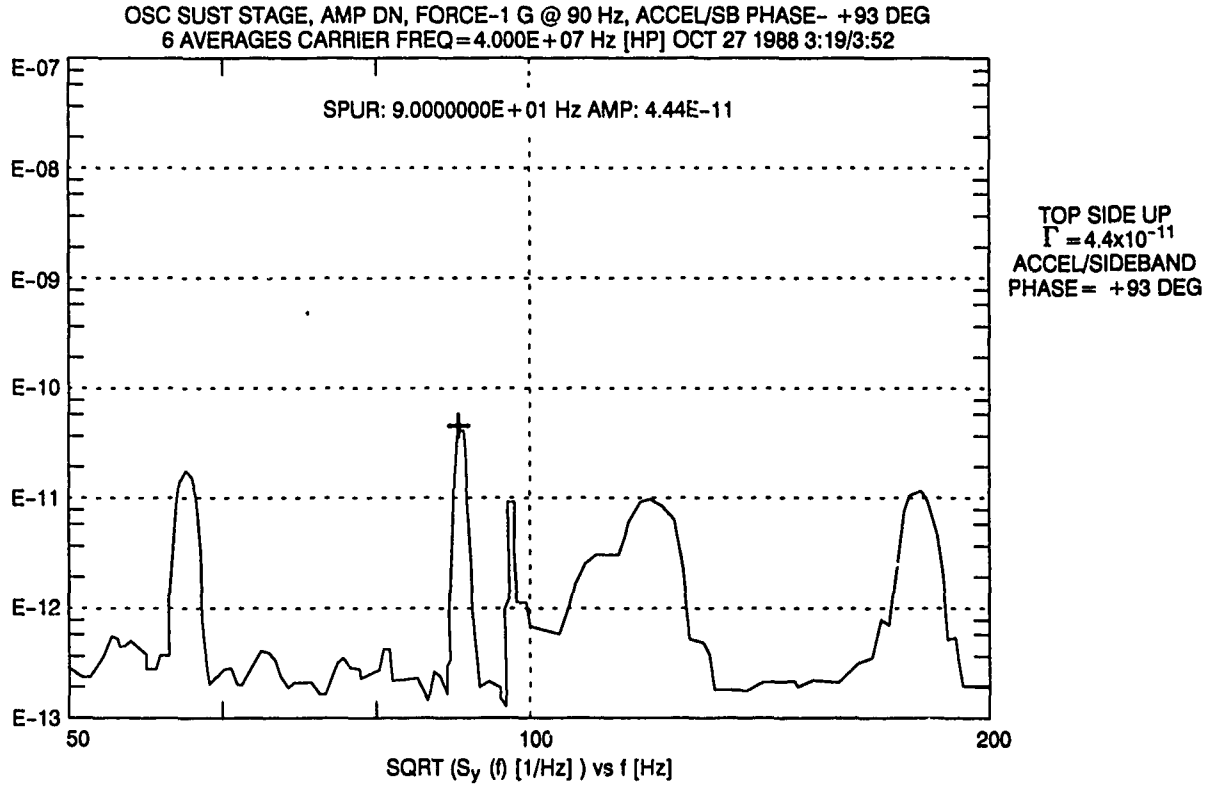
Figure 10 depicts this fact quite clearly. As shown in the figure, the level of the vibration-induced PM sideband is attenuated for vibration frequencies in excess of the 115-Hz crystal circuit half-bandwidth. Using the 35-Hz vibration frequency, -67 dBc sideband data, one obtains the correct corresponding crystal g-sensitivity value of 18×10^{-10}. Using the 800 Hz, -83 dBc sideband data, however, one would incorrectly obtain a crystal g-sensitivity value of 3×10^{-10}. The measurement setup of figure 9 is similar to that of figure 3 in that measured changes in sideband amplitude and sideband/accelerometer phase angle resulting from crystal position reversal can be used to isolate individual crystal and crystal-connecting cable vibration sensitivities.

Conclusions

A crystal oscillator design incorporating intentional use of a quarter-wavelength coaxial cable between the crystal resonator and the sustaining stage provides an excellent test vehicle for performing crystal g-sensitivity measurements.

Measurement of the change in vibration-induced oscillator FM sideband level and sideband/accelerometer phase angle due to crystal position reversal on the shake table allows the individual vibration sensitivities of the crystal and crystal-connecting cable to be determined.

For 40 MHz and 50 MHz, fifth overtone AT-cut and third overtone SC-cut resonators, larger magnitude but more uniform (magnitude and direction) vibration sensitivities were exhibited by the AT-cut resonators.



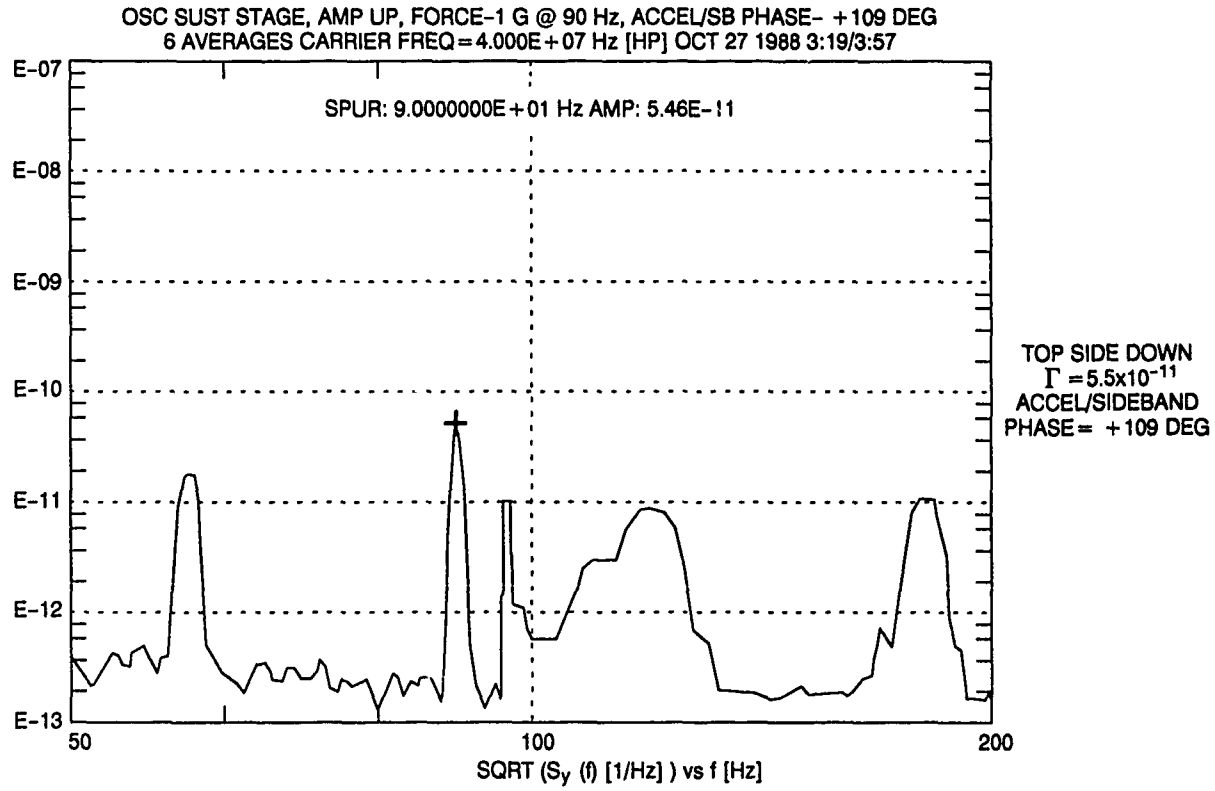


Figure 8. Sustaining Stage Vibration Test Results

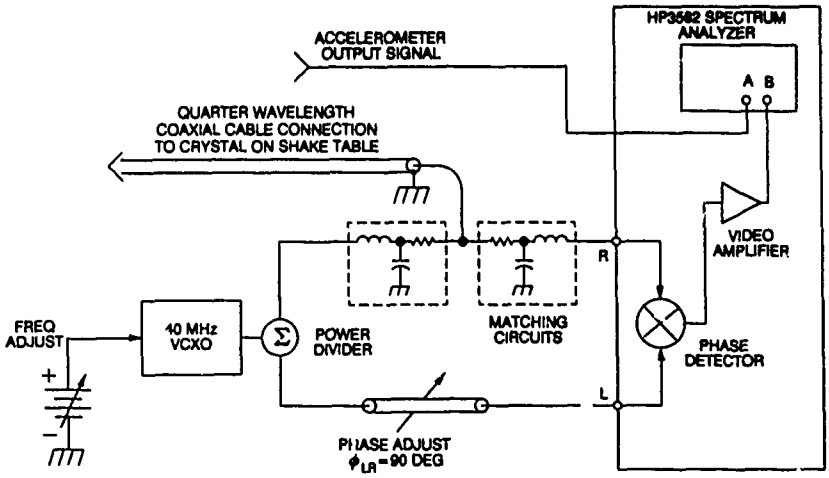


Figure 9. Passive Crystal G-Sensitivity Measurement

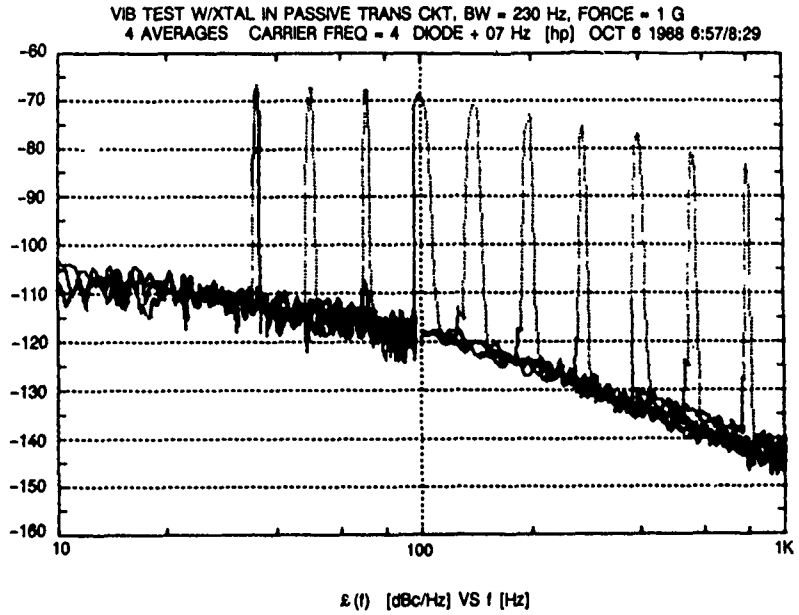


Figure 10. Passive Crystal Vibration Test Results

Typical quarter-wavelength cable effects at 40 MHz are characterized by vibration sensitivities of 5×10^{-11} per g for flexible coaxial cable and 1.5×10^{-11} per g for hard line. TO-8 amplifier oscillator sustaining stage vibration sensitivity is on the order of 5×10^{-12} per g at 40 MHz.

References

[1] M. M. Driscoll, "Low-Noise Crystal Oscillators Using 50-Ohm Modular Amplifier Sustaining Stages," Proc. 40th Annual Frequency Control Symposium, May 1986, pp. 329-335.

[2] M. M. Driscoll, "Low-Noise, Microwave Signal Generation Using Bulk and Surface Acoustic Wave Resonators," Proc. 42nd Annual Frequency Control Symposium, June 1988, pp. 369-377.

[3] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," Proc. 41st Annual Frequency Control Symposium, May 1987, pp. 398-408.

[4] V. J. Rosati, "Suppression of Vibration-Induced Phase Noise in Crystal Oscillators: An Update," Proc. 41st Annual Frequency Control Symposium, May 1987, pp. 409-412.

[5] P. C. Y. Lee and M. S. H. Tang, "Acceleration Effect on the Thickness Vibrations of Doubly Rotated Crystal Resonators," Proc. 41st Annual Frequency Control Symposium, May 1987, pp. 277-281.

[6] H. F. Tiersten and D. V. Shick, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," Proc. 1988 IEEE Ultrasonics Symposium, October 1988, pp. 357-364.

[7] R. L. Filler, J. A. Kosinski, and J. R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," Proc. 37th Annual Frequency Control Symposium, May 1983, pp. 265-271.

[8] M. H. Watts et al., "Technique for Measuring the Acceleration Sensitivity of SC-Cut Quartz Resonators," Proc. 42nd Annual Frequency Control Symposium, June 1988, pp. 442-446.

43rd Annual Symposium on Frequency Control - 1989

Acceleration Sensitivity as a Function of Temperature

William P. Hanson, Timothy E. Wickard
Piezo Crystal Company
100 K Street
Carlisle PA 17013

ABSTRACT

Acceleration sensitivity measurements on quartz resonators have been made as a function of temperature from -50°C to +100°C. One hundred megahertz and ten megahertz resonators were measured. Several cuts were tested including AT, SC, and IT's. Acceleration sensitivity is characterized by the vector Γ. The variation in Γ with temperature, is important in applications where Γ is specified over a wide temperature range or when the vectorial properties of Γ are being utilized to minimize the acceleration sensitivity in a particular direction. This may include ovenized crystal oscillators if Γ is not measured at the oven temperature.

Results indicate that the direction and magnitude of the acceleration sensitivity vector Γ can vary dramatically with temperature. The variations are greater at high frequencies than low frequencies. Likewise, the variations are greater in resonators with small Γ's than with large Γ's. Consequently, as resonators are manufactured with smaller Γ's, the variation of Γ with temperature can be expected to become a greater problem.

INTRODUCTION

Thirty years ago, Warner and Smith reported on the effects of acceleration in resonators for missile applications[1]. In 1987 Filler presented an excellent review of acceleration sensitivity in quartz crystal oscillators with a thorough reference listing[2]. The variation of acceleration sensitivity with temperature has been published in a paper Milliren, Martin and Emmons[3], and observed by Filler and Kosinski in an unpublished study.[4]

In recent years, the demand for resonators capable of maintaining frequency, while under acceleration, has grown. Resonator specifications now frequently include an acceleration sensitivity requirement in the range 0.5 to 2 x 10^{-9} /g. In ovenized oscillator applications, the variation of Γ with temperature is important to consider since the magnitude and direction of Γ may vary between the measurement temperature and the operating temperature of the oven. This problem can be eliminated by measuring Γ at the oven temperature. In non-ovenized applications, the variation of Γ with temperature is important, especially in applications where the vectorial properties of Γ are being used to minimize the acceleration effects. For example, in applications where a particular direction has a greater level of vibration, the Γ may be oriented 90 degrees from this direction to minimize the effects on phase noise.

EXPERIMENTAL PROCEDURE

The effect of temperature on the acceleration sensitivity was measured using an automated test system. Figure 1 is a block diagram of the automated measurement system. The system included an environmental test chamber which subjected the crystals to random vibration while simultaneously controlling the temperature from -5 °C to 100 °C. Data collection was done with a personal computer connected via the IEEE-488 bus to a dynamic signal analyzer. The system also included a ruggedized crystal oscillator, frequency counter and frequency synthesizer.

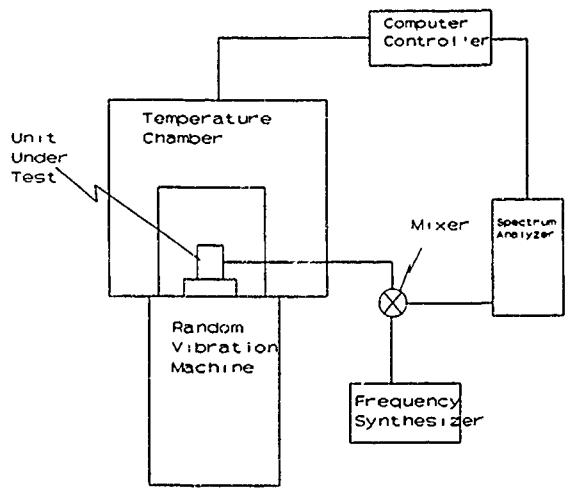


Figure 1: Diagram of Acceleration Sensitivity Measurement System

The oscillators used in this study, were enclosed in milled aluminum housings in order to achieve a 1:1 vibration transmissibility and were foamed to minimize the effects of vibration on the oscillator. The crystal measured was inserted into a socket in the oscillator and a milled aluminum crystal housing was secured to the top of the oscillator housing by four screws. The crystal can was held secure by the crystal can flange. The crystals used in the study were mounted on the X in a four point mount. 10 MHz crystals were sealed in coldweld C-holders and the 100 MHz crystals were sealed in coldweld TO5 cans.

CH2690-6/89/0000-427 $1.00 C 1989 IEEE

In order to minimize the effects of resonances in the test fixtures, transmissibility checks were made on all fixtures used. The checks looked for resonances between the range of 0 Hz and 5000 Hz. Random vibration was used to measure the acceleration sensitivity of the resonators. The random vibration system was set to subject the crystal resonators to a constant power spectral density (PSD) of either 1.0 or 2.0 G^2/Hz over the frequency range 40 Hz to 250 Hz.

The acceleration sensitivity data was collected in three mutually perpendicular axes defined in Figure 2 as A1, A2 and A3. Figure 2 is a drawing showing the definitions of the A1, A2 and A3 axis. The three axis in Figure 2 indicate the direction of the acceleration with respect to the configuration of the resonator. The acceleration sensitivity for the A1 axis, by definition, refers to the Γ measured while the direction of acceleration was normal to the blank.

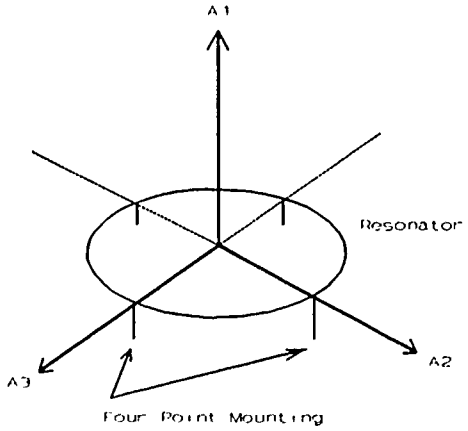


Figure 2: Definition of Axis System with respect to the Crystal Blank

Acceleration sensitivity measurements were taken by mixing a frequency synthesizer with the output of a ruggedized oscillator as shown in Figure 1. The output from the mixer was then put into a spectrum analyzer where L(f) could be measured. Figure 3 shows an example of the output on the crystal resonator while under random vibration of 40 Hz to 250 Hz. The spectrum analyzer was then used to take ten measurements per axis, each measurement consisting of twenty averages. Since the vibration was random over 40 Hz to 250 Hz, it was possible to collect data at 34 different offset frequencies between the range of 20 Hz and 210 Hz from the carrier simultaneously.

The temperature control needed for this investigation was accomplished by using a computer controllable temperature chamber which was modified to be used in conjunction with the random vibration system. Because of the aluminum housings containing the oscillators, the oscillator was first soaked at the temperature to be tested for an hour in order

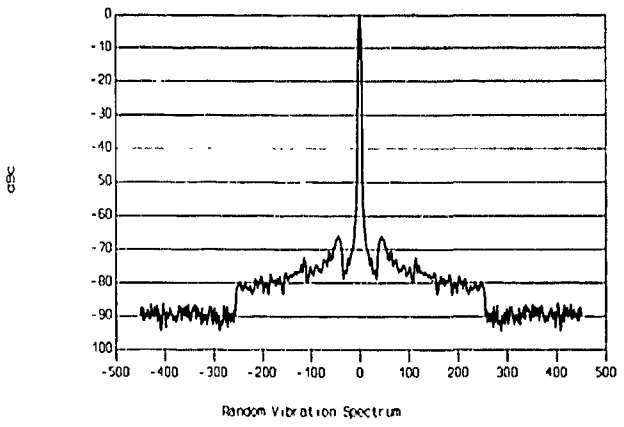


Figure 3: Example of Signal Analyzer Output

to achieve thermal stability. During the measurement, the temperature in the chamber was first dropped to -50 °C. After soaking at -50 °C, 10 acceleration sensitivity measurements were taken by using the method described previously. This process was repeated for each one of the temperatures of interest in the study: -50°C, -40°C, -25°C, 0°C, 25°C, 50°C, 75°C, and 100°C.

In the end, each crystal took a total of thirty six hours to collect the acceleration sensitivity data over temperature. The long test time was due mostly to the length of time needed to achieve thermal stability in the oscillator. As a means of assurance, the temperature was verified by measuring the crystal frequency at each temperature prior to the measurement of Γ. The frequency versus temperature curve for all three axis was plotted for each resonator as shown in Figure 4 for an AT cut 100 MHz resonator.

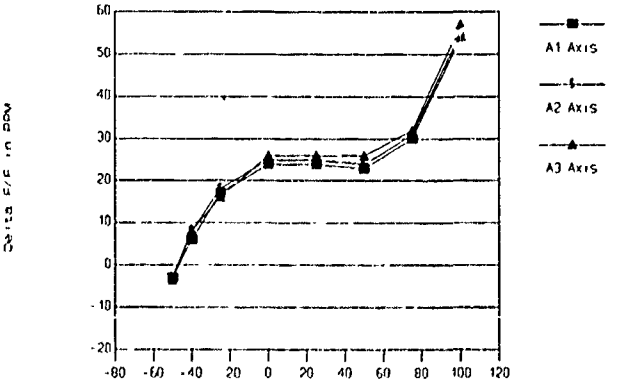


Figure 4: Example of Frequency versus Temperature Curve as Measured During Vibration (Serial Number 26, 100 MHz AT cut)

DATA TRANSFORMATIONS

The raw phase noise data collected for each axis was transformed into acceleration sensitivity data (Γ). After which, the Γ data for each axis was transformed into the magnitude and direction of the vector Γ. Further analysis included the calculation of the mean and standard deviations of each measured parameter.

In order to understand the transformations, it is first necessary to understand the definitions of the angles used. Figure 5 shows the location of Γ with respect to these axis and the definition of the angles α and β. These three axis were chose.ı to eliminate confusion with the crystallographic axis definitions while at the same time allowing the calculation of Γ, α, and β to be done more easily for each resonator measured.

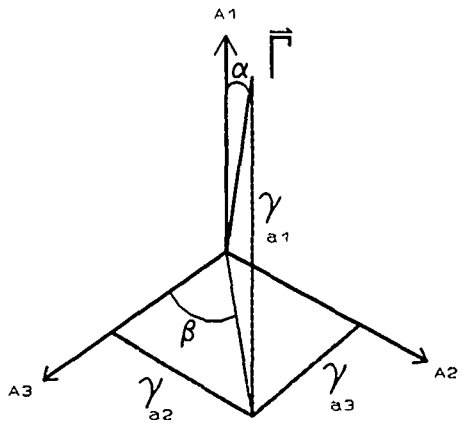


Figure 5: Definition of the Alpha and Beta Angles with respect to the Axis System

As mentioned earlier, random vibration was used to measure Γ because it allowed the automated collection system to collect a large number of data points easily and rapidly. The experimental system measured the phase noise induced by the random vibration, L(f), at various offset frequencies. Using the transformation described below, the acceleration sensitivity was calculated from the vibration induced phase noise. The transformation can be determined by knowing the relationship between the vector Γ (the acceleration sensitivity), the vector **A** (the acceleration) and the phase noise, L(f), as seen in 1.1[5].

$$1.1 \quad \mathcal{L}(f) = 20\mathrm{Log}\left[\frac{\bar{\Gamma}\cdot Af_o}{2f_v}\right]$$

Where:

Γ = acceleration sensitivity vector

|A| = 2 x PSD

f_o = frequency of resonator

f_v = frequency of vibration

$\mathcal{L}(f)$ = single sideband power ratio

The previous equation 1.1 relates L(f) with Γ. However Γ, the acceleration sensitivity vector, is what is of interest. Since L(f) can be measured directly as the power ratio of the sidebands at the frequency of vibration, the magnitude of Γ can be calculated by algebraic manipulation of 1.1 yielding the equation in 1.2.

$$1.2 \quad \Gamma = \frac{2f_v 10^{\left(\frac{\mathcal{L}(f)}{20}\right)}}{|A| f_o}$$

In other words, given a random vibration PSD for a random vibration spectrum the magnitude of the Γ vector can be determined by measuring the value of L(f) corresponding to a particular vibration frequency.

The calculation in equation 1.2 was done for each of the three perpendicular axis measured. This calculation gave the magnitude of the A1, A2, and A3 components. From the components, the magnitude of the total Γ could then be calculated from equation 1.4 by using the definition of axes A1, A2 and A3 as seen in Figure 5, where for the component of total Γ is shown to be parallel to the direction of acceleration.

$$1.3 \quad \bar{\Gamma}_{Total} = \gamma_{A1}\hat{i} + \gamma_{A2}\hat{j} + \gamma_{A3}\hat{k}$$

$$1.4 \quad \left|\bar{\Gamma}_{Total}\right| = \sqrt{\gamma_{A1}^2 + \gamma_{A2}^2 + \gamma_{A3}^2}$$

The next step in the evaluation of the acceleration sensitivity in this study was to calculate the change in the direction of the Γ vector over temperature. The magnitude of the individual Γ's calculated for each axis was used to calculate the variation in the direction of the vector Γ at each temperature measured. From Figure 5 it can be seen that the angle α was defined as the angle of declination from the axis normal to the resonator, A1. The β angle was defined as the angle from the A3 axis to the projection of the gamma vector on the A3-A2 plane. Therefore, using the magnitude of Γ for each axis the value of α and β can be determine and thus the direction of the acceleration vector by equations 1.5 and 1.6.

$$1.5 \quad \alpha = 90 - \mathrm{Tan}^{-1}\left[\frac{\gamma_{A1}}{\gamma_{A2}^2 + \gamma_{A3}^2}\right]$$

$$1.6 \quad \beta = \mathrm{Tan}^{-1}\left[\frac{\gamma_{A3}}{\gamma_{A2}}\right]$$

Data Analysis

Table I is a summary of all resonator data collected. Nine resonators were characterized for Γ as a function of temperature. The columns labeled "delta α/degrees" and "delta β/degrees" are the maximum changes in the α and β angles measured between -50 °C and +100 °C. The columns labeled Γ-low and Γ-high are the lowest and highest Γ measurements between -50 °C and +100 °C. The column labeled delta Γ is the difference between the highest and lowest Γ measurements taken.

The table also includes a statistical summary of the variations. Although, only three 100 MHz resonators were characterized, the differences in delta α and delta β are so large that they are statistically significant. Using the upper three sigma limits of these parameters, defined as the average value plus three times the standard deviation, we can determine the statistical significance of the differences. The upper three sigma limit is also useful in determining the expected variation in large lots of resonators, in the thousands, for a given parameter. The lower three sigma limit, defined as the average value minus three times the standard deviation, is a characterization of how small the parameter can be. In all cases, in this study, the lower three sigma limit is less than zero and can therefore be considered zero. The upper and lower three sigma limits are valid only when the distributions are normal. Normal distributions were assumed in this case even though the sample sizes were small.

The upper three sigma limit of the delta α variations in the 10 MHz resonators is 21.2 degrees which is less than the average delta α of the 100 MHz resonators of 28.8 degrees. The delta β upper three sigma limit of the 10 MHz resonators is 40.0 degrees which is less than the average of 41.4 degrees for the 100 MHz resonators. This indicates the distributions are different for delta α and delta β in the 10 MHz and 100 MHz resonators with better than a 99.97% confidence level. Therefore, it can be stated with confidence that the high frequency 100 MHz resonators tested showed significantly more change in Γ orientation with temperature than the 10 MHz resonators.

The delta Γ is also statistically different between the 10 MHz and the 100 MHz resonators. The upper three sigma limit of the delta Γ's in the 10 MHz resonators is 0.56 x 10^{-9}/g, almost third of the average in the 100 MHz resonators at 1.38 x 10^{-9}/g; an upper 10 sigma limit with essentially a 100% confidence.

Figure 6 is a plot of delta α variations for all 10 MHz resonators as a function of temperature. The key on the right side indicates the serial number and cut of each resonator. The differences between cuts are difficult to determine since only one SC and one IT were tested. The greatest change in α for the 10 MHz was 15.6 degrees. This occurred in serial number 7, a third overtone AT cut.

Table I Data Summary of Acceleration Sensitivity Measurements

10 MHz Resonators

Serial Number	Freq (MHz)	OT	Cut	delta α degrees	delta β degrees	Γ-low	Γ-high	delta Γ
						< 10^{-9}/g		>
350	10	3	AT	3.5	4.2	3.46	3.65	.19
107	10	3	AT	2.4	16.5	1.60	1.90	.30
7	10	3	AT	15.6	2.9	1.98	1.03	.05
B1	10	3	AT	10.0	30.2	0.91	1.11	.20
2401	10	3	IT	7.7	8.4	0.72	0.94	.22
C6	10	3	SC	3.0	4.8	2.17	2.58	.41
Average				7.0	11.2	1.64	1.75	.23
Standard Deviation				4.7	9.6	.95	.97	.11
Upper 3 Sigma				21.2	40.0	4.49	4.65	.56

100 MHz Resonators

Serial Number	Freq (MHz)	OT	Cut	delta α degrees	delta β degrees	Γ-low	Γ-high	delta Γ
						< 10^{-9}/g		>
114	100	3	SC	19.7	76.8	.85	3.28	2.43
35	100	5	SC	66.5	38.1	0.25	0.97	.72
26	100	3	AT	.3	9.2	1.10	2.10	1.00
Average				28.8	41.4	.73	2.12	1.38
Standard Deviation				27.8	27.7	.36	.94	.74
Upper 3 Sigma				112.4	124.4	1.80	4.94	3.63

Figure 7 is a plot of the delta β variations for all 10 MHz resonators. Again, little can be said about the differences between cuts because of the single samples for SC and IT cut resonators. The α angle varied to a greater extent than the β angle in the 10 MHz resonators. One resonator, serial number B1, varied 30.2 degrees in β.

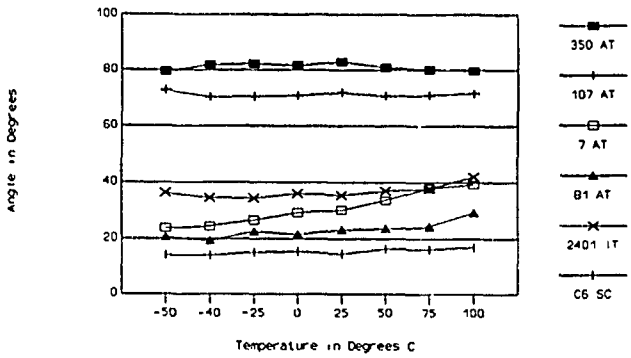


Figure 6: Alpha Angle vs Temperature for 10 MHz Resonators

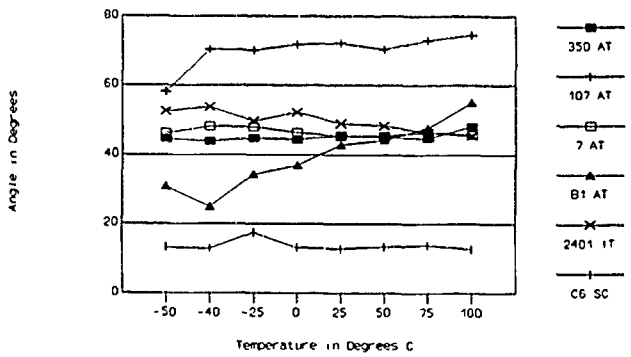


Figure 7: Beta Angle vs Temperature for 10 MHz Resonators

In the 10 MHz resonators, the variation in the magnitude of Γ was on the average 0.23 x 10^{-9}/g. Although this seems small, it is a 13.6% variation from the average Γ of 1.7 x 10^{-9}/g. Figure 8 is a plot of all the Γ magnitude changes as a function of temperature for the 10 MHz resonators.

Figure 9 is a plot of the variations in the α angles for the 100 MHz resonators. The variations in the SC cut resonators were much larger than the AT resonator in these measurements. The AT cut, serial number 26, varies only 0.3 degrees in α. The data was retaken and confirmed. More AT cuts will have to be measured to see if α always has a small variation.

Figure 10 is a plot of the variations in the β angles for the 100 MHz resonators. The AT cut varied only 9.2 degrees in α while the SC cuts varied 38.1 and 76.8 degrees. The drop at 75 °C in β, in serial number 114, is interesting and was confirmed in a second measurement.

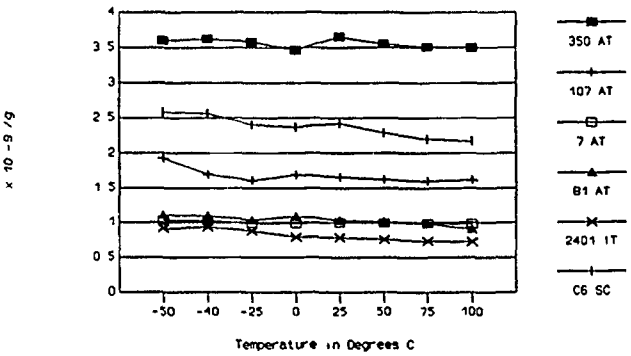


Figure 8: Magnitude of Gamma vs Temperature for 10 MHz Resonators

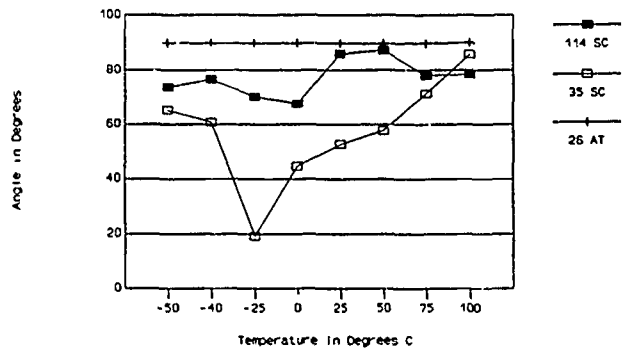


Figure 9: Alpha Angle vs Temperature for 100 MHZ Resonators

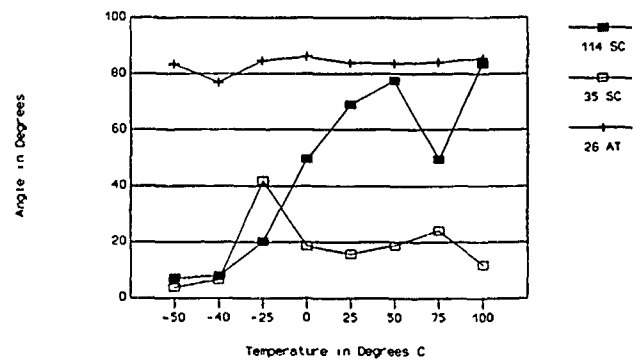


Figure 10: Beta Angle vs Temperature for 100 MHz Resonators

Figure 11 is a plot of the variations in Γ for the 100 MHz resonators. The AT cut resonator which had the least amount of variation in α and β varied about the same in Γ as the SC cuts. Serial number 114 has an interesting cyclic variation. The variation was confirmed in a second measurement. The AT cut, serial number 26 has an interesting peak in Γ and 0 °C

Figure 12 is a plot of all the α data for the 10 MHz and 100 MHz resonators versus the magnitude of Γ. This graph shows that as the

magnitude of Γ decreases the variation in α increases. Figure 13 is a similar plot of all β data for the 10 MHz and 100 MHz resonators versus the magnitude of Γ. This graph also shows the variations in β increasing as the magnitude of Γ decreases.

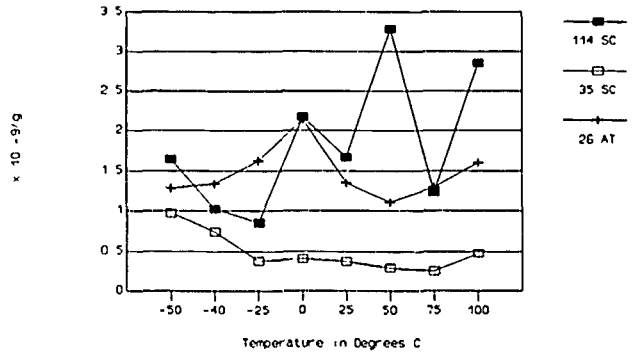


Figure 11: Magnitude of Gamma vs Temperature for 100 Mhz Resonators

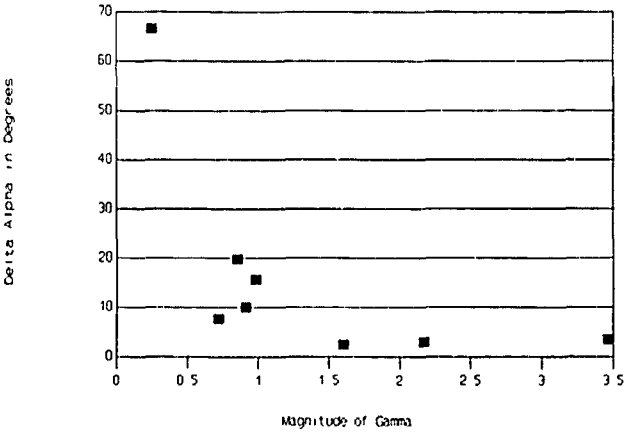


Figure 12: Delta α Angle vs the Magnitude of Γ

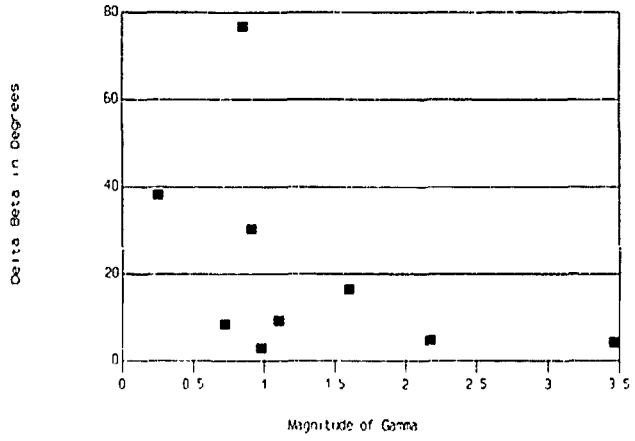


Figure 13: Delta β Angle vs Magnitude of Γ

Conclusions

The effect of temperature on acceleration sensitivity is greater in high frequency resonators than low frequency resonators. Resonators with small acceleration sensitivities have large variations in direction and magnitude.

As resonators with smaller acceleration sensitivities are manufacture,d the variation of acceleration sensitivity with temperature will become more of a problem. Ovenized oscillators at high frequencies will be greatly susceptible to differences in the acceleration sensitivity at the measurement temperature and the operating temperature of the oven. This will require the acceleration sensitivity to be measured at the oven temperature. Non-ovenized oscillators at high frequencies will obviously be the most susceptible to these variations, perhaps resulting in Γ being specified and measured over the operating temperature range of the oscillator.

Acknowledgements

We would like to thank Brian Milliren from Frequency and Time Systems for pointing the problem out to us and helping with the experiment design. We also want to thank Ray Filler from LABCOM for technical discussions and help with the experiment design. Thanks are also due Lee Bowermaster, Bill Taylor and Andy Deglas from Piezo for help in designing and building the test fixturing.

References

1. A. W. Warner and W. L. Smith, "Highly Stable Crystal Oscillators for Missile Applications" 13th ASFC*, 1959, pp 191-206.

2. R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," 41st ASFC*, 1987, pp 298-408.

3. B. T. Milliren, D. W. Martin and D. A. Emmons, "An Acceleration Compensated Precision Quartz Oscillator, " 42nd ASFC*, 1988, pp 334-341.

4. Private communications from R. Filler and J. Kosinski indicate the direction and magnitude of Γ varies with temperature. The results are unpublished, 1989.

5. John R. Vig, Quartz Crystal Resonators and Oscillators For Frequency Control and Timing Applications: A Tutorial, April 1989, pp 4-57 through 4-67.

* Annual Symposium on Frequency Control

THE VIBRATION-INDUCED PHASE NOISE OF A VISCO-ELASTICALLY SUPPORTED CRYSTAL RESONATOR.

R. D. Weglein
Hughes Aircraft Co., Canoga Park, CA 91304.

Abstract

The performance of a VHF crystal oscillator in the 100 MHz range is described under vibration and during acoustic noise excitation The 5th overtone resonator is supported on four massive posts and held in place by a precisely-controlled support area of a compliant adhesive. Because the adhesive is an insulator, electrical connections between crystal and header pins are made via Au-bonded wires. The behavior of the short term stability, composed of near-carrier phase noise and of spurious, flexure-induced resonances experienced under typical vibration and acoustic noise stresses will be reported.

Phase noise measurements, obtained at X-band after ≈100x multiplication, are described on a number of randomly-selected SC and some AT-cut crystals using the compliant mount. All results were obtained in the same foamed hybrid test oscillator excited with sinusoidal excitation of 2 g_{rms} at 1 KHz. These crystals yielded routinely a vibration sensitivity factor $\Gamma \leq 3 \times 10^{-10}$. Microphonics spectra as high as 100 KHz excited with random acoustic noise frequencies up to greater than 20 KHz, important in missile seeker radar applications, were obtained on the crystals themselves, some with compliant mount and some rigidly supported. The former were much less susceptible to the flexural mode-induced resonances.

Introduction

The performance of coherent pulse doppler radar used in tactical missile seekers is adversely affected by vibration and acoustic noise stresses that arise at various times during the mission. In large measure this degradation originates in the reference oscillator from which the radar frequency and timing functions are ultimately derived. The degradation affects the radar performance in several significant respects seen in the characteristic high-PRF transmission and doppler-shifted spectra of Figure I. The quiescent doppler spectrum shows five significant elements:

a) the carrier and its associated line width that is determined in large measure by the reference oscillator's quiescent near-carrier phase noise,

b) the side lobe clutter in the vicinity of the carrier due to unwanted ground return received by the side-lobes of the radar antenna,

c) the main lobe clutter that is offset from the carrier by the platform doppler of the seeker antenna,

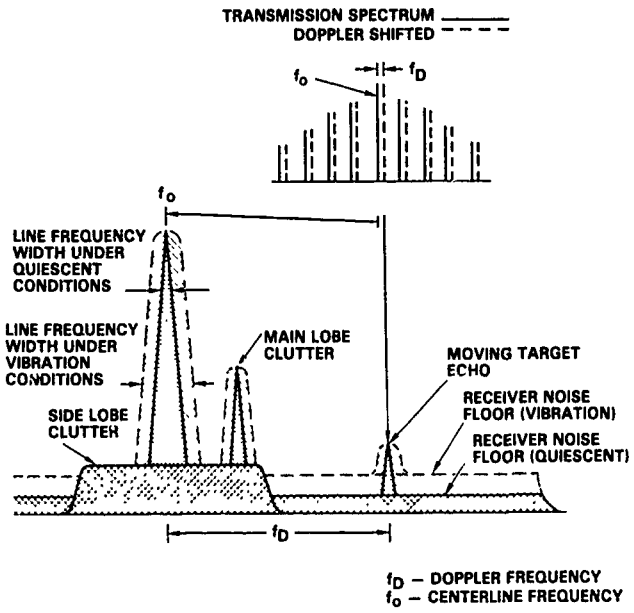


Figure 1 Pulse doppler frequency spectrum at high PRF and under vibration.

d) the moving target echo return that is the desired receiver signal, and

e) the receiver noise floor composed in equal measure of amplitude and phase noise.

Under vibrational stresses that may contains spectra up to perhaps several KHz, it is seen from Figure 1 that both the carrier line (a) and the target echo return (d) broaden substantially, largely due to the phase noise degradation of the reference oscillator. The receiver noise floor (e) is also raised in accordance with the degradation of the reference oscillator's behavior. The wider spectral lines introduce errors in relative target velocity estimation and the higher noise floor reduces the visibility of weak target returns. Considerable effort has been devoted in the recent past to the understanding and reduction of the effects of the vibrational stresses in crystal oscillators. [1] An earlier experimental study summarized the recent state of the art in the vibration sensitivity of commercially-available VHF quartz crystals and oscillators. [2]

Acoustic noise stresses on the reference oscillator, that arise from air turbulence and jet engine noise, with excitation

CH2690-6/89/0000-433 $1.00 C 1989 IEEE

spectra covering virtually the complete doppler range, elicit a different response and degrade the radar doppler spectrum in a manner, typically shown in Figure 2. The spectrum illustrates the phase noise contributions discussed above as well as spurious resonances in the otherwise frequency-independent noise floor. The extraneous spectral lines tend to clutter the desired doppler spectrum and may hide real target echo returns. These spurious, discrete signals may usually be traced to resonances in the crystal

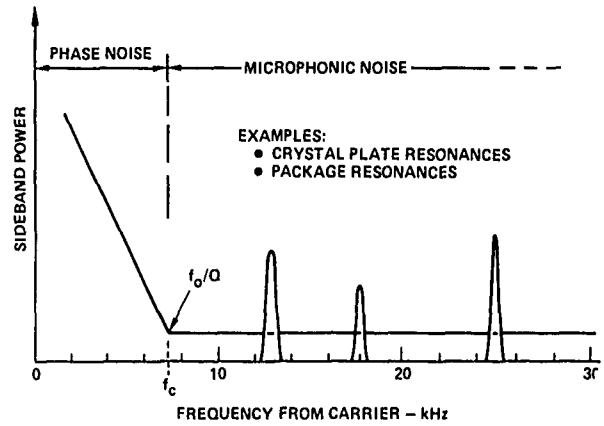


Figure 2 Schematic representation of the oscillator noise with spurious microphonic responses in the otherwise frequency-independent noise floor.

plate that determines the frequency of the reference oscillator. Plate resonances arise from flexural modes at characteristic mode frequencies that manifest themselves in FM sidebands of the oscillator carrier frequency. The resulting spectrum contains a number of discrete carrier spurious sidebands in the doppler range, shown typically in Figure 2. The origin of these flexural modes has been analyzed in the recent past. [3] However, a survey of the published literature shows that little effort has been devoted to the reduction of these spurious "microphonic" responses.

The Present Study.

The results of an environmental stress effects study due to vibration and acoustic noise on a number of 5th overtone SC-cut crystals are reported in this paper. The frequency was in the vicinity of 100 MHz for all crystals. The crystals units are unique in several respects.

The crystal plate is supported on a massive integral four-post structure and secured in place via a viscoelastic, insulating cement material. The term viscoelastic is meant to imply that after curing the cement retains the compliance, characteristic of the generic silicone class of materials. [4] The effect is to place the crystal plate on a "soft" support that relieves thermally-induced stresses that arise during the normal adhesive curing cycle and with subsequent temperature changes. The relative thermal expansion between the crystal place and the four-post support holder is minimized through material choice.

The supporting adhesive is also electrically insulating, requiring other means such as thermo-compression Au-bonded leads to make electrical connections between crystal electrodes and the TO-5 header pins. Finally, the cement contact area and position are precisely controlled in assembly. The completed unit, prior to the final cold-weld sealing operation, is compared in Figure 3 with a standard support structure, used widely throughout the industry, where hard-drying epoxy is employed to secure the crystal plate.

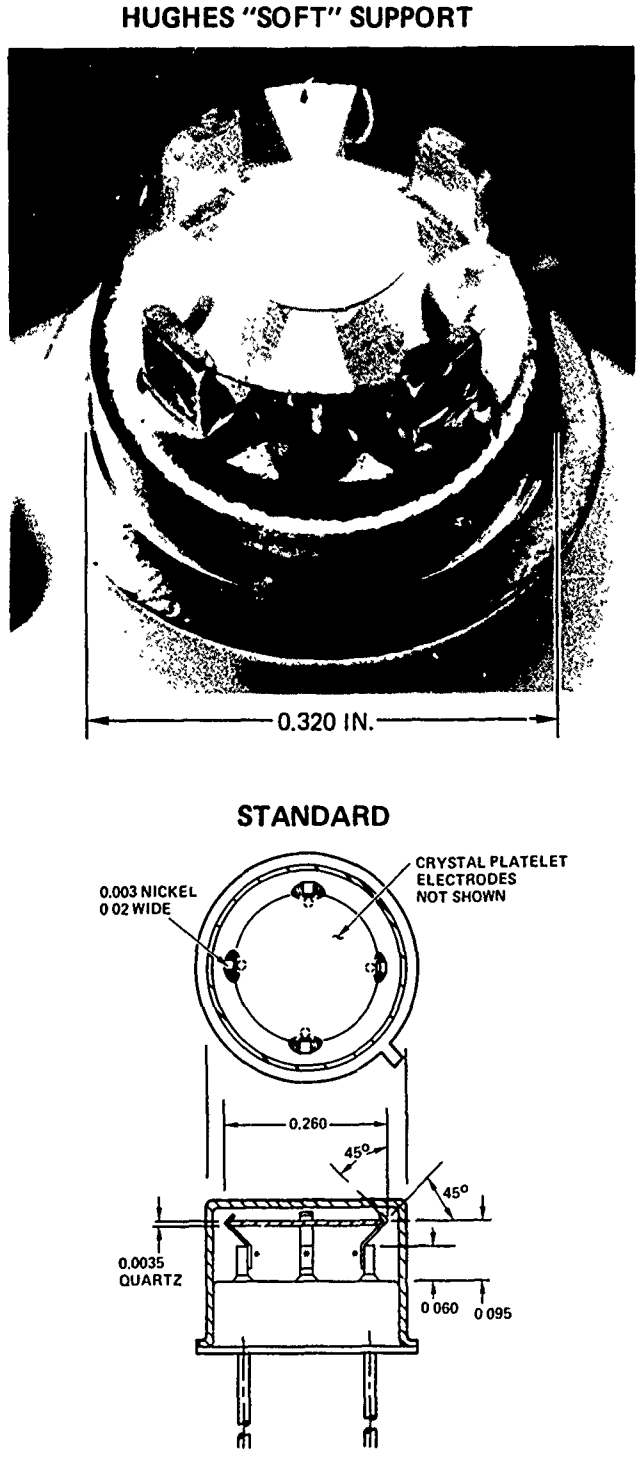


Figure 3 Four-post "soft"-supported crystal resonator plate on TO-5 header with enclosure removed and standard epoxy-supported four-post mount.

In the study, all crystals of the type just described were installed in the test oscillator and subjected to vibration and acoustic noise. The same test oscillator package used throughout these experiments was modified and adapted to replace crystals easily, while maintaining assurance against the possibility of introducing spurious vibration and microphonics effects. Sinusoidal excitation was applied at a vibration frequency of 1 KHz with a peak level of 2.8 g.

Random acoustic noise measurements were conducted on the crystals themselves (without the oscillator). In these experiments only the crystals were mounted on a small shake table with flexible lead connections to the oscillator that was located off the shake table. The purpose of these measurements was to observe the "microphonic" behavior of the crystals. The spectral response to acoustic noise excitation was observed directly at the crystal frequency in the 100 MHz frequency range without multiplication to X-band. This was possible due to the high level acoustic noise spectrum obtained from a 100 watt high-quality audio amplifier that was excited by a solid-state noise source at its input terminals. The resulting frequency response was nearly constant up to 20 KHz and rolled off by 10 dB per octave beyond 25 KHz. The spectrum of the shake table vibration level during such a measurement was obtained with an accelerometer on the table platform and is shown in Figure 4. The integrated acoustic excitation level up to 20 KHz is approximately $\leq 110\ g_{rms}$.

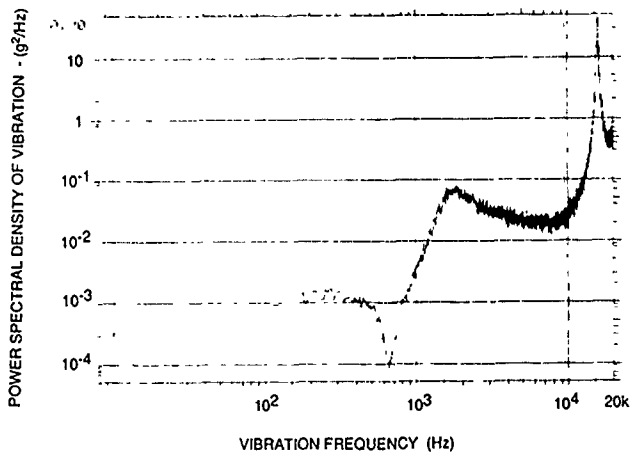


Figure 4 Acoustic noise spectrum excitation for the microphonics investigation.

Experimental Results: Vibration-Induced Phase Noise.

A number of these crystals were subjected to sinusoidal vibration at 1 KHz at a peak excitation level of 2.8 g and along each of the three major axes. A typical result for one crystal vibrated along the z-axis (normal to the plane of the crystal plate) is shown in Figure 5. The figure shows two spectra curves from zero to 25 KHz with the vertical axis in 10 dB steps is nearly superimposed at most places except at the frequency of vibration. The lower curve is measured in the quiescent condition. The upper curve shows the effect of the sinusoidal excitation. As indicated earlier these measurements were taken at approximately 10 GHz, after

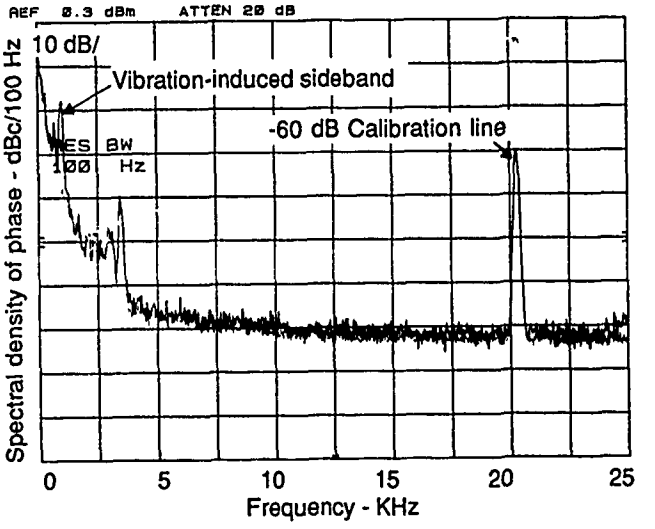


Figure 5 Measured oscillator phase noise along z-axis using crystal No. 303; the frequency of vibration is 1 KHz at a peak excitation level of 2.8 g.

appropriate multiplication from the nearly 100 MHz crystal frequency. This measurement, as all subsequent measurements, was taken with a resolution bandwidth of 100 Hz so that the power spectral density is in dBc/100 Hz. The level of the calibration line near the 20 KHz frequency is -60 dB for the 100 Hz noise bandwidth. It is seen that the peak of the 1 KHz vibration-induced phase noise is -48 dBc/Hz corresponding to a vectorial vibration sensitivity Γ_z (along z) of 2.3×10^{-10}. The irregular spikes, equally reproduced in both static and vibration curves, are attributed to the frequency synthesizer used in these experiments. Three-axis measurements on a number crystals of this type were performed and are summarized in Table 1. In each case, the vibration sensitivity vector Γ was computed from well-established principles[1] and is listed along with the mean $|\Gamma|$ amplitude. Also listed is $\mathcal{L}(f)$, the power spectral density peak level of the measured 1 KHz sideband. In a few cases the sideband magnitude was below the instrument threshold and is stated by "low". The predominantly worst-case vibration axis is perpendicular (along z) to the plane of the crystal plate. These results as well as the mean vibration sensitivity magnitude $|\Gamma| = 2.20 \times 10^{-10}$ from Table I correspond closely to the results obtained with AT-cut resonators in an earlier study[2].

Experimental Results:
Acoustic Noise-Induced Responses.

The crystal spectral response to the acoustic noise random excitation was observed on a spectrum analyzer with an available hard-copy output option. A total of 53 "soft-mounted" four-post crystal plates were subjected to the measurement routine described above. A typical spectral response of one of these crystals (SN 280) is shown in Figure 6. The linear frequency axis contains the spectrum from zero to 100 KHz and the vertical axis is in 10 dB per division increments. The heavy horizontal line denotes the calibration line at -50 dBc per KHz bandwidth, or -80 dBc/Hz.

Table I - Measured Vibration Sensitivity

Crystal S. No.	Axis	$\mathcal{L}(f)$ (dBc/Hz)	Γ_i (10^{-10})	$\lvert\Gamma\rvert$ (10^{-10})	Worst Case Axis
321	x	-58	0.9		
	y	-51	2.1		
	z	-50	2.3	3.2	z
303	x	-62	0.57		
	y	low	low		
	z	-48	2.3	0.96	z
330	x	-53	1.6		
	y	-57	1.0		
	z	-57	1.0	2.3	y/z
310	x	-58	0.9		
	y	-63	0.5		
	z	-54	1.4	1.7	z
321	x	-58	0.9		
	y	-51	2.0		
	z	-51	2.0	3.0	y/z
290	x	-60	0.7		
	y	-47	3.2		
	z	-54	1.4	3.6	y
301	x	-54	1.4		
	y	low	low		
	z	-63	0.51	0.64	x

The 5th overtone frequency of the SC-cut crystal at approximately 100 MHz is represented by the main response at the left. It is seen that there is a nearly complete absence of the effects of the high-level excitation to which the crystal is subjected. Three crystal responses of this lot - or 13 % - showed minor evidence of the acoustic noise excitation. In the worst case of this lot (SN 309) shown in Figure 7, there is evidence of a raised noise floor that gradually increases to nearly 15 dB above the quiescent noise floor.

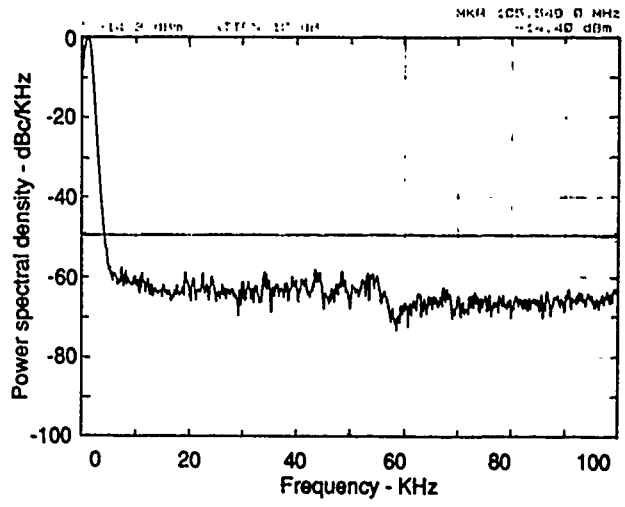


Figure 7 Worst case spectral response near 100 MHz of "soft-mounted" crystal SN 309 to acoustic noise spectrum of Figure 4; the ordinate is in units of dBc/KHz.

The spectral response of similar SC-cut 5th overtone crystals, but secured in place with hard-drying conductive epoxy cement, is shown in Figure 8. A total of 53 hard-mounted crystal plates were examined in this study. All crystals yielded spurious plate resonances at somewhat lower levels of the type shown in Figure 8. but well above the noise floor. The spurious response level of 19 of the 53 crystal plates or 36 % of the total showed spurious responses above the -50 dBc/KHz level.

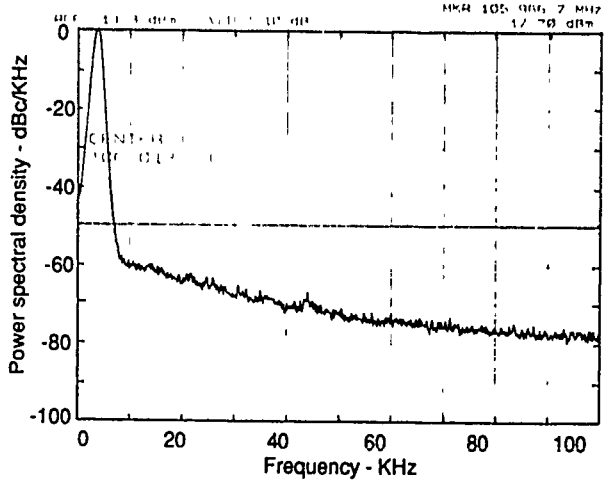


Figure 6 Typical spectral response near 100 MHz of "soft-mounted" crystal SN 280 to acoustic noise spectrum of Figure 4; the ordinate is in units of dBc/KHz.

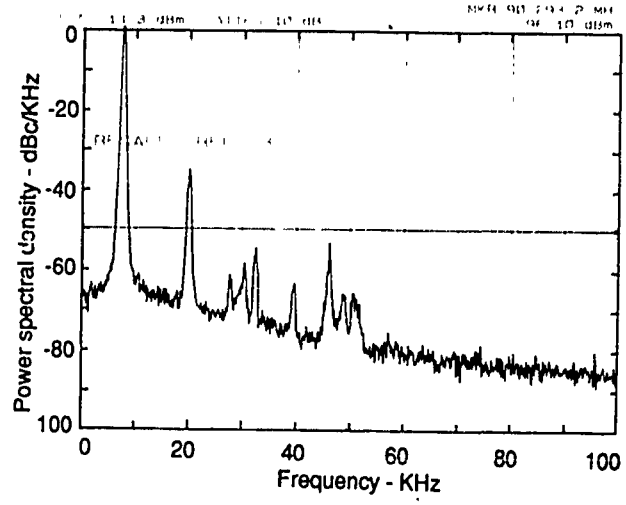


Figure 8 Typical spectral response near 100 MHz of hard-mounted crystal SN 61 to acoustic noise spectrum of Figure 4; the ordinate is in units of dBc/KHz.

Aging Results

The acceptance of the viscoelastic material as a reliable, stable adhesive for the crystal plate is predicated on the premise that no detrimental long-term efffects are introduced. The adhesive's effect on aging, that might possibly result from excessive outgassing of the compliant adhesive has been examined. Aging studies have been caried out on a continuing multi-year basis on the compliantly-supported crystal units and hard mounted samples as well. These studies conducted over a two-year period have revealed an average aging rate of less than 0.6 ppm per year for the compliantly-supported crystal units and approximately 0.3 ppm per year for the commercially-procured hard-mounted crystal units.

Discussion

The effect of two types of vibrational excitation on SC-cut crystals in the 100 Mhz frequency range, soft-mounted on a four-post support with a proprietary compliant adhevise, were investigated in this study. In the first evaluation, the crystal oscillator, containing these crystals were excited by sinusoidal vibration at 1 KHz at a peak level of 2.8 g. The oscillator output was multiplied to X-band and the phase noise measurements were recorded at approximately 10 GHz. These results show a vibration sensitivity magnitude $|\Gamma|$ in the $\leq 3.6 \times 10^{-10}$ range. Similar results had been obtained in an earlier investigation[2] with AT-cut crystals, but in otherwise identical circumstances, e.g., the identical crystal mounting method using the compliant adhesive, the same test oscillator and the same evaluation technique. The similarity of these results on different crystal cuts, but otherwise identical crystal mounting techniques suggests that the dominant factor of importance in achieving low $|\Gamma|$ is the crystal plate support structure and to a lesser extent the crystal cut. This is in agreement with earlier conclusions reached in the referenced study [2] and, therefore, constitutes a distinguishable parameter that controls to a large extent the vibration sensitivity of four-post supported crystal resonators [1].

The crystal support structure in the presently studied group is unique in several respects as was discussed earlier in this paper. The reduction of the relative thermal expansion coefficient between crystal holder and crystal in conjunction with the compliant adhesive tends to reduce thermal stresses in the plate, resulting in reduced stress bias at any temperature. Stress bias is likely to distort the symmetrical distribution of energy within the crystal that could adversely afffect phase noise.

Further, the adhesive crystal contact area at each of the four posts in these crystals is well controlled. The contact area is small, approximately circular, but sufficiently large to prevent loss of adhesion, even under the large vibrational forces that were used in this study. The circular support area at one post is shown in detail in Figure 9. The small, reproducible contact area and location at each of the four posts approaches the ideal support conditions - a point support - used in theoretical calculations that predict $|\Gamma|$ in the 10^{-11} range[5] The small, reproducible contact area is, therefore, a practically achievable compromise to the theoretical boundary condition that cannot be realized in practice.

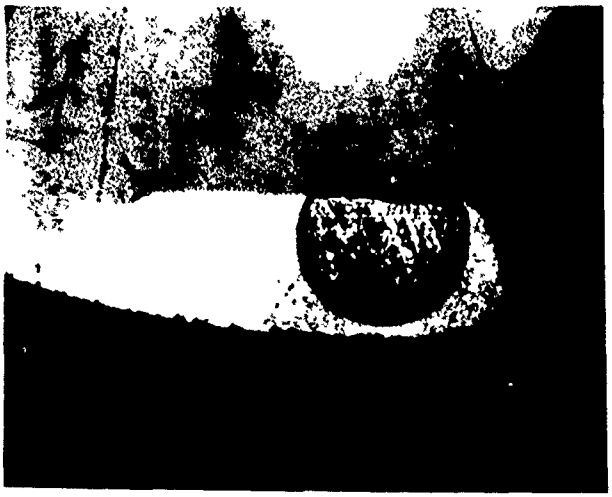


Figure 9 Detailed view of adhesive area (≈0.7 mm dia.) between the crystal plate (top) and the nearly rectangular support post (light area below plate).

There is a growing awareness of the importance of obtaining symmetry in the resonator support structure to effecting reduced phase noise under vibration [2,7]. In the earlier instance, symmetry in the suspension plane normal to the crystal plate was noted [2]. Recently, deviations from lateral symmetry in the resonator plane has been experimentally shown to contribute substantially to the phase noise degradation under vibrational stresses[7]. The well-controlled area and position of the adhesive in the Hughes-developed resonator contributes substantially to the symmetry of overall support at each of the four posts.

Finally, the adhesive that remains compliant and has viscoelastic damping properties, tends to act as an acoustic absorber thus reducing the vibrational energy that reaches the crystal plate. This latter effect is also thought to be dominant factor in reducing the excitation of flexural modes in crystal plates captivated with viscoelastic adhesive . It is the excitation of these flexural modes that result in FM sidebands in the vicinity of the crystal resonance frequency.[3] The energy in the excited flexural waves on the crystal plate is in all likelihood absorbed and converted into heat via the viscoelastic properties of the adhesive layer as earlier studies have shown.[6]

Conclusions

Resonators using visco-elastic supports tend to yield low and reproducible vibration sensitivity. The low Γ is attributed to a number of factors: a) small area contact and an accurately-controlled symmetric support position; b) reduced stiffness of the resonator capture produced by the compliant adhesive; and c) optimum crystal to post orientation that achieves minimum Γ over a wide angular range. The resonator support conditions appear to be of equal importance to the resonator crystal cut in maintaining low Γ. At acoustic noise frequencies, the visco-elastic adhesive behaves as an effective acoustic absorber that effectively reduces the vibrational stress on the resonator, thus minimizes the excitation of flexure-induced sidebands.

Acknowledgement

The author gratefully acknowledges the assistance of Mr. Charles Brancroft and his associates of the Hughes Radar Systems Group in carrying out the measurement reported in the paper.

References

1.) R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review", IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, pp. 297-305, vol. 35, No. 3, May 1988.

2.) R. D. Weglein, "The vibration sensitivity of VHF quartz crystals for missile applications", 38th Ann. Symp. Freq. Cont., Cat. No. 84CH2062-8, pp. 73-79, 1984.

3.) R. D. Weglein, W. C. Hu, A. F. Margagliano, "Microphonics in VHF Quartz Resonators and Oscillators", Proc. IEEE Ultrasonics Symp., vol. 1, Cat. No. 84CH2112-1, pp 378-382, 1984.

4.) The particular adhesive formulation is a propietary development of the Hughes Aircraft Company.

5.) P. C. Y. Lee and M. S. H. Tang, "Acceleration Effect on the Thickness Vibrations of Doubly-Rotated Crystal Resonators", 41st Ann. Symp. Freq. Cont., pp. 277-281, Cat. No. 87CH2427-3, 1987.

6.) E. M. Kerwin, Jr., "Damping of Flexural Waves by a Constrained Viscoelastic Layer", Journ. Acoust. Soc. Am., vol. 31, No. 7, pp. 952-962, July, 1959.

7.) E. EerNisse, et al, " Acceleration-Induced Frequency Shifts in Quartz Resonators", Session A3, this Symposium.

43rd Annual Symposium on Frequency Control - 1989

MECHANICAL AND ACOUSTIC EFFECTS IN LOW PHASE NOISE PIEZOELECTRIC OSCILLATORS

MM. Patrick RENOULT - Emmanuel GIRARDET - Louis BIDART
QK PIEZOELECTRONIQUE

ABSTRACT

Modern electronics evolve within more and more hostile environment.

The improvements in performances are altered by mechanical and acoustic strains imposed on the system.

Piezoelectric materials such as quartz or Lithium Tantalate are quite sensitive. Low phase noise oscillators OCXO, TCXO, VCXO must be operational without any performances losses in presence of mechanical vibrations or acoustic noise.

We propose solutions applied to the design and realization of oscillators called "hardened", in order to make the piezoelectric equipment insensitive to mechanical or acoustic disturbances.

INTRODUCTION

Modern electronic systems must work under more and more severe strains due to hostile environment. The system must keep constantly all its electrical characteristics in spite of thermal or mechanical strains.

Among the different operating conditions, mechanical vibrations and acoustic noise are very important. Starting at the design phase, an appropriate protection must be foreseen.

In the last few years, there has been much discussions about the importance of acceleration sensitivity (g sensitivity) for piezoelectric resonators and how to improve this parameter.

However, it is also necessary to consider the acoustic sensitivity of the frequence source.

It is obvious that electronic devices installed in a rocket, a missile or a fighter plane undergo both mechanical random vibrations and acoustic random noise at the same time. Both have an effect on the phase noise characteristic of the piezoelectric source.

In practice, the acceleration sensitivity of the resonator will disturb the phase noise figure for frequencies that are located close to the carrier. At the same time, acoustic sensitivity of the resonator will degrade the noise floor of the curve.

In order to keep the phase noise performances that have been obtained with a lot of difficulties, it is necessary to find appropriate solutions.

ACCELERATION SENSITIVITY AND PHASE NOISE

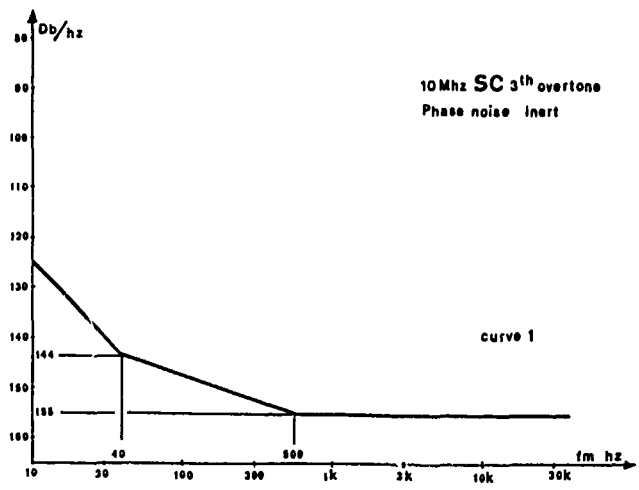
In order to preserve the phase noise performances during mechanical vibrations, it is necessary either to reduce or neutralize the g sensitivity of the used resonator.

For the demonstration, we consider a random vibration profile of 0,02 g^2/Hz in the frequency range 0-2000 Hz. Practical vibrations can have more complexe spectrums.

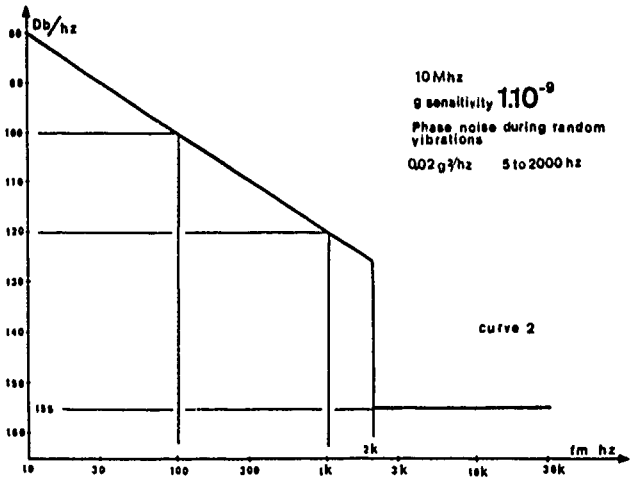
Let us consider a 10 MHz oscillator with a phase noise figure given by curve 1.

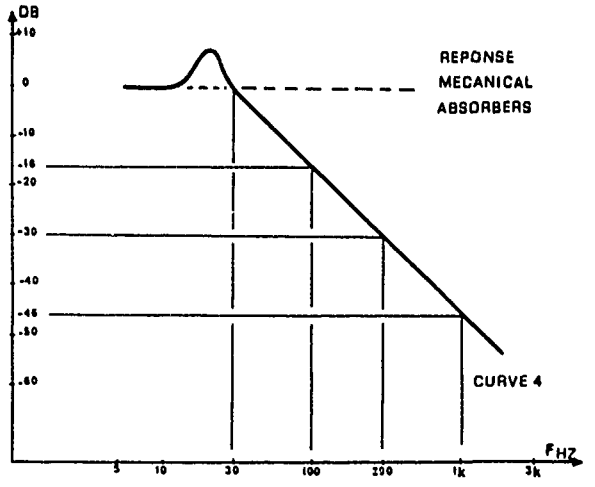
If the resonator has a g sensitivity of 1.10^{-9}/g the phase noise curve of the oscillator will be degraded by the random vibrations as shown on curve 2.

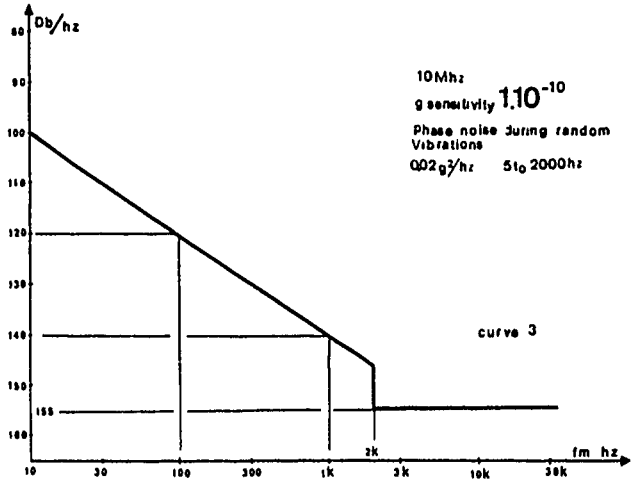
The same oscillator with a 1.10^{-10}/g sensitivity resonator will give in the same conditions curve 3.

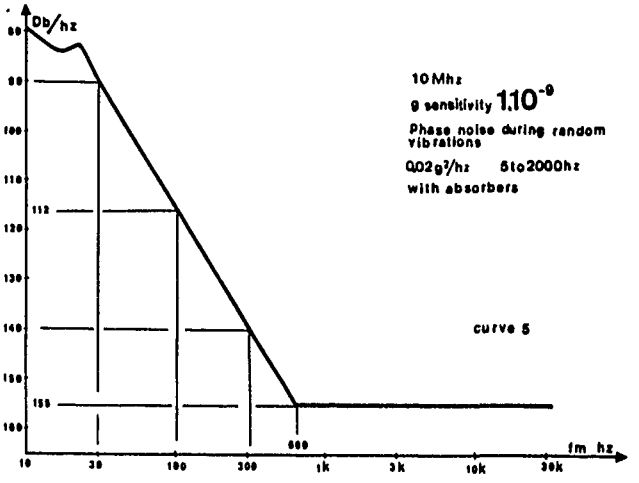


CH2690-6/89/0000-439 $1.00 C 1989 IEEE









It is possible to package the oscillator sub-assembly in a suspended enclosure. We use resilient rubber mounts working as a mechanical filter. Such filter has a typical frequency response as shown on curve 4, with a cutoff frequency of about 30 Hz. Of course, the shape of the curve depends both on the elasticity characteristic of the absorbers and on the suspended mass.

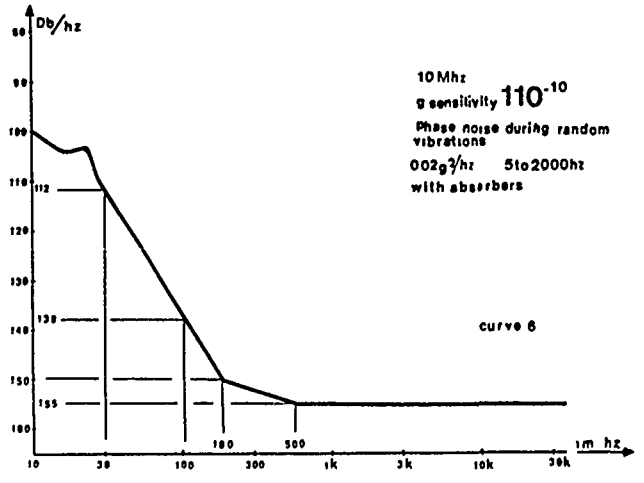
Using these absorbers, an oscillator driven by a resonator with a 1.10^{-9}/g sensitivity will have a phase noise curve during vibrations as shown on figure 5.

For the cut off frequency of the mounting structure, the phase noise will depend first on the g sensitivity of the resonator.

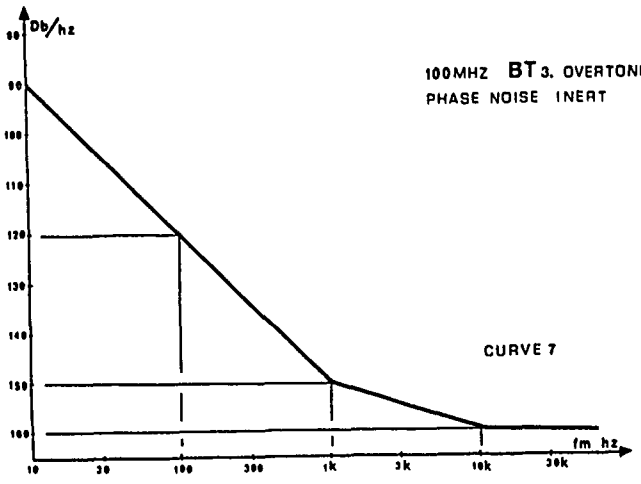
After the cut off frequency (30 Hz), the resulting phase noise is a combination of g sensitivity and mechanical filtering.

For Fourier frequencies greater than 600 Hz, the phase not will not be disturbed by the mechanical vibrations already described above.

The same oscillator and mechanical technology will give, with an 1.10^{-10}/g resonator phase noise as shown on curve 6.

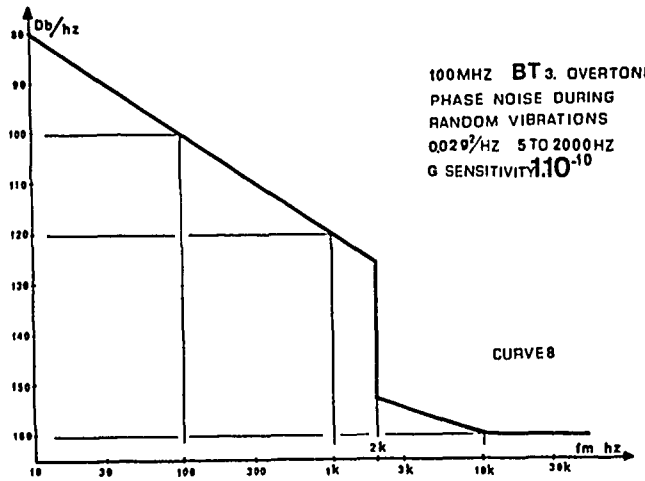


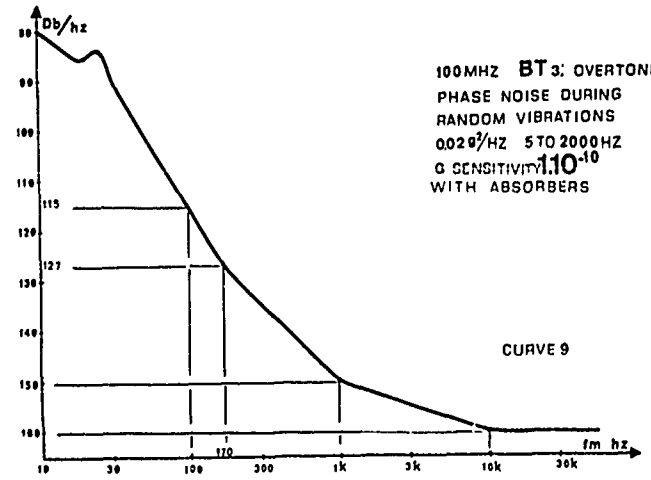
We next consider a 100 MHz oscillator with the phase noise figure as shown on curve 7.



A g sensitivity of $1.10^{-10}/g$ will give under vibrations (same spectrum) phase noise as shown on curve 8.

Using mechanical absorbers (curve 4), the phase noise under vibrations will be as shown on curve 9.





For the cut off frequency of the mounting structure (30 Hz), the phase noise curve depends still on the g sentivity of the resonator. After the cut off frequency, it depends on the combination of the mechanical filtering. Past 170 Hz, mechanical vibrations do not affect the phase noise curve.

Such mechanical filters reduce distorsions due to the g sensitivity of the resonators. The mechanical isolation becomes greater when the frequency vibration increases. The efficiency of such filtering is of course greater for random vibrations spectrum exceeding 2000 Hz.

For a lot of applications, the lower phase noise operational limit is more than a few hundred hertz (Avionic). Therefore, it would be more effective for industrial reasons to develop appropriate soft mounting structures, thereby reducing the required g sensitivity of the resonators ($1.10^{-9}/g$ instead of $1.10^{-10}/g$).

Inversely for lower Fournier frequencies, the phase noise depends directly on the g sensitivity of the resonator.

Table 10 gives verified acceleration sensitivity values (in parts per g) at different frequencies, with different piezoelectric materials and cuts.

g sensitivity (dF/F / g) — table 10

			5 MHz	10 MHz	20 MHz	50 MHz	100 MHz
QUARTZ	AT cut	P 1	$3\,10^{-9}$	$3\,10^{-9}$	$1\,10^{-9}$	$1\,10^{-8}$	$1\,10^{-8}$
		P 3			$5\,10^{-10}$	$5\,10^{-10}$	$3\,10^{-10}$
		BVA	$5\,10^{-11}$				
		P 5	$1\,10^{-9}$	$1\,10^{-9}$		$5\,10^{-10}$	$3\,10^{-10}$
	SC cut	P 3	$7\,10^{-10}$	$1\,10^{-9}$	$1\,10^{-10}$		
		BVA		$5\,10^{-11}$			
		QAS		$1\,10^{-10}$			
	BT cut	P 3		$1\,10^{-10}$	$1\,10^{-10}$	$1\,10^{-10}$	$1\,10^{-10}$
		P 5				$3\,10^{-10}$	$1\,10^{-10}$
$LiTaO_3$		P 1			$2\,10^{-9}$	$1\,10^{-8}$	$5\,10^{-9}$
		P 3 *					$5\,10^{-9}$

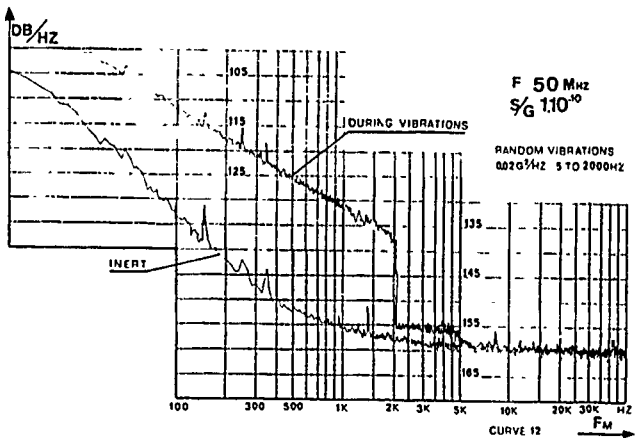
*doubly rotated cut

We can observe the degradation of performances for ion etched resonators, either Quartz or Lithium tantalate.

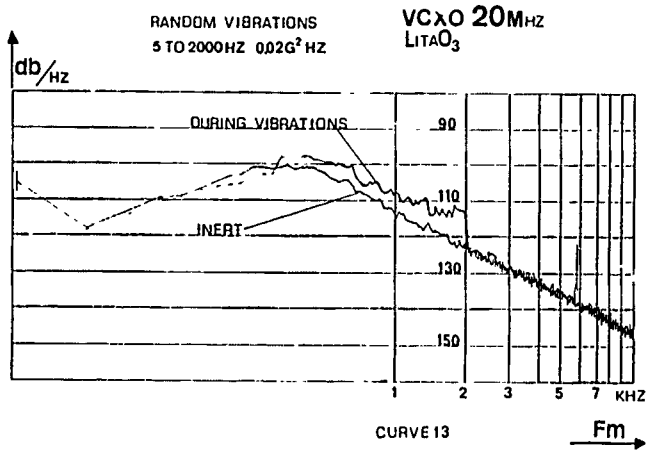
The g sensitivity measurements have been obtained by using measuring system as shown on Table 11 (see annex A3).

Figure 12 shows the behavior of a 50 MHz hand mounted OCXO under vibrations.

Figure 13 shows the phase noise figure of a 20 MHz lithium tantalate VCXO without mechanical absorbers. The influence of the acceleration sensitivity is easily observable. After 2000 Hz the phase noise curve will not be disturbed by perturbations due to acoustic sensitivity.



CURVE 12



CURVE 13

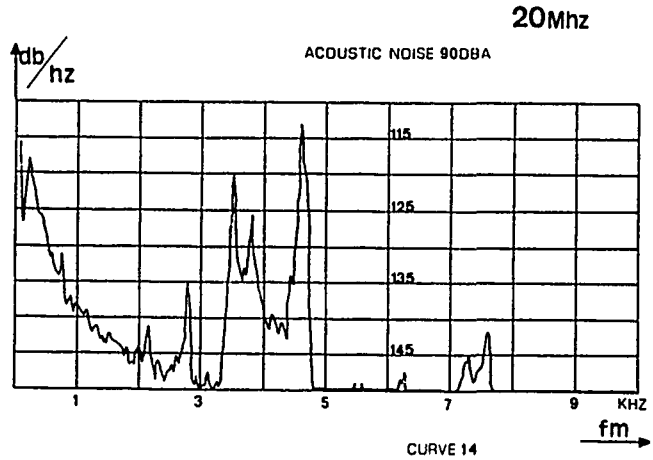
ACOUSTIC SENSITIVITY AND PHASE NOISE

As we know, it is possible to reduce the g sensitivity by appropriate design of the resonator. Likewise, it is possible to reduce its acoustic sensitivity. This reduction is important because an external isolation against acoustic perturbations is often difficult to obtain.

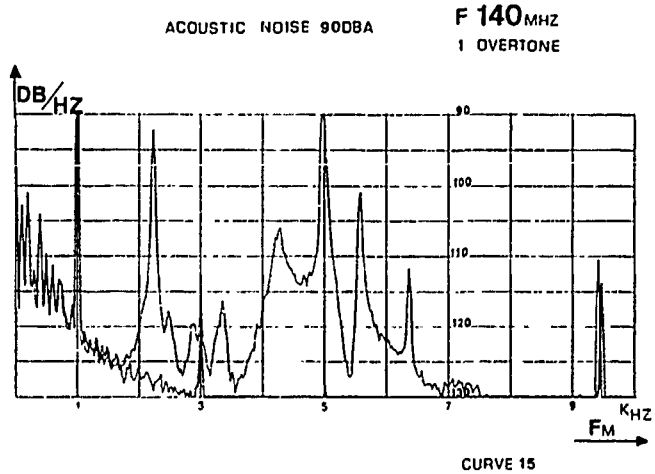
For a given frequency, the behavior of a resonator under severe acoustic vibrations depends on many parameters ; the most important are

- the piezoelectric material
- the cut
- the roughness
- the plate assembly characteristics :
  - . number of springs
  - . dimensions, thickness, stiffness

In addition, the acoustic sensitivity increases the frequency as we can see on figures 14 and 15.



CURVE 14



CURVE 15

In general, we can consider that every quartz resonator having a thickness lower than 50 m has a high acoustic sensitivity.

It appears that the use of ion beam milled or chemical etched resonators must be avoided when low acoustic sensitivity is required.

On the other hand, Lithium tantalate resonators have a real insensitivity to acoustic perturbations over a wide frequency range.

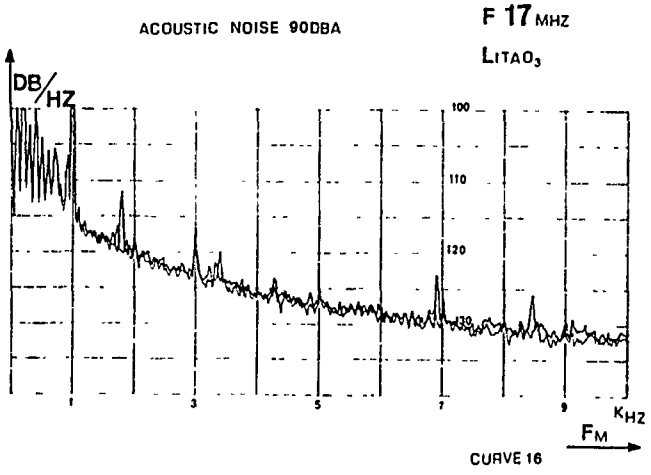
Curve 16 shows the behavior of a 17 MHz lithium tantalate resonator under acoustic noise. Curve 17 gives the phase noise characteristics of a 125 MHz ion beam milled lithium tantalate resonator. Therefore, for Lithium tantalate resonators, the reduction or neutralization of g sensitivity is the greatest technical difficulty.

The operating considerations generally define the frequency and piezoelectric cut of the resonator. The improvement of acoustic sensitivity requires the optimization of the mounting technology as well as adequate roughness and metallization.

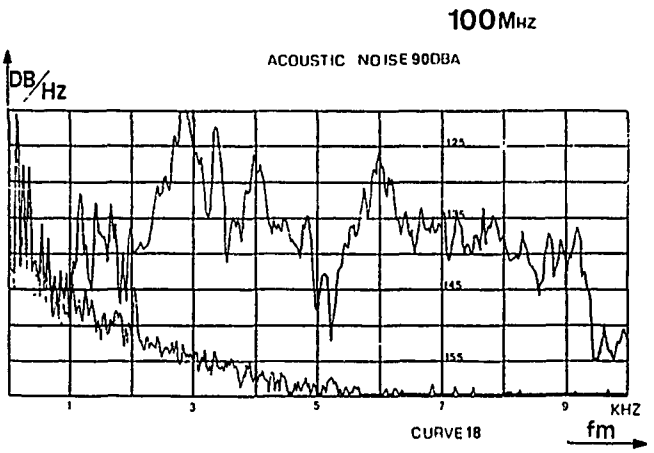
Unadequate roughness or deffects in metallization increases acoustic sensitivity. Curve 18 shows a 100 Mhz resonator. The electrodes of this resonator adhere poorly on the quartz plate.

Finally, the best way to reduce the acoustic sensitivity is to improve the mechanical fixation of the quartz plate on the holder. Unfortunatly, the g sensitivity also depends very much on the fixation technology.

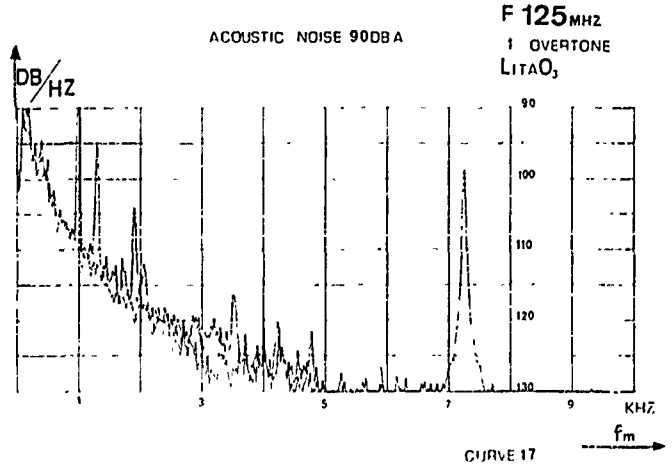
The improvement of both acoustic and g sensitivity requires different technological choices and therefore necessitates a compromise between these choices.



CURVE 16



CURVE 18



CURVE 17

All electrical measurements of acoustic sensitivity have been made following schema figure 11. This method is more qualitative than quantitative. The process must be quantified and possibly normalized.

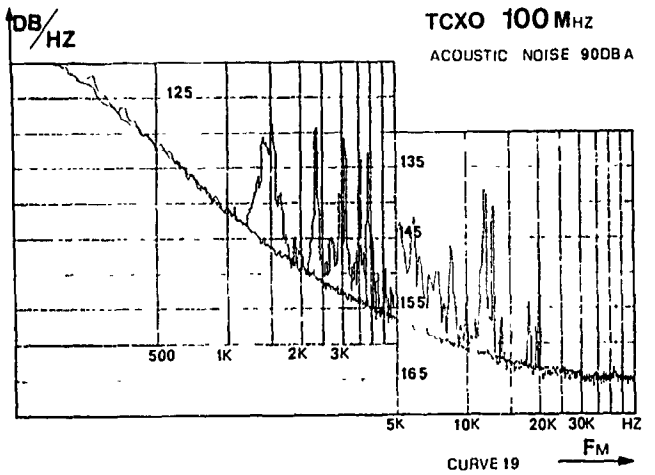
Another way for improving the acoustic sensitivity is to neutralize it.

In the same way that mechanical absorbers are used for g sensitivity, acoustic noise can be screened by using loaded elastomer foams. Such acoustic absorbers are fixed around the sensitive elements of the oscillator, especially the resonator.

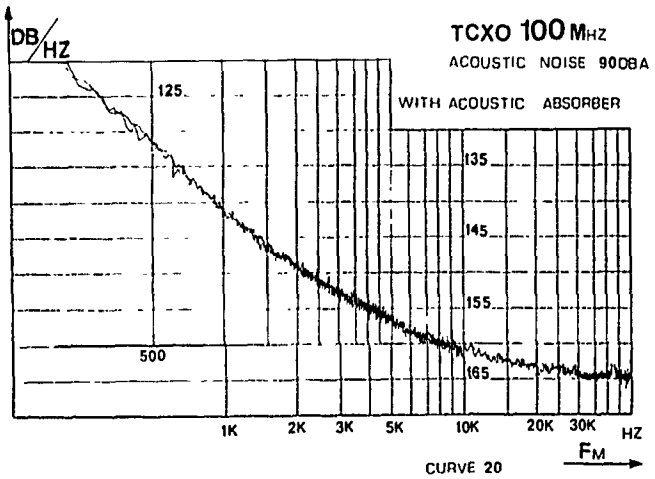
Curves 19 and 20 show phase noise measurements with 90 dBa acoustic noise in a frequency range 0-20KHz, with and without acoustic absorbers.

The same oscillator is measured during mechanical random vibrations 0,02 G^2/Hz 0-2000 Hz.

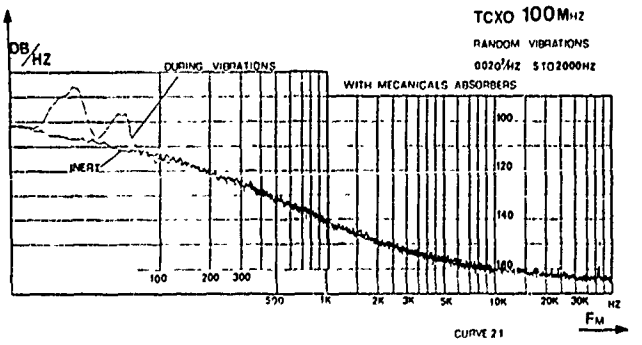
The increased phase noise below 100 Hz comes from the resonant frequencies of the two-stage mechanical absorbers system.



CURVE 19



CURVE 20



CURVE 21

CONCLUSION

After a lot of long experimentations, we can give some conclusions.

The improvement of g sensitivity is a long and expensive process.

For a given industrial state of the art, it is preferable to keep the phase noise characteristics during vibrations by using mechanical absorbers structures working as a filter.

In order to obtain the same results without absorbers, g sensitivity of some 1.10^{-12}/g would be necessary.

However, very close to the carrier frequency, the phase noise during vibrations depends directly on the g sensitivity of the resonator, either for quartz or Lithium Tantalate.

The next research and developpement efforts must be aimed at the improvement of acoustic sensitivity especially for quartz resonators. In the case of Lithium tantalate resonators acoustic sensitivity is not a problem area.

ANNEXES

A1 - ACCELERATION SENSITIVITY MEASUREMENT (See diagram 11)

Excitation devices
- sine generator
- power amplifier
- vibration exciter
- accelerometer- Load amplifier

Measurement device
- spectrum analyser (120 dBc dynamic)

The resonator is connected to an oscillator . It undergoes sine mechanical excitation of known frequency and level.

The calculation of the g sensitivity is done from measurement of the modulation lines on the oscillator phase noise spectrum.

Sine excitation

$$\mathcal{L}(f)_{dBc} = 20 \log \left[\frac{\Delta f}{f_o} \cdot g \cdot \frac{f_o}{2 f_m} \right]$$

$$\Delta = \frac{\Delta f}{f_o} = \frac{2 f_m}{f_o \cdot g} \cdot 10^{\frac{\mathcal{L}(f)}{20}}$$

with $\Delta = \frac{\Delta f}{f_o}$ g sensitivity of the resonator in parts per g.

g = sine vibration (number of g)
Fo = nominal oscillator frequency
Fm = sine vibration frequency

This formula can be used for modulation factors lower than 0.01. For modulation factors greater than 0.01, it is necessary to use Bessel functions.

Random excitation

The random excitation level is γ (g^2/Hz)

In a 1 Hz bandwidth $g = \sqrt{2\gamma}$

then $\mathcal{L}(f)dB_{/Hz} = 20 \log \left[\frac{\Delta f}{f_o} . \sqrt{2}\, g \frac{f_o}{2 f_m} \right]$

$$s = \frac{\Delta f}{f_o} = \frac{f_m}{\sqrt{2}\, f_o\, g} . 10^{\frac{\mathcal{L}(f)}{20}}$$

A2 - ACOUSTIC SENSITIVITY MEASUREMENTS (see diagram 11)

Excitation devices : - noise generator 20 Hz to 20 KHz bandwidth
- power amplifier
- anechoïc chamber
- microphone - amplifier

Measurement devices : - oscillator's phase noise measurement banch
- spectrum analyser

The vibration exciters can be used for maximum excitation frequencies of 2000 Hz to 10 000 Hz, depending on their size.

For constant excitation level, the amplitude of displacement decreases with the square of the frequency ( ). In this manner, the precision limits of the servo-control are reached.

For a 10 g excitation level, at a 10 KHz frequency, the peak to peak displacement is 0.05 m'

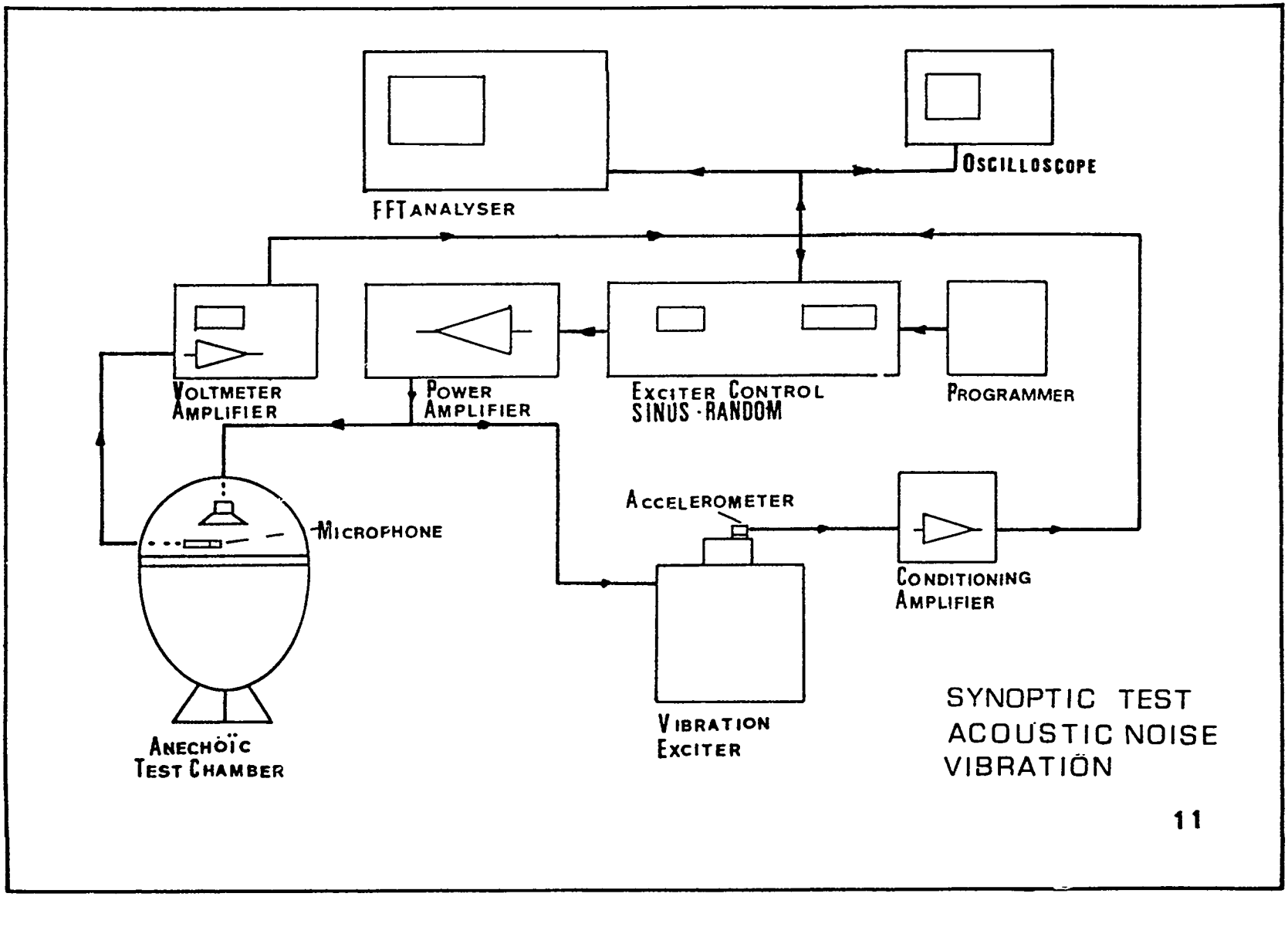
The acoustic excitation method allows us to obtain a regular white noise spectrum, for instance, 90 dBa $\pm$ 3 dB acoustic level in a 20 to 20 000 Hz bandwidth.

Acoustic noise is fed into a laboratory anechoïc chamber.

Such acoustic excitation spectrum allows to define the mechanical structure of resonators, their resonance frequencies and consequently, their influence on the phase noise oscillator's spectrum.

BIBLIOGRAPHY

- The effect of vibration on frequency standards and clocks - R.L. FILLER 35 th Frequency Control symposium
- Influence of environment conditions on a quartz resonator - M. VADOIS, J. BESSON, J.J. GAGNEPAIN Proceedings 28 th Annual Symposium on Frequency control
- Low " g " Sensitivity crystal Units and their testing - A. WARNER, B. GOLDFRANK, M.MEIRS and M. ROSENFELD 33rd Annual Symposium on Frequency control
- Design of a Bulk wave quartz resonator insensitive to acceleration - R. BESSON, JJ. GAGNEPAIN, D. JANIAUD and M. VALDOIS 33rd Annual Symposium on Frequency control
- Analyse des mécanismes de sensibilité accéléromètrique et barométrique des résonateurs à quartz de type QAS - J.P. AUBRY, J.C. CRAVEUR, F. DEYZAC, M. FRAGNEAU Revue technique THOMSON CSF - Vol. 19 n° 2, Juin 1987.


FFT ANALYSER
OSCILLOSCOPE
VOLTMETER AMPLIFIER
POWER AMPLIFIER
EXCITER CONTROL SINUS-RANDOM
PROGRAMMER
ACCELEROMETER
MICROPHONE
CONDITIONING AMPLIFIER
ANECHOÏC TEST CHAMBER
VIBRATION EXCITER
SYNOPTIC TEST
ACOUSTIC NOISE
VIBRATION
11

43rd Annual Symposium on Frequency Control - 1989

EXPERIMENTAL DETERMINATION OF THE RELATIONSHIP AMONG BAFFLE, TEMPERATURE DIFFERENCE AND POWER FOR THE HYDROTHERMAL GROWTH OF QUARTZ

Gary R. Johnson, Robert A. Irvine and Jonathan W. Foise

Sawyer Research Products, Inc.
35400 Lakeland Blvd.
Eastlake, Ohio 44094

Abstract

We have experimentally determined the relationship among baffle design, the temperature difference between the growing and dissolving chambers, and the power requirements. Increasing the opening of the baffle requires a higher power level to maintain a given temperature difference. The presence of lascas or grown crystals, and the type of baffle used have strong effects upon the temperature difference-power curve. Growing chamber temperature and top insulation have only minor effects.

Introduction

Two factors which greatly control the growth of crystals from solution are the transport of dissolved material to the region of the crystal-solution interface and the removal of the heat of crystallization from this region. While these processes can take place due to diffusion, they are accelerated by the free convection of solution about the growing crystals.[1] This free convection of solution (sometimes called natural convection) is produced and controlled by the establishment of a temperature difference across the volume of the solution. In an autoclave in particular, this can be modeled by the heat transfer system known as a closed thermosyphon.[2]

To aid in the establishment of a temperature difference, a flow restricting baffle is inserted, creating nearly isothermal growing and dissolving chambers within the autoclave. The design of this baffle largely determines the input power level necessary to establish a given temperature difference. Although the general relationship among baffle design, temperature difference and input power has long been understood, the exact details of this relationship have not been determined.

The purpose of these experiments is to numerically determine this relationship. This has two goals. The more immediate is to better be able to design run plans for quartz growth. Later, it is expected that such information will be part of a program to predict growth conditions required to obtain particular grades of quartz.

Experimental

Each experiment involved equipping an autoclave with a particular baffle. Once at operating conditions, various temperature differences are set and measurements are taken of the main and top heater power required to maintain this temperature difference at equilibrium.

Table 1 lists the variables examined and their ranges. Table 2 lists variables that were blocked for these experiments and the values selected.

Table 1

Variable	Range
Baffle	
Number of Holes	Single-hole vs. perforated (multiple holes).
Percentage Opening	Single-hole 10 to 30%, perforated 7.5 to 15%.
Temperature Difference	0 to 16°C
Top Insulation	Full, partial and no convection cooling allowed (bands open, one closed, both closed).
Growing Chamber Temp.	340 to 350°C, most data at 345°C.
Supply	350 lbs. lascas vs. none (one run).
Crystals	None vs. grown bars (one run).
Autoclave	H44 vs. H45.

Table 2

Variable	Value
Solution	0.8 M Na_2CO_3, no Li, unfiltered.
Pressure Target	12300 psi (fill calculated for maximum T_{diss} 360°C).
Bottom Heater	Originally 180 V, changed to 220 V (see below).

In general, the temperature difference (TD) was varied from low to high during an experiment, with two power readings taken at each of two or three intermediate values. However, for a number of runs, the temperature difference was reduced to a low value late in the run and the power was remeasured to check for hysteresis or other changes during the run. No hysteresis was found during any run.

Initially, all runs started with the blanket on the can around the top of autoclave (Figure 1). At higher values of TD, one would reach a point of no power to the top heaters and so first the blanket was removed, the top band was opened ad finally the bottom band was opened. As the bands are opened, the amount of convectional cooling around the top of the autoclave is increased. No other changes in top insulation were studied.

CH2690-6/89/0000-447 $1.00 C 1989 IEEE

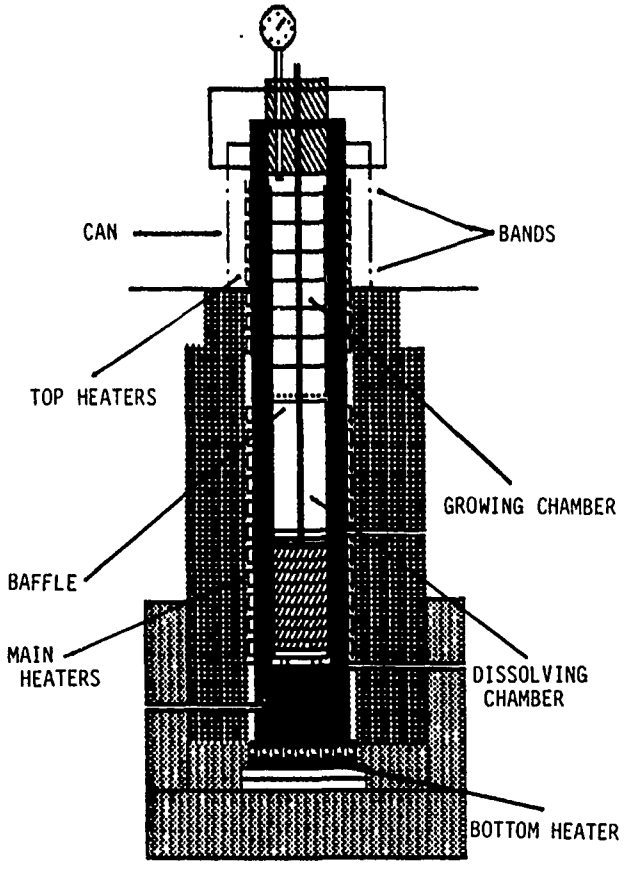


Figure 1 - Schematic diagram showing general design of the autoclaves used for these experiments.

Once a stable value of a TD was obtained, power measurements were made. Since power control is through zero-crossover SCRs, conventional methods of power measurement are not possible. Three different indirect techniques were used. The first involved measuring the frequency of the alternating current through one of the three legs from the SCRs to the heaters (in all cases the right leg was used) using a Fluke model 1910A Multicounter with a resolution of 0.1 Hz and the attenuator on. This gives a measure of power between 0 (at zero power) and 60 MHz (at full power). The second method also used the Fluke meter to get a frequency measurement, but used it in the totalizing mode to obtain a count of waveforms over a set period of time (2 minutes). The third method was to read the power of the control signal (using an ammeter) from the temperature controllers to the SCRs. This gives a measure of power from approximately 4 mamps at zero power to approximately 20 mamps at full power. For the top heaters, only the first and third methods were used.

Figure 2 compares the three measuring techniques. It was found that all three are linerally proportional to each other. In general, power reading shall be reported using the frequency method in this report.

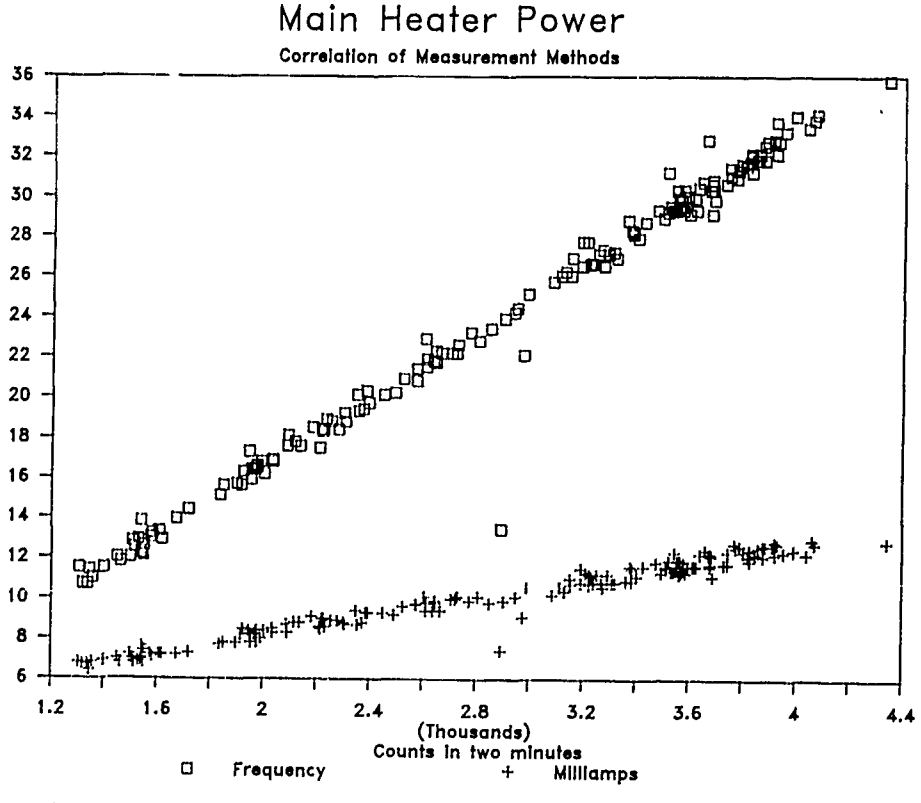


Figure 2 A - Power from frequency measurements and from miliamp readings versus power from counts for the main heaters using all data points from both autoclaves.

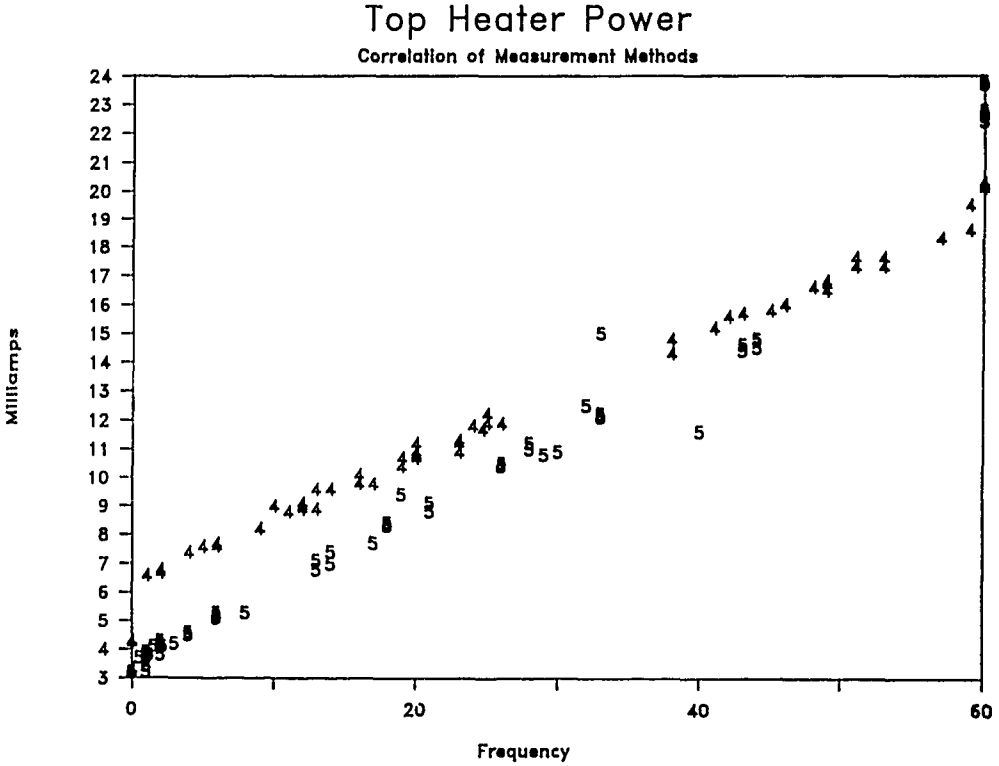


Figure 2 B - Power readings from milliamp readings versus readings by frequency measurements for the top heaters. Data for H44 ("4") and H45 ("5") is shown separately.

Power measurement calibrations were done to see if the power levels at full power were equal through each of the three legs of the autoclaves and between the two autoclaves. Measurements showed that on both autoclaves the three legs were virtually identical and that the power level readings between the two autoclaves differed by about 1%.

Most of the experiments were done at a constant growing chamber temperature of 345°C. However, during a number of runs, data was also collected at either 340°C or 350°C to determine the effect of growing chamber temperature upon power. In all cases temperatures were held to within ± 0.5° of target.

Two other measurements were also made during these experiments. Each run was profiled (temperature vs. height) at a TD of 5° (the 15% concentric baffle run was done at a TD of 2°) with a T_{grow} of 345°. For a number of runs, the thermocouples were moved to within a few inches above and below the baffle and the temperature was measured versus time with a strip chart recorder (Marlin model 4153) to try to determine the fluid velocity through the different baffles.

One problem encountered during these experiments occurred after the first pair of runs. Both autoclaves contained a large quantity of "heavy phase" (a polymeric silicate phase) at their bottoms, once they were cooled down and opened, that made cleanup and turnaround difficult. This was solved by raising the bottom temperature and by placing approximately 10 crystals into the rack so that the dissolved quartz would have something to grow onto, and not be left in a supersaturated solution at the end of a run.

Results

Figure 3 shows the results of power measurements for a typical run. The plot shows main heater power as a function of TD. The data is all for a growing chamber temperature of 345°C except where noted otherwise. The various letters indicate if the bands were closed, the top band open, or both open. All data at 350°C was taken with the bands open. One can see that it was necessary to open these bands to

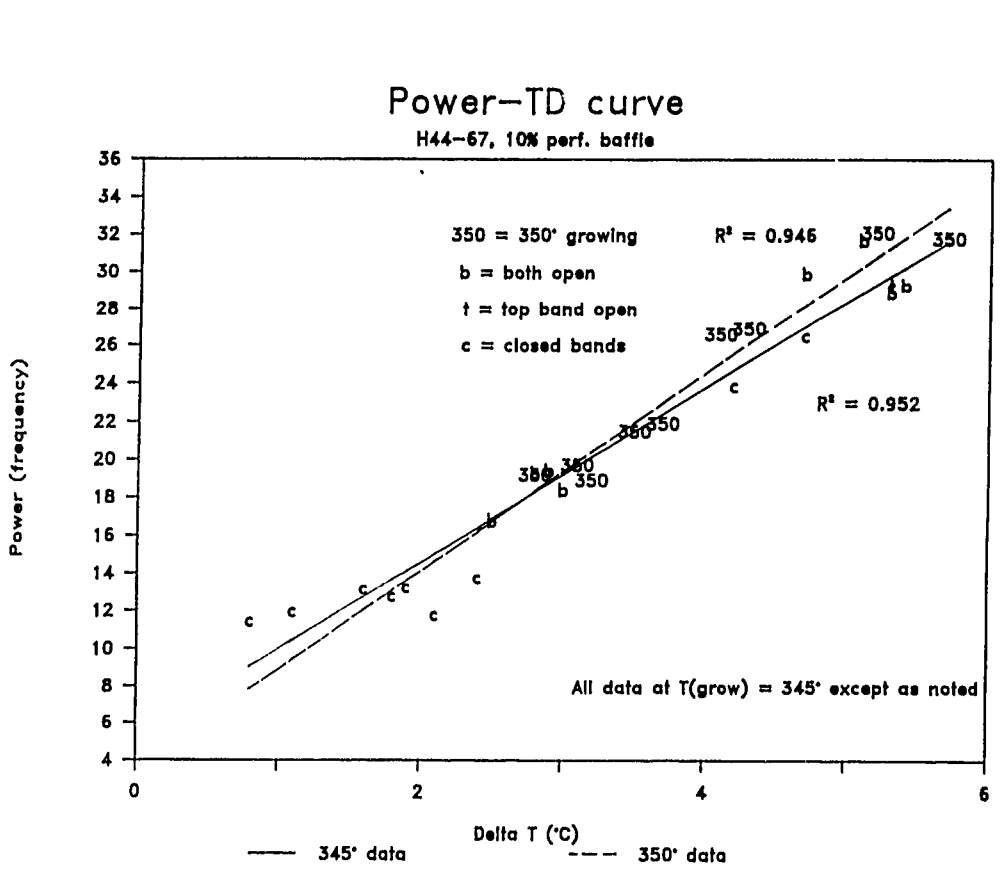


Figure 3 - Main heater power in frequency versus the temperature difference between the growing and dissolving chamber (TD) for the 10% perforated baffle. Data is shown for a growing chamber temperature of 345°. and 350°C.

achieve a higher TD. In all cases, the limits of TD were determined by this insulation and the top heat available. For example, the lowest TD for each baffle was achieved with the blanket on, the bands closed and the top heat at maximum power. The opposite case existed for the maximum TD.

The upper line shows the linear regression fit for the data at 345°C. For this example, the square of the regression coefficient was 0.946, which corresponds to a confidence limit of greater than 99.9%. This confidence limit also held for all of the 345°C data.

Figure 4 shows the data for the 17% single-hole baffle experiments. This pair of runs was conducted to examine this variation between the two autoclaves. As can be seen, there is good agreement between the two runs. At the highest TD, there is only about a 10% difference in power levels between the two.

The power-TD curves for various single-hole baffles are shown in Figure 5. The relationship among TD, percent opening of the baffle and power is as expected. The minimum and maximum TD increases with decreasing percent open area. Additionally, for a given TD, more power is required as the percent opening of the baffle increases. In other words, as the amount of opening in the baffle increases, there is greater fluid exchange between the two chambers and thus greater energy exchange between the two, so a greater amount of power is required to maintain a particular TD.

The same relationship is also seen in Figure 6 for the various perforated baffles. The change in the slope of the power-TD curve with the percent baffle opening is shown in Figure 7. There appears to be a general increase in slope (this slope can be thought of as the rate of increase of power with increasing TD) with increasing percent opening, though the most linear relationship is seen for the perforated baffles.

We have long postulated that a perforated baffle of a given percent opening behaves like a single-hole baffle of approximately twice that opening. In other words, a 15% perforated baffle behaves like a 30% single-hole baffle. Figure 8 shows the fitted power-TD curves for a number of single-hole and perforated baffles. It does appear that there is a grouping of the x% perforated baffles with the 2x% single-hole baffles, though it is difficult to determine the exact relationship from the limited data.

Top heater power was also measured during all of these experiments. As expected, it showed a general trend to decrease with increasing TD. It also showed an increase with decreasing top insulation for a given TD. However, it did not show the well behaved linear behavior of the main heaters.

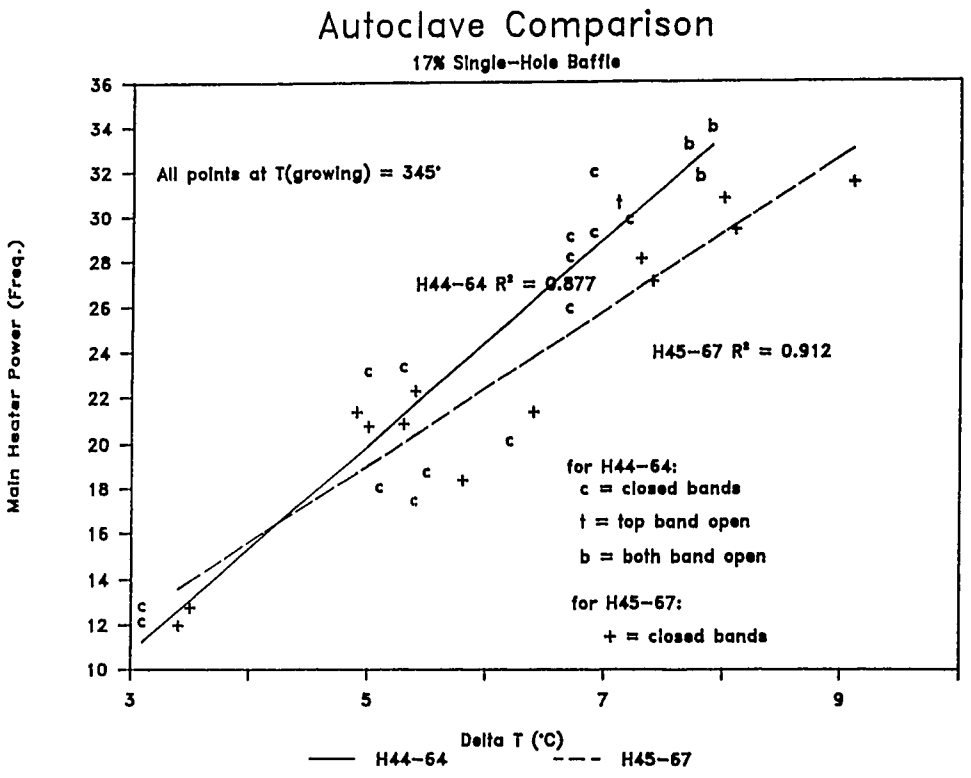


Figure 4 - Main heater power versus TD for the two 17% single-hole baffle experiments. The two autoclaves, H44-64 and H45-67 are compared.

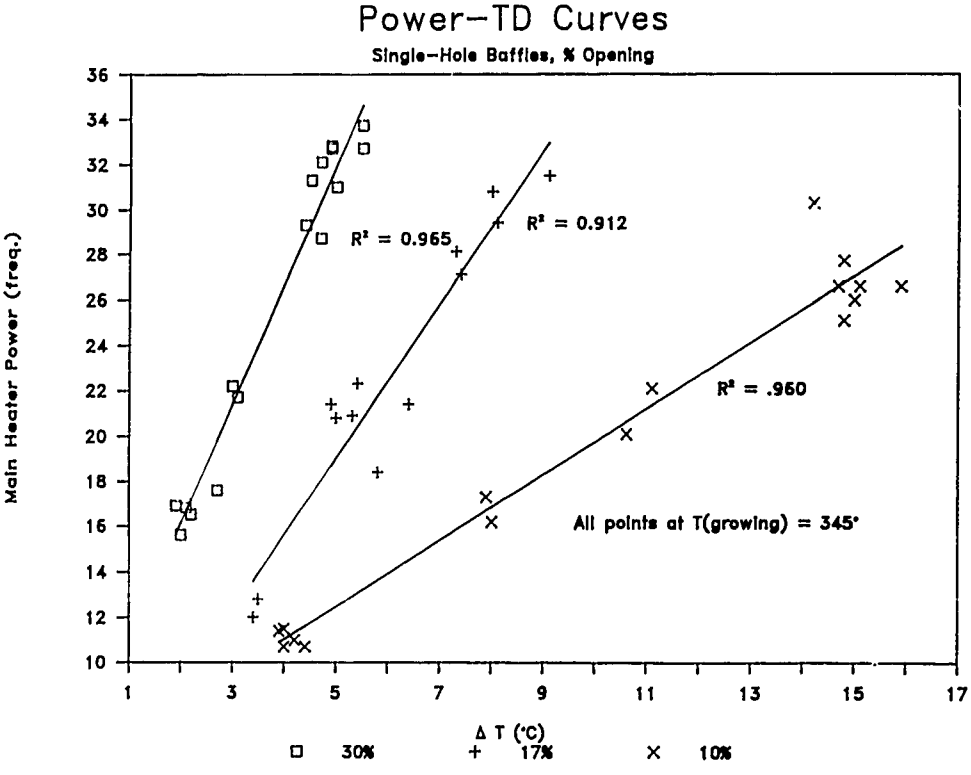


Figure 5 - Main heater power versus TD for the three single-hole baffle experiments. Only the data at $T_{growing}$ = 345° and the linear regressions of that data are shown.

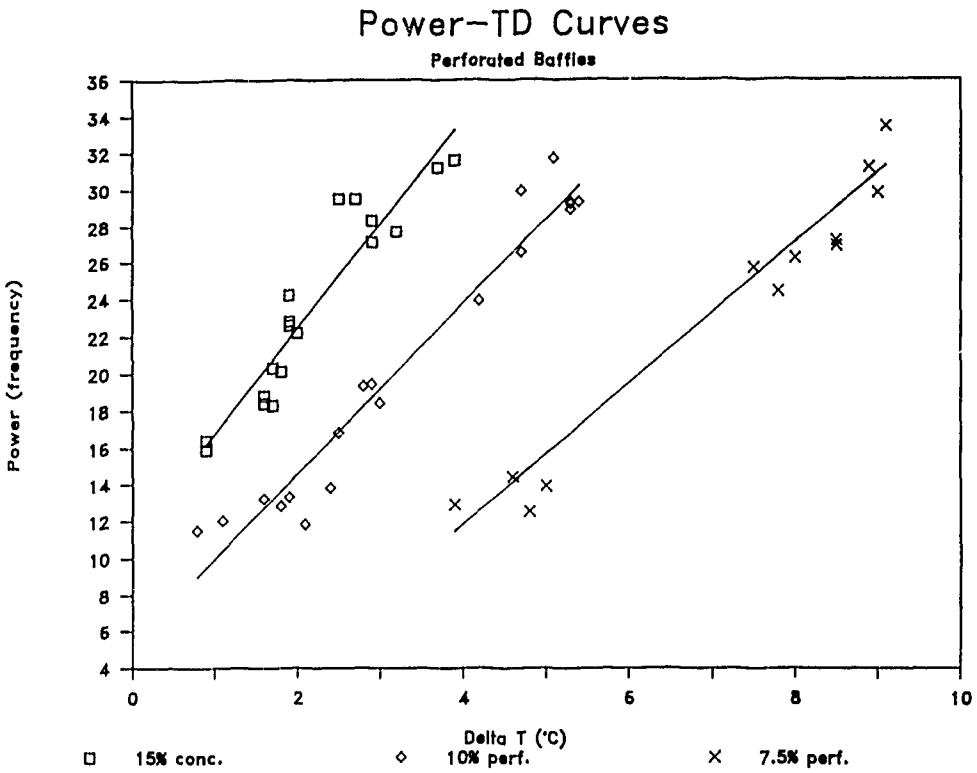


Figure 6 - Main heater power versus TD for the three perforated baffle experiments. Only the data at $T_{growing}$ = 345° and the linear regressions of that data are shown.

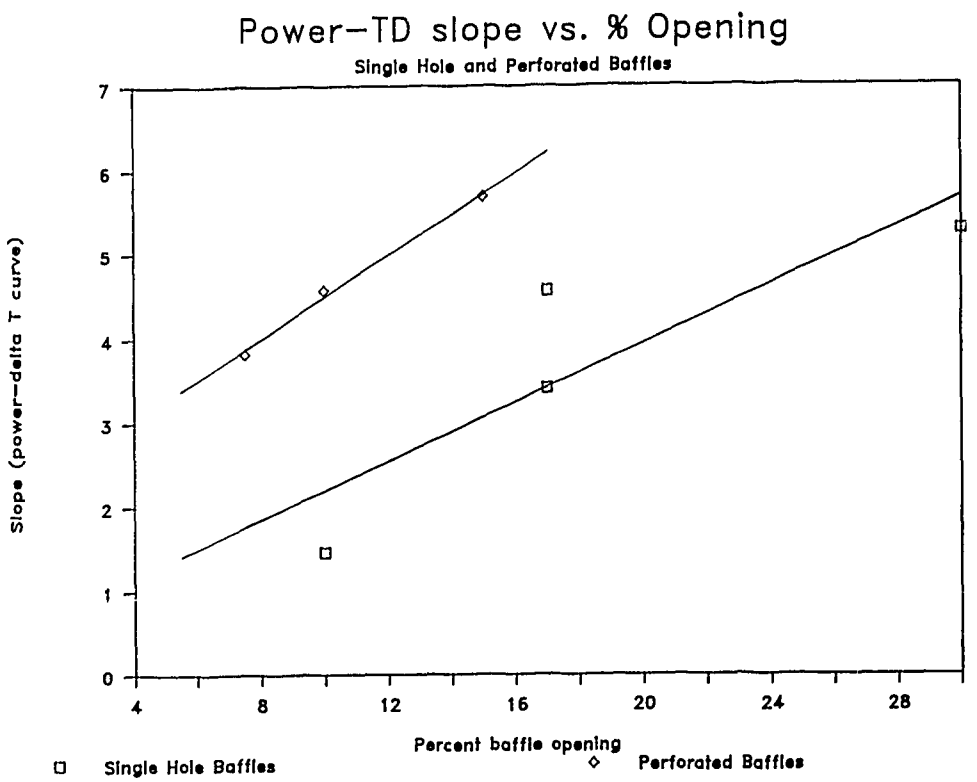


Figure 7 - Slope of the power-TD curves versus percent baffle opening for single-hole and perforated baffles (for $T_{growing}$ = 345°).

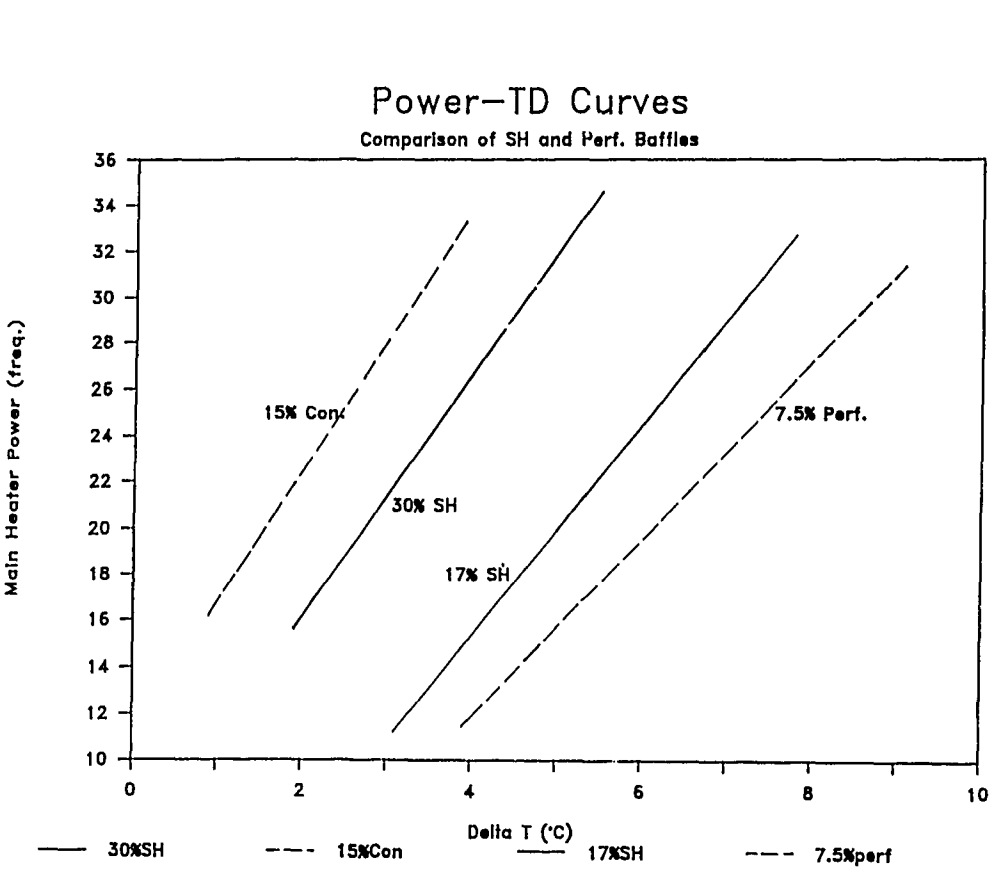


Figure 8 - Main heater power versus TD showing the comparison of the 2x% single-hole baffles with the x% perforated baffles. Lines shown are the linear regression fits for the data at $T_{growing} = 345°$.

During a number of runs, attempts were made to understand the dependence of main heater power on growing chamber temperature. An examination of the various runs shows no consistent relationship between power and growing chamber temperature. In fact, some runs showed an increase in power with increasing $T_{growing}$ (for a given TD); other runs showed the opposite. All runs did show that within the range examined the effect was small. Using the modeling programs which are part of ECHIP (Experiment in a Chip[3]), we have been able to fit the single-hole baffle data to a quadratic model. The results are shown in Figure 9. This model indicates that for a given TD and baffle, generally a higher growing chamber temperature would require a slightly higher power level, though the opposite may be true for very small TDs (this last result may be due to problems with the model based upon lack of data).

Experiments were conducted to compare conditions at the beginning of a run (350 lbs. of lascas and no grown quartz), conditions at the end of a run (approximately 30 lbs. of lascas and 300 lbs. of grown quartz), and conditions in an autoclave without quartz. Figure 10 shows the power-TD curves for these three conditions. From the beginning of a run ("no crystals") to the end of a run ("no lascas"), a significant change in both the slope of the curve and in the maximum TD is seen. The run without quartz ("no quartz") shows an almost identical slope in the end-of-a-run curve. This indicates that the amount of lascas present has a significant effect upon the power-TD curve, but that the amount of grown quartz has only minor effects. An implication of this is that the circulation, at least in the dissolving chamber, is largely uneffected by the amount of grown material. This is confirmed by the temperature profile data (see below).

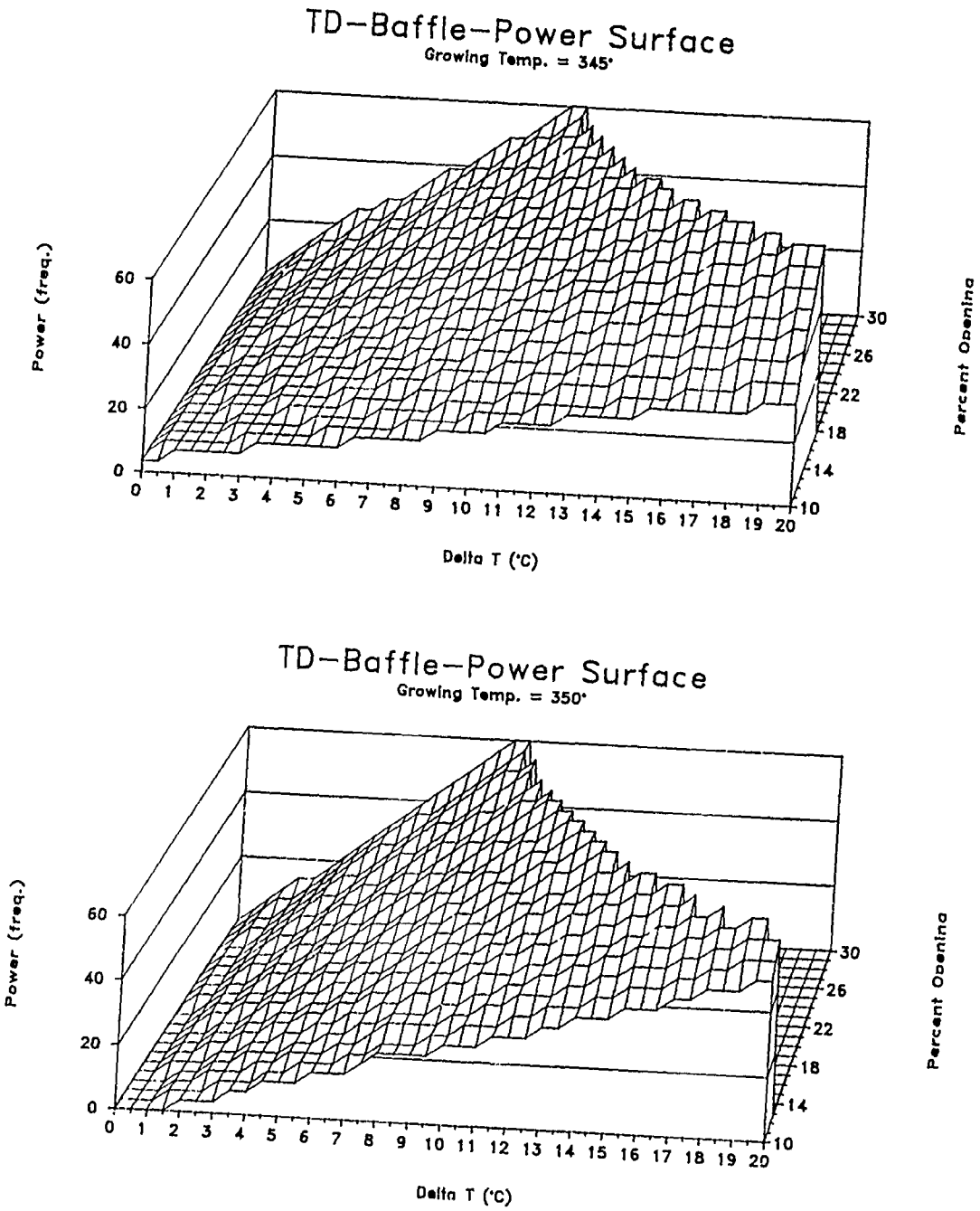


Figure 9 - Main heater power (in frequency) as a function of percent baffle opening and TD. The graphs do not show the actual data, but depict the surface generated by the ECHIP model. The top graph shows the results for a growing chamber temperature of 345°, the bottom for 350°.

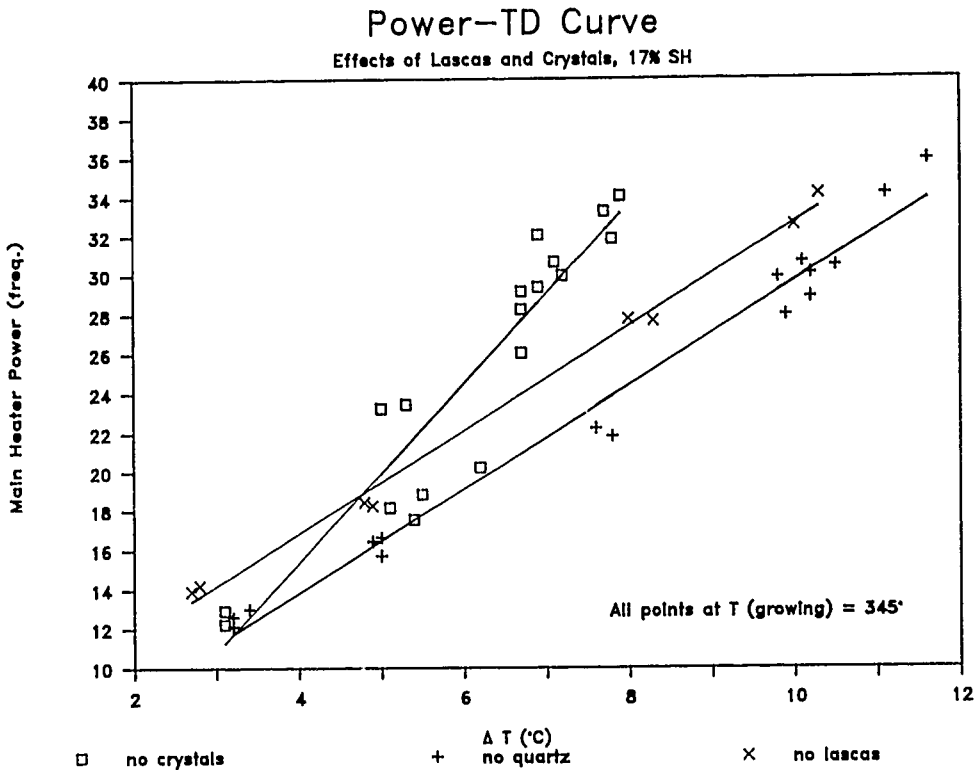


Figure 10 - Main heater power versus TD for the 17% single-hole baffle experiments with lascas (no crystals), without lascas or grown crystals (no quartz), and without lascas (no lascas).

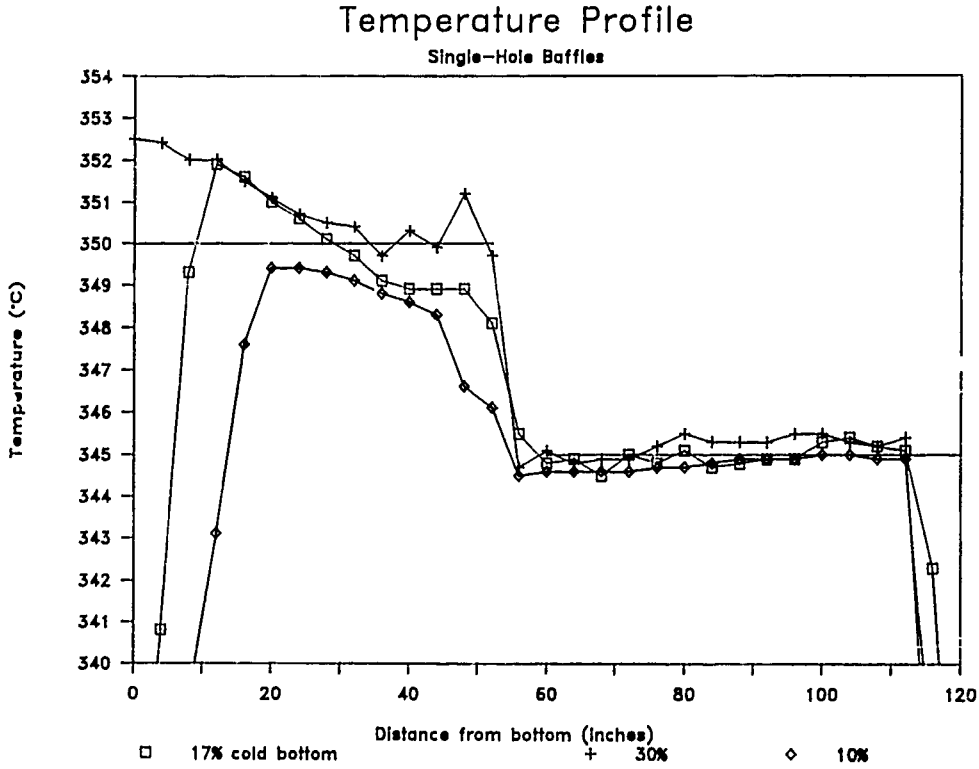


Figure 11 - Temperature (°C) versus distance from the bottom of the autoclave for the single-hole baffles. The target temperatures of 350° and 345° are shown.

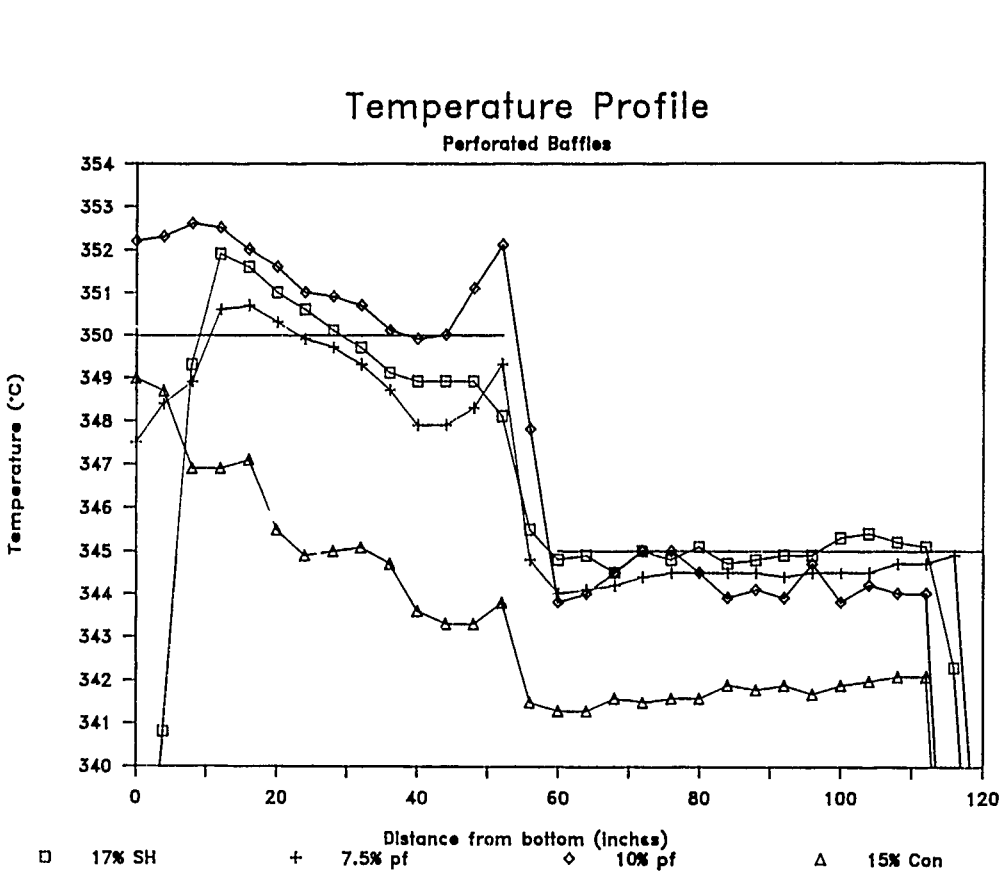


Figure 12 - Temperature (°C) versus distance from the bottom of the autoclave for the perforated baffles. The 17% single-hole baffle is shown for comparison. Also shown are the target temperatures of 350°C and 345°C. The profile for the 15% concentric baffle was done at a lower temperature because the higher temperatures could not be achieved with that baffle.

As discussed in the experimental section, temperature profiles were conducted for each run. Figure 11 shows the profiles for the single-hole baffle experiments. The 17% baffle run was conducted with the bottom voltage at 180 V, the 30% and 10% were done at 220 V. Despite the higher voltage, the 10% run shows a colder bottom than the 17% run. This same trend, with the tighter baffles showing a relatively colder bottom (relative to the temperature at the midpoint of the dissolving chamber), is shown in Figure 12 for the perforated baffles. It also appears that the perforated baffles show a peak in temperature just below the baffle. This peak cannot be attributed to noise. It may be due to an upward flow of fluid along the centerline just below the baffle.

Figure 13 shows the temperature profiles for the three runs containing different quantities of lascas and grown quartz, as described above. The run without quartz and the run without grown crystals both show a temperature gradation through the dissolving chamber, while the run with grown crystals shows a relatively uniform dissolving temperature. These differences confirm the results from the power-TD curves.

For a number of runs, temperature measurements versus time were conducted with the temperature monitored just above and below the baffle. For most of these experiments, the data was inconclusive. However, for the 15% concentric baffle, one was able to monitor the percolation of fluid through the baffle. It appears that the hot fluid generally comes up the middle of the baffle (though it appears that there may be occasional changes in the direction of flow) at a rate of about 5 cm per second or greater. Figure 14 shows a section of the trace.

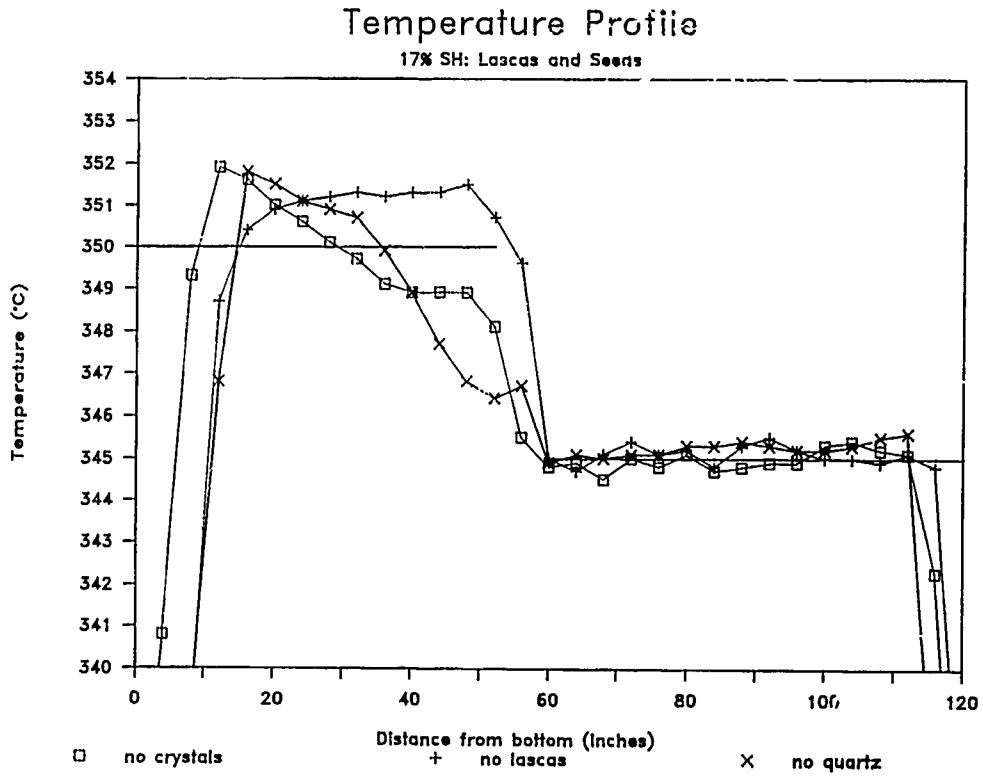


Figure 13 - Temperature (°C) versus distance from the bottom of the autoclave for the 17% single-hole baffle experiments with lascas (no crystals), without lascas or grown crystals (no quartz), and without lascas (no lascas).

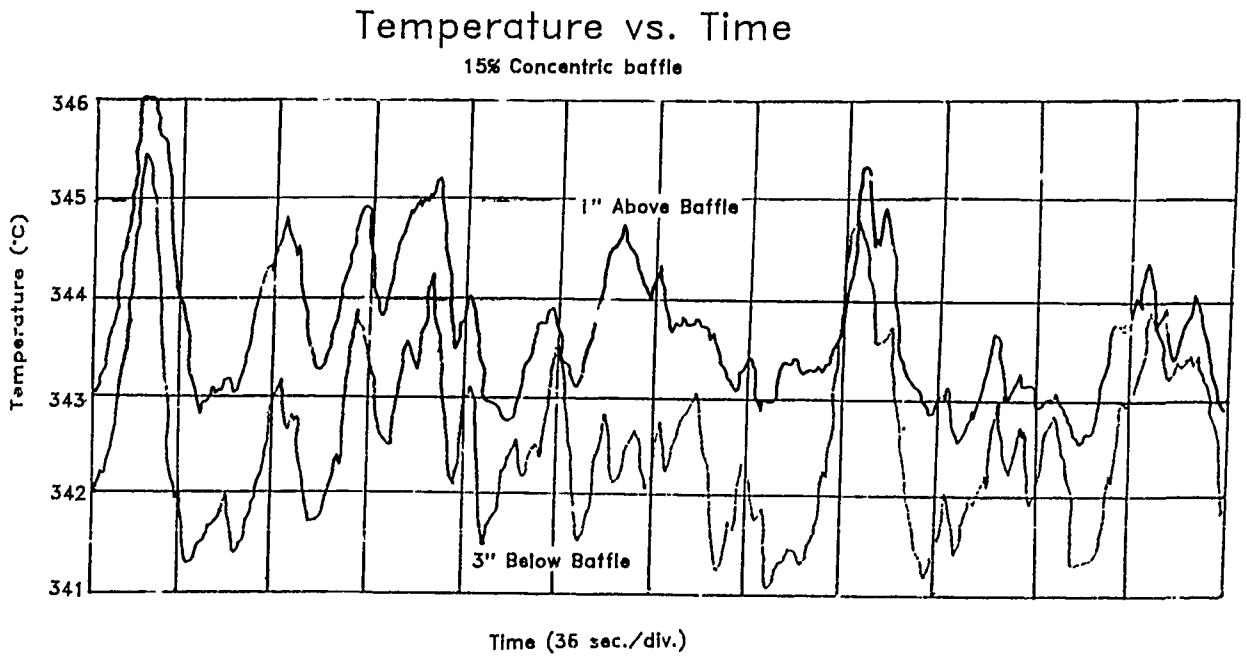


Figure 14 - Temperature versus time for the 15% concentric baffle. The darker line represents the trace for the thermocouple 3" below the baffle. the lighter line is 1" above the baffle.

Conclusions and Further Research

We have been able to form the following conclusions:

1. The power to the heaters could be measured by a variety of equivalent techniques.

2. The behavior of different autoclaves was the same.

3. Main heater power increases linerally with increasing temperature difference.

4. The limits of the temperature difference were determined by top insulation and the top heaters.

5. The rate of change of power with increasing temperature difference is dependent upon the percent of opening of the baffle and whether the baffle is single-hole or perforated.

6. The behavior of perforated baffles is equivalent to that of single-hole baffles of approximately twice the percentage opening.

This data will now be used in experiments to establish the exact effect of these relationships upon the quality of the grown quartz.

References

1. Petrov, T.G., Treivus, E.B., Kasatkin, A.P., Growing Crystals from Solution. Consultants Bureau, N.Y., c. 1969, translated by A. Tybulewicz.

2. Annamalai, N., Baksi, S., Saha, P., Indian J. of Technology. Vol. 21, p. 425, 1983.

3. ECHIP (Expert in a Chip), experimental deisgn and statistical software by ECHIP, Inc., Hockessin, DE.

GROWTH DEFECTS IN CULTURED QUARTZ; A MICROSCOPIC INVESTIGATION

P. CORDIER and J. C. DOUKHAN
Laboratoire de Structure et Propriétés de l' Etat Solide
Université de Lille-Flandres-Artois, 59655 Villeneuve d'Ascq-Cedex FRANCE
X. BUISSON and O. BIGNON
Société Industrielle des Combustibles Nucléaires 73104 Annecy-Cedex FRANCE

ABSTRACT

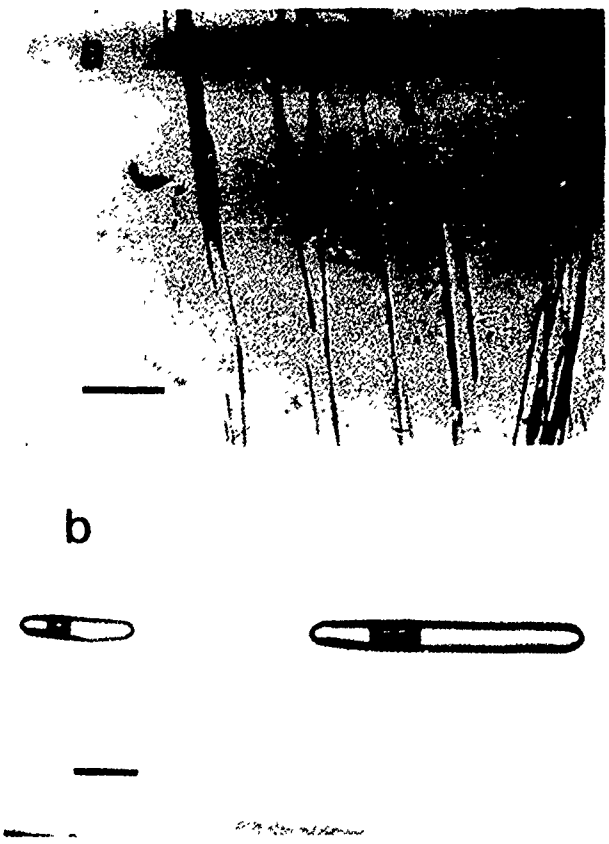
Although cultured quartz has now reached a high degree of purity and crystalline perfection, there remains, at least in some crystals, various types of defects which can limit the performances of piezoelectric devices. The most common are dislocations and chemical impurities; among these later ones "water" (or hydrogen or OH or hydroxyls) is generally assumed to be the most cumbersome defect. Two distinct modes of water incorporation have been recognized, molecular water (tiny fluid inclusions with diameters down to 30 Å) and isolated point defects like the so-called $4H_{Si}$ defects. The equilibrium concentration of this later type of defects is certainly extremely low at the usual T and P conditions of growth, however for relatively large growth rates, a supersaturation can occur. We have tried to determine the equilibrium concentration as a function of T and P. We have also tentatively estimated the diffusivity coefficient of the corresponding water species as a function of temperature.

I INTRODUCTION

Since 20 years numerous studies of defects in cultured quartz have been published (for recent reviews, [1] to [3]). The best characterized defects are probably dislocations which are currently visualized by X-Ray Lang topography [4,5]. In high quality crystals grown nowadays on natural seeds, the dislocation density can reach values as low as 10^1 cm^{-2} [6 to 9]. Lang topography on cross sections also reveals that the dislocation density is not homogeneous and various growth sectors have been defined [10 to 13], some corresponding to a low dislocation density could be the best material for new seeds.

Fluid inclusions have been observed by optical microscopy [14 to 17]; they are generally located in the immediate vicinity of the seed and are currently interpreted as resulting from preferential etching of the as-grown dislocations of the seed at the very beginning of the growth process [18]; this does not imply that far from the seed the crystal is wet. Smaller fluid inclusions have also been detected in the bulk by transmission electron microscopy (TEM) when the growth rate becomes large [19] or when the nutrient is impure [20]. It is now widely accepted that the amount of water incorporated in quartz during its growth strongly depends on its growth rate and small fluctuations of this parameter (the so-called breathing of crystal growth) leads to striations along the c axis which can be detected by light scattering tomography after heat treatment [21]. Fluid inclusions have also been found in some natural quartzes and have been interpreted as resulting from perturbed growth [22].

Solid inclusions have been detected by optical microscopy and by TEM for the smaller ones (with a few hundreds Angströms diameter); they are assumed to be particles of acmite $NaFe^{3+}Si_2O_6$ resulting from the slow corrosion of non lined autoclaves [17, 23, 24]; in fact other compositions are also found [25]. Fig. 1. illustrates these various defects. It is however to be remembered that in high quality crystals grown actually, such defects are very rare.



CH2690-6/89/0000-459 $1.00 C 1989 IEEE

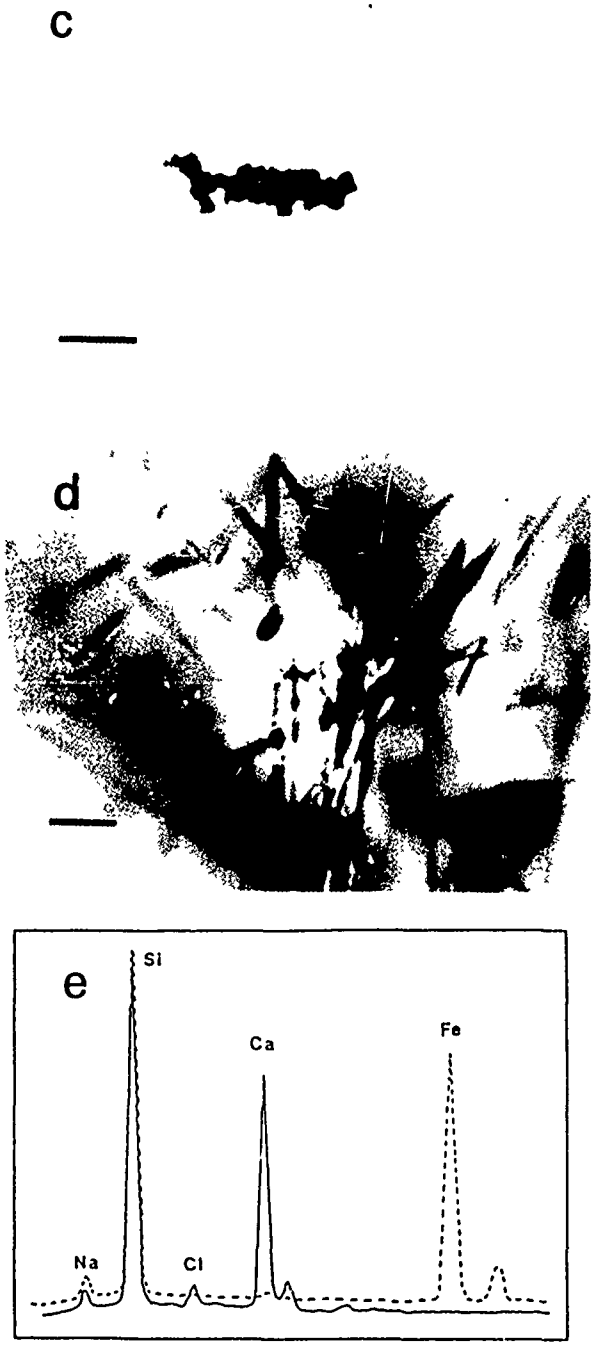


Figure 1. Various types of macroscopic defects in cultured quartz. a) Dislocations; X-Ray Lang topography (scale bar = 1 mm); the dislocations are approximately perpendicular to the seed and their density is ≈ 5 x 10^2 cm^{-2}. b) Fluid inclusions in the vicinity of the seed; optical micrograph (scale bar = 100 µm) c) solid inclusion, optical micrograph; (scale bar = 100 µm). d) Tiny solid inclusions observed by TEM; they have no epitaxial relationships with the matrix (scale bar = 1 µm). e) qualitative X-Ray energy-dispersive spectra of two different inclusions.

Substitutional Aluminum (Al_{Si}) and water or hydroxyls (OH) are the most frequent chemical impurities in cultured quartz [26 to 29]. The occurrence of water or OH's is known to affect dramatically the piezoelectric performances of resonators and it has been shown that the quality factor Q is strongly correlated with the content of water [30]; as the stretching vibrations of OH groups absorb infrared light in the vicinity of 3000 cm^{-1}, infrared spectroscopy in the range 4000-2500 cm^{-1} (IRS) is currently used for the characterization of the water content [31 to 33]. As shown below water can enter in quartz under various forms: OH compensated impurities, molecular water or/and isolated water point defects. We report in this article our own investigations on these modes of water incorporation. We have also studied the equilibrium solubility of water point defects and their diffusivity as a function of temperature and pressure.

II WATER RELATED DEFECTS

The water content of cultured quartz is currently estimated by the ratio of its infrared absorbance at two wave numbers: 3500 cm^{-1} (at which water is active) and 3800 cm^{-1} (at which it is not) [31 to 33]. Mineralogists who are also interested by the presence of water in quartz (because it dramatically affects the rheological properties of this mineral) use to consider the whole IR spectrum which contains more information that a unique ratio; they estimate the water content by the area under the absorbance spectrum [review in 34]. The IR spectrum of a relatively wet quartz (Fig. 2) presents a broad absorption band from approximately 3600 to 2500 cm^{-1} with a number of sharp absorption peaks superimposed to it; some of these peaks are clearly attributed to OH compensated impurities, for instance the peak at 3580 cm^{-1} (at room temperature) has been attributed sometimes to substitutional aluminum defects (Al_{Si}). As alkali ions (Na^+ for instance) can play the same role of charge compensator, the concentration of such OH's is at most equal to the concentration of the trivalent impurities (Al + Fe^{3+}). One usually considers at least three types of water related defects in quartz. The above peaks correspond to the first type The two other ones (molecular water under the form of tiny fluid inclusions and truly dissolved water point defects) have both an IR signature contained in the broad band; they can be distinguished only by near infrared spectroscopy (NIRS) in the range 7000-4000 cm^{-1} [35,36]. NIR spectra exhibit several absorption bands (Fig. 3). The one centered on 4500 cm^{-1} results only from the vibration modes of isolated OH point defects (i.e. of Si-OH groups) [37] while the band centered on 5200 cm^{-1} is results from a combination of the bending and stretching vibration modes of the H_2O molecule [38]; both types of defects contribute to the other bands which are thus useless.

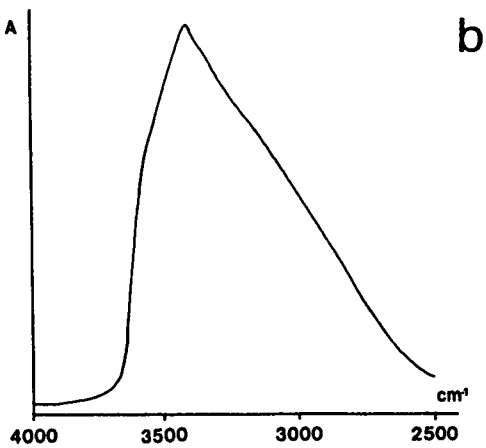


Figure 2: a) Conventional IR spectrum of a wet quartz at room temperature and b) broad absorption band obtained by substracting the spectrum of a dry region to the one of a wetter one in the same material (the vertical scale is in arbitrary units).

As the NIR absorbance is markedly lower than the IR one, appreciably thicker samples must be used as well as much longer recording times.

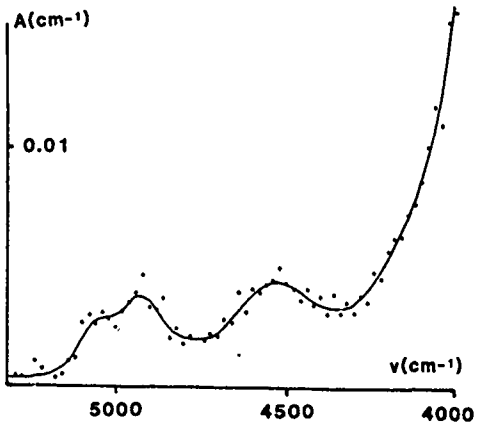


Figure 3: NIR spectrum of a wet quartz (Bell Tel crystal far from the seed) containing at a time water point defects and molecular water. The total water content estimated by conventional IRS is 150 at. ppm approximately.

The nature of the water point defects still is controversed. Nutall and Weil [39, 40] have characterized by Electron Spin Resonance (ESR) the substitutional defect $(4H)_{Si}$ which is formally equivalent to the substitution of a SiO_2 group by two H_2O. Other possible point defects cited in the litterature are interstitial water molecules in the c channels [41] and hydrolysed Si-O-Si bonds i.e. water molecules having chemically reacted with quartz leading to Si-OH — Si-OH (this is the first model of water incorporation proposed by Griggs in 1967 [42]).

III EXPERIMENTAL

A number of crystals have been analysed by NIRS. Some have been grown by GEC[1] in nutrients deliberately doped with a view to get doped quartzes; as shown below the prominent impurity in these materials is water. Another crystal which has been extensively used for various types of experiments in the laboratory has been grown twenty years ago by Bell Tel[2]. All the others have been grown by SICN in the usual growth conditions. The one labelled L was grown in a small autoclave (10 litres) while the others labelled 229 and 140 where grown in the large autoclaves used for production.

All the conventional IR spectra have been recorded with a 882 Perkin-Elmer spectrometer with unpolarized light, at room temperature, in air. The double beam of the spectrometer eliminates the perturbations which could be

[1] GEC Wembley, Great-Britain; samples kindly provided by Dr V. Cheel. Oxford

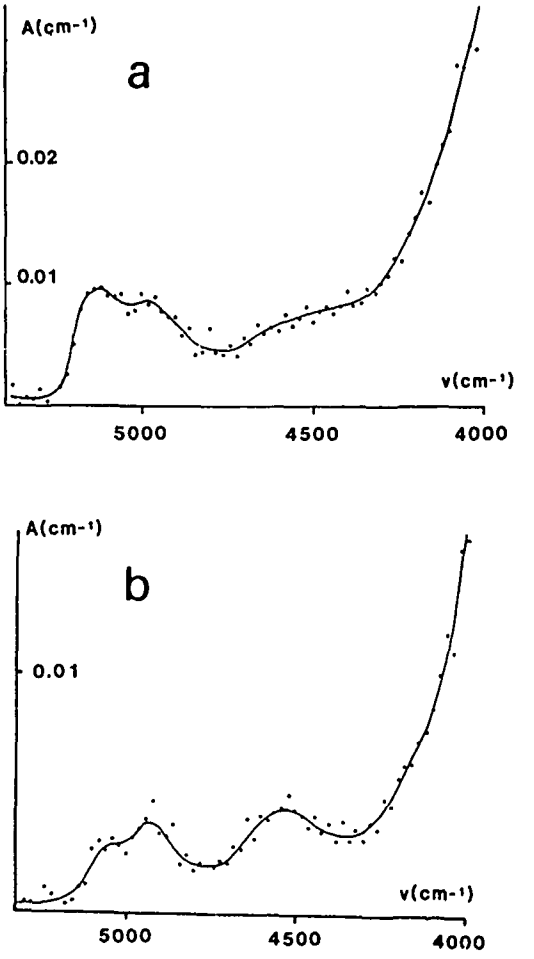
[2] Sample kindly provided by Prof Chaklader, UBC Vancouver, Canada.

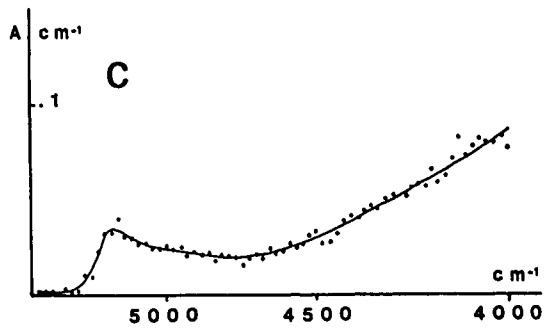
caused by the water vapor in air. The chosen resolution was 2.4 cm^{-1}. The NIR spectra (wave number range 5500-4000 cm^{-1}) have been recorded in air, at room temperature, with a spectrometer built in the laboratory. This spectrometer has only one beam; a blank spectrum (without sample) must thus be recorded for calibration before experiments. Two detectors measure, for each wave number, the incident and the transmitted intensities i.e. before and after the sample. Most of the experiments have been done with unpolarized light but for some experiments requiring polarized light (see section V) a Cambridge IGP 228 polarizer has been used. The resolution of the spectrometer varies with the wave length from 15 cm^{-1} for the largest wave number (5500 cm^{-1}) to 35 cm^{-1} for the lowest one (4000 cm^{-1}). Recording a NIR spectrum with the above conditions requires approximately 2 hours. Furthermore, for most of the samples we have also recorded another spectrum with a better resolution (18 to 25 cm^{-1}) in the range 4760 and 4255 cm^{-1} which corresponds to the absorption band of the point defects; under these circumstances the recording time reaches 4 hours. Because of the very low absorbance in the NIR spectral region, 200 measurements were accumulated for each wave number with a view to increase the signal/noise ratio; further noise reduction was obtained in eliminating the high frequencies of the Fourier transforms of the spectra. Comparison of the OH content of the various samples requires precise spectra in the range 4760-4255 cm^{-1}; the base line due to the feet of other peaks has thus been substracted; finally the absorption band of the point defects has been tentatively decomposed in a series of sharp peaks with gaussian profiles. This operation has been performed with the help of a "top-hat" numerical filter [43].

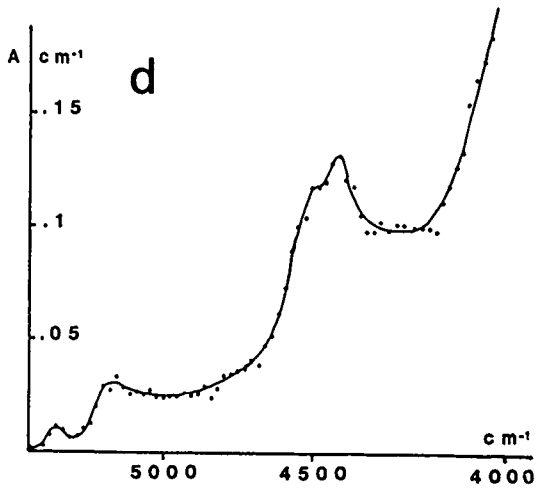
We have tried to determine the solubility of water point defects as a function of pressure and temperature by performing a series of annealings of wet quartzes under various T and P conditions, and then by checking, by TEM, wether water has precipitated or not. The assembly used for annealings under pressure is described in [44], it derives from the so-called Griggs deformation apparatus. The solid medium used for the confining pressure is a cylinder of talc and the pressure is transmitted owing to a rigid piston of sintered alumina. The furnace is constituted by a small cylinder of graphite surrounding the sample. As the solubilitity of water increases with pressure, the procedure consists in increasing first the pressure at room temperature, then the temperature is increased up to the required value and the pressure is readjusted; both parameters are then kept unchanged during the imposed annealing time; in most cases IR spectra have been recorded before and after annealings. These experiments have also allowed the measurement of the diffusivity of water point defects in measuring on the TEM micrographs the mean distance between the small water precipitates. Assuming a random walk diffusion for the point defects, this mean distance is $X = (2D\ t)^{1/2}$ where t is the annealing time and D the diffusion coefficient of the water point defects.

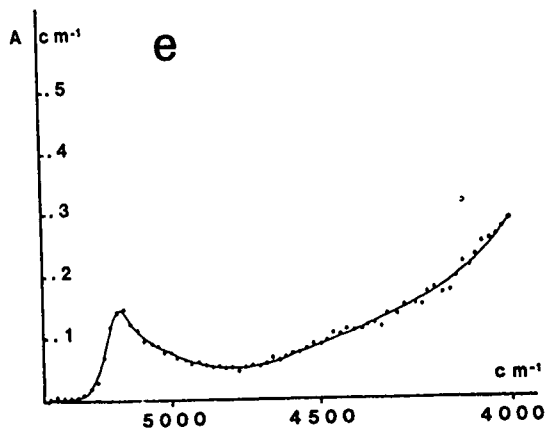
IV NIR SPECTROSCOPY AND WATER SPECIATION

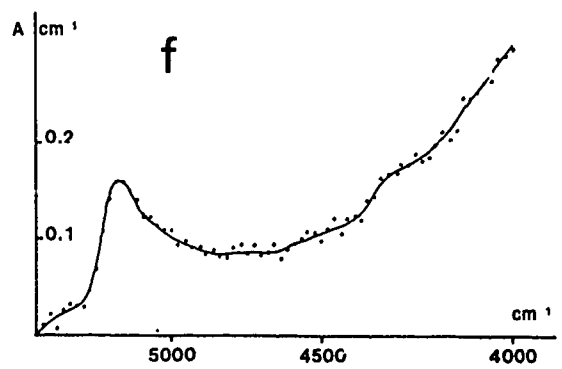
As mentioned above, only NIRS allows the distinction between water point defects (truly dissolved water) and molecular water incorporated during growth under the form of tiny fluid inclusions. Fig. 4 shows the NIR spectra of the as-grown crystals. The Bell Tel crystal is a wet but otherwise pure material; its total water content (determined by conventional IRS) varies from H/Si = 1000 at. ppm in the vicinity of the seed down to 150 ppm far from it at the end of growth. The corresponding NIR spectra indicate that near the seed only molecular water occurs (this is confirmed by the observation of numerous tiny bubbles on TEM micrographs); in contrast far from the seed both types of OH's occur, the amount of molecular water is lower (but tiny bubbles still are detected on TEM micrographs) while the absorption band of point defects becomes clearly visible.

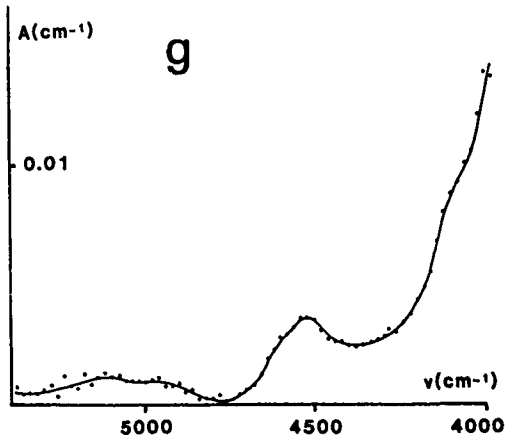












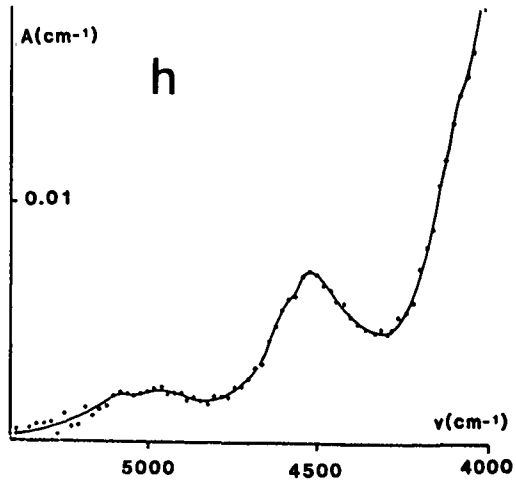


Figure 4: NIR spectra of the as-grown crystals. a) Bell Tel crystal; basal slab in the vicinity of the seed (H/Si = 600 ppm); thickness = 2.5 cm. b) same crystal far from the seed (H/Si = 150 ppm); same thickness. c) d) and e) crystals grown in nutrients doped with P, Al and Ti respectively. f) SICN crystal labelled L synthesized with a high growth rate. g) and h) SICN crystals labelled 140 and 229; they contain only point defects. i) TEM micrograph of a wet zone probably due to a brief fluctuation of the growth rate in an otherwise very dry crystal (scale bar = 1 µm).

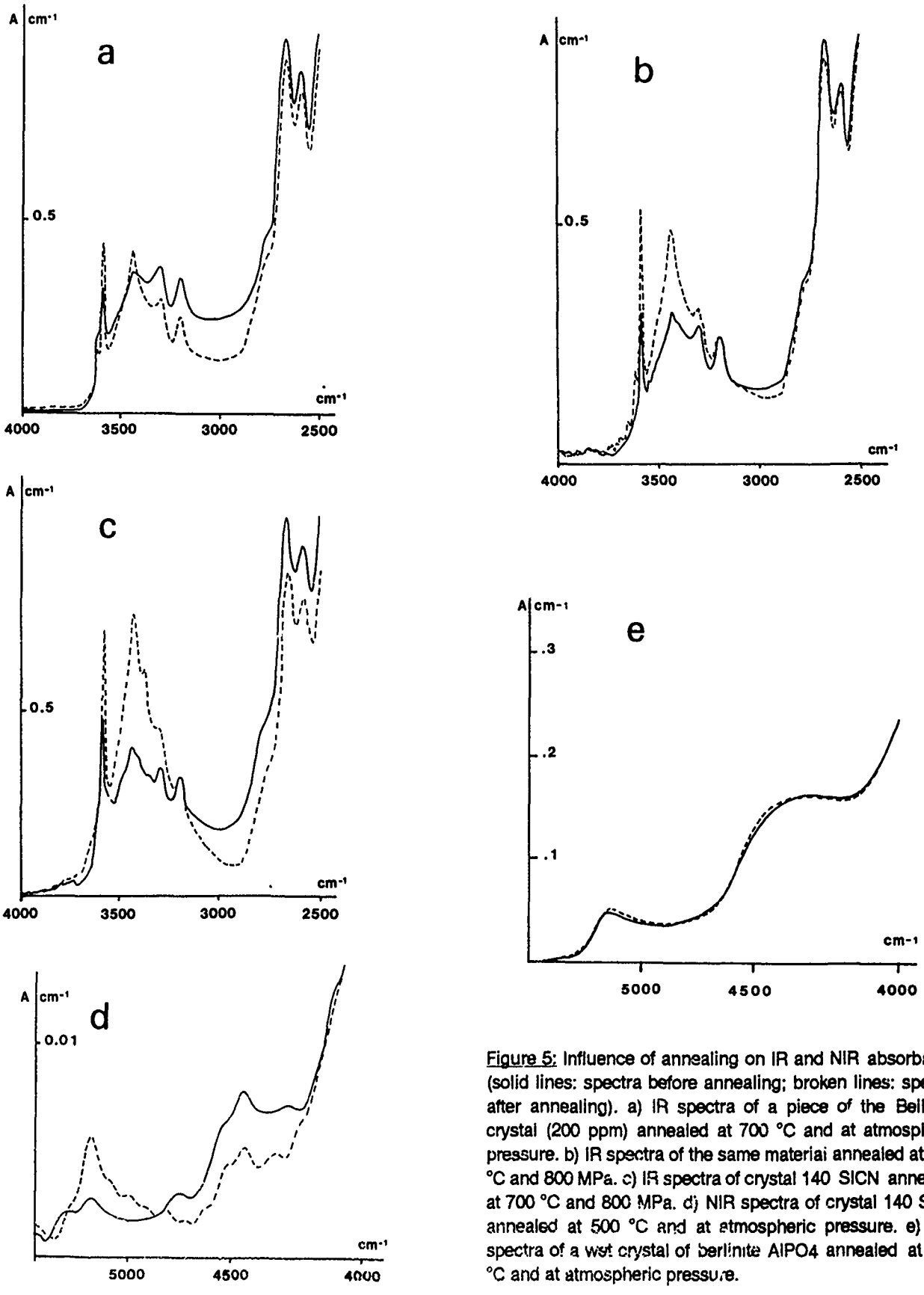
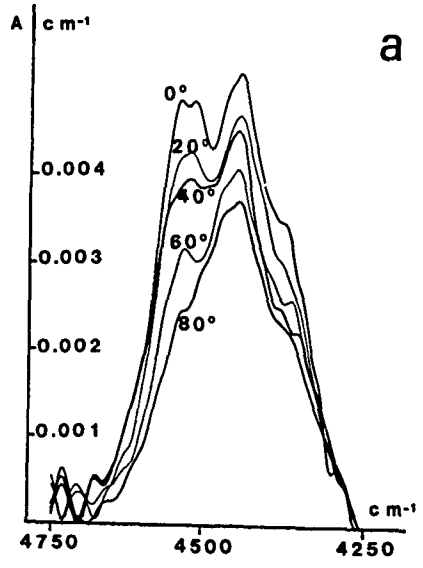
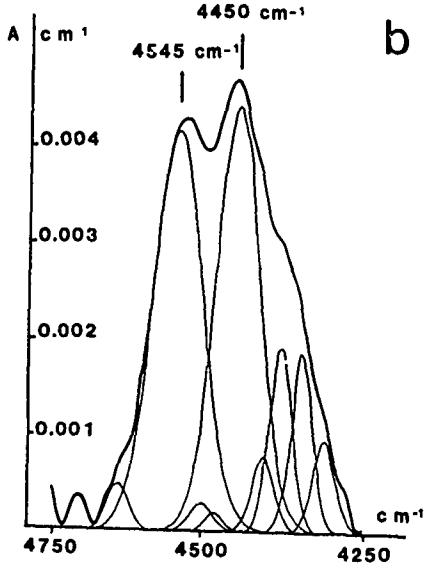
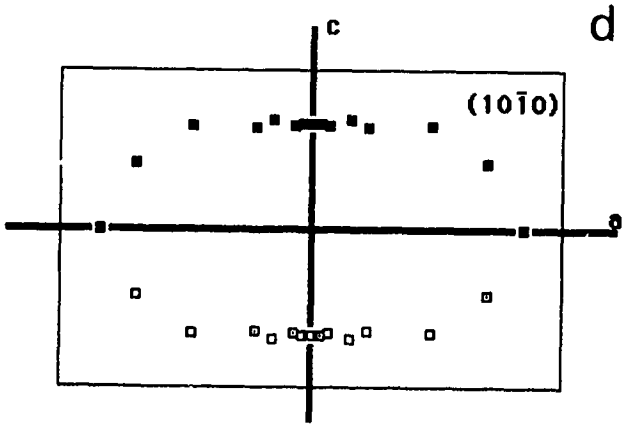


Figure 5: Influence of annealing on IR and NIR absorbance (solid lines: spectra before annealing; broken lines: spectra after annealing). a) IR spectra of a piece of the Bell Tel crystal (200 ppm) annealed at 700 °C and at atmospheric pressure. b) IR spectra of the same material annealed at 700 °C and 800 MPa. c) IR spectra of crystal 140 SICN annealed at 700 °C and 800 MPa. d) NIR spectra of crystal 140 SICN annealed at 500 °C and at atmospheric pressure. e) NIR spectra of a wet crystal of berlinite $AlPO_4$ annealed at 315 °C and at atmospheric pressure.







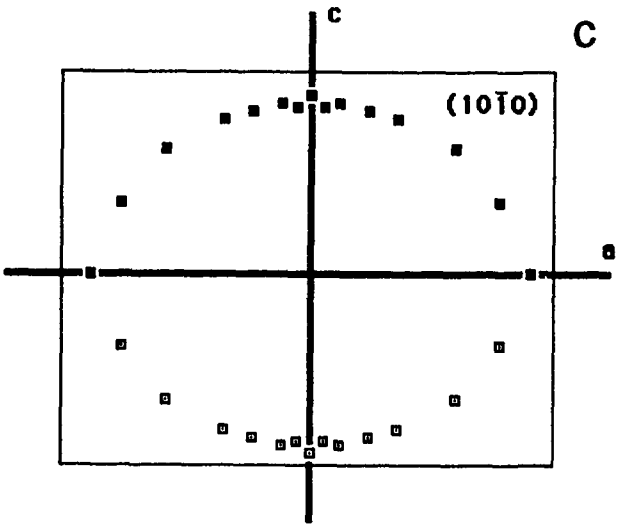


Figure 6: polarized NIR spectra of the 140 SICN crystal, beam // Y. Variation of the absorbance with the polarization orientation. a) deconvolution of the absorption band into gaussian peaks. b) variation with the polarization direction of the absorption band. c) evolution of the peak at 4450 cm^{-1}. d) evolution of the peak at 4545 cm^{-1}.

The crystals grown in nutrients deliberately doped are wet (their IR spectra indicate 500, 2400 and 1400 ppm respectively for P, Al and Ti doped nutrients). Their NIR spectra show only molecular water except for the Al doped crystal which also contains point defects; TEM investigations confirm the presence of numerous tiny bubbles in all these crystals. the SICN crystal labelled L which has been synthesized with a high growth rate contains only molecular water homogeneous concentration of water of the order of 1000 ppm). Finally the other SICN crystals labelled 140 and 229 contain only water point defects with total amounts of 180 and 100 ppm respectively. The corresponding quality factors as determined by conventional absorption at 3900 and 3500 cm^{-1} are of the order of 1.5×10^6. These results show that crystals synthesized with a relatively high growth rate are wet and in most cases water enters under the form of molecular water, however there seems to be a restricted range of growth rate (not markedly larger than the one used for industrial production) which rather leads to supersaturated concentrations of point defects. Accidental fluctuations of this important parameter generates zones of wet material as shown on Fig. 4h.

The influence of annealing at atmospheric pressure on the IR spectra is shown on Fig. 5. For the Bell Tel crystal (sample cut in a region containing 200 ppm i.e. in a region with at a time point defects and molecular water), one observes an increase of some absorption peaks (at 3580 and 3430 cm^{-1} for instance) while the absorbance decreases in the region

comprised between 3400 and 2900 cm^{-1}. A similar evolution is observed for the wet quartz 140 which contains initially only water point defects. The situation remains essentialy the same for annealings at moderate pressure (< 500 MPa). In NIRS one clearly observes after annealing at atmospheric pressure a decrease of the absorption band related to point defects while the band related to molecular water increases. It is interesting to compare this evolution to the one of wet berlinite $AlPO_4$ (Fig. 5e). Berlinite is a structural analog of quartz synthesized by hydrothermal growth and it presents very similar water related defects (see for instance [45]); one also observes by NIRS in this material two absorption bands related to molecular water and water point defects respectively but the band corresponding to water point defects does not desappear after annealing although TEM investigations clearly show that the water precipitates (bubbles) grow by a mechanism of Oswald rippening (growth of the larger bubbles at the expense of the smaller ones) [46]. We interpret this later result in assuming that the water point defects of berlinite (which are related to the absorption band centered on 5200 cm^{-1}) occur essentially under the form of immobile OH's associated to aliovalent impurities (probably antisites Al_P), a situation which compares to the one of OH's attached to substitutional Al_{Si} impurities in quartz.

As noted above, conventional IRS does not provide precise information on the mode of incorporation of water in quartz; we have thus performed a NIRS study with polarized light on the crystal 140 SICN with a view to characterizing the orientations of the OH bonds of the water point defects contained in this material. It is clear that, for such experiments, the orientation of the beam versus the crystal axes is of particular importance; however in our case, because of the very low absorbance, thick crystals have to be used; a not too small absorption could be obtained only with the beam parallel to the Y axis (the studied crystal being cut in two pieces along its long dimension and both pieces being put one above the other along the beam direction, this provides a crystal with a total length of 18 cm). The polarized absorbance spectra are shown on Fig. 6. Point defects with a well defined structure should give rise to sharp absorption peaks rather than to a relatively broad band; we have assumed that the observed band results from the convolution of several sharper peaks and we have tried to deconvoluate them in assuming they all have gaussian profiles. Such a mathematical treatment shows two main peaks and at least eight smaller ones; this is in qualitative agreement with the model of substitutional $4H_{Si}$ defects which could give rise to twelve different directions for the OH dipoles in the unit cell. We have studied in detail the variation of the two main peaks with the polarization orientation. For polarizations in the (10$\underline{1}$0) plane these peaks never vanish but an anisotropic effect is clearly detected. At the moment it is unfortunately not possible to investigate other polarization directions.

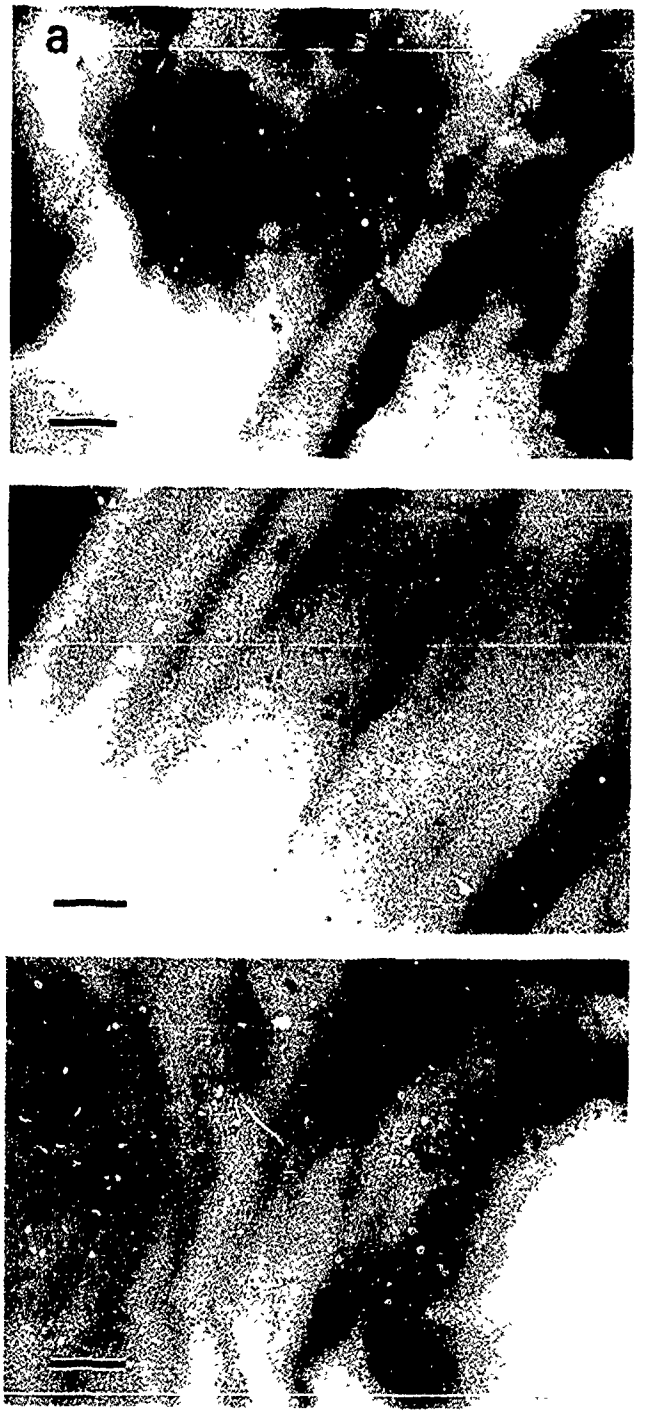


Figure 7: Topology of the tiny water precipitates formed by annealing (scale bar = 1µm). a) Bell Tel crystal (region containing H/Si ≈ 200 ppm); annealing of 90 min at 700 °C and at atmospheric pressure. b) same material annealed 240 min at 700 °C and at P = 800 MPa. c) 229 SICN crystal (H/Si ≈ 100 ppm) annealed 360 min at 700 °C and at P = 1000 MPa.

V WATER SOLUBILITY AND DIFFUSIVITY

It has been mentioned above that the annealing experiments allow the determination of the solubility of water point defects and of their diffusion coefficient. Recent theoretical models predict that the intrinsic equilibrium concentration of water point defects in quartz increases dramatically with pressure [50, 51]. To get at least one experimental value of this solubility we have annealed some specimens of the wet crystal 229 SICN which contains only point defects; these annealings have been performed under various confining pressures up to 1200 MPa and at the fixed temperature of 700 °C. We have then checked by TEM wether water has precipitated or not. The corresponding TEM micrographs are shown on Fig. 7. It is clear that at low or moderate pressure (say below 300 MPa) water precipitation is rapid and abundant, similar to what happens at atmospheric pressure. The internal pressure in the tiny precipitated bubbles is relaxed by the nucleation of small sessile dislocation loops attached to the bubbles (to allow the increase of the bubble volume some material is removed from its surface and this material forms the extra half plane of the dislocation). We believe that the energy necessary for the nucleation of these dislocation loops is provided by the gain of Gibbs energy associated to the decrease of the supersaturated concentration of point defects. At higher pressure precipitation still occurs; the density of tiny bubbles remains qualitatively the same but their size is smaller and the whole process requires more annealing time. This means that the nucleation rate of the precipitates decreases because this rate depends strongly of the supersaturation of point defects which is much lower than previously. At still higher pressure (order of 1000 MPa) precipitation becomes extremely slow and scarce; in a few regions one can observe very tiny precipitates while thev major part of the crystal is free of precipitates; such regions were probably slightly wetter than the mean concentration measured by NIRS in a large piece of material. The solubility of water point defects in quartz is thus H/Si ≈100 at. ppm at 700 °C and 1000 MPa. It is much lower at lower T and P and it is clear that experimental values for different T and P conditions still are necessary to characterize properly this important parameter.

The diffusion coefficient of water point defects has been estimated in measuring on TEM micrographs the mean distance X between tiny water bubbles which precipitated during annealing. This distance is assumed to be simply related to the diffusion coefficient of water point defects D(T) by the relation $X = (2Dt)^{1/2}$ where t is the annealing time. This is a very crude hypothesis which assumes that the time necessary for the nucleation of the precipitates is negligible as compared to the one necessary for their growth; alternatively one can say that it is assumed that critical embryos already exist and the precipitation process consists only in the growth and coalescence of the larger embryos (which become visible precipitates) at the expense of the smaller ones (which redissolve). The Fig. 8 summarizes the various values of D measured in the temperature range 350-1000 °C and at confining pressures up to 700 MPa. It can be seen that this apparent diffusion coefficient fits quite satisfactorily an Arrhenius law whatever the confining pressure and the type of quartz (α or β). We find

$$D = D_0 \exp(-Q/RT)$$

with $Q = (95 \pm 10)$ KJ mole^{-1} and $D_0 = (2 \pm 1) \times 10^{-12}$ m^2 s^{-1}.

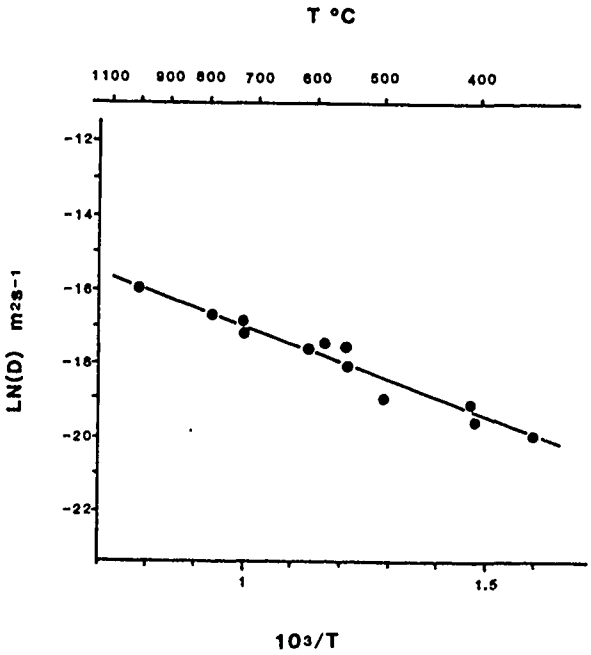


Figure 8: Plot of Ln D (T) versus 1/T; D (T) is the diffusion coefficient of the water point defects.

This corresponds to a very low mobility at room temperature but the drift velocity could be no more negligible at T = 500 to 550 °C and under an electric field of the order of 5000 V cm^{-1} which are the usual sweeping conditions. Furthermore it is to be remembered that diffusion in quartz is highly anisotropic; the diffusion coefficient of a number of species is several orders of magnitude larger along the **c** axis (in the channels of the structure) than in the basal plane. This is probably the case for water point defects too and the D value measured here should correspond to the lowest diffusion direction. In a sweeping experiment with an electric field parallel to **c**, water and/or other foreign species could diffuse much more rapidly than predicted by the above law.

VI CONCLUDING REMARKS

The above results are now briefly discussed.

i) The sharp absorption peaks observed on conventional IR spectra are generally attributed to OH's associated to impurities [32]; their increase after annealing has already been observed by other authors [47] who attributed this phenomenon to the formation of microcracks during annealing under pressure; in the case of our annealing experiments at atmospheric pressure no microcracks occur and the above explanation must probably be discarded. A more satisfactory explanation could be the following: in the as-grown material some impurities were compensated by holes, after annealing they become compensated by H's thus increasing the corresponding absorption peaks.

ii) In contrast with the previous case the decrease of IR absorbance in the range 3400-2900 cm^{-1} corresponds to a change of the broad absorption band which is generally assumed to characterize the presence of water in an otherwise chemically pure material; after annealing this broad band becomes narrower, more symmetric and very similar to the one of pure water [36, 48]. The shape of the IR spectrum and the area under the curve change after annealing and this questions the methods used for measuring the total content of water in quartz. It would perhaps be necessary to consider two distinct types of molecular water with absorbances at slightly different wave numbers, the true liquid phase in the tiny bubbles and water molecules chemisorbed on their surfaces; the first type would increase at the expense of the second one during annealing when the bubbles grow by Oswald rippening.

iii) The nature of the point defects in quartz still is debated. Several models have been proposed [39 to 42, 49]; the above NIR spectra are consistent with the $4H_{Si}$ substitutional model of Nutall and Weil [39] but no definitive conclusion can be drawn yet.

REFERENCES

[1] BRICE J.C. Crystals for quartz resonators Reviews of modern physics **57** 105-146 (1985)

[2] HALLIBURTON L. E., MARTIN J. J. and KOEHLER D. R. Properties of piezoelectric materials. In E. A. GERBER and A. BALLATO (eds) Precision frequency control Vol. 1 Academic Press p. 1-45 (1985)

[3] BESSON R. J. Quartz crystal resonators Proc. First Eur. Time Freq. Forum. p. 20-30 (1987)

[4] LANG A. R. and MUISKOV V. F. J. Appl. Phys. **38** 2477 (1967)

[5] McLAREN A. C., OSBORNE C. F. and SAUNDERS L. A. Phys. Stat. Sol. **4** 235 (1971)

[6] BARNES R. L., FREELAND P.E., KOLB E. D., LAUDISE R. A. and PATEL J. R. J. Crys. Growth **43** 676 (1978)

[7] CROXALL D. F., CHRISTIE I. R., HOLT J. M., ISHERWOOD B. J. and TODD A. G. 36th Ann. Freq. Contr. Symp. p.62 (1982)

[8] ARMINGTON A. F. and LARKIN J. J. J. Cryst. Growth **71** 799 (1985)

[9] CAPELLE B. and ZARKA A. Private communication

[10] PARPIA D. Y. Philos. Mag. **33** 715 (1976)

[11] PARPIA D. Y. Philos. Mag. **37** 375 and 401 (1980)

[12] YOSHIMURA J. and KHORA K. J. Cryst. Growth **33** 311 (1976)

[13] ZARKA A., LIN L. and BUISSON X. J. Cryst. Growth **54** 394 (1981)

[14] SPENCER W. J. and HARUTA K. J. Cryst. Growth **33** 311 (1976)

[15] AUVRAY P. and REGRENY A. Bull. Minéral. **96** 267 (1973)

[16] BARNS R. L., KOLB E. D. and LAUDISE R. A. J. Cryst. Growth **34** 198 (1976)

[17] JOHNSON G. R. and IRVINE R. A. 41st Ann. Freq. Contr Symp. p. 175 (1987)

[18] IWASAKI F. J. Cryst. Growth **39** 291 (1977)

[19] IWASAKI F. and KURASHIGE M. Jap. J. Appl. Phys. **17** 917 (1978); also Ferroelectrics **43** 43 (1982)

[20] CORDIER P., BOULOGNE B. and DOUKHAN J. C. Bull. Minéral. **111** 113 (1988)

[21] MORIYA K. and OGAWA T. J. Cryst. Growth **44** 53 (1978) and J. Cryst. Growth **58** 115 (1982)

[22] SER A., BIDEAU J. P., CLASTRE J. and ZARKA A. J. Appl. Crys. **13** 50 (1980)

[23] HOMMA S. and IWATA M. J. Cryst. Growth **19** 125 (1973)

[24] FLICSTEIN J. and SCHIEBER M. J. Cryst. Growth **24** 603 (1974)

[25] CORDIER P., DOUKHAN J. C., BIGNON O. and BUISSON X. 2nd Eur. Time Freq. Forum p. 921 (1988)

[26] CHAKRABORTY D. J. Cryst. Growth **36** 188 (1976): also J. Cryst. Growth **41** 177 (1977) and J. Cryst. Growth **44** 599 (1978)

[27] BALASCIO J. F. and ARMINGTON A. F. 38th Ann. Freq. Contr. Symp. p. 5 (1985)

[28] KAHAN A., EULER F. K. and LIPSON H. G. 41st Ann. Freq. Contr. Symp. p. 216 (1987)

[29] LIPSON H. G. and KAHAN A. 38th Ann. Fre , Contr. Symp. p.10 (1985) and J. Appl. Phys. **58** 963 (1985)

[30] DODD D. M. and FRASER D. B. J. Phys. Chem. **26** 673 (1965)

[31] KATS A. and HAVEN Y. Phys. CHem. Glasses **1** 99 (1960)

[32] KATS A. Philips Res. Report **17** 133 and 201 (1962)

[33] BRICE J. C. and COLE A. M. 32nd Ann. Freq. Contr. Symp. p. 1 (1978) and J. Appl. Phys. **12** 459 (1979)

[34] PATERSON M. S. Bull. Minéral. **105** 20 (1982)
[35] AINES R. S., KIRBY S. H. and ROSSMAN G. R. Phys. Chem. Minerals **11** 204 (1984)
[36] AINES R. D. and ROSSMAN G. R. J. Geophys. Res. **89** 4059 (1984)
[37] STOPLER E. Contrib. Mineral. Petrol. **81** 1 (1982)
[38] HUNT G. R. and SALISBURY J. W. Modern Geology **1** 283 (1970)
[39] NUTALL R. D. H. and WEIL J. A. Solid Sate Comm. **33** 99 (1980)
[40] WEIL J. A. Phys. Chem. Minerals **10** 149 (1984)
[41] HAGON J. P., STONEHAM A. M. and JAROS M. Philos Mag. **B 55**(2) 211 and 225 (1987)
[42] GRIGGS D. T. Geophys. J. R. Astr. Soc. **14** 19 (1967)
[43] McCARTHY J. J. and SCHAMBERS F. H. Proc. Nal Bureau of Standards Spec. Pub. **604** p.273 (1981)
[44] CORDIER P. and DOUKHAN J. C. Eur. J. Mineral. **2** 221 (1989)
[45] DETAINT J., PHILIPPOT E., JUMAS J. C., SCHWARTZEL J., ZARKA A., CAPELLE B. and DOUKHAN J. C. 39th Ann. Freq. Cont. Symp. p. 234 (1985)
[46] BOULOGNE B., CORDIER P. and DOUKHAN J. C. Phys. Chem. Minerals **16** 250 (1988)
[47] ROVETTA M. R., HOLLOWAY J. R. and BLACIC J. D. Geophys. Res. Lrt. **13** 145 (1986)
[48] MACKWELL S. J. and PATERSON M. S. In Point Defects in Minerals, SCHOECK R. N. (ed) Amer. Geophys. Union Washington p.141 (1985)
[49] BRUNNER G. O., WONDRATSCHEK and LAVES F. Z. Elektrochem. **65** 735 (1961)
[50] J. C. DOUKHAN and PATERSON M. S. Bull. Minéral **109** 193 (1986)
[51] M. S. PATERSON Phys. Chem. Minerals **13** 245 (1986)

43rd Annual Symposium on Frequency Control - 1989

Stroboscopic X-ray topography of quartz resonators

B.CAPELLE, A.ZARKA, Y.ZHENG
Laboratoire de Minéralogie-Cristallographie,
Universités P. & M. CURIE (Paris VI) et Paris VII, unité associée au CNRS,
4 place Jussieu, 75252 Paris Cedex 05, France

J.DETAINT and J.SCHWARTZEL
Centre National d'Etudes des Télécommunications,
PAB/BAG/MCT, 92220 Bagneux, France

Abstract

AT-cut quartz resonators containing dislocations, etch channels or growth bands have been studied by the stroboscopic X-ray topography using the synchrotron radiation. Experimental images and simulated ones presented here show that the acoustic wave is perturbed in the vicinity of dislocations and indicate that non-linear terms have to be considered to describe the acoustic deformation field near dislocations. Similar experimental results were obtained in vibrating resonators containing etch channels and suggest that the influence of etch channels may be more important than for the case of dislocations. The effect of growth bands on acoustic waves is much less localized than for the previous cases, but important too, since the images of vibration amplitude enveloppes are mostly perturbed.

Classic X-ray topographs using the synchrotron radiation are also presented here to show coupled modes in AT-cut quartz resonators. In particular the u_2 and u_3 components of the acoustic displacement were visualized simultaneously with the main thickness-shear component u_1. An interesting mechanism of the mode coupling has been evidenced.

I. Stroboscopic X-ray topography on crystal defects

One nanosecond time resolved X-ray topography has been performed to study acoustic vibrational states in quartz resonators using the synchrotron radiation (LURE, France). The principle of the experimental set-up consists on the exact synchronization between the acoustic vibration and the synchrotron radiation [1,2]. Results already obtained have shown that this technique is very performant to determine low vibration amplitudes (0-7Å), through the *Pendellösung* fringes in section topographs and through equi-level curves of vibration amplitudes in traverse topographs [2,3]. The proposal here is to characterize the effects of dislocations, etch channels and growth bands by this technique.

When an acoustic wave and a dislocation are both present in a crystal, the total displacement field can be written in a general form as following:

$$\mathbf{u}(\mathbf{r},t) = \mathbf{u}_a(\mathbf{r},t) + \mathbf{u}_d(\mathbf{r}) + \mathbf{u}_i(\mathbf{r},t)$$

where $\mathbf{u}_a(\mathbf{r},t)$ is the acoustic displacement without dislocation, $\mathbf{u}_d(\mathbf{r})$ the static dislocation displacement without acoustic wave and $\mathbf{u}_i(\mathbf{r},t)$ the displacement resulted of the interaction between the acoustic wave and a dislocation. For linear piezoelectricity theory, no interaction is expected and $\mathbf{u}_i(\mathbf{r},t)$ is equal to zero.

To clarify the existence or not of the interaction term $\mathbf{u}_i(\mathbf{r},t)$, a dislocation was selected from a synthetic AT-cut quartz resonator operating on the fifth overtone thickness-shear mode. The Burgers vector $\mathbf{b}$ of the dislocation was identified by X-ray section topographs, $\mathbf{b}=[100]$ for the present case. It is an almost edge dislocation. Stroboscopic X-ray section topographs were taken with the reflection vector $\mathbf{g}=2\bar{1}0$ which is parallel to the dislocation Burgers vector $\mathbf{b}=[100]$ and to the acoustic displacement vector $\mathbf{u}_a(\mathbf{r},t)$ which is reduced essentially to the thickness-shear component u_1. The Bragg angle θ_B used is 8.3406° which corresponds to a seclected X-ray wavelength λ of 0.7127Å. The X-ray extinction distance Λ_0 is equal to 96.396μm.

Figure 1a represents a section topograph of the dislocation without acoustic vibration. Two

CH2690-6/89/0000-470 $1.00 C 1989 IEEE

arrows indicate roughly the zone over which the dislocation displacement $u_d(r)$ (more precisely the induced misorientation of the reflecting planes) exhibits effects in the X-ray diffraction pattern. The corresponding simulated image is represented in figure 1b where the dislocation displacement $u_d(r)$ was calculated from the linear anisotropic elastic theory [4]. It is noticed that most characteristic contrasts of the dislocation image can be simulated. Some differences, namely near the dislocation core image, exist and this is probably due to the polychromatism of the X-ray incident beam [5,6] and other experimental conditions not taken into account in the calculation.

When an acoustic wave is excited in the resonator, the stroboscopic section topograph (Fig. 1d) shows two types of contrasts:

- on the upper part of the image far from the dislocation, interference fringes due to the acoustic wave can be seen and they seem to be not altered by the presence of the dislocation;
- near the dislocation on the lower part, a drastical change is recorded in comparison with figure 1a and fringes dues to the acoustic wave are completely destroyed.

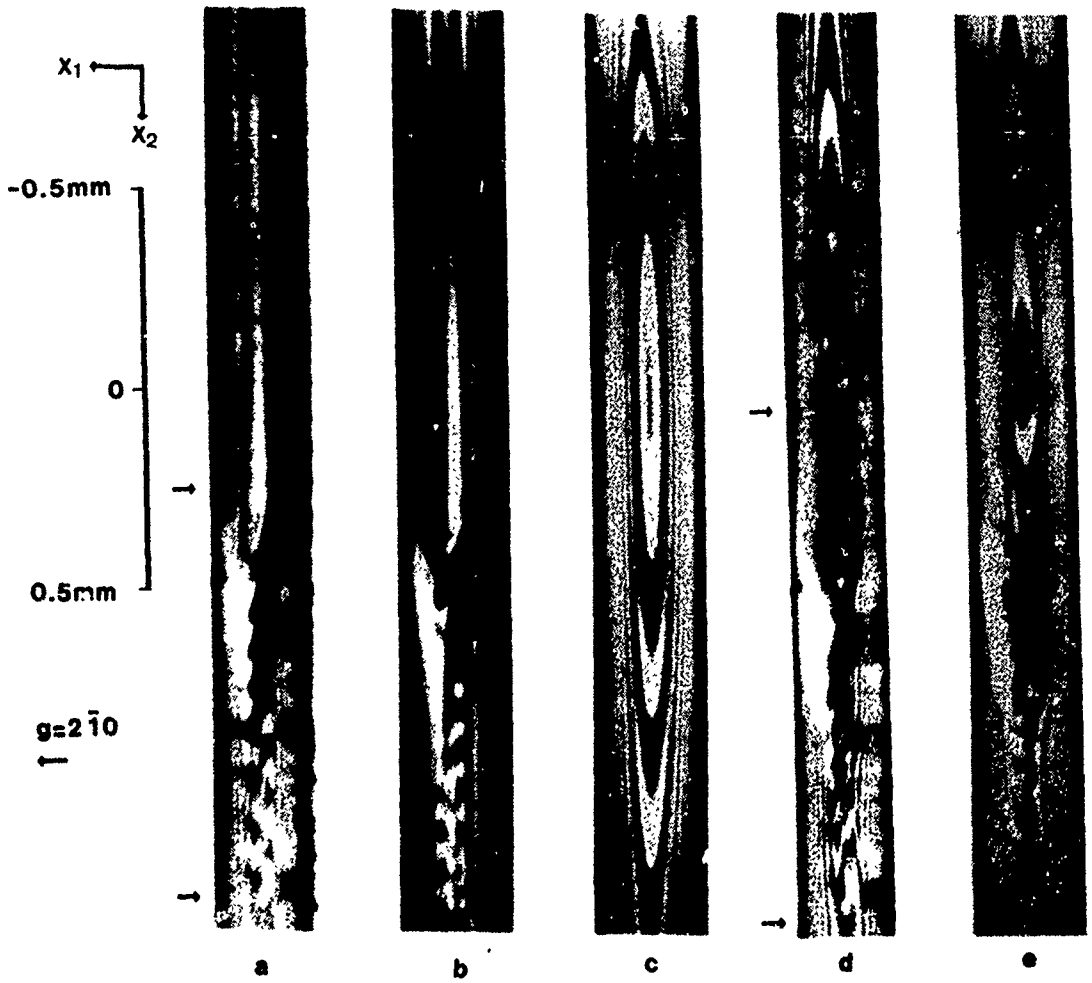


Fig.1 Experimental and simulated stroboscopic X-ray section topographs of a dislocation with increasing vibration amplitudes.

a.-b. Experimental and simulated images of the dislocation in absence of acoustic vibration;

c. Simulated image for the virtual case where the dislocation is removed out and only an acoustic vibration exists (with the vibration amplitude u_o equal to 2.41Å at the resonator centre);

d.-e. Experimental and simulated images of the dislocation in presence of a vibration with an amplitude u_o of 2.41Å at the centre. In the simulated image a linear addition of the dislocation and vibration displacements was made.

In order to analyse this image, two simulations were performed. The first simulated image (Fig. 1c) corresponds to the virtual case where only the acoustic wave is considered. This case permits to know the zone where the acoustic wave is not affected by the presence of the dislocation. The interesting point is that, in comparison with the dislocation alone case (Fig. 1a), the dislocation influence zone is enlarged (indicated by arrows in figure 1d). The acoustic displacement $u_a(r,t)$ used in the calculation was from the linear piezoelectricity theory [7,8]. It can be seen that this acoustic deformation model is a very good approximation far from the dislocation. The second simulated image (Fig. 1e) took into account both the dislocation displacement $u_d(r)$ and the acoustic displacement $u_a(r,t)$ used previously. This calculation allows to know whether a simple linear addition $u(r,t) = u_a(r,t) + u_d(r)$ of two displacements permits or not to reproduce most features of the experimental image. It can be easily seen that this image is very different of the experimental one over all the zone in the vicinity of the dislocation. It means that the interaction term $u_i(r,t)$ must be considered near the dislocation and linear theories are not suffisant for the present case. This point is important because it indicates a different behaviour of the acoustic wave around the dislocation and it suggests the existence of a dislocation associated vibration component in the resonator. An exact calculation of $u_i(r,t)$ should be difficult and its action range around the dislocation can be roughly

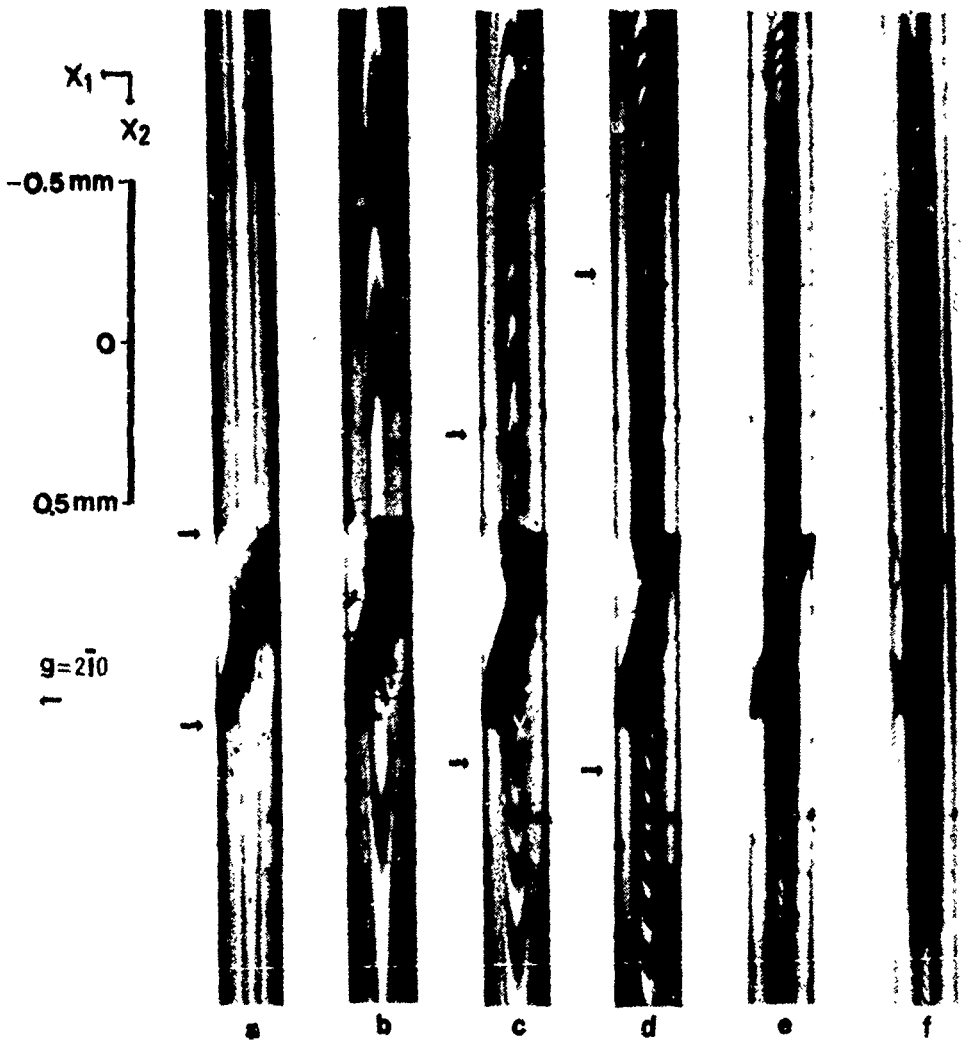


Fig.2 Experimental stroboscopic X-ray section topographs of an etch channel with increasing vibration amplitudes.

a. Image of the channel without acoustic vibration;

b-d. Images of the channel with acoustic vibrations of low amplitude;

e-f. Images of the channel with acoustic vibrations of high amplitude.

estimated to be 90μm for the present case. In a general case, $u_1(r,t)$ should depend also on the dislocation geometry.

Stroboscopic experiments were also performed on synthetic AT-cut quartz resonators containing etch channels and operating on the third overtone thickness-shear mode. The X-ray reflection parameters θ_B, λ and Λ_0 taken were similar to previously and the electrical excitation were increased progressively. Figures 2a-f show a series of stroboscopic X-ray section topographs on a channel with increasing vibration amplitudes. For the case of low amplitude vibrations ($u_0 \leq 10$Å), interference fringes due to the acoustic wave can be seen (Fig. 2b-d) and they are strongly deformed in the vicinity of the etch channel. The channel influence zones are indicated by arrows and it can be seen that these zones are enlarged in function of vibration amplitudes. This means that an interaction between the acoustic wave and the channel should exist and the range of this interaction term can be roughly estimated to be 800μm around the etch channel for the present case. This value is approximate because of the existence of many different channel types and the ranges of the channel influence may vary more largely than for the case of dislocations.

In case of growth bands which are extent defects, stroboscopic experiments were performed on traverse topographs to get a global view over a resonator. The reflection parameters used are similar to that of the previous cases and resonators studied operated on the fifth overtone thickness-shear mode. Figures 3a-d show a series of traverse topographs of growth bands with increasing vibration amplitudes. It can be observed in figure 3c, where the vibration amplitude is low, that characteristic elliptic curves of the acoustic vibration exist even in presence of growth bands and they are deformed in the present case. Figure

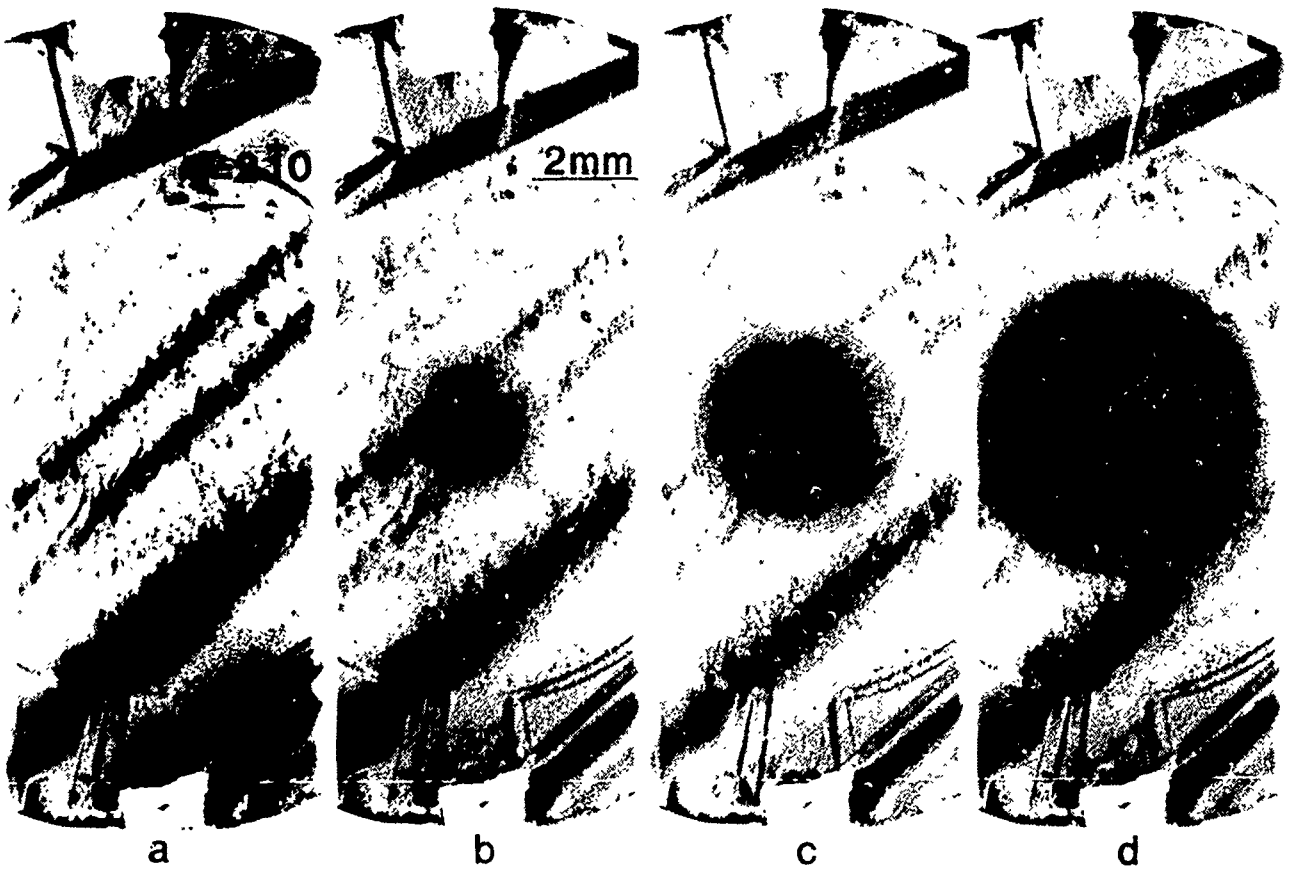


Fig.3 Experimental stroboscopic X-ray traverse topographs of growth bands with increasing vibration amplitudes.

a. Image of the growth bands without acoustic vibration;

b.-c. Images of the channel with acoustic vibrations of low amplitude.

d. Image of the channel with an acoustic vibration of high amplitude.

3d taken with a higher vibration amplitude shows a dark contrast on the central part of the resonator, as expected. But on the edges of this contrast which represents the enveloppe of vibration amplitudes, contrasts of irregular forms can be seen. This indicates that growth bands have an important effect and they perturb probably the spatial coherence of the acoustic wave.

II. Classic X-ray topography on coupled modes

The identification of vibration modes by the X-ray topography is quite common. The particularity of the investigation carried out here is the utilization of the synchrotron radiation which permits the observation of several diffraction patterns simultaneousiy, thus the observation of the three spatial components u_1, u_2 and u_3 of the acoustic displacement. More precisely the three spatial components concerned should be the three eigen displacements u_1', u_2' and u_3' which are very close to u_1, u_2 and u_3 for the case of AT-cut quartz [8]. For the sake of simplicity this conversion will be omitted here.

Two particular diffraction vectors $g_1=2\bar{1}0$ and $g_2=0\bar{1}3$ were chosen in the case of AT-cut quartz resonators to separate the main thickness-shear component u_1 from u_2 and u_3. As it can be recalled here, the sensibility of the X-ray diffraction to a deformation field $\mathbf{u}(r)$ is determined by the product $\mathbf{g}.\mathbf{u}(r)$. For the present case, one has:

$$\mathbf{g}_1.\mathbf{u}(r) \propto u_1(r) \quad \text{and}$$
$$\mathbf{g}_2.\mathbf{u}(r) \propto 0.213\ u_2(r) + 0.977\ u_3(r)\ .$$

So, only the main component u_1 is visualized with $\mathbf{g}_1$. The two components u_2 and u_3 are visualized together with $\mathbf{g}_2$ and they cannot be completely separated because they do not correspond to simple crystallographic directions. This u_2 and/or u_3 component will be noted as u_2/u_3 below.

Using these two diffraction vectors, high quality natural AT-cut quartz resonators were analysed in function of resonance frequencies. Three examples are shown below to illustrate different mechanisms of the mode coupling. The resonator concerned here has the following parameters:

- maximal thickness: $2h_0 = 1.3234$ mm
- curvature radius: $R = 175$ mm
- resonator diameter: $D = 15$ mm
- electrode diameter: $d = 8$ mm
- electrode mass loading: $2\rho'h'/\rho h_0$ => ajusted for $f_{500}=6.338565$MHz

Figures 4a,b show a resonator operating on the fifth overtone thickness-shear mode (nmp=500 and f=6.338565MHz) which is a well trapped mode. According to the theories well established for trapped modes [7,8], one may estimate that the magnitudes of u_2 and u_3 are of respectively one and two orders less than that of u_1. The u_2/u_3 component revealed by figure 4b has a non-negligeable value and it is interesting to note the amplitude form of this component with two supplementary nodal lines in comparison with the main component u_1 (Fig. 4a). It will be useful to check whether the value and form of the u_2/u_3 component correspond to the theoretical prediction.

Figures 5a,b show a case where the excited mode is untrapped (nmp=120 and f=1.310256MHz). The extension of the mode until the resonator edges implies a strong coupling of modes. The u_2/u_3 component observed in figure 5b has a high value and its geometric form is mostly complicated. The coupling for untrapped modes has always been observed experimentally, but its theoretical modelling is quite difficult.

Figures 6a,b reveal a surprising case of a strong coupling for a well trapped mode (nmp=720 and f=8.967894MHz). For this mode, it appears that the u_1 component is strongly coupled with the u_2/u_3 one of a similar amplitude. This u_2/u_3 component is well trapped and has a amplitude form similar to an antisymetrical anharmonic with two nodal lines along the x direction and three nodal lines along the z' direction. Similar coupled modes involving u_2/u_3 were also observed for all other anharmonic modes of the seventh overtone.

A strong coupling for well trapped modes is not predicted by the theories. As it can be remarked that these strong coupled modes are of the real operating ones in practice for AT-cut quartz

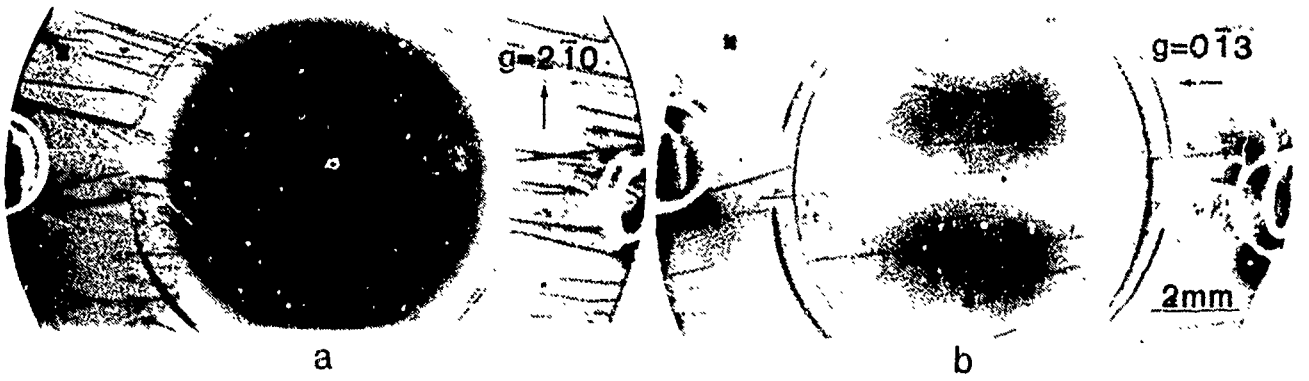


Fig 4 Classic X-ray topographs showing a mode coupling for the case of a well trapped mode (nmp=500). a). the main thickness-shear component u_1, b). the u_2/u_3 component.

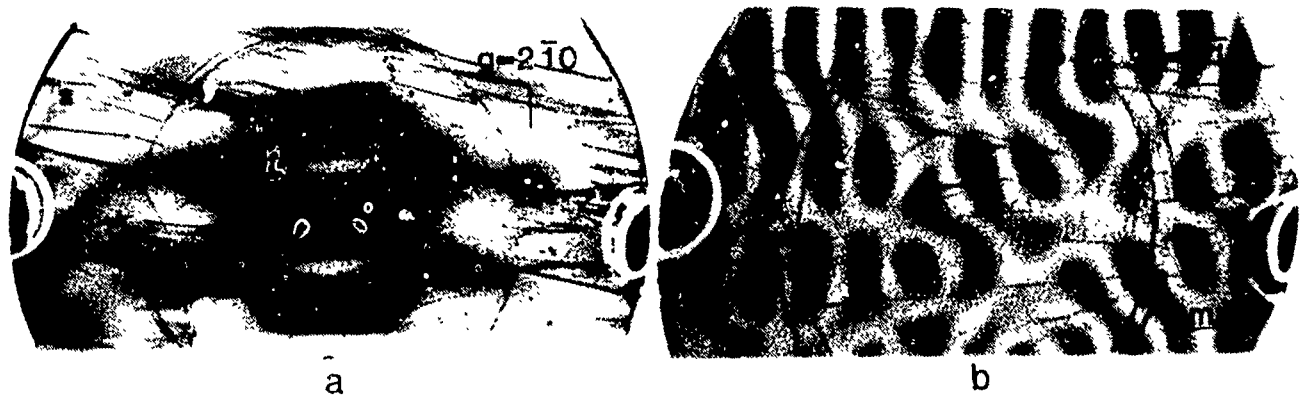


Fig 5 Classic X ray topographs showing a strong mode coupling when the excited mode is untrapped (nmp=120) a) the main thickness-shear component u_1, b). the u_2/u_3 component

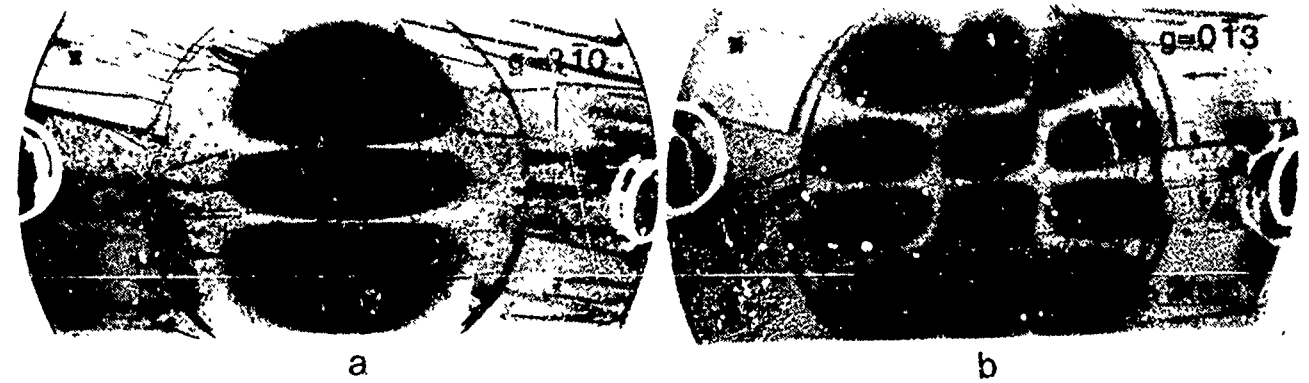


Fig 6 Classic X ray topographs showing a strong mode coupling when the excited mode is well trapped (nmp=720) a) the main thickness shear component u_1 b) the u_2/u_3 component

resonators, a new theoretical modelling seems to be necessary for the present case in order to clarify this mechanism of the mode coupling. The idea of the simultaneous excitation of an u_1 eigenmode and an u_2 or u_3 eigenmode could be suggested. In such case, the geometric parameters of a resonator, which determine the eigen frequencies of u_1, u_2 and u_3, should play an essential role, i.e. by varing these parameters the mode coupling might be stronger or weaker, or even avoided.

Conclusion

Stroboscopic experiments performed on quartz resonators containing dislocations and channels have provided interesting informations, namely on the existence of an interaction between the acoustic wave and the crystal defects. It suggests that a non-linear treatment should be carried out to describe a more exact acoustic behaviour in the vicinity of defects. This treatment will be useful to get local acoustic characteristics such as the deviation of the acoustic wave propagation and global acoustic characteristics such as the energy dissipation through crystal defects.

Classic X-ray topography using the synchrotron radiation has revealed the existence of strong coupled modes for certain well trapped modes. A further theoretical analysis concerning this mechanism of the mode coupling is very interesting for the basic understanding of the resonator theory and important for the practice.

Acknowledgment

The authors wish to acknowledge SICN (France) and Mr. Euler of RADC ESE (USA) for kindly providing crystals. This work was supported by DRET (France) through the contract no. 87 34 13 80 04 70 75 01.

References

[1] ZARKA,A., CAPELLE,B., DETAINT,J. and SCHWARTZEL,J. Proceedings of the 41st Annual Symposium on Frequency Control, p.236-240, 27-29 May 1987, Philadelphia, U.S.A.

[2] ZARKA,A., CAPELLE,B., ZHENG,Y., DETAINT,J. and SCHWARTZEL,J. Proceedings of the 42nd Annual Symposium on Frequency Control, p.85-92, 1-3 June 1988, Baltimore, U.S.A..

[3] ZHENG,Y., ZARKA,A., CAPELLE,B., DETAINT,J. and SCHWARTZEL,J. (1989). Acta Cryst., A**45**, 275-285.

[4] HIRTH,J. and LOTHE,J. (1968). *Theory of dislocations*, MacGraw-Hill, New-York, p.411.

[5] AUTHIER,A. and BALIBAR,F. (1970). Acta Cryst., A**26**, 647-654.

[6] MILTAT,J.E.A. and BOWEN,D.K. (1975). J. Appl. Cryst., **8**, 657-669.

[7] TIERSTEN,H.F. and SMYTHE,R.C. (1979). J. Acoust. Soc. Am., **65**, 1455-1460.

[8] STEVENS,D.S. and TIERSTEN,H.F. (1986). J. Acoust. Soc. Am., **79**, 1811-1826.

43rd Annual Symposium on Frequency Control - 1989

FUNDAMENTAL LATTICE VIBRATIONS IN QUARTZ

Robert Aaron Murray and John G. Gualtieri
U.S. Army Electronics Technology and Device Laboratory
SLCET-EO
Fort Monmouth, New Jersey 07703-5000

Abstract

The fundamental lattice vibrations in quartz (normal modes) were examined in an attempt to find better test methods to characterize impurities. A computer model of alpha quartz was constructed using the atomic positions in the quartz unit cell and a Keating-type interatomic potential. Using this model, the infrared (IR) absorption was calculated for the fundamental modes. This perfect (no defect) model was compared with experimental IR absorption results that we obtained and to experimental data taken from the literature, and was found to be in qualitative agreement with both. Another model was also constructed which simulated an aluminum substitutional defect by changing the mass of one of the three silicon atoms in the unit cell to the mass of aluminum. This represents an Al to Si concentration of 33 percent. The IR absorption results of the "defect" model showed very little change in the absorption peaks. The bands were shifted to higher wave numbers in the defect model, with the maximum movement at the 800 cm^{-1} band, which shifted about 6 cm^{-1}, but no separate peaks due to Al were formed. A shift of this magnitude would not be resolvable when applied to the Al concentrations found in high-grade quartz. Experiments on the IR absorption of electronic grade quartz with different aluminum concentrations were conducted and in fact showed no differences. Even though this method was found to be impractical for aluminum detection, it did result in the construction of a relatively good numerical model for studying the fundamental atomic vibrations in quartz.

Introduction

For users and growers of quartz the problem of accurate determination of the aluminum content has proved to be a demanding task. The standard chemical methods, such as atomic absorption and mass spectroscopy, are difficult to perform because the samples must be digested in hydrofluoric acid. This introduces the possibility of contamination during sample preparation. At present, a favored method to accurately test for Al is the electron spin-resonance method. This procedure looks for the electron hole signal of an oxygen atom at the Al site. However, the sample must be irradiated to activate the Al hole centers and so requires special sample preparations. With this in mind, it was felt that a simple infrared test for Al would be very desirable. It is generally believed that when the Al substitutes for Si in the quartz lattice, the number of oxygen bonds remains constant at four. We felt that even if the bonding environment of the Al and Si were exactly the same (which they are not), the fundamental vibrational modes involving Al and Si would occur at different frequencies, based on the mass difference of the two atomic species. These vibrations could then be seen in the far IR. Additionally, changes due to Al would be easier to interperet in the normal modes than in the overtone or combination bands.

Several groups had conducted far IR studies of quartz in the frequency range 0-1500 cm^{-1}.[1-4] Some had tried to determine the fundamental frequencies from the optical data, however, there were too many peaks in the IR to pin down the normal modes exactly using this method. One group also looked for evidence of Al in quartz without success.[2] We decided to approach the problem from a different perspective. In situations such as this, computer modeling can be of tremendous help. If a model could be developed that reproduced the available data correctly, then this model could be extended and changed to provide information on situations for which there is no good data. In this case, if we could construct a model that reproduced the infrared spectra of pure quartz then, possibly, this model could be modified to predict the IR response of quartz containing the Al impurity. Since the Al peak (if it existed) was likely to be small in cultured quartz due to the low level of the impurity (a few parts per million), it was felt that a good computer model would give us an edge in deciding where to look for it.

Methods

The experimental IR measurements were performed on a Perkin-Elmer 1760 Fourier transform spectrometer. Due to instrument limitations, the lowest wave number accessible to us was 450 cm^{-1}. Thus, part of the spectrum of fundamental modes was unavailable for our study. All samples were Motorola AT-cut blanks from two lot numbers. The first set of samples was from lot GC2-48 with an Al content of 3 ppm as

determined by atomic absorption (AA) tests performed by Motorola and by mass spectrometer (MS) methods performed at Fort Monmouth. These samples are hereafter referred to as the low Al samples. The second set of samples was from lot GC1-7. The samples were determined to contain 20 ppm Al by the AA method and 9 ppm by the MS procedure. This difference underscores the difficulties encountered in attempting accurate Al analyses. Due to the large absorption coefficient of quartz throughout the 0 - 1500 cm^{-1} region, the samples had to be made extremely thin. All the samples used in this study were physically lapped and then chemically etched to a thickness of less than 20 μm. This necessitated careful handling with a vacuum chuck to reduce breakage of the blanks. The IR readings were taken at room temperature with the beam perpendicular to the sample face. For all of the results reported here, the spectrometer beam was unpolarized. The results are reported in the "Results" section.

The starting point for the computer model for perfect quartz began with the atomic positions of Si and O in the crystallographic unit cell, as determined by X-ray diffraction.[5] The fractional atomic positions for the non-orthogonal unit cell are given in Table 1.

Table 1
Atomic Positions of Alpha Quartz

atom type	x	y	z
Si	.535	.535	.333
Si	.465	.000	.000
Si	.000	.465	.667
O	.415	.272	.120
O	.857	.585	.453
O	.728	.143	.787
O	.143	.728	.880
O	.272	.415	.547
O	.585	.857	.213

The cell dimensions are A = B = 4.913 Å and C = 5.405 Å. The angle between A and B is 120°. In the normal modes calculation, the unit cell was configured with periodic boundaries. This effectively simulates an infinite bulk of perfect quartz with no guesswork necessary for the surface terms. The Al defect model used the same atomic positions listed in Table 1. We felt that, given the similarities of the Al and Si bonding patterns (both bond to four O atoms), the major difference would be due to the differences in atomic masses. Consequently, in the vibrational calculations the atomic mass of Al was substituted at one of the Si sites in the unit cell. Because of the periodic nature of the cell, this amounted to a 33 percent Al to Si concentration. The mass change was the only difference between the two models.

The atomic positions alone are useless without an accurate interatomic potential. In ionic solids, a simple, centrally symmetric potential (two-body potential) is sufficient to mimic the Coulomb forces on the atoms. However, in quartz the atomic bonding is mostly covalent hence, a three-body potential (one that incorporates bond bending as well as simple bond stretching) must be used to obtain acceptable results. In the calculations discussed here we used a Keating-type potential, which is given in Eqs. (1) and (2). This potential has been shown to provide good results in modeling the vibrational modes of silicate glass.[6]

$$V^{Si}_{\ell} = \frac{3}{16}\sum_{i}\frac{\alpha}{d^2_i}\left(|r_{\ell i}|^2-d^2_i\right)^2 + \frac{3}{8}\sum_{ij}\frac{\beta_1}{d_i^2}\left(r_{\ell i}\cdot r_{\ell j}-d_i^2\cos\phi\right)^2 \qquad (1)$$

$$V^{O}_{n} = \frac{3}{16}\sum_{i}\frac{\alpha}{d^2_i}\left(|r_{ni}|^2-d^2_i\right)^2 + \frac{3}{8}\sum_{ij}\frac{\beta_2}{d_i^2}\left(r_{ni}\cdot r_{nj}-d_i^2\cos\theta\right)^2 \qquad (2)$$

Here, V^{Si}_{ℓ} is the potential of the ℓth Si atom and V^{O}_{n} is the potential of the nth O atom. The first term on the right side of both Eq. (1) and Eq. (2) is the bond-stretching potential summed over the covalent bonds of the nearest-neighbor atoms. For Si, the sum is equal to four (due to the fourfold coordination of the Si site). For the O sites, the sum is two, because in quartz, each O atom bonds two Si atoms. The constant d_i is the "ideal" Si-O bond length. In this calculation, it was taken to be the mean Si-O bond distance in alpha quartz, 1.609 Å. The second term in Eq. (1) is the O-Si-O bond-bending potential; the "ideal" O-Si-O angle ϕ is 109.4°, which is the perfect SiO_4 tetrahedral angle. The second term in Eq. (2) is the Si-O-Si bond-bending term and θ is equal to 144°. This is the mean bridging angle the O atoms make in alpha quartz. The remaining constants are the potential strength constants for each term of the equations. These were determined empirically in modeling studies of silicate glasses[6] to optimize the fit to in-

frared absorption curves to be $\alpha = 371$ N/m, $\beta_1 = .14\alpha$, and $\beta_2 = .057\alpha$. These same values were found to provide the best fit with the experimental data in our calculations, and were used to generate all the results discussed in this paper.

After the potential equations have been determined, the calculation of the fundamental modes can be accomplished by employing eigenvalue methods for classical harmonic oscillators. These can be found in most texts on classical mechanics. The application of this procedure for atoms in a solid can also be found in the literature.[6-11] The Lagrange equation is,

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = 0 \quad , \tag{3}$$

where L=T-V. The kinetic energy is T and V is the potential energy. These are given by

$$T = \frac{1}{2}\sum_{kl} M_{lk}\dot{x}_l\dot{x}_k \tag{4}$$

$$V = \frac{1}{2}\sum_{kl} K_{lk}x_l x_k \quad . \tag{5}$$

The x_k are the generalized atomic coordinates and the M_{lk} is the mass matrix. If the x_k are orthogonal, then M_{lk} is a diagonal matrix. The K_{lk} term is given by

$$K_{lk} = \frac{\partial^2 V}{\partial x_l \partial x_k}\bigg|_{x_1=x_2=\ldots=x_f=0} \quad . \tag{6}$$

Note that the K_{lk} terms are evaluated at the equilibrium x_k positions. We can let the vector x be

$$x = ae^{-i\omega t} \quad . \tag{7}$$

Now, the Lagrange equations can be solved by setting the following determinant to zero.

$$\det \begin{vmatrix} K_{11}-\omega^2 M_{11} & \cdots & K_{1f}-\omega^2 M_{1f} \\ \vdots & \cdots & \vdots \\ K_{f1}-\omega^2 M_{f1} & \cdots & K_{ff}-\omega^2 M_{ff} \end{vmatrix} = 0 \tag{8}$$

This is solved for the eigenvalues ω_i (which are the normal mode frequencies) and the eigenvectors a_i. Thes a_i terms are the direction and relative magnitude of the atomic movements in real space.

After the eigenvalue problem has been solved, the quantum mechanical coupling of incident electromagnetic radiation to the vibrating atoms must be evaluated in order to calculate the infrared absorption. The derivation of this interaction can be found in the work of Bell[11] and in reference texts on vibrational spectroscopy.[12,13] For infrared light, the absorption interaction potential is u, the intrinsic dipole moment of the system, and u can be expanded to show the dependence of the atomic displacements as

$$u_\alpha = u^0{}_\alpha + \sum_{j\beta} \frac{\partial u_\alpha}{\partial x_{j\beta}}\bigg|_{\text{at equilibrium}} x_{j\beta} + \ldots \quad , \tag{9}$$

where u_α is the Cartesian component of u and $x_{j\beta}$ is the displacement of atom j in the ß direction. Using the linear term, the interaction matrix can be represented as

$$\langle E'|u|E\rangle \propto \int \psi^*_{E'} \left(\sum_{j\beta} \frac{\partial u_\alpha}{\partial x_{j\beta}}\bigg| x_{j\beta}\right) \psi_E \, dr \quad , \tag{10}$$

where $\psi_{E'}$ and ψ_E are vibrational eigenstates. The dipole strength of the ith eigenstate is

$$u'_\alpha \approx \sum_{j\beta} Q_j \, a^i{}_{j\beta} \quad , \tag{11}$$

where Q_j is the effective charge of atom j and is substituted for the derivative of u. The $a^i{}_{j\beta}$ term is the ith normalized eigenvector in direction ß for atom j. The Q values were taken from Mulliken-type[14] calculations made on the amorphous quartz system,[6] and were $Q_{Si} = 0.9$ and $Q_O = -0.5$ in units of electron charges. The electromagnetic energy absorbed in the

solid I is proportional to[11]

$$I(\omega) \propto \sum_i g(\omega_i) |u^i|^2 . \qquad (12)$$

In Eq. (12), $g(\omega_i)$ is the vibrational density of states for the system. The results of the calculations are given in the next section.

Results and Discussion

An example of the IR absorption of the low Al samples is shown in Fig. 1. There appear to be three main bands of absorption. These are approximately located at 450 cm^{-1}, 800 cm^{-1}, and 1100 cm^{-1}. The peaks within the two lower bands are rather sharp, whereas the band at 1100 cm^{-1} is broad and relatively featureless. The experimental IR results that we obtained were very similar to the various experimental work that has been published on this system over the years.[1-4] Typical peak locations for the low Al samples are as follows: 478, 536, 697, 790, and 807 cm^{-1}. The band at 1100 cm^{-1} is too broad for assigning a peak value.

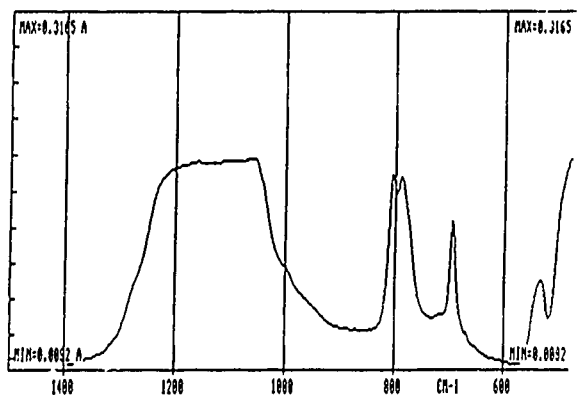


Fig. 1. Infrared absorption plot for low Al, Motorola AT-cut.

The IR absorbance plot in Fig. 2 is typical of the high Al samples tested. Due to the difficulty in sample preparation (chemical etching to less than 20 μm thickness), it was hard to achieve uniform thickness for different sample blanks. All of the high Al samples were slightly thicker than the low Al blanks; therefore, all of the peaks in the 450 cm^{-1} band were cut off. Also, all of the high Al samples had several sharp features superimposed on the 1100 cm^{-1} band. These sharp features also occurred in some of the low Al samples at the same wave numbers. We feel that these are possibly surface contaminants. The peak positions of the 800 cm^{-1} band (which, for the high Al samples were the only peaks that were unambiguously measurable) are identical to those found for the low Al samples. It was at this point that the computer models were brought into play to determine where to look for the Al peak.

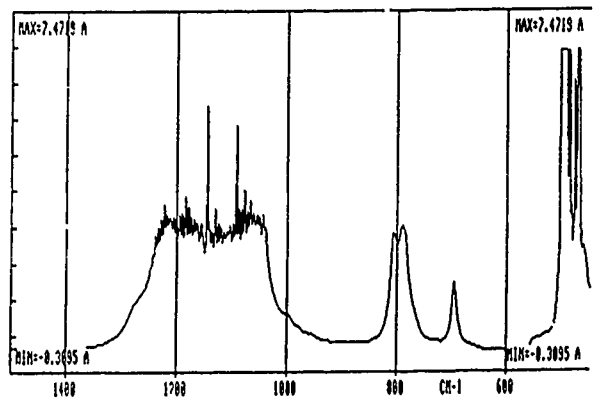


Fig. 2. Infrared absorption for high Al, Motorola At-cut.

The first stage of the vibrational calculations yields the fundamental frequencies. Since there are nine atoms in the alpha quartz unit cell there should be twenty-seven vibrational eigenvalues (fundamental modes). However, three of these are equal to zero and correspond to simple translational motion. Also, the remaining twenty-four non zero eigenvalues can be divided into three symmetry types.[3] There are four type A frequencies which are Raman active only (no IR absorption). There are four type B frequencies which are IR active only. The remaining sixteen frequencies are type E, Raman and IR active. However, these frequencies are degenerate which means that there are only eight different frequencies of this type. Our calculated frequencies, along with those experimentally determined by Spitzer and Kleinman[2] and Kats[3] are given in Table 2.

Table 2
Fundamental Frequencies (cm^{-1})

type	Spitzer& Kleinman	Kats	this work
A		207	221
Raman		357	368
active		467	527
		1082	1240
B	364	370	355
IR	495	49[illegible]	499
active	778	546	783
	1080	787	1243
E	128	127	133
Raman	256	266	276
and	394	404	407
IR	450	496	475
active	697	695	690
	797	787	813
	1072	1065	1133
	1163	1162	1237

For the type A symmetry, the fit of our results to those given by Kats is good for the two lowest frequencies; however, for the other two values, our frequencies are higher than Kats'. For the type B symmetry, all three sets of results agree for the two lowest frequencies, but the third value of Kats' differs not only from our calculated results but is different from the experimental value of Spitzer and Kleinman. For the highest frequencies, the results of all three sets disagree with each other. This emphasizes the problem of using experimental absorption results exclusively to determine the fundamental modes in quartz. There are more peaks in the absorption plots than fundamental modes to account for them. As a result, those working from the experimental data have had to guess which peaks were fundamental and which were combinational. In our work, we know that the results listed in Table 2 are the fundamental modes. However, at this point, it is unclear whether the potentials used are accurate enough that the frequencies obtained are correct. For type E symmetry, all three sets of results are in agreement, except for the two highest frequencies. For those two values, our results were somewhat higher than either Spitzer's or Kats'.

The fact that the model for perfect quartz could successfully reproduce reasonable results for the fundamental frequencies was no guarantee that the IR absorption could be modeled. In Fig. 3 the computed IR absorption for the perfect quartz model is shown superimposed on our experimental IR results for a low Al sample. The computed 450 cm^{-1} band has all the features of the experimental band, although the placement of the peaks is slightly shifted. The computed 800 cm^{-1} band, on the other hand, reproduces the experimental peak shape and peak position rather well. Finally, the 1100 cm^{-1} computed band is not a good representation of the experimental curve. It is much too sharp and the peak placement is not really central to the broad observed band; the reason for this is not clear. But on examination of the computed curve it seems that, if the small peak at approximately 1140 cm^{-1} were larger, the two peaks in this band might merge and better approximate the broad experimental band. The problem may simply be due to a deficiency in the potentials used.

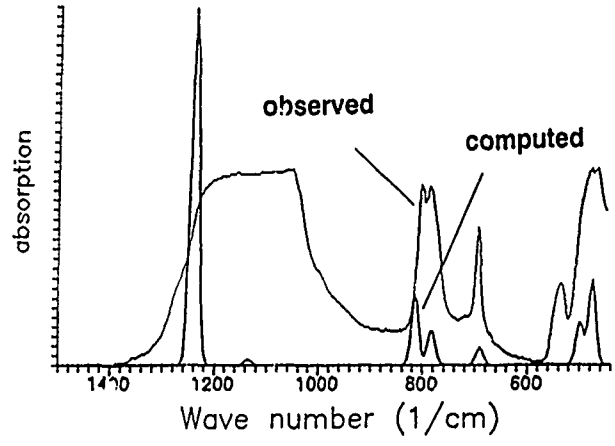


Fig. 3. Infrared absorption of computed perfect quartz model and observed low Al sample.

Given that our perfect quartz model could reproduce the observed IR spectra reasonably well, the next step was to do the same calculations for our Al defect model. As discussed in the previous section, the only difference between the perfect and defect models was that the mass at one of the Si sites was changed from that of Si to the mass of Al. All of the atomic potentials were unchanged. Since the unit cell is periodic, this represents a 33 percent Al to Si ratio in the model. The IR absorption of the Al defect model is shown in Fig. 4, where the IR plot looks remarkably like that of the perfect quartz model in Fig. 3. All of the peaks in the perfect model are present in the defect model, but on closer examination, the defect model peaks are shifted to higher wave numbers. We were very puzzled by this result, because we had anticipated that the Al atoms in the defect model would form a separate set of peaks that would be easy to see at this concentration. We examined the difference of the fundamental frequencies of the two models and found that the maximum shift occurred in the 800 cm^{-1} band.

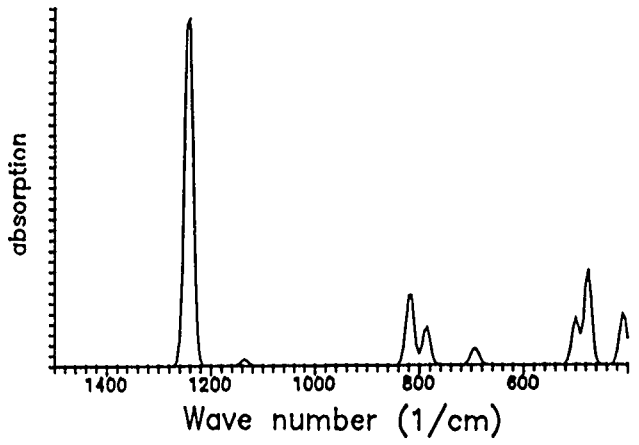


Fig. 4. Computed infrared absorption for Al defect model.

The defect frequency was approximately 6 cm^{-1} higher than the perfect model frequency. An enlarged section of the two computed IR plots is shown in Fig 5.

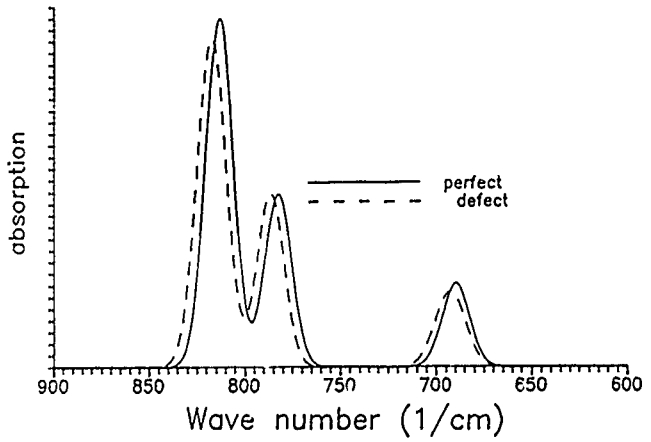


Fig. 5. Computed infrared absorption of 800 cm^{-1} bands in perfect and Al defect quartz models.

The question remains, is this Al model IR result real? Gervais et al.[15] conducted a systematic study of the IR spectra of the alumino-silicate glass system. They were specifically looking for the presence of an Al peak. Given the fact that the bond angles and bond lengths are similar in silicate glass and silicate crystals[6], the glass can be used to roughly approximate the crystalline behavior. Even at 40 percent Al, the Gervais group found no separate Al peak, but they did find that the peaks shift as the Al concentration is increased and that the largest shift is at the 800 cm^{-1} band. This supports our results for the Al defect model except that their peak shift with increasing Al was to lower wave numbers, whereas ours was to higher wave numbers. In our model this would mean that, in addition to the mass difference, the bonding potential of the Al would also be different (smaller) than the Si potential. They explained the suppression of the localized Al peak by considering that, like Si, the Al atom works as a network-former, and so they see similar bonding environments. This, coupled with the fairly small difference in the atomic mass of Al and Si, resulted in the perturbation of the existing fundamental frequencies instead of the formation of new Al modes.[15]

Since it appeared, from our model computations, that the most sensitive band to the effects of Al was the 800 cm^{-1}, we concentrated on just that region of the experimental curves. The IR absorption of the high and low Al samples in the 800 cm^{-1} band is plotted in Fig 6. We examined this set of data closely, but there was no discernable difference in the IR plots at the small Al concentrations in the samples. This means that, while higher Al concentrations of Al may be detectable in the far IR, this method is not viable as a test for Al in quartz for electronic applications.

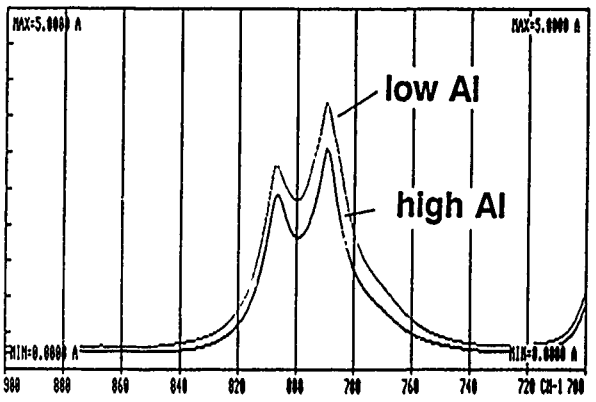


Fig. 6. Experimental infrared absorption of 800 cm^{-1} bands in high and low Motorola At-cut samples.

If higher Al concentrations can be seen in alpha quartz, it might be possible to develop the Al defect model into a more accurate tool. Aside from the glass results, other publications in the literature exist on the effects of Al on the physical properties of alpha quartz. For example, a Signal Corps report[16] from 1955 shows changes in the crystal lattice spacing of alpha quartz that correlate with Al concentration. More recent work by James[17] has also pointed to changes in the quartz elastic moduli, which can be attributed to Al. All in all, it appears there is enough data in the

literature to construct a better Al model in the near future.

Conclusions

Our original aim of using the fundamental modes to test for Al in electronic grade quartz does not seem to be possible. The lack of a separate peak that can be associated with Al and the low concentrations of Al found in the Z-growth areas of cultured quartz combine to put any Al-induced changes in the in the 0-1500 cm^{-1} region below the threshold of detection of our IR instruments. However, larger concentrations, such as those found in colored natural quartz or in the X-zones of cultured quartz, might be detectable. If the shifts that we found in the Al defect model could be seen in alpha quartz, it would provide crucial data needed to fine-tune the defect model and its interatomic potential. At the same time, other physical changes in the quartz due to Al content, such as lattice spacing and elastic constants,[16, 17] can also be analyzed to help refine the model.

On the brighter side, our model seemed to be able to accurately determine the fundamental modes in alpha quartz. In fact, only the highest eigenfrequencies were out of line with the values obtained from experimental studies.[2, 3] If further studies can make some slight improvements in this model, the discrepancies found in the experimental results can be resolved and the exact fundamental modes determined. The determination of the fundamental modes by model calculations has the additional advantage that the eigenvectors are automatically generated, consequently, the atomic movement of each of t e modes can be simulated and studied. This is not possible in the experimental work. In studying the fundamental modes we are attempting to pin down the most basic vibration signature of this solid. If we can accurately generate vibrations at this level, it may be possible to model the macroscopic changes (elastic moduli, Q, etc.) resulting from the incorporation of defects.

Acknowledgments

The authors thank Dr. J. Balascio of Motorola for his help in providing the samples used in this investigation, and Mr. Atwater, of our Laboratory for the ICP determinations of the samples. We also thank Dr. Halliburton of Oklahoma State University for his discussions on the ESR method for determining the Al content of quartz.

References

[1] B. D. Saksena, "The Infra-Red Absorption Spectra of α-Quartz Between 4 and 15 Microns," Proc. Phys. Soc. London, vol. 72, pp. 9-16, Feb. 1958.

[2] W. G. Spitzer and D. A. Kleinman, "Infrared Lattice Bands of Quartz," Phys. Rev., vol. 121, pp. 1324-1335, March 1961.

[3] A. Kats, "Hydrogen in Alpha-Quartz," Philips Research Reports, vol. 17, pp. 201-279, June 1962.

[4] J. N. Plendl, L. C. Mansur, A. Hadni, F. Brehat, P. Henry, G. Morlot, F. Naudin, and P. Strimer, "Low Temperature Far Infrared Spectra of SiO_2 Polymorphs," J. Phys. Chem. Solids, vol. 28, pp. 1589-1597, 1967.

[5] H. E. Swanson, R. K. Fuyet, and G. M. Ugrinic, NBS Circular 539, Vol. III.. 1954, pp. 24-26.

[6] R. A. Murray and W. Y. Ching, "Electronic and Vibrational Structure Calculations in Models of the Compressed SiO_2 Glass System," Phys. Rev. B, vol. 39, pp. 1320-1331, Jan. 1989.

[7] R. J. Bell and P. Dean, Nature, vol. 212, p. 1354, 1966.

[8] R. J. Bell and P. Dean, Discuss. Faraday Soc.,vol. 50, p. 55, 1970.

[9] R. J. Bell and D. C. Hibbing-Butler, J. Phys. C, vol. 9, p. 1171, 1976.

[10] R. J. Bell, A. Carnnevale, C. R. Kurkjian, and G. E. Peterson,J. Non-Cryst. Solids, vol. 35-36, p. 1185 1980.

[11] R. J. Bell, Methods in Computational Physics, Vol. 15. New York: Academic, 1976, p. 215.

[12] D. Steele, Theory of Vibrational Spectroscopy. Philadelphia: Saunders, 1971, p. 114.

[13] E. B. Wilson, Jr., J. C. Decius, and P. C. Cross, Molecular Vibrations., New York: McGraw-Hill, 1955, p. 289.

[14] R. S. Mulliken, J. Chem. Phys., vol. 23, p. 1833, 1955.

[15] F. Gervais, A. Blin, D. Massiot, J. P. Coutures, M. H. Chopinet, F. Naudin, J. "Infrared Reflectivity Spectroscopy of

Silicate Glasses," J. Non-Cryst. Solids, vol. 89, pp. 384-401, 1987.

[16] D. L. Hammond, A. R. Chi, and J. M. Stanley, Signal Corps Engineering Report E-1162. Fort Monmouth: Signal Corps Engineering Labs., pp. 1-8, 1955.

[17] B. J. James, "A New Measurement of the Basic Elastic and Dielectric Constants of Quartz," in Proc. of the 42nd Annual Freq. Cont. Symp., pp. 146-154, 1988.

43rd Annual Symposium on Frequency Control - 1989

DETERMINATION OF THE THIRD-ORDER PIEZOELECTRIC CONSTANTS AND ELECTROSTRICTION OF ALPHA QUARTZ USING THE THICKNESS MODES OF PLATES

Carl K. Hruska and Remi Brendel*

Piezoelectricity Research Laboratory
Department of Mathematics, York University
Toronto, Ontario M3J 1P3, Canada

*Laboratoire de Physique et Metrologie
des Oscillateurs du C.N.R.S.
associe a l'Universite de Franche-Comte
25000 Besancon, France

Summary

Ten linear combinations of the fundamental material constants of third-order piezoelectricity and electrostriction of quartz have been determined; they were obtained by the least-squares fit to over one hundred observations of the change in the resonance frequency of the thickness modes of quartz plates caused by a dc field (the resonator method).

With one exception, there is a very good agreement between our results and their counterparts obtained independently by the transit-time method (pulse propagation in bulk quartz); the differences between comparable constants are no longer out of proportion with their standard errors.

Introduction

Interactions between the dc electric field and quartz are a source of valuable information about the material nonlinearities of quartz.

Typically, the interactions are observed by the resonator method or by the transit-time method. The former consists in changes in the resonance frequency of quartz resonators, the latter in the variations in the transit time of acoustic pulses propagating through bulk quartz. Both changes are induced by a dc electric field acting on quartz material.

Large bodies of experimental data have been collected as part of the effort to study the nonlinear piezoelectricity (or electroelasticity) and electrostriction which are the subject of this work. An interesting and potentially fruitful objective is to show that their reduction leads to mutually consistent and harmonious sets of nonlinear material constants. Such a favourable result would offer a reassurance concerning the present values of these material constants and reflect positively on the existing nonlinear theory of piezoelectrics and its applications.

The present values of the third-order piezoelectric and electrostrictive constants of quartz have been determined using the transit-time method [1]. This paper aims at an independent determination of the same nonlinear material constants using the largest body of data ever assembled for this purpose and obtained by the resonator method using the thickness vibrations of plates.

The most recent attempt [2] at the solution of the problem dealt only with the electroelastic constants, it successfully eliminated large discrepancies of several orders of magnitude; otherwise the agreement reached between the transit-time method and the resonator method was substantially poorer than the one attained earlier [3,4,5] for the effective electroelastic constants. As the opposite should be true on theoretical grounds, such results were of little credit to the nonlinear theory.

This work builds on the results of [6,7,8,5]. Its advantages over the previous work [2] are the incorporation of electrostriction and the calculation of the pre-strain due to the dc field bias altered to correspond better to the conditions of the analyzed experiment.

The numerical values of all quantities (generated or referenced) in this paper are stated for right-hand quartz and the frame of reference according to the IRE 1949 Standard [9]. The third-order piezoelectric and electrostrictive constants calculated here are valid at room temperature.

Experimental data used

The determination of the nonlinear constants of quartz by means of the resonator method is based on the observations of the change Δf of the resonance frequency f of piezoelectric resonators on the dc electric field ΔE applied to their body. This phenomenon has also been known as the polarizing effect or the electroelastic effect.

For any quartz resonator the above dependence can be described in terms of its respective linear and quadratic coefficients, $L = (1/f).(df/dE)_{E=0}$ and $Q = (1/2f).(d^2f/dE^2)_{E=0}$, and can be recorded as

$$\Delta f/f = L \,.\, \Delta E + Q \,.\, (\Delta E)^2$$

It is the linear coefficient L of this dependence that is used for the calculation of the nonlinear constants of quartz. It is known to depend primarily on the crystallographic orientation of the resonator, on its mode of vibration and on the direction of the applied dc field relative to the resonator.

The data for this work were collected using the fundamental frequency of the thickness longitudinal mode and the two thickness shear modes of fifty-six plates made of natural quartz. The orientations of the plates were chosen to cover the primitive region of quartz as effectively as possible even including its parts hitherto unexplored. The frequency of the resonators was between 2.6 MHz and 6.5 MHz. The dc electric field was always applied in the thickness direction of the plates. The frequency and its changes were measured under the condition of zero phase shift.

CH2690-6/89/0000-485 $1.00 C 1989 IEEE

The measurements were conducted at a temperature of 25 °C. A detailed description of the experiments can be found in [5,10].

One hundred and thirty-four values of the linear coefficient L were collected, ranging in magnitude between 0.07×10^{-12} and 17.04×10^{-12} m/V. The bulk of the experimental values used here is listed in [5] with an additional twelve values published in [2].

It appears that the largest source of errors in L are unexplained variations in this quantity among physically different samples of sufficiently similar orientation to be expected to yield the same value of L. A proper statistical analysis of the experimental errors in L would be a monumental task exceeding the material resources usually available for this type of work. Simplifying assumptions that have been made lead to a global estimate of the standard deviation of L equal to $\pm\ 0.30 \times 10^{-12}$ m/V.

The linear coefficient L and the material constants of quartz

The measured linear coefficient of the frequency-dc field dependence L is equivalent to the quantity $\Delta\omega/(\omega.E)_H$ in Eq.(92) of [7]. This equation presents the linear coefficient L as a function of the crystallographic orientation of the plate, mode of its thickness vibrations, and, of the linear and nonlinear material constants of quartz It is the current model of the linear coefficient L in the nonlinear theory of piezoelectric materials.

Three alterations are made in this model before it is used in this paper. The algebra of the modifications is not shown as it has nothing interesting to offer; but the changes made and their resulting effect are described in full detail.

Alteration 1. The representation of electrostriction. The electrostrictive constants e_{ijkl} in Eqs. (30) and (31) of [7] have been replaced by the electrostrictive constants l_{ijkl} according to the equation (Eq.(54), [11])

$$e_{ijkl} = \varepsilon_o l^{rel}_{ijkl} = \varepsilon_o l_{ijkl} + \varepsilon_o (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \delta_{ij}\delta_{kl}),$$

where i,j,k,l = 1,2,3, and where ε_o is the dielectric permittivity of vacuum and δ_{ik} is the Kronecker delta. The introduction of l_{ijkl} to replace e_{ijkl} generally simplifies the record of algebraic expressions (eg. Eqs.(31) and (32), [7]); later, when these electrostrictive constants and the third-order piezoelectric constants form combinations both types of the constants have a conveniently similar order of magnitude (Tabs. I and II).

Alteration 2. Calculation of the static strain. The development of the model of L requires the calculation of static displacement gradients in the resonator plate due to a dc electric field applied in the plate thickness.

In [7] this calculation is done under the conditions analogous to those used in the dynamic part of the problem and corresponding to a simplified treatment of the resonator as an infinite plate.

For the static case this simplification is not needed. As a result, the static displacement gradients are calculated in this paper as

$$\bar{u}'_{m,n} = d'_{2mn} E'_2, \qquad m,n = 1,2,3,$$

reflecting the boundary condition that the whole surface of the plate is free of traction. Here $\bar{u}'_{m,n}$ represents the displacement gradient, d'_{2mn} the piezoelectric strain constants and E'_2 the dc field intensity in the plate thickness. The prime indicates that these quantities are related to the reference frame of the plate as defined in [7]. This alteration makes Eqs.(81), (82), (87) and (88) in [7] redundant.

Alteration 3. Inclusion of electrostriction in the model.
Although the electrostrictive constants were duly included during its development stages, they have been dropped from the final formulation of the model [7] as their importance was secondary to the third-order piezoelectric constants (the electroelastic constants) which were retained. Aiming now at a more refined model of the linear coefficient L, this omission is being corrected by adding the term

$$G = (1/2\lambda)\, G_{ijkl}\, l_{ijkl}$$

Incorporating the described changes into [7] while progressing from Eqs.(31) and (32) on towards Eq.(92), an improved model of the linear coefficient L is obtained and is adopted in this paper

$$(1)\quad L = (1/2\lambda)\,(M + N + P + R) + (1/2\lambda)\,(F + G) .$$

Whereas the symbols used in this section for the quantities of the same physical meaning are those used in [7], their definitions have in many cases changed in consequence of the above alterations. In order to make sure that no change is overlooked, and to make the use of Eq.(1) possible without further references to [7], all symbols appearing in Eq.(1) are defined anew

$$(2)\quad M = n_i n_e (n_j \delta_{km}\delta_{ln} + n_l \delta_{kn} l_j l_m + n_l \delta_{jn} l_k l_m) d_{emn} c_{ijkl} ,$$

$$(3)\quad N = n_i n_l n_r l_j l_k d_{rmn} c_{ijklmn} ,$$

$$(4)\quad P = n_j n_l (2 r n_r l_i d_{rik} - n_k) e_{jkl} ,$$

$$(5)\quad R = - r^2 n_i n_j n_k \kappa_{ijk} ,$$

$$(6)\quad F = n_i (2 r n_k n_r l_j d_{rlm} - n_j n_m n_k l_l) e_{ijklm} ,$$

$$(7)\quad G = \varepsilon_o n_i n_j (-r^2 n_r d_{rkl} + 2 r n_l l_k) l_{ijkl} ,$$

$$(8)\quad r = n_i n_k l_j e_{ijk} / (n_a n_b \varepsilon_{ab}) ,$$

The quantities on the right-hand side of Eqs.(2)-(8) are all defined in the basic frame of reference in quartz according to the IRE Standard of 1949 [9]. In the order of their appearance above they are:

- n_i - the direction cosines of the thickness of the resonator plate,
- δ_{km} - the Kronecker delta,
- l_j - components of the amplitude vector of the plate vibrations,
- d_{emn} - the piezoelectric strain constants,
- c_{ijkl} - the elastic stiffnesses,
- c_{ijklmn} - the (nonlinear) third-order elastic constants,

e_{jkl} - the piezoelectric stress constants,

κ_{ijk} - the (nonlinear) third-order dielectric permittivities,

e_{ijklm} - the (nonlinear) third-order piezoelectric constants,

ε_0 - the permittivity of free space,

l_{ijkl} - the electrostrictive constants, and

ε_{ab} - the dielectric permittivities.

The quantity λ in Eq.(1) and the amplitude vector (l_1, l_2, l_3) are the respective eigenvalue and eigenvector of the matrix

$$\begin{bmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{21} & \Gamma_{22} & \Gamma_{23} \\ \Gamma_{31} & \Gamma_{32} & \Gamma_{33} \end{bmatrix}$$

where

$$(9) \quad \Gamma_{ik} = n_j n_l c_{ijkl} + n_n n_j n_m n_l e_{nij} e_{mkl} / (n_a n_b \varepsilon_{ab}).$$

The three generally existing eigenvalue-eigenvector pairs of the above matrix correspond to the three thickness modes of vibrations of the plates under consideration.

All indices in Eqs.(1)-(9) take on the values of 1,2,3. Einstein summation applies to all pairs of repeating indices.

Calculation of the third-order piezoelectric and electrostrictive constants

$$(10) \quad L = C + (1/2\lambda) F_{ijklm} e_{ijklm} + (1/2\lambda) G_{ijkl} l_{ijkl} ,$$

$$i,j,k,l,m = 1,2,3 ,$$

where $C = (1/2\lambda) (M + N + P + R)$. Both C and the coefficients $(1/2\lambda) F_{ijklm}$ and $(1/2\lambda) G_{ijkl}$ are easily calculable for a plate of any orientation and any mode of vibrations.

The material constants of quartz necessary for their actual evaluation - the linear material constants, the third-order elastic constants and the third-order permittivity - are taken from [12,13,14].

The only unknown quantities left are the third-order piezoelectric constants e_{ijklm} and the electrostrictive constants l_{ijkl}: Eq.(10) represents a linear equation for their determination. Before this is done Eq.(10) should be simplified by making the usual transition from the tensor index notation to the matrix notation, by taking into account the interchange symmetry among pairs of indices and the crystal symmetry of quartz, and by retaining and subsequent factoring out only the traditional independent material constants of third-order piezoelectricity and electrostriction. The simplified equation is

$$\begin{aligned} (11) \quad Y_i &= L_i - C_i \\ &= + a_{11} \cdot e_{111} + a_{12} \cdot e_{113} + a_{13} \cdot e_{114} \\ &\quad + a_{14} \cdot e_{122} + a_{15} \cdot e_{124} + a_{16} \cdot e_{134} \\ &\quad + a_{17} \cdot e_{144} + a_{18} \cdot e_{315} \\ &\quad + b_{11} \cdot l_{11} + b_{12} \cdot l_{12} + b_{13} \cdot l_{13} \\ &\quad + b_{14} \cdot l_{14} + b_{15} \cdot l_{31} + b_{16} \cdot l_{33} \\ &\quad + b_{17} \cdot l_{41} + b_{18} \cdot l_{44} , \end{aligned}$$

where

$$(12) \quad e_{111}, e_{113}, e_{114}, e_{122}, e_{124}, e_{134}, e_{144}, e_{315}$$

are the eight independent third-order piezoelectricity constants and

$$(13) \quad l_{11}, l_{12}, l_{13}, l_{14}, l_{31}, l_{33}, l_{41}, l_{44}$$

are the eight independent electrostrictive constants; all sixteen constants related to the basic reference frame of alpha-quartz [9].

The index i in Eq.(11) has been added to all relevant terms to indicate their dependence on the crystallographic orientation of the plate and its mode of vibration. The details of the transition from Eq.(10) to Eq.(11) are not shown as the process is uneventful, tedious and well known.

Eq.(11) can be now applied to all one hundred and thirty-four available observations of the linear coefficient L_i ($i = 1,2,\ldots,134$) of the frequency-dc field dependence.

The obtained system of one hundred and thirty-four linear equations is overdetermined and its solution set

(14)

$$e_{111}, e_{113}, e_{114}, e_{122}, e_{124}, e_{134}, e_{144}, e_{315},$$
$$l_{11}, l_{13}, l_{14}, l_{31}, l_{33}, l_{41}, l_{44}$$

can only be estimated; this has been attempted using the least-squares method.

It should be noted that solution set (14) consists of only fifteen nonlinear material constants. It has been found that in the entire linear system the coefficient $b_{12} = 0$ ($i = 1,2,\ldots,134$). As a result the electrostrictive constant l_{12} cannot be determined from our linear system and must be dropped from solution set (14).

The attempt to determine the least-squares estimates of the fifteen nonlinear constants (14) was unsuccessful. It was found that the matrix of the linear system which consists of two concatenated matrices, a_{ij} and b_{ik}, ($i = 1,2,\ldots,134$, $j = 1,2,\ldots,8$, $k = 1,3,4,\ldots,8$) is heavily plagued by near multicolinearity. This phenomenon is generally known to severely inconvenience the algebraic operations necessary to perform the least-squares fit and to be the cause of heavily distorted solutions [8,15,16].

The cure of the problem has been described earlier [16]. Here it leads to the following transformation of linear system (11)

$$\begin{aligned} (15) \quad Y_i &= L_i - C_i \\ &= + a_{11} \cdot c_{111} + a_{13} \cdot c_{114} + a_{15} \cdot c_{124} \\ &\quad + a_{17} \cdot c_{144} + a_{18} \cdot c_{315} \\ &\quad + b_{13} \cdot c_{13} + b_{14} \cdot c_{14} + b_{16} \cdot c_{33} \\ &\quad + b_{17} \cdot c_{41} + b_{18} \cdot c_{44} , \end{aligned}$$

where $i = 1,2,\ldots,134$. The least-squares process will ultimately yield reliable estimates for a new, transformed solution set

(16) c_{111}, c_{114}, c_{124}, c_{144}, c_{315},
c_{13}, c_{14}, c_{33}, c_{41}, c_{44} ,

consisting of only ten elements, each being a linear combination of the original material constants (14).

Results obtained

The calculable combinations (16) of the third-order piezoelectric constants and electrostrictive constants are listed in Table I.

$$
\begin{aligned}
c_{111} &= e_{111} + 0.079\, l_{11} - 0.012\, e_{122} \\
c_{114} &= e_{114} + 0.006\, l_{11} - 0.170\, e_{113} \\
&\quad - 0.190\, e_{122} - 0.715\, e_{134} \\
c_{124} &= e_{124} + 0.046\, l_{11} - 0.170\, e_{113} \\
&\quad - 0.700\, e_{122} - 0.717\, e_{134} \\
c_{144} &= e_{144} - 0.040\, l_{11} + 0.506\, e_{122} \\
c_{315} &= e_{315} - 0.052\, l_{11} - 0.129\, e_{113} \\
&\quad + 0.666\, e_{122} - 0.543\, e_{134} \\
c_{13} &= l_{13} + 2.012\, l_{11} - 1.000\, l_{31} \\
&\quad +27.043\, e_{113} - 12.168\, e_{122} \\
&\quad + 6.940\, e_{134} \\
c_{14} &= l_{14} - 0.098\, e_{122} \\
c_{33} &= l_{33} + 0.981\, l_{11} + 28.842\, e_{113} \\
&\quad - 12.444\, e_{122} + 7.341\, e_{134} \\
c_{41} &= l_{41} - 0.019\, l_{11} + 0.220\, e_{122} \\
&\quad - 0.124\, e_{134} \\
c_{44} &= l_{44} - 0.550\, l_{11} + 0.500\, l_{31} \\
&\quad + 0.279\, e_{122} - 0.145\, e_{134}
\end{aligned}
$$

TABLE I. Calculated combinations of the fundamental third-order piezoelectric and electrostrictive constants. The numerical coefficients are rounded off to three decimal places.

Their calculated values together with the standard errors are shown in the second column of Table II. They are given for right-hand quartz and the frame of reference according to the IRE 1949 Standard [9].

The experimental values and material constants needed for the preceeding calculations represent a mixture of quantities referred to 20 or 25 °C or just room temperature. Consequently, the third-order piezoelectric and electrostrictive constants calculated here are understood as valid for room temperature. Considering the existing accuracy of the determined material constants and their temperature dependence the temperature inconsistency of several degrees C is of no practical consequence.

The order of magnitude of the combinations c_{ijk} and c_{ij} in Table II is consistent with that of the other quantities in the system (15) which is typically between 10^{-11} and 10^{-12} m/V, 10^{-12} and 10^{-13} m^2/N, and between 10^{-13} and 10^{-15} m/V, for Y_i a_{ij} and b_{ik}, respectively. The standard errors in the combinations c_{ijk} and c_{ij} are also in harmony with the standard errors in the measured quantity L_i (globally estimated to be $\pm$ 0.30 x 10^{-12} m/V). As the principal data used in this work originate from resonance frequency observations, the method of determination of the nonlinear constants in this paper is called the resonator method.

component combination	this work	Kittinger et al. original	Kittinger et al. recalculated
c_{111}	1.96 ± 0.04	1.81	1.81 ± 0.03
c_{114}	1.33 ± 0.07	1.15	1.20 ± 0.08
c_{124}	1.22 ± 0.08	1.03	1.06 ± 0.08
c_{144}	-0.47 ± 0.03	-0.32	-0.30 ± 0.04
c_{315}	-0.41 ± 0.06	-0.43	-0.41 ± 0.08
c_{13}	-11.2 ± 4.5	-11.7	-12.6 ± 4.2
c_{14}	2.3 ± 0.5	2.3	2.2 ± 1.4
c_{33}	-24.3 ± 7.9	-21.3	-22.6 ± 6.2
c_{41}	4.7 ± 0.7	4.2	3.6 ± 2.4
c_{44}	-4.9 ± 1.7	4.3	4.8 ± 1.7

TABLE II. Combinations of the third-order piezoelectric and electrostrictive constants. The errors are standard errors. Values obtained independently by Kittinger, Tichy and Friedel [1] - both original and recalculated in this work - are added for comparison. The values of c_{ijk} are in N/(V.m); c_{ij} are dimensionless. Given for room temperature, right-hand quartz and the frame of reference according the IRE Standard of 1949 [9].

In 1986 Kittinger, Tichy and Friedel [1] published a complete set of sixteen values for the third-order piezoelectric and electrostrictive constants (12)-(13). They were determined from the dc field-induced variations in the transit time of pulses propagating in bulk quartz (the transit-time method). Their values were used to produce the combinations defined in Table I which are placed in the third column of Table II next to the results of this work for comparison. Unfortunately, no similarly simple procedure exists to provide the corresponding standard errors.

To obtain the proper standard errors the work by Kittinger et al. [1] had to be recomputed starting from the raw data. This has been done and the results - now with their standard errors - are placed in the last column of Table II.

The raw data by Kittinger et al. [1] were treated here in a manner consistent with all other calculations done in this work which do not fully agree with the

original work [1] This and possible numerical reasons account for the differences between the third and the fourth columns in Table II.

The differences that exist among authors in how they execute the data reduction process undoubtedly deserve some attention. Depending on the size of the standard errors they may cause statistically significant and unnecessary contradictions. However, the question is beyond the scope of this paper and will have to be addressed later.

Conclusion

This analysis of the earlier observations of the linear coefficient of the frequency-dc field dependence resulted in the determination of ten values of linear combinations of the third-order piezoelectric and electrostrictive constants of quartz.

With the exception of one case, where we find a significant discord in sign, the agreement between this work (the resonator method) and [1] (the transit-time method) is spectacularly good. Moreover, it is the first time for these material constants and the nonlinear theory that the differences between the two methods become comparable with the size of the calculated standard errors.

The participation of the individual material constants in the combinations calculated in this paper may be of different magnitude. Consequently, an agreement attained for a particular combination should not be understood as a universal confirmation of the value of all the participants in it as substituted from [1].

The reason for the improved agreement between the results of the transit-time and the resonator method - quite a dramatic one when compared with the previous attempt [2] - is to be seen in the incorporation of electrostriction into the model used but also, and more importantly, in the employed method of calculation of the static strain of the resonators due to the dc field.

Acknowledgements

The authors are obliged to Miss P. Hruska for her assistance with the final stages of preparation of this paper. This work was supported by the York University-Faculty of Arts Fellowship and a Minor Research Grant (1988) and the National Sciences and Engineering Research Council of Canada.

References

[1] Kittinger E., Tichy J., and Friedel W., "Nonlinear piezoelectricity and electrostriction of alpha quartz". J. Appl. Phys. 60(1986), pp. 1465-1471.

[2] Hruska C. K, and Brendel R., "The electroelastic constants of quartz determined by the resonator method", J. Appl. Phys. 65(1989), pp. 715-717.

[3] Reider G. A. Kittinger E., Tichy J. , "Electroelastic effect in alpha quartz', J. Appl. Phys. 53(1982), pp. 8716-8721.

[4] Kittinger E., private communication.

[5] Hruska C. K., Merigoux H., Kucera M., "Linear frequency-dc field effect with the thickness modes of alpha-quartz plates", Journal of the Canadian Ceramic Society 57(1988), pp.53-57.

[6] Baumhauer J. C., Tiersten H. F., Nonlinear electroelastic equations for small fields superimposed on a bias", J. Acoust. Soc. Am. 54(1973), pp. 1017-1034.

[7] Brendel R., "Material nonlinear piezoelectric coefficients for quartz", J.Appl.Phys. 54(1983), pp. 5339-5346.

[8] C. K. Hruska, "Independent verification of the values of electroelastic constants of alpha quartz", J. Appl. Phys. 61(1987), pp. 1127-1129.

[9] IRE Standards on Piezoelectric Crystals, IEEE Standard 176; see also Proc. IRE 37(1949), pp. 1378-1395.

[10] Hruska C. K., Kucera M., "The dependence of the polarizing effect with the thickness modes of alpha-quartz plates on their resonance frequency", Journal of the Canadian Ceramic Society 55(1986), pp. 31-41.

[11] son D. F., "Theory of nonlinear electroacoustics of dielectric, piezoelectric, and pyroelectric crystals", J.Acoust. Soc. Am. 63(1978), pp. 1738-1748.

[12] Bechmann R., "Elastic and piezoelectric constants of alpha-quartz", Phys. Rev. 110(1958), pp. 1060-1061.

[13] Thurston R.N., McSkimmin H.J. and Andreatch, "Third-order elastic coefficients of quartz", J. Appl. Phys. 37(1966), pp. 267-275.

[14] Besson R. and Gagnepain J.-J., "Determination des coefficientes non lineaires de polarisation electrique du quartz", C. R. Acad. Sci. Paris B 274(1972), pp. 835-838.

[15] Mendenhall W. and Sincich T., "A second course in business statistics: Regression analysis", Dellen Publishing Company, San Francisco, Collier Macmillan Publishers, London, 2nd edition, 1986, p. 168, pp. 222-223, p. 228, pp. 234-235.

[16] Hruska C. K., "On the disagreement in the order of magnitude of the electroelastic constants of alpha-quartz", Proc. 40th Annual Frequency Control Symposium, U.S.Army Electronics Command, Ft. Monmouth, May 1986, pp. 54-62.

43rd Annual Symposium on Frequency Control - 1989

DEFECT DISTRIBUTION MAPPING IN QUARTZ

Pierre Zecchini* and Herbert G. Lipson**
* Université de Franche-Comté, Laboratoire de Cristallographie et Chimie Minérale, 25030 Besancon CEDEX France
** Rome Air Development Center, Solid State Sciences Laboratory, Hanscom AFB, MA 01731-5000

SUMMARY

Infrared and 632.8 nm laser absorption measurements have been used to characterize changes in aluminum compensation for irradiated and air-swept cultured and natural quartz. These measurements demonstrated irradiation dose effects and the competition between Al-OH and Al-h (aluminum-hole) compensation mechanisms. The laser technique was also used to localize mechanical and growth defects. By combining the results of X- and Y-direction observations for a given XY plane, information on both bulk growth defects and surface imperfections from polishing and other mechanical operations can be obtained. A complete map of defects can be obtained from measurements made over a range of Z-positions. The effect of ^{60}Co gamma irradiation, air sweeping, re-irradiation and annealing on grown-in OH, Al-OH and Al-h has been shown. While localized Al-h defects produced by irradiation were no longer observable after air sweeping, re-irradiation introduced a small 632.8 nm absorption over the entire crystal. A large decrease in 632.8 nm absorption occurs from annealing at 250°C, a temperature well below that of Al-h removal. These absorption changes may arise from another defect possibly associated with another trivalent impurity.

INTRODUCTION

Impurity-associated defects [1] can introduce detrimental frequency offsets [2-4] in resonators fabricated from cultured quartz when used in a radiation environment. It is therefore important to establish convenient non-destructive techniques to characterize the effect of radiation and electrodiffusion (sweeping) on these defects. The principal impurities introduced during hydrothermal growth of cultured quartz are an aluminum-metal center, Al-M, formed when an aluminum substitutional in a silicon site is compensated with an alkali ion (Li,Na) and grown-in OH, a hydrogen atom adjacent to an oxygen site. Germanium substitutional for silicon is also found in quartz as well as the possibility of other trivalent ions in silicon sites. Ionizing radiation dissociates Al-M and OH and forms either Al-OH or Al-h, a hole trapped at a non-bonding oxygen ion, depending on the amount of hydrogen available in the crystal. It is also desirable to have a convenient method to locate growth defects such as bubbles, pipes and cracks which introduce problems in the fabrication and electroding of resonators.

Experimental techniques used to characterize the various defects in quartz [5-11] are usually applicable to a specific section of quartz bar, but are not adapted to conveniently determine variations across the bar. The established commercial method for cultured quartz quality evaluation and infrared Q determination [12] uses scans across the crystal at specific room temperature OH absorption regions or peaks. While grown-in OH distributions can be specified from this method, low temperature infrared measurements provide improved sensitivity for determining OH and Al-OH distributions before and after irradiation and sweeping. We have developed low temperature infrared scanning techniques for monitoring grown-in OH and Al-OH over an entire bar [13,14]. More recently absorption at the 632.8 nm He-Ne laser wavelength has been used to measure Al-h concentration and irradiation dose dependence for localized regions over an entire quartz crystal [15]. A 632.8 nm absorption coefficient calibration of 0.1 cm^{-1} / ppm Al-h [16] was determined from absorption and electron spin resonance measurements (ESR) made on the same crystal section.

In this paper we apply both the low temperature infrared and He-Ne laser absorption techniques to investigate changes in the distribution of OH, Al-OH and Al-h produced by irradiation, air sweeping, re-irradiation and annealing. Multi-directional He-Ne laser scans are also used to localize bulk mechanical defects and separate them from surface imperfections.

EXPERIMENTAL PROCEDURES

The origin, type, aluminum concentration, and treatment of quartz samples used for these investigations are listed in Table 1. The samples were rectangular in shape, between 1 and 2 cm in size with faces normal to the X-, Y-, and Z-axis directions (IEEE convention). All faces were polished flat and parallel to within 10 seconds of arc or better.

Table 1. Quartz samples.

sample	type	Al ppm	treatment
E42-21	SARP High-Q	6-8 (*)	accumulated irradiation dose (krad): 20, 50, 80, 100, 200, 250, 500, 1000; dose rate: 0.5 krad/min.
BH-A 1	SARP Premium-Q	3.7-7.6 (*)	accum. dose (krad): 50, 60 80, 100, 150, 200, 500, 1000; dose rate: 0.5 krad/min; annealed at 370 C for 4 h.
H33-31	SARP Premium-Q	3 (**)	accum. dose (krad): 200, 400, 800, 1600, 2600, 5000; dose rates (krad/min): 10 up through 1600 krad, 50 for 2600 and 5000 krad; swept in air 14 days; re-irradiated with 6000 krad at 10 krad/min; annealed at 250°C for 1 h; annealed at 370°C for 2 h.
NB1	natural		irradiated with 2000 krad.

SARP: Sawyer Research Products (*) ESR measurement (**) ICP measurement

Infrared spectra were obtained between 3100 and 3700 cm^{-1} using a Nicolet 170SX Fourier spectrometer. The sample mounted inside a Dewar cooled to 85 K was scanned normal to the Z-axis between the +Z and -Z faces. Measurements were made with a 3 mm diameter unpolarized infrared beam parallel to the Y-axis, at one or two beam positions along the X-axis. The peak absorptions of the strong narrow 3581 cm^{-1} and 3366 cm^{-1} bands are used to monitor relative changes in grown-in OH and irradiation- or sweeping-produced Al-OH, respectively.

The experimental laser absorption set-up consisted of a randomly polarized 2 mW He-Ne laser, a sample holder and a silicon detector. The laser beam width was 0.67 mm at 1/e^2 with a divergence of 1.2×10^{-3} radians. The sample was mounted on a linear x,y positioning stage driven by a 1 micron/step motor. The positioning stage was actuated by an Aerotech Unidex IIIA motion controller operated from Hewlett-Packard series 310 computer The transmitted laser beam was measured with a calibrated E.G. and G. Gamma Scientific silicon photovoltaic detector with a circular sensitive area of 1.0 cm^2 and a surface uniformity of better than 1%. The power meter had a range of 10^{-8} to 10^{-2} watts with a readout accuracy of 1% and stability and repeatability of better than 1%. Absorption values were calculated from transmitted power I/I_0 ratios using the crystal reflectivity at the

CH2690-9/89/0000-490 $1.00 C 1989 IEEE

laser wavelength. The sample absorption was measured with the laser beam aligned in one of the two equivalent optical directions, Y or X, by scanning along the Z-axis at different positions along the X- or Y- directions, as shown in Figures 1a and 1b. Absorption was also measured in the Y- and X- directions scanning along the X- or Y-axis at different positions along the Z-axis as shown in Figures 1c and 1d.

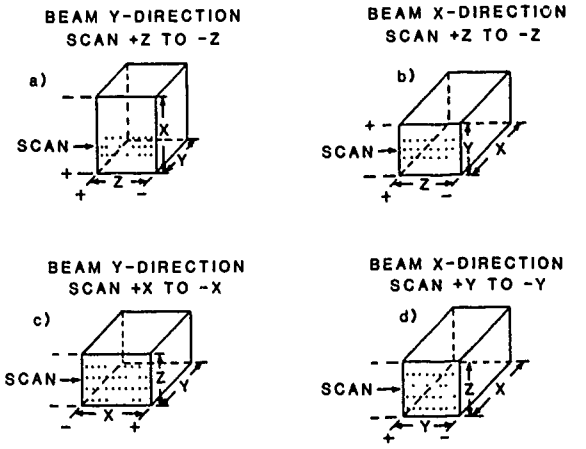


Fig. 1 Laser beam directions and scans relative to quartz orientation.

The samples were irradiated at room temperature with the Rome Air Development Center (RADC) ^{60}Co source using successive doses to obtain the total doses of Table 1 for samples E42-21, BH-A 1 and H33-31. Natural crystal N1B had a single 2000 krad dose. Sample H33-31 was re-irradiated with 6000 krad at the University of Lowell Radiation Facility ^{60}Co source. Air sweeping was performed at RADC at 500°C using an electric field of 7000 V/cm.

EXPERIMENTAL RESULTS

QUARTZ E42-21

Figure 2 shows the results of 632.8 nm absorption scans with beam orientations similar to those of Figures 1a, b, c and d for sample E42-21 up to 1000 krad dose. With the beam along the Y-axis, absorption saturation is nearly reached at 1000 krad with the same dose dependence for all positions between the +X and -X faces. Similar dose dependence behavior but with a lower saturation value is observed with the beam along the X-axis for positions between the +Y and -Y faces. For scans made along the Z-axis with the beam in either the X- or Y-direction higher saturation values are observed closer to the -Z face.

Figure 3 shows the 632.8 nm absorption coefficient measured in the X-direction at 1000 krad for 2 to 11 mm beam heights from -Z face. The areas designated with vertical lines demonstrate the increase from the +Z to -Z face, and the relatively uniform absorption between the +Y and -Y faces. Figure 4a shows absorption coefficient values obtained at 1000 krad for scans between the -Y and +Y faces with the beam aligned along the X-axis, at Z-positions 3, 6, 9 and 12 mm from the -Z face. Figure 4b shows absorption coefficient values obtained at 1000 krad for scans between the -X and +X faces with the beam aligned along the Y-axis at the same Z-positions. At the 3 mm Z-position both X- and Y- direction measurements give nearly equivalent absorption. The difference between X- and Y-direction absorption increases for Z-positions 6 and 9 mm with a pronounced difference at 12 mm. The X-direction absorption which varies between 0.13 and 0.15 cm^{-1} is actually an average of the larger variation between 0.12 and 0.18 cm^{-1} measured in the Y-direction at the 12 mm Z-position.

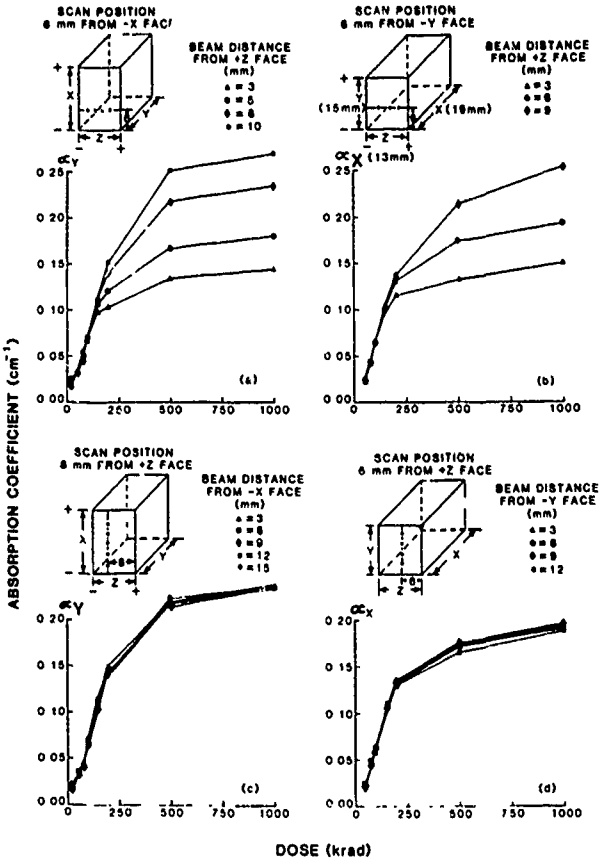


Fig. 2 Sample E42-21, 632.8 nm absorption coefficient irradiation dose dependence with position for different crystal directions.

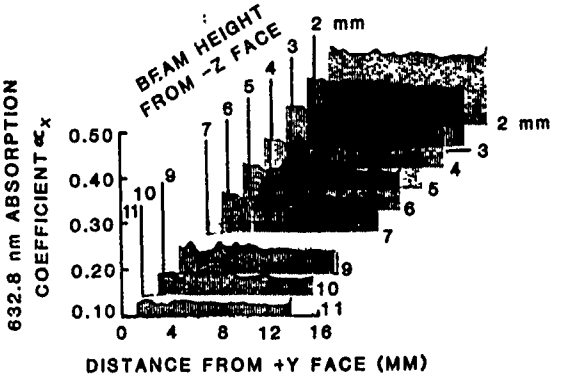


Fig. 3 Sample E42-21, 632.8 nm absorption coefficient dependence with position. 1000 krad dose, beam along the X-direction, scan from +Y to -Y faces, for beam positions measured from the -Z face.

In Figure 5, an expanded plot of the 6 mm Z-position of Figure 4, two strong peaks are found at 4.9 and 6.3 mm from the +Y face with the beam aligned in the X-direction, and at 8.4 and 10.1 mm from the -X face with the beam aligned in the Y-direction. These peaks in absorption are the result of localized scattering of the laser beam due to inclusions or bubbles within the crystal.

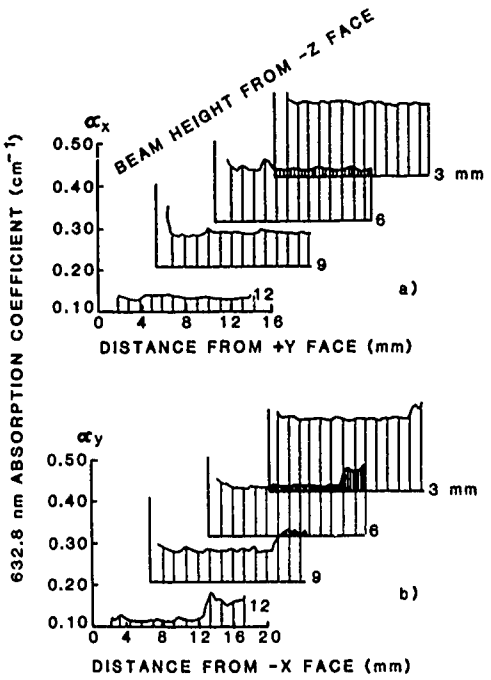


Fig. 4 Sample E42-21, 632.8 nm absorption coefficient dependence with position. 1000 krad dose, beam along X- and Y-directions, scans from +Y to -Y and -X to +X, respectively, beam positions measured from -Z face.

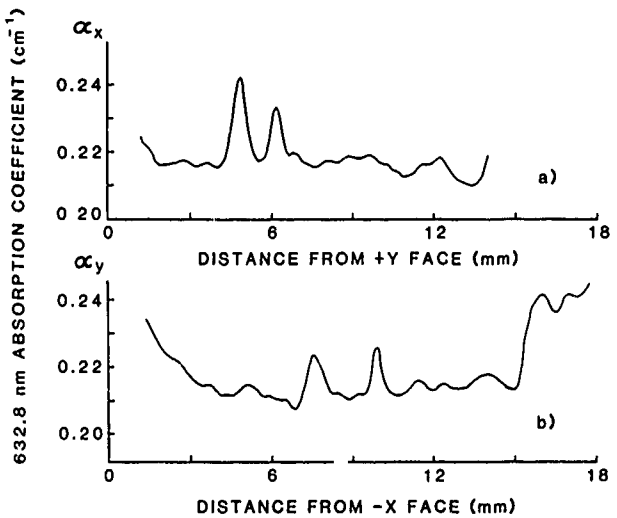


Fig. 5 Sample E42-21, 632.8 absorption coefficient dependence with position. 500 krad dose, 6mm beam position from -Z face, (a) beam X-direction, scan +Y to -Y, (b) beam Y-direction, scan -X to +X.

QUARTZ BH-A

In Figure 6, absorption variation of the 3581 cm^{-1} grown-in OH and the irradiation-induced 3366 cm^{-1} Al-OH band peaks measured along the Z-axis for sample BH-A 1 is shown for 0 to 1000 krad dose range. OH decreases and Al-OH increases up to 250 krad. Between 250 and 1000 krad, no further change in Al-OH and OH absorptions is observed. At the 1000 krad irradiation dose, some OH still remains between 0 and 10 mm from the +Z face, while OH is depleted between 10 and 15 mm. 632.8 nm absorption increases slowly with dose in the OH-containing regions and increases rapidly in the OH-depleted regions. This sample has a high overall OH concentration relative to aluminum, and contains sufficient OH to compensate the Al-M centers dissociated by irradiation over most of the crystal. With higher aluminum concentration toward the -Z end of the crystal, a pronounced increase in Al-h occurs after Al-OH is saturated.

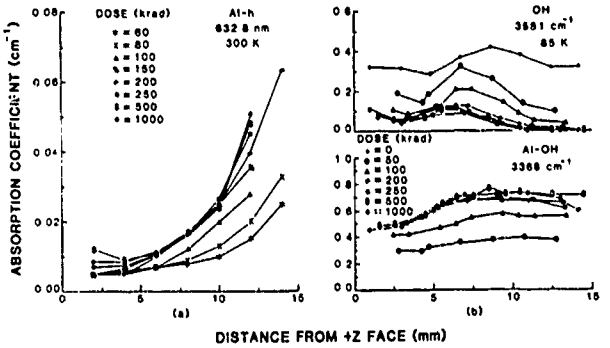


Fig. 6 Sample BHA-1, beam Y-direction, 12 mm from the +X face, scan +Z to -Z. Irradiation dose dependence and variation with Z-axis position of (a) 632.8 nm absorption coefficient and (b) 3581 cm^{-1} grown-in OH and 3361 cm^{-1} irradiation-produced Al-OH absorption coefficient.

Figure 7 shows the 632.8 nm absorption coefficient variation observed in the X- and Y- directions for various beam positions measured from the -Z face, after 500 and 1000 krad irradiation doses, and after annealing. The irradiation-induced absorption level is relatively constant across the sample in both X- and Y-directions in each XY plane, at the 6, 9 and 12 mm beam positions except for isolated strong peaks. Similar to the increase observed for a single X-position in Figure 6, an increase over the entire crystal toward the -Z face is shown here. The absence of this overall absorption in either X- or Y-direction after annealing is evidence that all Al-h centers have been removed. The strong absorption peaks which still remain and are observed only in the Y-direction are attributed to surface scratches on one or both of the Y-faces.

NATURAL QUARTZ

A natural quartz sample fabricated for an AT-cut was measured with the beam normal to the cut face, and scanned at 90° to the beam direction. Figure 8 shows absorption variation across the sample of OH band peaks characteristic of natural quartz. All peaks show strong variation with position and indicate that OH is essentially depleted at the 6.5 mm scan position. Irradiation-induced Al-OH is relatively uniform across the crystal. Al-h absorption and coloration are strongest in the region where OH is depleted.

QUARTZ H33-31

Figure 9, for Y-direction measurements between the +Z and -Z faces for a specific X-position, shows the variation of each of the principal synthetic quartz OH absorption peaks across the unirradiated crystal, and those of the 3581 cm^{-1} OH and 3366

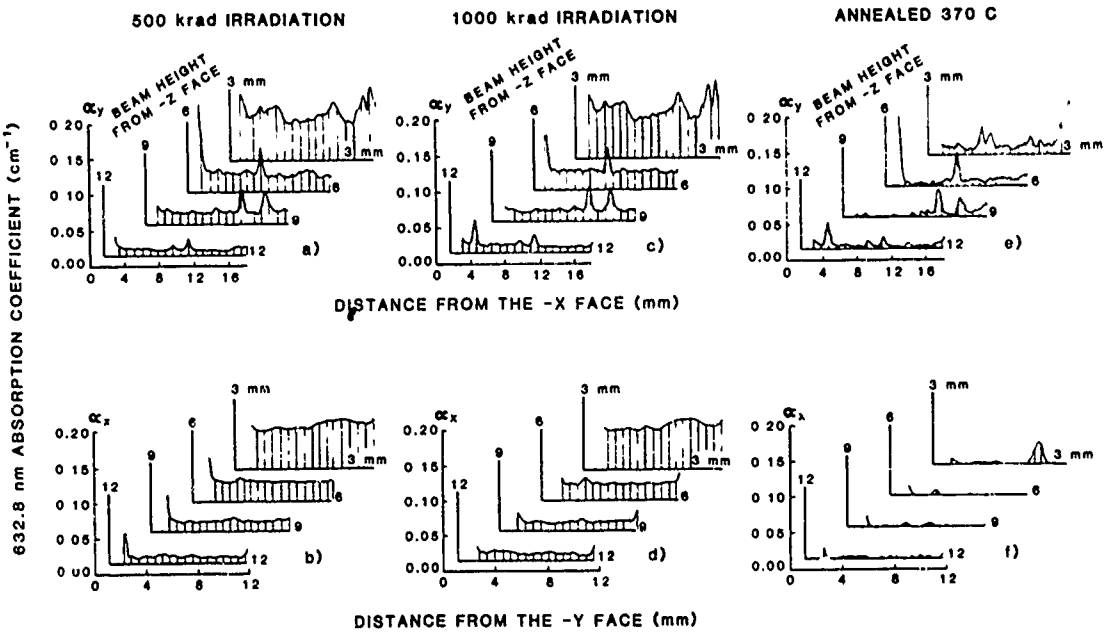


Fig. 7 Sample BHA-1, 632.8 nm absorption coefficient, X- and Y- directions, beam positions measured from the -Z face, (a) and (b) 500 krad irradiation, (c) and (d) 1000 krad irradiation, (e) and (f) 370°C anneal.

cm^{-1} peaks after a series of irradiation doses. After a 400 krad dose, OH reaches a stable level and Al-OH saturates. Even after a 5000 krad dose, OH still remains over the entire crystal providing most of the aluminum compensation. Correspondingly a low overall 632.8 nm absorption is measured between the +Z and -Z faces in the Y-direction as shown in Figure 10a and b.

Air sweeping restored grown-in OH to its initial level with an additional increase between 4 and 8 mm from the anode side as seen in Figure 11a. From Figure 11b, it can be seen that air sweeping produced no change in Al-OH, a further indication of Al-OH saturation.

Figure 11a and b show that the 6000 krad re-irradiation produced no appreciable change in OH, but reduced overall Al-OH. This effect was previously observed in an incompletely air-swept sample [17], but is puzzling in this fully swept sample. The 6000 krad irradiation restored the 632.8 nm absorption to the level reached before air sweeping, Figure 10b. A large reduction in 632.8 nm absorption was achieved by heat treatment at 250°C, a temperature well below that of Al-h removal. Most of the remaining absorption annealed at 370 °C, Figure 10c.

Figures 12a and b show 632.8 nm absorption variations observed between the +Z and -Z faces with the beam aligned along the X-axis at a specific Y-position after various treatments. The broad absorption seen between 10 and 14 mm along the Z-axis after the 5000 krad irradiation was completely eliminated by air sweeping and was not restored by the 6000 krad re-irradiation. Re-irradiation did produce an overall absorption increase. This broad band was also observed between 10 and 14 mm along the Z-axis from measurements made with the beam aligned along the Y-axis. Different irradiation doses, air sweeping and annealing did not remove a sharp peak at a distance of about 11 mm from the +Z face, Figures 12c and d. This peak is attributed to a surface defect.

Figure 13 shows 632.8 nm absorption coefficient variation after 5000 krad irradiation for sets of XY planes at different positions along the Z-axis. Al-h peaks can be observed with the beam in the X-direction between 12 and 15 mm from the +Y

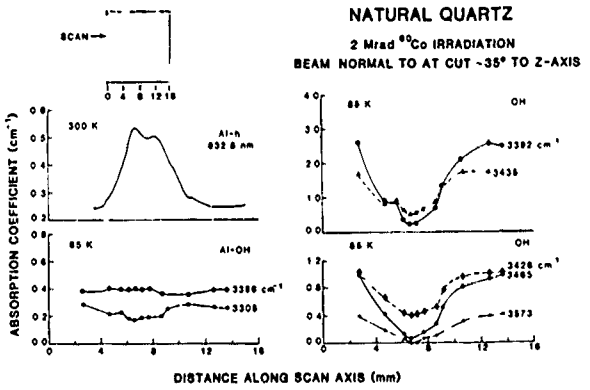


Fig. 8 Natural quartz, coloration, 632.8 nm, OH and Al-OH band peak absorptions after 2000 krad irradiation.

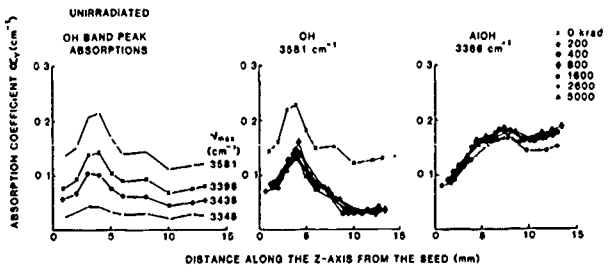


Fig. 9 Sample H33-31, beam Y-direction, 8 mm from -X face, scan +Z (seed) to -Z. Grown-in OH band peak absorptions before irradiation, 3581 cm^{-1} OH and 3366 cm^{-1} Al-OH band peak absorption dose dependence.

face and 0 to 4 mm from the +X face with the beam in Y-direction. A relatively uniform absorption along either the X- or Y-axis is observed for the regions where Al-h peaks are not present.

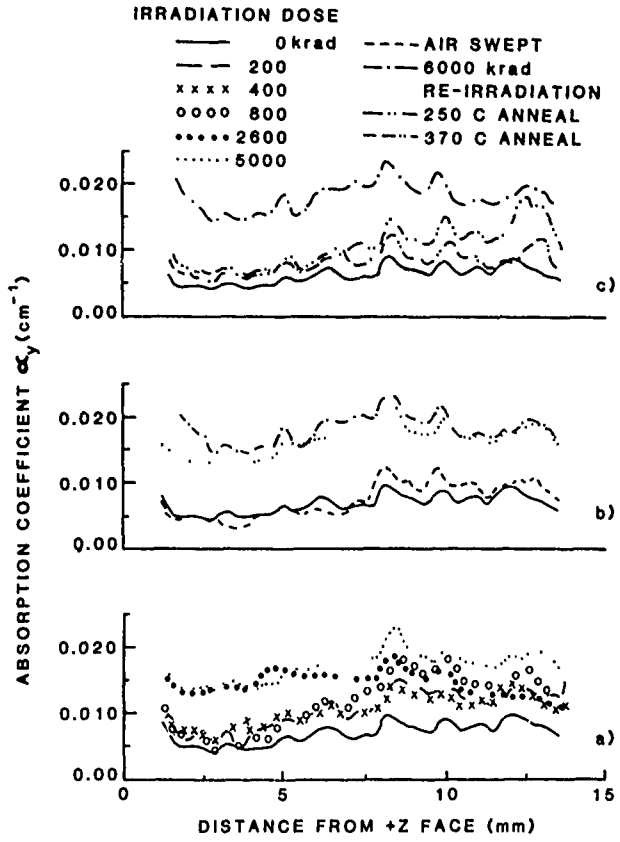


Fig. 10 Sample H33-31, 632.8 nm absorption coefficient, beam Y-direction, 9 mm beam position measured from -X face, scan +Z to -Z, (a) irradiation dose dependence, (b) after air sweeping and re-irradiation, (c) after annealing at 250°C and 370°C.

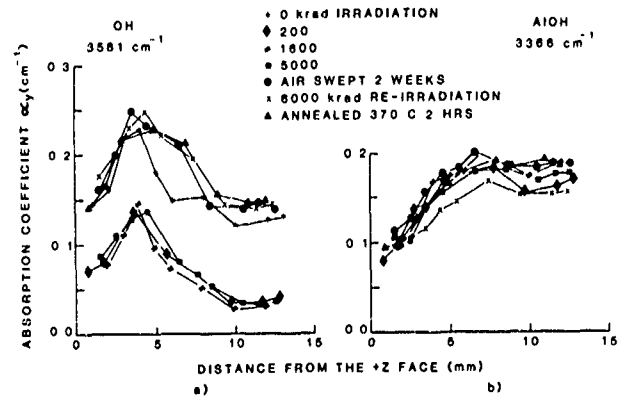


Fig. 11 Sample H33-31, beam Y-direction, 8 mm beam position from -X face, 3581 cm^{-1} OH and 3366 cm^{-1} Al-OH band peak absorption dependence on irradiation dose, air sweeping, re-irradiation and annealing.

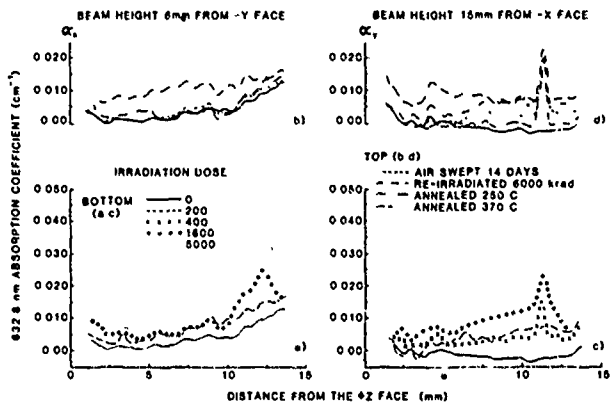


Fig. 12 Sample H33-31, beam X-direction, 6 mm beam position from -Y face, and beam Y-direction, 15 mm beam position from -X face, scan -Z to +Z. (a) and (c) 632.8 nm absorption irradiation dose dependence, (b) and (d) effect of air sweeping, re-irradiation and annealing.

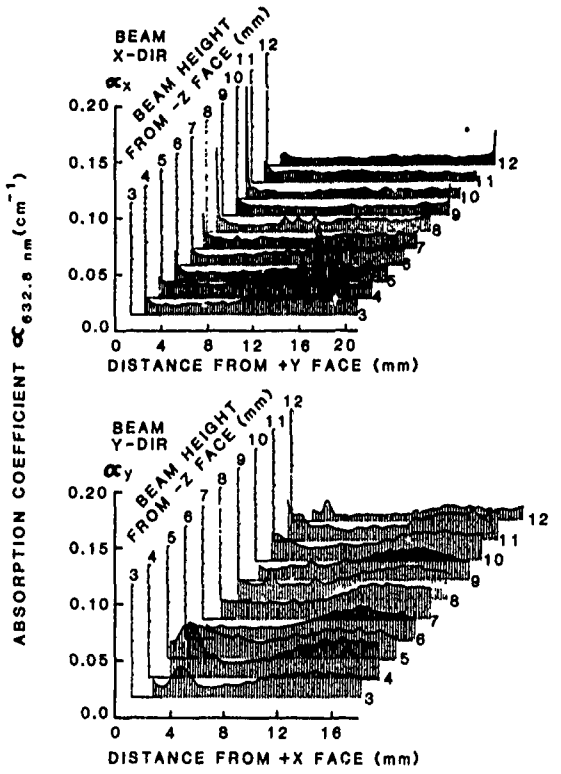


Fig. 13 Sample H33-31, 632.8 nm absorption coefficient dependence with position, 5000 krad irradiation dose, X- and Y- directions, scans from +Y to -Y and +X to -X, respectively, beam positions measured from -Z face.

Figure 14, for only the 3, 4, 5 and 6 mm Z-planes, shows the effect of 14 days of air sweeping and 6000 krad re-irradiation after the 5000 krad initial irradiation. Elimination of Al-h peaks by air sweeping and their failure to recur with re-irradiation, as well as an increase in background with the re-irradiation, are demonstrated in this Figure. The strong peak close to the -Y face, observed only in the X-direction for the 3mm Z-plane,

Figures 14e and f, is due to laser scattering produced by a surface defect generated during the sweeping process.

After air sweeping, infrared measurements made at 8 and 12.5 mm from the -X face with the beam aligned along the Y-axis show essentially the same variation of 3581 cm^{-1} absorption along the Z-axis for both positions, Figure 15. Figure 15b shows larger Al-OH absorption at the 12.5 mm position, close to the region of the sample where the largest Al-h production was detected after the initial 5000 krad irradiation. This indicates this region contains a high localized aluminum concentration, producing the coloration shown by an arrow in Figure 16.

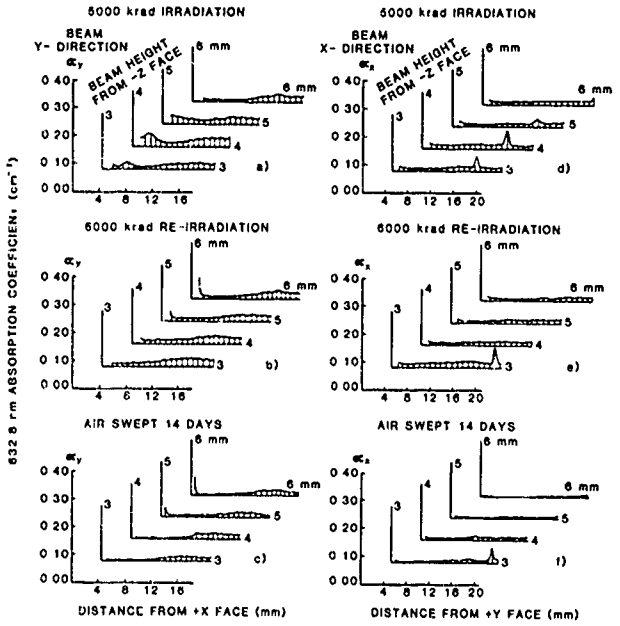


Fig. 14 Sample H33-31, 632.8 nm absorption coefficient dependence with irradiation, air sweeping and re-irradiation. Beam positions measured from -Z face. (a), (b) and (c) beam Y-direction, scans from +X to -X; (d), (e) and (f) beam X-direction, scans from +Y to -Y.

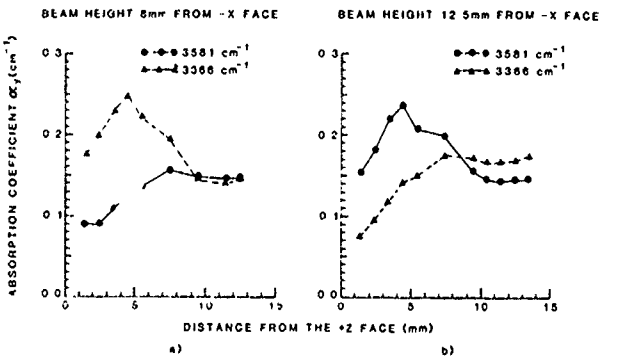


Fig. 15 Sample H33-31, beam Y-direction, beam positions 8 mm and 12.5 mm from -X face, scan +Z to -Z, 3581 cm^{-1} and 3366 cm^{-1} band peak absorptions after air sweeping and 6000 krad re-irradiation.

ANALYSIS AND DISCUSSION

These experimental results, using the combination of infrared spectroscopy and laser absorption measurements, demonstrate the expected dependence of Al-OH and aluminum-hole formation by irradiation on the initial aluminum concentration and OH distribution in the crystal. Both aluminum-compensating mechanisms compete, but Al-OH formation predominates in samples containing large concentrations of grown-in OH. Al-h production proceeds more rapidly in samples with lowered or depleted OH. After irradiation, air sweeping restores OH, increases Al-OH and eliminates Al-h.

Re-irradiation after air sweeping for sample H33-31 does not change OH, reduces Al-OH slightly, and does not restore Al-h, but produces a low 632.8 nm absorption level over the entire sample. Absorption measurements after annealing at a temperature below that of Al-h removal indicate that another defect center may contribute to this low level 632.8 nm absorption. Chemical analysis of the section adjacent to the H33-31 sample investigated gave the following results listed in Table 2.

Table 2: Chemical analysis of H33-31 by inductively coupled plasma spectrometry (ICP).

element	K	Li	Na	Ca	Al	B	Fe
ppm weight	0.81	0.07	2.44	1.90	3.08	2.67	1.14
atomic*	1.24	0.60	6.36	2.84	6.84	17.77	1.22

* atoms per 10^6 Si atoms.

The amount of boron detected in this sample by ICP probably has to be taken into account. This trivalent element, found in many other quartz crystals [18], could be equivalent to aluminum or iron in silicon substitution. This suggestion must be confirmed by more chemical analyses, laser and infrared absorption measurements.

The competition between the production of Al-OH and Al-h is also a controlling factor in the dose dependence of 632.8 nm absorption for sample E42-21. This sample, with a relatively high aluminum-to-OH impurity ratio, is OH-depleted between 50 and 100 krad and Al-OH-saturated above 100 krad [16]. The function

$$\alpha = \alpha_{sat}\ (1\text{-}e^{-kd})^2$$

where α is the absorption coefficient, k is the rate constant and d is the dose in krad, was found to best fit the Al-OH dose dependence [19] which had an inflection point between 10 and 30 krad. An inflection point is also observed here between 60 and 100 krad for Al-h dose dependence. Figure 17 shows 632.8 nm absorption variation between 0 and 1000 krad for three different positions along the Z-axis (points) fitted with (1-exp) and (1-exp)^2 functions. The squared function fits best for the 0 to 100 krad range where both Al-OH and Al-h are produced. Above the 100 krad Al-OH saturation region, a simple exponential function provides a better fit for Al-h production. The 632.8 nm α_{sat} values and rate constants for the Z-positions measured from the +Z face of the sample are given in Table 3.

Table 3. 632.8 nm absorption saturation values and rate constants for sample E42-21.

distance from +Z face (mm)	α_{sat} (cm^{-1})	k ($krad^{-1}$)
3	0.142	0.012
6	0.185	0.0089
9	0.242	0.0069

A decrease in rate constant k is observed for the more strongly absorbing Al-h regions.

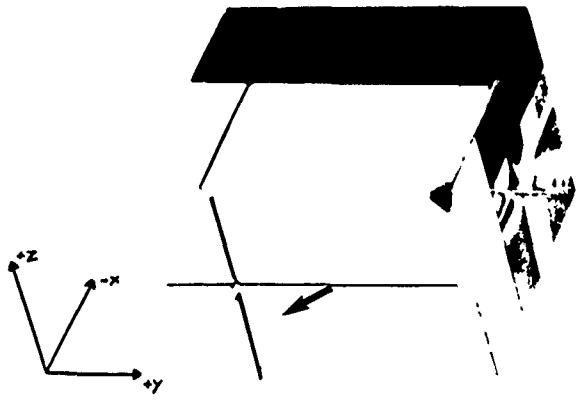


Fig. 16 Localized coloration (designated by arrow) from Al-h in H33-31 sample

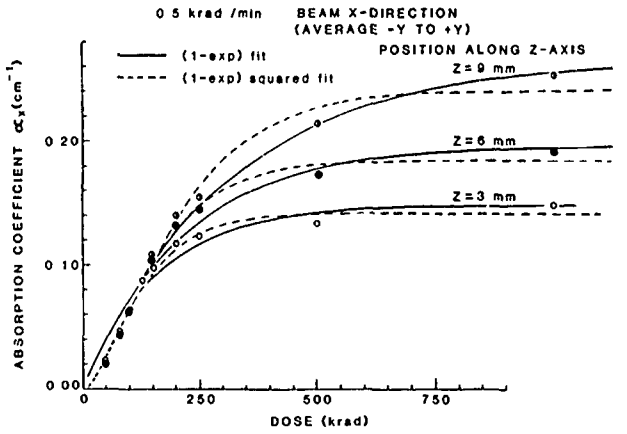


Fig. 17 Sample E42-21, curve fitting of Al-h production dose dependence.

632.8 nm He-Ne scans can be used to locate large localized concentrations of aluminum in a crystal after irradiation with gamma or X-rays. The technique can be used to map Al-h over an entire crystal with polished faces. While an exact determination of aluminum concentration cannot always be made in this way due to the competing process of Al-OH formation, aluminum distributions can be obtained for regions of depleted OH by the laser technique. Annealing above 280°C removes all irradiation-produced Al-h.

Growth defects, such as bubbles, pipes and cracks, are characterized by strong apparent absorptions resulting from laser scattering. These defects are observed from measurements made in an XY plane with Z-axis scans in both the X- and Y- directions.

Mechanical surface defects also give a strong apparent absorption but are observed in only one direction. The absorption intensity of growth and mechanical defects does not change during irradiation and annealing processes.

CONCLUSION

The results of this paper demonstrate that the techniques of low temperature infrared spectroscopy and He-Ne laser absorption can be used to follow changes in OH, Al-OH and Al-h defect centers produced by irradiation, air sweeping, re-irradiation and annealing. Using preferentially chosen observation directions, the laser technique can be used to map the distribution of Al-h, bulk growth and surface defects. Removal of Al-h centers by annealing can be used to differentiate this defect from bulk or surface defects.

Al-h production data, in the dose region of competing Al-OH and Al-h formation fits an exponential squared function similar to that obtained for Al-OH formation [19].

An assumption can be made from results obtained for one of the samples investigated, H33-31, that another type of defect center, associated with boron rather than aluminum, may be present in quartz.

ACKNOWLEDGMENTS

The authors would like to thank F. Euler for helpful discussions on both the experimental investigations and preparation of the manuscript, L. Diaco for competent technical assistance, S. Mittleman and J. Cappelli for performing the irradiations, M. Monte Salvo of the University of Lowell for additional sample irradiations, L. Halliburton of Oklahoma State University for performing the ESR measurements, and M. Khayar and B. Viard of Université de Franche-Comté for ICP analysis.

REFERENCES

1. L.E. Halliburton, J.J. Martin and D.R. Koehler in *Precision Frequency Control*, E.A. Gerber and A. Ballato, eds., Academic Press, 1985, Vol. 1, pp. 32-36.

2. J.C. King and H.H. Sander, IEEE Trans. Nucl. Sci., *NS-19*, 23 (1972); Rad. Effects, *26*, 203 (1975).

3. P. Pellegrini, F.E. Euler, A. Kahan, T.M. Flanagan and T. Wrobel, IEEE Trans. Nucl. Sci. *NS-25*, 1267 (1978).

4. H.G. Lipson, F.E. Euler and P.A. Ligor, Proc. 33rd Ann. Frequency Control Symp., available from Electronic Industries Association, 2001 Eye St., Washington, DC 20006; 122 (1979).

5. A. Kats, Phillips Res. Reports, *17*, 133 (1962).

6. D.M. Dodd and D.B. Fraser, J. Phys. Chem. Solids, *26*, 673 (1965).

7. D.B. Fraser, *Physical Acoustics*, W.P. Mason ed., New York Academic Press, Vol. V, ch. 2, pp. 59-110.

8. B.R. Capone, A. Kahan, R.N. Brown, and J.R. Buckmelter, IEEE Trans. Nucl. Sci., *NS-17*, 217 (1970).

9. K. Nassau and B.E. Prescott, Phys. Stat. Sol.(a), *29*, 659 (1975).

10. M.E. Markes and L.E. Halliburton, J. Appl. Phys., *50*, 8172 (1979).

11. H. Jain and A.S. Nowick, J. Appl. Phys., *53*, 485 (1982).

12. B. Sawyer, IEEE Trans. Ultrason., Ferroelec. and Freq. Control, *UFFC-34*, 558 (1987).

13. H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci., *NS-31*, 1223 (1984).

14. H.G. Lipson and A. Kahan, J. Appl. Phys., *58*, 963 (1985).

15. H.G. Lipson, Proc 40th Ann. Freq. Control Symp., available from IEEE, 445 Hoes Lane, Piscataway, NJ 088544, doc. no. 86CH2330-9, 63 (1986).

16. H.G. Lipson and P. Zecchini, Mat. Res. Soc. Proc., 1988 Fall Meeting, in publication.

17. A. Kahan and H.G. Lipson, Proc. 40th Ann. Freq. Control Symp., available from IEEE, 445 Hoes Lane, Piscataway NJ 08854, doc. no 85CH2186-5, 255 (1985).

18. M. Khayar, B. Viard and H. Merigoux, Proc. 3rd European Frequency and Time Forum, (1989), in publication.

19. H.G. Lipson and F.E. Euler, Proc. 42nd Ann. Freq. Control Symp., available from IEEE, 445 Hoes lane, Piscataway NJ 08854, doc. no 88CH2588-2, 169 (1988).

43rd Annual Symposium on Frequency Control - 1989

LATTICE DEFECTS IN LITHIUM TANTALATE

J. C. DOUKHAN, P. CORDIER and N. DOUKHAN
Laboratoire de Structure et Propriétés de l'Etat Solide (URA CNRS n° 234)
Université de Lille Flandres Artois, 59655 Villeneuve d'Ascq Cedex FRANCE
M. FRAGNEAU and J. P. AUBRY
Compagnie d'Electronique et de Piézo Electricité BP 165 94100 Argenteuil FRANCE

ABSTRACT

The lattice defects in lithium tantalate single crystals and their possible role on the performances of piezoelectric devices are investigated. Synthetic crystals grown by a Czochralski process are non stoichiometric (with an atomic ratio [Li]/[Li+Ta] ≈ 48%). This induces a first kind of lattice defects: the point defects associated to this non stoichiometry. The other possible point defects are the chemical impurities; the substitution of Li by H seems to be the main one.

During cooling from high temperature $LiTaO_3$ suffers a phase transition (R3c -> R3c). A second kind of lattice defects (ferroelectric domains) is generally nucleated at the transition. These defects constitute a poison because the polarization vector **c** is reversed in the domains. One can in principle prevent them by a poling process (cooling under a static electric field).

Dislocations and twins occur as as-grown lattice defects but they can also be introduced by standard machining processes (sawing, grinding ...). Furthermore because of the very high values of the piezoelectric constants the stress field of the dislocations can induce around them ferroelectric domains, even at room temperature, and such domains cannot be removed by poling.

These defects have been characterized by various techniques and especially by transmission electron microscopy (TEM). It is found that the density of as-grown defects should not induce any appreciable decrease of the piezoelectric performances excepted for some dislocations which generate non removable ferroelectric domains around them. Such dislocations can also be accidentally introduced by machining operations as well; in this latter case their density can be much larger and an appreciable decrease of the piezoelectric performances may result. The techniques used for sawing and polishing are thus very important for getting performant devices.

INTRODUCTION

Lithium tantalate is one of the rare piezoelectric materials which can lead to thermally compensated resonators and its very strong coefficients of electromechanical coupling are highly appreciated for specific applications (see a recent review in [1]). This material is synthesized since 20 years by a Czochralski process; to get homogeneous crystals it is necessary to start with a non equimolar mixture of Li_2O and Ta_2O_5 leading to an uncongruent melt and the resulting crystals are non stoichiometric. The crystal structure of $LiTaO_3$ has been refined by X-Ray and neutron diffraction techniques [2,3]. It is isostructural with $LiNbO_3$ and the structural models for the ferroelectric R3c -> paraelectric R3c phase transition have thoroughly been discussed for both materials [4, 5]. The only noticible difference is the value of

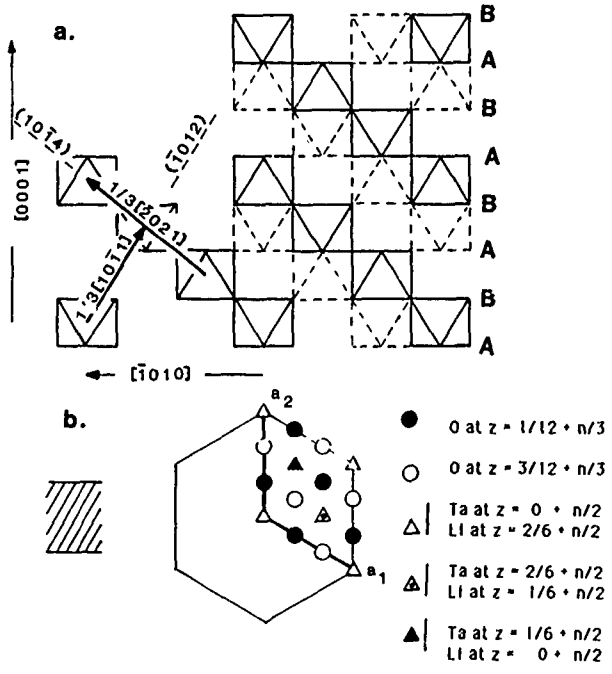


Figure 1: Simplified view of the crystal structure of $LiTaO_3$. The HCP oxygen sublattice is indicated by the sequence A,B,A,B...; the TaO_6 octahedra are drawn with solid lines and the LiO_6 ones with broken lines. a) projection on (1$\bar{2}$10); only a slab corresponding to the width of an octahedron (hatched region in b) is represented, a few short lattice repeats which are possible Burgers vectors are reported. b) projection on (0001), the levels of the various atoms along the **c** axis are reported on the right, n is an integer ($0 \geq n \geq 3$).

CH2690-6/89/0000-497 $1.00 C 1989 IEEE

the transition temperature which is of the order of 620 °C for $LiTaO_3$ while it is 1100 °C for $LiNbO_3$. The ferroelectric domains in the low temperature phase correspond to an inversion of the polarity and are thus an electric poison for piezoelectric applications. In principle such domains can be eliminated by poling (cooling under static electric field).

A working approach of the crystal structure is easily obtained in assuming that the oxygen sublattice is a perfect hexagonal closed packed structure (HCP) with the Li^+ and the Ta^{5+} cations in the octahedral sites. 1/3 of these sites is empty, 1/3 is occupied by the Li^+'s and the last third by the Ta^{5+}'s. As pointed out by Megaw [5], the rigid framework of the structure is constituted by the skeleton of edge-sharing TaO_6 octahedra. This framework in shown on Fig. 1 and the configuration of a ferroelectric domain is shown on Fig. 2 for a single column of octahedra parallel to the c axis. Such a domain is generated by the cooperative jump of the Li ions from their initial position beneath the Ta's to the adjacent empty octahedral site. The intermediate configuration corresponds to the high temperature paraelectric phase.

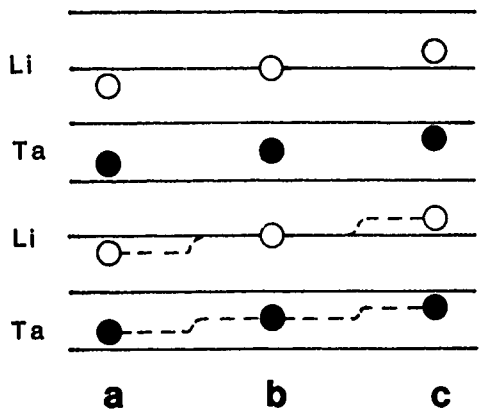


Figure 2: Formation of a ferroelectric domain by jump of the Li's in the octahedral empty sites. a) matrix. b) intermediate configuration (paraelectric phase). c) ferroelectric domain.

We have investigated and characterized by chemical etching and by transmission electron microscopy (TEM) the lattice defects occurring in this material, either the as-grown defects or the ones induced by machining (simulated by plastic deformation under various temperature and confining pressure conditions). The crystals being Li deficient (with an atomic ratio Li/(Li+Ta) of the order of 48%) they contain an appreciable density of point defects associated to this non-stoichiometry. As long as these point defects are randomly distributed they cannot be detected by TEM; in contrast, if they cluster along some surfaces, they form crystallographic shear structures which can be detected by TEM [6,7]. Among the chemical impurities expected in $LiTaO_3$, hydrogen which can be substituted to Li in an appreciable proportion could be the most abundant; we have estimated its concentration in various crystals by infrared spectroscopy (IRS) in the range 4000-2500 cm^{-1} which corresponds to the absorption bands for O-H stretching vibrations [8]. We first report our results on the as-grown materials (section II). We then report the tests and TEM observations related to the simulation of the damages induced by machining (section III). Finally we briefly discuss the possible influence of these defects on the performances of piezoelectric resonators.

II DEFECTS IN AS-GROWN MATERIALS

II-1 Point defects

Crystals from various origins have been investigated, however most of the experiments have been performed on two crystals only. One was grown 10 years ago by CNET[1] and the other one has been grown recently by CRISMATEC[2]. The chemical composition, i.e. the ratio Li/(Li+Ta), can in principle be determined by X Ray diffraction but the variation of the lattice paramerters *a* and *c* with composition is so small than reliable measurements would require a very high precision diffractometer [9,10]. We have used another method based on the variation with composition of the ferroelectric-paraelectric transition temperature [9-12].

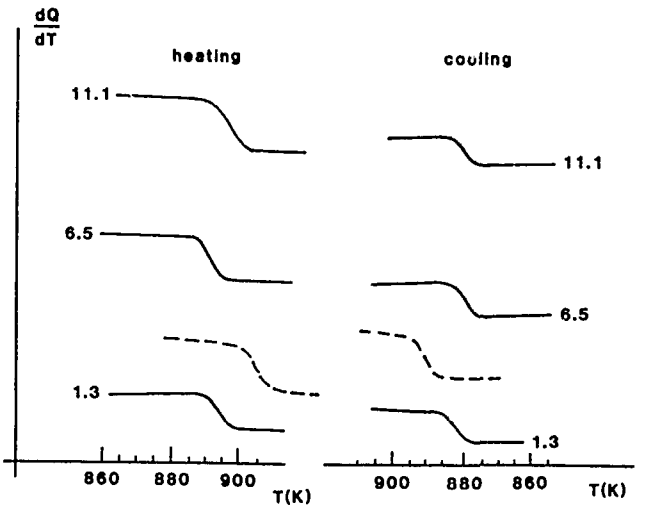


Figure 3: Typical DSC curves for a scan rate 40 °C min^{-1}. Labels 11.1, 6.5 and 1.3 correspond respectively to the bottom, the middle and the top of a CRISMATEC boule. The broken line corresponds to a HITACHI boule with a different stoichiometry.

This temperature is easily measured on very small pieces of

[1] Centre National d'Etudes des Télécommunications, Lannion France.

[2] CRISMATEC, Grenoble France.

material by differential scanning calorimetry (DSC). For fast scans (40°C/min) one records transition temperatures which are slightly different when temperature increases or when it decreases; therefore one does not obtain absolute compositions (the calibration curve itself is not very precise) but comparison between different samples is easy. Some typical DSC curves are reported on Fig. 3. Choosing as transition temperature Tt the turning point of the increasing temperature curve, a change ΔTt of 4 °C (which is detected unambiguously) corresponds to a change of 0.1% of the atomic composition ratio Li/(Li+Ta). As can be seen on Fig. 3 the crystals grown by CRISMATEC have a rather homogeneous composition but it is not strictly constant from boule to boule (Tt =621 ± 3 °C in a recent boule, Tt = 628 ± 1 °C in an older one). The japanese material (HITACHI) has a different composition (Tt = 633 °C) nearer to stoichiometry and it might be less homogeneous.

It has been experimentally shown that H or OH ions are soluble in $LiNbO_3$ [13-16]. We have thus performed a series of annealings in air with a view to test the possibility of Li-H exchange owing to the traces of H_2O which are always present in air. No change of Tt is detected after 100 hours at 800 °C but the OH content increases appreciably (Fig. 4). To get standardisation we have first characterized by IRS a cube cut in the as-grown CNET crystal with its faces parallel to (0001), (2$\bar{1}$10) and (10$\bar{1}$0) respectively. The absorbance curves with non-polarized light clearly show some dissolved hydrogen or hydroxyl in this material (characteristic absorption at 3400 cm^{-1} due to the stretching vibrations of the OH groups in well defined sites); the largest absorbance occurs for a beam parallel to the **c** direction. This is the geometry adopted for the other tests done on annealed samples. To estimate quantitatively the H content we have used the calibration used for quartz [17]. The atomic concentration H/Li (in at. ppm) is assumed to be approximately equal to the area (in cm^{-2}) under the curve absorbance A (cm^{-1}) versus wave number ν (cm^{-1}). This leads to an atomic concentration H/Li ≈ 30 at. ppm in the as-grown CNET crystal. In the as-grown CRISMATEC crystals this concentration is approximately 3 times larger and it is found to be homogeneous in the various boules which have been tested. After annealing at 500 °C no change is detected, but after annealing at 800 °C the H content increases significantly for long annealing times and H/Li is roughly multiplied by a factor 3 after 48 hours annealing. The solubility of H in $LiTaO_3$ is known to be large [18] and the occurrence of appreciable amounts of H can affect seriously a number of physical properties. However H-Li exchange at high temperature and in the absence of electric field appears to be very low and to have a very low cinetics as compared

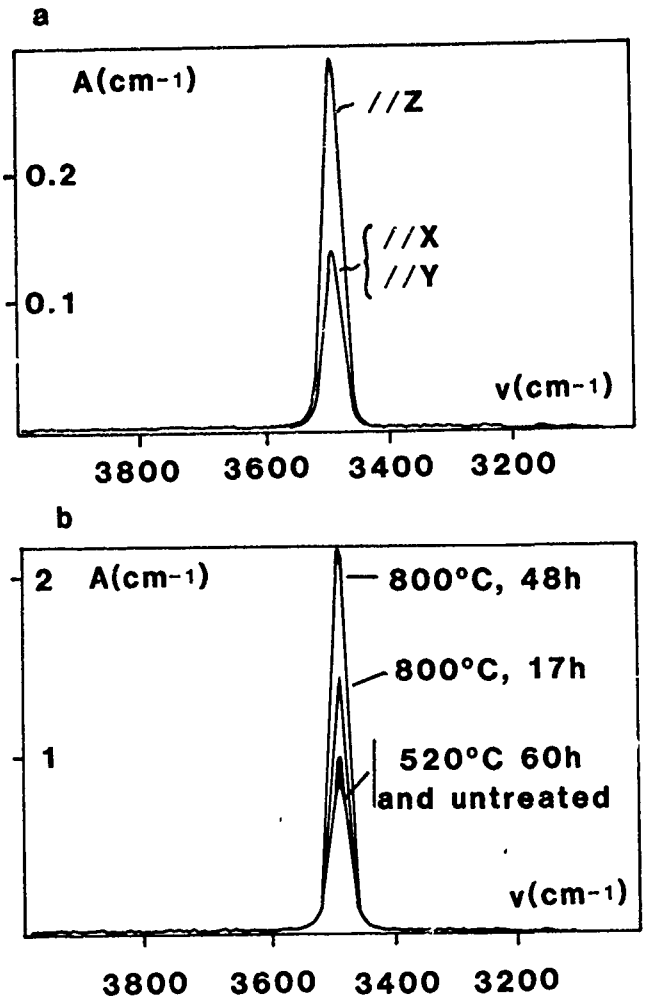


Figure 4: Infrared spectra. a) Influence of the crystallographic orientation; as-grown CNET crystal with the incident beam parallel to Z, X, and Y respectively. b) Influence of annealing in air ; CRISMATEC crystal, beam parallel to Z; the annealing conditions are indicated on the curves.

to the Ag-Li exchange observed by Bursill and Peng [19].

The few characterizations reported here show that there are significant differences of chemical composition from boule to boule which probably stem from slightly different growth conditions, however every CRISMATEC boule appears to be homogeneous.

II-2 Dislocations

a) chemical etching: $LiTaO_3$ like $LiNbO_3$ is chemically inert and the only known chemical etchant at moderate temperature for $LiTaO_3$ is a solution of 1/3 of concentrated HF and 2/3 of concentrated HNO_3 [10,20 to 22]; the etching process requires 1 to 4 hours at 110 °C which is the boiling

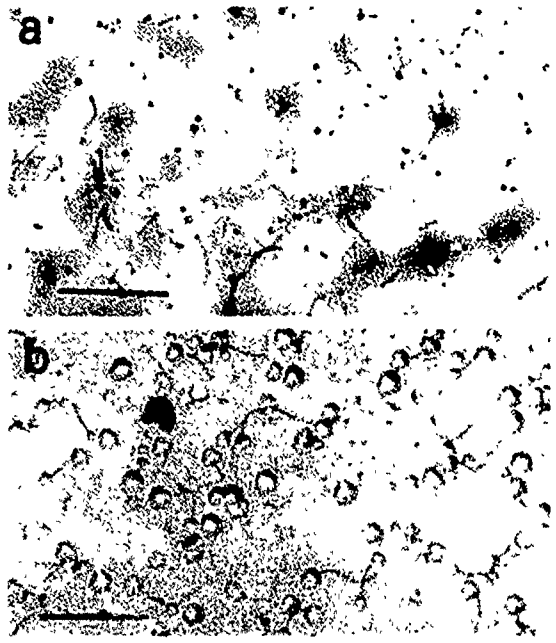


Figure 5: CNET crystal, basal slab mechanically polished with Syton W30 then etched two hours at 110 °C. a) and b) show the two opposite faces which are etched at different rates. One also distinguishes clearly the channels emerging at the etch pits. Scale bar = 100 µm.

temperature of the solution. For optical observations it is necessary to start from a well polished surface. The best mechanical polishing has been obtained with Syton[3] (grade W 30). Fig. 5 shows some typical results. On basal slabs the etching rate is different on both faces and the corresponding etch pits have different sizes. Furthermore, on one basal face, one clearly observes tunnels which start from some of the etch pits; it is tempting to assume that they result from preferential dissolution of the dislocation cores. Scratches and other residual polishing defects are easily detected; it seems that they never penetrate very deeply beneath the surface in $LiTaO_3$ (less than 10 µm). Finally the density of dislocations, as revealed by chemical etching, is of the order of 10^{10} m^{-2} and the density of etch pits prolongated by tunnels is only 10^8 m^{-2} approximately. These later ones could correspond to the as-grown dislocations while all the other ones (which represent 99% of the total) would correspond to dislocations nucleated by machining and which do not penetrate deeply beneath the polished surface in such a way that no tunnel is visible. This would mean that the operation of mechanical polishing with Syton is not perfect. We have performed a few longer polishings (three to four hours) and verified that the density of etch pits not related to tunnels is dramatically decreased. This comforts the above hypothesis about the origin of these two kinds of dislocations.

[3] Sold by Monsanto. It is a suspension of very small spheres of silica (1000 Å) in an acqueous solution of concentrated KOH (pH = 10.4) to prevent accretion between the spheres.

b)Transmission Electron Microscopy (TEM): thin foils (30 µm) polished with Syton on both faces have been ion thinned in a Gatan apparatus. We have remarked that ion thinning at moderate accelerating voltage (4 to 4.5 kV) reveals the surface defects (scratches for instance) which had not been completely removed by mechanical polishing

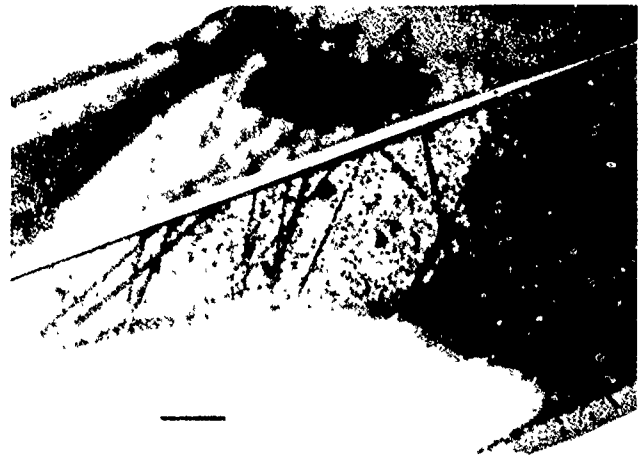


Figure 6: Defects revealed by preferential ion thinning on a thin foil (acceierating voltage 4.5 kV); scale bar = 0.1 mm.

although they were not detected by optical observation of the corresponding regions. These remaining defects are thinned more rapidly and holes in the thin foil generally result. A typical example is shown on Fig. 6. For higher accelerating voltage (6 to 7 kV) the samples become dark. It has already been pointed out that annealing under vacuum induces a similar darkening which has been attributed to the nucleation of point defects associated with loss of oxygen; a further annealing in air retores the initial transparency [10]. We presume that the argon beam generated by a high accelerating voltage of 6 to 7 kV induces locally an appreciable temperature increase which could induce in turn some loss of oxygen. We have therefore chosen an accelerating voltage of 5.5 kV and an incident angle of 15° which appeared to be the best compromise; however even with these conditions it appeared very difficult to obtain good thin foils (i.e. transparent over large areas). Most of our samples were irregularly thinned and only sparse areas were found to be convenient for TEM observations. It seems that quite similar difficulties have been met by other authors (Venables [21] and Bursill and Peng [19] also show electron micrographs with restricted transparent areas). Very few as-grown dislocations are observed in the thin foils, their density is lower than the one detected by chemical etching and the reason for such a discrepancy may stem from the fact that an appreciable part of the dislocations have disappeared during the ion thinning process either because they are replaced by holes (preferential ion thinning) or

because these dislocations were confined into the external layers of the foil which are removed during the thinning process (such dislocations would have been introduced during mechanical polishing). The rare dislocations observed in TEM are sometimes pinned by some kind of precipitates. Fig. 7a and 7b show an as-grown dislocation pinned by tiny spherical precipitates which could not be characterized because of their too small size. Standard diffraction contrast analysis indicates that the Burgers vector of this dislocation is 1/3[10$\underline{1}$1] while its line lies in (10$\underline{1}$2). Furthermore this dislocation is slightly dissociated (but the fringe contrast associated to the fault can only be seen in weak beam conditions). In the vicinity of this dislocation one observes a number of circular contrast features; the corresponding lattice defects are not yet fully elucidated. They might correspond to crystallographic shear faults due to the clustering of point defects (oxygen vacancies?). Fig. 7c shows another as-grown dislocation linked to a platelet shaped precipitate lying in the (10$\underline{1}$4) plane; this later dislocation is apparently not dissociated; its Burgers vector is 1/3 [2$\underline{1}$1$\underline{0}$]. A few other as-grown dislocations have been observed; some are surrounded by ferroelectric domains (Fig. 7d) which are probably induced by the stress field of the dislocation; such domains cannot be suppressed by poling because they are automatically renucleated at the transition temperature. These complex defects constituted by a ferroelectric domain surrounding a dislocation are probably the ones which affect most seriously the performances of piezoelectric devices.

II-3 Extended defects

a) twins: No as-grown twins have been observed in the studied samples although the occurrence of twins in the {1012} planes has already been reported in $LiTaO_3$ as well as in $LiNbO_3$ [5,10,19,23]. Bursill and Peng [19] have also observed "chemical twins" in the {10$\underline{1}$4} planes (the concept of chemical twins is developed in [24]) but no such defects have been observed in our material.

b) precipitates: We have heated a thin sample in the microscope owing to a heating sample holder with a view to investigate its evolution with temperature. After one hour at 550 to 600 °C numerous tiny dark spots appear (Fig. 8a). They are too small to be conveniently analysed by microdiffraction. We assume that they are tiny particles of Ta_2O_5 which have precipitated to compensate the initial non stoichiometry of the material. Very similar precipitates have been observed in $LiNbO_3$ after heating at 550 °C in air [25]. The dislocation shown on Fig. 7a is probably pinned by such tiny precipitates formed during crystal growth. A prolongated electronic irradiation leads to visible irradiation damage but heating above T ≈ 700 °C restores completely the crystal structure. This is well illustrated by Fig. 8a to 8c which show

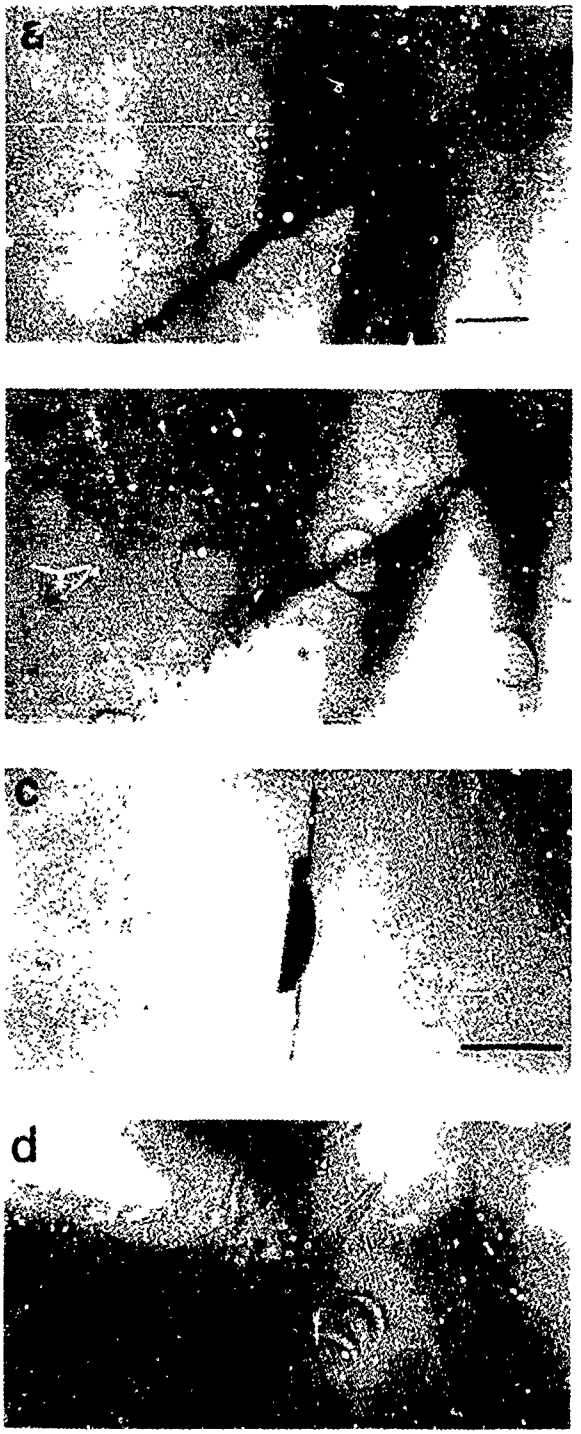


Figure 7: Poled CRISMATEC crystal; TEM micrographs of as-grown defects in a (2$\underline{1}$1$\underline{0}$) thin foil. Scale bar = 1 µm. a) scalloped dislocation. b) same dislocation, it is out of contrast for $\vec{g}$ = 12$\underline{1}$0 and the small pinning defects are now clearly visible; they are probably very small inclusions (< 500 Å) of tantalium oxide; note also the circular defects in the vicinity of the dislocation. c) dislocation pinned by a small platelet shaped inclusion. d) dislocation surrounded by a ferroelectric domain.

the evolution of a given area containing initialy a dislocation (seen edge on) surrounded by a ferroelectric domain. After moderate heating, irradiation damage begins to occur, mostly on one side of the ferroelectric domain; after further heating at higher temperature the damage extends to a wider region but above 700 °C a complete restoration occurs within minutes while the ferroelectric domain has disappeared; furthermore a new dislocation loop is nucleated.

c) ferroelectric domains: Ferroelectric domains are quite common in non poled materials. When a sample is "depoled" by heating it above the transition temperature and cooling it down without applied electric field, numerous ferroelectric domains are spontaneously nucleated at the transition; the boundaries have no preferred orientation (see for instance Fig. 8d), no chemical twins are formed. Ferroelectric domains are easily characterized owing to the differences of the structure factors of the two variants.

d)crystallographic shear(CS): This particular surface defect is known to occur in transition metal oxides; it generally accomodates the oxygen deficiency in changing along some surface the organization of the corner (or edge) sharing octahedra into edge (or face) sharing ones (see for instance [6,7]). In lithium tantalate it is generally assumed that non stoichiometry is accomodated by randomly distributed point defects [26, 27]. However such defects could also cluster and form crystallographic shear defect surfaces. The circular defects shown on Fig. 7b could be such defects because the associated fault vector (which could not be completely characterized) seems to lie out of the fault plane as it is the case for CS defects. It is to be mentionned that these defects are very rare in high quality materials and they have been observed only in one crystal (the resonators cut in this material presented an anomalous behaviour). Our in-situ TEM investigations with a heating sample holder have shown that these CS defects, or whatever they are, cannot be suppressed by heat treatment even above the transition temperature; in contrast, new ones are nucleated with n minutes at T ≈ 600 °C, then their population becomes stable. In numerous areas they are randomly distributed but in some other places they are roughly distributed along some directions which might correspond to regions with an initial stoichiometric ratio slightly lower than the mean one (Fig. 9).

III DEFORMATION EXPERIMENTS

The aim of these experiments was to simulate the defects induced by mechanical polishing. Indeed it is not possible to ion thin samples which have been ground but not carefully polished because this would lead to TEM samples with irregular thicknesses and therefore with very restricted transparent areas We have performed three kinds of tests:

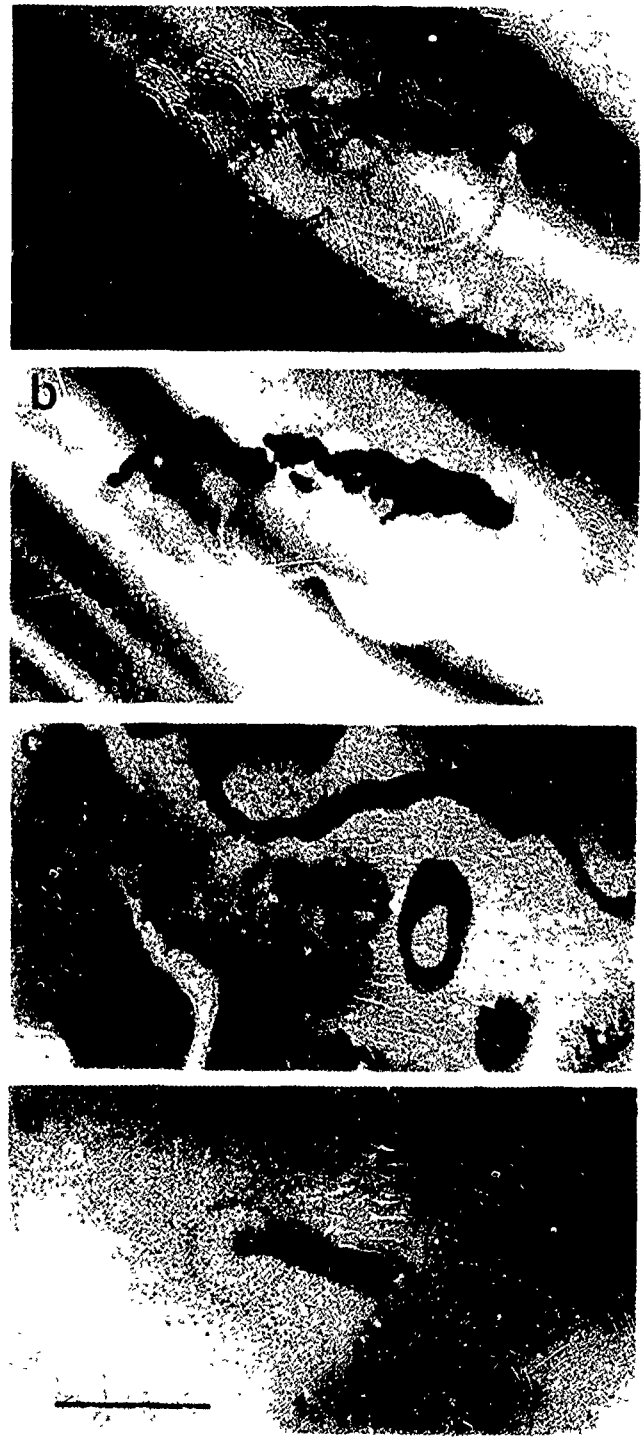
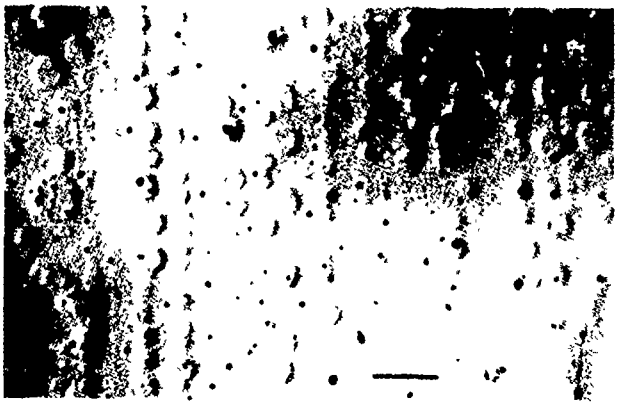


Figure 8: Evolution of the damage caused heating and electronic irradiation; scale bar = 1 µm. a) initial situation; the ferroelectric domain is more sensitive that the matrix and it is already slightly irradiated at room temperature. b) after one hour heating at 650 °C c) after 15 minutes at 780 °C the damage is completely restored while a new dislocation loop has been nucleated. d) the sample has been cooled down to room temperature and numerous ferroelectric domains have been nucleated.



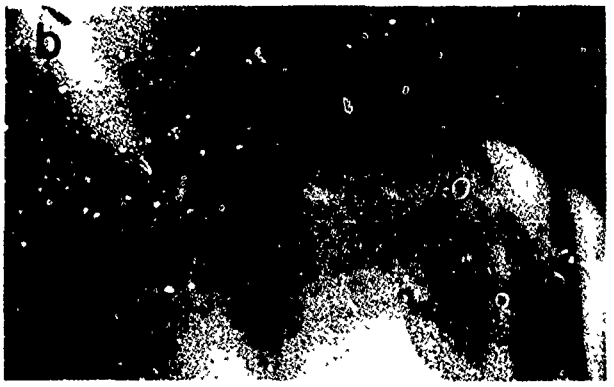


Figure 9; Thin sample heated in-situ in the microscope; scale bar = 1 µm. a) new CS defects formed after one hour heating at 600 °C. They are randomly distributed. b) high concentration of CS defects in an elongated zone; they might reveal an initial lower composition ratio.

-The first one consists in micro-indentation followed by TEM investigation of the induced defects. This technique has been successfully used for a number of stong materials, especially ceramics (see a review in [28]). Micro-indentation is a highly heterogeneous deformation performed under a large confining pressure (of the order of 2000 MPa) and with an appreciable deviatoric stress; the resulting strain rate is also very heterogeneous and in the immediate vicinity of the indent it is probably as large as 10^{-2} s^{-1}. Such conditions are somewhat similar to the ones occurring during mechanical grinding. This is a very simple and rapid technique but it has to be performed on well polished surfaces. Subsequent TEM analysis of the indented region allows, in principle, the determination of the activated deformation mechanism (indeed, if dislocations have been activated, characterizing their Burgers vectors and their line orientation allows the determination of thecorresponding activated glide systems).

Figure 10; CRISMATEC crystal; (10$\bar{1}$0) thin foil; scale bar = 1 µm. Dislocations induced by scratching. a) Dislocations with their surrounding ferroelectric domains. b) Same area, the domains are out of contrast. c) Other region; dislocations highly entangled in such a way that their complete characterization could not be achieved. No domains are detected around them.

In the case of very low dislocation mobilities, scratches (rather than static indents) are to be preferred because they lead to larger dislocation loops Fig. 10a and b show the defects induced by such a scratch performed with a low load to prevent fracture (lithium tantalate is brittle; this could be either an intrinsic property of the material due to weak bonds along some planes, or this could result from a high level of internal residual stresses). Some scratch induced dislocations are surrounded by cylindrical ferroelectric domains; their cross-section, 1 to 2 µm wide, has roughly the shape of a butterfly; they have certainly been nucleated by the strees fields of the dislocations. Such domains do not systematically occur; in a number of other cases no domains are found around the dislocations (see for instance Fig.10 c).

-We have also performed some compression tests at room temperature and under a confining pressure of 600 MPa. This type of experiment still is rather similar to the case of mechanical grinding and its main advantage stems in the control of the orientation of the deviatoric stress. The deformation equipment used is derived from the so-called "Griggs deformation apparatus" and the confining medium is a synthetic rubber (for details, see [29]). A rapid investigation of the crystal structure (Fig. 1) shows that the shortest lattice repeats which are good candidates as Burgers vectors i.e. as glide directions are 1/3 <2$\underline{1}$1$\underline{0}$> (I = 5.15 Å), 1/3 <1$\underline{0}$1$\underline{1}$> (I = 5.19 Å), 1/3 <20$\underline{2}$1> (I = 6.68 Å) and <1$\underline{0}$1$\underline{0}$> (I = 8.92 Å); assuming that the associated glide planes are the densest planes containing these vectors, the possible easy glide systems could thus be: a glide in the dense planes {0001}, {1012} and {10$\underline{1}$4}; {10$\underline{1}$2}1/3<1$\underline{0}$1$\underline{1}$>, {1$\underline{2}$10}1/3<1$\underline{0}$1$\underline{1}$>, {10$\underline{1}$4}1/3 <20$\underline{2}$1>, {10$\underline{1}$2}1/3 <0$\underline{2}$21> and <1$\underline{0}$1$\underline{0}$> in {0001}, {1210} or any of the dense pyramidal planes {12$\underline{1}$n}. Four compression orientations parallel to the c axis (noted Z), to the a_1 axis (X), perpendicular to these axes (Y) and at 45° to them (noted O+), respectively have been tested. The Schmid factors of these various glide systems are reported in Table I and the resulting stress-strain curves are shown on Fig. 11a.The experiments corresponding to the compression axes X and Y exhibit a linear stress-strain relationship up to a large deviatoric stress of 2000 to 2500 MPa; the experiments end by an explosive fracture and an abrupt drop of the applied load (indicated by an arrow on Fig. 11a). This is typical of a purely elastic behaviour; it thus seems that the most favoured glide systems for these orientations, i.e. {1012}1/3 <1011>, {2110}<0110>, {1014} 1/3 <2021>, {10$\underline{1}$2}1/3 <20$\underline{2}$1> and {01$\underline{1}$0}1/3 <2$\underline{1}$10> could not be activated. For the two other orientations the recorded curves are typical of a ductile deformation with a progressive decrease of the slope $d\sigma/d\varepsilon$; the maximum flow stresses are approximately 500 and 1000 MPa respectively. For orientation Z, the possible glide systems (i.e. with non vanishing Schmid factors) are the same as the ones which could not be activated in the previous experiments, therefore the recorded plastic strain can only result from twinning. For the O+ orientation, basal a glide is the best candidate (smallest Burgers vector and maximum Schmid factor = 0.50). The yield phenomenon at the beginning of the deformation curve could correspond to a process of multiplication of the initial dislocation density like in covalent semi conductors [30]. As shown below the TEM observations comfort these hypotheses.

TABLE I: Schmid factors of the various possible glide systems for the four compression orientations tested. For each family of equivalent glide systems only the largest Schmid factor is indicated.

	Z	X	Y	O+
{0001}1/3<2$\underline{1}$1$\underline{0}$>	0	0	0	0.50
{0001}<1$\underline{0}$1$\underline{0}$>	0	0	0	0.43
{01$\underline{1}$0}1/3<2$\underline{1}$1$\underline{0}$>	0	0.43	0.46	0.22
{1$\underline{2}$10}1/3<1$\underline{0}$1$\underline{1}$>	0	0.23	0.23	0,32
{1$\underline{2}$10}1/3<20$\underline{2}$1>	0	0.34	0.34	0.32
{10$\underline{1}$2}1/3<1$\underline{2}$10>	0	0.36	0.36	0.31
{10$\underline{1}$2}1/3<1$\underline{0}$1$\underline{1}$>	0.46	0.34	0.46	0.24
{11$\underline{0}$2}1/3<20$\underline{2}$1>	0.33	0.50	0.33	0.38
{10$\underline{1}$4}1/3<1$\underline{2}$10>	0	0.26	0.26	0.33
{10$\underline{1}$4}1/3<2021>	0.48	0.36	0.48	0.24

-Finaly we have also performed a few compression tests at constant strain rate without confining pressure and at various temperatures. We have tested only the orientation parallel to c (Z). Below 700 °C the samples are brittle; they break for relatively low stresses without any plastic deformation. From this temperature and above, the samples exhibit a ductile behaviour with some unstabilities on the stress-strain curves while the corresponding flow stress remains quite low (Fig. 11b). No yield phenomenon is observed.

In TEM the prominent feature in the samples deformed at 700 °C without confining pressure is a large and homogeneous density of ferroelectric domains which have been nucleated during the cooling stage without poling electric field. One also observes numerous twins in three of the twinning planes {10$\underline{1}$2} (see Fig. 12a). They correspond to the ones with a large Schmid factor. No free dislocations

are obseved; plastic deformation thus occurred by mechanical twinning only. This accounts for the observed unstabilities on the stress-strain curves. These experiments also show that twins can be nucleated at rather low stresses at elevated temperature (700 °C corresponds roughly to half the melting temperature in Kelvin). The potential glide system {0001}1/3<2110> which presents the highest Schmid factor (0.50) for this orientation could not be activated. Its elastic limit is thus larger than the stress necessary to activate twinning.

Among the samples strained at room temperature and under confining pressure, the ones compressed along the X and Y directions present a large and pervasive density of micro-cracks; this has rendered still more difficult the preparation of thin and transparent foils for TEM. In the rare transparent areas we have generally found no defects except at crack tips. This confirms that the recorded deformation was purely elastic up to fracture as expected from the stress-strain curves. The other samples compressed along the Z and O^+ directions are also microcracked (to a less extent) and it has also been difficult to prepare thin foils for TEM. In the samples compressed along the c axis, we mostly find thin twin lamellas (see Fig. 12a). This result is in complet agreement with our previous analysis of the Schmid factors of the various possible glide systems. In the samples compressed along the O^+ direction we observe a dislocation content which appears to be extremely heterogeneous; relatively large regions are dislocation free while, in some others, one observes bands of entangled dislocations; their density is so large that individual contrasts cannot be resolved (see Fig. 12b) and it has not been possible to fully characterize them. However the orientation of the bands is consistent with a glide plane {1014}. No twinning has been detected in these samples. Again one time these observations are consistent with our previous analysis based on the Schmid factors of the potential glide systems. It thus seems that only the glide system {1014 }1/3<1210> could be activated at room temperature; its elastic limit is quite high (1000 MPa). It is not the same as the one activated by micro-indentation. Twinning is another possible deformation mode at room temperature which is activated for similar stresses. At high temperature, twinning on the {1012} planes appears to be an easier deformation mode than dislocation glide. This is an uncommon behaviour, twinning being generally considered as a low temperature deformation mode (occurring under high stress) while dislocation glide occurs at higher temperature and at a lower stress level. For comparison recent deformation experiments on lithium niobiate single crystals at elevated temperature have shown that dislocation glide is activated only for temperatures of the order of 0.8 times the melting temperature [31].

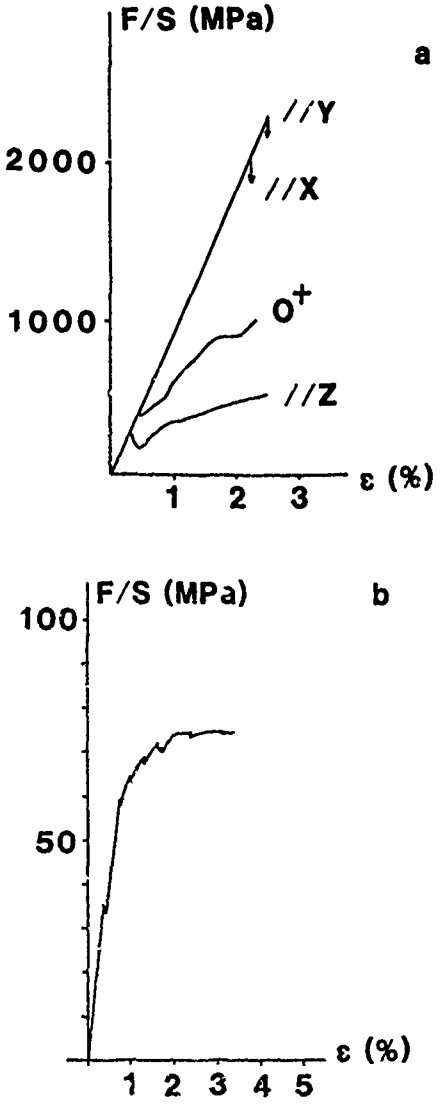


Figure 11: Stress-strain curves. a) deformation at room temperature under a confining pressure of 600 MPa; the orientation of the compression axis is reported on each curve; the imposed strain rate is 10^{-5} s^{-1}. b) deformation at 700 °C at atmospheric pressure; same strain rate; compression orientation c.

IV DISCUSSION

It is clear that the above experimental characterisation of the lattice defects in lithium tantalate still is incomplete. However a few points have been clearly established. The occurrence of point defects is inherent to the non stoichiometry of the uncongruent melt composition and cannot be suppressed. The density of as-grown dislocations is generally reasonably low and the preparation techniques used in this study induce defects (mostly dislocations) into superficial layers not thicker thant 10 µm approximately. Finally the poling process

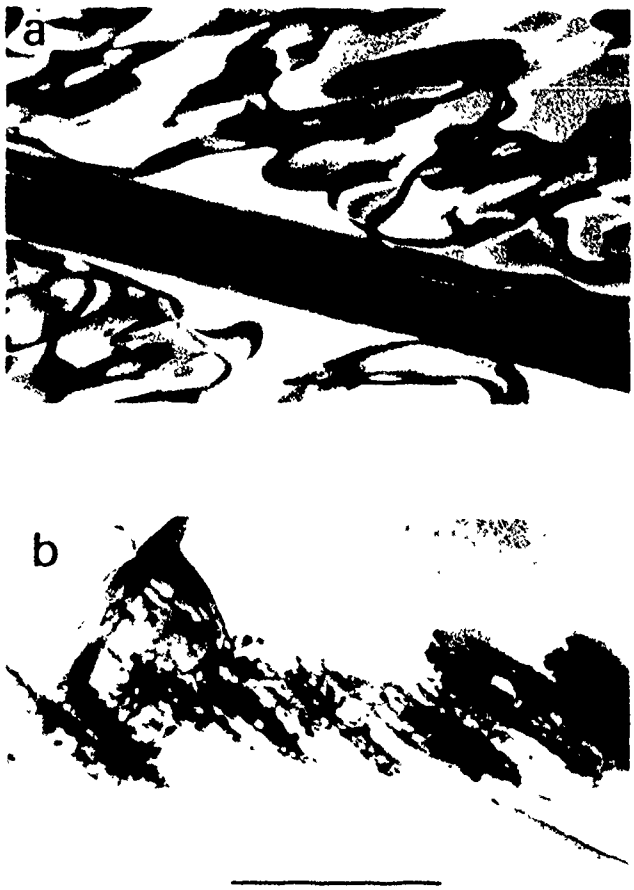


Figure 12: Dislocation micostructures in the deformed samples; scale bar = 1 μm. a) deformation at 700 °C at atmospheric pressure; the prominent features are ferroelectric domains and twin lamellae, no free dislocations are observed. b) deformation at room temperature under a confining pressure of 600 MPa; sample compressed along the Z direction, dislocation band in (10$\bar{1}$4).

prevents the occurrence of ferroelectric domains in the bulk. Therefore the performances of $LiTaO_3$ piezoelectric devices should not be strongly affected by any of these lattice defects as long as their density remains moderate. It is however to be remembered that at least some types of dislocations are able to nucleate ferroelectric domains around them which cannot be removed by poling. If the density of as-grown dislocations is large or if a large density of dislocations is introduced during machining, a large density of ferroelectric domains connected to these dislocations can affect dramatically the piezoelectric performances of the corresponding devices. This ability for the dislocations to nucleate ferroelectric domains is briefly considered now.

Let us consider a straight dislocation lying along the axis of a cylinder; the elastic stress and strain fields σ and ε of this dislocation generate the electric fields **E** and **D** coupled to them by the well known Voigt equations [32].

$$\sigma_{ij} = \sum_{kl} C_{ijkl} \frac{\partial U_k}{\partial x_l} + \sum_{m} e_{mij} \frac{\partial V}{\partial x_m} \qquad (1)$$

$$D_m = \sum_{kl} e_{mkl} \frac{\partial U_k}{\partial x_l} + \sum_{j} \chi_{mj} \frac{\partial V}{\partial x_j}$$

C_{ijkl}, e_{mkl} and χ_{mj} are the components of the stiffness tensor, of the piezoelectric tensor and of the dielectric tensor of the material respectively. **u** (components u_k), **D** (components D_m) and V are the displacement vector, the electric exitation field and the electrostatic potential respectively. Within a ferroelectric domain C_{ijkl}, e_{mlk} and χ_{mj} are constant but when one crosses a domain boundary they change abruptly. As the dislocation lies along the axis of the cylinder σ, **u** and V are independent of the corresponding coordinate (along the axis). The equilibrium conditions are

$$\sum \frac{\partial \sigma}{\partial x} = 0 \text{ (mechanical equilibrium)} \qquad (2a)$$

$$\sum \frac{\partial D}{\partial x} = 0 \text{ (electric neutrality)} \qquad (2b)$$

The boundary conditions have also to be considered. The external surfaces of the cylinder are stress free On the lateral surface with unit normal n_j, one has

$$\sum \sigma_{ij} n_j = 0 \qquad (3\,a)$$

Furthermore the Burgers circuit theorem imposes the following condition

$$\oint_{\Gamma} du_i = b \quad (i = 1 \text{ to } 3) \qquad (3b)$$

where the b_i's are the components of the Burgers vector and Γ is any closed circuit enclosing the dislocation line.
This electro-elastic problem can be completely solved, in principle, if one knows the displacement vector **u** and the electrostatic potential V at any point, the stress tensor σ and the electric exitation **D** being thus computed with eqs (1) The shape of the domain boundary has to be computed in minimizing the Gibbs energy of the system constituted by the cylinder with the dislocation along its axis. The elastic energy part of the Gibbs energy G is given by the volume integral

$$G_{elast} = \iiint_V \frac{1}{2} \sum_{ij} \sigma_{ij} \frac{\partial u_i}{\partial x_j} dV \qquad (4a)$$

where V is the volume of the cylinder. This integral has in fact to be divided into two parts, one corresponding to the "normal" material which occupies the volume V_n while the other one which occupies the volume $V-V_n$ corresponds to the ferroelectric domain.

The electrostatic energy part of G is also given by a volume integral extended to the total volume V (which is also to be divided into two parts V_n and $V - V_n$)

$$G_{electrostat} = \iiint_V \frac{1}{2} \sum_j D_j \frac{\partial V}{\partial x_j} dV \qquad (4b)$$

The last part of G is the surface energy of the domain boundary; this later term is very difficult to calculate because the boundary energy density can vary with the surface orientation. A rough estimate may be obtained in assuming that it is isotropic; under these conditions the surface energy is simply the product of the total domain boundary surface times some constant representing the mean surface energy density. The next step of the calculation consists in minimizing the total Gibbs energy. The complete calculation would be extremely complicated and would not be very precise because the surface energy density of the domain boundary is not known.

There exists however a simplified method. We assume that there is no domain and we compute the fields generated by the dislocation, σ, u and V; They can be computed in using a method developed by a number of authors [33-39]. The electric field $E = -\text{grad } V$ is then easily computed at any point around the dislocation. If, in some region, E is found to be large and roughly opposite to the spontaneous polarization, we assume that a ferroelectric domain is nucleated without any appreciable change of the stress and the strain fields of the dislocation. Computations of E around various types of dislocations have recently been published by Cerva et al. [40] for $LiNbO_3$. Their results can be used here as a correct approximation of the expected electric field around dislocations in $LiTaO_3$ which is a material with very similar properties and physical constants. These authors find that E can become huge in the immediate vicinity of some dislocation cores (the modulus of E reaches values as high as 10^7 V cm^{-1} at distances of the order of 1000 Å). Similar values are probably reached in $LiTaO_3$ for the same dislocation configurations and we suppose that the cases with the largest E values opposite to the polarisation (i.e. opposite to the c axis) lead to ferroelectric domains around the corresponding dislocations. Examination of Fig. 2 of Cerva et al. [40] shows that very large E's occur for edge dislocations with a Burgers vector 1/3<1$\underline{1}$01> and the domain boundaries defined as the locus of |E|= some constant value would exhibit a butterfly shape similar to the one experimentally observed (Fig. 10b). In contrast E around the screw dislocations with the same Burgers vector is more symetrically distributed and its modulus is markedly lower; it would thus not lead to domains. It is clear that the above considerations render account of the TEM observations only at a rough and qualitative level. However they explain how dislocations introduced by mechanical grinding can stongly affect the piezoelectric properties of a device although these dislocations do not penetrate deeply beneath its surface.

REFERENCES

[1] HALLIBURTON L.E. and MARTIN J.J. "Precision frequency control" ed. by E.A. GERBER and A. BALLATO Vol. I Acoustic resonators and filters Academic Press, Newyork p 1 (1985)

[2] ABRAHAMS S. C. and BERNSTEIN J. L. J. Phys. Chem. Sol. **28** 1685 (1967)

[3] ABRAHAMS S. C., HAMILTON W. C. and SEQUEIRA A. J. Phys. Chem. Sol. **27** 1693 (1967)

[4] MEGAW H.D. Acta Cryst. **A 24** 583 (1968)

[5] MEGAW H.W. Crystal structures; a working approach Saunders p.223 (1973)

[6] HEYRING L. and O' KEEFFE M. The chemistry of extended defects in non metallic solids North Holland (1970)

[7] DE RIDDER R., VAN LANDUYT J. and AMELINCKX S. Phys. Stat. Sol. (a) **9** 551 (1972)

[8] AINES R.D. and ROSSMAN G.R. J. Geophys. Res. **89** 4059 (1984)

[9] BARNS R.L. and CARRUTHERS J.R. J. Appl. Cryst. **3** 395 (1970)

[10] NOLF M., COQUILLE R., LEBRET A., AUVRET P., LANCON R. and DETAINT J. Rapport technique CNET CPM/DEF/ 1592 (1976)

[11] BALLMAN A. A., LEVINSTEIN H.J., CAPIO C.D. and BROWN H. J. Amer. Ceram. **50** 657 (1967)

[12] MIYAZAWA S. and IWAZAKI H. Review Elec. Comm. Lab. **21** 374 (1972)

[13] HERRINGTON J.R., DISCHLER B., RAUBER A. and SCHNEIDER J. Solid State Commun.**12** 351 (1973)

[14] BOLLMANN W. and STOHR H.J. Phys. Stat. Sol. (a) **39** 477 (1977)

[15] BOLLMAN W. Cryst. Res. Technol. **22** (**4**) K67 (1987)

[16] FORSTER A., KAPPHAN S. and WOHLECKE M. Phys. Stat. Sol.(b) **143** 755 (1987)

[17] PATERSON M. S. Bull. Minéral. **105** 20 (1982)

[18] SMITH R.G., FRASER D.B., DENTON R.T. and RICH T.C. J. Appl. Phys. **39** 4600 (1968)

[19] BURSILL L. A. and PENG J. L. Phil. Mag. **A 46** 1 (1982)

[20] NASSAU K., LEVINSTEIN H.J. and LOIACONO G.M. J. Phys. Chem. Sol. **27** 983 (1966)

[21] VENABLES J.D. J. Appl. Phys. Letters **25** 254, (1974)

[22] OHNISHI N. and LIZUKA T. J. Appl. Phys. **46** 1063 (1975)
[23] BLISTANOV A.A., NOSOVA I.V. and TAGIEVA M.M. Soviet. Phys. Crystallogr. **20** 409 (1975)
[24] ANDERSON S. and HYDE B. G. J. Solid State Chem. **9** 92 (1974)
[25] LEE W.E., SANFORD N.A. and HEUER A.H. J. Appl. Phys. **59** 2629 (1986)
[26] NASSAU K. and LINES M.E. J. Appl. Phys. **41** 533 (1970)
[27] FENG D. Cryst. Latt. Def. and Amorph. Mat. **12** 229 (1985)
[28] WESTBROOK J.H. and CONRAD H. (eds) "The science of hardness testing and its reseach applcations" . Amer. Soc. Metals, Metals Park (1971)
[29] BOULOGNE B., FRANCOIS P., CORDIER P., DOUKHAN J.C., PHILIPPOT E. and JUMAS J.C. Phil. Mag. A **57** 411 (1988)
[30] ALEXANDER H. and HAASEN P. Solid State Physics **22** 27 (1968)
[31] FRIES E. and PETER A. Revue Phys. Appl. **22** 1353 (1987)
[32] CADY W.G. "Piezoelectricity" McGraw Hill (1946)
[33] ESHELBY I.D., READ W.T. and SHOCKLEY W. Acta Metall. **1** 251, (1953)
[34] MERTEN L. Phys. Kondens. Mater. **2** 53 (1964)
[35] MERTEN L. Phys. Kondens. Mater. **2** 66 (1964)
[36] KOSEVICH A.M., PASTUR L.A. and FELDMAN E.P. Soviet Phys. Cristallogr. **12** 797 (1968)
[37]SAADA G. Phys. Stat. Sol.(b) **44** 717 (1971)
[38] FAIVRE G. and SAADA G. Phys. Stat. Sol. (b) **44** 127 (1972)
[39] TREPIED L. J. Physique **39** L433 (1978)
[40] CERVA H., PONGRATZ P. and SKALIKI P. Phil. Mag. **54** 199 (1986)
[41] DOYLE P.A. and TURNER P.S. Acta Crystallogr. A **24** 390 (1968)

43rd Annual Symposium on Frequency Control - 1989

UPDATE ON POSSIBLE ELECTRODE MECHANISMS IN THE SWEEPING OF ALPHA QUARTZ

J. G. Gualtieri, L. Calderon, and R. T. Lareau
U.S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

Abstract

In the air-sweeping of alpha quartz, porous electrodes can facilitate electrochemical reactions which enhance the indiffusion of hydrogen [1]. The use of thick nonporous Au/Cr electrodes can lead to a change in the sweeping mechanism. Under these conditions, a color-center front advances from anode to cathode in quartz that is swept in air at temperatures below the α - β transition.

Evidence is presented which indicates that: 1) color centers form beneath nonporous portions of the anode, that correlate with the formation of interface Cr-oxides and silicides; 2) clear striae are created along the c-axis where color centers either do not form or are subsequently bleached along cylindrical volumes which correlate with openings in the anode; 3) the clear cylindrical volumes have diameters much larger than the metal openings; 4) a uniformly colored region can form in areas close to but not covered by the anode. Also, a clear band forms between the uniformly colored region and the striated region associated with the anode.

These observations suggest that: 1) openings in the anode metallization are required for hydrogen introduction during air-sweeping; 2) the electrochemical charge-transfer reaction, responsible for proton introduction into the quartz, takes place at the gas-electrode-quartz boundary and spreads to the electrode-quartz boundary by interfacial diffusion of hydrogen species; 3) the electrode mechanism responsible for the formation of color centers in air-sweeping is a 2-phase (electrode-quartz interface) reaction, influenced by the suppression of interfacial H-diffusion; 4) hydrogen can also be introduced at the gas-quartz interface adjacent to the anode. The extension of both the uniform color center and hydrogen-introduction areas to regions close to the anode suggests electron diffusion from these regions to the anode.

Introduction

Very little is known about the mechanisms occurring at the anode and cathode during the sweeping of alpha quartz.

All solid-state electrode materials used in anodes and cathodes for electrochemical processes, e.g., fuel cells, chemical power sources, and electrolyzers make use of porous electrodes to raise their operating efficiency. The porosity forms a multitude of gas--electrode-electrolyte interfaces. These interfaces are required because transport of reactants through a metal/metal oxide electrode would be too slow.

The study of interfacial phenomena in fast-ion transport materials, such as the solid-oxide-based electrolytes, is instructive for the understanding of the quartz sweeping-mechanisms problem. In these electrolytes, the conduction at high temperature in air or oxygen is ionic and is the result of the migration of oxygen ions from cathode to anode by a vacancy mechanism. On the other hand, quartz may be considered a fast-ion conductor in the c-axis direction at high temperature. Electrical conduction is also ionic, but is due to the migration of proton and mainly alkali-metal interstitials from anode to cathode. Nevertheless, similarities with respect to electrode (interfacial) phenomena are evident and one can draw upon the abundant scientific literature associated with the solid-oxide-based electrolytes [2-7].

Previously, we reported that, in air-swept quartz, hydrogen indiffusion was enhanced by porous electrodes that could facilitate 3-phase (gas-electrode-quartz) interfacial electrochemical reactions. We also reported that the use of thick Cr-electrodes resulted in a change in the mechanism of sweeping. Under this condition, color-center striae were observed (along the c-axis) in quartz being swept in air at temperatures below the α - β transition. We suggested that the atmosphere was not penetrating the thick Cr-metallization, i.e., there was less electrode porosity [1].

The accepted model of the color center in alpha quartz is that of a trivalent aluminum impurity in a substitutional silicon site which is charge compensated by a positive electron-hole [8]. Interstitial monovalent positive ions such as alkali-metal ions or hydrogen (protons) usually provide the charge compensation. Under ionizing radiation, the $[AlO_4]^-$ traps a hole at an adjacent nonbonding oxygen p orbital. The $[AlO_4/h^+]^0$ aluminum-hole center correlates with optical absorption bands at 620 nm and 460 nm; [9] these bands are responsible for the smoky coloration produced. The released electron and interstitial ions do not drift very far, because annealing at 350°C results in recombination of electrons and holes and recapturing of interstitial compensators [8]. The color centers are bleached in this process.

King [10] reported that crystals that have been vacuum swept produce resonators that have negligibly small frequency transients under intense ionizing radiation. Vacuum sweeping induces color centers that are thermally stable into the quartz, that is, resistant to bleaching up to 1000°C [11]. This stability is probably due to a permanent removal of the extra electron and compensating ion from the Al center.

In this report, we present experimental observations that reveal: 1) a correlation between anode nonporosity and color-center formation; 2) the relationship of porosity, or openings, in the anode to hydrogen introduction, which result in clear striae in color-center regions; 3) that surface diffusion along the electrode-quartz interface contributes to the enlargement of the hydrogen introduction area; 4) the extension of the color-center introduction area to regions close to but not covered by the anode; and 5) the relationship of the latter to the introduction of hydrogen in the extended region.

Observations of clear striae in color-center regions and the extension of the color-center introduction area have been reported by others [11,12]. The objective of this study was to increase our understanding of the influence of the electrode-quartz interface on sweeping. We hope that increased knowledge of the mechanisms of hydrogen and color-center introduction will lead to improvements in sweeping technology.

Experimental Methods

Cultured quartz was used exclusively in this investigation. The quartz was lumbered into Y-bars after seed removal [13]. The Z-surfaces of the sample bars were lapped and polished as outlined previously [14]. All surfaces were cleaned using trichloroethane, hot detergent, alcohol and distilled water. As a final step, samples were cleaned in an oxygen plasma for 2 minutes prior to electroding. For electrodes, evaporated metallizations were applied to the cleaned and polished Z-surfaces of the lumbered bars. Temperature-ramped sweeping was carried out in a flowing-air atmosphere [15]. The flow rate was 100 cc/min. A Perkin-Elmer model 1760 Fourier transform infrared spectrophotometer was used to obtain room-temperature OH-vibrational spectra.

Analysis of metallized quartz, before and after air-sweeping, was performed utilizing secondary ion mass spectroscopy (SIMS) in an attempt to obtain depth profiles. The SIMS analyses were obtained using a quadrupole-based ion microanalyzer (PHI 6300) at ETDL. The primary Cs ion beam was rastered over a 400 μm by 400 μm area, with ejected secondary ions collected from the center 9 percent of the sputtered crater. The ion impact energy of the primary Cs ions was 9 keV, with an angle of incidence of 60 degrees to the surface normal; electron neutralization was used to compensate for electrical charging when profiling through the insulating substrate. The system vacuum was approximately 8 X 10^{-10} torr when performing the sputter depth profiling, hence the ambient background for H and O was much lower than the levels observed in the analyzed materials.

Experimental Observations

Color Centers Generated Using Air-Sweeping

Normally, when sweeping is performed in air, color centers do not form. Instead, interstitial alkali ions trapped adjacent to substitutional aluminum centers are replaced by the formation of an Al-OH center. We have observed color-center formation when sweeping with thick (1000 Å) evaporated Cr, Al, and V metallizations. In Fig. 1(A), we show a photomicrograph of a Au/Cr (1000 Å/1000 Å) anode as viewed from the cathode side, looking along the c-axis through the quartz material.

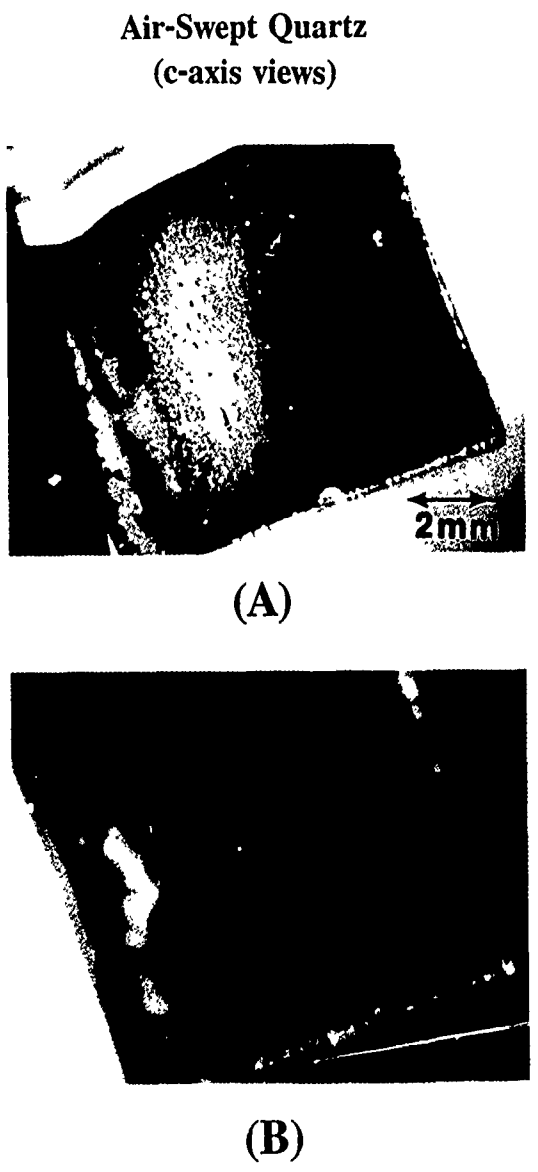


Fig. 1 Photomicrograph of an evaporated Au/Cr anode surface after sweeping. (A) Backlighted view through the quartz bar in the c-axis direction. (B) The same area as in (A), using side illumination to reveal color centers.

The anode is back-lighted to reveal both dark and light areas. The light areas indicate that extensive porosity has developed during the sweeping process.

In Fig. 1(B), the illumination is perpendicular to the viewing direction, making the color centers more visible. Comparison of Figs. 1(A) and (B), reveals a definite correlation between electrode nonporosity and color-center formation. Examination of the Au/Cr anode after sweeping, using powder x-ray diffraction, revealed the formation of amorphous Cr_2O_3 in electrode surface areas which were colored green. These green areas were bonded strongly to the quartz surface and could not be removed with a razor blade. Polishing with a diamond abrasive compound was required to remove these strongly bonded areas. After polishing, the underlying quartz surface was found to be pitted extensively. Areas not showing Cr_2O_3 formation (brown areas) were removed easily with a rubber eraser. In this case, the underlying surface was smooth. The reason for the formation of patterns of strongly adherent and weakly adherent Au/Cr films on Z-surfaces, after air-sweeping, has not been determined.

Color-Center Areas Containing Clear Striae

Large preexisting openings in the metallization are related to the observation of clear striae (along the c-axis) in smoky colored regions. The striae appear as large-diameter clear cylinders. In Fig. 2(A) and (B), a one-to-one correspondence of preexisting anode openings (A), clear cylindrical regions in color-center areas (B) and clear cylindrical regions continuing through areas exhibiting light scattering or gray haze, (C), can be seen. The clear cylindrical regions are much larger than the openings in the metal, varying in diameter from about 0.1 mm to about 0.6 mm. Clear striae can also be observed in regions of the quartz which were probably covered by anodic areas suffering metal recession during the prolonged high temperature phase of the sweeping process. Metal recession was especially noticeable on cathode surfaces in oxidizing environments. The nonuniform appearance of the color-center regions in Fig. 1(B) are indicative of porosity-induced bleaching of color centers.

Uniform Color-Center Areas

We have observed color centers forming outside the anode region, i.e., in regions close to but not covered by the anode. These regions are uniformly colored and contain no clear or color-center striae. Between these regions and the anode boundary a clear band develops which, in the sample shown in Fig. 3, is about 0.6 mm wide. The X-axis view shows the same clear band and color-center regions. Note that the striated color-center region on the left (under the anode) is separated from the uniformly colored region on the right by a clear band and that the uniform color-center front displays a pronounced curvature.

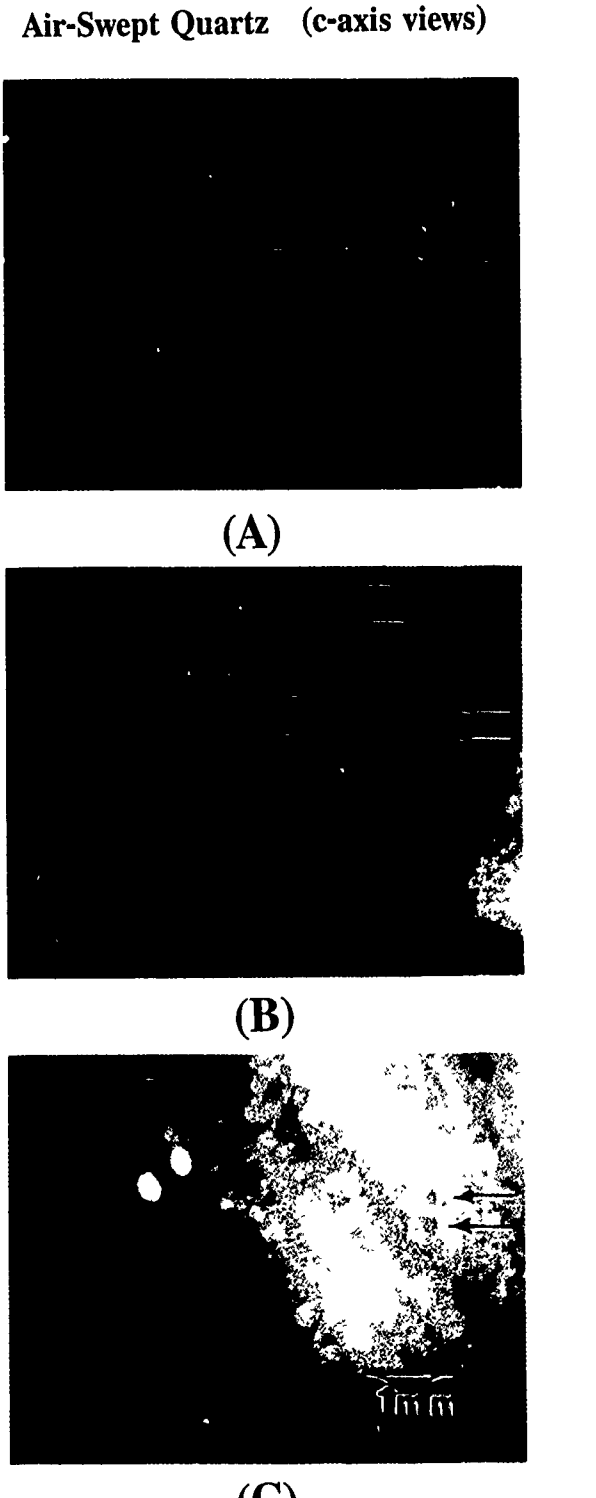


Fig. 2 Photomicrograph of (A), an evaporated Au/Cr anode surface after sweeping. This view shows holes in the metallization (indicated by arrows). (B) The same view as in (A), using side illumination to reveal color centers and clear cylindrical striae. Note the correspondence of holes in (A) with clear areas in (B). (C) The same view as in (A) and (B), using intense side illumination to reveal light-scattering haze. Note, that two of the clear cylindrical regions continue through haze above the color centers.

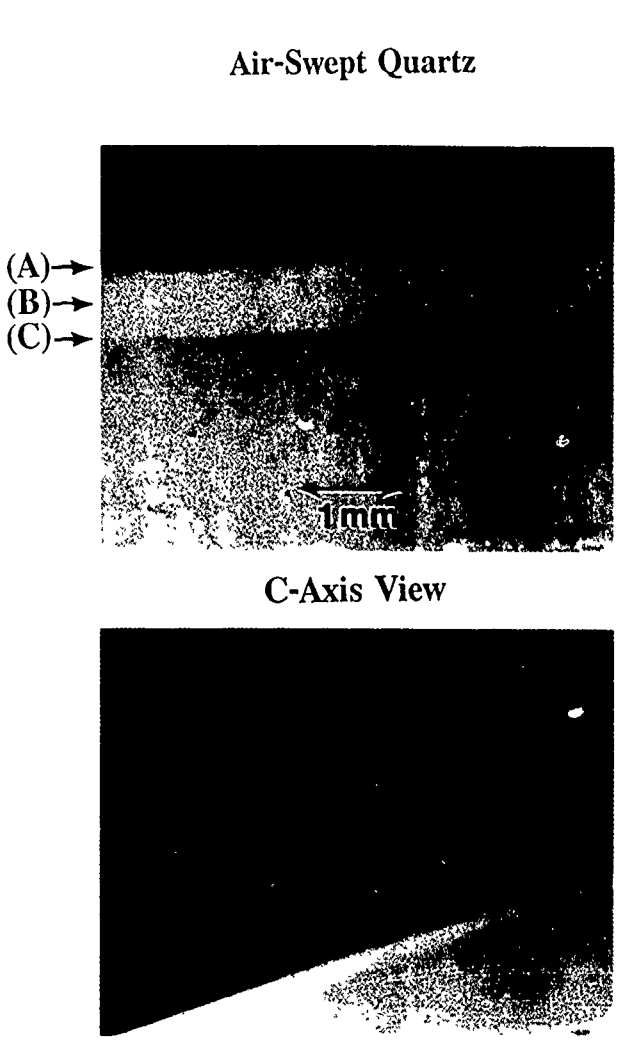


Fig. 3 Photomicrograph of a Au/Cr anode boundary on a quartz Z-surface. The c-axis view shows a clear band (border) adjacent to the metal boundary. (A) is the electrode boundary, (B) is the clear band, and (C) marks the start of the uniformly colored region. The X-axis view shows the uniformly colored region (small dark area on the right). This region has a curved front and is separated from the striated region on the left by a clear band.

Discussion of the Observations

Hydrogen Introduction

The suggested mechanism for the introduction of hydrogen during air-sweeping involves the electrolysis of water. The overall anodic reaction:

$$H_2O\ (g) \text{ -----> } 2\ H^+ + 1/2\ O_2\ (g) + 2\ e^-\ ,$$

requires a 3-phase boundary [1]. The appearance of clear striae in otherwise colored regions of the quartz bar, when thick Cr-metallizations are used, indicates that hydrogen is being introduced at areas of the electrode which are open to the atmosphere (electrode porosity). The correspondence of clear cylindrical regions with openings in the metallization supports this view. If an infrared probe beam, much larger than one of these clear cylinders is used for spectral analysis, the region analyzed will include both electron-hole and H^+ types of Al-compensation. To show that hydrogen is being introduced at the position of the clear cylinder, the IR beam was apertured down to approximately the size of the largest clear cylinder we could find (0.6 mm). Fortunately, FTIRs have circular cross-sectional beams, high optical throughput, and low beam divergence, permitting reduced aperture analysis.

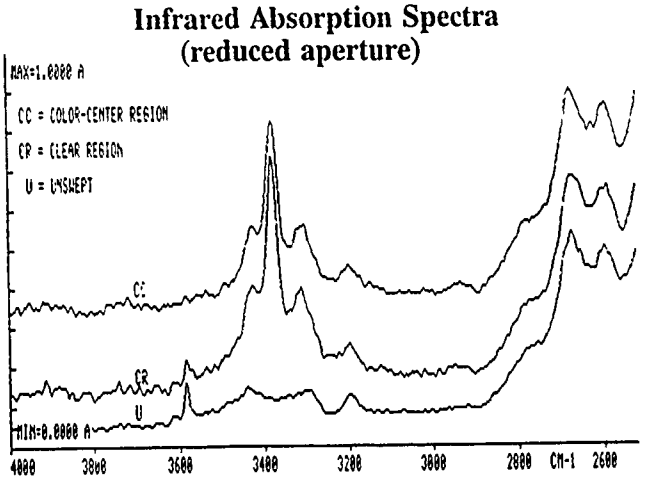
The required aperturing was accomplished with some tradeoff in the signal-to-noise of the IR spectrum. We were able to compare the IR spectrum of one of these clear cylinders to an adjacent mixed (colored and clear) region. The spectra are shown in Fig 4. In the clear region, sweeping occurred without much depletion of as-grown OH at 3585 cm^{-1}. In the mixed region, depletion of the 3585 cm^{-1} band is evident. Electrolytic formation of color centers implies that an external source of hydrogen was not available. Under this condition, the crystal will use as-grown OH to compensate Al centers. An infrared spectrum of electrolytically colored quartz will show an increase in the Al-OH band at 3378 cm^{-1}, but will also show depletion of the as-grown OH band at 3585 cm^{-1} [1,16]. When an external source of hydrogen is available, depletion of the as-grown band ceases and the Al-OH band continues to increase [1,16]. Also, we have found that sweeping in air at 300°C, depletes as-grown OH, with a very slight increase in Al-OH. However, sweeping at 400°C results in a similar decrease in as-grown OH and a noticeable increase in Al-OH. (See Fig. 5.) It seems that the introduction of hydrogen, probably beginning at 400°C, arrests the depletion of as-grown OH. Thus, we conclude that hydrogen has entered the quartz only in the clear (cylindrical) regions.

Interfacial Diffusion of Hydrogen Species

The clear-striae cylinders are much larger in diameter than the openings in the metallization. (See Fig. 2(A) and (B).) This indicates that some hydrogen species is diffusing interfacially to enlarge the hydrogen-introduction area. The hydrogen species could be molecular H_2O, H_2, or, even, OH^- and H. Exclusive three phase introduction would be too constricting, since it is essentially one-dimensional, that is, the electrode, quartz, and gas phase meet along a linear region. Two-phase introduction, through interfacial diffusion in all directions of the electroded quartz Z-surface, enlarges the (two-dimensional) area over which hydrogen can enter the quartz. Apparently, this mode of hydrogen introduction is noticeable only when strongly adherent electrode films are employed and color centers develop. Weakly adherent films would be expected to favor interfacial

diffusion of hydrogen species and suppress the formation of color centers. However, these films will suffer metal recession at high sweeping temperatures [7].

Clear-striae cylinders form inside the quartz because of the highly directional nature of the interstitial electrodiffusion of H^+ and alkali-metal ions. Thus, the cylindrical bleaching of color centers in the quartz bulk suggests that the principal hydrogen-introduction mechanism is through interfacial diffusion of hydrogen species.



(A)

(B)

Fig. 4 A) Infrared absorption spectra (displaced for clarity) of a clear cylindrical region and an adjacent, mixed clear and colored region compared to unswept. Room temperature spectra were taken along the c-axis using a 0.6 mm diameter circular aperture. B) Photomicrograph of the clear cylindrical region (CR) and the adjacent clear and colored region (CC) analyzed.

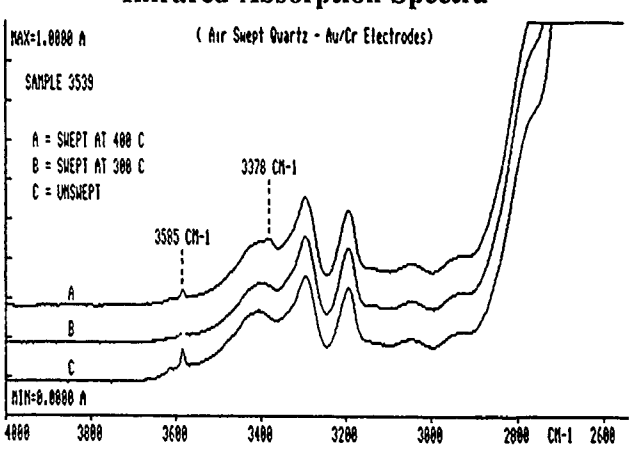


Fig. 5 Room temperature infrared absorption spectra of: C) unswept quartz and similar samples swept in air using Au/Cr electrodes, B) at a maximum temperature of 300°C, and A) at a maximum temperature of 400°C. The Al-OH band at 3378 cm^{-1} is increased and the as-grown OH band at 3585 cm^{-1} is not changed by sweeping at 400°C. The spectra were taken in the X-direction.

This probably occurs after many pores are created by metal recession, exposing the quartz to the ambient gas.

Studies of the oxidation behavior of chromium between 300° and 600°C, revealed a transition at 400°C. At the transition, chrome oxide nodules and ridges develop to relieve stress, providing a high density of mass transport paths [17]. This oxidation transition could be related to the onset of metal recession and the noticeable increase in hydrogen introduction at 400°C, cited above.

It is possible to estimate the interfacial diffusion coefficient from the mean free path of adsorbed species (clear-cylinder radius) from $x \cong (Dt)^{1/2}$. The largest clear cylinder we found had a radius of 0.03 cm, and the sweeping time logged (over 400°C) was 36 hours. The apparent diffusivity is approximately 7×10^{-9} cm^2/s. We use the term apparent, because the real diffusivity must be greater, whereas the diffusivity we observe is retarded by the draining of the interfacial diffusant [18] into the bulk for volume diffusion (sweeping).

Color-Center Introduction

The observations suggest that, in air-sweeping, even at temperatures below the α - β transition, color centers form because the diffusion of hydrogen species is suppressed along the electrode-quartz interface.

R. Pretorius, et. al., [19] have found that when metals react strongly with SiO_2, silicides as well as oxides are formed. The silicides

form on the SiO_2 substrate and oxides form on the metal surface. The suppression of the diffusion of H-species probably results from the reaction of metals such as Cr, Al, and V with the SiO_2 surface. The formation of compounds (in the Au/Cr case) such as Cr_3Si, Cr_5Si_3 [19], and, possibly, Cr_2O_3 act as diffusion barriers.

High Cr content Au/Cr films have been shown to produce amorphous Cr_2O_3 under strong annealing and oxidizing conditions. The green reflectance color of the films after annealing is determined by the oxidation state of the chrome rather than the gold [20]. We have observed that when Au/Cr films are annealed during the sweeping process, Cr migrates to the surface and into the interface and Si migrates out of the quartz and into the interface. The amount of hydrogen in the interface seems to be related to the amount of Si in the interface. Thus, silicide formation relates to the suppression of interfacial diffusion of H-species to entry points in the quartz surface. (See Figs. 6-9.) We have found that the green reflecting areas adhere strongly, are more resistant to metal recession, and contain chrome oxides and silicides (compare Figs. 8 and 9). The strongly adherent interface-silicides are expected to inhibit the surface diffusion of hydrogen species with the likely result that sweeping would proceed by introduction of hydrogen predominantly at or near three-phase points. Presumably, gas-phase polarization builds as a result of the nonporosity and reduced interfacial diffusion. To reduce the polarization, current then forms over the much larger two-phase region (metal-silicide + oxide/quartz interface) as a result of the transition of electrons from the $[AlO_4]^-$ to the anode, with the formation of electron-holes $[AlO_4/h^+]^0$ (smoky color centers) in the quartz.

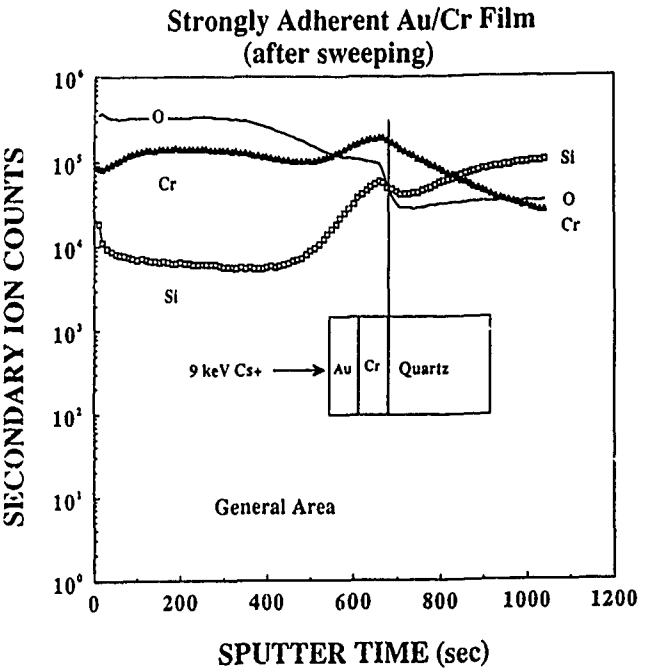


Fig. 7 SIMS of a Cr-quartz interface after sweeping. Note the high yields of Cr, O, and Si in the interface.

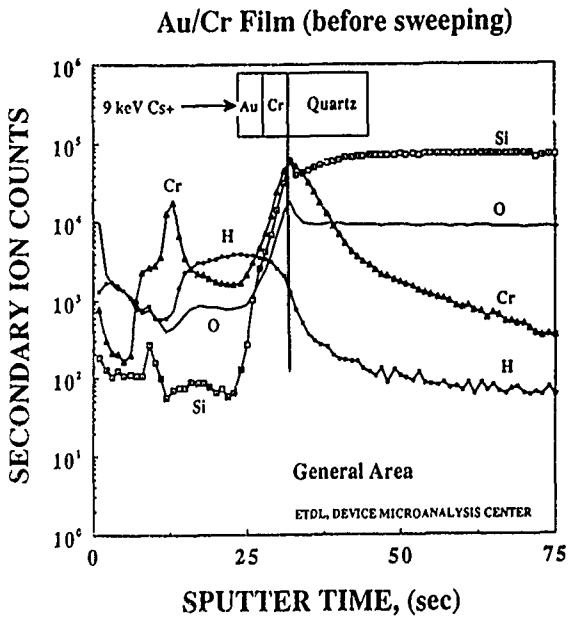


Fig. 6 SIMS of a Cr-quartz interface before sweeping. Note the low yield of Si, O, and Cr in the interface region. The H profile indicates that H-species are present at the interface. The vertical line locates the Cr-quartz interface. A randomly selected (general) area was profiled.

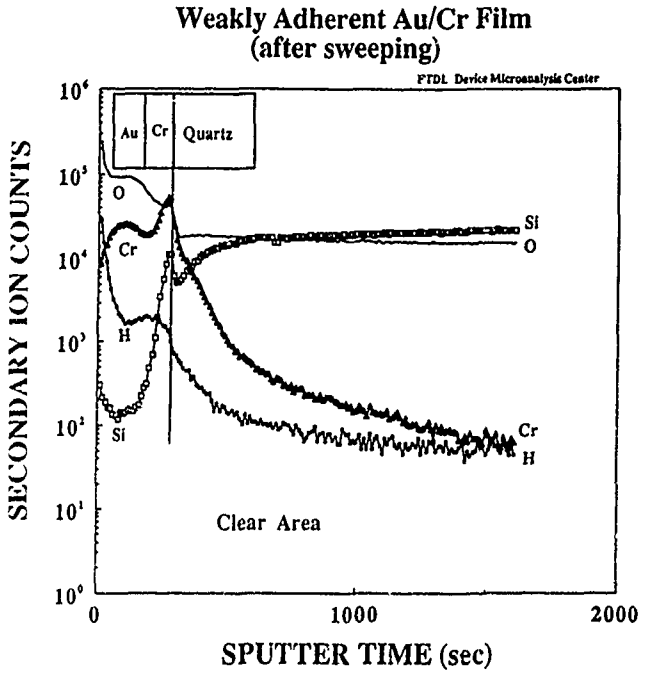


Fig. 8 SIMS of a weakly adherent Au/Cr film after sweeping. Note the low yield of Si and the low yield of H in the interface. The electrode area profiled was located above a clear area of the underlying quartz.

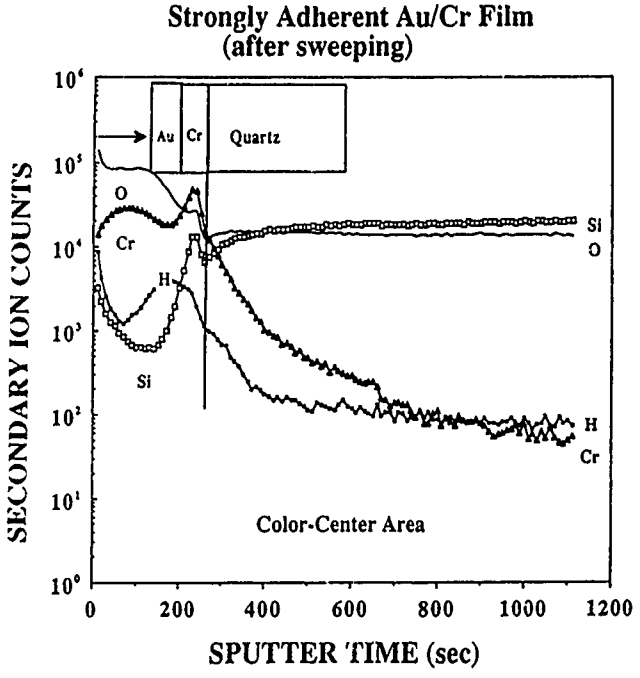


Fig. 9 SIMS of a strongly adherent Au/Cr film after sweeping. Note the high yield of Si and the high yield of H in the interface. The electrode area profiled was located above a color-center area of the underlying quartz.

The color-center front probably advances through the exchange of electrons and holes among the nonbonding 2p orbitals of oxygen atoms [21]. One would not expect this advancement to be as highly directional, as the electro-diffusion of H^+ and alkali-metal ions. That is, the diffusional component of the electron flux may influence the color--center advance more than the electron migration in the electric field. The curvature of the uniform color-center front and the complete absence of any striations in the uniform region also attest to an isotropic type of advancing color-center front. Thus, at any anode-open quartz boundary the color-center front could advance laterally as well as from anode to cathode. If the lateral advance includes any open-quartz surface area, that area could become conducting and active for the introduction of hydrogen directly from the gas phase into the quartz. The sharp boundary that develops between the clear region and the uniformly colored region may be another indication of the highly directional bleaching of color centers by the electro-diffusion of H^+. The sharp drop in infrared absorption due to Al-OH at 3378 cm^{-1} at the boundary, supports this view. (See Fig. 10.)

Conclusions / Recommendations

It appears that electrode porosity and interfacial diffusion of H-species are required for efficient H-sweeping. We have found that some metal films, used as sweeping electrodes, can form areas of strongly adherent metal oxides and metal silicides during

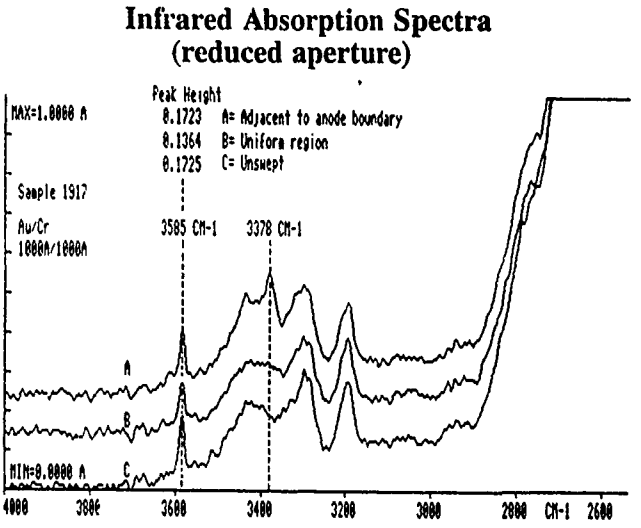


Fig. 10 Room temperature infrared absorption spectra (displaced for clarity) of: A) the clear region adjacent to the anode boundary, B) uniform color-center region, and C) unswept. A 0.6 mm circular aperture was used to analyze along the c-axis direction. The peak heights above background are indicated for each spectrum at 3585 cm^{-1}.

the high-temperature pnase of the sweeping process. These formations act to reduce interfacial diffusion of H-species and promote color-center formation. We have found that color centers form when strongly adherent films of Cr, V, and Al are used. We expect color-centers to form with other strongly adherent metal films, such as, Hf, Zr, Ti, and Nb. Color centers have not been observed when weakly adherent films of either Au, Pt, or Ni were used. We would expect similar results if other weakly adherent films, such as, Cu, Mn, Fe, or Co, were used. Weakly adherent films are subject to recession during the high-temperature phase of the sweeping process. In some cases, the entire metal electrode was lost because of extensive metal recession.

For hydrogen introduction during sweeping, we recommend the use of thin (50-100 Å) strongly adherent base films on polished Z-surfaces in order to maintain electrode porosity and physical and electrical contact with the quartz. Sweeping at lower temperatures (400 - 450°C) will reduce oxide and silicide formations which suppress interfacial diffusion of H-species.

Thick, strongly adherent base films should result in reduced porosity and reduced interfacial diffusion of H-species leading to the formation of hole-compensated Al color centers during air-sweeping at higher temperatures (500-550°C).

Acknowledgements

The authors thank Mr. William Washington for the cutting and polishing of samples, Mr. M.J. Wade for metallizations, Mr. D.W. Eckart for x-ray diffractions, and Drs. J.R. Vig and R.A. Murray for a number of helpful discussions.

References

[1] J.G. Gualtieri, "Possible Electrode Mechanisms for the Introduction of Hydrogen into Alpha Quartz During Sweeping," Proc 42nd AFCS, pp 155-161, 1988.

[2] Da Yu Wang and A.S. Nowick, "Diffusion-Controlled Polarization of Pt, Ag, and Au Electrodes with Doped Ceria Electrolyte," J Electrochem Soc 128, pp 55-63, 1981.

[3] Da Yu Wang and A.S. Nowick, "Cathodic and Anodic Polarization Phenomena at Platinum Electrodes with Doped Ceria as Electrolyte," J Electrochem Soc 126, pp 1155-1165, 1979.

[4] Da Yu Wang and A.S. Nowick, "Polarization Phenomena Associated With Reduction of a Doped Ceria Electrolyte," J Electrochem Soc 127, pp 113-122, 1980.

[5] Hiroaki Yanagida, R.J. Brook, and F.A. Kroger, "Direct Current-Voltage Characteristics of Calcia-Stabilized Zirconia," J Electrochem Soc 117, pp 593-602, 1970.

[6] R.J. Brook, W.L. Pearlman, and F.A. Kroger, "Platinum Electrodes and Calcia-Stabilized Zirconia," J Electrochem Soc 118, pp 185-192, 1971.

[7] C.S. Tedmon, Jr., H.S. Spacil, and S.P. Mitoff, "Cathode Materials and Performance in High-Temperature Zirconia Electrolyte Fuel Cells," J Electrochem Soc 116, pp 1170-1175, 1969.

[8] M.C.M. O'Brien, "The Structure of the Color Centers in Smoky Quartz," Proc Roy Soc (London) A231, pp 404-414, 1955.

[9] E.W. Mitchell and E.G.S. Paige, Proc Phys Soc B67, 262 1954.

[10] J.C. King, "Vacuum Electrolysis of Quartz," U S Patent 3,932,777, 1976.

[11] Gerda B. Krefft, "Effects of High - Temperature Electrolysis on the Coloration Characteristics and OH-Absorption Bands in Alpha Quartz," Radiation Effects 26, pp 249-259, 1975.

[12] J. Leitz and M.R. Hanisch, Naturwissenschaften 46, pp 67-68, 1959.

[13] J.G. Gualtieri and J.R. Vig, "Sweeping and Irradiation Studies in Quartz," Proc 38th AFCS, pp 42-49, 1984.

[14] John G. Gualtieri and Donald W. Eckart, "The Influence of Surface Finish and Metallization on Electrode Electromigration in Alpha Quartz During Sweeping," Proc 40th AFCS, pp 115-120, 1986.

[15] 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.

[16] F. Euler, H.G. Lipson, A. Kahan, and A.F. Armington, "Characterization of Alkali Impurities in Quartz," Proc 36th AFCS, pp 115-123, 1982.

[17] D.J. Young and M. Cohen, "Oxidation Behavior of Chromium between 300 and 600°C," J Electrochem Soc 124, pp 769-774, 1977.

[18] Paul G. Shewmon, "Surface Diffusion from a Point Source," J Appl Phys 34, pp 755-757, 1963.

[19] R. Pretorius, J.M. Harris, and M-A. Nicolet, "Reaction of Thin Metal Films with SiO_2 Substrates," Solid-State Electronics 21, pp 667-675, 1975.

[20] P. Gadenne, C. Sella, M. Gasgnier, and A. Benhamou, "Structural, Optical, and Electrical Properties of Evaporated and Sputtered Au-Cr Thin Films," Thin Solid Films 165, pp 29-48, 1988.

[21] R.C. Hughes, "Time-Resolved Hole Transport in a-SiO_2," Phys Rev B 15, pp 2012-2020, 1977.

43rd Annual Symposium on Frequency Control - 1989

STUDY OF QUARTZ SWEEPING IN α AND β PHASES

J. P. BACHHEIMER and G. DOLINO
Laboratoire de Spectrométrie Physique (associé au C.N.R.S.)
Université Joseph Fourier (Grenoble I)
Boite postale 87 - 38402 Saint-Martin-d'Hères Cedex (France)

ABSTRACT

Defects of quartz are very important both for application considerations and fundamental aspects. In applications, it has been recognized for a long time that the performance of piezoelectric devices are limited by impurities, while in the fundamental field recent studies have shown that the incommensurate phase of quartz is very sensitive to the nature and concentration of impurities. Thus, from both points of view it is important that the defects be controlled. This can be achieved by sweeping of the impurities from electrolysis at high temperature, now an industrial process used to improve the radiation hardness of quartz oscillators. It is usually performed for several days at 500°C far away from the α-INC-β transition (573°C) which could induce fracture. Due to thermal activation of the process, it would be interesting to operate at higher temperature, either in the α phase nearer to the transition temperature or even in the β phase, if the problems related to the α-β transition can be solved.

We present the results of our first investigation on synthetic samples in the α phase at 550°C and in the β phase at 650°C with air and vacuum atmosphere and Pt foil electrodes. The electrical current variation was measured both during sweeping and final cooling, giving activation energies in the α and β phase.

The samples were characterized by I.R. measurements at 77K : a complete disparition of the OH lines was observed after vacuum sweeping in the α phase (550°C - 3500 V/cm) or in the β phase (650°C - 1500 V/cm). Considerable modification of the birefringence variations within the incommensurate phase was found.

The samples which are heated in the β phase become twinned. However we checked that a torque can detwin the major part of a AT cut twinned sample.

INTRODUCTION

The study of the influence of quartz defects is very important both for application considerations and fundamental aspects. It is well known (1) that the performance of piezoelectric devices, particularly resonators, are limited by impurities. The major impurities present in synthetic quartz are alkali ions (M^+) associated to aluminium (Al-M^+ center) which exists as a substitution for silicon and hydrogen related as grown OH defects. It is also known (2) that the α-β transition temperature (~ 573°C) of quartz is sensitive to impurities. In fact this transition is an α-incommensurate-β phases transition. The incommensurate (INC) phase (3) is an intermediate phase which exists on a small temperature interval (~ 1.4K) between the α (low temperature) and the β (high temperature) phases. Very recently (4) a correlation was established between the OH content and the properties of the INC phase. Varying the impurities concentration is therefore interesting in order to study the INC-phase and vice versa. This can be obtained from electrolysis (sweeping) which is now a commercial process used to dissociate and eliminate impurities by means of their displacements under an applied electric field (1000 to 2500 V/cm along the Z direction). Sweeping in an ambiant air atmosphere replaces the alkali ion by an hydrogen to form Al-OH (e_1 - e_2 peaks in I.R. spectra) while sweeping in a vacuum atmosphere forms Al-hole center (5). This is explained by the depletion of as grown OH^- which is not compensated by the introduction of external hydrogen (from atmospheric water). Although both processes improve radiation hardness, the best is not yet known. Usually sweeping is performed for several days in a phase at 500°C away from the transition at 573°C which could induce Dauphiné twinning or fracture. Due to the thermal activation of the process, it would be interesting to operate at higher temperature, either in the α phase nearer to the transition temperature or even in the β phase, if the problem of electrical twinning induced in cooling during the transition can be solved. Indeed Wooster et al. (6) have shown the possibility of detwinning quartz crystals by applying proper stresses (a torque at high temperature seemed to be the most effective). From a process in phase one can expect new effects on defect accomodation and diffusion due to higher temperature and to the more open structure. Dodd et Fraser (7) found that narrow hydrogen lines of the IR spectra are eliminated by heating at T > 1000°C and Krefft (8) demonstrated a strong hydrogen depletion and coloration in the anode side region of natural quartz electrolysed at 700°C. We present the results (sweeping currents, I.R. spectra, effects on the incommensurate phase) from our first electrolysis investigations in the α-phase at 550°C and in the β phase at 650°C with air and vacuum atmosphere and we compare with a 500°C "classical sweeping". The presence of electrical twinning was checked by means of second harmonic generation of light (9).

EXPERIMENTAL

The samples were successively cut from a single cultured (grown several years ago by SICN corporation) quartz bar of medium quality. The impurities content was determined by torch-plasma analysis :

Al	Li	Na	K	Ca	Fe	
1,5	0	2,7	1,6	1,9	0,4	ppm

Typical dimensions were ~ (1//X) x 2 (// Y) x (1//Z) cm^3 for all the samples (pure Z-seed removed) except for the air swept sample at 650° which was ~ 1 cm(//Y). The pressure of Pt electrodes foils on sample Z faces was about 1.8 N/cm^2 (in an air experiment it was checked that multiplying by 10 the pressure has negligible effect on the intensity of the sweeping current). Electrolysis in air was carried out in an open furnace while the vacuum electrolysis was performed in a vacuum furnace operating at 10^{-6} mm Hg. Electrical field was increased slowly, once the sweeping temperature was reached, to limit the current to about 10 mA/cm^2 (except for 650° in air) across the sample. A Lambda 9 Perkin Elmer spectrometer was used to obtain I.R. transmission spectra with non polarized light. Birefringence variations in the incommensurate phase were continuously recorded by means of the de Senarmont method.

RESULTS

- *Vacuum sweeping*

Z faces were finely grounded

Fig. 1 shows the sweeping currents in a phase (500°C) - (a) and in b phase (650°C) (b). At 500°C the current is reduced ~ 100 after 18 days of sweeping. Both currents reach not steady state value.

Fig. 2a shows the as grown spectrum which displays the s_1, s_2, s_3, s_4 peaks of the OH bonds (10). The absorption coefficient at 3500 cm^{-1} is $\alpha(3500 = 0.06 cm^{-1}$.

Fig. 2b shows the result of a "classical" 500°C - 1500 V/cm sweeping for 18 days : s_1, s_2, s_3, s_4 decreases while e_2(Al-OH) appears (10) which demonstrates the effectiveness of the sweeping process.

Fig. 2d shows the spectrum from the second sample swept in the b phase at 650°C : all the s peaks and e peaks disappear, furthermore a strong reduction of the broad OH absorption occurs ($\alpha(3500 = 0.024$ cm^{-1}). This process appears to be stable because no change of this spectrum was observed after the sample was reheated for two weeks in the incommensurate phase. Unfortunatly, after cooling the sample is twinned. We also looked for a better sweeping in the α phase by a subsequent electrolysis of the first sample : after the Z faces were optically polished it took about 2 days to reobtain the final current of fig. 1a, then the temperature was increased to 550°C (1500 V/cm).

CH2690-6/89/000-517 $1.00 C 1989 IEEE

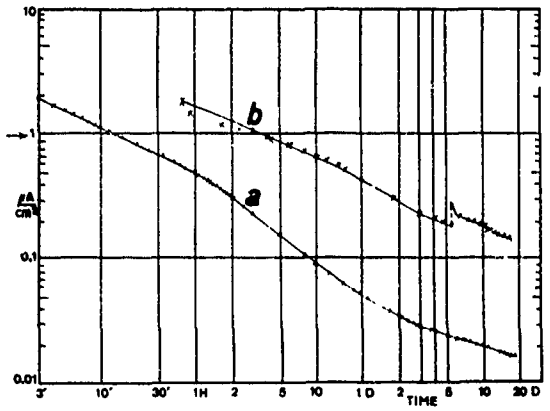


Fig. 1 : Vacuum sweeping :
Sweeping current density vs time for two samples from the same synthetic bar
a) :α phase 500°C-1500 V/cm
b) βphase 650°C - 1270 V/cm till 6 days - 1500 V/cm afterwards

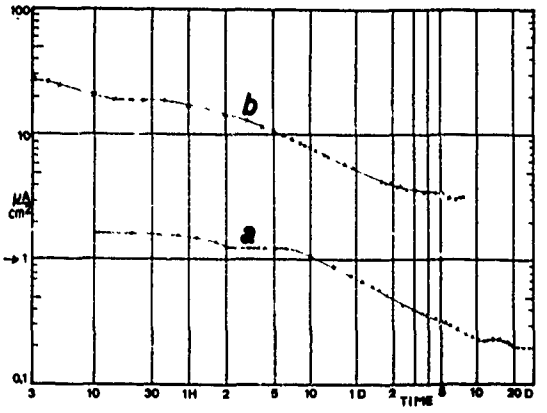


Fig. 3 : air sweeping :
Sweeping current density vs time for two samples from the same synthetic bar as fig. 1.: a :α-phase : 500°C - 2000 V/cm, b :β phase 650°C - 1270 V/cm.

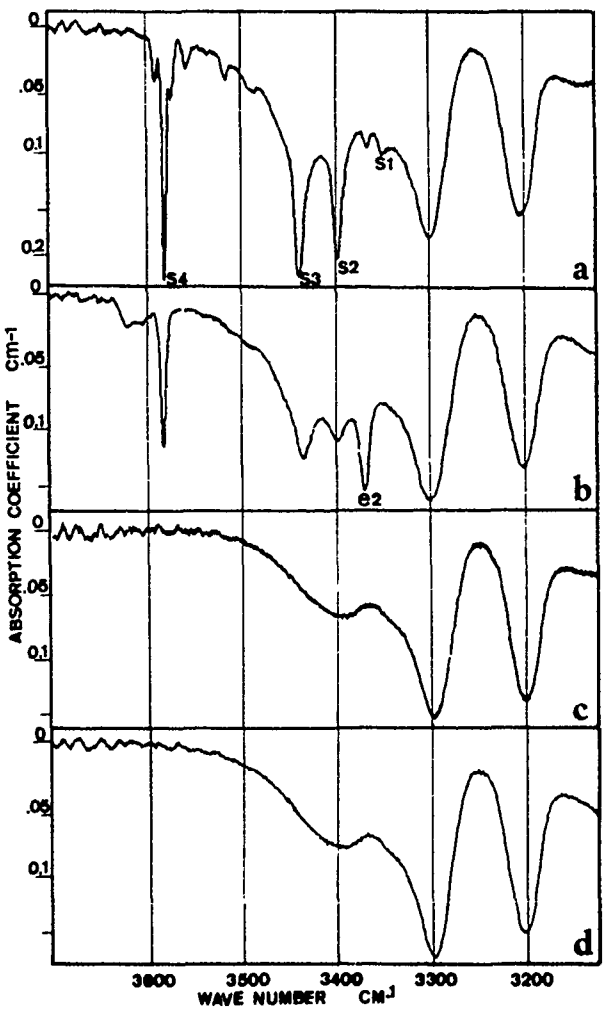


Fig. 2 : Vacuum sweeping : IR spectra (beam along Z - 77 K)
a) as grown
b) 500°C -1500 V/cm 18 days and with 550°C - 1500 V/cm
c) 550°C - 3500 V/cm 6 days (in addition to b)
d) 650°C - 1500 V/cm (fig. 1) 18 days.

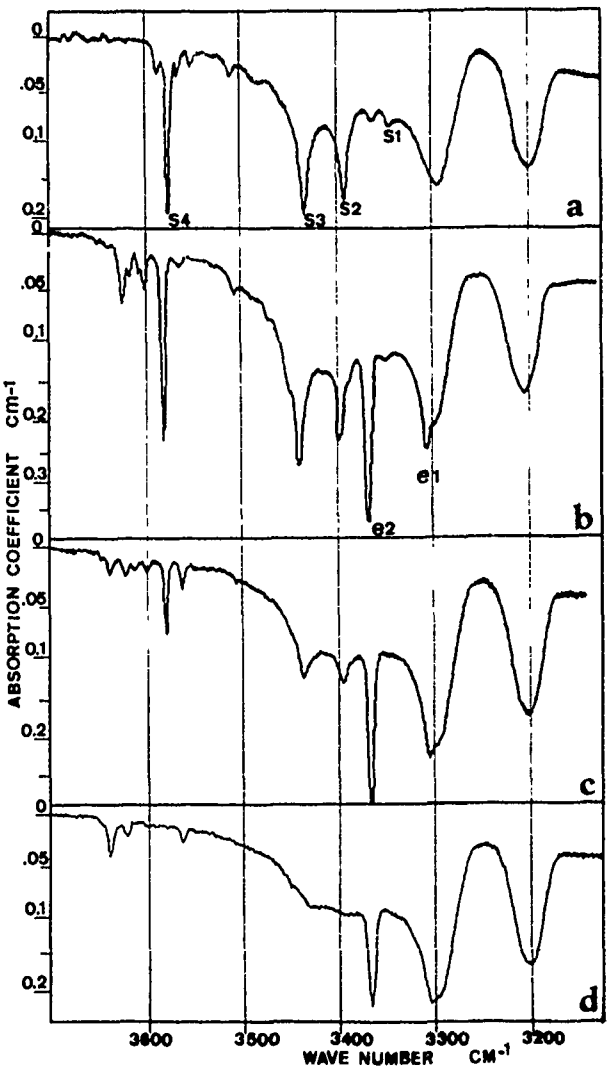


Fig. 4 : air sweeping : IR spectra (beam along Z - 77 K)
a) as grown
b) 500°C - 2000 V/cm 30 days.
c) idem b + 550°C - 3500 V/cm - 7 days
d) 650°C - 1270 V/cm - 8 days.

After 8 days the spectrum was identical to the former one (fig. 2b) but the activation energy rose (2.0 eV instead of 1.46 eV). The electrolysis was then extended for 6 days at 550°C with stronger electric field of 3500 V/cm. The result is the spectrum of fig. 2c which is practically identical to the spectrum obtained in β phase. Furthermore, no twinning appears as the temperature transition was not reached. Both processes seem to induce microfissure (1.5/cm^3., area 1 mm^2), no visual coloration appears.

- Air sweeping

The two samples were optically polished.

Fig. 3 shows the sweeping currents in α (500°C) and β (650°C) phases. Both reach a nearly steady state value which is about 1/9 of their initial value. But the steady state is obtained in ~ 2 days at 650°C instead of 10 days at 500°C.

Fig. 4 shows the different spectra obtained with similar conditions to those used in vacuum sweeping. The "classical" process at 500°C - 2000 V/cm gives (fig. 4b) the e_1 - e_2 (Al - OH) peaks and a slight increase of the broad OH spectrum. Subsequent sweeping on the same sample at 550°C - 3500 V/cm reduces all the sharp peaks and the broad band which becomes weaker than the as grown spectrum. We found about twice more and larger microfissures than in vacuum. In α phase (650°C - 1270 V/cm) s_2, s_3, nearly vanish, s_1, s_4 are completely eliminated while e_1, e_2, decrease. The sample also becomes twinned with about the same quantity of microfissure as in vacuum. No coloration appears.

- Activation energies

The table lists the activation energies E (according to the law $\sigma T = A \exp(-E/kT)$ measured during the cooling at the end of the different sweepings and the α_{3500} coefficient of absorption. ($\alpha = d^{-1} \log_{10} T\,3800/T\,3500$), measured along the Y direction, thickness d, at room temperature).

Table : Activation energies and absorption coefficients after sweeping

VACUUM SWEEPING			AIR SWEEPING		
Process T°C-V/cm	E (eV)	$\alpha_y(3500)$ (cm^{-1})	Process T°C-V/cm	E (eV)	$\alpha_y(3500)$ (cm^{-1})
as grown		0.06	as grown		0.06
500-1500	1.45	0.060	500-2000	1.87	0.064
+ 550 - 1500	2.00	0.060	+550-3500	1.95	0.026
+ 550-3500	1.70	0.020			
	2.24(β)			1.77(β)	
650-1500	1.90(α)	0.024	650-1270	1.97(α)	0.030

- Effect on the incommensurate phase

We show on the insert of fig. 5 the abrupt, 1st order α-β transition as it appears from birefringence or dilatation (11) measurements with the INC phase region indicated by a dashed rectangle. By cooling from the β phase we obtain in the INC phase the birefringence staircase variation (12) showed in fig. 5a for the as grown sample. The cooling is stopped before the transition to the α phase. The sample is then reheated and a thermal hysteresis of about 0.15 °C is observed. Fig. 5b shows the same cycle of temperature variation for the sample after vacuum sweeping in β phase (fig. 2d) The air swept sample of fig. 4d exhibits roughly the same behavior : a spectacular effect of the sweeping is observed with an hysteresis which becomes 0.5°C and the steps width which increases from 0.02 to 0.1°C.

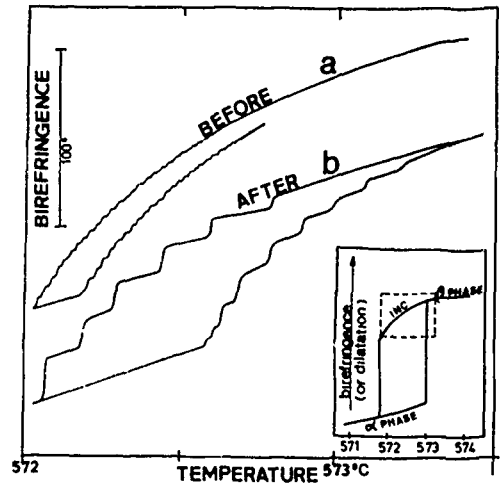


Fig. 5

Birefringence variations during a cooling-heating cycle in the INC phase before a) and after b) vacuum sweeping in β phase. Insert : the whole α-INC-β transition of quartz as seen from birefringence or dilatation measurements. The dashed rectangle shows the region observed here.

DISCUSSION

There is a noticeable difference between the curves (fig. 1 and 3) of swept current as a function of time : the final currents are about ten times higher in air than in vacuum. (Fig. 1) This fact supports recent evidence (13, 14) on the role of hydrogen (profusely supplied by atmospheric water) in the electrical conduction of quartz. At 500°C the initial currents are practically identical in vacuum and in air sweeping as expected for an alkali bulk conduction . At 650°C we do not get in vacuum the initial current in this experiment but in a second experiment (with electrode contact condition closer than that used in air vacuum we also found an initial current close to the air conditions.. In air the two curves at 500 and 650°C are roughly identical after adequate scaling : the 650°C curve is obtained from the 500° curve with a reduction of ~ 5 on the time scale and a multiplication of ~ 15 of the currents. Taking into account the different electric fields used the electrolysis process is accelerated as expected. The ratio of the initial current gives an activation energy of 1.1 eV typical of alkali conduction (15). However the final ratio 1.4 eV should be greater to be consistent with an activation energy of ~ 1.8 eV (typical of conduction by proton H^+ (16) which have been measured directly on the sample after sweeping. In vacuum the curves are much more difficult to interpret. No steady state current was reached . Maybe the lack of hydrogen slows down the process which becomes more dependent on the nature of electrode contact and on the existence of space charge. This space charge can vary considerably with temperature and electric field (1). However at the end of these different sweepings, the activation energy is also typical of hydrogen conduction except in the 500°C - 1500 V/cm process where the smaller value is perhaps due to a part of conduction by remaining alkali ions.

The evolution of infrared spectra may also qualitatively be explained (5) by the role of hydrogen.

i) under vacuum there is mainly the as grown H and "classical electrolysis" is not sufficient (fig. 2b) to dissociate and displace all the M^+ and optically active OH. This can be obtained either with the help of high electric field near the transition temperature in α phase (fig. 2c) or in the β phase at middle field by means of thermal activation (fig. 2d). The complete disparition of the s and e peaks is thus observed. This fact has not been already mentioned as far as we know for samples with a strong initial OH content. We find on the whole sample a result found on the anode side by Krefft (8) in β phase..

ii) In air with the conditions (500°C - low field) of fig. 4b there is enough H available to compensate the loss of alkali ion giving strong e_1-e_2 lines due to Al-OH. When the temperature and the field are increased, the rate of supply of hydrogen from the anode is not sufficient and there is a depletion of H in the sample (fig. 4c).

This phenomenon is more marked in β phase (fig. 4d) with low field. Maybe a still more intense field in air might allow us to reach the results obtained in a phase in vacuum. Thus with respect to the infrared spectra we found sweeping conditions under vacuum which allow similar results in α and β phases to be obtained.

As β swept samples are twinned we have looked for detwinning processes. The study of twinning and detwinning was the object of much work some decades ago (6, 18). Frondel (18) found that thin sample can pass through the transition without twinning. Up to now we were not successful in obtaining a β phase swept thin sample (AT cut 1 mm thick) free of twins. Perhaps because sweeping can introduce strains as recently evidenced (19) or because of initial stresses at the periphery of the sample as evidenced by X-ray topography (20)

Wooster et al (6) discovered that the application of adequate stresses can eliminate electrical twins (based on the fact that twins have different elastic modulus due to the s14 elastic coefficient which changes of sign). Among the various possibilities they investigated, a torque applied at high temperature was the most powerful. Fig. 6 shows the effect of such a treatment on an AT cut quartz which was twinned after heating in the β phase : In second harmonic generation of laser light (SHG) twins give strong variations of the recorded light intensity (fig. 6b). After a torque applied at 550°C (~ 5N.cm) around the X axis we observe (fig. 6c) that the twins are eliminated in the major part of the crystal except at the borders probably due to an edge effect.

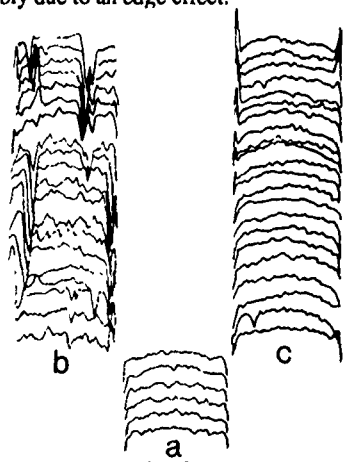


Fig. 6

Detwinning effect of a torque applied around the X-axis of an AT cut twinned sample (15 x 15 x 1 mm³)
The scans give variations of the intensity of S.H.G. from a 1.06m laser.
a) initial b) after heating in β phase c) after torque application.

The incommensurate phase of quartz appears on a microscopic scale as a slight modulation of the periodicity of the atomic positions, with a wavelength which varies (150 to 250 Å) with the temperature. Impurities act as a friction force (12) which pins the displacement of the modulation. It is thus possible to explain the stair case behavior of the birefringence and the increase of the hysteresis. Both vary with impurities concentrations displaying the sweeping effect in a manner which has still to be elucidated.

CONCLUSION

We have shown that it is possible to obtain samples totally free of as grown OH. This can be obtained either in β phase (due to high temperature) or in α phase (due to the high electric field) with vacuum-sweeping while in air-sweeping such an effect will be more difficult. In order to compare the efficiency of α and β phase electrolysis, other measurements (dielectric absorption, E.P.R., radiation hardness) are necessary.. Sweeping appears to have considerable effects on the properties of the incommensurate phase of quartz. These effects can be useful for appreciating the sweeping effectiveness.

ACKNOWLEDGEMENTS

The authors wish to thank P. PALLEAU for the building of the vacuum furnace apparatus and for his assistance particularly to obtain I.R. spectra. B. VIARD (Besançon) for impurities determination, X. BUISSON (SICN) and J.P. AUBRY (CEPE) for providing samples and R. BOURQUIN (Besançon) for communication of his unpublished results.

This work is supported by a DRET contract. 87-126

REFERENCES

[1] J. C. BRICE, "Crystals for quartz resonators" Rev. of Modern Physics, 57, 105-146 (1985)

[2] M.L. KEITH and O.F. TUTTLE, "Significance of variations in the high-low inversion of quartz" Am. J. of Science, 203-253, (1952)

[3] G. DOLINO, "The incommensurate phase of quartz" in Incommensurate phases in dielectrics R. BLINC and A.P. LEVANYUK, eds., Elsevier (1986)

[4] F. MOGEON, G. DOLINO, "Birefringence study of the irreversible behavior of the inc. phase of quartz : effects of chemical impurities" To be published

[5] H. G. LIPSON and A. KAHAN, "Effect of vacuum sweeping and radiation on defect distribution in quartz", IEEE Trans. Nucl. Sci. NS 31, 1223-1229 (1984)

[6] W.A. WOOSTER, N. WOOSTER, J. L. RYCROFT and L.A. THOMAS "The control and elimination of Electric (Dauphiné) twinning in quartz", Proc. Inst. Elect. Eng. 94, 927-937 (1945)

[7] D.M. DODD, D.B. FRASER "The 3000-3900 cm^{-1} absorption bands and anelasticity in crystalline α-quartz" J. Phys. Chem. Solids, 26, 673-686, (1965)

[8] G.B. KREFFT "Effects of high temperature electrolysis on the coloration characteristics and OH-absorption bands in alpha-quartz". Radiation Effects, 26, 249-259 (1975).

[9] G. DOLINO, J. P. BACHHEIMER, M. VALLADE "Direct observation of Dauphiné twins in quartz with Second harmonic light", Appl. Phys. Lett., 22, 623-625 (1973)

[10] H. G. LIPSON, F. EULER and A. A. ARMINGTON "Low temperature infrared absorption of impurities in high grade quartz", Proc. 32nd Ann. Freq. Control Symp. EIA, 11-23, (1978).

[11] J. P. BACHHEIMER "An anomaly in the β phase near the α-β transition of quartz" J. Phys. Lett. 41, L 555-561, (1980)

[12] F. MOGEON, G. DOLINO and M. VALLADE "Kinetic crossover from continuous to discontinuous behavior in the inc. phase of quartz", Phys. Rev. Lett, 62, 179-182 (1989)

[13] J. G. GUALTIERI "Possible mechanism for the introduction of hydrogen into alpha-quartz during sweeping" Proc. 42nd. Ann. Freq. Control Symposium IEEE, 155-161, (1981)

[14] A. K. KRONENBERG, S. H. KIRBY "Ionic conductivity of quartz . DC time dependence and transition in charge carriers" Am. Mineralogist, 72, 739-747, (1987)

[15] J. VERHOOGEN "Ionic diffusion and electrical conductivity in quartz", Am. Mineralogist, 37 637-655, (1952).

[16] J.J. MARTIN "Electrodiffusion (sweeping) of ions in quartz - a review" IEEE Trans on ultrasonics Ferroelectrics and Frequency Control, 35, 288-296, (1988)

[17] J. FELSCHE unt J. LIETZ "Störstellenunter Suchengen am quartz II Feldverteilung beider elektrolysuschen Rauch quartz verfärbung", N. Jb. Miner. Abb. 109, 250-258 (1968)

[18] C. FRONDEL "Secondary Dauphiné twinning in quartz", Am. Mineralogist, 30, 447-460, (1945)

[19] M.T. SEBASTIAN, A. ZARKA, B. CAPELLE "A new X-ray topographic defect contrast on swept quartz crystals" J. Appl. Cryst., 21, 326-329, (1988)

[20] B. CAPELLE Private communication

43rd Annual Symposium on Frequency Control - 1989

Frequency Drift For Quartz Resonator Irradiated By CO Gamma Ray

Shigeru Obara, Mitsuaki Koyama,Akio Chiba, Hideaki Fukuda

Nihon Dempa Kogyo CO., Ltd.

Hideo Ohba

National Space Development Agency Of Japan

Summary

In recent years the effect of radiation on components carried by artificial satellites has come into question. This is why we heve conducted an investigation and research into the radiation environment of a geostationary satellite the radiation resistance of quartz resonatore to be installed in artificial satellites.

The total absorbed dose of a synchronous satellite in a geostationery orbit reaches about $1x10^4$ Gy[1] in ten years. For example, the rete of change of the frequencyof a quartz resonator manufactured from a unswept cryatal with an Q-value of about 2.8 million was about $0.2x\ 10^{-6}$ at a total absorbed dose of $1x10^5$.

Introduction

The space-use quartz resonator is used as the heart of artificial satellites, rocket inertial guidance systems,etc. When used in an artificial satellite, it is always kept in outer space and therefore operate under conditions of exposure to radiation (cosmic rays). For instance, a communication satellite, which is launched into a geostetionary orbit at an altitude of 36,000 km, is directly subjected to the effects of the solar flare proton and the high-energy electrons captured by the geomagnetic field.

Therefore,it is important to determime the magnitude of the quartz resonator frequency change caused by radiation and clarify the mechanism of that change.[1][2]

Cosmic rays which can be mentioned are captured cosmic rays, solar flare protons and galactic coamic rays. Of these the first and second are the cosmic rays that effect artificial satellite equipment. Captured cosmic rays consist of highenergy electrons and protons captured by the geomagnetic fild. Solar flare protons are high-energy protons emitted by solar flares,i.e., local eruptions of the sun.

The total absorbed dose of satellite-borne equipment is estimated to reach about $1x10^4$ GY in ten years, according to the NASA's radiation environment model.

1. Radiation Source and Measurement Samples

In radiation tests of quartz resonators, gamma-rays (^{60}co:about 1.25 MeV), electron beams (10MeV pulse) or x-rays(10 MeV) are mainly used as the radiation source. An electron beam is a mass of charged particles and attenuates so rapidly through interaction with matter that its premeability is limited.

In contrast, gamma-rays and X-rays are electromagnetic radiation and have a great permeability. Therefore, hot an electron beam with its limited permeability but gamma-rays were used as the radiation source.

The sample used in the test was not a swept quartz crystal (quartz rid of alkali ion in the crystal in a high-temperature electric field) but an unswept crystal made into an HC-36/U type quartz resonator with a fundamental wave of two megahertz.[3]

CH2690-6/89/0000-521 $1.00 C 1989 IEEE

2. Measuring Conditions

As shown in Table 1, we used five quartz crystal different in Q value to producs as many quartz resonators, which were placed in a constant-temperature oven of 80 ℃±0.01℃(Fig. 1a) while oscillation circuits were separately kept in a constant-temperature oven of 60℃±0.5 ℃ (Fig. 1b)

The oscillation circuits were prevented from direct exposure to gamma-rays by shielding them with a 200x100x50mm lead block

Furthermore, as heat generation at the time of exposure to radiation had been expected, we produced two Y-cut 10-megahertz quartz resonators as temperature change monitors and put each of them in the 80℃ constant-temperature oven shown in Fig 1 However, one of the baths was left without heater power supply so as to see how much temperature rise results directly from exposure to radiation The other constant-temperature oven supplied with heater power was used to see how much the temperature in the oven changes.

Used as the radiation source was the ^{60}Co gamma-rays of Irradiation Room No.6 of ^{60}Co Gamma-ray Irradiation Bldg. No.2 in the Takasaki Laboratory of the Japan Atomic Energy Research Institute. Gamma-ray irradiation was adjusted to give a total absorbed dose of $1x10^4$ Gy/H and continued for ten hours so as to obtain an integrated total absorbed dose of $1x10^5$ Gy.

In the measuring system used a 10-megahertz highstability crystal oscillator is connected with a frequency counter as shown in Fig 2 so that the frequency could be measured with an accuracy of about $1x10^{-6}$ This is because the rate of frequency change caused by exposure to gamma-rays to the quartz resonators manufactured from quartz crystal different in Q value had been expected to be $1x\ 10^{-6}$ to $1x10^{-7}$

Table 1 Q Values of Classification in Terms (Normal-temperature infrared absorption method)

Sample group	Q value (million)
Group A	2.8 to 3.0
Group B	2.0 to 2.2
Group C	1.3 to 1.8
Group D	0.6 to 1.0

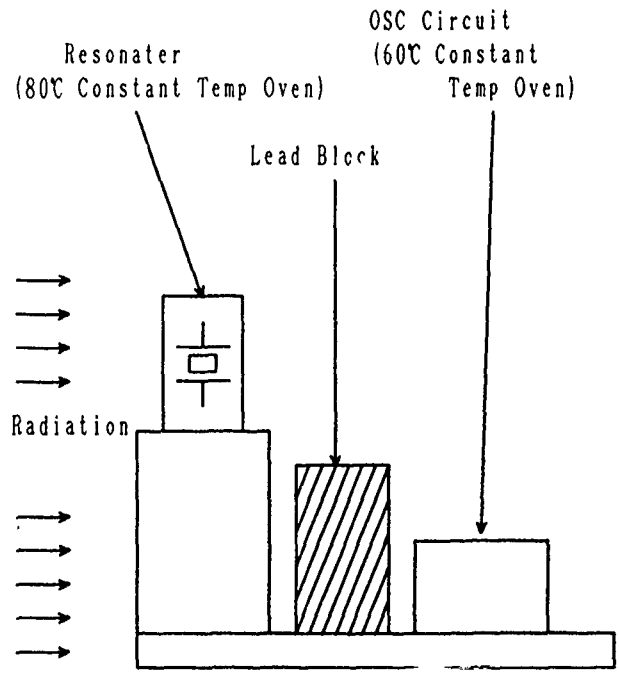


Fig 1 Measuring System In the Irradiation Room

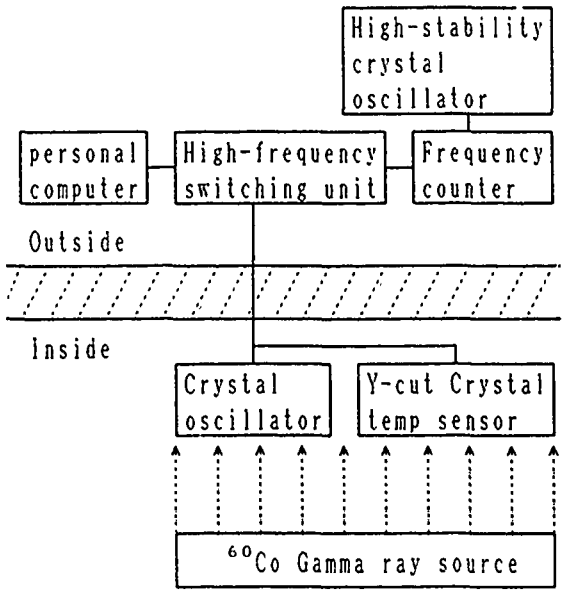


Fig 2 Block Diagram of Gamma-ray Irradiation

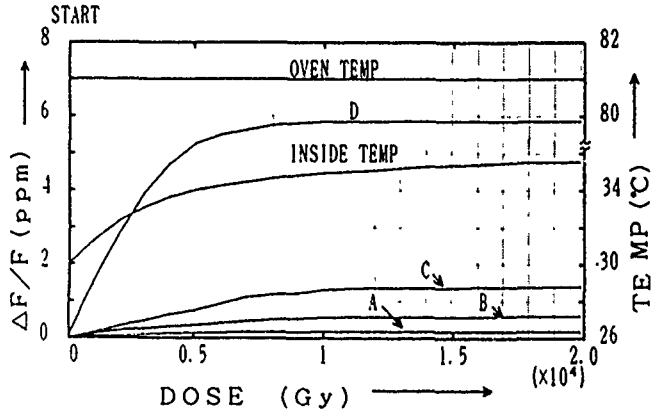


Fig.3 Frequency Change Two Hours After the Start of Irradiation

3. Results and a Review of the Results

The rate of crystal oscillator frequency change due to exposure to gamma-rays is shown in Figs. 3. 4. 5. 6 and 7.

Fig.3 shows an oscillator frequency change of $2x10^4$ Gy. i.e., the frequency change during the first two hours in ten-hour irradiation. At the absorbed dose of $1x10^4$ Gy, the rate of frequency change is beginning to be saturated

Room temperoture begins to rise as irradiation starts, rising by about 6 ℃ on the positive side

In spite of this, the temperature in the 80 ℃ constant-temperature oven doas not change, but remains constant. It can be said from this that the room temperature rise of about 6℃ results from genuine heat generation due to exposure to gamma-rays.

Fig.4 shows $1x10^4$, i,e., the crystal oscillator frequency chage due to exposure to gamma-rays near the end of continuous 10 -hour irradiation

Test samples A to D show no frequency change. remaining constant in a saturated, state at $1x10^4$ Gy.

Both the temperature within the constant-temperature oven and room temperature are constant.

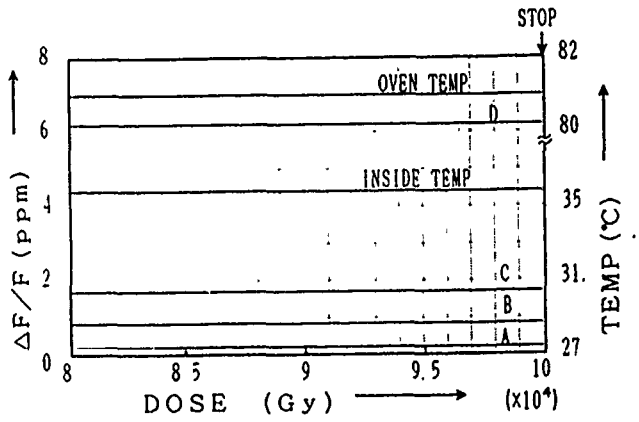


Fig.4 Frequency Change Two Hours Before the End of Irradiation

Fig.5 showe the crystal oscillator frequency change occurring after the oscillator is left to stand for two hours following the completion of irradiation. While room temperature drops rapidly upon completion of irradiation, the constant-temperature oven temparature remains constant.

The frequency changes of test samples A to D does not change but remains constant after the end of gamma-ray irradiation. No frequency return is observable.

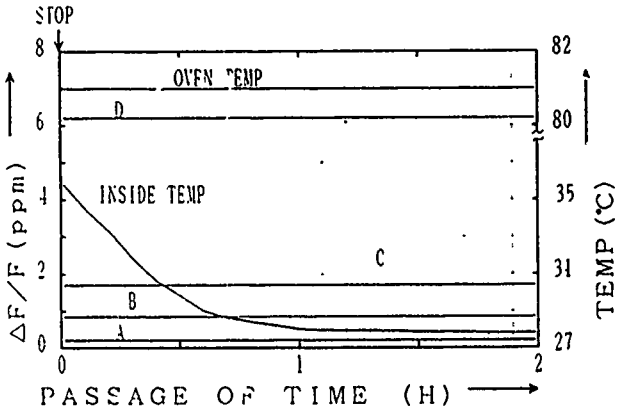


Fig.5 Frequency Change After the End of Irradiation

Fig.6 shows the rate of frequency change for each Q-value in the case of $1x10^4$ Gy irradiation.

The lower the Q value, the greater the rate of frequency change. The frequecy change rate of a Q value of 600,000 can be said to be extremely large as compared with that of a Q value of 1,300,000 or over.

Fig 7 shows the rate of frequency change for etch-channel number at the time of $1x10^4$ Gy irradiation. The larger the number of etch-channeles, the greater the rate of frequency change. The D group shows an extremely great rate of frequency change because the number of etch-channels of the group is far larger than those of groups A to C.

The number of etch-channels shows a crystal defect.

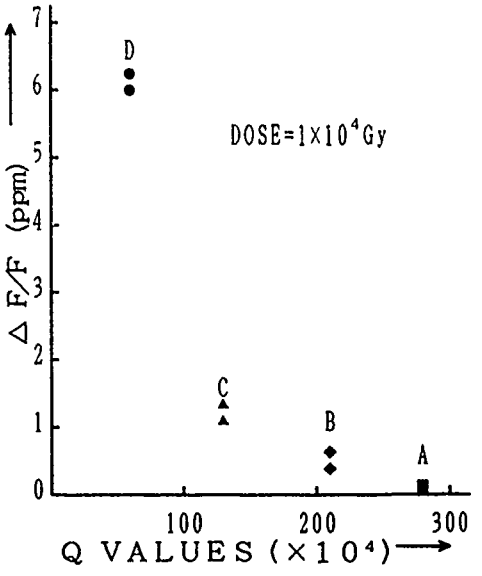


Fig 6 Q Values and the Rate of Frequency Change Due to Radiation

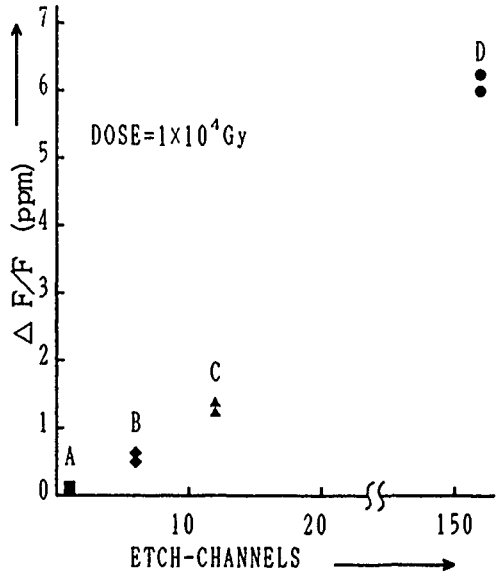


Fig.7 Etch-channels and the Rate of Frequency Change Due to Radiation

Shown in Table 2 are the results of an impurity analysis of these three groups: group D with a particularly great frequency change rate, group A with the lowest such rate and special group S with a Q value of 3 million or over. Group S was not subjected to the test.

The results revealed that group D, which had a particularly great frequency change, contained the least quantities of impurities such as Li and Na centered around Al. For this reason, it is inconceivable that the frequency should change with impurities.

Table 2 Concentration of Impurities in Man-made Crystals

GROUP	Q VALUES ($\times 10^4$)	IMPURITIES ($\times 10^{-6}$)			
		Li	Al	Na	K
A	280~300	0.04	0.22	0.03	<0.03
D	60~100	<0.01	0.09	0.04	<0.03
S	300~380	0.04	0.22	0.07	0.03

4. Conclusion

A roundup of the test results is shown in Table 3, which makes possible the following statement.

1) Frequency change due to radiation is propertional to the number of etch-channels i.e., the number of cryatal defects.
2) Frequency change due to radiation is propertional to the Q value.
3) Frequency change due to radiation is not propertional to the kind of quantity of impurities.
4) The Frequency change of a quartz resonator using an unswept quartz crystal with a Q value of 2.8 million is 0.2×10^{-6} at a maximum at 1×10^5 Gy irradiation.
5) The space-use quartz resonator using an unswept quartz crystal produced by us is considered to stand use for at least ten years.

Table 3 Comparison of Test Results

GROUP	Q VALUES ($\times 10^4$)	IMPURITIES ($\times 10^{-6}$)				ETCH-CHANNELS	$\Delta F/F$ ($\times 10^{-6}$)
		Li	Al	Na	K		
A	280~300	0.04	0.22	0.03	<0.03	2	0.2MAX
B	200~220	—	—	—	—	8	1.0MAX
C	130~180	—	—	—	—	12	2.0MAX
D	60~100	<0.01	0.09	0.04	<0.03	150	7.0MAX
S	300~380	0.04	0.22	0.07	0.03	—	—

References

1. H.L.Olesen; "Radiation Effects on Electronic Systems" PLENUM PRESS, 1966
2. E.G Stassinopoulos; "The Geostationary Radiation Environment" J.SPACECRAFT, Vol 17, No.2 pp145-152, 1980
3. J.R.VIG; "Radiation Effects in Quartz Osillators" 6th Quartz Devices Conference and Exhibition, pp49-83, 1984
4. T.J.Young, D.R.Koehler, R.A.Adams; "Radiation Induced Frequency and Resistance Changes in Electrolyzed High Purity Quartz Resonators" Proc.32th Annual Frequency Control Symposium, pp34-42.1978
5. P.Pellegrini, F.Euler, A.Kahan "Steady State and Radiation Effects in Precision Quartz Ocillator" IEEE Trans.Nuclear Science, Vol NS-25, NO.6, pp1267-1273, 1978
6. J.J.Martin, H.B.Hwang, H.Bahadur "Radiation Effects in The Acoustic loss Spectra of AT-Cut Quartz Crystals" Proc 39th Annual Frequency Control Symposium, 1985

43rd Annual Symposium on Frequency Control - 1989

RADIATION INDUCED FREQUENCY OFFSETS AND ACOUSTIC LOSS IN AT-CUT QUARTZ CRYSTALS

J. J. Martin
Department of Physics
Oklahoma State University,
Stillwater, OK 74078-0444

ABSTRACT

Ionizing radiation modifies point defects already present in quartz. For example, the Al-Li center present in as-grown high quality cultured quartz is converted into a mixture of Al-OH and Al-hole centers. While neither the Al-Li nor the Al-OH center have acoustic loss peaks, peaks at 23K, 100K and 135K are associated with the presence of the Al-center. In unswept quartz, the ratio of stable Al-hole centers to Al-OH centers is determined by the availability of hydrogen trapped in the growth defect sites. Since these peaks occur at temperatures below the temperature of a crystal oscillator their production causes a negative frequency offset. The magnitude of the radiation-induced offset due to the production of the Al-hole center should match the sum of the heights of these three loss peaks. We have measured the acoustic loss and frequency versus temperature spectra as a function of radiation dose of a number of matching Na-swept, unswept, and H-swept AT-cut quartz crystals. As expected, large positive frequency offsets at both the upper and lower turnovers were observed for the Na-swept crystals. However, the magnitude of the offset was consistently much less than the change in height of the strong 53K (5MHz) Al-Na loss peak. This smaller offset was caused by a change in "apparent" angle of the crystal that comes about when the Al-Na center is replaced the Al-OH or Al-hole center. King first noted that AT-cut crystals with large Al-Na concentrations behave as if they were cut at a smaller angle than natural or swept crystals. Both the unswept crystals and the H-swept crystals usually showed negative frequency offsets at the turnover temperatures. The offsets of the swept crystals were usually smaller than those of the unswept samples. Only part of the observed fractional frequency offset at the turnovers could be explained by summing the heights of the loss peak changes An additional radiation induced offset was observed at temperatures below 15 K where the the frequency versus temperature curves flatten out. This offset does not appear to be related to any acoustic loss peaks; instead, it is caused by shifts in the elastic moduli that take place when a defect is modified. Once this low temperature offset is subtracted from the offset measured at the turnover the remaining frequency shift matches the value predicted by the changes in the acoustic loss peaks.

INTRODUCTION

The effects of ionizing radiation on precision quartz oscillators must be considered in a number of aerospace applications. Both transient and steady-state frequency offsets and series resistance changes of the crystal have been observed following the exposure to radiation [1-10]. King and Koehler [1] have reviewed the radiation response of quartz with respect to frequency control applications. Norton, Cloeren, and Suter, Suter and Maurer; and Suter have recently reported low-dose frequency offset studies [11,12,13]. Several mechanisms for low-dose effects have been discussed by Flanagan, Leadon, and Shannon [14]. Palkuti and Truong [15] have described an automated radiation test facility. The frequency of the thickness-shear mode used in rotated Y-plates and similar crystals is given by

$$f = (n/2t)(c'/\rho)^{1/2} \qquad (1)$$

where n is the overtone, t is the thickness of the crystal, ρ is the density, and c' is the rotated elastic modulus. The elastic modulus and the density are the primary "materials" factors that determine the crystal frequency. Other factors such as the piezoelectric coupling, resonator contour, and mounting structure also play a role in the actual frequency of the quartz device. The elastic modulus represents a type of average of the interatomic force constants; thus, it should be sensitive to the substitution and/or modification of impurities and defects in the crystal lattice. If a defect has two or more equivalent orientations it has the possibility of thermally hopping between the two orientations [16]. When this hopping comes into resonance with the operating mode of crystal it gives ise to a shift in the elastic modulus, Δc, given by

$$\Delta c' = -Dc_o'/(1 + \omega^2\tau^2) \qquad (2)$$

where D is the strength of the interaction, c_o' is the eleastic modulus at absolute zero, ω is the angular frequency of the mode, and $\tau = \tau_o \exp(E/kT)$. τ_o contains the "jump-time" and an entropy term for the defect, and E is the barrier height between the equivalent orientations. The resonance between the defect and the crystal vibration also removes energy from the vibration; this gives rise to an acoustic loss, $\Delta(Q^{-1})$ given by

$$\Delta(Q^{-1}) = D\omega\tau/(1 + \omega^2\tau^2). \qquad (3)$$

Suppose that the only temperature dependence of the elastic modulus, c', caused by the defect giving rise to equations 2 and 3. Then as the temperature is scanned from T = 0 K to T = ∞ the loss $\Delta(Q^{-1})$ reaches a maximum value of D/2 at T = T_m when $\omega\tau = 1$ and then decreases to zero. At the same time, the elastic modulus decreases by an amount Dc_o'. By using equation 1 we see that there is a corresponding fractional frequency shift $\Delta f/f = 0.5D$. Of course, the elastic modulus and, therefore, the frequency is not independent of temperature and the shifts predicted by equation 4 should be sumperimposed upon upon the actual f vs T curve of the crystal. As discussed below, several defects that have anelastic loss are often observed in quartz and are either produced or destroyed by ionizing radiation. Therefore, the radiation induced frequency shift, $\Delta f/f$, should be given by

$$\Delta f/f = -\sum\Delta(Q^{-1})_{mi} \qquad (4)$$

where $\Delta(Q^{-1})_{mi}$ is the radiation induced change in the height of the ith loss peak. For example, the Al-Na center has a strong loss peak at 53 K in 5 MHz crystals; it is destroyed by irradiation so $\Delta(Q^{-1})_m$ is negative and a positive offset is observed.

Cultured quartz contains a number of interesting point defects. Fraser [16] has discussed the early work on defect-related acoustic loss. Weil [17,18] has reviewed the paramagnetic defects which are observed after irradiation. Halliburton, Martin, and Koehler [19] have also discussed the point defects present in culutured quartz. Quartz has a number of defects which can trap hydrogen. Often the hydrogen forms an OH molecule by bonding to one of the intrinsic oxygen atoms forming an infrared active center. Kats [20] carried out a classic study covering the OH-related

CH2690-6/89/0000-526 $1.00 C 1989 IEEE

defects in both cultured and natural quartz. Cultured quartz has a set of "grwoth-defects" which trap hydrogen and give rise to infrared bands at 3350, 3398,3437, and 3581 cm^{-1}. The defects responsible for these bands have not been identified. Recently, Martin, et al [21] have reported that these defects can also trap alkalis. While the hydrogen trapped at these "growth-defect" sites does not cause any anelastic loss, lithium or sodium trapped at the site produces loss peaks at 305 K or 335 K respectively. These loss peaks grow rapidly with irradiation and then in some crystal go out completely at higher doses. The OH-related growth defects are the source of the hydrogen that is transferred to the aluminum during the radiation process. The "growth-defects" also seem to act as one of the traps for the alkalis that are released from the aluminum. In semiconductor terminology, aluminum substituting for silicon is an acceptor. Quartz is grown in an alkali rich environment and lithium or sodium are trapped intersititially next to the aluminum with the valence electron providing the compensation. The sodium sits "off center" in the Z-axis channel; and the resulting Al-Na center causes a strong acoustic loss peak that is observed at 53 K in 5 MHz crystals. The Al-Na center also has a much weaker loss peak at 135 K. Lithium, however, sits "on center" when trapped intersitially by the aluminum and, consequently, the Al-Li center shows neither acoustic [22] nor dielectric loss peaks [23]. Lopez, Hwang, and Martin [23] discussed the conversion by irradiation of the Al-Li (and Al Na) centers into a mixture of Al-OH and Al-hole centers. The infrared active Al-OH center does not produce an acoustic loss peak. However, acoustic loss peaks at 23 K, 100 K, and 135 K are associated with the presence of the Al-hole center [8,21].

Other point defects are also present in cultured quartz. Germanium can substitute for silicon in the quartz lattice and act as an electron trap. Weil [18] has given a complete discussion of the paramagnetic centers associated with germanium in quartz. Euler and Kahan [24] have reported an anelastic loss peak at 250 K (5 MHz) caused by the Ge-Li center in germanium doped quartz. Martin, et al, [20] observed the same loss peak in natural quartz; however, they were unable to find it in resonators taken from eight different bars of conventionally grown cultured quartz. The Ge-Li loss peak at 250 K grows quicly with irradiation and then decays at higher doses. Both oxygen and silicon vacancies are present in cultured quartz. These centers become paramagentic under irradiation and can be observe by using epr techniques. However, no anelastic loss peaks have been observed that can be attributed to these vacancy centers. Table 1 summarizes the acoustic loss peaks normally observed in conventionally grown cultured quartz.

Table 1. Point defects and their acoustic loss peaks in cultured quartz.

Center	Peak	Comments and peak height per/ppm
Al-Li	NONE	the dominant Al center in as-grown quartz.
Al-Na	53 K	22×10^{-6}/(ppm Al-Na) [25]
	135 K	much weaker than the 53 K peak
Al-OH	NONE	present after sweeping or irradiation.
Al-hole	23 K	may be stronger or weaker than the 100 K peak.
	100 K	1.7×10^{-6}/(ppm Al-hole) [26]
	135 K	weaker than the 100 K.
OH-GD+	NONE	
Li-GD	305 K	weak, present in Li-doped as-grown quartz.
Na-GD	335 K	stronger than the Li version.
Ge-Li	250 K	only in germanium doped.

+GD = growth-defect

For a number of years, the Oklahoma State group has been studying the "materials properties" of quartz using a variety of spectroscopic techniques. Of these techniques, acoustic loss measurements provide the most direct connection to crystal oscillator performance. We report here a comparison between changes produced in the acoustic loss spectra by irradiation or electrodiffusion and the frequency at which the crystal would normally be operated (upper-turnover). We find that while the relationship between the fractional-frequency offset and change in the loss peaks given in Eq. 4 usually holds that both the magnitude and temperature dependence of the elastic modulus may also change with irradiation.

EXPERIMENTAL PROCEDURE

Cultured quartz from a number of different sources was used for these experiments. With one exception, all of the crystals were taken from the Z-growth region of the stone. The exception was a R-growth stone, HA-A, which was the source of the high aluminum content material. The aluminum content of the quartz used in the study ranged from less than 1 ppm up to the 60-70 ppm of HA-A. Samples designated M----, such as MH11-19 were grown at Motorola. Martin, et al, [25] reported a number of evaluation tests on these samples. Samples from our Sawyer Premium Q bar PQ-E and Toyo Supreme Q bar SQ-B were also included in this study. EPR measurements were used to determine the aluminum concentration of all of our samples [26]. Low temperature IR absorption scans were also made on all samples. Samples from bars PQ-E, SQ-E, and MH11-19 all had about 10-12 ppm of aluminum. However, the IR scans showed that the growth-defect bands were much larger in the MH11-19 samples than in either PQ-E or SQ-B. All of the data reported here is for AT-cut crystals. Most of the resonator blanks were 5 MHz 5th overtone devices fabricated by Piezo Crystal. However, 10 MHz 3rd overtone crystals were made from bar MH11-19 by Piezo. Frequency Electronics prepared 5 MHz 3rd overtone blanks from the Toyo SQ-B bar. K-W Mfg. fabricated the blanks from the Sawyer PQ-E bar. After fabrication, the blanks were then sodium, lithium, or hydrogen swept using our standard process [27]. Finished crystals were made "in-house." The blanks were electroded and mounted in HC-40 holders using a silver-filled polyimide adhesive.

Acoustic loss versus temperature measurements were made using an automated transmission method system [28]. Most of the measurements were made on finished crystals; however, a number of blanks mounted in a gap holder were also measured. The crystals or blanks were mounted on the cold-head of a closed-cycle helium refrigerator. After the samples were cooled to 8 K the loss and frequency were recorded as system was warmed to about 400 K. The heating rate was approximately 20 K/hr.

Room temperature irradiations were performed using a 18.8 kRad(Si) per hour ^{60}Co gamma source. Successive room temperature irradiations were usually carried out until the samples reached saturation. Low temperature IR absorption measurements of the Al-OH center production were made on matching samples that were irradiated at the same time as the resonator crystals. Similar visible-uv absorption measurements of the Al-hole center production were also made on the high-aluminum content sample. Low temperature irradiations were made "in-situ" with the refrigerator cold-head placed in-line with the 1.75 MeV electron beam from our Van de Graaff accelerator. Sample heating was overcome by chopping the beam so as to hold the sample temperature between 50 K and 90 K as indicated by the resonant frequency of the crystal.

RESULTS AND DISCUSSION

Sodium Compensated Aluminum

Since the effects of the Al-Na center are larger than any other point defect we will first consider Na-swept crystals. Figure 1 shows the acoustic loss versus temperature spectrum over the 9 to 400 K temperature range for a Na-swept 10 MHz 3rd overtone AT-cut crystal taken from bar MH11-19. The height of the large low-temperature peak in the as-Na-swept curve is consistent with the 12 ppm aluminum concentration of this sample. Since this is a 10 MHz crystal the sodium peak occurs at a temperature slightly above the 53 K observed in 5 MHz samples. A Na-related growth-defect loss peak at 350 K is present in the as-Na-swept curve.. Again, the higher frequency causes the peak to appear at a higher temperature than the 335 K observed for 5 MHz samples. Note also that the low temperature peak is several hundred times as strong as the 350 K peak. Successive room temperature gamma irradiations destroyed the low temperature peak and caused the 350 K peak to grow. Figure 2 shows the change in the height of the low temperature Al-Na loss peak, $\Delta(Q^{-1})_{Na}$, and the fractional-frequency offset, $\Delta f/f$, measured at the upper turnover temperature of 326 K. Because the upper turnover is below the peak temperature (350K) the effect of the Na-related-related growth-defect peak should be small. Equation 4 predicts that we should see a $\Delta f/f$ of 270 ppm at saturation. Instead, an offset of only 35 4 ppm is found at saturation. We consistently find that the radiation-induced $\Delta f/f$ is much less than the reduction in the height of the Al-Na loss peak. The results for four crystals with Al-Na contents ranging from 0.2 ppm to 70 ppm are given in Table 2. A positive frequency offset at the upper turnover was observed for all Na-swept AT-cut crystals. A smaller but still positive offset was also observed at the lowest temperature reached (about 9 K) in the data runs, this low temperature offset is listed in Table 2.

Table 2. Change in Al-Na peak and $\Delta f/f$ for Na-swept crystals.

Sample	Al-Na (ppm)	$\Delta(Q^{-1})_{Na}$	$\Delta f/f$(UF)* (ppm)	$\Delta f/f$(LTF)+ (ppm)
ME24-48	0.2	4.4×10^{-6}	0.57	0.23
MH26-39	2.3	58×10^{-6}	7.0	0.86
MH11-19	12	270×10^{-6}	35.4	2.8
HA-A	70	2400×10^{-6}	406	103

*UF = upper turnover frequency.
+LTF = low temperature frequency.

Figure 3 compares the frequency versus temperature curves for an as-grown MH11-19 10 MHz crystal with a H-swept crystal from the same batch and the Na-swept crystal in the as-Na-swept condition and after a 656 kRad irradiation. The blanks for these crystals were fabricated by Piezo Crystal to an angular tolerance of plus or minus 1'. The Na-swept crystal appears to be cut to a smaller angle than either the unswept or the H-swept samples. King [29] first noted that AT-cut crystals with large Al-Na concentrations behave as if they are cut to a smaller angle. After irradiation the frequency versus temperature curve for the Na-swept sample approaches same "angle" as that of the unswept and H-swept samples.

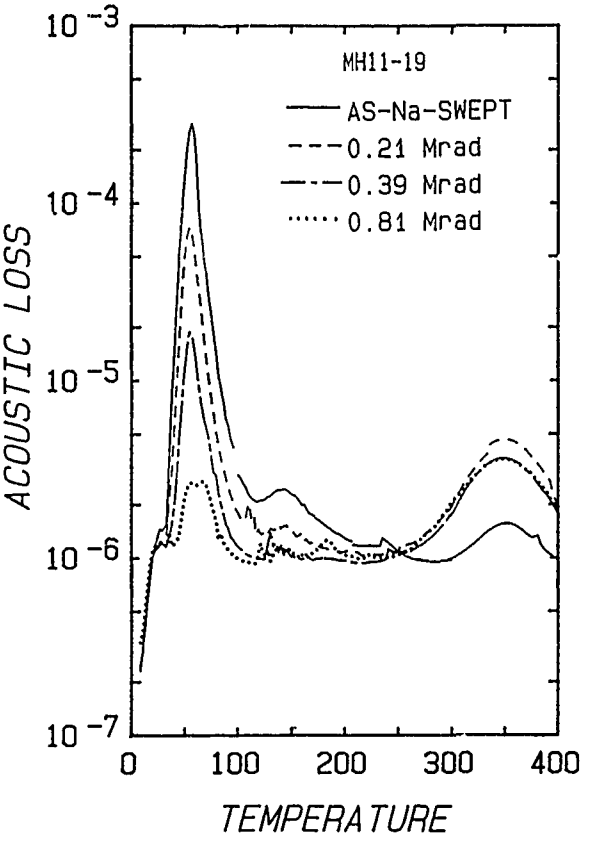


Figure 1. The reduction of the large Al-Na loss peak in a Na-swept 10 MHz 3rd overtone MH11-19 crystal is shown. The Na-related peak at 350 K grows with irradiation out to 0.21 Mrad then decreases slighlty.

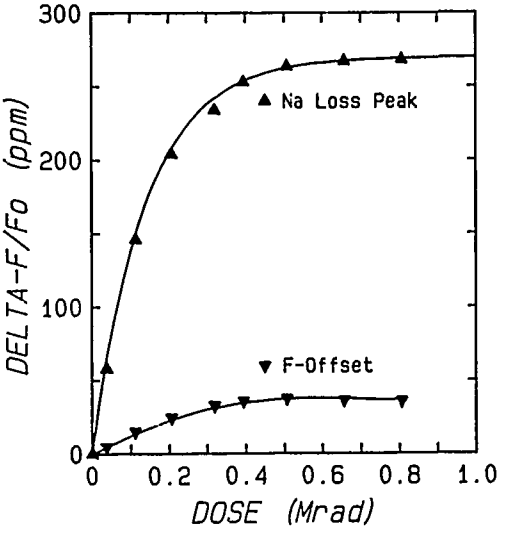


Figure 2. The fractional-frequency offset at the upper turnover of the Na-swept MH11-19 crystal versus radiation dose is shown by the deltas. The offset is much less than the amount predicted by the change in the Al-Na loss peak which is shown by the triangles.

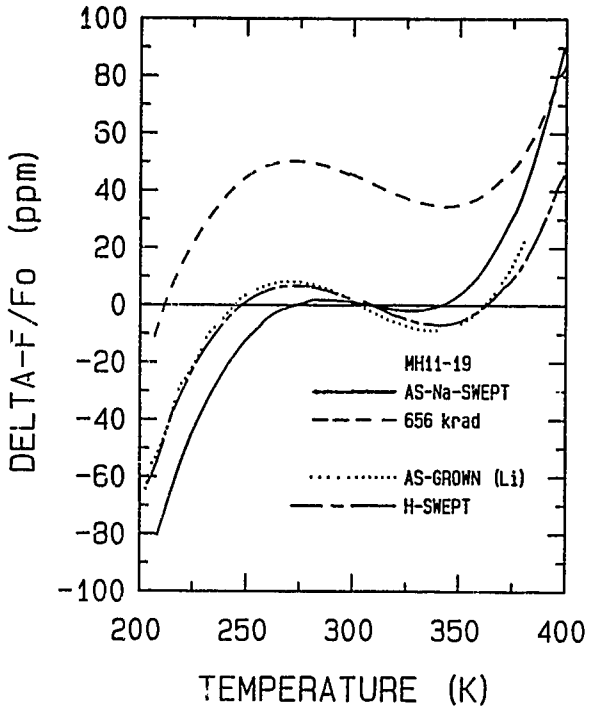


Figure 3. The frequency versus temperature curve for the Na-swept MH11-19 crystal appears to be for a smaller angle than the unswept and H-swept crystals. All three were cut to the same angle. Irradiation brings the Na-swept sample back towards the angle of the unswept and H-swept units.

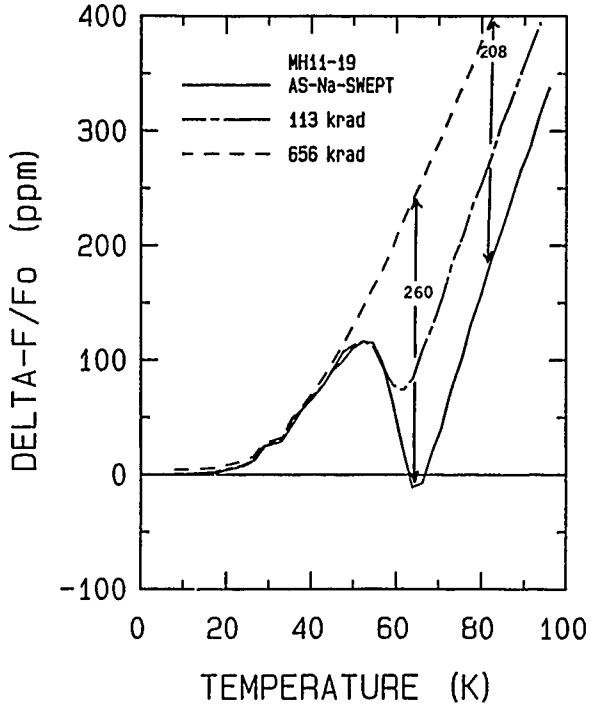


Figure 4. The f vs T curves for the MH11-19 crystal are shown for the as-Na-swept condition and for several irradiations. Between 60 and 70 K the radiation-induced frequency offset agrees with the change in the Al-Na loss peak. At higher temperatures, the f vs T curves start to come together.

Figure 4 shows the frequency versus temperature curves for the Na-swept MH11-19 crystal in the as-Na-swept condition and after two different irradiations. The maximum radiation-induced frequency offset of 260 ppm nearly matches that predict using Eq. 4 and the change in the Al-Na loss peak. However, at higher temperatures the offset decreases as the "apparent angle" shifts. The cause of this "angle shift" must be an alteration of the elastic moduli by the Al-Na center that is separate from the anelastic contribution given by Eq. 2.

Lithium Compensated Aluminum

Both positive and negative radiation-induced offsets were observed in the unswept and Li-swept samples that we tested. Our Li-swept HA-A quartz can be considered as a worst case example. The solid curve in Fig. 5 shows the acoustic loss versus temperature for the as-Li-swept condition of this 60-70 ppm aluminum content sample. There is a small AL-Na center loss peak at 53 K; the height of this peak corresponds to about 0.1 ppm of Al-Na centers [25]. The triangles in Fig. 5 shows the 305 K loss peak that was produced by a 38 krad irradiation. This Li-growth-defect loss peak continued to grow with irradiation up to about 0.15 Mrad and then decayed for higher doses. The deltas show the acoustic loss versus temperature curve measured after a total radiation of 4.73 Mrad. At this stage of the irradiation, a matching optical sample showed a distinct smoky coloration. The 23 K, 100 K, and 135 K peaks associated with the Al-hole center are clearly seen in this curve. The 305 K loss peak disappeared at these very high doses.

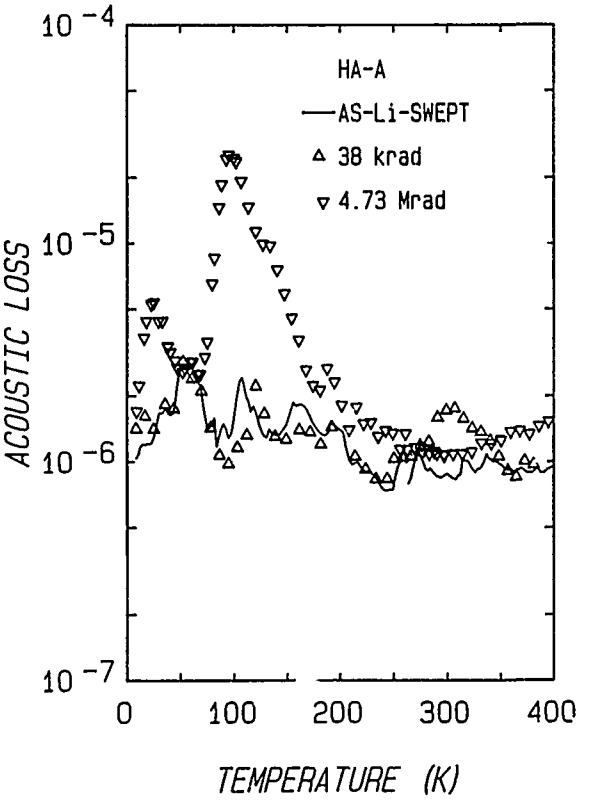


Figure 5. The solid curve shows the as-Li-swept acoustic loss versus temperature spectrum for the HA-A crystal. The 305 K Li-growth-defect loss peak was produced by a 38 krad dose. Higher doses destroyed this peak and produced the Al-hole related peaks at 23 K, 100 K and 135 K shown by the deltas.

The curve marked by the triangles in Fig. 6 shows the growth and decay of Li-growth-defect loss peak (305 K) in the Li-swept HA-A sample. The production (and destruction) of this loss peak caused an associated frequency offset. Since the 305 K peak "straddles" the upper and lower turnovers of the frequency versus temperature curve lower temperature effects can be removed by considering the difference, UF-UL, between the radiation-induced offsets at the upper and lower temperatures. As shown by the deltas in Fig. 6 this difference matches the growth and decay of the loss peak.

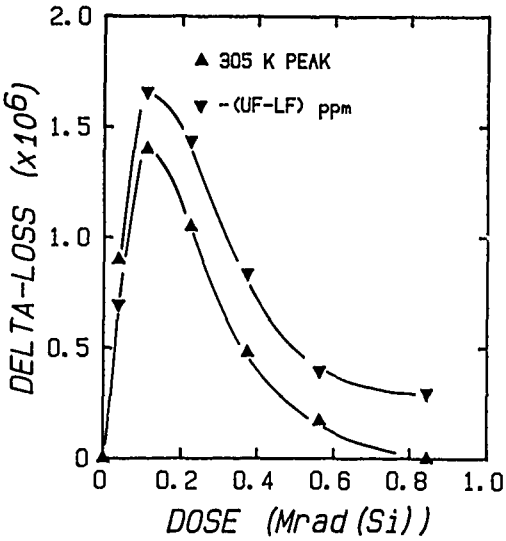


Figure 6. The fractional-frequency offset calculated as the difference (UF-LF) between the upper and lower turnovers for the 5 MHz Li-swept HA-A crystal tracks the growth and destruction of the 305 K loss peak.

We have previously reported that the peaks at 23 K, 100 K and 135 K shown by the deltas in Fig. 5 grew linearly in this sample with the production of the Al-hole center [23]. The sum of the heights of the three loss peaks predicts that a frequency offset of -32 ppm should be observed for the total dose of 4.73 Mrad. The curve marked by the squares in Fig. 7 shows that the fractional-frequency offset at the upper turnover temperature goes slightly positive and finally stabilizes at +3.5 ppm for the 4.73 Mrad. An unexpected large positive offset was observed at the lowest temperatures reached (9 K). This offset grew systematically with radiation dose as shown by the squares in Fig. 7. Apparently, the low temperature radiation-induced offset compensates for the negative shift due to the the production of the Al-hole center loss peaks. If there is not a large radiation-induced shift in the "apparent angle" of this AT-cut crystal the difference between the offset at the upper turnover and that at low temperature should then track with the frequency shift predicted from Eq. 4. The curves shown by the triangles (UF - LTF) and by the deltas (Eq. 4) in Fig. 7 show that this assumption holds reasonably well.

To determine that this unexpected low temperature offset was not an artifact of the mounting structure, the resonator blank was removed from the HC-40 holder and the electrodes stripped off. An isochronal anneal study was carried out as follows. The acoustic loss and frequency of the blank mounted in a gap holder were measured from 9 K to 380 K. The blank was then successively annealed for 5 minutes at temperatures up to 790 K. After each anneal the loss and frequency versus temperature were were recorded. The 100 K center loss peak was used to track the Al-hole center. As shown by the curve marked with the deltas in Fig. 8 the Al-hole center anneals out in the expected 500 k to 600 K temperature region [26]. Frequency offsets at both the upper turnover, UF, and at low temperatures, LTF, relative to the annealed values are plotted as the diamonds and the squares respectively in Fig. 8. The initial low temperature offset of the blank is in reasonable agreement with that of the fully irradiated crystal. The low temperature offset shows a small decrease when the Al-hole center decays; however, its major decay is at higher temperatures where the Al-OH center is being converted back into Al-Li centers and the hydrogen is returning to the growth-defect sites [21]. The offset at the upper turnover increases by about 20 ppm in the 500-600 K range where the Al-hole center is decaying. At the higher temperatures it tracks the low temperature offset.

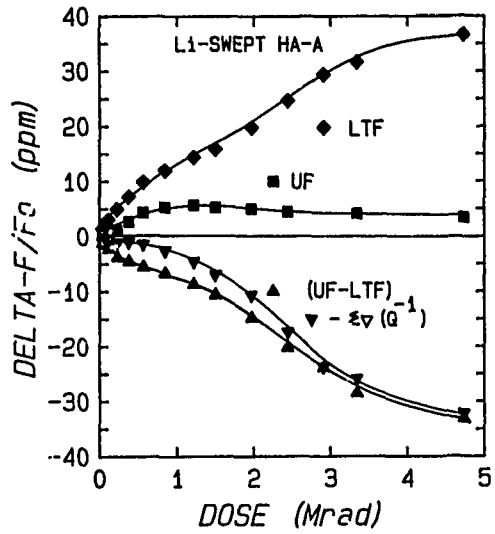


Figure 7. The deltas show the sum of the changes in the loss peaks for the Li-swept HA-A crystal. In steady of following this predicted curve (Eq. 4) the fractional-frequency offset at the upper turnover versus dose is positive as shown by the squares. The frequency at low temperatures (9 K) increased with dose as shown by the diamonds. The difference between the turnover offset and the low temperature offset does follow the predicted curve.

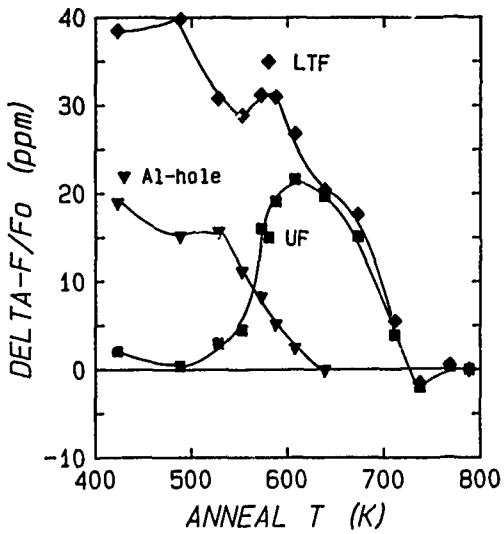


Figure 8. The anneal of the irradiated HA-A blank is shown. The initial offsets match those of the crystal.

The acoustic loss versus temperature curves for an unswept 10 MHz 3rd overtone AT-cut crystal taken from bar MH11-19 is shown in Fig. 9 for the as-grown condition and after two successive irradiations. The complete absence of the low temperature Al-Na loss peak and the absence of the Al-OH infrared band show that all of the aluminum in the as-grown MH11-19 material is lithium compensated. The acoustic loss curve for the sample in the as-grown condition shows a weak loss peak at 130 K of unidentified origin. The 0.11 Mrad irradiation removed the 130 K peak and produced the Li-growth-defect peak near 330 K [20]. This peak is the lithium-analog of the stronger peak seen in Fig. 1 for the Na-swept MH11-19 crystal. This irradiation also produced a weak peak near 180 K. As the sample was irradiated to higher doses the Li-growth-defect peak decreased slightly and 180 K peak disappeared. Since they disappeared when the crystal was operated on its 5th overtone, the small peaks near 240 K and the sharp peak near 320 K in the curve for the as-grown condition are due to mode crossings. In contrast to the results for the HA-A sample, no Al-hole center loss peaks were observed following irradiation at room temperature in the MH11-19 crystals. There is enough hydrogen present at the growth-defect sites to compensate nearly all of the aluminum so that very few room-temperature stable Al-hole centers are produced.

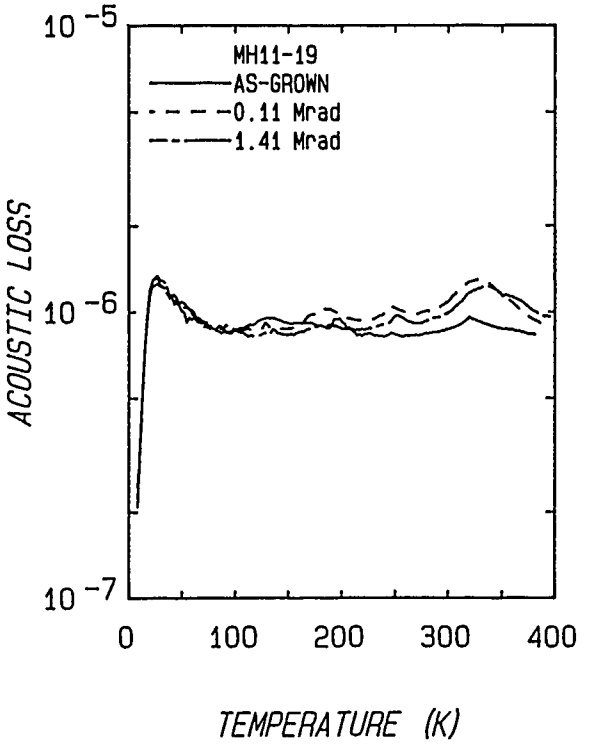


Figure 9. The acoustic loss spectrum of unswept the MH11-19 crystal is shown.

A negative fractional-frequency offset at the upper turnover, UF, versus radiation dose curve as shown by the squares in Fig. 10 was observed for this as-grown MH11-19 crystal. This offset is larger than is predicted by the changes in the observed acoustic loss peaks. As shown by the diamonds in Fig. 10 this as-grown MH11-19 crystal had a negative radiation-induced offset at the lowest temperatures reached. If we again plot the difference between the offsets at the upper turnover and at low temperatures, UF-LTF, we get the curve marked by the triangles in Fig. 10. Its magnitude is still somewhat larger than predicted from the changes in the loss peaks. Table 3 gives the fractional-frequency offsets at the upper turnover and at low temperatures that we have observed in four different samples where the aluminum is compensated by lithium.

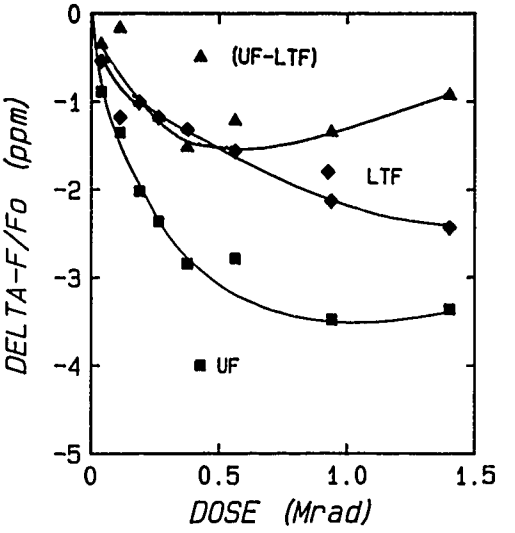


Figure 10. The fractional-frequency offsets at the upper turnover, UF; at low temperatures, LTF; and their difference (UF-LTF) for the unswept MH11-19 crystal are shown.

Table 3. Fractional-frequency offsets at saturation for Li-compensated crystals.

Sample	Al (ppm)	Δf/f(UF) (ppm)	Δf/f(LTF) (ppm)
ME24-48	0.2	+1.55	+1.2
MH11-19	12	-3.36	-2.43
PQ-E	10-12	-.054	+0.4
HA-A	60	+3.18	+32

In all of the samples that we have tested, the offset at low temperatures grew systematically with radiation dose. In these unswept/Li-swept samples the low temperature offset was usually much smaller than in the Na-swept samples.

Hydrogen Compensated Aluminum

Swept quartz is usually preferred for crystals that may be exposed to ionizing radiation. Commercially, the process is done in air; we have found that carrying out the electrodiffusion process in a hydrogen atmosphere yields the same result [30]. In swept quartz, the alkalis that charge compensate the aluminum have been replaced with hydrogen. The room temperature mechanical Q is also improved because the alkalis trapped in the growth-defects have also been replaced with hydrogen. Fig. 11 shows the acoustic loss versus temperature curves for a hydrogen swept MH11-19 10 MHz 3rd overtone AT-cut crystal. This crystal is a mate to the Na-swept and unswept MH11-19 crystals discussed above. The Li-growth defect loss peak which occurs at 328 K for 10 MHz crystals has been removed by the electrodiffusion process; consequently, the mechanical Q at the operating temperatures is improved. The dashed curve in Fig. 11 shows that no loss peaks were produced by a 1 Mrad irradiation. However, we did observe frequency offsets at both the upper turnover, UF, and at low temperatures, LTF, as shown by the squares and diamonds respectively in Fig. 12. The low temperature offset accounts for nearly the entire frequency shift observed at the upper turnover. Sweeping did improve the radiation resistance of the MH11-19 crystal; the unswept sample showed a -3.36 ppm offset while in the swept crystal the offset was -1.05 ppm. The low temperature offset was reduced in the swept crystal as well.

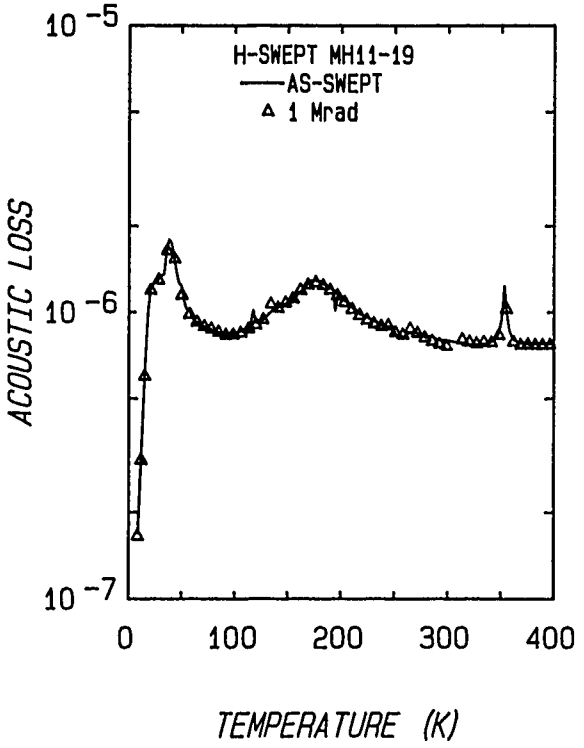


Figure 11. The acoustic loss versus temperature curves for the H-swept MH11-19 are shown. No radiation-induced changes were observed.

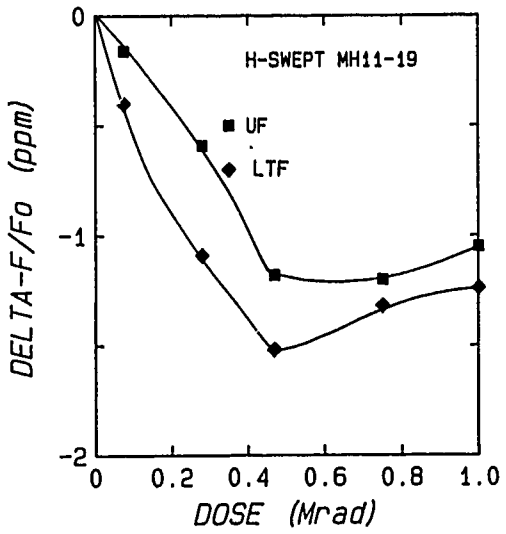


Figure 12. While no changes were seen in the acoustic loss radiation did cause frequency offsets at both the upper turnover, UF, and at low temperatures, UTF, in the H-swept MH11-19 crystal. Most of the offset seems to be associated with the low temperature shift.

A hydrogen swept 5 MHz AT-cut blank from the high-aluminum content bar HA-A was also tested. Figure 13 shows the fractional-frequency offsets versus radiation dose measured at the upper turnover (squares) and at low temperatures (diamonds). Just as in the H-swept MH11-19 crystal both offsets are negative and the low temperature offset seems to be the controlling factor.

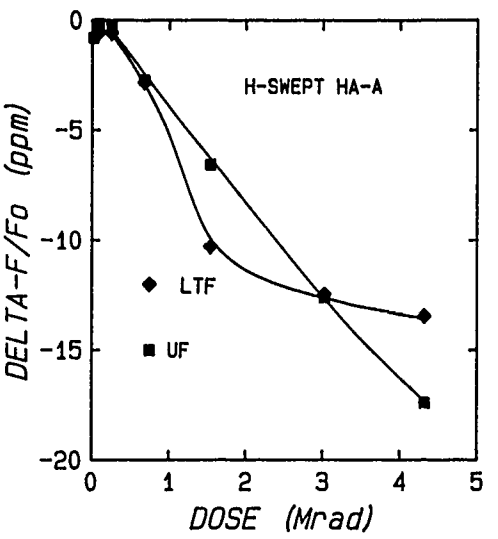


Figure 13. The radiation-induced offsets at the upper turnover, UF, and at low temperatures, LTF, in a H-swept HA-A crystal are shown. The offsets are much larger in this 60-70 ppm Al sample than in MH11-19.

Table 4 lists the observed maximum offsets at the upper turnover and at low temperatures that we have found in swept quartz following irradiation at room temperature.

Table 4. Frequency offsets in swept quartz.

Sample	Al (ppm)	Δf/f(UF) (ppm)	Δf/f(LTF) (ppm)
ME24-48	0.2	0.1	0.3
MH11-19	12	-1.24	-1.05
HA-A	60	-13.5	-17.4

The radiation-induced offsets are larger for the less pure crystals. With the exception of the HA-A sample sweeping improved the radiation hardness of the crystals tested. HA-A is somewhat unusual as the low temperature offset compensated the offsets produced by the production of the Al-hole centers in the unswept sample.

Low Temperature Offset

The origin of the low temperature frequency shift is most probably caused by small changes in the elastic moduli due to the modification of the defects present in quartz. As shown in Table 2, we consistently observed a positive low temperature shift in the Na-swept crystals with irradiation. This shift increased with the aluminum content of the crystal. In unswept crystals the offset was smaller; and both positive and negative shifts were observed. In two of the three H-swept samples the low temperature offset was negative and increased with increasing aluminum content. We describe below two additional observations of low-temperature offsets.

When swept quartz is irradiated at room temperatures the Al-OH center is converted into Al-hole centers. However, the hydrogen returns to the aluminum site and annihilates the hole. This recovery is the basic cause of the smaller steady-state offsets usually seen in swept quartz and of the transient frequency shift reported by King and Sander [7,8,9]. If instead the radiation is carried out at temperatures below 100 K the hydrogen remains trapped as an interstitial atom [18, 26] and the Al-hole center is stable as long as

the sample remains cold. King and Sander reported that irradiation at 80 K produced a loss peak at 100 K that they attributed to the Al-hole center. Later, Martin et al. [21,23] presented further evidence that this 100 K peak and additional peaks at 23 K and 135 K were associated with the presence of the Al-hole center. While all three peaks are associated with the presence of the Al-hole center symmetry arguments show that, in quartz, a given defect can have only two loss peaks [31]. The solid line in Fig. 14 shows the as-swept acoustic loss versus temperature spectrum for a hydrogen-swept Toyo Supreme Q 5 MHz 3rd overtone AT-cut crystal. The peak near 30 K in the as-swept curve is due to an interfering mode. The smaller peaks near 100 K and 130 K are probably also interfering modes. After obtaining the as-swept curve the sample was cooled below 90 K and given a strong electron irradiation. The crystal was then cooled to 9 K and the acoustic loss curve shown by the dashed line in Fig. 14 was taken. The Al-hole associated peaks at 23 K, 100 K and 135 K were produced by the low temperature irradiation. After the crystal was warmed to 300 K overnight the acoustic loss spectrum returned to that shown by the original curve.

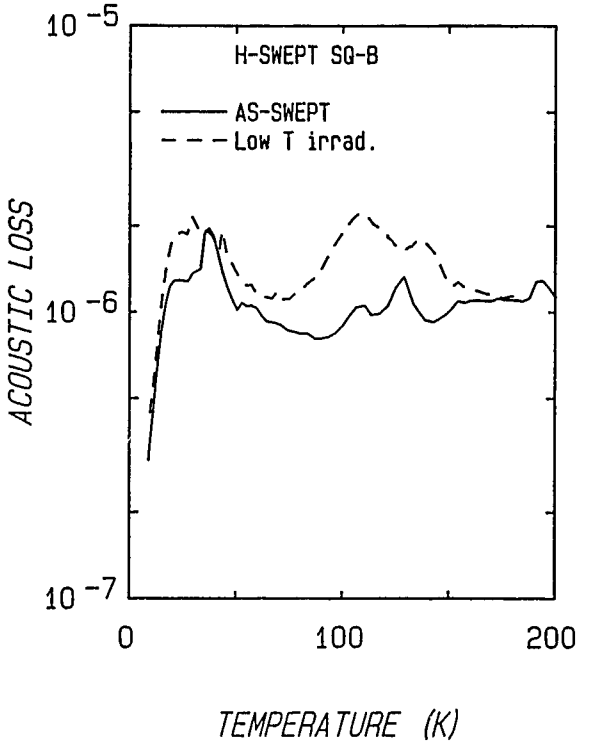


Figure 14. The as-swept acoustic loss versus temperature spectrum for a H-swept Toyo SQ-B crystal is shown. An irradiation at temperatures below 90 K produced the 23 K, 100 K, and 135 K Al-hole loss peaks. The three peaks disappeared after the sample was warmed to room temperature.

Figure 15 shows the the frequency versus temperature curves that were taken at the same time as the loss. The heavy solid curve shows that the irradiation produced a negative offset of about 1 ppm a the lower temperatures. After the crystal was annealed at 300 K the frequency recovered to the original curve as shown by the dots superimposed on top of the light solid line.

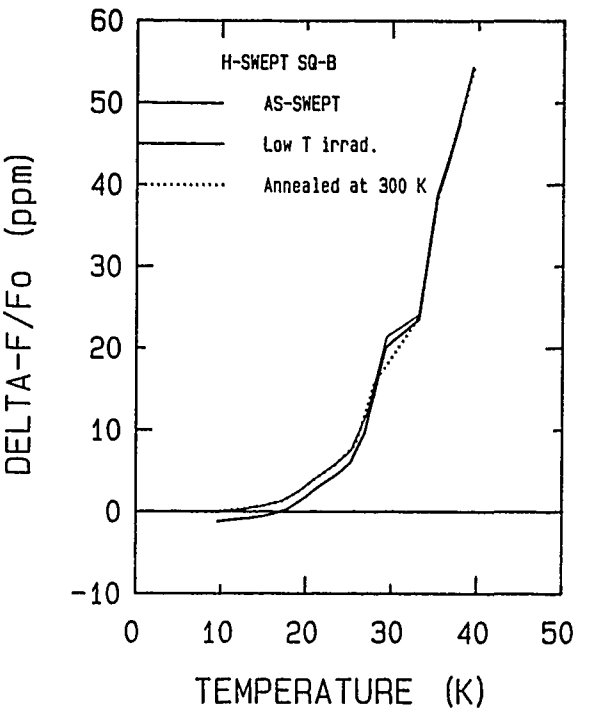


Figure 15. The frequency versus temperature curves for the H-swept SQ-B crystal taken at the same time are shown. The irradiation produced a -1 ppm offset at the lowest temperatures. This offset disappeared after the crystal was annealed at room temperature.

Measurements made on a blank in a gap holder of fixed dimensions allow a direct comparison of the frequency of the blank as a function of electrodiffusion with different ions. The Li-swept HA-A blank used in the experiment described above was then swept with sodium to convert the Al-Li centers into Al-Na centers and remeasured. This was the Na-swept HA-A blank used for the radiation test given in Table 2. Another Li-swept blank from our Sawyer Premium Q bar PQ-E was also measured over the 9-400 K range, then swept with sodium and remeasured. A second Li-swept HA-A blank was measured and then swept with hydrogen to convert the Al-Li centers into Al-OH centers. Table 4 summarizes the fractional-frequency offsets at the upper turnover and at low temperatures induced by these sweeping experiments.

Table 5. Sweeping induced offsets in blanks.

Sample	Al (ppm)	Δf/f(UT) (ppm)	Df/f(9K) (ppm)
Li replaced by Na			
PQ-E	12	-45	-12
HA-A	60	-500	-97
Li replaced by H			
PQ-E	12	-2.4	2.8
HA-A	60	-22	5.9

The replacement of interstitial lithium adjacent to the aluminum by sodium produces a negative frequency offset at both the upper turnover, UF, and at low temperatures. In the HA-A sample, the lithium to sodium conversion produced an offset at the upper turnover slightly larger in magnitude than than the 406 ppm radiation-induced offset in the Na-swept sample. However, the offset is still much smaller than one would predict from the strength of the Al-Na 53 K loss peak. The replacement of lithium by sodium also caused changes in the frequency versus temperature curve similar to those shown in Fig. 3 for the MH11-19

crystals. Air or electrodiffusion sweeping replaces the alkalis with hydrogen. The sweeping induced offsets shown in Table 5 for the HA-A blank suggest that the replacement of lithium with hydrogen causes a small change in the "apparent" angle of the crystal.

Low temperature fractional-frequency offsets were observed following both irradiation and alkali replacement by electrodiffusion. In the HA-A samples these offsets were larger than could be reasonably produced by additional anelastic relaxation loss peaks at temperatures below 8 K or 9 K. It seems reasonable then to suggest that they are caused by changes in the elastic moduli. The elastic moduli represents a type of average of the interatomic force constants and, therefore should be sensitive to the substitution and/or modification of impurities and defects. Recently, James [32] reported new measurements of the elastic and dielectric constants of α-quartz. He observed differences between the elastic moduli at 25°C of GEC grown very-high-purity quartz and of Russian quartz with 0.8 ppm Al. James reported that most of these differences could be attributed to the anelastic relaxation of the Al-Na present. For rotated Y-plates such as the AT-cut the elastic modulus, c', used in Eq. 1 is given by

$$c' = c_{66}\cos^2\theta + c_{44}\sin^2\theta - c_{14}\sin 2\theta \qquad (5)$$

where θ is the angle of rotation about the X-axis [16]. Consequently, measurements on crystals cut with three different angles such as an AT-cut, BT-cut and a Y-plate would be needed to determine the effect on a specific modulus. Aoki, Norisawa and Sakisaka [32] measured the radiation-induced frequency offset at room temperature of three plates cut at three different angles varying slightly from that of the AT-cut so as to minimize the variation with sample inhomogeneity. They reported that ratio of the fractional change in c_{44} to that in c_{14} to that in c_{66} was 1.2:3.1:1.0. Such shifts in the moduli would probably cause changes in the apparent "angle" of the crystal. It would be interesting to repeat our alkali exchange using the gap-holder technique on AT-cut, BT-cut, and a Y-plate blanks taken from one bar of quartz such as MH11-19 or SQ-B.

SUMMARY AND CONCLUSIONS

As expected, the large positive radiation-induced frequency offsets were observed in Na-swept quartz; and the magnitude of the offsets increased with increasing aluminum concentration. However, the size of the offset in the Na-swept material was consistently smaller than one would predict from the change in the height of the Al-Na 53 K loss peak. This smaller than predicted magnitude was caused by radiation-induced changes in the frequency-temperature curve such that Na-swept AT-cut crystals behave as if they are cut at a smaller angle. Radiation returned the f-T curve to its regular angle. An unexpected positive frequency offset was also observed at the lowest temperatures reached. This offset grew with increasing dose and was larger for crystals with higher aluminum contents.

Both positive and negative radiation-induced offsets at the upper turnover were observed in unswept and Li-swept crystals. Again, the crystals with the largest aluminum content showed the largest offsets. A low temperature radiation-induced offset was also found in these crystals. In one case, this low temperature offset partly compensated the offset caused by the Al-hole center loss peaks. If the low temperature offset is subtracted from that at the upper turnover most of the observed frequency shift can be explained by the changes in the Li-growth-defect loss peak and the Al-hole center loss peaks.

The negative fractional-frequency offsets at the upper turnover were observed in all but one H-swept crystal. With one exception, these offsets were always smaller than in an unswept (or Li-swept) crystal taken from the same quartz bar and they were smaller in low aluminum crystals. The exception was the high aluminum content crystal where the low temperature offset compensated offset due to the Al-hole center loss. A radiation-induced low temperature offset was also found in the H-swept crystals.

An unexpected radiation-induced frequency offset was found in all of the crystals tested. This offset was also produced when a H-swept crystal was irradiated low temperatures. This low temperature irradiation produced the Al-hole center loss peaks and induced a -1 ppm frequency shift at 9 K. The frequency shift and the loss peaks disappeared after the crystal was warmed to room temperature. When a Li-swept blank with 60-70 ppm aluminum was swept with sodium its frequency at 9 K decreased by 97 ppm as measured in a gap holder of fixed geometry. When a similar Li-swept blank was swept with hydrogen its low temperature frequency increased by 5.9 ppm. The origin of the low temperature offset is probably due to changes in the elastic moduli brought about by the radiation- or electrodiffusion-induced modification of the defects.

ACKNOWLEDGMENTS

This work was supported in part by the Solid State Sciences Division, Rome Air Development Command, USAF; and by Sandia National Laboratories. The author thanks J. F. Balascio of Motorola for supplying the MH series quartz stones. The author also thanks A. F. Armington, A. Kahan, J. C. King, D. R. Koehler, L. E. Halliburton, H. B. Hwang, H. Lipson, A. Lopez, and A. S. Nowick for a number of helpful discussions.

REFERENCES

1. J. C. King and D. R. Koehler, "Radiation effects on resonators, in Precision Frequency Control, E. A. Gerber and A. Ballato, eds., Orlando, FL: Academic Press, 1985, vol. 1. pp. 147-159.

2. B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckmelter, "Quartz crystal radiation effects," IEEE Trans. Nucl. Sci., vol. NS-17, pp. 217-221, 1970.

3. R. A. Poll and S. L. Ridgway, "Effects of pulsed ionizing radiation on some selected quartz oscillator crystals," IEEE Trans. Nucl. Sci., vol. NS-13, pp. 130-140, 1965.

4. T. J. Young, D. R. Koehler, and R. A. Adams, "Radiation induced frequency and resistance changes in electrolyzed high purity quartz," in Proc. 32nd Ann. Freq. Control Symposium, EIA, May 1978, pp. 34-42.

5. A. Kahan, F. Euler, H. Lipson, C. Chen and L. E. Halliburton, "Radiation effects in vacuum swept quartz," in Proc. 41st Ann. Freq. Control Symposium, IEEE, May 1987, pp. 216-221.

6. P. Pelligrini, F. Euler, A. Kahan, T. M. Flanagan, and T. F. Wrobel, "Steady state and transient radiation effects in precision quartz oscillators," IEEE Trans. Nucl. Sci. vol. NS-25, pp. 1267-1273, 1976.

7. J. C. King and H. H. Sander, "Transient changes in quartz resonators following exposure to pulse radiation," Radiat. Eft. vol. 26, pp. 203-212 1975.

8. J. C. King and H. H. Sander, "Rapid annealing of frequency change in crystal resonators following exposure to pulse x-rays," IEEE Trans. Nucl. Sci., vol. NS-19, pp. 23-32, 1972.

9. J. C. King and H. H. Sander, " Transient change in Q and frequency of AT-cut crystal resonators following exposure to pulse x-rays," IEEE Trans. Nucl. Sci., vol. NS-20, pp. 117-125, 1973.

10. D. R. Koehler and J. J. Martin, "Radiation-induced transient acoustic loss in quartz crystals," J. Appl. Phys. vol. 57, pp. 5205-5209, 1985.

11. J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results from gamm ray and proton beam testing of quartz resonators," in Proc. 38th Ann. Symposium on Freq. Control, IEEE, May 1984, pp. 63-72.

12. J. Suter and R. Maurer, "Low and medium dose radiation sensitivity of quartz crystal resonators with different Al-impurity content," in Proc. 40th Ann. Symposium on Freq. Control, IEEE, May 1986, pp. 134-139.

13. J. J. Suter and R. H. Maurer, "Low- and medium-dose radiation sensitivity of quartz crystal resonators with different aluminum impurity content," IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, vol. UFFC-34, pp. 667-673, 1987.

14. T. Flanagan, R. Leadon, and D. Shannon, "Evaluation of mechanisms for low-dose frequency shifts in crystal oscillators," in Proc. 40th Ann. Symposium on Freq. Control, IEEE, May 1986, pp. 127-133.

15. Leslie J. Palkuti and Quang T. Truong, "An x-ray irradiation system for total-dose testing of quartz resonators," in Proc. 38th Ann. Symposium on Freq. Control, IEEE, May 1984, pp. 55-62.

16. D. B. Fraser, "Impurities and anelasticity in crystalline quartz," in Physical Acoustics, W. P. Mason, ed. New york: Academic Press, 1968. vol. V, chap. 2, pp. 59-110.

17. J. Weil, "The aluminum centers in alpha-quartz," Radiat. Eff. vol. 26, pp. 261-265, 1975.

18. J. A. Weil, "A review of electron spin resonance spectroscopy and its application to the study of paramagnetic defects in crystalline quartz," Phys. Chem. Minerals, vol. 10, pp. 149-165, 1984.

18. L. E. Halliburton, J. J. Martin, and D. R. Koehler, "Properties of Piezoelectric materials," in Precision Frequency Control, E. A. Gerber and A. Ballato, eds. Orlando, FL: Academic Press, 1985, vol. 1, pp. 1-45.

19. A. Kats, "Hydrogen in alpha-quartz," Phillips Res. Rpts., vol. 17, pp. 133-279, 1962.

20. J. J. Martin, Ho. B. Hwang, H. Bahadur, and G. A. Berman, "Room temperature acoustic loss peaks in quartz," J. Appl. Phys. in press.

21. J. J. Martin, "Aluminum related acoustic loss in AT-cut quartz crystals," J. Appl. Phys., vol. 56, pp. 2536-2540, 1984.

22. J. Toulouse and A. S. Nowick, "Dielectric loss of quartz crystals electrodiffused with either Na^+ or Li^+," J. Phys. Chem. Solids, vol. 46, pp. 1285-1292, 1985.

23. Augusto R. Lopez, Ho B. Hwang, and J. J. Martin, "A production study of acoustic loss related defects in quartz," in Proc. 42nd Ann. Symposium on Freq. Control, IEEE, June 1988, ppl76-183.

24. F. Euler and A. Kahan, "Radiation effects and anelastic loss in germanium-doped quartz," Phys. Rev. B, vol. 35, pp. 4351-4359, 1987.

25. J. J. Martin, Augusto R. Lopez, A. F. Armington, and J. F. Balascio, "Evaluation of resonators fabricated fron high quality quartz," in Proc. 42nd Ann. Symposium on Freq. Control, IEEE, June 1988, pp. 162-169.

26. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, "Radiation effects in crystalline SiO_2: the role of aluminum," J. Appl. Phys., vol. 57, pp. 3565-3674, 1981.

27. Augusto R. Lopez, J. D. West, and J. J. Martin, "Electrodiffusion of ions in alpha-quartz," Mat. Res. Soc. Symp. Proc., vol. 60, pp. 451-457, 1986.

28. J. J. Martin, Ho B. Hwang, and T. M. Wilson, "A radiation qrowth study of acoustic loss related defects in alpha-quartz," Proc. 40th Ann. Symposium on Freq. Control, IEEE, May 1986, pp. 32-38.

29. J. C. King, "Electrolysis of synthetic quartz; effect upon resonator performance," Proc. IEEE (London), vol. B109, Suppl. 22, pp. 295-301, 1959.

30. J. J. Martin, "Electrodiffusin (sweeping) of ions in quartz-a review," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 35, pp. 288-296, 1988.

31. A. S. Nowick and B. S. Berry, Anelastic Relaxation in Crystalline Solids, Academic Press, NY, 1972.

32. B. James, "A new measurement of the basiec elastic and dielectric constants of quartz (invited paper)," in Proc. 42nd Ann. Symposium on Frequency Control, IEEE, June 1988, pp. 146-154.

33. T. Aoki, K. Norisawa, and M. Sakisaka, "Frequency change and elastic constants of quartz irradiated by 1 MeV electrons," Japanese J. Appl Phys. vol. 15, pp. 749-754, 1976.

43rd Annual Symposium on Frequency Control - 1989

STACKED CRYSTAL FILTERS IMPLEMENTED WITH THIN FILMS

K.M. LAKIN,* G.R. KLINE,* R.S. KETCHAM,*

J.T. MARTIN,* and K.T. McCARRON

Microelectronics Research Center
Iowa State University
1925 Scholl Road
Ames, Iowa, 50011

ABSTRACT

High performance ultraminiature cointegrated bandpass filters are required for fully integrated RF and microwave subsystem developments in communication, navigation, and other signal processing systems. The use of microwave acoustics allows a 10^5 reduction in wavelength and a corresponding reduction in the size of wavelength scaled devices. For the frequency range 1 to 2 GHz, conventional filtering techniques have size and/or performance characteristics that prevent their cointegration with active devices onto a common chip or die. Filters based on the piezoelectric Thin Film Resonator (TFR) technology are being studied for applications requiring discrete as well as integrated filters.

The thin film implementation of the Stacked Crystal Filter (SCF) has produced low insertion loss filters at 1 to 2 GHz frequencies. Consequently new applications are being addressed that were formerly beyond the performance range of microwave acoustic devices. These applications involve the cointegration of frequency control devices with active devices on substrates such as GaAs and Si for increased systems performance and in order to minimize parasitic effects. Specifically, the SCF addresses the need for low insertion loss receiver front-end filters by achieving 2-pole filter insertion losses near 1.5 dB.

This paper reviews the SCF technology and reports on the latest experimental results on low insertion loss filters for the 1 to 2 GHz frequency range as fabricated on Si and GaAs substrates and on the overmoded devices on spinel substrates.

INTRODUCTION

Filter Requirements

In modern communication systems there is an ever increasing demand for high performance passive frequency control elements that are complementary to the active circuit devices in terms of size and performance. The demand is particularly pressing for relatively narrow bandwidth (less than 3%) filters and high Q resonators that can be integrated onto the active circuit chip or are at least of comparable size for hybrid circuits. Considerable emphasis is being placed on the UHF through 2 GHz frequency range where cellular radio, telemetry, and navigation systems are expanding.

* now with TFR Technologies, Inc.

Filters for use in the front-end of a receiver have particular size and performance requirments that are difficult to meet if they are to be integrated onto the semiconductor chip. Candidate filters for use in a GPS navigation receiver are shown in the size versus performance diagram of Fig. 1. For on-chip integration a two-pole filter should be considerably smaller than the typical integrated circuit die, with device area to the left of the vertical dotted line in Fig. 1, and the insertion loss for the front-end filter must be less than 2 dB below the horizontal line in Fig. 1. Filters made by electromagnetic transmission line sections give satisfactory performance although their size is too large even when fabricated on high dielectric constant substrates. Lumped element filters suffer from high inductor losses at microwave frequencies and are therefore unsatisfactory. Filters based on SAW devices suffer from high insertion losses and are generally not integrable due to materials compatibility constraints.

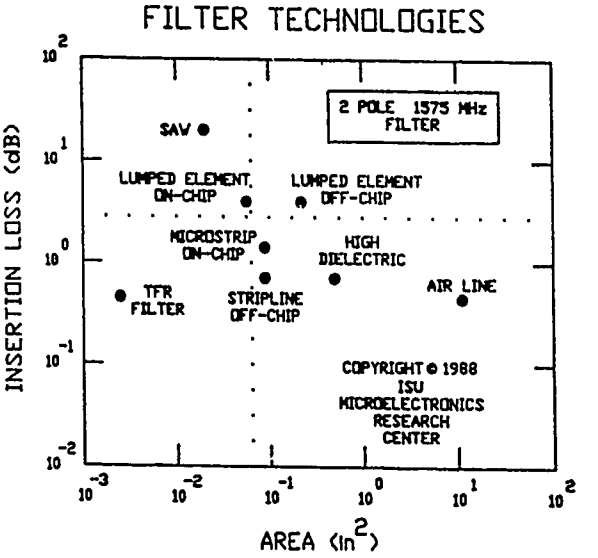


Fig. 1. Performance versus size for a class of two-pole filters for frequencies in the 1-2 GHz range. Integration of receiver front-end filters requires the filter area to be substantially to the left of the vertical dotted line, and insertion loss must be less than 2 dB. The stacked crystal filter, (SCF), is one kind of TFR device.

The issues of filter size and performance can be addressed with microwave acoustic devices because of wave and material properties [1].

CH2690-6/89/0000-536 $1.00 C 1989 IEEE

Filter size is reduced because of the significantly shorter wavelength of sound waves and insertion losses are low due to high material Q at UHF and low microwave frequencies. Only those filters implemented with microwave acoustics and high quality thin piezoelectric films currently meet size, performance, and integrability conditions.

There are three general classifications of bandpass filters that have been implemented with electromechanical resonators using piezoelectric materials; the conventional ladder filter, the Monolithic Crystal Filter, (MCF), and the more recent stacked crystal filter (SCF) [2-4]. All of these have been implemented with thin piezoelectric films with the greatest emphasis being placed on the SCF [5].

The Basic Stacked Crystal Filter.

The SCF, shown schematically in Fig. 2, is characterized by two resonator plates, or transducers, in direct contact separated only by a common ground plane electrode. The SCF differs from the ladder filter in that the resonators are acoustically coupled and not simply electrically circuit coupled. The SCF achieves resonator coupling via direct acoustic radiation in the thickness direction whereas the MCF resonators are laterally coupled. The SCF is generally implemented as a one-dimensional wave propagation device with a large diameter-to-thickness ratio while the MCF is clearly two dimensional with nearly equal thickness and width dimensions. It will be shown that the SCF performs as a single two-port resonator rather than two coupled resonators.

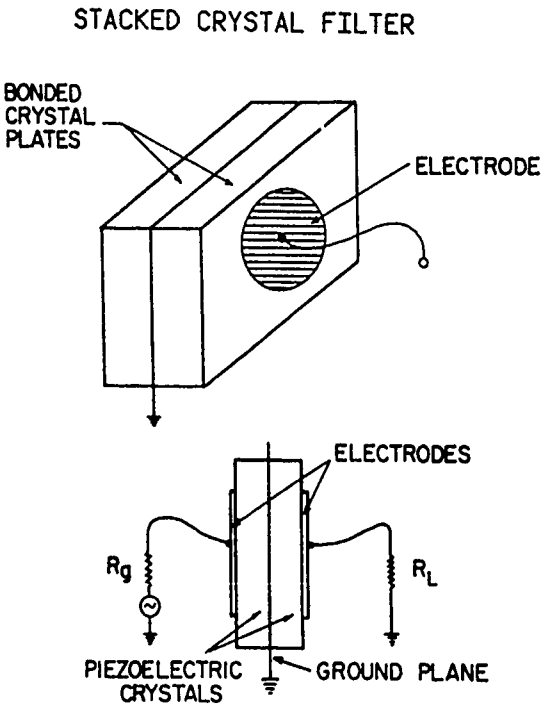


Fig. 2. Schematic representation of the SCF showing the two piezoelectric plates (films), common ground plane, and two I/O electrodes.

The one-dimensional nature of the SCF leads to some interesting operating features of importance to system design. As will be apparent in the experimental results, the single mode wave propagation feature leads to a filter with low spurious responses beyond the normal mode resonances. The general nature of the filter passband is illustrated in Fig. 3 for a theoretical calculation of the filter response for a frequency range covering the three dominate resonances. The inset for Fig. 3 illustrates the three resonances associated with the three main filter responses. The lowest frequency response corresponds to the fundamental thickness resonance where there is a half wavelength across the entire structure. The response is weaker in this case because there is only a quarter wavelength across each piezoelectric plate. The second overtone corresponds to a half wavelength across each plate and therefore has the strongest coupling. Consequently the filter exhibits the widest bandwidth and lowest insertion loss under these conditions. The experimental devices reported later employ this mode. Finally, the third overtone is a weaker coupling because of the poorer phase matching conditions between the wave and applied fields. Similarly, the fourth overtone response is entirely missing because the fields entirely cancel.

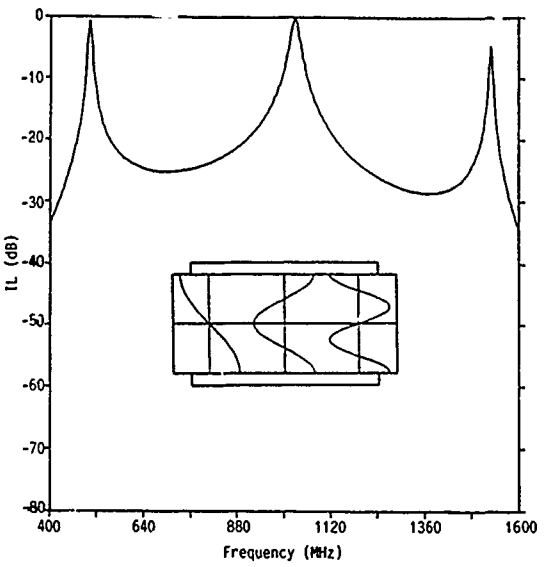


Fig. 3. Representive theoretical response of SCF showing the three main resonances and, inset, the corresponding internal acoustic displacements for fundamental, second, and third resonances.

As a one-dimensional device the SCF offers some flexibility in its design area for purposes of impedance matching. The SCF can be constructed so that the area is large enough to scale impedances to 50 ohm input/output impedance levels as required for front-end filters in receivers. Practical filters could be scaled for lower impedance levels for power handling at UHF and low microwave frequencies. In general, the weak spurious responses of the filters is reduced for larger diameter-to-thickness ratios.

Single Crystal Plates

The first experimental work on the SCF was the work of Ballato and Lukaszek [2] who bonded single crystal quartz plates together and demonstrated several novel features of the device. Unique in the crystal plate implementation is the ability to simply rotate the shear wave polarization of one plate relative to the another in order to couple two quasi-shear wave modes to produce a multipole response. This early work also demonstrated the relatively low spurious response of the SCF compared to the MCF under development in the same time period.

The practical difficulty of bonding quartz plates to obtain a reproducible shear wave bond proved the major stumbling block in further development of the SCF. It appeared that the manufacturibility of the device suffered relative to the MCF and the SCF was abandoned.

However, the unique features of the device were sufficient to stimulate an effort in thin film synthesis of the SCF where the problem of crystal plate bonding would be solved by the direct deposition of the piezoelectric resonators one on another in a five layer structure.

DEVICE MODELING

A detailed multi-mode modeling of the SCF was carried out by Ballato [6,7] using distributed transmission line models. However, modeling of the SCF reviewed here is based on the Mason lumped element transmission line model and the extension of the model into an equivalent circuit PI network representation [8]. The modeling is accurate because the Mason model is an exact representation of the one dimensional plane wave excitation boundary value problem. The modeling of the SCF is simplified further by reducing the three-port Mason model to a two-port network where one of the acoustic ports is passively terminated as shown in Fig. 4. In this modeling the left and right transmission line sections represent metal electrodes.

The overall network is analyzed using the ABCD matrix cascading technique common to electrical network analysis. A typical computed response using this method is shown in Fig. 4 for a single section SCF and for two sections connected in series electrically. In the computed results finite material Q's are accounted for in the Mason model of the piezoelectric film and metal transmission line sections.

For design purposes the direct Mason model approach is not satisfactory and a lumped element circuit model is more useful. In order to support design efforts an equivalent circuit model was derived from the full Mason model by a process illustrated in Fig. 5. Here two Mason models are arranged back-to-back to simulate an ideal SCF having zero metal electrode thicknesses and identical piezoelectric films. Then by a series of PI-to-T and reverse transforms a complex PI network is obtained. Finally using

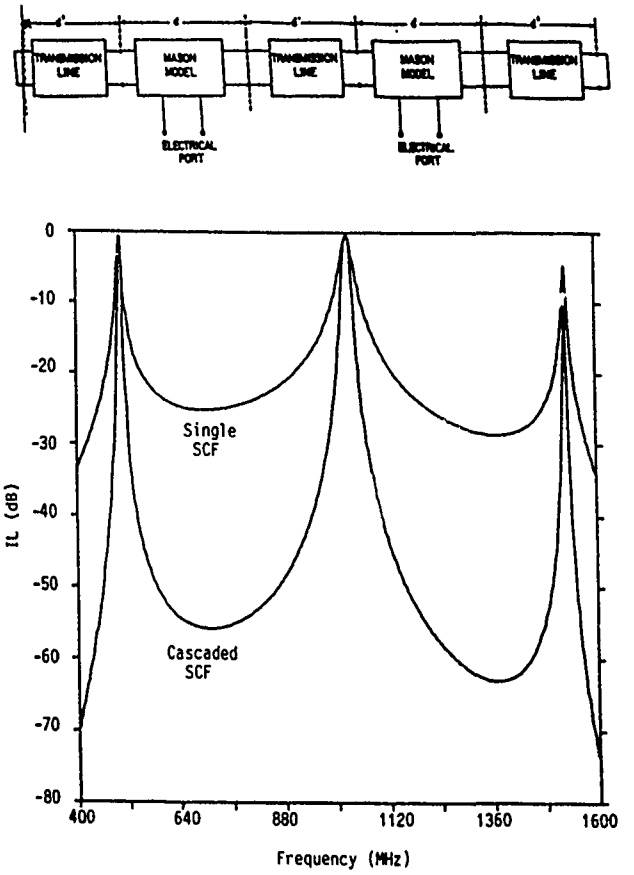


Fig. 4. Modeling schematic of the SCF using the Mason plane wave model. Calculated results are for one and two pole filters taking into account metal thicknesses and material losses.

trigonometric expansions about the operating frequency a simple lumped element equivalent circuit for the SCF is obtained, Fig. 6. This model has been found to be accurate across the filter passband and well into the skirt region.

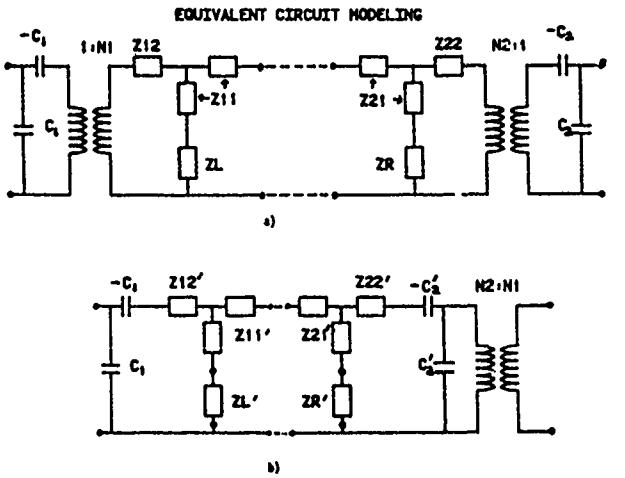
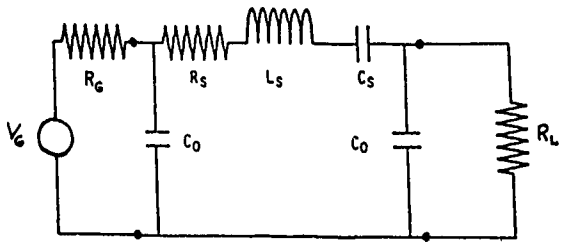


Fig. 5. Schematic representation of the lumped element Mason model approach to the PI network circuit model. Two Mason models are used back-to-back.



C_O = individual resonator capacitance
$C_S = \alpha K^2 C_O$
α = .454
K^2 = piezoelectric coupling coefficient
L_S = inductor series resonant with Cs
$R_S = 1/(\omega C_S Q)$
Q = resonator Q

$$IL = 10\log_{10}\left[(1+b)^2+a^2(\Omega-\Omega^{-1})^2\right]$$

where $a = (q+q^{-1})/(2\alpha K^2)$
$q = \omega C_O R_g$
R_g = load/source resistance
$b = a/Q$
Ω = normalized frequency = 1 series resonance.

$$IL_{min} = 20\log_{10}\left(1 + \frac{1}{\alpha K^2 Q}\right)$$

Fig. 6. PI network model of the SCF and associated circuit parameters and derived results. The model assumes zero electrode thicknesses and identical piezoelectric regions. The model is accurate across the filter passband and down the steep skirts.

Insertion loss and other parameters of the SCF can now be derived in terms of material prameters from the PI network model and placed in a form that is useful for filter disign. The minimum insertion loss occurs when the magnitude of the reactance of the static film capacitance, Co, is equal to the source and load resistances. It can also be shown that this is the condition for maximum bandwidth for the network. The fractional bandwidth under these conditions is approximately $.45K^2$.

It is apparent from the PI network model that a significant feature of the SCF is that the resonator area can be adjusted to meet impedance matching conditions. Since the device area is scaled impedance matching can be accomplished without incurring excessive electrode resistance losses. This feature allows the electrode area to be designed for a particular frequency and impedance level. Electrode areas corresponding to 5 to 100 ohms seem practical at UHF and microwave frequencies.

THIN FILM IMPLEMENTATION

The thin film approach to the implementation of the SCF was driven by the need for the success of the Thin Film Resonator, (TFR) [9,10] efforts were directed toward the SCF using the film growth process as a means of eliminating the resonator interface bonding problems. The key elements in the SCF implementation are the basic TFR microelectronics fabrication processes and the production of high quality piezoelectric films.

The basic TFR structures [5] shown in Figs. 7 and 8, consist of piezoelectric membranes supported in some manner by the surrounding substrate or wafer. Essential elements common to these structures are the piezoelectric film, the metal electrodes used to apply the electrical signals, and two reflecting surfaces which allow the resonance to occur. The method of fabricating these structures has roots in the basic microelectronics materials processing technology.

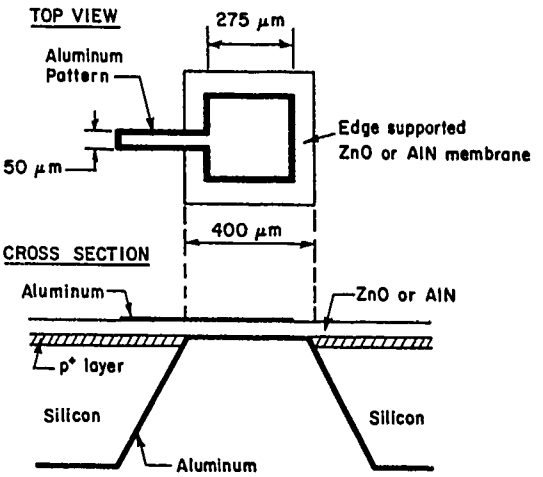


Fig. 7. Schematic representation of device topology used for thin film resonator and SCF construction. Films are approximately 5 micrometers thick for 1GHz frequencies. Device area is a function of the desired impedance level.

In the case of silicon substrates, Fig. 7, the aluminum nitride film is first grown on the boron doped p+ layer and then the piezoelectric film is isolated by etching away the p+ silicon. When the piezoelectric is AlN, the material strength is such as to allow a free standing piezoelectric "plate" suitable for further device processing.

A similar "via etch" process is shown for GaAs substrates in Fig. 8a. Here the membrane structure is formed by well known GaAs microwave integrated circuit processing techniques. The via may be formed by plasma processing or by wet chemical processing. However, the vias required for TFR devices are generally larger than those required for microwave integrated circuit grounding purposes.

A more universal structure, illustrated by the bridge resonator shown in Fig. 8b, is formed by an undercutting etch which forms the air

gap region [11]. This general technique has its roots in integrated circuit cross-over metalization processing and is used here to fabricate SCF devices on substrates where the via etch process is not possible or difficult for materials compatibility or processing related reasons.

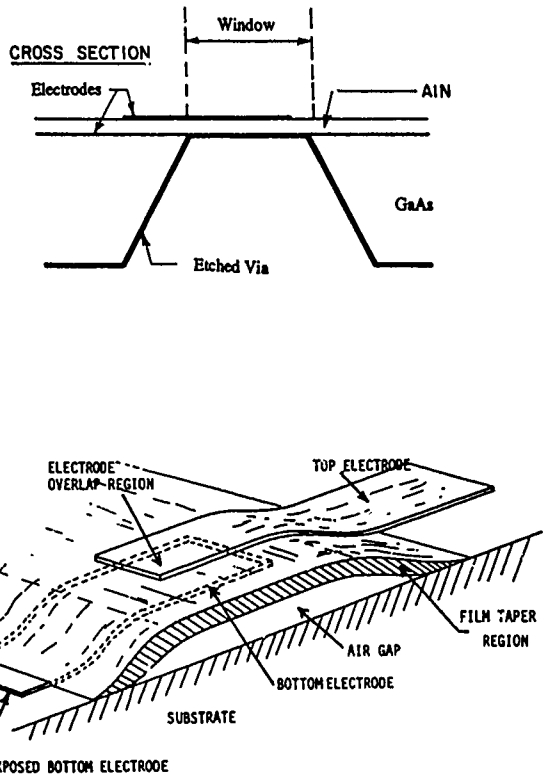


Fig. 8. Schematic views of device geometries as fabricated by, a). the via etch process and, b), the bridge process. The bridge process is the least substrate dependent of the various processes.

Piezoelectric film deposition is a key element in the fabrication of the thin film SCF. Depositiion conditions for ZnO and AlN have been discussed in the literature at length [12-14]. The most successful technique for obtaining high quality piezoelectric films is the reactive deposition process using a metal target, mixture of reactive and noble gasses, and a planar magnetron target configuration. In addition, for purposes of integration with active devices, film deposition must take place at relatively low temperatures compared with active device processing.

Film properties which must be controlled and optimized are piezoelectric coupling coeficient, in-plane mechanical strain, density, crystallographic orientation, and thickness uniformiy. Some experimental results are shown in Fig. 9 for coupling coefficient as a function of die position on a 3 inch diameter Si wafer. In this case the film was grown in a ultra-high vacuum chamber to insure minimum contamination from oxygen and water vapor. Measurements of coupling coefficient and resonator Q are mapped onto a common graph, Fig. 10, where single section SCF insertion loss is a parameter. Ideally the cluster of experimental points should be close together in order to give highest device yield and performance, as will be discussed below in relation to filter results.

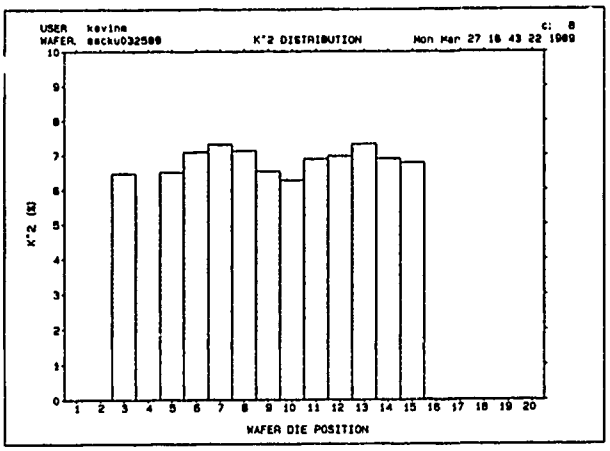


Fig. 9. Piezoelectric coupling coefficient variation as a function of die position across a 3 inch silicon wafer. The film was grown in an ultra-high vacuum system.

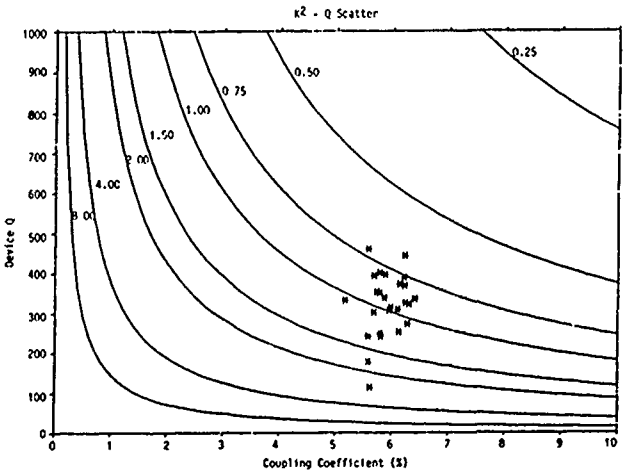


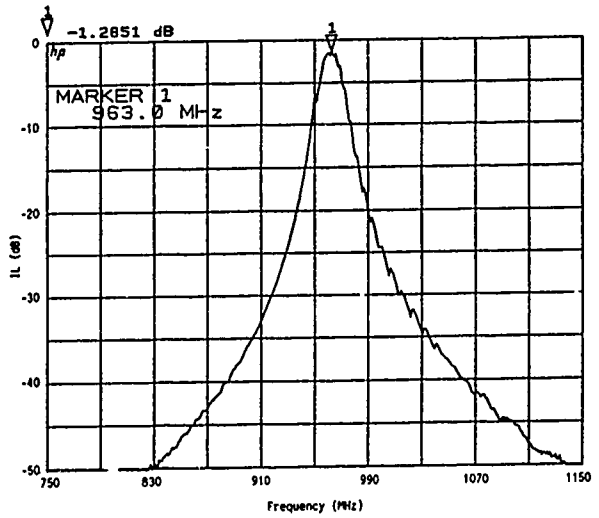
Fig. 10. Material parameter map for coupling coefficient and Q. The running parameter lines are for various levels of single pole insertion loss. The data points are for material parameters measured on a particular silicon wafer.

EXPERIMENTAL FILTER RESULTS

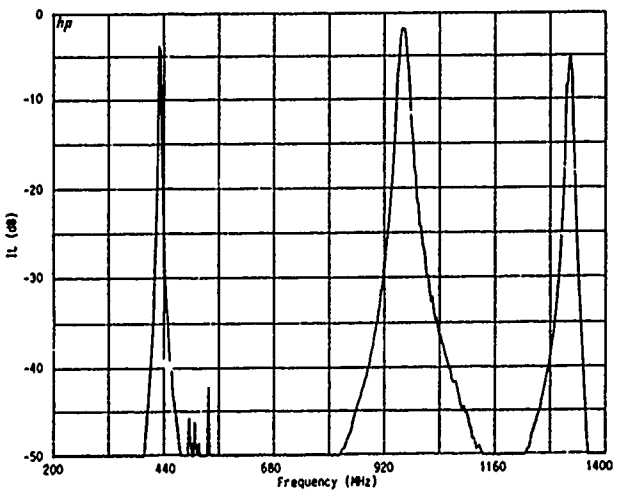
Three types of SCF have been investigated in feasibility studies of low insertion loss filters depending on the initial substrate used for support of the film. Most of the work has been done on Si substrates because of the advanced processing already available and low cost of the starting material.

Demonstrations of SCF device fabrication were also carried out on GaAs using processing nearly identical to that used in Si. Finally, the over-moded SCF is reviewed for completeness.

Representive responses from filters fabricated on silicon substrates are shown in Figs. 11 and 12 that confirm or illustrate various features of the SCF device. The device of Fig. 11a is a low insertion loss, IL<1.3 dB,two pole filter with a center frequency of 963 MHz. The filter bandwidth is 11 MHz (1%), A wide frequency scan of the filter, Fig. 11b, illustrates the three main filter responses and the clean out-of-band response expected for the SCF. The reduced bandwidth is due to mismatch since the filter geometry was designed for 50 ohms at just over 1 GHz.



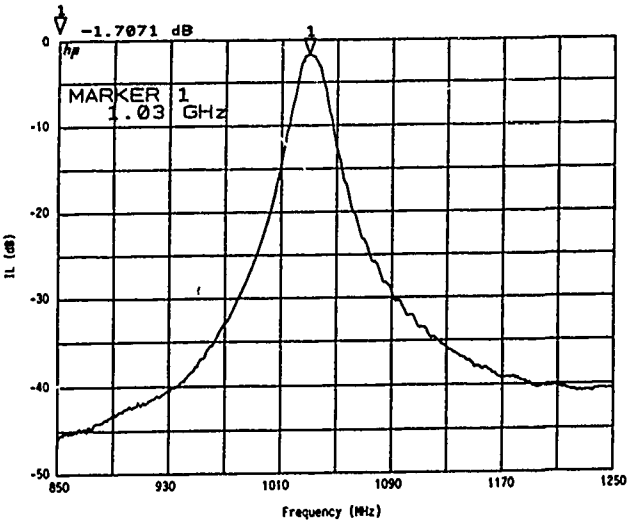
a)



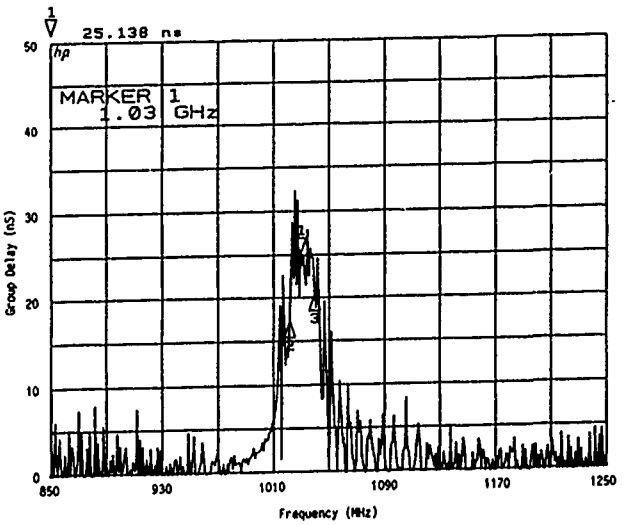
b)

Fig. 11. Experimental results for a two-pole SCF of silicon: a). Near-in response, 1.3 dB IL, 11MHz bandwidth. b). Wide band response showing the expected three resonances and negligible spurious responses.

The filter in Fig. 12 was designed for the 1030 MHz frequency range. The filter has an insertion loss of 1.7 dB and a bandwidth of 19 MHz (1.8%). The filter group delay characteristics are shown in Fig. 12b to have a spread of less than 15 ns across the 3 dB bandwidth.



a)



b)

Fig. 12. Wider bandwidth SCF filter on a silicon substrate: a). Near-in response, 1.7 dB IL, 19 MHz bandwidth. b). Group delay response, less than 15 ns in the passband.

Fabrication of the SCF on silicon substrates for large volume production requires a high degree of film thickness and material property uniformity throughout the wafer surface. Of most importance is the coupling coefficient and crystallographic orientation uniformity. Thickness non-uniformity affects the filter center frequency directly but the filter may be trimmed to frequency within the fabrication tolerances. However, film quality non-uniformities cannot be corrected for

the device has been processed and measured. A sample probing of a Si wafer is shown in Fig. 13. The dark solid areas of the wafer represent filters whose insertion loss was less than 1.95 dB (a specification threshold). Other filters on the wafer were operational but did not conform to specifications. Changes in processing are being studied that will give a higher yield of acceptable devices for production than that obtained from the research deposition system.

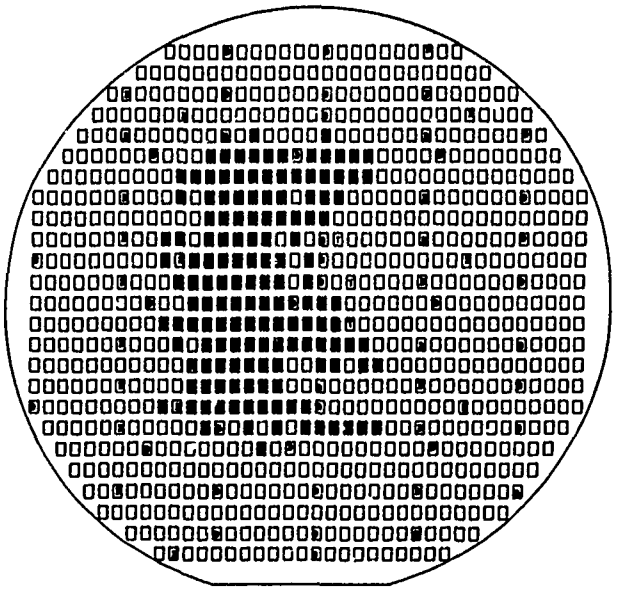


Fig. 13. Test-select pattern for a 3 inch diameter silicon wafer of two-pole SCF's. The solid black areas are for filters with insertion losses below 1.95 dB IL.

For complete integrated microwave subsystems at the chip level the greatest interest is in GaAs substrates. The filter results shown in Fig. 14 are the results of a preliminary study on the fabrication of the thin film SCF on GaAs substrates. The filter is of identical design as those fabricated on silicon and the device processing was essentially very similar. The filter insertion loss is approximately 1.6 dB and the bandwidth is 15 MHz (1.5%). In this case, and as occurs in some Si based filters, the parasitic plate modes are more apparent as seen in the increased ripple in the passband.

Although reported on previously [8,14], the Over-Moded SCF (OMSCF) is an important device whose features are predicted in the modeling [8] are even more impressive in experimental results. The OMSCF device was demonstrated on a number of substrates including Si, GaAs, sapphire, and spinel. Fig. 15 shows the response for an OMSCF fabricated on spinel as a two-pole high Q filter. The wide band scan of Fig. 15a is under-sampled and accordingly does not show the relatively uniform peak response of each resonance although the wider bandwidth rejection band is well illustrated.

Fig. 15b is a narrow band frequency scan of the device about one of the central resonances. The filter bandwidth is 180KHz, representing a very high Q resonance. The phase slope Q, relevent for oscillator work, is approximatily 9000 as measured in a 50 Ohm loaded system.

The OMSCF will prove to be an important device for narrow band receivers and for microwave oscillators requiring low phase noise [15].

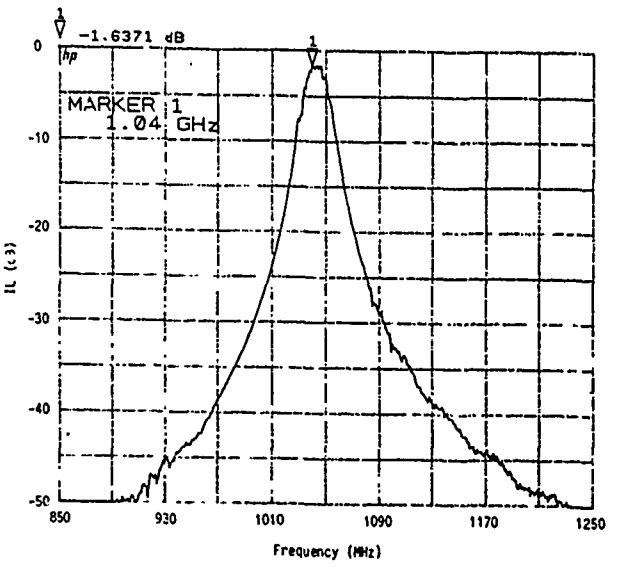


Fig. 14. Response of a preliminary two-pole SCF fabricated on GaAs substrate by the via process. Results are nearly as good as for devices fabricated on Si having the same device geometry.

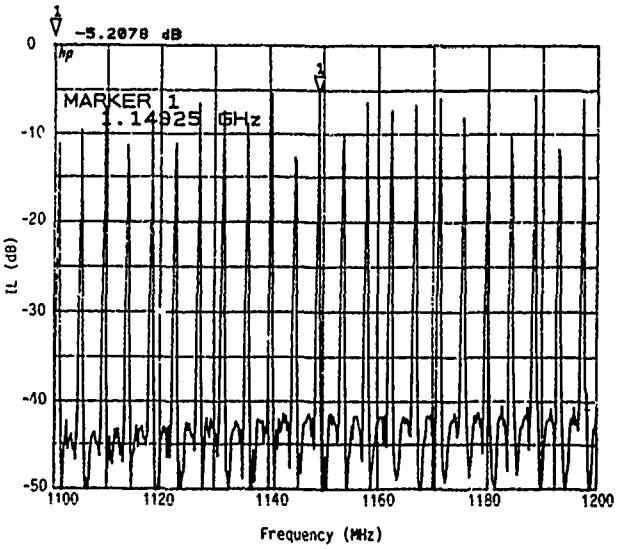


Fig. 15. Experimental results from a two-pole high loaded-Q over-moded SCF fabricated on a 1 mm thick spinel substrate. Resonant peaks are actually at approximately the same IL, the network analyzer was under sampling frequency.

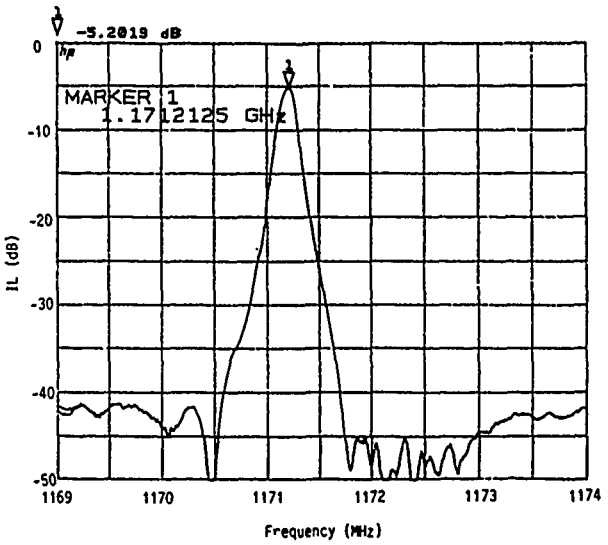


Fig. 16. Narrow band insertion loss plot of filter in Fig. 15. Bandwidth of 180 KHz, loaded Q of 9000.

SUMMARY

Low insertion loss 1GHz stacked crystal filters have been implemented with thin films of piezoelectric AlN. Two-pole filters having insertion losses of less than 2 dB have been fabricated on Si and GaAs substrates using microelectronic processing techniques. These devices can serve as high performance front-end filters for receivers or for other frequency control applications. The over-moded stacked crystal shows considerable promise as a two-port 50 ohm high Q feedback element for oscillators or for other narrow band applications.

Continued research and development on the SCF will involve advanced device modeling, design of unique sub-systems, and improved processing for volume production of piezoelectric film based devices and their co-integration with microwave integrated circuits.

REFERENCES

1. "Thin Film Resonators and Filters", K.M. Lakin, J.S. Wang, G.R. Kline, A.R. Landin, Y.Y. Chen, and J.D. Hunt, 1982 Ultrasonics Symposium Proceedings, San Diego, CA, Oct. 27-29, vol. 1, p. 466.

2. "A Novel Frequency Selective Device: The Stacked Crystal Filter", A. Ballato and T. Lukaszek, Proc. 27th Annual Frequency Control Symposium, June 1973, pp. 262-269.

3. "Stacked-crystal Filters", A. Ballato, T. Lukaszek, Proc. IEEE, Vol. 51, October 1973, pp. 1495-1496.

4. "The Stacked-crystal Filter", A. Ballato, Proc. IEEE Intl. Symposium on Circuits and Systems, April 1975, pp. 301-304.

5. "Thin Film Resonator Technology", K.M. Lakin, G.R. Kline, A.R. Landin, W.A. Burkland, K.T. McCarron, S.G. Burns, and S.D. Braymen, 41st Annual Frequency Control Symposium Proceedings, Philadelphia, PA, May 27-29, 1987, p. 371.

6. "Transmission-line Analogs for Stacked Piezoelectric Crystal Devices", A. Ballato, 26th Annual Frequency Control Symposium Proceedings, June 1972, pp. 86-91.

7. "Systematic Design of Stacked-crystal Filters by Microwave Network Methods", A. Ballato, H.L. Bertonia, and T. Tamir, IEEE Trans., Vol. MTT-22, January 1974, pp 14-25.

8. "Equivalent Circuit Modeling of Stacked Crystal Filters", K.M. Lakin, 35th Annual Frequency Control Symposium Proceedings, Philadelphia, PA, May 27-29, 1981, pp. 257-262.

9. "Fundamental Mode UHF/VHF Miniature Acoustic Resonators and Filters", Applied Physics Letters, T.W. Grudkowski, J.F. Black, T.M. Reeder, D.E. Cullen, and R.A. Wagner, Vol. 37 (11), 1980, pp. 993-995.

10. "Acoustic Bulk Wave Composite Resonators", Applied Physics Letters, K.M. Lakin and J.S. Wang, Vol. 39 (3), Feb. 1, 1981, pp. 125-128.

11. "An Air Gap Type Piezoelectric Composite Resonator", H. Satoh, Y. Ebata, H. Suzuki, and C. Narahara, Proceedings of the 39th Annual Symposium on Frequency Control, 1985, pp. 361-366.

12. "Low-Temperature Growth of the Piezoelectric AlN Film and its Optical and Acoustical Properties", T. Shiosaki, K. Harada and A. Kawabata, 1981 Ultrasonics Symposium Proceedings, pp. 506-509.

13. "Aluminum Nitride Thin Film and Composite Bulk Wave Resonators," K.M. Lakin, J.S. Wang and A.R. Landin, 36th Annual Symposium on Frequency Control Proceedings, Philadelphia, PA, June 2-4, 1982, p. 517.

14 "Thin Film Resonator (TFR) Low Insertion Loss Filters," G.R. Kline, K.M. Lakin and R.S. Ketcham, IEEE Ultrasonics Symposium Proceedings, Chicago IL, Oct. 2-5, 1988, Vol. 1, p. 339.

15. "Design and Performance of Voltage Controlled Oscillators Using TFR Stacked-Crystal Filters", P.H. Thompson, S.G. Burns, R.J. Weber, 43rd Annual Symposium on Frequency Control, Denver, CO., May31-June 2, 1989.

43rd Annual Symposium on Frequency Control - 1989

A NEW DESIGN FOR HIGH FREQUENCY BULK RESONATORS

A. Lepek and U. Maishar

T.F.L., Holon 58117, Israel

Abstract

TFL is in a process of developing a novel but simple method for production of high frequency bulk resonators with frequencies up to 200 MHz operating in the fundamental mode. So far frequencies up to 170 MHz were manufactured. The thin blanks may be used as resonators at these frequencies or be used as a starting point for higher frequencies achieved by using other technologies.
This method is developed both for our AT and SC crystals.
The principle of the present method is to start with a relatively thick blank bonded to a support, then to thin it using conventional polishing methods, to the desired frequency, the support being a part of the final crystal unit. In this way the problem of the handling and protecting thin blanks is soleved easily. The method allows for controlling the resonance frequency during the polishing phase.
The present paper describes a patent application and some measurements made on resonators manufactured using this method.

Project Targets

The project targets can be summarized as follows

- Develop a technique to obtain AT and SC cut resonators having thickness of 10 microns.
- Overcome handling problems with such thin resonators.
- No investment in new or exotic equipment.
- Frequency control to a high degree.
- Easy transfer to manufacruring.

At the present stage of development we believe that the adopted schemes will actualy lead to a solution of all these targets and will be suitble for efficient manufacturing.

Crystal Unit Configuration and Preparation

The first stage of the crystal unit preparation is shown in Fig. 1.
A thin blank (thickness of about 70 microns), having polished and parallel faces to a high degree, is bonded onto a thick blank (0.3 mm). the thin plate will become the resonator and the thick plate is the support.
The bonding has a ring shape so that the central portion (2 mm) of the resonator is

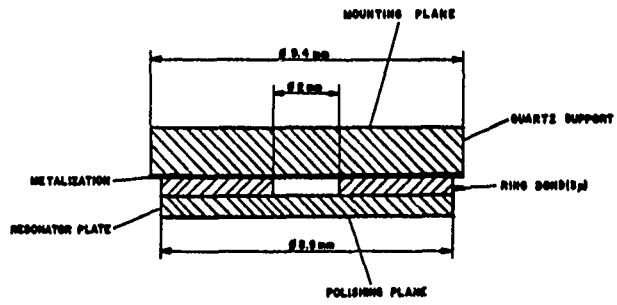


Fig.1 A schematic diagram of an initial crystal unit configuration

free. The bonding glue has a thickness of about 3 microns. This was found to be sufficient to support the thin resonator when it is bended under external pressure during subsequent polishing. The metalization on the support blank will be used as an air-gap electrode. It has the shape of a strip, 1 mm wide. The two blanks, i.e. the resonator and its support are made of same cut and are bonded with their crystalographic axes in parllel to minimize thermal stresses.
The support plate is mounted on a polishing button and the resonator's outer face is thinned by polishing to the desired thickness using standard polishing procedures.
Whenever desired, the polishing is interupted and the frequency is measured by inserting the assembly into an air-gap electrode device, which is a part of a S-parameter/network-analyzer (HP35677A + HP3577A). It is very convinient to trace the resonant frequency on a large frequency span by measuring the transmission delay. Fig. 2 shows such measurement on an AT cut resonator in process when reaching a thickness of about 24 microns (69.6 MHz).
The thinning rate is stable but its magnitude depends on many parameters. Stable rates of 0.3 to 1 microns per minutes are common. Once the rate is measured, it is easy to predict the thickness as a function of time to a high degree, and then there is no need for measurments until the frequency is close to its target.
It is interesting to show some of the resonators characteristics at this intermediate stage. Fig. 3 shows the transmission amplitude spectrum of the resonator of Fig. 2. Fig. 4 shows the transmission amplitude of the main mode and Fig. 5 the real and imaginary parts keeping the cursor at the frequency of maximum transmission. Note the narrow characteristic line shapes even at this early stage.

CH2690-6/89/0000-544 $1.00 C 1989 IEEE

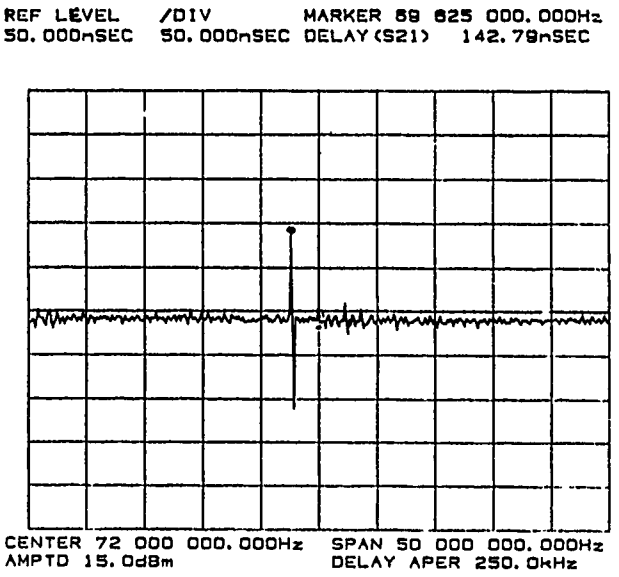


Fig.2 A transmission delay spectrum of a crystal mounted on a polishing button

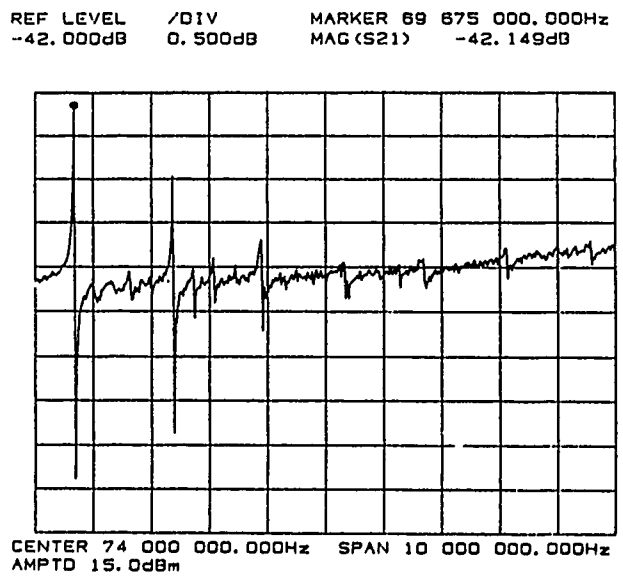


Fig.3 Amplitude transmission spectrum of the crystal of Fig.2

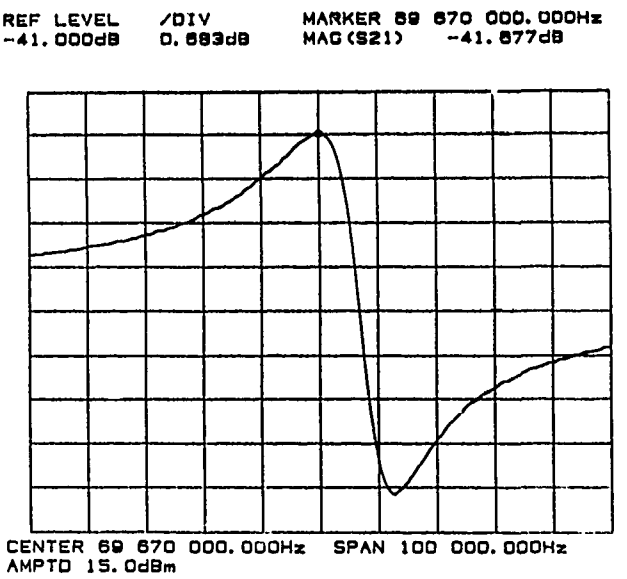


Fig.4 Transmission amplitude of the main mode of Fig.2

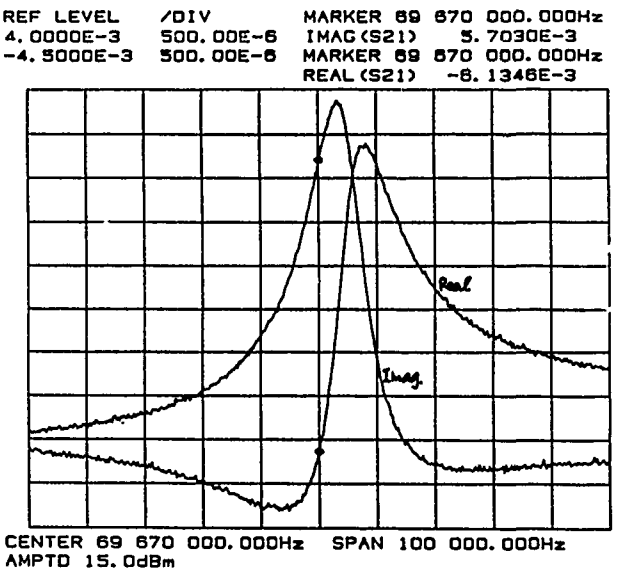
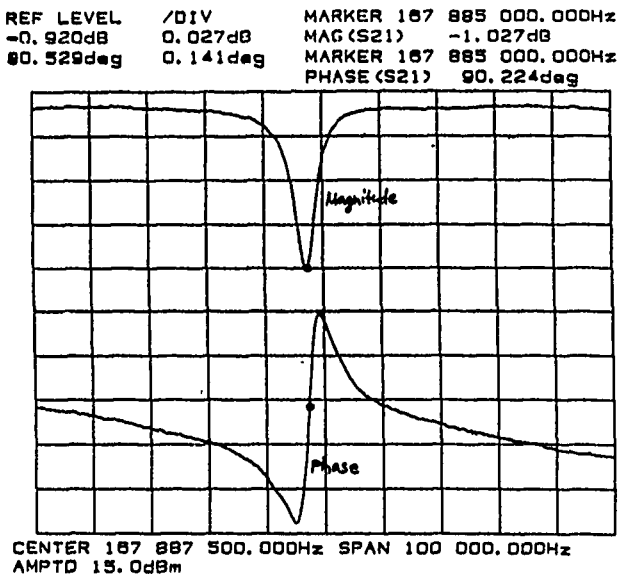


Fig. 5 Real and imaginary parts of the transmission of the main mode of Fig.2

During the present study we have used other configurations then the one of Fig. 1 as well. One such variant was prepared with the electrode coated on the inner face of the resonator rather than on the support. The vacuum deposition of the electrode was done either before the bonding or after it. In the latter case a through hole in the support blank was ulrasonicaly drilled and the deposition was enabled through this hole. This hole was used as well to make an electrical connection to the deposited electrode.



Fif. 6 Transmission amplitude and phase response of a 10 micrometers crystal still mounted on a polishing button

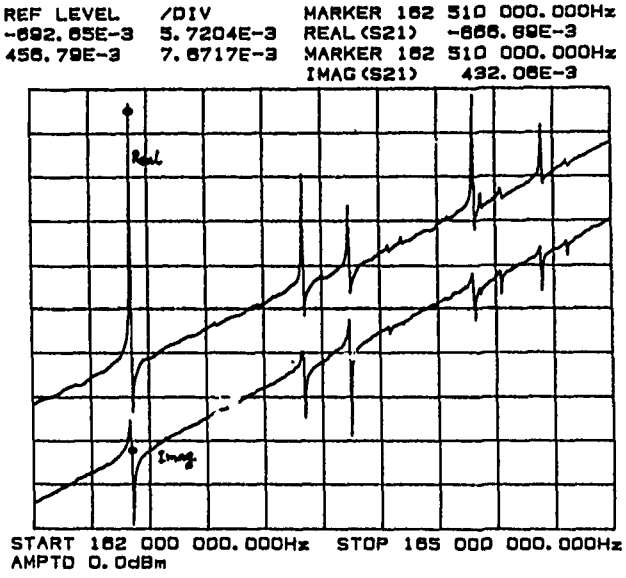


Fig.7 The spectrum of the crystal of Fig.6 after final calibration and encapsulation

Results

So far we were successful in achieving frequencies up to 170 MHz in the fundamental mode. Fig. 6 shows one such resonator still on the polishing button. This sample has a thickness of 10 microns (167.9 MHz) which was our primary development target. Note the characteristic line shapes of few kHz. This crystal was mounted in a TO-8 case and a counter electrode, made of silver, was deposited on the outer surface of the resonator. Although the deposition was manualy controlled in a direct plating facility, steps of 10 kHz were easily obtained. (We estimate that automation will lower the steps and enable precise final calibration).
We decided to stop editional deposition when a slight degradation in performsnce was detected. The case was then evacuated and cold welded and additional measurements were done. Fig.7 shows the spectrum of this crystal unit and Fig. 8 the stucture of its main mode. The drive power in these graphs was reduced to get the best line shapes. The best power level was obtained from Fig. 9 were the dependence of the line shape on the drive power was investigated.

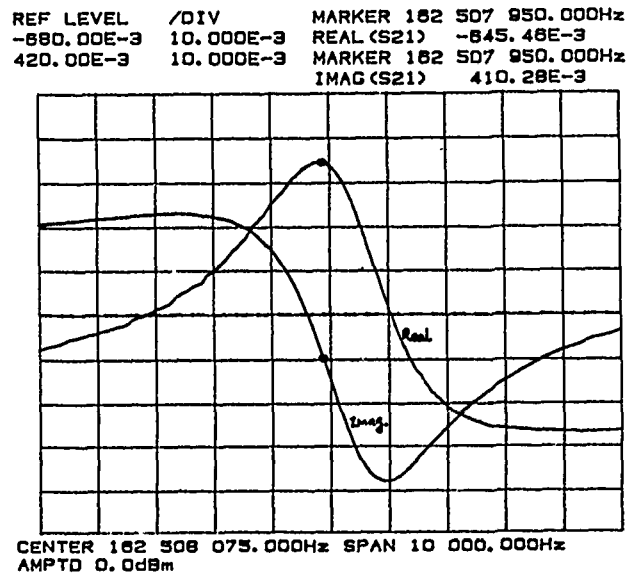


Fig.8 Detail of the main mode of Fig.7

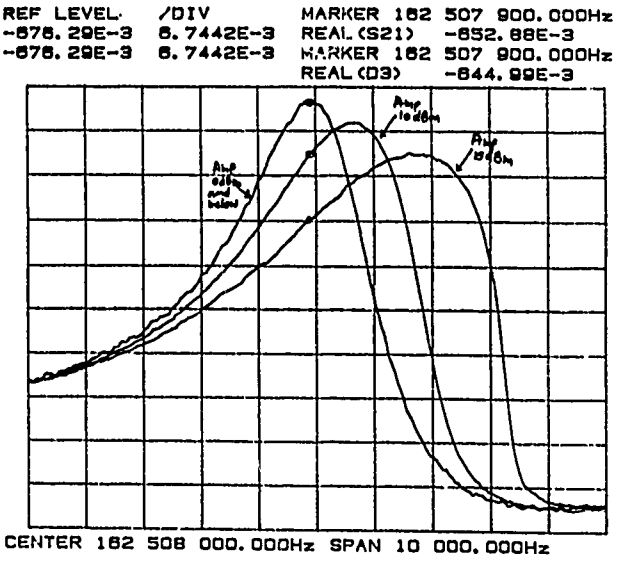


Fig. 9 The dependence of the line shapes on the power drive level for the crystal of Fig. 7

Summary

We have shown in this paper that conventional polishing methods are suitable for achieving high frequency, fundamental mode, bulk resonators, provided appropriate support is assembled as a part of the crystal unit.

Based upon this work, we believe that frequencies up to 200 MHz are feasible. Beyond this frequency other technologies may be employed, using the thin blanks obtained by this technique as a starting point rather then thinning relativly thick blanks.

The narrow line widths and ease of frequency control, almost in real time, make this technology very attractive.

THICKNESS-FIELD EXCITATION OF DOUBLY-ROTATED PIEZOELECTRIC CRYSTAL

Janpu Hou* and P. C. Y. Lee**
*Allied-Signal Inc.
**Princeton University

Abstract

Two-dimensional equations of motion of piezoelectric crystal plates, developed previously by Lee, Syngellakis and Hou, from the three-dimensional series of the thickness coordinate of the plate, are employed to study the vibrations driven by the thickness electric field applied to the major surfaces of doubly-rotated quartz strips.

Straight-crested vibrational solutions in closed form are obtained for the first eight coupled equations which constitute the two-dimensional, first order approximate equations of motion of piezoelectric crystal plates.

Dispersion curves, frequency spectrum for electrically shorted and traction free surface conditions, and mode shapes at different resonances and for various width-to-thickness ratios of the plate are computed and plotted for SC-cut quartz strips.

Introduction

Forced vibrations of piezoelectric crystal plates with finite width (or length) were studied by Tiersten and Mindlin [1], Mindlin [2],[3]. In these studies, two dimensional plate equations based on power series expansion were employed to investigate thickness-field excitation.

In the present paper, the two-dimensional plate theory based on trigonometric series expansion by Lee Syngellakis and Hou [4] is employed to study the thickness-field excitation of doubly-rotated quartz strip.

General solution in closed form for the eight coupled equations of motion is obtained as the sum of the solutions of homogeneous equations and the solution of the inhomogeneous equations. The electrically forced and traction-free boundary conditions are satisfied.

Two-Dimensional Plate Equations

Consider a doubly-rotated crystal plate, e.g. SC-cut quartz plate as shown in Fig. 1., referred to a rectangular cartesian coordinate system, with the width denoted by 2a in x_1 direction, thickness 2b in x_2 direction, and length 2c in x_3 direction, as shown in Fig. 2. The plate is called a strip when the length 2c is approaching infinite. The faces of the plate are coated with conducting electrodes and the edges , at $x_1 = \pm a$, are traction-free.

When the two-dimensional first-order equations for the vibration of piezoelectric crystal plates are applied to the strips, the mechanical displacements and electric potential are expanded in a two-term trigonometrical series:

$$u_1(x_1, x_2, t) = u_1^{(0)}(x_1, t) + u_1^{(1)}(x_1, t)\sin(\pi x_2/2b)$$

$$u_2(x_1, x_2, t) = u_2^{(0)}(x_1, t) + u_2^{(1)}(x_1, t)\sin(\pi x_2/2b)$$

$$u_3(x_1, x_2, t) = u_3^{(0)}(x_1, t) + u_3^{(1)}(x_1, t)\sin(\pi x_2/2b)$$

$$\phi(x_1, x_2, t) = \phi^{(0)}(x_1, t) + \phi^{(1)}(x_1, t)\sin(\pi x_2/2b) \quad (1)$$

Above six displacement components $u_i^{(n)}$ and two potential components $\phi^{(n)}$, $i = 1,2,3$ and $n = 0, 1$ accommodate the vibrations of flexure, extension, face-shear, and fundamental thickness-shear, thickness-twist, and thickness-stretch modes and their anharmonic overtones, and interactions with electric fields $E_1^{(0)} = -\phi^{(0)}{}_{,1}$ and $E_2^{(1)} = -(\pi/2b)\phi^{(1)}$.

The stress equations of motion and charge equations which govern the eight components in Equation (1) are:

$$T_{11,1}^{(1)} - (\pi/2b)T_{21}^{(0)} = \rho(1+2R)u_{1,tt}^{(1)},$$

$$T_{21,1}^{(1)} - (\pi/2b)T_{22}^{(0)} = \rho(1+2R)u_{2,tt}^{(1)},$$

$$T_{31,1}^{(1)} - (\pi/2b)T_{23}^{(0)} = \rho(1+2R)u_{3,tt}^{(1)},$$

$$D_{1,1}^{(1)} - (\pi/2b)D_2^{(1)} = (2/b^2)[e_{26}u_1^{(1)} + e_{22}u_2^{(1)} + e_{24}u_3^{(1)} - \varepsilon_{22}\phi^{(0)}],$$

$$T_{11,1}^{(0)} = 2\rho(1+R)u_{1,tt}^{(0)},$$

$$T_{21,1}^{(0)} = 2\rho(1+R)u_{2,tt}^{(0)}/\alpha_2^2,$$

$$T_{31,1}^{(0)} = 2\rho(1+R)u_{3,tt}^{(0)},$$

$$D_{1,1}^{(0)} = 0. \quad (2)$$

In equation (2), the mechanical effect of platings and the electrical effect of conducting platings on the faces of the plate, i.e. $T_{2j} = +2\rho' b' u_{j,tt}$ and $\phi = \pm\phi_0 e^{i\omega t}$ at $x_2 = \pm b$, have been taken into account. The R is defined as the ratio of the mass of conducting platings per unit area to the mass of the plate per unit area.

The stresses and electric displacements are related to the mechanical displacements and electric potential by the constitutive equations:

$$T_{11}^{(1)} = c_{11}u_{1,1}^{(1)} + c_{16}u_{2,1}^{(1)} + c_{15}u_{3,1}^{(1)} + e_{11}\phi_{,1}^{(1)},$$

$$T_{12}^{(1)} = \alpha_2[c_{61}u_{1,1}^{(1)} + \alpha_2 c_{66}u_{2,1}^{(1)} + c_{65}u_{3,1}^{(1)}] + e_{16}\phi_{,1}^{(1)},$$

$$T_{13}^{(1)} = c_{51}u_{1,1}^{(1)} + c_{56}u_{2,1}^{(1)} + c_{55}u_{3,1}^{(1)} + e_{15}\phi_{,1}^{(1)},$$

$$D^{(1)} = e_{11}u_{1,1}^{(1)} + e_{16}u_{2,1}^{(1)} + e_{15}u_{3,1}^{(1)} - \varepsilon_{11}\phi_{,1}^{(1)},$$

$$T_{21}^{(1)} = (4\alpha_1/\pi)[c_{61}u_{1,1}^{(0)} + c_{66}u_{2,1}^{(0)} + c_{65}u_{3,1}^{(0)}] + (4/\pi)e_{16}\phi_{,1}^{(0)} + (\pi/2b)[c_{66}u_1^{(1)} + c_{62}u_2^{(1)} + c_{64}u_3^{(1)}] + (\pi/2b)e_{26}\phi^{(1)}$$

$$T_{22}^{(1)} = (4\alpha_1/\pi)[c_{21}u_{1,1}^{(0)} + c_{26}u_{2,1}^{(0)} + c_{25}u_{3,1}^{(0)}] + (4/\pi)e_{12}\phi_{,1}^{(0)} + (\pi/2b)[c_{26}u_1^{(1)} + c_{22}u_2^{(1)} + c_{24}u_3^{(1)}] + (\pi/2b)e_{22}\phi^{(1)}$$

CH2690-6/89/0000-548 $1.00 C 1989 IEEE

$$T_{23}^{(1)} = (4\alpha_1/\pi)\,[c_{41}u_{1,1}^{(0)} + c_{46}u_{2,1}^{(0)} + c_{45}u_{3,1}^{(0)}] + (4/\pi)e_{14}\phi_{,1}^{(0)} + (\pi/2b)\,[c_{46}u_1^{(1)} + c_{42}u_2^{(1)} + c_{44}u_3^{(1)}] + (\pi/2b)e_{24}\phi^{(1)}$$

$$D_2^{(1)} = (4\alpha_1/\pi)\,[e_{21}u_{1,1}^{(0)} + e_{26}u_{2,1}^{(0)} + e_{25}u_{3,1}^{(0)}] - (4/\pi)\varepsilon_{21}\phi_{,1}^{(0)} + (\pi/2b)\,[e_{26}u_1^{(1)} + e_{22}u_2^{(1)} + e_{24}u_3^{(1)}] - (\pi/2b)\varepsilon_{22}\phi^{(1)}$$

$$T_{11}^{(0)} = 2[c_{11}u_{1,1}^{(0)} + c_{16}u_{2,1}^{(0)} + c_{15}u_{3,1}^{(0)} + (\alpha_1/b)(c_{16}u_1^{(1)} + c_{12}u_2^{(1)} + c_{14}u_3^{(1)})] + 2e_{11}\phi_{,1}^{(0)} + (2/b)e_{21}\phi^{(1)}$$

$$T_{12}^{(0)} = 2[c_{61}u_{1,1}^{(0)} + c_{66}u_{2,1}^{(0)} + c_{65}u_{3,1}^{(0)} + (\alpha_1/b)(c_{66}u_1^{(1)} + c_{62}u_2^{(1)} + c_{64}u_3^{(1)})] + 2e_{16}\phi_{,1}^{(0)} + (2/b)e_{26}\phi^{(1)}$$

$$T_{13}^{(0)} = 2[c_{51}u_{1,1}^{(0)} + c_{56}u_{2,1}^{(0)} + c_{55}u_{3,1}^{(0)} + (\alpha_1/b)(c_{56}u_1^{(1)} + c_{52}u_2^{(1)} + c_{54}u_3^{(1)})] + 2e_{15}\phi_{,1}^{(0)} + (2/b)e_{25}\phi^{(1)}$$

$$D_1^{(0)} = 2[e_{11}u_{1,1}^{(0)} + e_{16}u_{2,1}^{(0)} + e_{15}u_{3,1}^{(0)} + (\alpha_1/b)(e_{16}u_1^{(1)} + e_{12}u_2^{(1)} + e_{14}u_3^{(1)})] - 2\varepsilon_{11}\phi_{,1}^{(0)} - (1/b)\varepsilon_{21}\phi^{(1)} \qquad (3)$$

Conditions of traction-free and charge-free without electrodes for edges at $x_1 = \pm$ a require,

$$T_{11}^{(0)} = T_{12}^{(0)} = T_{13}^{(0)} = T_{11}^{(1)} = T_{12}^{(1)} = T_{13}^{(1)} = 0$$

and

$$D_1^{(0)} = D_1^{(1)} = 0 \qquad (4)$$

Integration of the last equation of (2) with respect to x_1, leads to the conclusion that $D_1^{(0)}$ must be a function of time only, i.e.,

$$D_1^{(0)} = -\varepsilon_{11}A(t) \qquad (5)$$

By considering the edge conditions of $D_1^{(0)} = 0$ on $x_1 = \pm a$, we note that $A(t) = 0$ and then $D_1^{(0)} = 0$ throughout the plate. By the use of constitutive relation on $D_1^{(0)}$ in (3) and solving for $\phi_{,1}^{(0)}$ one has

$$\phi_{,1}^{(0)} = 1/\varepsilon_{11}[e_{11}u_{1,1}^{(0)} + e_{16}u_{2,1}^{(0)} + e_{15}u_{3,1}^{(0)} + (\alpha_1/b)(e_{16}u_1^{(1)} + e_{12}u_2^{(1)} + e_{14}u_3^{(1)}) - (1/b)\varepsilon_{21}\phi^{(1)}] \qquad (6)$$

Substitution of (6) into the rest equation in (3) then into (2) yields the governing equation of strip resonator on the seven functions $u_i^{(0)}$, $u_i^{(1)}$, and $\phi^{(1)}$ (i = 1, 2, 3), as follows:

$$c_{11}u_1^{(1)},_{11} + c_{15}u_3^{(1)},_{11} + c_{16}u_2^{(1)},_{11} + e_{11}\phi^{(1)},_{11} - (2\alpha_1/b)[c'_{16}u_1^{(0)},_1 + c'_{65}u_3^{(0)},_1 + c'_{66}u_2^{(0)},_1] - (\pi/2b)^2[c'_{62}u_2^{(1)} + c'_{64}u_3^{(1)} + c'_{66}u_1^{(1)} - e'_{26}\phi^{(1)}] = \rho(1 + 2R)\,u_1,_{tt}^{(1)},$$

$$\alpha_2[c_{61}u_1^{(1)},_{11} + \alpha_2 c_{66}u_2^{(1)},_{11} + c_{65}u_3^{(1)},_{11}] + e_{16}\phi^{(1)},_{11} - (2\alpha_1/b)[c'_{21}u_1^{(0)},_1 + c'_{26}u_2^{(0)},_1 + c'_{25}u_3^{(0)},_1] - (\pi/2b)^2[c'_{22}u_2^{(1)} + c'_{24}u_3^{(1)} + c'_{26}u_1^{(1)} - c'_{22}\phi^{(1)}] = \rho(1 + 2R)\,u_2,_{tt}^{(1)}$$

$$c_{51}u_1^{(1)},_{11} + c_{56}u_2^{(1)},_{11} + c_{55}u_3^{(1)},_{11} + e_{15}\phi^{(1)},_{11} - (2\alpha_1/b)[c'_{41}u_1^{(0)},_1 + c'_{45}u_3^{(0)},_1 + c'_{46}u_2^{(0)},_1] - (\pi/2b)^2[c'_{42}u_2^{(1)} + c'_{44}u_3^{(1)} + c'_{46}u_1^{(1)} - e'_{24}\phi^{(1)}] = \rho(1 + 2R)\,u_3,_{tt}^{(1)},$$

$$e_{11}u_1^{(1)},_{11} + e_{15}u_3^{(1)},_{11} + e_{16}u_2^{(1)},_{11} - \varepsilon_{11}\phi^{(1)},_{11} - (2/b)[e'_{21}u_1^{(0)},_1 + e'_{25}u_3^{(0)},_1 + e'_{26}u_2^{(0)},_1] - (\pi/2b)^2[e'_{22}u_2^{(1)} + c'_{24}u_3^{(1)} + e'_{26}u_1^{(1)} - \varepsilon'_{22}\phi^{(1)}] - (2/b^2)[e_{26}u_1^{(1)} + e_{22}u_2^{(1)} + e_{24}u_3^{(1)}] = -(2/b^2)\,\varepsilon_{22}\phi_0,$$

$$c'_{11}u_1^{(0)},_{11} + c'_{15}u_3^{(0)},_{11} + c'_{16}u_2^{(0)},_{11} + (\alpha_1/b)[c'_{12}u_2^{(1)},_1 + c'_{14}u_3^{(1)},_1 + c'_{16}u_1^{(1)},_1] + (1/b)e'_{21}\phi^{(1)},_1 = \rho(1 + R)\,u_1,_{tt}^{(0)}$$

$$c'_{61}u_1^{(0)},_{11} + c'_{65}u_3^{(0)},_{11} + c'_{66}u_2^{(0)},_{11} + (\alpha_1/b)[c'_{62}u_2^{(1)},_1 + c'_{64}u_3^{(1)},_1 + c'_{66}u_1^{(1)},_1] + (1/b)e'_{26}\phi^{(1)},_1 = \rho(1 + R)\,u_2,_{tt}^{(0)}/\alpha_2^2,$$

$$c'_{51}u_1^{(0)},_{11} + c'_{55}u_3^{(0)},_{11} + c'_{56}u_2^{(0)},_{11} + (\alpha_1/b)[c'_{52}u_2^{(1)},_1 + c'_{54}u_3^{(1)},_1 + c'_{56}u_1^{(1)},_1] + (1/b)e'_{25}\phi^{(1)},_1 = \rho(1 + R)\,u_3,_{tt}^{(0)}, \qquad (7)$$

where

$$c'_{pq} = c_{pq} + (e_{1p}e_{1q}/\varepsilon_{11}) \qquad e'_{iq} = e_{iq} - (\varepsilon_{1i}e_{11}/\varepsilon_{11})$$

$$\varepsilon'_{ij} = \varepsilon_{ij} - (\varepsilon_{1i}\varepsilon_{1j}/\varepsilon_{11}) \qquad \alpha_1^2 = \pi^2/8.$$

In Eq. (7), c'_{pq}, e'_{iq}, and ε'_{ij} are modified material constants, and α_1, α_2 are correction factors introduced in order to match the dispersion relations obtained from the two-dimensional equations more closely to those obtained from the three-dimensional equations of piezoelectricity.

The α_2 can be calculated from the velocity of surface wave.[5] It depends on the orientation of the plate as well as the direction of the wave normal of the surface wave. For example, a surface wave propagating along the x_1 - in an SC-cut quartz plate should have a speed of 3250 m/sec. Then $\alpha_2 = 0.8714$.

Forced Vibrations of Piezoelectric Crystal Strips by Thickness-Field Excitation

We seek the solution of forced vibrations driven by a steady alternating voltage applied to the electrodes coated on the surfaces of the plate at $x_2 = \pm$ b.

The edge is free of traction and charge, at $x_1 = \pm$ a:

$$T_{11}^{(0)} = T_{12}^{(0)} = T_{13}^{(0)} = D_1^{(0)} = 0$$
$$T_{11}^{(1)} = T_{12}^{(1)} = T_{13}^{(1)} = D_1^{(1)} = 0 \qquad (8)$$

The general solution will be obtained as the sum of the solution of homogeneous equations by letting $\phi = 0$ and the particular solution of the inhomogeneous equations by letting $\phi = \phi_0 e^{i\omega t}$. Then the boundary conditions are imposed on the general solution.

Solutions of Homogeneous Equations and Dispersion Relations

We let $\phi = 0$ and assume the steady vibrational form of solution:

$$u_1^{(1)} = b\,A_1 \cos\xi x_1 e^{i\omega t}$$

$$u_2^{(1)} = b\,A_2 \cos\xi x_1 e^{i\omega t}$$

$$u_3^{(1)} = b\,A_3 \cos\xi x_1 e^{i\omega t}$$

$$\phi^{(1)} = (\hat{c}_{66}/\varepsilon_{22})^{1/2}\, b\, A_4 \cos \xi\, x_1\, e^{i\omega t}$$

$$u_1^{(0)} = b\, A_5 \sin \xi\, x_1\, e^{i\omega t}$$

$$u_2^{(0)} = b\, A_6 \sin \xi\, x_1\, e^{i\omega t}$$

$$u_3^{(0)} = b\, A_7 \sin \xi\, x_1\, e^{i\omega t} \qquad (9)$$

where $\hat{c}_{66} = c_{66} + e_{26}^2 / \varepsilon_{22}$

Substitution of (9) into (7) leads to seven simultaneous equations on seven amplitudes A_n (M.1)

where

$\tilde{c}_{pq} = c'_{pq} / \hat{c}_{66}$ $\qquad$ $\bar{c}_{pq} = c_{pq} / \hat{c}_{66}$

$\tilde{e}_{iq} = e'_{iq} / (\hat{c}_{66}\varepsilon_{22})^{1/2}$ $\qquad$ $\bar{e}_{iq} = e'_{iq} / (\hat{c}_{66}\varepsilon_{22})^{1/2}$

$\tilde{\varepsilon}_{ij} = \varepsilon'_{ij} / \varepsilon_{11}$ $\qquad$ $\bar{\varepsilon}_{ij} = \varepsilon_{ij} / \varepsilon_{22}$

$\Omega = \omega / [(\pi/2b)/\sqrt{(c_{66}/\rho)}]$ $\qquad$ $z = \xi / (\pi/2b)$

The dispersion relation can be computed by setting the determinant of the coefficient matrix to zero. For a given value of Ω, it yields seven pairs of roots $\pm z_n$, $n = 1, 2, 3, \ldots 7$. And for each root z_n there are seven amplitudes, A_{in}, $i = 1, 2, 3, \ldots 7$, for the solution (9). Then the ratio among these amplitudes are defined by

$$\alpha_{in} = A_{in} / A_{1n}, \; i = 1, 2, \ldots 7 \qquad (10)$$

and they can be solved from (8). We note that $\alpha_{1n} = 1$.

The seven branches of the dispersion relation are computed for an SC-cut quartz plate and illustrated in Fig. 3. The predominance of amplitude of each branches is indicated by the letter symbols:

P = Electrical Potential

E = Extension

F = Flexure

FS = Face-Shear

TT = Thickness-Twist

TSh = Thickness-Shear

TSt = Thickness-Stretch

Particular Solution and Cut-Off Frequencies

For the particular solution of the inhomogeneous equations, we assume $\phi = \phi_0\, e^{i\omega t}$ and

$$u_1^{(1)} = B_1\, e^{i\omega t}$$

$$u_2^{(1)} = B_2\, e^{i\omega t}$$

$$u_3^{(1)} = B_3\, e^{i\omega t}$$

$$\phi^{(1)} = (\hat{c}_{66}/\varepsilon_{22})^{1/2}\, B_4\, e^{i\omega t}$$

$$u_1^{(0)} = u_2^{(0)} = u_3^{(0)} = 0 \qquad (11)$$

Substitution of (11) into (7) with $\phi_0 \neq 0$ yields matrix (M.2). For frequencies not equal to the thickness frequencies of infinite plate, i.e., the frequencies which will not make the determinant of coefficient matrix of (M.2) equal to zero, we can solve for amplitudes B_i, (i = 1, 2, 3, 4) from (M.2) in terms of ϕ_0.

General Solution and Frequency Spectrum

The appropriate form of the solution of the strip vibrator is the sum of the solution of homogeneous equations and the particular solution of the inhomogeneous equations.

$$u_1^{(1)} = \{ \textstyle\sum_n^7 = 1\, \alpha_{1n} A_{1n} \cos [2a/(\pi b z_n)] + B_1 \}\, b\, e^{i\omega t},$$

$$u_2^{(1)} = \{ \textstyle\sum_n^7 = 1\, \alpha_{2n} A_{1n} \cos [2a/(\pi b z_n)] + B_2 \}\, b\, e^{i\omega t},$$

$$u_3^{(1)} = \{ \textstyle\sum_n^7 = 1\, \alpha_{3n} A_{1n} \cos [2a/(\pi b z_n)] + B_3 \}\, b\, e^{i\omega t},$$

$$\phi^{(1)} = \{ \textstyle\sum_n^7 = 1\, (\hat{c}_{66}/\varepsilon_{22})^{1/2} \alpha_{4n} A_{1n} \cos [2a/(\pi b z_n)] + (\hat{c}_{66}/\varepsilon_{22})^{1/2} B_4 \}\, b\, e^{i\omega t},$$

$$u_1^{(0)} = \textstyle\sum_n^7 = 1\, \alpha_{5n} A_{1n} \sin [2a/(\pi b z_n)]\; b\, e^{i\omega t},$$

$$u_2^{(0)} = \textstyle\sum_n^7 = 1\, \alpha_{6n} A_{1n} \sin [2a/(\pi b z_n)]\; b\, e^{i\omega t},$$

$$u_3^{(0)} = \textstyle\sum_n^7 = 1\, \alpha_{7n} A_{1n} \sin [2a/(\pi b z_n)]\; b\, e^{i\omega t}, \qquad (12)$$

where α_{iq}, i = 1, 2, 3, ...7, q = 1, 2, 3, ... 7 are defined in (10). Upon substituting the displacement and potential (12) into the constitutive relation (3) and then into the edge conditions (4) we have, in matrix form (M.3).

where

$$M_{1n} = [\, \bar{c}_{11}\alpha_{1n} + \bar{c}_{16}\alpha_{2n} + \bar{c}_{15}\alpha_{3n} + \bar{e}_{11}\alpha_{4n} \,]\, z_n \sin[2a/(\pi b z_n)],$$

$$M_{2n} = [\, \bar{c}_{61}\alpha_{1n} + \alpha_2 \bar{c}_{66}\alpha_{2n} + \bar{c}_{65}\alpha_{3n} + \bar{e}_{16}\alpha_{4n} \,]\, z_n \sin[2a/(\pi b z_n)],$$

$$M_{3n} = [\, \bar{c}_{51}\alpha_{1n} + \bar{c}_{56}\alpha_{2n} + \bar{c}_{55}\alpha_{3n} + \bar{e}_{15}\alpha_{4n} \,]\, z_n \sin[2a/(\pi b z_n)],$$

$$M_{4n} = [\, \bar{e}_{11}\alpha_{1n} + \bar{e}_{16}\alpha_{2n} + \bar{e}_{15}\alpha_{3n} + \bar{\varepsilon}_{11}\alpha_{4n} \,]\, z_n \sin[2a/(\pi b z_n)],$$

$$M_{5n} = \{(2\alpha_1/\pi)\, [\bar{c}_{11}\alpha_{1n} + \bar{c}_{12}\alpha_{2n} + \bar{c}_{14}\alpha_{3n} + \bar{e}_{21}\alpha_{4n}\,] + [\bar{c}_{11}\alpha_{5n} + \bar{c}_{16}\alpha_{6n} + \bar{c}_{15}\alpha_{7n} + \bar{e}_{11}\alpha_{8n}\,]\, z_n\} \cos[2a/(\pi b z_n)],$$

$$M_{6n} = \{(2\alpha_1/\pi)\, [\bar{c}_{66}\alpha_{1n} + \bar{c}_{62}\alpha_{2n} + \bar{c}_{64}\alpha_{3n} + \bar{e}_{26}\alpha_{4n}\,] + [\bar{c}_{61}\alpha_{5n} + \bar{c}_{66}\alpha_{6n} + \bar{c}_{65}\alpha_{7n} + \bar{e}_{16}\alpha_{8n}\,]\, z_n\} \cos[2a/(\pi b z_n)],$$

$$M_{7n} = \{(2\alpha_1/\pi)\, [\bar{c}_{56}\alpha_{1n} + \bar{c}_{52}\alpha_{2n} + \bar{c}_{54}\alpha_{3n} + \bar{e}_{25}\alpha_{4n}\,] + [\bar{c}_{51}\alpha_{5n} + \bar{c}_{56}\alpha_{6n} + \bar{c}_{55}\alpha_{7n} + \bar{e}_{15}\alpha_{8n}\,]\, z_n\} \cos[2a/(\pi b z_n)].$$

For a given value of frequency Ω, the dimensionless wave number z_n, and amplitude ratios aij can be computed from (M.1) and B_m

from (M.2). If the width-to-thickness ratio, a/b, is also given, then each member of the coefficient matrix of (M.3) can be calculated, and amplitude A_{1n}, n = 1, 2, 3, ...7, can be solved in terms of f_o for forced vibrations. For A_{1n} not all equal zero, the determinant of the coefficient matrix (M.3) must be zero. Therefore,

$$\det | M_{nm} | = F(\Omega, \theta, \phi, R, a/b) = 0 \qquad (13)$$

The above is the equation for resonance frequencies. Mass loading effect can be studied by change the value of R in the equations. The roots Ω of (13) is a function of a/b are computed for SC-cut of quartz with R = 0.03 as shown in Fig. 4. Also numerical values of resonance frequencies for a/b = 17, 20, 23 and for various values of R are listed in Table 1.

Vibrational Mode Shapes and Frequency Separation Factors

Once the resonance frequencies are determined, it is of interest to identify the mode shape corresponding to each resonance and to examine the effect of a/b ratio on the change of mode shapes as well as on the frequencies. For a particular value of resonance from (M.1), we can calculate the amplitude ratios among A_{1n}. By further substituting these amplitude ratios back into (12) then (13), mechanical displacement and electric potentials are determined as functions of x_1 and t within a constant of multiplication. Mode shapes are calculated at different frequencies and they are shown in Figs. 5, 6, 7, 8, 9 for a/b = 20.

From experimental measurements and design experiences, it is observed that difference of the resonance frequencies of the same mode corresponding to open and shorted face conditions is proportional to the piezoelectric strength of that particular mode. Therefore, we may define the frequency separation factor of thickness-shear mode as

$$C_{TSh\text{-}i} = (\Omega^{o}_{TSh\text{-}i} - \Omega^{s}_{TSh\text{-}i}) / \Omega^{s}_{TSh\text{-}i}$$

where $\Omega^{o}_{TSh\text{-}i}$ is the frequency of ith anharmonic thickness-shear mode of the plate with electrically open surfaces and $\Omega^{s}_{TSh\text{-}i}$ is the frequency of the same plate with electrically shorted surfaces.

Similarly, we can define the frequency separation factors of flexural, face-shear, extension thickness-twist and thickness-stretch modes by

$$C_{M\text{-}i} = (\Omega^{o}_{M\text{-}i} - \Omega^{s}_{M\text{-}i}) / \Omega^{s}_{M\text{-}i},$$

where M may present E, F, FS, TT or TSt.

By using the present two-dimensional plate equations, both open and shorted electric surface conditions can be easily accommodated in computing the dispersion relations. In order to calculate Ω^{o} resonant frequency for plate with electrically open face condition, in (2), we have

$$D_{1,1}^{(1)} - (\pi / 2b) D_2^{(1)} = 0.$$

The frequency separation factors can be easily obtained.

For ease of comparison, we define the normalized frequency separation factors to be:

$$C_{M\text{-}i} = 20 \log (C_{M\text{-}i} / C_{TSh\text{-}1})$$

Table 2. gives the resonant frequencies of an SC-cut quartz plates with electrically open surfaces. Figures 10, and 11 show the mode spectrum of an SC-cut quartz strip with different width/thickness ratios of 17 and 20, respectively.

Discussions

The slope of a particular point on the dispersion curves is related to the group velocity or the propagation of the energy. These information can be used to do simplified energy-trapped resonator design and analysis. [6] The coupling between different modes can be identified from the curves as well and can be used for delay line design and analysis. [7]

Frequency spectrum will provide some useful information. We can examine the influence of width-to-thickness ratio on the coupling between modes. For example:

Weakly Coupled: Face-Shear(FS-15) and
Thickness-Shear(TSh-1)

Moderately Coupled: Extension(E-11) and
Thickness-Shear(TSh-1)

Strongly Coupled: Flexural(F-26) and
Thickness-Shear(TSh-1)

We may select the width-to-thickness ratio to avoid the unwanted modes For example, for the fundamental thickness-shear mode (TSh-1), a/b between 19.0 to19.4 will be a good choice.

Vibrational Mode Shapes will give further information. We can examine the influence of width-to-thickness ratio on the predominance of the mode. We can also study the influence of width-to-thickness ratio on the coupling between the modes. For example, flexural mode is always coupled with thickness-shear mode and face-shear mode is always coupled with thickness-twist mode.

Frequency separation factors have their own usefulness. We can use them to examine the influence of width-to-thickness ratio on the strength of the piezoelectric coupling of each mode. For example, for width-to-thickness = 17 the fundamental thickness TSh-1 is the predominate mode and the strength reduced for anharmonic overtones TSh-3 and TSh-5. It is easier to design the energy-trapped resonator under this condition. For width-to-thickness = 20, the anharmonic overtones of thickness-shear modes TSh-3 and TSh-5 have equal strength as the fundamental one TSh-1. It will be more difficult to obtain the energy-trapped resonator under this condition.

Summary

(1) By solving the plate equations of motion, we obtain the dispersion relations of SC-cut quartz plates.

(2) By taking into account the edge conditions, we obtain the frequency spectra, frequency separation factors and vibrational mode shapes for SC-cut quart strips with finite width.

(3) The coupling between different modes can be identified from the frequency spectra and vibrational mode shapes.

(4) The piezoelectric strength of each mode can be examinated from the frequency separation factors which is one of the advantage of using the present two-dimensional plate equations.

References

[1] H. F. Tiersten and R. D. Mindlin, " Forced vibrations of piezoelectric crystal plates", Quat. Appl. Math., Vol. 20, pp. 107 - 119, 1962.

[2] R. D. Mindlin, " High frequency vibration of piezoelectric crystal plates", Int. J. Solids Structures, vol. 8, pp. 895 - 906, 1972.

[3] R. D. Mindlin, " Frequencies of piezoelectrically forced vibrations of electroded, doubly rotated, quartz plates", Int. J. Solid Structures, Vol. 20, pp. 141 - 157, 1984.

[4] P. C. Y. Lee, S. Syngellakis and J. P. Hou, " A two-dimensional theory for high frequency vibrations of piezoelectric crystal plates with or without electrodes, J. Appl. Phys, 61(4), 15, 1987.

[5] P. C. Y. Lee and J. P. Hou, "Vibrations of doubly-rotated piezoelectric crystal strip with a pair of electrode-plated, traction-free edges", Proc. 39th Ann. Freq. Control Symp., pp. 453 - 461, 1985.

[6] T.R. Meeker, "Plate constants and dispersion relations for width-length effects in rotated Y-cut quartz plates" Proc. 29th Annu. Freq. Control Symp., pp. 54-64, 1975"

[7] J. Hou, "Approximation method for plate modes in surface acoustic wave devices" Proc. 41st Annu. Freq. Control Symp., pp. 365-370, 1987

$$\begin{bmatrix}
\bar{c}_{11}z^2+\tilde{c}_{66}-\Omega^2(1+2R) & \bar{c}_{16}z^2+\tilde{c}_{62} & \bar{c}_{15}z^2+\tilde{c}_{64} & \bar{e}_{11}z^2+\tilde{e}_{26} & \frac{4\alpha_1}{\pi}\tilde{c}_{61}z & \frac{4\alpha_1}{\pi}\tilde{c}_{66}z & \frac{4\alpha_1}{\pi}\tilde{c}_{65}z \\
\alpha_2\bar{c}_{61}z^2+\tilde{c}_{26} & \alpha_2^2\bar{c}_{66}z^2+\tilde{c}_{22}-\Omega^2(1+2R) & \alpha_2\bar{c}_{65}z^2+\tilde{c}_{24} & \alpha_2\bar{e}_{16}z^2+\tilde{e}_{22} & \frac{4\alpha_1}{\pi}\tilde{c}_{21}z & \frac{4\alpha_1}{\pi}\tilde{c}_{26}z & \frac{4\alpha_1}{\pi}\tilde{c}_{25}z \\
\bar{c}_{51}z^2+\tilde{c}_{46} & \bar{c}_{56}z^2+\tilde{c}_{42} & \bar{c}_{55}z^2+\tilde{c}_{44}-\Omega^2(1+2R) & \bar{e}_{15}z^2+\tilde{e}_{24} & \frac{4\alpha_1}{\pi}\tilde{c}_{41}z & \frac{4\alpha_1}{\pi}\tilde{c}_{46}z & \frac{4\alpha_1}{\pi}\tilde{c}_{45}z \\
\bar{e}_{11}z^2+\tilde{e}_{26}-\frac{8}{\pi^2}\bar{e}_{26} & \bar{e}_{16}z^2+\tilde{e}_{22}-\frac{8}{\pi^2}\bar{e}_{22} & \bar{e}_{15}z^2+\tilde{e}_{24}-\frac{8}{\pi^2}\bar{e}_{24} & -\bar{\varepsilon}_{11}z^2-\bar{\varepsilon}_{22} & \frac{4\alpha_1}{\pi}\tilde{e}_{21}z & \frac{4\alpha_1}{\pi}\tilde{e}_{26}z & \frac{4\alpha_1}{\pi}\tilde{e}_{25}z \\
\frac{2\alpha_1}{\pi}\tilde{c}_{16}z & \frac{2\alpha_1}{\pi}\tilde{c}_{12}z & \frac{2\alpha_1}{\pi}\tilde{c}_{14}z & \frac{2\alpha_1}{\pi}\tilde{e}_{21}z & \tilde{c}_{11}z^2-\Omega^2(1+R) & \tilde{c}_{16}z^2 & \tilde{c}_{15}z^2 \\
\frac{2\alpha_1}{\pi}\tilde{c}_{66}z & \frac{2\alpha_1}{\pi}\tilde{c}_{62}z & \frac{2\alpha_1}{\pi}\tilde{c}_{64}z & \frac{2\alpha_1}{\pi}\tilde{e}_{26}z & \tilde{c}_{61}z^2 & \tilde{c}_{66}z^2-\frac{1}{\alpha_2^2}\Omega^2(1+R) & \tilde{c}_{65}z^2 \\
\frac{2\alpha_1}{\pi}\tilde{c}_{56}z & \frac{2\alpha_1}{\pi}\tilde{c}_{52}z & \frac{2\alpha_1}{\pi}\tilde{c}_{54}z & \frac{2\alpha_1}{\pi}\tilde{e}_{25}z & \tilde{c}_{51}z^2 & \tilde{c}_{56}z^2 & \tilde{c}_{55}z^2-\Omega^2(1+R)
\end{bmatrix}
\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ A_5 \\ A_6 \\ A_7 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Matrix (M.1)

$$\begin{bmatrix}
\bar{c}_{66}-\Omega^2(1+2R) & \bar{c}_{62} & \bar{c}_{64} & \bar{e}_{26} \\
\bar{c}_{26} & \bar{c}_{22}-\Omega^2(1+2R) & \bar{c}_{24} & \bar{e}_{22} \\
\bar{c}_{46} & \bar{c}_{42} & \bar{c}_{44}-\Omega^2(1+2R) & \bar{e}_{24} \\
\bar{e}_{26}\left[1-\frac{8}{\pi^2}\right] & \bar{e}_{22}\left[1-\frac{8}{\pi^2}\right] & \bar{e}_{24}\left[1-\frac{8}{\pi^2}\right] & -\bar{\varepsilon}_{22}
\end{bmatrix}
\begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ B_4 \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{8}{\pi^2}\left[\frac{\bar{c}_{22}}{\hat{c}_{66}}\right]^{1/2}\phi_0 \end{bmatrix}$$

Matrix (M.2)

$$\begin{bmatrix} \\ \\ \\ M_{mn} \\ \\ \\ \\ \end{bmatrix}
\begin{bmatrix} A_{11} \\ A_{12} \\ A_{13} \\ A_{14} \\ A_{15} \\ A_{16} \\ A_{17} \end{bmatrix}
=
\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{2\alpha_1}{\pi}(\bar{c}_{16}B_1+\bar{c}_{12}B_2+\bar{c}_{14}B_3+\bar{e}_{21}B_4) \\ \frac{2\alpha_1}{\pi}(\bar{c}_{66}B_1+\bar{c}_{62}B_2+\bar{c}_{64}B_3+\bar{e}_{26}B_4) \\ \frac{2\alpha_1}{\pi}(\bar{c}_{56}B_1+\bar{c}_{52}B_2+\bar{c}_{54}B_3+\bar{e}_{25}B_4) \end{bmatrix}$$

Matrix (M.3)

a/b	R = 0.000	R = 0.005	R = 0.010	R = 0.020
17.0	0.924876	0.921895	0.918921	0.912991
17.0	0.948760	0.944171	0.939660	0.930872
17.0	0.978479	0.975219	0.971664	0.963929
17.0	0.983588	0.979866	0.976512	0.970643
17.0	1.013210	1.009818	1.007533	0.999104
17.0	1.018059	1.012562	1.011514	1.004672
17.0	1.035755	1.030791	1.025914	1.009012
17.0	1.046497	1.041863	1.037311	1.016432
17.0	1.057212	1.052729	1.048331	1.028466
20.0	0.928263	0.925289	0.922318	0.916370
20.0	0.947168	0.942567	0.938045	0.929250
20.0	0.980439	0.977789	0.975154	0.969923
20.0	0.984316	0.981536	0.978792	0.973373
20.0	1.004202	1.000324	0.996478	0.988892
20.0	1.030762	1.012562	1.007533	0.997141
20.0	1.031501	1.026692	1.022451	1.009012
20.0	1.043599	1.030382	1.025594	1.013761
20.0	1.062259	1.038576	1.033628	1.016636
23.0	0.930817	0.927849	0.924878	0.918899
23.0	0.946152	0.941542	0.937014	0.928227
23.0	0.954349	0.951842	0.949351	0.944417
23.0	0.966320	0.960477	0.957471	0.948859
23.0	0.984656	0.981623	0.978552	0.972252
23.0	0.995882	0.992195	0.988582	0.981627
23.0	1.016690	1.013508	1.008500	1.001478
23.0	1.032644	1.027780	1.022951	1.009012
23.0	1.037722	1.032811	1.027964	1.013408
23.0	1.046902	1.042406	1.037949	1.018460
23.0	1.053781	1.049846	1.045872	1.029171

Table 1. Resonant frequencies of SC-cut quartz strip with shorted surfaces

a/b	R = 0.000	R = 0.005	R = 0.010	R = 0.020
17.0	0.925084	0.922112	0.919149	0.913244
17.0	0.949716	0.945117	0.940595	0.931783
17.0	0.978802	0.975067	0.972255	0.964705
17.0	0.984303	0.980457	0.976958	0.970902
17.0	1.013546	1.011008	1.006813	0.999718
17.0	1.019132	1.015453	1.011878	1.004981
17.0	1.036584	1.031695	1.026713	1.011512
17.0	1.047307	1.042644	1.038061	1.017197
17.0	1.057773	1.053293	1.048896	1.029152
20.0	0.928459	0.925496	0.922539	0.916624
20.0	0.948130	0.943518	0.938985	0.930161
20.0	0.980468	0.977818	0.975184	0.969957
20.0	0.984563	0.981777	0.979033	0.973621
20.0	1.004810	1.000945	0.997111	0.989543
20.0	1.031738	1.010249	1.021999	0.995016
20.0	1.033435	1.011008	1.023257	1.011512
20.0	1.044328	1.031139	1.026359	1.014584
20.0	1.063140	1.039300	1.034348	1.017341
23.0	0.931005	0.928050	0.925095	0.919162
23.0	0.947118	0.942497	0.937957	0.929132
23.0	0.954365	0.951859	0.949369	0.944439
23.0	0.971096	0.967765	0.961562	0.952308
23.0	0.984965	0.981953	0.978911	0.972687
23.0	0.996457	0.992769	0.989149	0.982157
23.0	1.017249	1.010249	1.009520	1.002022
23.0	1.038565	1.028657	1.021999	1.011512
23.0	1.047642	1.033654	1.028807	1.014272
23.0	1.054090	1.043155	1.038704	1.019296
23.0	1.066514	1.050161	1.046201	1.029935

Table 2. Resonant frequencies of SC-cut quartz strip with open surfaces

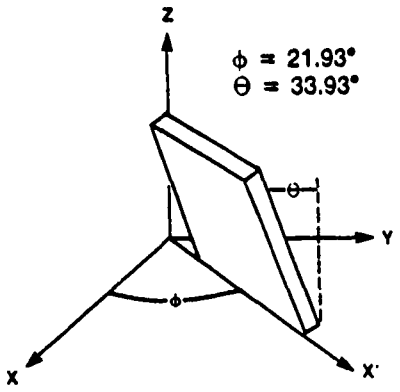


Figure 1. Illustration of an SC-cut Quartz Plate

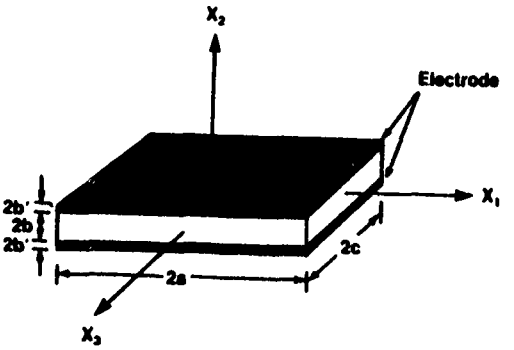


Figure 2. Thickness Field Excitation

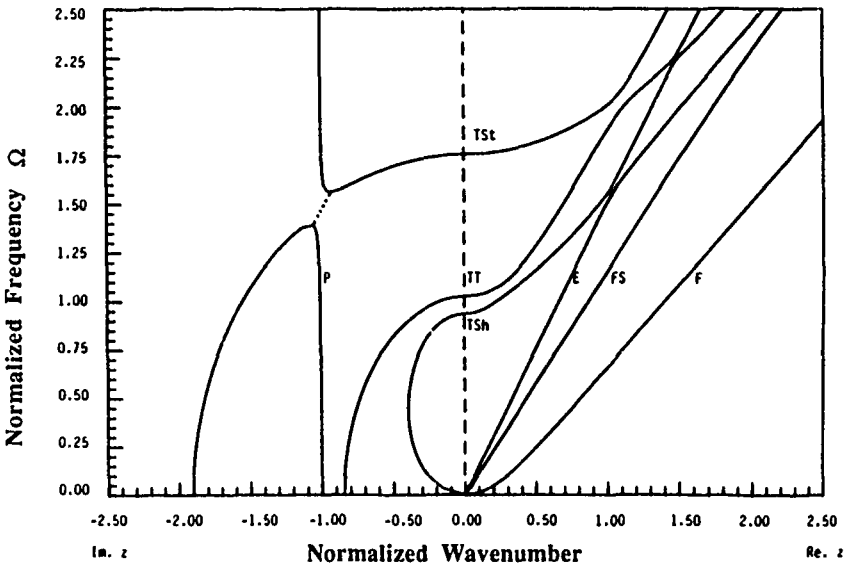


Figure 3. Dispersion Curves of an Infinite SC-cut Quartz Plate

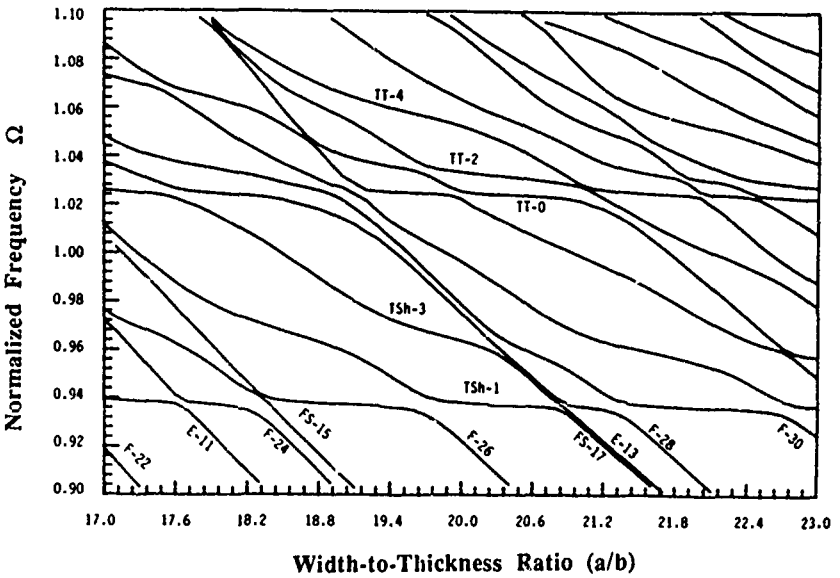


Figure 4. Frequency Spectrum of SC-cut Quartz Strips (Shorted Surfaces)

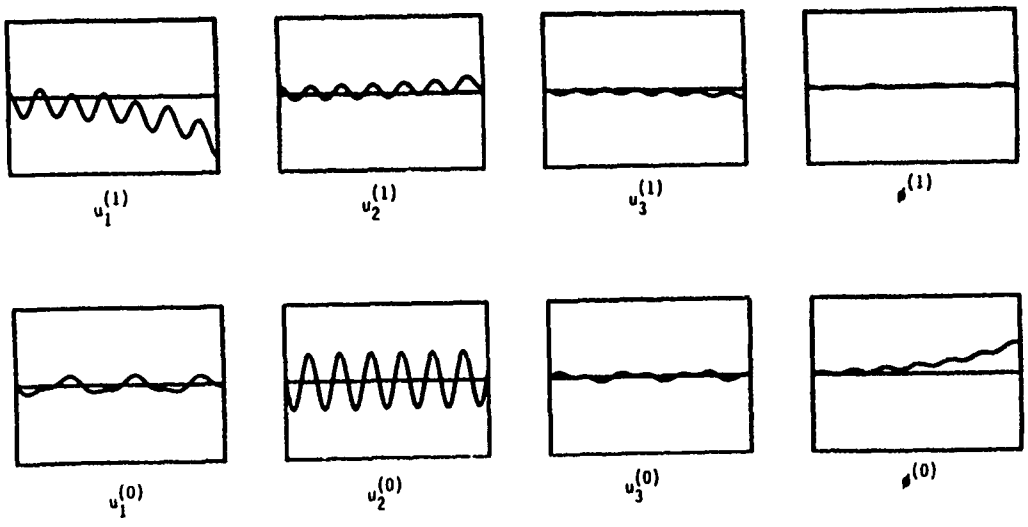


Figure 5. Modes for Predominately Flexural Vibration (F-26) at $\Omega = 0.922$, a/b = 20

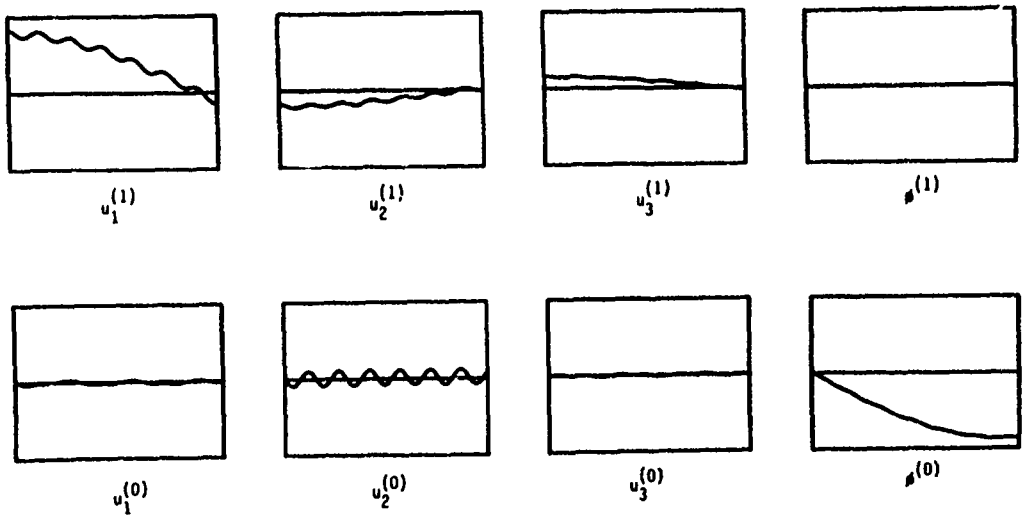


Figure 6. Modes for Predominately Thickness-Shear Vibration (TSh-1) at $\Omega = 0.938$, a/b = 20

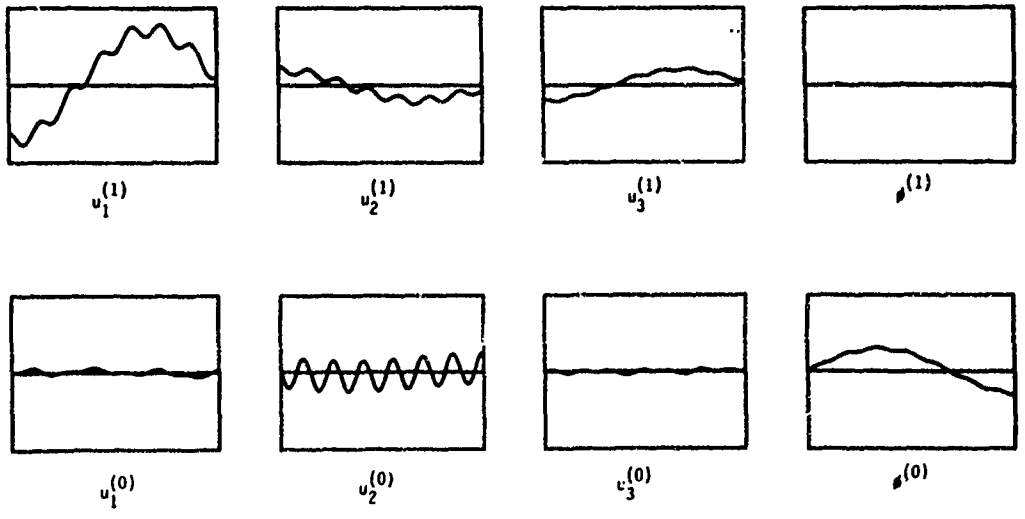


Figure 7. Modes for Predominately Thickness-Shear Vibration (TSh-3) at $\Omega = 0.963$, a/b = 20

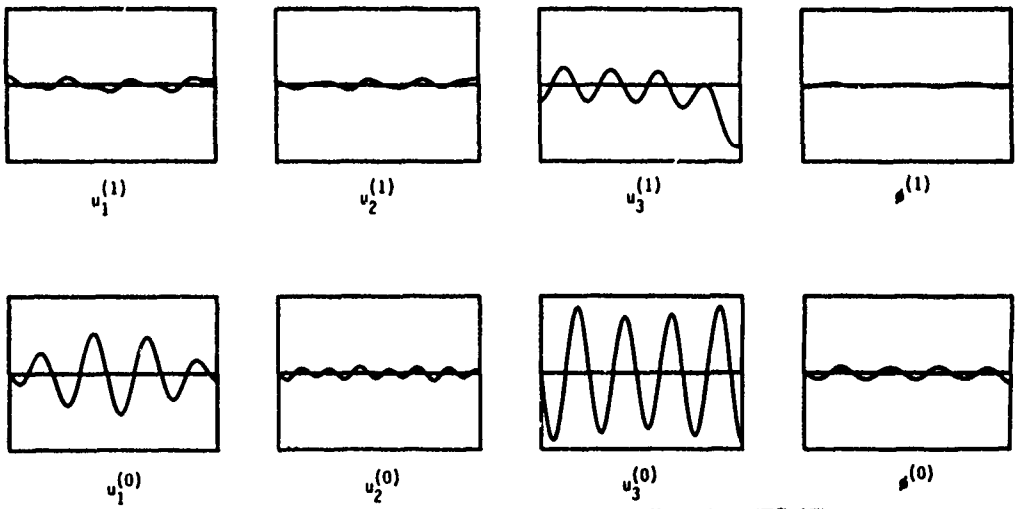


Figure 8. Modes for Predominately Face-Shear Vibration (FS-17) at $\Omega = 0.975$, a/b = 20

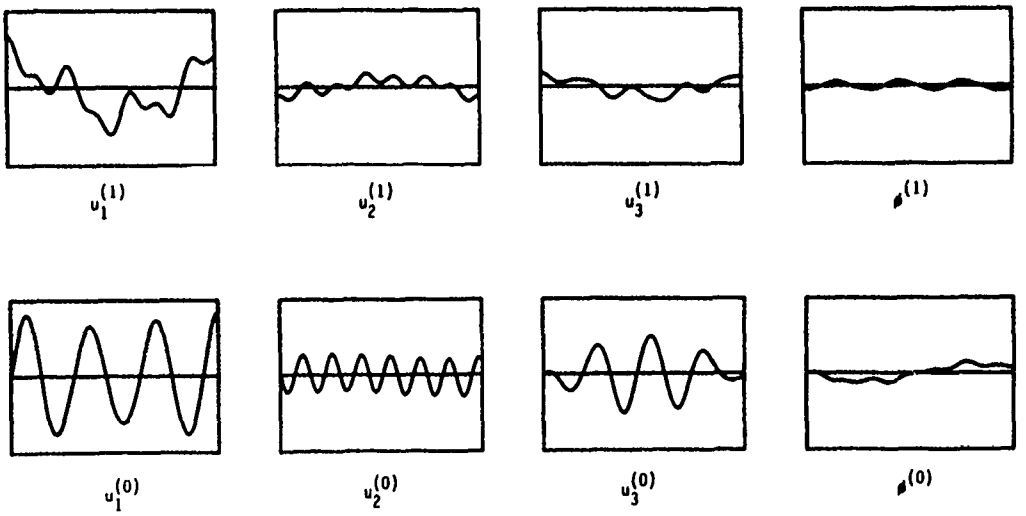


Figure 9. Modes for Predominately Extension Vibration (E-13) at $\Omega = 0.979$, a/b = 20

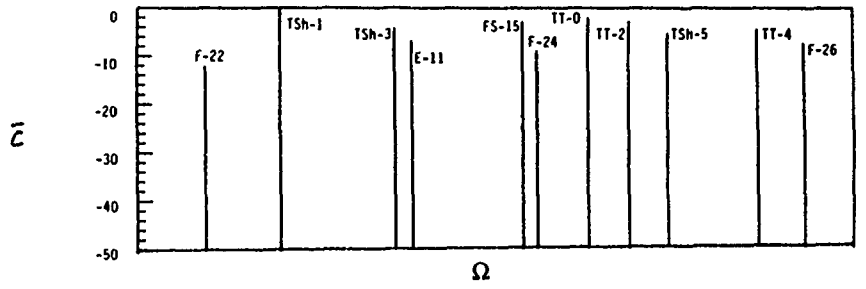


Figure 10. Frequency Separation Factors of SC-cut Quartz Strips, a/b = 17

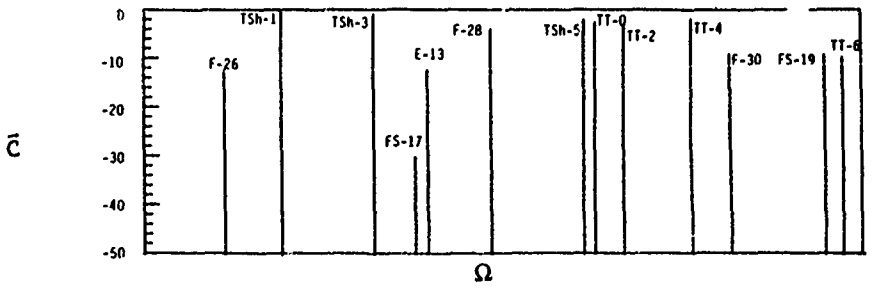


Figure 11. Frequency Separation Factors of SC-cut Quartz Strips, a/b = 20

43rd Annual Symposium on Frequency Control - 1989

LATERAL- AND THICKNESS-FIELD COUPLING IN LITHIUM TETRABORATE

A. Ballato, J. Kosinski, M. Mizan, and T. Lukaszek

US Army Electronics Technology and Devices Laboratory
LABCOM, Fort Monmouth, NJ 07703-5000

ABSTRACT

Lithium tetraborate is a tetragonal material of considerable promise for frequency control and signal processing applications. It exhibits piezoelectric coupling values that fall between those of lithium niobate and quartz, but possess orientations for which the temperature coefficient of frequency or delay time is zero for both bulk and surface acoustic waves.

Calculations have previously been made for rotated y-cuts, including the regions where the quasi-extensional and quasi-shear thickness modes have zero temperature coefficients of frequency. In this paper we extend the calculations to doubly rotated bulk wave resonators, and compute the coupling factors for the three simple thickness modes driven by thickness (TE) and lateral (LE) quasistatic electric fields as a function of the orientation angles phi and theta, and the direction of the applied lateral field psi. Because of the temperature coefficients of the piezoelectric coupling factors, the temperature coefficient of a resonator will depend not only upon orientation, but also upon harmonic number and location of the resonator operating point on the immittance circle. These considerations are addressed.

INTRODUCTION

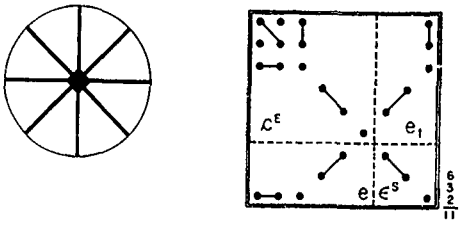
Lithium tetraborate (LBO) is a tetragonal material in crystal class 4mm (C_{4v}). As such it possesses a single 4-fold polar axis, and four symmetry planes containing the 4-fold axis; these are shown in Fig. 1. The primitive region is 1/8 th of a hemisphere, which we comprise as the angle ranges (yxwl)Φ/Θ, with $0 \leq \Phi \leq \pi/4$ and $0 \leq \Theta \leq \pi/2$. As a consequence of its symmetry, LBO is

- Pyroelectric
- Optically uniaxial (LBO is negative)
- Piezoelectric
- Not enantiomorphic (no twinning)
- Nonferroelectric (poling not required)

Figure 1 shows the pertinent Van Dyke matrix: there are 6 independent linear elastic constants, three independent linear piezoelectric constants, and two independent linear dielectric constants.

Particularizing to the substance lithium tetraborate (LBO), we find from the literature [1]-[26] the following specific properties and virtues:

LITHIUM TETRABORATE



TETRAGONAL SYSTEM IVb CRYSTAL CLASS 4mm [C_{4v}]

Figure 1. Symmetry elements and Van Dyke matrix for lithium tetraborate, class 4mm.

Properties of $Li_2B_4O_7 = Li_2O \cdot 2B_2O_3$

- Congruently melting phase in the lithium oxide-boron oxide system; transparent and colorless.
- Low melting point: 917°C.
- Czochralski growth, Pt crucibles, diameters > 50 mm along [100], [001], or [110]; sensitive to thermal shock (cooling).
- Lattice spacings: a = b = 9.479 Å, c = 10.280 Å
- Mohs hardness = 6 (between $LiTaO_3$ and quartz = 7).
- Low density = 2451 kg/m^3, but acoustic velocities only slightly greater the those in $LiNbO_3$ and $LiTaO_3$.
- Solubility: 1) dissolves rapidly in acids, slowly in bases. 2) hot water used as etchant. 3) insolvent in organic "solvents."
- Relatively high piezocoupling k and $\underline{k}$ values.
- Surface acoustic wave (SAW) reflectivity per stripe > 5 times that for $LiNbO_3$, $LiaO_3$, and quartz, leading to miniaturization.
- Zero temperature coefficients of frequency and time delay for BAW and SAW.

DETERMINATION OF CONSTANTS

The elastic, piezoelectric, and dielectric constants of 4mm crystals may be determined from the simple thickness modes of thin plates driven by thickness excitation [TE] and lateral excitation [LE].

Orientation (yx); Y-cut = X-cut

[TE]: pure shear along X_3
c_{44}^E, e_{15}, ϵ_{11}^S
[LE]: pure stretch along X_2, driven by X_3 field
c_{11}^E, e_{31}

Orientation (zx); Z-cut

[TE]: pure stretch along X_3
c_{33}^E, e_{33}, ϵ_{33}^S

Orientation (yxl)θ; rotated Y-cut

[TE]: coupled shear-stretch
c_{13}^E
[LE]: pure shear along X_1, driven by X_1 field
c_{66}^E

Orientation (yxw)Φ

[LE]: coupled shear-stretch, field along X_3; X_1, X_2 motion
c_{12}^E

COMPUTATIONAL SCHEME

Input data are taken from Ref. [19], and used as follows:

- s^E, d, ϵ^T are converted to c^E, e, ϵ^S
- c^E, e, ϵ^S, density, and thickness are given at reference temperature T_o = 25°C
- first- and second-order temperature coefficients TC(1) and TC(2) are used to compute c^E, etc., at two other temperatures, T_c and T_h
- For assumed angles Φ,θ, and psi, the conventional eigenvalue problem is solved to yield N_m, k_m, $\underline{k}_m$(psi), etc., for each temperature, T_c, T_o, and T_h
- TC(1) and TC(2) of f_R, f_A, etc., are computed for each mode, harmonic, and excitation type; for further details, see Refs. [27]-[29].

VELOCITY AND COUPLING

The frequency constants, N_m, are given in Fig. 2 for rotated Y-cuts; the corresponding [TE] coupling factors, k, are shown in Fig. 3. Figures 4-7 present the [LE] coupling factors, $\underline{k}$(psi), for rotated Y-cuts having applied field direction psi = 0°(30°)90°, respectively.

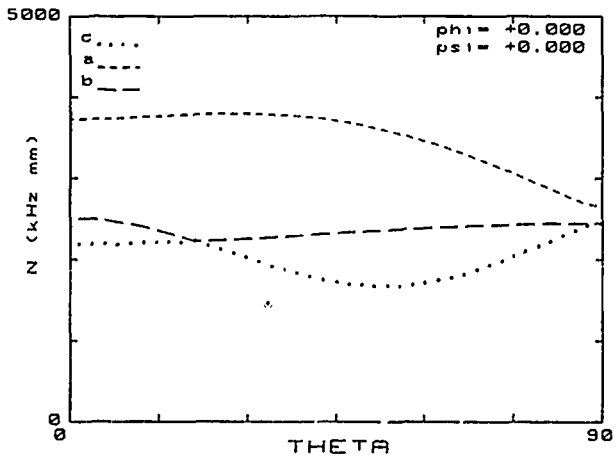


Figure 2. Frequency constants for (yxwl)Φ = 0°/θ open-circuited plates; shear mode degeneracy occurs at approximately 20°.

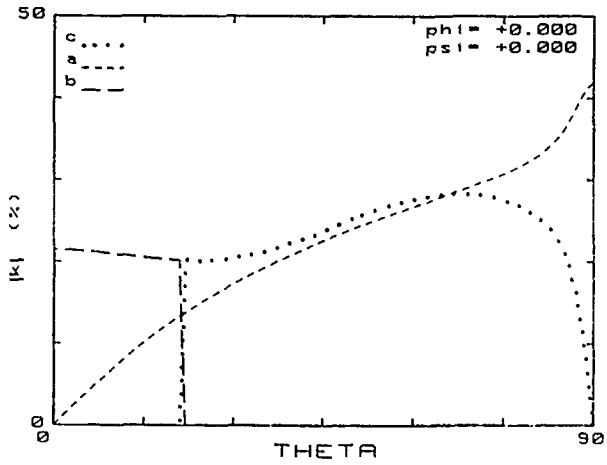


Figure 3. Thickness excitation piezocoupling factors for (yxwl)Φ = 0°/θ plates; these are weak functions of angle Φ.

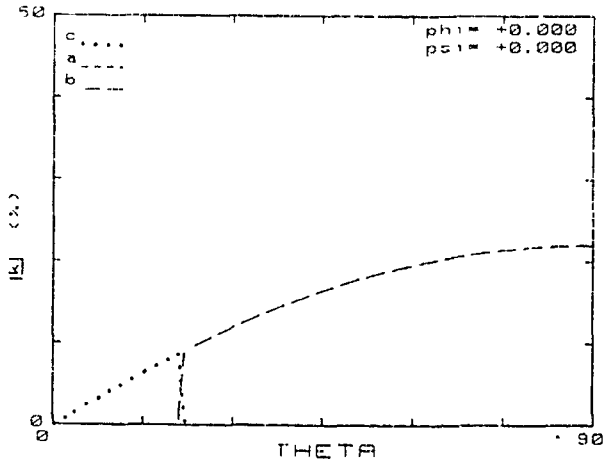


Figure 4. Lateral excitation piezocoupling factors for (yxwl)Φ = 0°/θ plates.

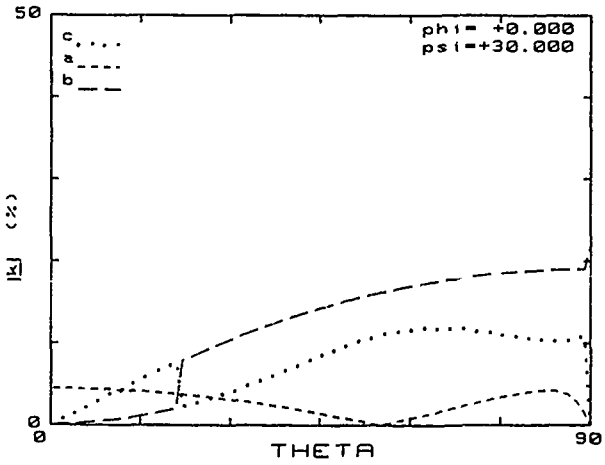


Figure 5. Lateral excitation piezocoupling factors for (yxwl)Φ = 30°/θ plates.

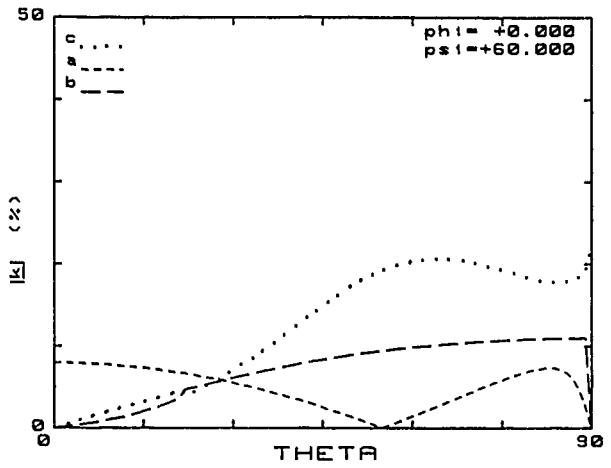


Figure 6. Lateral excitation piezocoupling factors for (yxwl)Φ = 60°/θ plates.

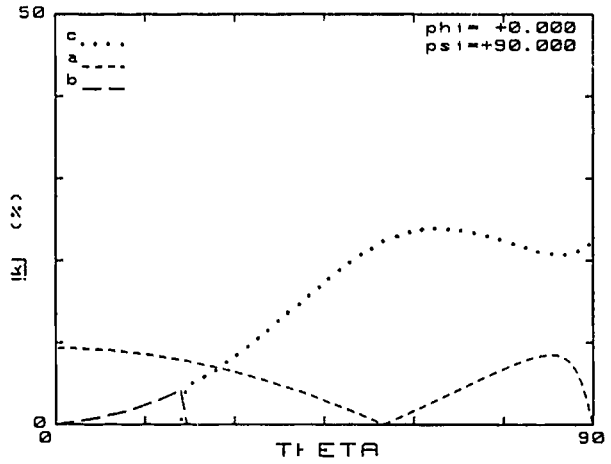


Figure 7. Lateral excitation piezocoupling factors for (yxwl)Φ = 90°/θ plates.

TEMPERATURE COEFFICIENTS

The first-order temperature coefficients of antiresonance frequency, for the c mode of rotated Y-cuts driven by lateral fields is shown in Fig. 8 for the fundamental harmonic. Angle psi is the parameter. For values of angle θ > about 20°, it is seen from Fig. 4 that the c mode is undriven by [LE] for psi = 0°; in Fig. 8 this corresponds to the upper curve, which coincides with the [LE] resonance frequency TC. The influence of k(psi) and its TC is seen in the change of TC(1) with psi in Fig.8.

Figure 9 presents the [TE] case paralleling Fig. 8, which means the c mode TC(1) of resonance frequency. Figure 3 shows that the c mode is driven for θ > about 20°. The influence of coupling is most readily observed in [TE] plates by varying harmonic, and one sees in Fig. 9 that as harmonic number becomes indefinitely large, the TC(1) of resonance frequency becomes equal to the TC(1) of antiresonance frequency in Fig. 8 with psi = 0°.

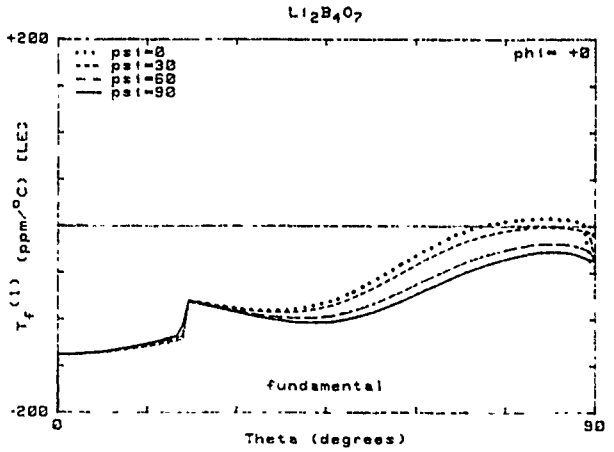


Figure 8. Lateral excitation, first order temperature coefficients of slow shear "c" mode frequency for (yxwl)Φ = 0°/θ plates for various values of azimuthal angle psi. Psi = 0° corresponds to electric field along the X_1 axis.

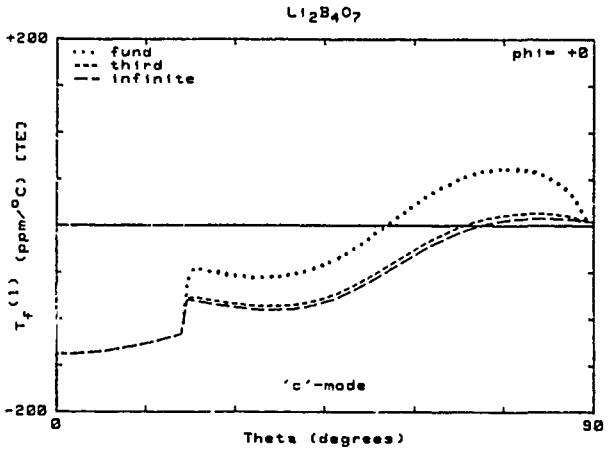


Figure 9. Thickness excitation, first order temperature coefficients of slow shear "c" mode frequency for (yxwl)Φ = 0°/θ plates operating at the fundamental, third, and infinite harmonics.

In LBO the fast shear mode (b mode) is not of great interest with respect to its TC values, and is not discussed here. Figures 10 and 11, are the a mode equivalents of Figs. 8 and 9, respectively. Although zeros of TC(1) exist for all harmonics, the [LE] coupling for this mode is weak; the [TE] coupling is much stronger, leading to a considerable displacement in the angles at which the zero TC(1) occurs as function of overtone.

Loci of TC(1) = 0 for [TE], a-mode, doubly rotated plates at several harmonics are given in Fig. 12. The plot shows only a weak sensitivity to variations with angle Φ. In Fig. 13 are shown values of TC(2) for rotated Y-cuts with harmonic as parameter. In Fig. 14 the TC(2) = 0 loci for [TE], a-mode, doubly rotated plates are shown. Here the variation with angle Φ is much greater than with TC(1).

Superposition of Figs. 12 and 14 discloses that a unique doubly rotated orientation in LBO exists for which TC(1) = TC(2) = 0. It occurs for $\Phi/\theta \approx 40°/33°$ at the fundamental harmonic, driven in [TE]. The corresponding N_a and k_a values may be read approximately from the graphs in Figs. 15 and 16, respectively.

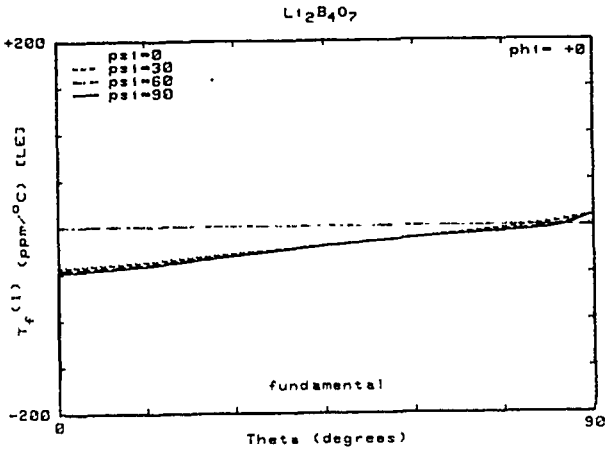


Figure 10. Lateral excitation, first order temperature coefficients of extensional "a" mode frequency for (yxwl)Φ = 0°/θ plates for various values of azimuthal angle psi. Psi = 0° corresponds to electric field along the X_1 axis.

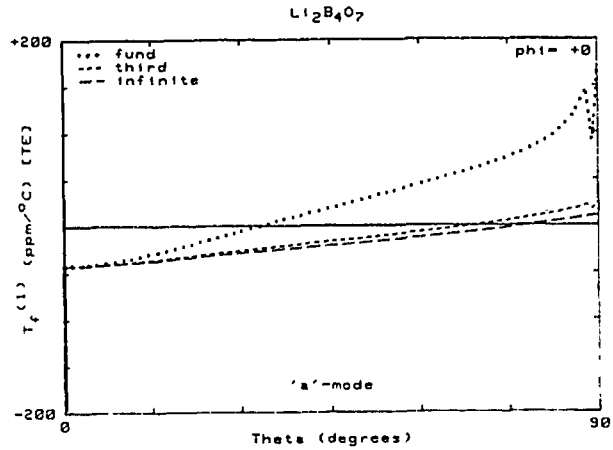


Figure 11. Thickness excitation, first order temperature coefficients of extensional "a" mode frequency for (yxwl)Φ = 0°/θ plates operating at the fundamental, third, and infinite harmonics.

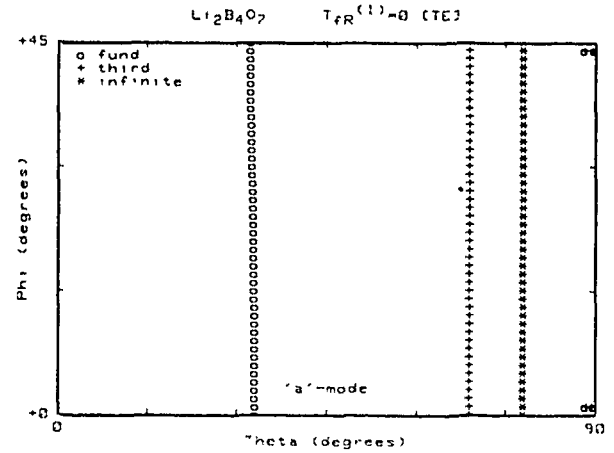


Figure 12. Loci of zeros of thickness excitation, first order temperature coefficients of extensional "a" mode resonance frequencies of (yxwl)Φ/θ plates operating at the fundamental, third, and infinite harmonics.

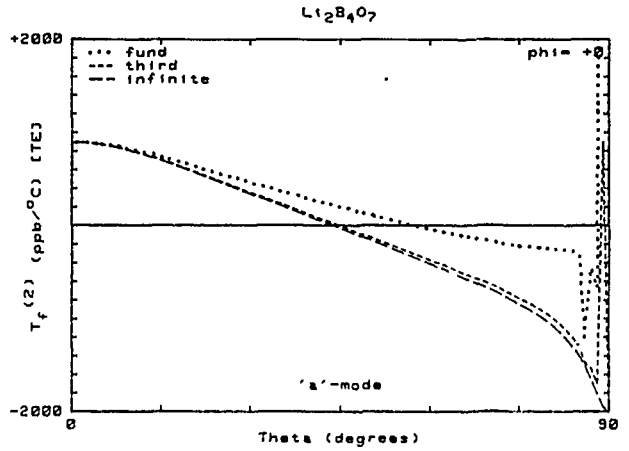


Figure 13. Thickness excitation, second order temperature coefficients of extensional "a" mode frequency for (yxwl)Φ = 0°/θ plates operating at the fundamental, third, and infinite harmonics.

REFERENCES

[01] G. S. Smith and G. E. Rindone, "High-Temperature Energy Relations in the Alkali Borates: Binary Alkali Borate Compounds and Their Glasses," J. Am. Ceram. Soc., Vol. 44, No. 2, 1961, pp. 72-78.

[02] J. Krogh-Moe, "The Crystal Structure of Lithium Diborate, $Li_2O{\cdot}2B_2O_3$," Acta Cryst., Vol. 15, 1962, pp.190-193.

[03] J. Krogh-Moe, "Refinement of the Crystal Structure of Lithium Diborate, $Li_2O{\cdot}2B_2O_3$," Acta Cryst., Vol. B24, Pt.2, 1968, pp. 179-131.

[04] S. R. Nagel, L. W. Herron, and C. G. Bergeron, "Crystal Growth of $Li_2B_4O_7$," J. Am. Ceram. Soc., Vol. 60, No. 3-4, March-April 1977, pp. 172-173.

[05] J. D. Garrett, M. N. Iyer, and J. E. Greedan, "The Czochralski Growth of $LiBO_2$ and $Li_2B_4O_7$," J. Cryst. Growth, Vol. 41, 1977, pp. 225-227.

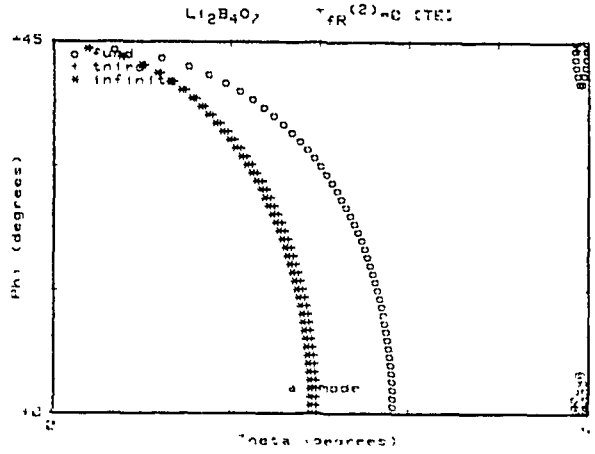


Figure 14. Loci of zeros of thickness excitation, second order temperature coefficients of extensional "a" mode resonance frequencies of (yxwl)Φ/θ plates operating at the fundamental, third, and infinite harmonics.

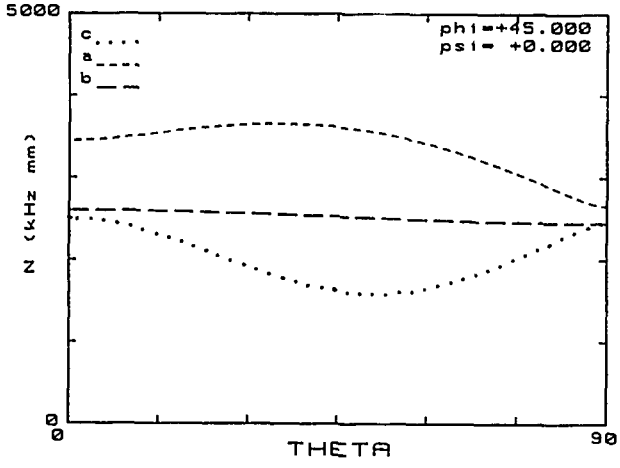


Figure 15. Frequency constants for (yxwl)Φ = 45°/θ open-circuited plates.

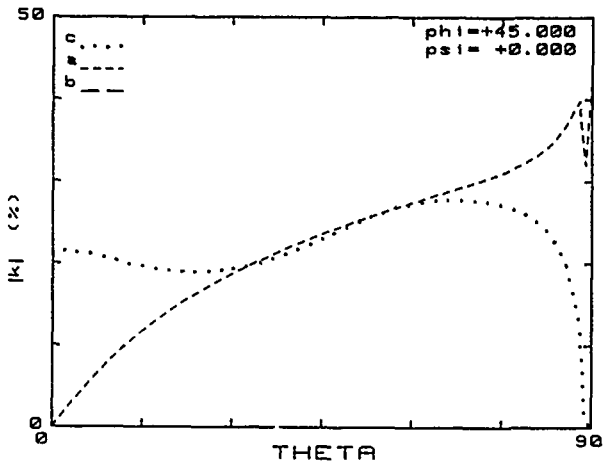


Figure 16. Thickness excitation piezocoupling factors for (yxwl)Φ = 45°/θ plates.

[06] R. W. Whatmore, N. M. Shorrocks, C. O'Hara, F. W. Ainger, and I. M. Young, "Lithium Tetraborate: A New Temperature-Compensated SAW Substrate Material," Elx. Lett., Vol. 17, No. 1, 8th January 1981, pp. 11-12.

[07] N. M. Shorrocks, R. W. Whatmore, F. W. Ainger, and I. M. Young, "Lithium Tetraborate: A New Temperature Compensated Piezoelectric Substrate Material for Surface Acoustic Wave Devices," IEEE Ultrasonics Symp. Proc., 1981, pp. 337-340.

[08] B. Lewis, N. M. Shorrocks, and R. W. Whatmore, "An Assessment of Lithium Tetraborate for SAW Applications," IEEE Ultrasonics Symp. Proc., 1982, pp. 389-393.

[09] D. S. Robertson and I. M. Young, "The Growth and Growth Mechanism of Lithium Tetraborate," J. Mater. Sci., Vol. 17, 1982, pp. 1729-1738.

[10] Y. Ebata, H. Suzuki, S. Matsumura, and K. Fukuta, "SAW Propagation Characteristics on $Li_2B_4O_7$," Proc. 3rd Symp. Ultrasonic Elx., Tokyo, 1982; Jpn. J. Appl. Phys., Vol. 22, 1983, Suppl. 22-3, pp. 160-162.

[11] C. D. J. Emin and J. F. Werner, "The Bulk Acoustic Wave Properties of Lithium Tetraborate," Prcc. 37th Annual Frequency Control Symp., June 1983, pp. 136-143.

[12] R. C. Peach, C. D. J. Emin, J. F. Werner, and S. P. Doherty, "High Coupling Piezoelectric Resonators Using Lithium Tetraborate," IEEE Ultrasonics Symp. Proc., 1983, pp. 521-526.

[13] K. Fukuta, J. Ushizawa, H. Suzuki, Y. Ebata, and S. Matsumura, "Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Applications," Proc. 4th Meeting on Ferroelectric Materials and Their Applications, Kyoto, 1983; Jpn. J. Appl. Phys., Vol. 22, 1983, Suppl. 22-2, pp. 140-142.

[14] T. Shiosaki, M. Adachi, H. Kobayashi, K. Araki, and A. Kawabata, "Elastic, Piezoelectric, Acousto-Optic and Electro-Optic Properties of $Li_2B_4O_7$," Proc. 5th Symp. Ultrasonic Elx., Tokyo, 1984, pp. 5-8; Jpn. J. Appl. Phys., Vol. 24, 1985, Suppl. 24-1, pp. 25-27.

[15] M. Adachi, T. Shiosaki, and A. Kawabata, "Crystal Growth of Lithium Tetraborate ($Li_2B_4O_7$)," Proc. 5th Meeting on Ferroelectric Materials and Their Applications, Kyoto, 1985; Jpn. J. Appl. Phys., Vol. 24, 1985, Suppl. 24-3, pp. 72-75.

[16] Y. Fujiwara, M. Ono, M. Sakai, and N. Wakatsuki, "Strip Type Resonator of Lithium Tetraborate," Proc. 39th Annual Frequency Control Symp., May 1985, pp. 351-355.

[17] M. Adachi, T. Shiosaki, H. Kobayashi, O. Ohnishi, and A. Kawabata, "Temperature Compensated Piezoelectric Lithium Tetraborate Crystal for High Frequency Surface Acoustic Wave and Bulk Wave Device Applications," IEEE Ultrasonics Symp. Proc., 1985, pp. 228-232.

[18] A. S. Bhalla, L. E. Cross, and R. W. Whatmore, "Pyroelectric and Piezoelectric Properties of Lithium Tetraborate Single Crystal," Jpn. J. Appl. Phys., Vol. 24, 1985, Suppl. 24-2, pp. 727-729.

[19] T. Shiosaki, M. Adachi, and A. Kawabata, "Growth and Properties of Piezoelectric Lithium Tetraborate Crystal for BAW and SAW Devices," IEEE Intl. Symp. Applications Ferroelectrics (ISAF) Proc., Lehigh University, Bethlehem, PA, June 1986, pp. 455-464.

[20] A. Ballato, E. R. Hatch, T. Lukaszek, and M. Mizan, "Lateral-Field Coupling of Rotated BAW Plates with 3m, 4mm, & 4 bar 3m Symmetries," IEEE Ultrasonics Symp. Proc., 1986, pp. 339-342.

[21] H. Suzuki, Y. Ebata, S. Matsumura, and J. Ushizawa, "Surface Acoustic Wave Device," U.S. Patent 4,672,255, issued June 9, 1987.

[22] Y. Ebata and M. Koshino, "SAW Resonator and Resonator Filter on $Li_2B_4O_7$ Substrate," Jpn. J. Appl. Phys., Vol. 26, 1987, Suppl. 26-1, pp. 123-125.

[23] H. Abe, H. Saitou, M. Ohmura, T. Yamada, and K. Miwa, "Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators," IEEE Ultrasonics Symp. Proc. 1987, pp. 91-94.

[24] W. S. Ishak, C. A. Flory, and B. A. Auld, "Acoustic Modes in Lithium Tetraborate," IEEE Ultrasonics Symp. Proc., 1987, pp. 241-245.

[25] S. Matsumura, T. Omi, N. Yamaji, and Y. Ebata, "A 45^{o} X Cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators," IEEE Ultrasonics Symp. Proc., 1987, pp. 247-250.

[26] M. Murota and Y. Shimizu, "Characteristics of Leaky Surface Waves Propagating on $Li_2B_4O_7$," IEEE Ultrasonics Symposium Proc., Montreal, Canada, October 1989.

[27] M. Onoe, "Relationship between Temperature Behavior of Resonant and Antiresonant Frequencies and Electromechanical Coupling Factors of Piezoelectric Resonators," Proc. IEEE, Vol. 57, No. 4, April 1969, pp. 702-703.

[28] T. Yamada and N. Niizeki, "Admittance of Piezoelectric Plates Vibrating under the Perpendicular Field Excitation," Proc. IEEE, Vol.58, No.6, June 1970, pp. 941-942; "Formulation of Admittance for Parallel Field Excitation of Piezoelectric Plates," J. Appl. Phys., Vol. 41, No. 9, August 1970, pp. 3604-3609; "A New Formulation of Piezoelectric Plate Thickness Vibration," Rev. Elec. Comm. Lab. NTT (Tokyo), Vol. 19, Nos. 5-6, May-June 1971, pp. 705-713.

[29] A. Ballato and T. Lukaszek, "Frequency-Temperature Coefficients of Mass-Loaded Plate Vibrators," Technical Report ECOM-4356, US Army Electronics Command, Fort Monmouth, NJ 07703, September 1975, 49pp.

43rd Annual Symposium on Frequency Control - 1989

AIR-GAP DETERMINATION OF THE CUT-OFF FREQUENCY OF UNELECTRODED PLATES AND EFFECTS OF LOCAL THICKNESS MODIFICATIONS IN PLANE RESONATORS

J. Détaint*, H. Carru*, J. Schwartzel*, C. Joly*, B. Capelle**, A. Zarka**.

*Centre National d'Etudes des Télécommunications
Dept PAB/BAG/MCT/CMM 92220 Bagneux France

**Laboratoire de Minéralogie et de Cristallographie
Université P. et M. Curie 75005 Paris France

SUMMARY

In the course of a study of VHF resonators for an application to the intermediate frequency filters of the new Pan-European numerical radiotelephone system, it was observed that a very precise control of the resonators parameters was needed. The cut-off frequency of the unelectroded flat plate, the electrical frequency lowering, the mass loading and the electrode geometry are the most important parameters of the plane resonators. They determine the modes and the electrical response of the resonators. The exact knowledge of the first quantity is of prime importance for the modelization and to optimize several steps of the elaboration of the resonators. An approximate determination of this quantity is usually made by an air gap frequency measurement of the unelectroded plates. To obtain a greater precision for these measurements, electrodes of size much smaller than the plate diameter are generally used, since they lead to a simpler frequency response characteristic of a trapped mode. The frequency so obtained differs from the cut-off frequency by a quantity that can be quite important and is a function of the gap, of the electrodes dimensions and of the coupling coefficient of the material. In the first part of this communication we present a method to extract the cut-off frequencey from such air gap measurements with a model of the corresponding resonators.
Then, we consider the interest of modifications of the cut-off frequencies of selected regions of a plane resonator. This can be achieved by local thickness modifications, and permits to obtain new degrees of freedom in the design of resonators. The interest of this possibility has already been demonstrated in one case by Lukaszek who has proposed the use of resonators with electrodes embedded in the plate to gain a better control of the unwanted response and of the impedance level of V.H.F. resonator. A theory of the corresponding resonators is presented and the possibility given by this design explored. If three regions of the resonator are allowed to have different thicknesses, new possibilities are gained, such as to eliminate a selected unwanted anharmonic mode or to obtain a faster decrease of the principal mode across the plate. A theory of such resonator with 3 regions is presented, As the theories corresponding to the two previous cases, it is based on the approximate equations to the thickness vibrations of piezoelectric plates established by H.F. Tiersten and coworkers. These equations are solved either by a semi-algebraical method or numerically by the finite elements method. Examples of computed modes are given and compared to experimentally observed ones.

I - INTRODUCTION

More and more modern Telecommunication systems make use of very high frequency piezoelectric devices for filtering or for frequency generation. This will be the case for the new Pan-European numerical radiotelephone system, now in a development phase, to be operational in the early nineties. It will make use of a very large number of V.H.F devices (Antenna filters, intermediate frequency filters etc..).In the course of a study concerning bulk wave versions of the latter filters, we have observed that significative conceptual and technical advances were needed to establish designs for these filters that permit to obtain the wanted characteristics together with the possibility of a large volume production at a cost compatible with the objectives of a very large diffusion of the mobile and portable stations.
The first point observed was that a very tight control and a very high reproductibility of electrical parameters of the piezoelectric components of the filter were required at a high frequency (near 70 MHz). Since the electrical parameters (equivalent scheme, unwanted responses) of a plane resonator are mostly dependant on three parameters (the thickness [or the cut-off frequency of the unelectrode plate], the electrodes geometry and the plate back) and since the two latter ones parameters can be accurately controled using the modern technology (masquing techniques and precise quartz thin film thickness monitoring), it was deduced that a critical aspect will be to achieve a very precise control of the plate thickness (in the order of 20 micron for the berlinite version of the filter). A method, based upon the theory of the thickness vibrations of the piezoelectric plates (1)(2)(3), permitting to obtain very accurate measurements of the cut-off frequency of very high frequency plates was investigated. This method will be described in the first part of this paper.

In order to obtain a better control of the anharmonic spectrum of the high frequency resonators (the rejection of the anharmonic modes has to be very high for this application) we have theoretically investigated, a method originally proposed par Lukaszek (4), which consists in depositing the electrodes in a previously etched region of the plate. This design allows also to obtain a much lower impedance level for the resonators. The corresponding theory and results will be given in paragraph III.

Other local modifications of the plate thickness can allow to minimize the lateral dimensions of the devices or to integrate several resonators on a single plate. To investigate this point a theoretical model of resonators having three regions with different thickness was established, it will be described in paragraph IV.

CH2690-6/89/0000-563 $1 00 C 1989 IEEE

In paragraph V a finite elements solution of the Tiersten equations, relative to the cases presented in paragraphs II,III and IV, is described and the results obtained by this method are compared to those of the semialgebraical solutions and to experimental ones.

All these investigations are based upon the theory established by H.F. Tiersten and coworkers of the thickness vibration of the piezoelectric plates. They extend previous works concerning the modelization of plane resonators with electrodes of arbitrary geometry (5) and of corrugated resonators (6).

II - Accurate determination of the cut-off frequency of unelectroded plane piezoelectric plates

II.1 - Air gap measurements and the cut-off frequency

This quantity (f_{cl}) which is the antiresonance frequency of an infinite plate having a zero mass perfectly conductive electrode is, as shown by the theory, with the mass loading and the electrode geometry, a fundamental parameter of the thickness mode plane resonators. It is the parameter that include the thickness effect and most of the material properties of the resonator.

In practice, the exact knowledge of this quantity is required for the modelization of the resonator and also to make resonators at the wanted frequency with a precise mass loading. Errors on these quantities can have detrimental effects on the equivalent scheme and on the parasitic response of the resonator.

The frequency directly measured from air gap measurements differs from f_{cl} by a quantity which is, in relative value, of the order of $4\,k^2/\pi^2$ (2,6 % for AT quartz and 4 % for AT Berlinite). This error which is very dependent of several experimental parameters (gap, electrode dimension) is an important fraction of the mass loading used for high frequency resonators (about 0,25 to 1 %) so that this frequency cannot be used directly when a precise control of the mass loading is required.

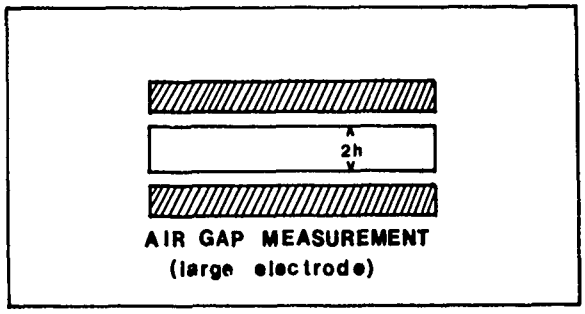


Figure 1 : Air gap measurements with a large electrode

To obtain from air gap measurements a quite exact value of f_{cl} it is necessary to use a theoretical model of the corresponding experimental set up. Two type of experimental set up are of common use. In the first one the plate is situated between two plates having the same dimensions at the plate *(figure 1)* and the quantity measured is either the anti resonance frequency at zero gap or the resonance frequency with a large gap. f_{cl} can be determined by a theory of the vibration of unelectroded square or circular plates made by H.F. Tiersten (7). However, with this experimental set up, the electrical response is very often complicated by plates modes coupled at the edge and is perturbated (antiresonance measurements) by the parasitic capacitance.

The second experimental set up is represented in figure 2, it uses an electrode with a diameter much lower than the lateral dimensions of the plate. A small gap is generally used and transmission measurements of the resonance frequency most generally employed.

With this set up the electrical frequency lowering in the region under the electrodes is generally sufficient to induce the trapping of the vibration mode at the center of the plate so that "clean" frequency responses with low losses are obtained. This permits to have simpler modes (by avoiding the coupling to plates modes due to a non negligeable value of the displacement at the plate edge) and thereby to make precise frequency measurements. A typical exemple of the obtained response is given in figure 3. It concerns a 20 MHz AT quartz plate, the electrode diameter was 2,000 mm and the gap varied from 10 µm to 130 µm with 10 µm increments. On figure 3 it is possible to observe that the antiresonance frequency varies with the gap.

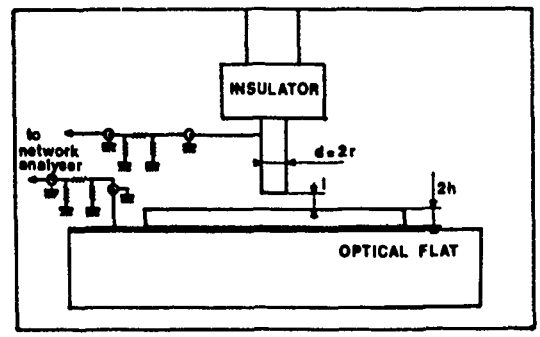


Figure 2 : Air gap measurements with a small electrode

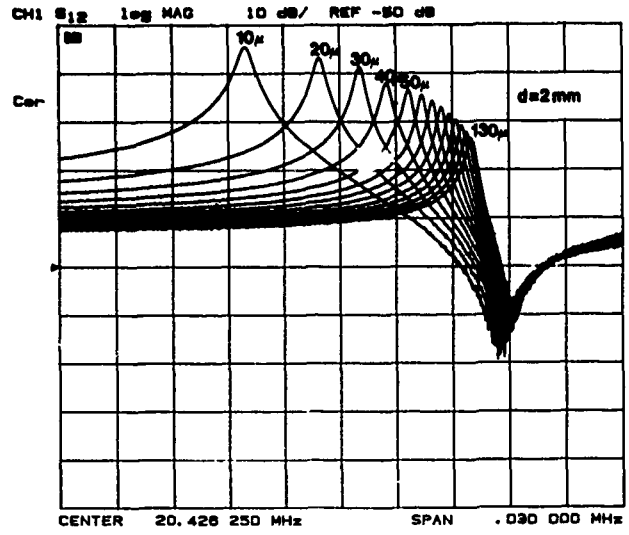


Figure 3 : Example of the frequency responses obtained with a variable gap

If a zero gap condition is realized, without applying any stresses on the plate, then, the resonance frequency is that of a resonator having perfectly conductive electrodes with a zero mass. Thin can be accurately simulated by the mathematical extrapolation of the resonances frequencies measured with different gaps.

II.2 - Modelization of a resonator where the energy trapping results only of the electrical lowering

As a basis of this model, we use the scalar equation governing the lateral dependance of the anharmonics in the vicinity of one overtone (number n) of a given thickness mode. This equation was first established for monoclinic plates (2) then for plates of an arbitrary orientation (3). In this case the equation has, in a coordinate system with x_2 normal to the thickness and x_1 x_3 choosen to eliminate the mixed derivative (3) the following form :

$$M_n \frac{\partial^2 \bar{u}_1^n}{\partial x_1^2} + P_n \frac{\partial^2 \bar{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 C^*}{4h^2}\bar{u}_1^n + \rho\omega^2\bar{u}_1^n = \rho\omega^2(-1)^{(n-1)/2}\frac{e_{26}4V_0 e^{j\omega t}}{C^{(1)}n^2\pi^2} \quad [1]$$

In this equation, $\bar{u}_2(x_1,x_3,t)$ is the lateral dependance of $\hat{u}_1^n(x_1,x_2,x_3,t) = \bar{u}_1^n \sin(n\pi x_2/2h)$.

$\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of ϕ the potential , made to replace the inhomogeneous boundary conditions $\phi = \pm V_0 e^{j\omega t}/2 at x_2 = \pm h$ by the homogeneous conditions $\hat{\phi} = 0$. The inhomogeneous term of the equation results from this transformation.

c* is either $\bar{c}^{(1)}$ for the unelectroded part of the resonator or $\acute{c}^{(1)}$ 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).

$$u_1^n = \hat{u}_1^n - \frac{e_{26} V_0 e^{j\omega t}}{c^{(1)} 2h} x_2;$$

$$\phi = \hat{\phi} + \frac{x_2 V_0 e^{j\omega t}}{2h};$$

$$u_1 = \sum_n u_1^n$$

$$\acute{c}^{(1)} = \bar{c}^{(1)}(1 - 8k^{(1)2}/n^2\pi^2)$$

$\acute{c}^{(1)}$ is a constant that includes the electrical effect of the metallization. $4k^{(1)2}/n^2\pi^2$ is approximatively the relative frequency lowering due to the electrical effect of the metallization $c^{(1)} = \bar{c}^{(1)}(1 - k^{(1)})^2$ .is a pseudo ordinary elastic constant. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode. M , P are intricate functions of the material constants and of the plate orientation that can be obtained from the method described in reference (3) or in reference (8).

- Boundary and continuity conditions : The traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish the dispersion relation. On the surfaces limiting the electroded and the unelectroded regions of the resonator we have to specify the continuity of u_1 and of its normal derivative.

In this paper we suppose that the amplitude of the vibration mode near the edges of the plate are negligible so that no boundary condition at the plate edge are taken into account. It was previously shown that, in this case, they induce a negligible modification of the resonances frequencies (9)(10).

Eigen mode analysis : The eigen modes of the resonator can be obtained following the simultaneous solution of the equations for the electroded and unelectroded region taking into account the continuity conditions at the electrode edge. At V=0 the homogeneous form of the approximate equation for the two regions reduces to :

$$M'_n u_1^n, 11 + P'_n u_1^n, 33 + 4\pi^2 \rho (f^2 - f'^2) u_1^n = 0 \qquad [2]$$

with :

$$f' = f_{ce1} = \frac{n}{4h}\sqrt{\frac{\acute{c}^{(1)}}{\rho}}$$

for the electroded region.

or:

$$f' = f_{c1} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$$

for the unelectroded region.

The coordinate transformations :

$$X_1 = \left(\frac{C^*}{M'_n}\right)^{1/2} \cdot \gamma x_1 \qquad X_3 = \left(\frac{C^*}{P'_n}\right)^{1/2} \cdot \gamma x_3$$

$$X_1 = r\cos t \qquad X_3 = r\sin t \qquad [3]$$

with: $\gamma^4 = \frac{n^2\pi^2}{2h^4}$

$c^* = \acute{c}^{(1)}$ (electroded region)

$c^* = \bar{c}^{(1)}$ (unelectroded region)

leads to the equation :

$$u_{1,rr} + \frac{1}{r}u_{1,r} + \frac{1}{r^2}u_{1,tt} + A u_1 = 0 \qquad [4]$$

$$Where: A = \frac{n\pi}{4}\left(\frac{f^2 - f'^2}{f'^2}\right)$$

that can be separated as $u_1(r,t) = R(r).T(t)$ in :

$$r^2 R'' + rR' + R(r^2 A - \upsilon^2) = 0$$

$$T'' + \upsilon^2 T = 0$$

For the electroded region the symmetrical solution bounded at r = 0 can be expressed as (6) :

$$u_1 = \sum_{m=0}^{\infty} A_m J_m(r\sqrt{A})\cos mt \qquad [6]$$

with m = v even integer

For the unelectroded region the symmetrical solution bounded at r =∞ is (since A is then negative)

$$u_1 = \sum_{m=0}^{\infty} B_m K_m(r\sqrt{A'})\cos mt \qquad [7]$$

with: m = υ even integer and A' = -A

Jm and Km are respectively the Bessel function of first kind and the modified Bessel function of second kind, of order m.

At all points of the electrode edge we have to specify the continuity of u_1 and of its normal derivative. As previously discussed (6)(11)(12), this can be approximated by the expression of these conditions at only a discrete number q_1 of points (figure 4). It the solutions for the two regions are :

$$u_1^I = \left(\sum_{m=0} A_m J_n(r\sqrt{A})\cos mt \sin(n\pi x_2/2h)\right.$$

$$u_1^{II} = \left(\sum_{m=0} B_m K_m(r\sqrt{A'})\cos mt \sin(n\pi x_2/2h)\right. \qquad [8]$$

at the point $P(x_1^P, x_2^P, x_3^P)$ we have :

$$u_1^I(x_1^P, x_2^P, x_3^P) = u_1^{II}(x_1^P, x_2^P, x_3^P)$$

$$n_1\frac{\partial u_1^I(P)}{\partial x_1} + n_3\frac{\partial u_1^I(P)}{\partial x_3} = n_1\frac{\partial u^{II}{}_1(P)}{\partial x_1} + n_3\frac{\partial u_1^{II}(P)}{\partial x_3} \qquad [9]$$

If we choose to express these conditions at q points at the electrode edge, we obtain 2q linear relations between the coefficients A_0, A_2, A_{2q-2},..., B_0, B_2, B_{2q-2}, to obtain a solution we have to truncate the series to q terms so that the 2q relations constitute an homogeneous linear system in the 2q coefficients. The determinant of this system has to vanish to have a non trivial (0) solution in the coefficients. This condition constitutes a frequency equation that is solved numerically for the eigen frequencies. For each eigen frequency the coefficients of the eigen modes are found as the solution of the homogeneous linear system. The forced modes can then be obtained as described in reference [6]. In figure 4 are represented the computed modes of two air gap resonators (with a zero gap) for an AT quartz plate 23 micron thick. On figure 4a we have the case of a small electrode with 2r=20.2h; it can be observed that the lateral extension of the displacement (the iso-level curves are spaced by one tenth of the maximum amplitude) is very important outside the electrodes. It may be noticed that the lateral anisotropy of the mode is also very important. On figure 4b the computed zero-gap mode of a resonator with 2r/2h=60 is represented at the same scale.

In this case, the total lateral extension of the mode is only a little more important than in the previous one, so that most of the energy is confined under the electrodes. The rate of decrease of the displacement outside the electrodes is also much faster in this case, so that it can be expected that the measurements made with this kind of electrodes will be less sensitive to small defects of parallelism of the plates situated outside the electrodes.

EXAMPLE OF MODES OF ZERO GAP RESONATORS AT QUARTZ, 2h= 23 10^{-3} mm, 72MHz Fundamental

SMALL SIZED ELECTRODES 2r ≈20.2h

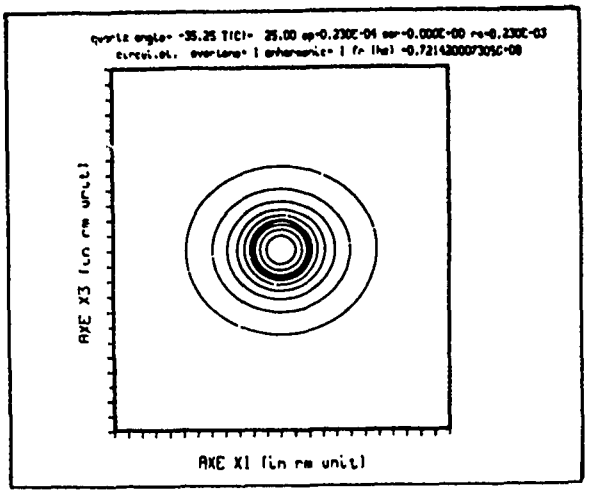


Figure 4a: Mode with 2r/2h=20

MEDIUM SIZED ELECTRODES 2r ≈ 60.2h

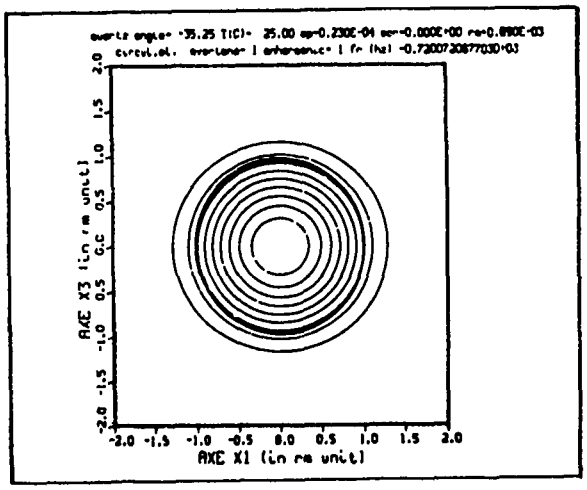


Figure 4b: Mode with 2r/2h=60

SAME SCALE FOR ALL FIGURES
-60.2h< X_1 <+60.2h ; -60.2h< X_3 <+60.2h

Figure 4 : Computed zero-gap modes

II.3 - Application of the model to determine f_{cl}

As previously indicated, the airgap frequencies are measured as a function of the gap and then numerically extrapolated to a zero gap. The zero gap frequency is considered as resonance frequency of a resonator having a zero mass electrode of diameter d = 2r. By an iterative process an apparent (or corrected) thickness of the resonator is numerically searched, using the described model, so that the computed resonance frequency corresponds to the experimental one. The f_{ce1} and f_{cl} frequency are then determined.

Since the resonators are designed using similar models making use of the f_{ce1} and f_{cl} frequencies (determination of the electrode radius and of the mass loading to have a spurious free response and a given value of inductance), this global process eliminates a large part of the modelization approximations and of experimental uncertitudes.

II.4 - Experimental and computed results

This method was applied to quartz and berlinite plates with thickness ranging from about 200 to 20 microns that were obtained by conventional polishing techniques.

The experimental set-up is represented on *figure 5*. An HP 8753 network analyser is used for the electrical measurements; on *figure 6 and 7* are represented the relative variations of the resonance frequencies measured as a function of the gap for a 51 MHz and a 71 MHs resonator ; the corresponding zero gap frequencies are given on the figures.

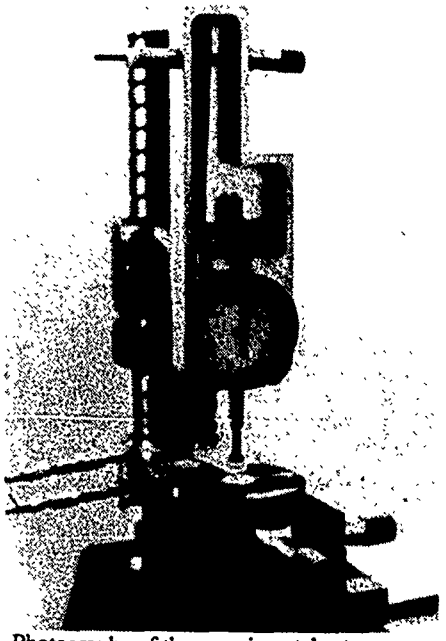


Figure 5 : Photography of the experimental set up

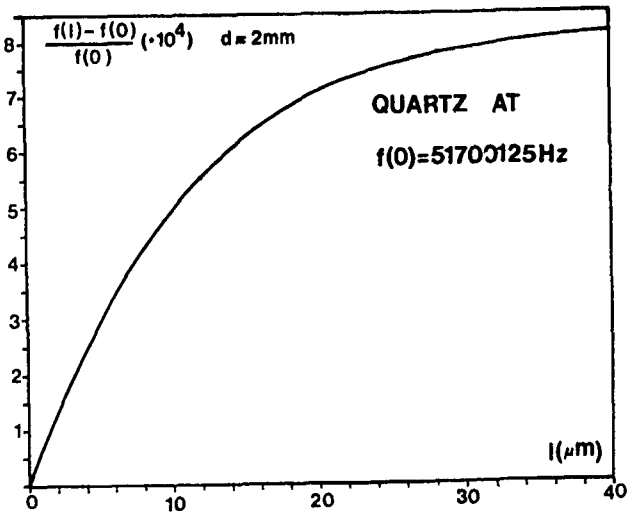


Figure 6 : Relative variations of the air gap frequency with l (2 mm probe, 2 h = 32 μ)

The frequency response of a 71 MHz resonator measured with a probe having a diameter of 2.000 mm and a gap of a few microns is given on *figure 8*. It can be observed in this case that with this large electrode the trapping is not sufficient to have no unwanted modes (the spurious modes observed near the main resonance result most probably of a coupling at the edge of the plate of the thickness mode to plate modes) . In *figure 9* is given the frequency response of a similar plate with a probe having a diameter of 1,50 mm ; we have now a "clean" response. This illustrates the importance of the choice of the probe diameter that must be such that the vibration mode have a negligible amplitude at the edge of the plate.

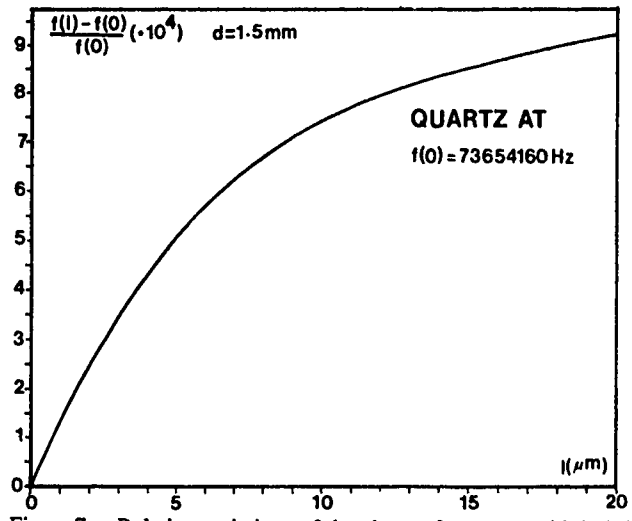


Figure 7 : Relative variations of the air gap frequency with l (1,5 mm probe, 2 h = 22,5µm)

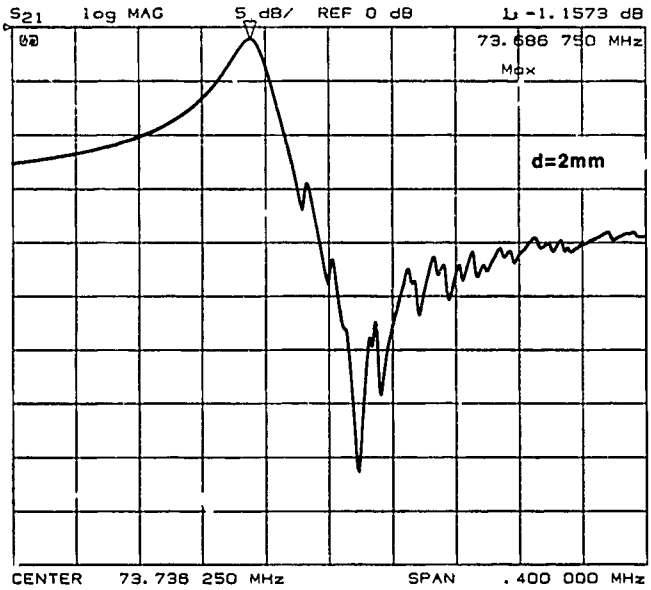


Figure 8 : Response of a 73 MHz resonator with a 2 mm probe

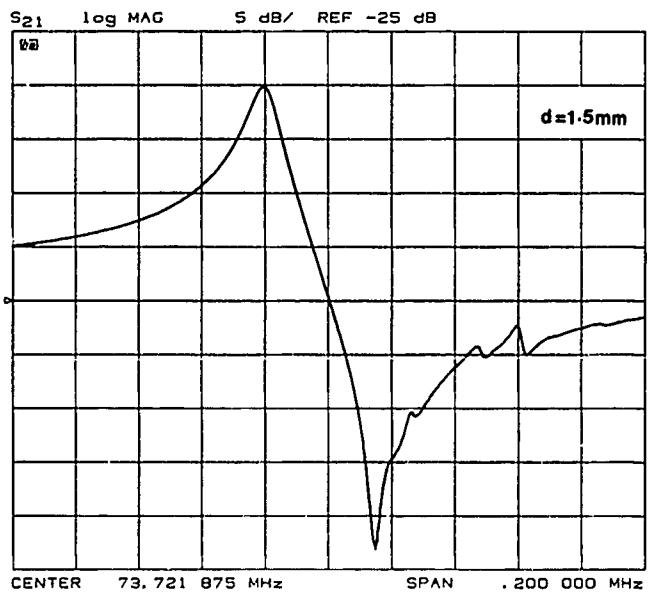


Figure 9 : Response of the same resonator with a 1,5 mm probe

With this experimental set up it was possible to make measurements at much higher frequencies, the 3rd overtone of a 71 MHz plate is displayed in *figure 10*. Due to the very small value of the reactance of the static capacitance at very high frequency, reflection measurements (S_{11}) are to be prefered above say 200 MHz. The S_{11} frequency response of the fifth overtone of a 71 MHz plate is given on *figure 11*.

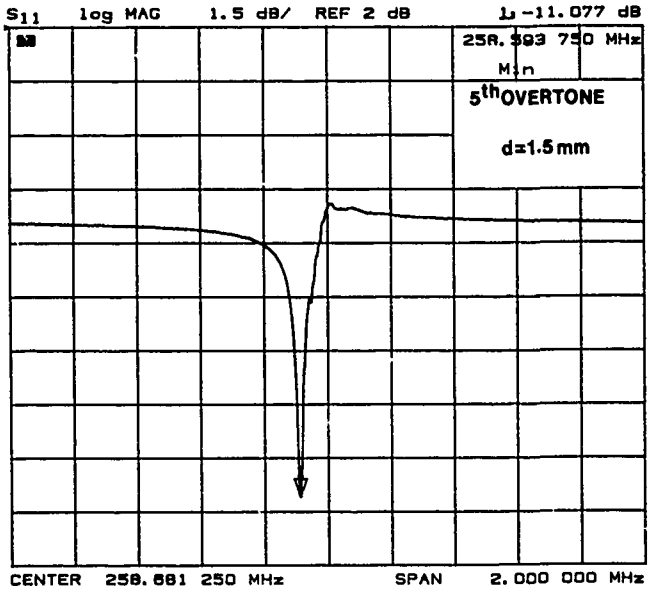


Figure 10 : Air gap response of the 5rd harmonic of a 51 MHz plate (S_1 measurement)

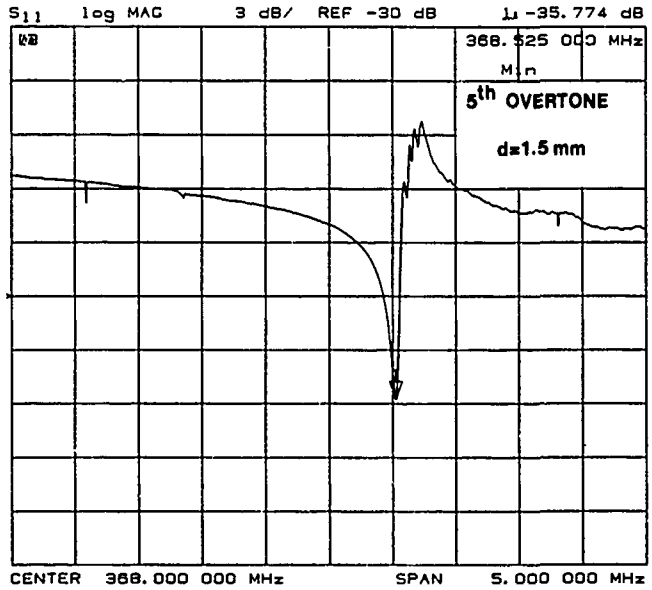


Figure 11 : Air gap response of the 5th harmonic of a 71 MHz plate (S_{11} measurement)

Typical exemples of application of this method are displayed *in table 1 and 2* were are given the thickness of the plate measured by a conventional method, the zerogap frequencies and the extracted values of f_{c1}, f_{c1} and of the computed "apparent" thickness (relative to the values of the material constants used in the computation).

TABLE 1 : AIR GAP MEASUREMENTS

RESULTS OBTAINED FROM MEASUREMENTS WITH r = 1 mm
AT QUARTZ EXEPT SAMPLE N° 5 (Y+3° BERLINITE)

$2h_{measur}$	$f_{airgap=0}$	f_{ce1}	f_{c1}	$2h_{calc.}$
32.0	51569793.	51662672.	51825912.	32.0579
80.9	20466933.	20424233.	20488768.	81.0899
81.1	20414451.	20371758.	20436127.	81.2988
87.5	18977731.	18935085.	18994914.	87.4672
187.2	9244542.	9195947.	9252228.	186.68

TABLE 2 : AIRGAP MEASUREMENTS

RESULTS OBTAINED FROM MEASUREMENTS WITH r = 0,75 mm

$2h_{measur}$	$f_{airgap=0}$	f_{ce1}	f_{c1}	$2h_{calc.}$
31,9 F	51799083.	51750620.	51914138	32.0034
31,9 P3	155453330	155439444	155493787	32.0546
22,5 F	73654160	73614188	73846788	22,4983

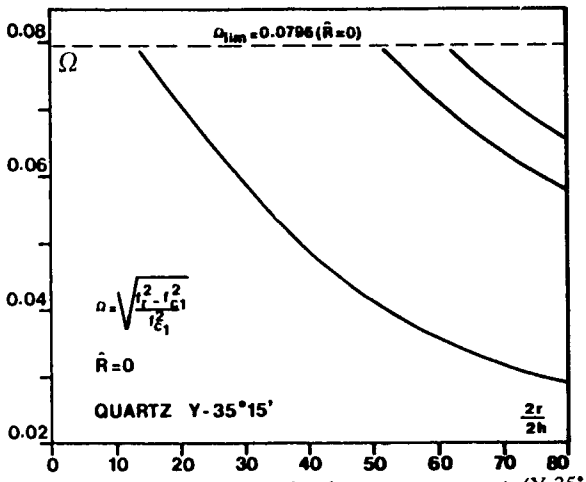


Figure 12 : Correction diagram for air gap measurements (Y-35°15' Quartz)

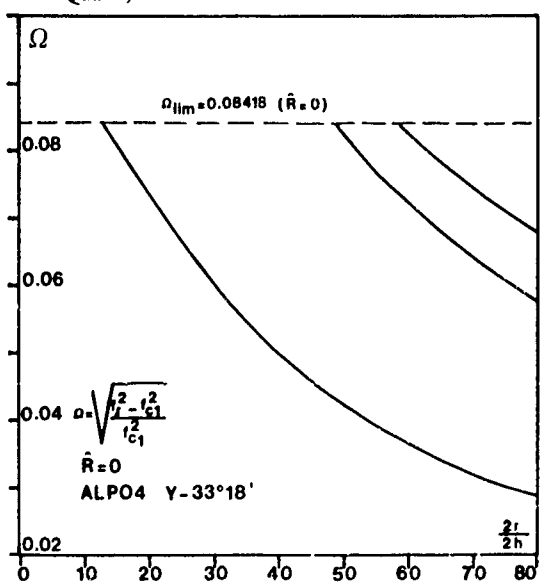


Figure 13 : Correction diagram for airgap measurements (Y-33°18' Berlinite)

To avoïd the calculation process made to extract the values of f_{ce1} and f_{c1} in the cases where only a moderate precision is required, abaci giving in reduced coordinates the dependence of the zero gap resonance frequency upon the electrode diameter were calculated. From the reduced frequency $\Omega^2 = (f_r^2 - f_{ce1}^2)/f_{ce1}^2$ for a given 2r/2h ratio, the f_{ce1} frequency can be calculated from the measured value f_r ; the f_{c1} frequency can then be determined knowing the value of $k^{(1)2}$. They are given on figure 12 (AT cut of quartz) and on figure 13 (AT cut of Berlinite).

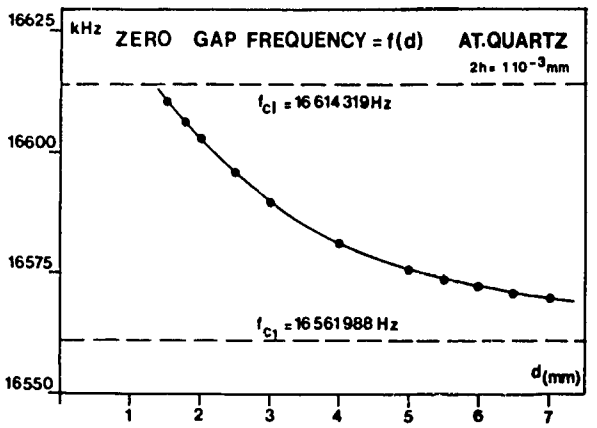


Figure 14 : Resonance frequency (l=0) of an AT Quartz (2h = 0,100 mm)

To illustrate the dependence upon the electrode dimension of the resonance air gap frequency in a simpler manner we have plotted on figure 14 the zero gap frequencies for a plate of AT quartz, 0.1 mm thick as a function of the electrode diameter.

III - Resonators with electrodes embedded in the plate

This type of resonator was proposed by Lukaszek (4) to control the unwanted responses and the impedance level of VHF resonators for filter applications. As shown in figure 15 the electrodes are deposited in an etched region of the plate. These resonators are of the energy trapping type if the cut-off frequency of the electroded central region (f_{ce1}) is lower than f_{c1}. These frequencies differ by a quantity that depends upon the electrode thickness, the depth of the grooved region and upon the electrical frequency lowering. If we define a new apparent mass loading $\hat{R}'$ by :

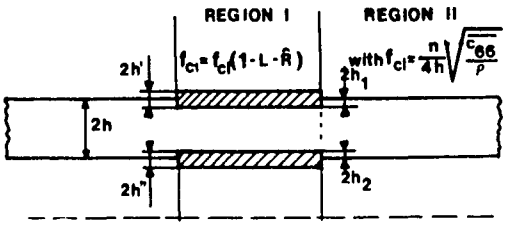


Figure 15 : Resonator with embedded electrodes

$$f_{ce1} = f_{cl}\left(1 - \frac{4k^2}{n^2\pi^2} - \hat{R}'\right)$$

$$\hat{R}' \# \frac{\rho_m(2h' + 2h'') - (2h_1 + 2h_2)\rho_q}{\rho_q . 2h}$$

where :

h, h', h", h_1, h_2 are defined on figure 15. ρ_m and ρ_q are the densities of electrodes and quartz.

The theory of paragraph II can be applied to these resonators if $\left(\frac{4k^2}{n^2\pi^2} + \hat{R}'\right)ts > 0$ to ensure that the trapping of the energy occurs with this new value of f_{ce1}).

Several calculations were made for the case of 72 MHz resonators using quartz and berlinite (the thickness is about 23 µm for AT quartz and 20 µm for AT Berlinite). We have assumed that the interesting values of $\hat{R}'$ were between - 1.10^{-3} and 8.10^{-3}. For a total thickness of aluminium electrodes of about 2 x 1000 Å, this corresponds to an etching ranging from about 2 x 1100 Å to nearly 2 x 200 Å. The results are given in the form of a diagram giving for AT Quartz, the reduced resonance frequency Ω as a function of 2r/2h. On this diagram we can observe (figure 16) that, for a given value of $\hat{R}'$, the trapped modes can exist between $\Omega = 0$ and a limit Ω_{lim} which is a fonction of $\hat{R}'$.

$$\Omega_{lim} \# 8k^2/n^2\pi^2 + 2\hat{R}'$$

The principal effect of reducing $\hat{R}'$ is to reduce the number of modes that can exist, so that it is possible to obtain resonators with no anharmonics for much larger values of 2r/2h than with conventional values of the mass loading.

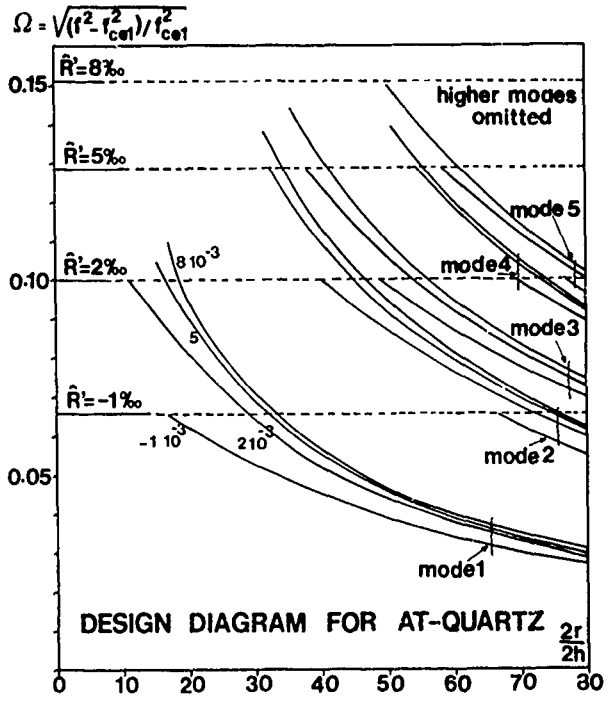


Figure 16 : Diagram for the design of resonators with embedded electrodes

The effect on the computed values of the inductance (6) is displayed on Figure 17 and in table 3. For this computation a thickness of 23,015 µ was used and the values of the electrode diameter were choosen to obtain round values of 2r/2h. Two important facts can be observed. First, as expected, the inductance decreases when 2r is increased, but, not, in following a linear low. The second fact is relative to the influence of $\hat{R}'$, for the low values of 2r/2h, the inductances decreases strongly when $\hat{R}'$ is increased, whereas, for the higher values of 2r/2h, the inductance is quasi independant of $\hat{R}'$. The combination of figure 16 and 17 permits to find a solution to the problem of designing filter resonators with no anharmonics and a low value of the inductance. However, in practice, the possibility of using small values of $\hat{R}'$ is limited by an eventual lack of parallelism of the plates ; so that this parameter must be also taken in account.

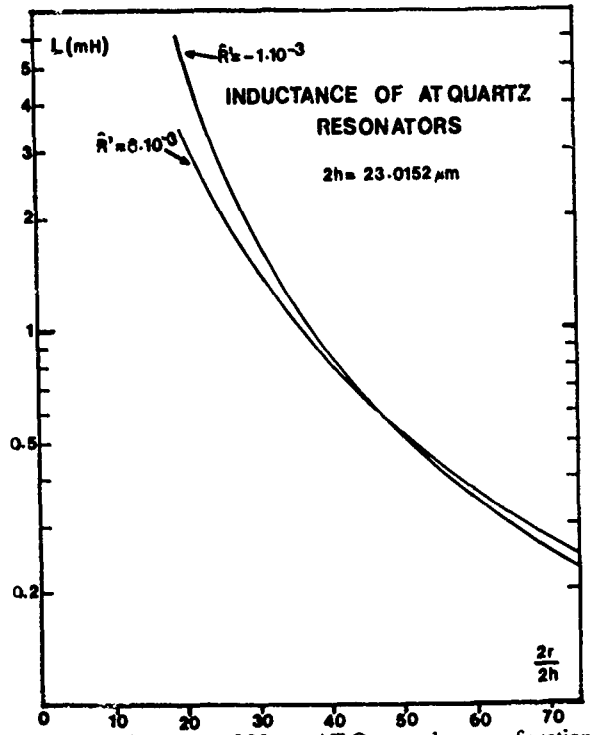


Figure 17 : Inductances of 23 µm AT Quartz plate as a function of $\hat{R}'$ and 2r/2h

TABLE 3 : INDUCTANCE

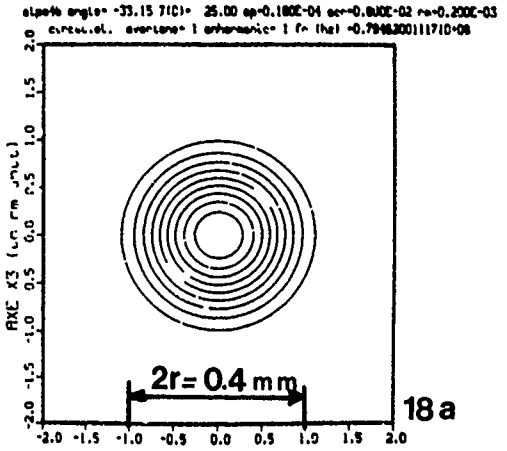
COMPUTED VALUES OF L for 72 MHz AT QUARTZ
2 h = 23,01524 µm - Values in mH

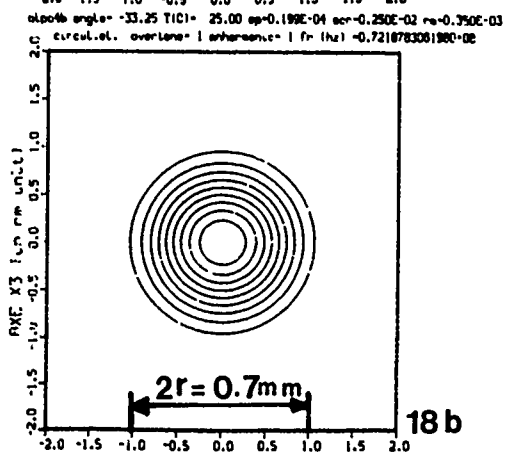
$\hat{R}'$ / 2r/2h	-1.10^{-3}	2.10^{-3}	5.10^{-3}	8.10^{-3}
20	5.507*	3.891*	3.492*	3.3349*
30	1,761*	1,471*	1.430*	1.430*
40	0,8598*	0,7985	0,8025	0,8135
50	0,5206*	0,5110	0,5205	0,5304
60	0,3539*	0,3582	0,3671	0,3748
80	0,1988	0,2065	0,2124	-

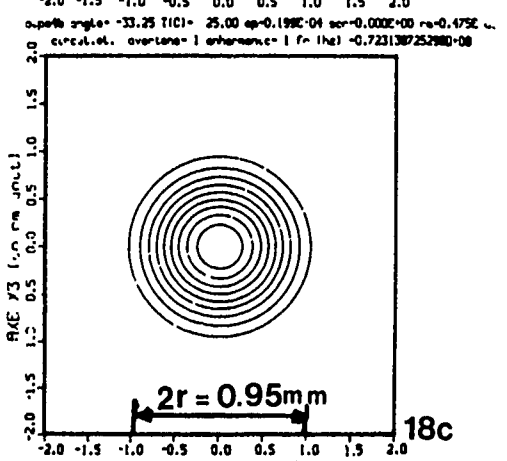
(*) monomode response

Another important parameter to consider, since it determines the required lateral dimensions of the plate, is the lateral extension of the mode. This parameter is closely related to the value of the inductance since this quantity is proportional to the integral of u_1^2 on the surface of the resonator divided by the square of the integral of u_1 on the electroded surface. For a given 2 r/2h ratio, the value of L indicates roughly the fraction of the energy which is not confined under the electrodes, but it is always more accurate to compute the vibration mode $u(x_1, x_3)$.

Examples of modes computed for AT Berlinite resonators near 70 MHz are given an Figure 18. On figure 18a we have the mode of a conventional resonator with $\hat{R}' = 1$ % design to thave nearly the minimum inductance together with no anharmonics (this has lead to use an electrode diameter of 0,40 mm). On figure 18b the value of $\hat{R}'$ was choosen to be 0,25 %, so that it is now possible to have an electrode diameter of 0,70 mm without any anharmonic mode. On Figure 18c a value of $\hat{R}' = 0$ is choosen, it is now possible to use an electrode diameter of 0,95 mm. With the same value of $\hat{R}'$ it is still possible to have a lower inductance (a greater electrode surface) using an other geometry. This is demonstrated in Figure 18d where an elliptical electrode permits to have an electroded surface equal to that given by a 1 mm electrode without any anharmonics (6).







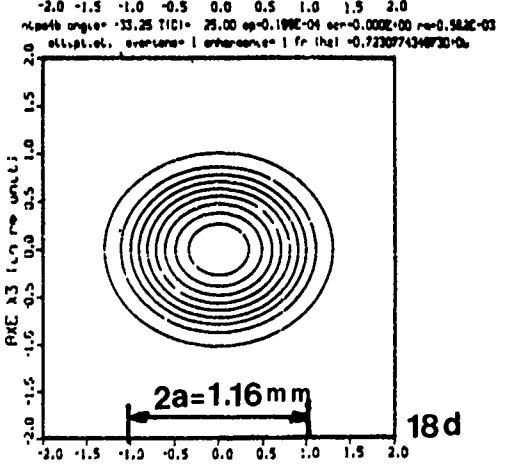
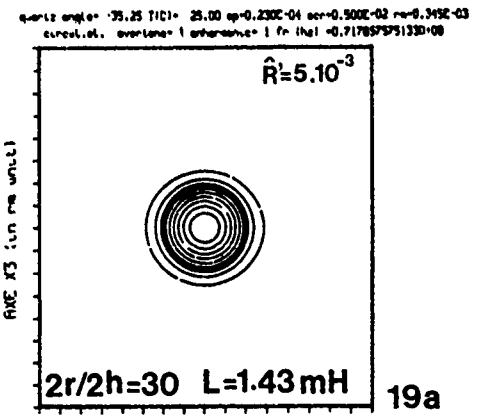
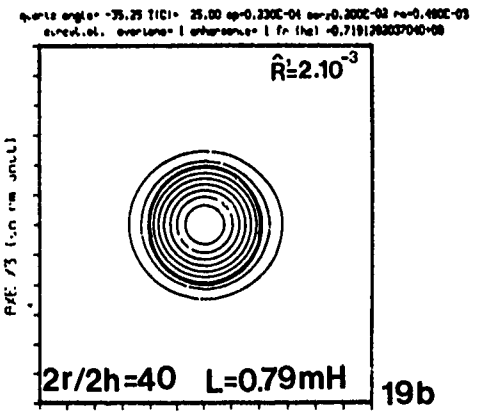


Figure 18 : Computed vibration mode of 72 MHz berlinite resonators

Other examples of computed modes concerning 72 MHz (fundamental) quartz resonators designed so to have no anharmonics (filter application) are given in figure 19 where they are drawn at the same scale. In figure 19a is represented the mode of a quite conventional resonator with an apparent mass loading of .5% and 2r/2h=30; the inductance is then equal to 1.43mH. Decreasing the apparent mass loading to .2% by grooving under the electrodes permits to have a monomode response with 2r/2h=40 and to reduce the inductance nearly by a factor of two (.79mH) (Figure 19b). A further decrease of the apparent mass loading to -0.1% permits to obtain a monomode response with 2r/2h=60 and an inductance of .39 mH (Figure 19c). These examples illustrate the interest of this type of resonators, however, for very small apparent mass loadings, the sensitivity to defects of parallelism of the plates can become too important for practical applications. Then the solution can be to use the design described in paragraph IV that permits to enhance the rate of decrease of the displacement outside the electrodes.





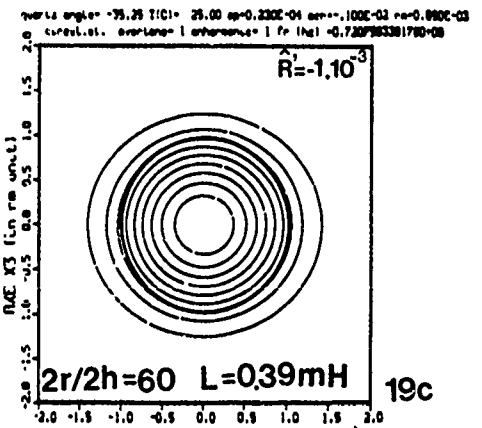


Figure 19: Computed modes of 72MHz AT Quartz resonators.

IV - Theory of resonators having 3 regions with different thickness

...nted on *figure 20* we consider a resonator that have 3 ...h different thickness, the central region noted 1 being electroded.

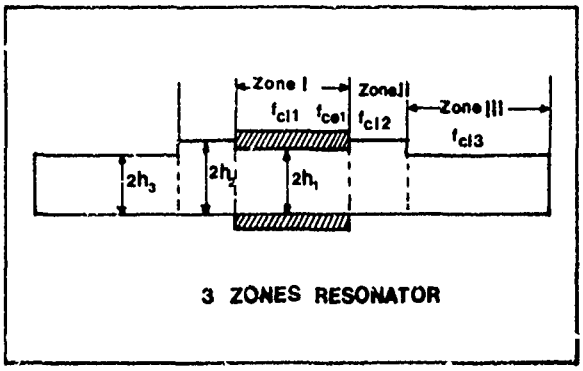


Figure 20 : Resonator with 3 regions of different thickness

In each region the lateral dependance of u_1 is governed by the approximate equation and at each boundary separating 2 regions we have to express the continuity of the displacement and of its normal derivative we also suppose that the lateral dimension of region III are such that the trapped mode considered has a negligible amplitude at the edge of the plate. We suppose that the difference of thickness between two regions is small so that no boundary conditions are to be expressed on the small step existing there.

Eigen mode Analysis : For each region we have to solve the homogeneous Helmoltz equation [2]. In each region the quantity A* which carry the frequency dependance has a different expression

$$A^* = \frac{n\pi}{4}\left(\frac{f^2 - f^{*2}}{f^{*2}}\right)$$

f^* being either f_{ce1}, f_{cl2}, f_{cl3}

These cut-off frequencies being defined in regions 1, 2 and 3 by the same formula as f_{cl1} and f_{c1} in paragraph II, in function of the thickness, the mass loading, the electrical lowering, and the appropriate constant C*. To have trapped modes, we have also to suppose that f_{cl2} has the highest value among (f_{cl1}, f_{cl2}, f_{cl3}). The same method of resolution by separation given in paragraph II is used then for the three Helmoltz equations relative to the 3 regions.

The solution for region 1 is now either

$$u_1^I = \sum_m A_m J_m(r\sqrt{A})\cos mt \; if A > 0$$

or

$$u_1^I = \sum_m A_m I_m(r\sqrt{|A|})\cos mt \; if A < 0$$

The solution for region 2 is :

$$u_1^{II} = \sum B_m F_m(r\sqrt{|A|})\cos mt + \sum_m C_m G_m(r\sqrt{|A|}\cos mt$$

where F_m and G_m are either J_m and Y_m or I_m and K_m depending on the sign of A.

The solution for region 3 is :

$$u_1^{III} = \sum_m D_m K_m(r\sqrt{|A|})\cos mt$$

The continuity conditions relative to the continuity for u_1 and of its normal derivative at the boundary of regions 1 and 2 and of regions 2 and 3 are, as in paragraph 2, expressed at a discrete number of points (p). Again this conduct to truncate the series to p terms so that we have 4 p linear relations between the coefficients A_i, B_i, C_i and D_i in the form :

$$\begin{cases} \alpha A = \beta B + \gamma C \\ \alpha' A = \beta' B + \gamma' C \end{cases}$$

$$\begin{cases} \beta'' B + \gamma'' C = \delta D \\ \beta''' B + \gamma''' C = \delta' D \end{cases}$$

where : A, B, C, D are vectors containing each the p coefficients of the tuncated series $\alpha, \beta, \gamma, \ldots, \delta'$ are pxp matrix

From these relations it is possible to obtain the homogeneous system

$$\begin{pmatrix} \alpha & -\varepsilon \\ \alpha' & -\varepsilon' \end{pmatrix}\begin{pmatrix} A \\ D \end{pmatrix} = 0 with \begin{pmatrix} \varepsilon \\ \varepsilon' \end{pmatrix} = \begin{pmatrix} \beta & \gamma \\ \beta' & \gamma' \end{pmatrix}\begin{pmatrix} \beta'' & \gamma'' \\ \beta''' & \gamma''' \end{pmatrix}^{-1}\begin{pmatrix} \delta \\ \delta' \end{pmatrix}$$

The determinant of this system must vanish to have a solution in the A and D coefficients. This condition constitutes a frequency equation which is solved numerically to obtain the eigen frequencies. For each of the eigen frequencies, the corresponding eigen mode is obtained as the solution of the homogeneous linear system . This leads to the A and D coefficients. The B and C coefficients are obtained by back substitution in one of the following relations.

$$\begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} \beta'' & \gamma'' \\ \beta''' & \gamma''' \end{pmatrix}^{-1}\begin{pmatrix} \delta \\ \delta' \end{pmatrix}\begin{pmatrix} D \\ D \end{pmatrix}$$

$$\begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} \beta & \gamma \\ \beta' & \gamma' \end{pmatrix}^{-1}\begin{pmatrix} \alpha \\ \alpha' \end{pmatrix}\begin{pmatrix} D \\ D \end{pmatrix}$$

V FINITE ELEMENTS SOLUTION AND COMPARISON TO EXPERIMENTAL RESULTS:

The homogeneous form of the approximate equation is solved by a finite element technique (14)(15), with in each region of the resonator, the appropriate values of 2h, R , and C* . The natural boundary condition implies the normal derivative of u_1 equal to zero, so that the continuity relations between the different regions are automatically satisfied and the conditions at the edge of the plate is of Neuman type. This latter one can also be made of Dirichlet type. The domain which comprises the different circular and annular regions is divided into triangular elements with the displacement at the nodes as unknown (100X4 to 400X4 nodes per resonator). The problem is solved as an eigen value one. The matrix resulting from the assemblage of these elements are stored on a "skyline" basis. The inductance and the dynamic capacitance can also be calculated by the finite element program implemented at C.N.E.T.

The overal accuracy can be established by considering the variations of the frequency when one adds a few elements. The size of the external region has to be kept to a sufficient value, and the domain has to be enough subdivised to give a fixed accuracy.

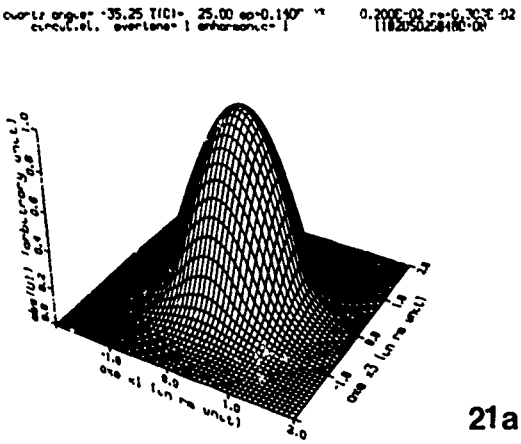
On table 4 are given examples of results obtained by the finite element method and the semialgebraical method concerning resonators having 2 or 3 zones. The first one is relative to an AT quartz with grooved electrodes 2h=0.140mm, 2r=6mm, R'=0.2%; the second concerns also a similar AT quartz with 3 zones of different thickness 2h1=0.13888mm, 2h2=0.140mm, 2h3=0.13888mm, 2r1=6mm, 2r2=8mm, k=1%.

TABLE 4: EXAMPLES OF COMPUTED RESULTS

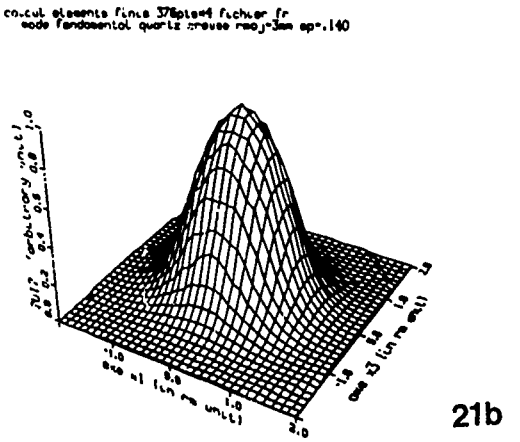
	2 ZONES	2 ZONES	3 ZONES
	Semi-algebraical	F.E.M.	F.E.M.
Fr (kHz)	11820.	11816.	11818.
L (mH)	4.4	4.2	4.2

The computed frequencies of table 4 are nearly identical ; a very small difference (5%) exists between the computed inductances found by the two calculation methods. The most important constatation is that it possible to have, with a resonator with three zones, electrical properties very similar to those of a resonator with a grooved electrode. The corresponding computed modes are represented in figures 21 and 22.

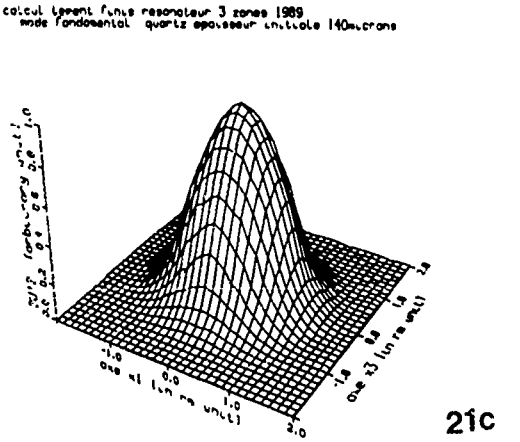
The mode represented in figures 21a and 22a corresponds to the 2 zones resonator, it was computed assuming that the plate has infinite lateral dimensions (Paragraph 3). The mode represented in figures 21b and 22b was computed using the F.E.M. for a circular plate with a finite radius of 6.1mm and natural (Neumann) boundary conditions at the edge of the plate. The two results are very similar except for a slightly different lateral anisotropy that explains the small difference of the inductances in table 4.



21a



21b

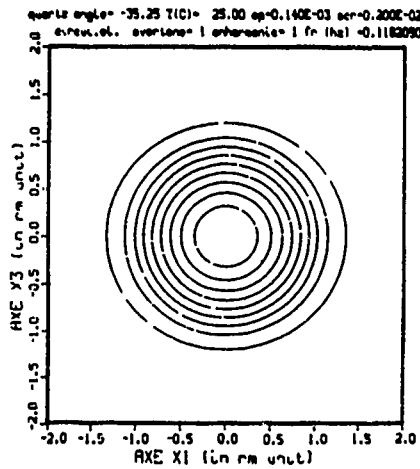


21c

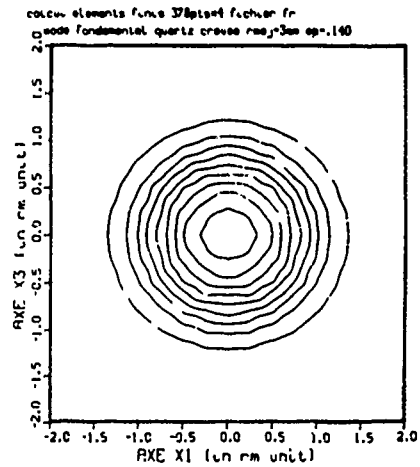
Figure 21 : Computed modes of resonators with 2 and 3 zones.

In figures 21c and 22c is represented the mode of the resonator with 3 zones computed by the finite element method assuming a circular plate plate with a finite radius of 6.1mm and Neumann conditions at the edge. A careful comparison with the preceeding case would reveal a faster decrease of the amplitude of the displacement in the external region of the resonators. This is one of the interest of using such a design.

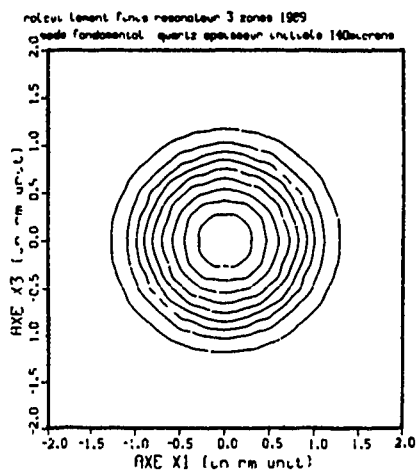
In figure 23 are represented the observed mode of a resonator with grooved electrodes which has design parameters very similar to those used for the computations of table 4. The observed mode is very close to the computed one with however a slightly faster rate of decrase of the displacement outside the electroded region.



22a



22b



22c

Figure 22 : Computed modes of resonators with 2 and 3 zones.

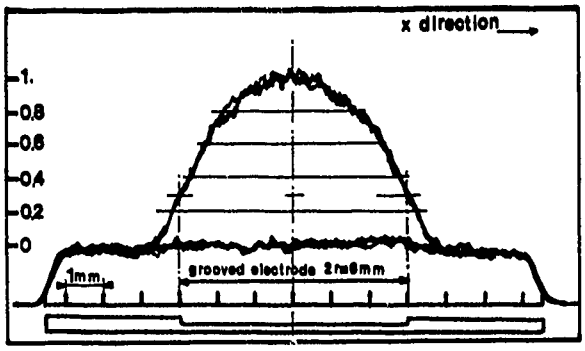


Figure 23 : Experimental mode of a resonator with 2 zones.

VI - CONCLUSION

A method is proposed to obtain precise measurements of the cut-off frequencies of unelectroded plates. These quantities can then be used in a model, similar to those which has permitted the extraction of the cut-off frequency, to determine the other parameters of the resonator (electrode geometry and mass loading). Two sucessive applications of these models permit to remove most of the uncertainties relative to the material constants and to obtain precise values of the equivalent scheme and, if necessary, a response free from anharmonic modes.

A modelization of the resonators with embedded electrodes proposed, by Lukaszek has been made, and, the properties of this type of resonators computed. It appears that they are an efficient solution to obtain filter or VCXO resonators with more favourable values of the equivalent scheme.

A modelization of resonators having 3 regions with different thicknesses was made and is now being used to find designs that permit to have the advantages of the resonators with grooved electrodes together with a much reduced sensitivity to the defects of parallelism of the plates and the possibility of reducing the lateral dimensions of the resonators.

The models made for the three type of resonators described in this paper are based upon the approximate equations governing the thickness vibrations of piezoelectric plates established by H.F.Tiersten and co-workers. Two methods of resolution of these equations were used: a semialgebraical one and the finite elements method.

We expect that the designs found using the model of the grooved resonator or of the 3 regions resonators will permit to have much better solutions for VHF filters and VCXOs and particularly for the intermediate frequency filter of the new numerical radiotelephone system.

Aknowledgments : The authors whish to acknowledge R. Lefèvre for kindly lending the airgap apparatus of figure 5. They are indebted to MM. Chenebault, Douis and Mrs Daniel for the elaboration of the many samples used during this study, they thank particularly Mrs M.P. Louis for the preparation of this manuscript.

REFERENCES

[1] H.F. Tiersten
J. Acoust. Soc. Am vol 59 p 879 (1976)

[2 H.F. Tiersten, R.C. Smythe
J. Acoust. Soc. Am vol. 65 n° 6 p 1455 (1979)

[3] D.S. Stevens, H.F. Tiersten
J. Acoust. Soc. Am vol. 79 n° 6 p 1811 (1986)

[4] T.J. Lukaszek
IEEE trans. Sonics vol. S.U. 16 n° 4 p 238 (1971)

[5] J. Détaint, A. Zarka, B. Capelle, Y. Toudic, J. Schwartzel, E. Philippot, J.C. Jumas, A. Goiffon, J.C. Doukhan
Proc. 40th Annual Frequency Control Symp. p 101 (1986)

[6] J. Détaint, J. Schwartzel, C. Joly, E. Philippot
Proc. 41st Annual Frequency Control Symp. p 314 (1987)

[7] H.F. Tiersten, R.C. Smythe
Proc. 41st Annual Frequency Control Symp. p 230 (1981)

[8] R.C. Peach
Proc. 42nd Annual Frequency Control Symp. p 38 (1988)

[9] Dulmet, H.F. Tiersten
Proc. 1986 Ultrasoncis Symp. p 355 (1986)

[10] J. Détaint, H. Carru, J. Schwartzel, B. Capelle, A. Zarka
Proc. 42nd Annual Frequency Control Symposium p 19 (1988)

[11] K. Nakamura, M. Shimizu
Proc. 1976 Ultrasonics Symposium p 606 (1976)

[12] H. Seikimoto
IEEE trans. Sonics Vol. SU 31 n° 6 p 664 (1984)

[13] H. Carru, R. Lefèvre
Proc. 42nd Annual Freq. Control Symp p 101 (1988)

[14] O.C. Zienkiewiz
The Finite Element Method Mc Graw Hill (1977)

[15] K.J. Bathe, E.L. Wilson
Numerical Methods in Finite Element Analysis
Prentice Hall (1976)

LOW-COST HIGH-PERFORMANCE RESONATOR AND COUPLED-RESONATOR DESIGN: NSPUDT AND OTHER INNOVATIVE STRUCTURES*

P.V. Wright†
RF Monolithics
4441 Sigma Road
Dallas, TX 75244

Abstract

A design philosophy for low-cost, high-volume SAW resonators and coupled resonators is presented. In view of the low-cost constraint, only configurations requiring a single-level of metallization are considered. To illustrate the design concepts involved, examples are given of alternate resonator topologies and their important performance characteristics discussed. Design emphasis is placed on resonators having good out-of-band characteristics. The latter is particularly important if the resonator is to be used as a building block for a coupled-resonator filter. For such an application, a symmetric resonator response is especially desirable. Two novel resonator structures that have been developed with essentially symmetric out-of-band responses are described. Both employ only single-level metallization and no grooves. The fundamental difference between the two approaches is that the first structure requires Natural Single-Phase Unidirectional Transducer (NSPUDT) orientations, while the second can be built on conventional crystal orientations. The NSPUDT resonator is a fascinating example of the properties of these interesting crystal orientations which were originally investigated for SAW filter applications.

1 Introduction

SAW resonators are used extensively in frequency-stable sources in the typical frequency range of 100 MHz to 2 GHz. Among their principal advantages are that they are small, very stable, and can be fabricated exceedingly cheaply in large volumes using techniques borrowed from the semiconductor industry. The market for SAW resonators can roughly be divided into two; that for commercial applications which consumes very large volumes at a low unit cost, and that for military or specialist applications which consumes many fewer devices, but at a much higher unit cost. The design requirements for these two markets are typically very different. For example, in the latter to achieve optimum performance, individual device tuning is frequently performed and special high-cost packaging can be employed [1]. However, the cost constraints of the commercial market do not permit such "luxuries". This paper considers primarily the design of commercial resonators.

In order for commercial resonators to be mass-producible at a low cost, a very important requirement is that they require only a single-level of metallization. In addition, the response of the resonator must not be overly sensitive to production variables such as metal thickness and line widths. This requirement stems from the need to eliminate critical production trimming which would significantly increase the cost of the resonator. The fabrication of grooves in a resonator is also expensive as they cannot be fabricated reliably by large-volume production techniques. This eliminates the freedom to use grooved reflectors or recessed electrodes in the design. Furthermore, cost, and therefore chip size, are of critical importance. In order to maximize the number of devices per wafer, the resonator design must be the smallest size possible, capable of meeting the design specifications. Together, these and other design constraints demand that a commercial resonator is designed with a thorough understanding of the sensitivity of the structure to all the production variables.

*This work was supported by National Science Foundation Grants ECS-8460847 and ISI-8521286.

†Present address: Etudes et Productions Schlumberger, 26 rue de la Cavee, Clamart 92140, France

CH2690-6/89/0000-574 $1.00 C 1989 IEEE

It is essential, therefore, that an accurate and efficient theoretical modeling of the resonator is available.

While the market for resonators is much larger than that for coupled resonators, the latter demand a much higher unit cost. Not infrequently, coupled resonators are used in oscillator circuits because of their more desirable phase characteristics. However, more commonly they are employed in narrow-band, low-loss, front-end filtering applications. For this application they have the dual advantages of both small size and stability over competing technologies. The design of a coupled resonator for filter applications is significantly more difficult than the design of a resonator intended for frequency stabilization. Additional design constraints are generally imposed. Very low in-band insertion loss and high out-of-band rejection are usually required. Consequently, the problems of device reproducibility also become more critical. In order to meet these requirements two novel resonator structures were devised.

2 Conventional Structures

In this section we consider the performance and the sensitivities to topological changes of conventional resonator structures. The theoretical responses shown were all obtained using a new, very general Coupling-Of-Modes (COM) analysis described in an accompanying paper [2]. This analysis predicts measured resonator performance very accurately and is computationally very efficient. The COM analysis allows the designer complete freedom to change any resonator parameter at will and to see immediately the impact on the response. We shall begin by considering the design and performance of a good commercial resonator for frequency-stabilization applications. Then, we shall consider the ways in which its performance might be modified to make it more suitable for filter or coupled-resonator applications.

2.1 The Synchronous Resonator

A two-port, one-pole resonator comprises two transducers and two gratings as shown schematically in Fig. 1. For maximum reflectivity in the grating regions there are generally two stripes per wavelength (λ) and the metallization ratio is nominally 0.5. The stripe and gap widths in the grating regions are, therefore, all nominally $\lambda/4$. Assuming that the same type of metal is employed throughout, a condition for resonance is that the separation between the innermost stripes of the two gratings be $n\lambda/2$, where n is an integer. The precise value of n, and the positioning of the two transducers within the cavity, are the principal parameters available to the designer for adjusting the precise characteristics of the resonator.

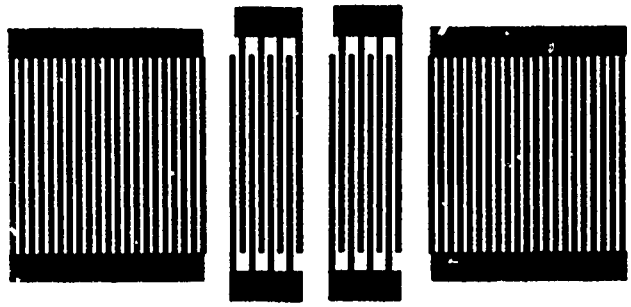


Figure 1: Two-port one-pole SAW resonator schematic.

In high-performance resonator designs, the transducer electrodes are typically recessed in grooves to eliminate any localised reflections within the transducer regions [3]. Eliminating such reflections has distinct advantages which will be described shortly . However, the recessed-electrode technology is too expensive to implement for high-volume, low-cost manufacture. Alternate schemes for eliminating transducer electrode reflections, such as using sixth- or eight-wavelength-wide electrodes also have disadvantages. First, having significantly different line widths within the resonator significantly increases manufacturing difficulties, especially for high-frequency devices. Second, having different electrode widths in the transducer and grating regions results in differing SAW velocities in the two regions. This complicates the design and the manufacturability of the resonator. For these reasons most commercial resonators employ $\lambda/4$ electrodes in both the transducer and grating regions.

If two-electrodes-per-wavelength transducers are used in a resonator, the internal reflections in the transducers must be properly accounted for in the design. These reflections are particularly important since the transducers are in the middle of the resonant cavity (Fig. 1) and thus they can significantly affect the device response. One placement option with significant advantages over alternate approaches is synchronous placement [4], [5]. In this case, the transducers are placed "synchronously" or periodically with respect to the adjacent gratings. Thus, as regards reflection, the transducers form a continuous

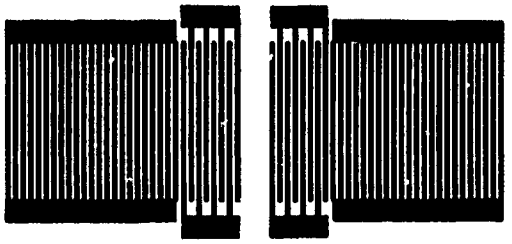


Figure 2: Synchronous placement of transducers.

extension of the gratings. This concept is illustrated in Fig. 2.

The length of the resonant cavity in a synchronous resonator is determined by the separation of the innermost transducer electrodes, rather than the grating electrodes. In a synchronous resonator, therefore, the resonant cavity length can be quite small, of the order of a few wavelengths. Generally, having a small cavity length is an advantage since it results in spurious cavity resonances being well removed from the pass band.

If the transducers are not placed synchronously with the gratings, additional close-in spurious resonances arise. These additional resonances are the result of interference between the mechanical reflections of the transducers and the gratings. While these spurious responses may present no problems for oscillator applications, they are highly undesirable for filter applications.

The major advantage of synchronous placement, however, comes in regard to manufacturability. With synchronously placed transducers, providing the gratings are sufficiently long, the resonator response has a minimal sensitivity to the acoustic reflection coefficient per unit length $\mathcal{K}(x)$ [2]. This allows the metal thickness to be adjusted in manufacture for frequency trimming, while the shape of the response remains otherwise unchanged. With non-synchronous configurations, both the device frequency and response shape will be affected by small changes in the metal thickness. The response of a typical synchronous resonator with $\lambda/4$ aluminum electrodes is shown in Fig. 3.

In summary, the following characteristics are highly desirable if a resonator is to be manufacturable at a low unit cost.

- Single-level metallization.
- No grooves or recessed electrodes.
- Uniform $\lambda/4$ electrode widths in both the transducer and grating regions.

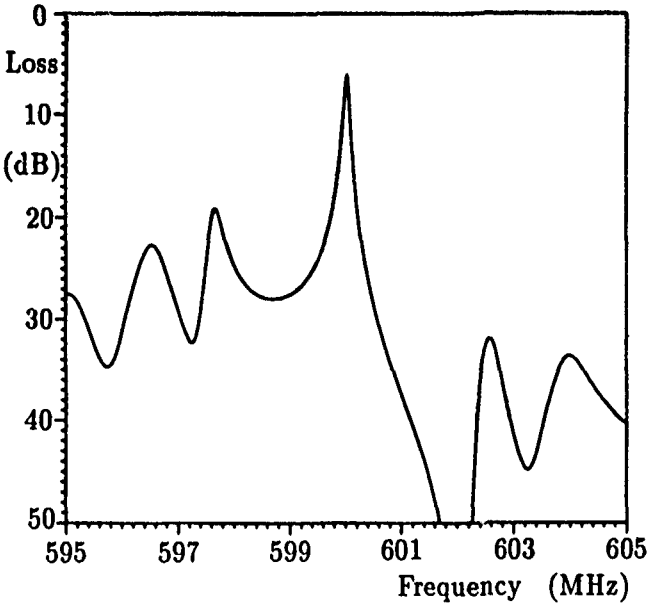


Figure 3: Response of a typical synchronous, two-port resonator.

- synchronous placement of the transducers with respect to the gratings.

2.2 The Problem of an Asymmetric Response

The structure of a simple two-port resonator with synchronous transducer placement is totally symmetric about the centre of the cavity. However, the response of such a resonator, as shown in Fig. 3, is unsymmetric about the resonant frequency. On the low side there is an adjacent spurious peak only 13 dB below the maximum resonance. At a corresponding distance away on the high side, however, there is actually a null. The reason for the asymmetry is that the transducers are not situated to couple maximally to the standing waves in the cavity. In fact, at resonance, they are situated exactly midway between the positions of maximum and minimum coupling. The result is that when moving to one side of the resonance, the coupling increases, while when moving to the other side the coupling decreases. This results in the asymmetric response of Fig. 3.

For oscillator applications the asymmetric response of Fig. 3 is generally of no consequence. However, this is not always the case. The response shown is for a resonator in a 50 Ω system. In many oscillator configurations the resonator may actually be in a much higher-impedance environment. In a high-impedance environment the amplitude of the low-side spurious response generally increases at a faster rate

than the main resonance. In some cases it may actually approach or exceed the amplitude of the latter. If this happens the oscillator can have a frequency-hopping problem, jumping backwards and forwards between the two resonant peaks. To prevent this problem it is necessary to ensure that the undesired resonance is sufficiently suppressed with respect to the desired resonance. One way to achieve this is to reduce the beamwidth of the resonator. However, this increases the resonator insertion loss. A compromise must therefore be made between passband insertion loss and spurious rejection. A resonator response with lower spurious sidelobes would thus also be advantageous from the viewpoint of reducing resonator insertion loss.

2.3 Effects of Topological Changes

A resonator with a near-symmetric, out-of-band response, and enhanced sidelobe rejection compared with that in Fig. 3 would be advantageous in many applications. However, for it to be manufacturable at low cost we would like to retain all the advantages of the synchronous structure discussed above. In this section, we examine the effect of simple topological changes to the resonator. We begin by examining in more detail the problem of the transducer coupling to the standing waves.

The acoustic impedance under an aluminum stripe on quartz is generally lower than that on the free surface [6]. In this case, the reflection coefficient of a SAW incident on the stripe will be negative at the first edge and positive at the second. In a periodic grating, as in a resonator, all the stripes act in synchronism at resonance. Thus, the inner edges of the grating stripes in a resonator will act as low-impedance points while the outer edges will act as high-impedance points. The resulting, associated electric potential of the standing wave at resonance is shown in Fig. 4

In a resonator with synchronous placement of the transducers the innermost grating stripes are the transducer electrodes. Referring to Fig. 4 we see that at the resonant frequency of the gratings the transducers are not coupled to the peaks of the standing wave. This increases the insertion loss of the resonator. If the transducers are re-located to the positions of maximum coupling, the problems associated with asynchronous placement arise that were discussed above. The problems of asynchronous can be avoided if the transducer electrodes are made reflectionless. This can be achieved by recessing the electrodes [3]. Both transducers can then be located for maximum coupling in order to achieve minimum insertion loss. However, this approach is not amenable to large-volume, low-cost manufacturing techniques.

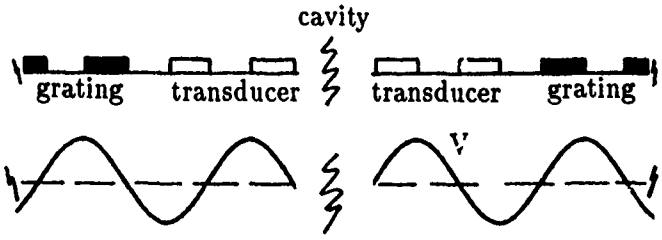


Figure 4: Electrical potential of standing wave under electrodes at resonance.

Referring to Fig. 4, below resonance the standing-wave peaks of the electrical potential move in towards the cavity. This results in increased coupling of the transducers to the standing wave. However, above resonance the peaks of the potential move out from the cavity. Thus the transducers become less tightly coupled to the standing wave. This behavior results in the low-side spur and the high-side null in the resonator response shown in Fig. 3. The exact frequencies at which the transducers are maximally and minimally coupled depend on the value of $\mathcal{K}(x)$. The greater the value of $\mathcal{K}(x)$, i.e. the greater the metal thickness, the further away from resonance are these frequencies. Let us now examine in more detail the effects of some topological changes that might be attempted in order to obtain a more symmetrical resonator response.

First, it should be evident from the above arguments that if the effective transmission-line impedance under an electrode stripe is greater than that of the free surface, rather than less as is the case for aluminum, the sense of the response asymmetry should reverse. To confirm this prediction a resonator was fabricated on quartz with gold metallization. Its response is shown in Fig. 5. The effective SAW impedance under a gold electrode on quartz is higher than that of the free surface, thus the sign of $\mathcal{K}(x)$ is opposite to that of aluminum. As predicted, the response asymmetry in Fig. 5 is reversed as compared with that in Fig. 3. With the gold resonator, the null appears on the low side of the resonance and the maximum spurious peak is on the high side.

As a result of this observation, the possibility immediately presents itself of being able to achieve a

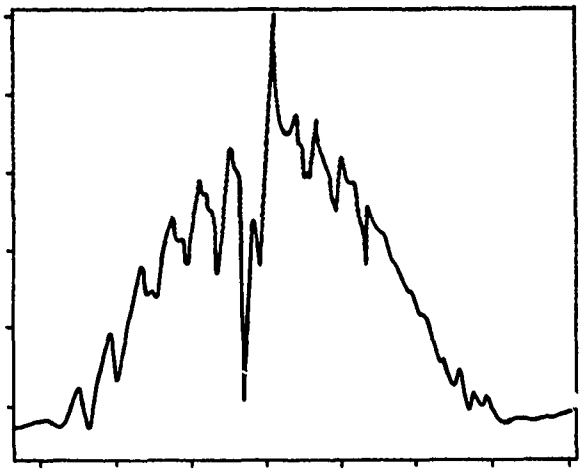


Figure 5: Measured response of a 674 MHz gold resonator on quartz (vertical 10 dB/div., horizontal 1.5 MHz/div.).

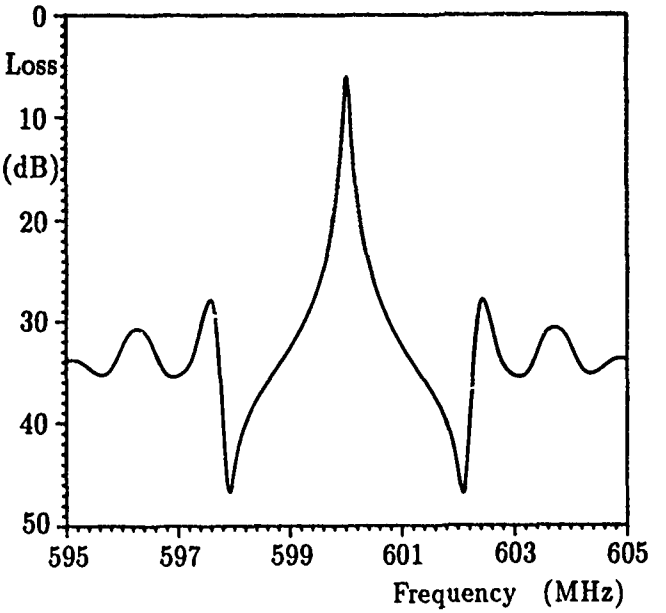


Figure 6: Theoretical response of a resonator on quartz with aluminum and gold electrodes.

symmetric resonator response by constructing a resonator from an aluminum/gold combination. That is, one grating-transducer pair constructed from aluminum and the other pair from gold. Since the reflection coefficients of the two halves of the resonator will be of opposite sign the resonant cavity length must be reduced by $\lambda/4$ compared to a resonator employing a uniform metallization. To achieve a symmetric response the distributed reflectivity $\mathcal{K}(x)$ in the two halves of the resonator must be equal in magnitude but of opposite sign. The transmission zero of each transducer-grating combination will then precisely null out the near-in spurious peak of the other half of the resonator. The theoretical response of such a resonator is shown in Fig. 6. The response is totally symmetric with the highest sidelobe levels now being approximately 22 dB below the main resonance. Such a response would be ideal for filter and coupled-resonator applications. Unfortunately, it is entirely impractical for a number of reasons.

1. Gold is not a useful material for resonators because it has high viscous damping resulting in a high insertion loss.

2. It would be extremely difficult trying to keep $|\mathcal{K}(x)|$ identical in the gold and aluminum regions.

3. The SAW velocity would be different in the gold and aluminum regions.

4. The turnover temperatures in the gold and aluminum regions would be different.

5. Fabrication and trimming of a such dual-metal resonator would be extremely difficult.

The symmetric response of Fig. 6, however, does give us some idea of what might be achievable from an optimally-configured resonator. The question is how it might be achieved by a practical low-cost structure?

Referring to Fig. 4, another solution that presents itself for achieving a more symmetric response is to displace one of the transducers $\lambda/4$ away from its neighboring grating. The resulting 90° phase shift in the standing wave under the displaced transducer, causes it to be become less-tightly coupled to the standing wave below the resonance, and more tightly coupled above. Exactly the opposite behavior to the other transducer-grating combination which remain synchronous. Thus, we would expect both low- and high-side nulls.

The theoretical response of such a resonator with one transducer displaced $\lambda/4$ from synchronism and the other synchronously placed is shown in Fig. 7. As expected the resonator does have adjacent zeros on both sides of the resonance. However, it has significantly increased insertion loss in the main resonance and there is a low-side spurious peak that is actually greater in amplitude than the desired resonance. Furthermore, the frequency of the desired resonance has been shifted upward by the presence of the low-side null to above 600 MHz, the frequency of maximum grating reflection. The cause

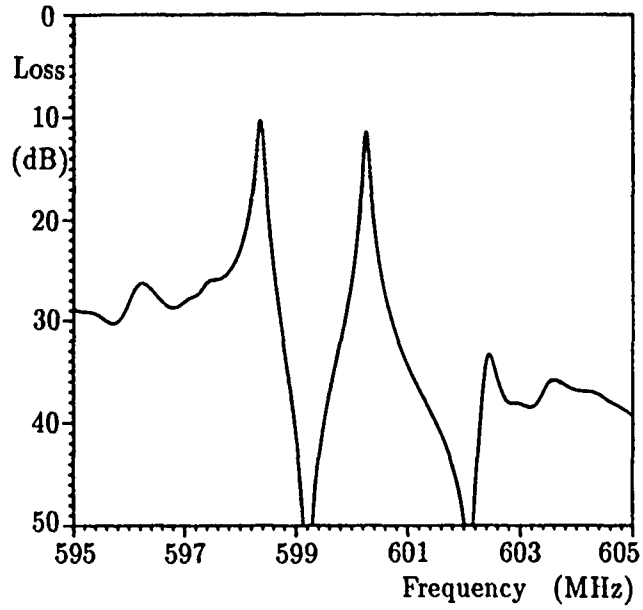


Figure 7: Theoretical response of a resonator with one transducer shifted $\lambda/4$ from synchronism.

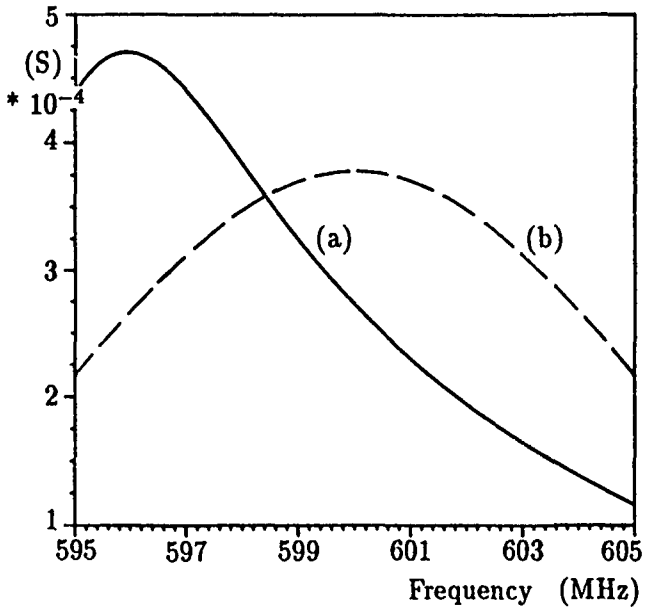


Figure 8: Input conductance of a SAW transducer on quartz, (a) aluminum electrodes with $K(x) = -0.02$, (b) $K(x) = 0$.

of the problem is that by asynchronously placing one of the transducers we have introduced another pole into the response. We essentially have a very non-optimum two-pole coupled resonator. The reason that the low-side peak is so high is that the conductance of a transducer with distributed internal reflections is asymmetric. In the case of aluminum electrodes on quartz, the conductance is a maximum below the frequency corresponding to the periodicity of the transducer. The input conductance of an aluminum two-electrodes-per-wavelength transducer on quartz is shown in Fig. 8(a). The periodicity of the transducer corresponds to 600 MHz. For reference the input conductance of an identical transducer but with the electrode reflectivity suppressed is shown in Fig. 8(b). Note, that the conductance peak is shifted down in frequency by approximately 4 MHz and the conductance at 600 MHz is reduced by the electrode reflections.

Various other simple topological changes can be made to a two-port resonator resulting in a near limitless number of resonator responses. However, none comes close to achieving the near-optimal response of the gold-aluminum resonator of Fig. 6. In addition, all have the disadvantages of asynchronous placement.

In the following sections, two new resonator structures are described that have essentially symmetric resonant responses with substantially reduced spurious peaks. At the same time, however, both have all the advantages of synchronous placement as desired for high-volume, low-cost manufacture. To achieve this it was necessary to make some fundamental changes to the conventional-type resonator configurations considered above.

3 The NSPUDT Resonator

The concept of Natural Single-Phase Unidirectional Transducer (NSPUDT) crystal orientations was first described with a view to low-loss filter applications [7]. On these special crystal orientations there is a natural phase shift of 45° between the effective centres of transduction and reflection in a two-electrodes per-wavelength transducer. The result is that even a single-level, two-electrodes-per-wavelength transducer has finite directivity, i.e. radiates preferentially in one direction. The particular preferred radiation direction in such a device depends only on the substrate and the metal employed for the electrodes. The sense of the directivity cannot be reversed by changes to the structure or to the drive to the transducer as, for example, with a three-phase transducer. Hence the term "natural".

The construction of a low-loss filter comprised of two NSPUDT's requires two metals with opposite-signed directivities for the specific crystal orientation. In order to avoid this problem, initial low-loss filter designs on NSPUDT orientations have used only one

NSPUDT transducer and another SPUDT or bidirectional transducer.

After confirming the existence of NSPUDT crystal orientations, it quickly became evident that these orientations had advantages for other classes of devices than just low-loss filters. The symmetric input conductance and flat input susceptance obtainable from a single two-electrodes-per-wavelength transducer on an NSPUDT orientation is ideal for notch-filter applications [8], [9]. In this case, the problem of being able to reverse the directivity does not arise. Less obvious, perhaps, is the fact that these crystal orientations also have significant advantages for resonator applications. To see why, we return to Fig. 4 showing the phase of the electrical potential of the standing waves under the synchronously-placed transducers at resonance.

Each transducer is located 45° away from the position of maximum coupling to the standing wave. This increases the resonator insertion loss and also causes the asymmetric response. Thus, to eliminate both problems we need to devise a structure that will result in both transducers being maximally coupled at the resonance. As we have already seen this cannot be accomplished by simply moving the transducers. The grating and transducer reflections are then asynchronous which results in other problems. If the transducers cannot be moved, then the standing wave must be displaced instead, ideally by 45°. This can be achieved by fabricating the resonator on an NSPUDT orientation.

On an NSPUDT orientation, the electrical potential of a standing under a two-electrodes-per-wavelength grating is displaced 45° as compared to that under an identical grating on a conventional orientation. If, as desired, only one metal preferably aluminum is used throughout, the standing waves under each grating-transducer pair will be displaced in the same direction. Thus, if a standard synchronous resonator configuration is deposited on an NSPUDT orientation, at resonance one of the transducers will be maximally coupled to the standing wave while the other will be totally uncoupled. The result will be a null rather than a peak at the expected resonant frequency. This "anti-resonance" phenomenon is useful for identifying NSPUDT orientations but not very useful as a resonator!

The solution to the problem is to move both transducers to the same side of the resonant cavity on an NSPUDT orientation [10]. Then, both transducers are maximally coupled at resonance which results

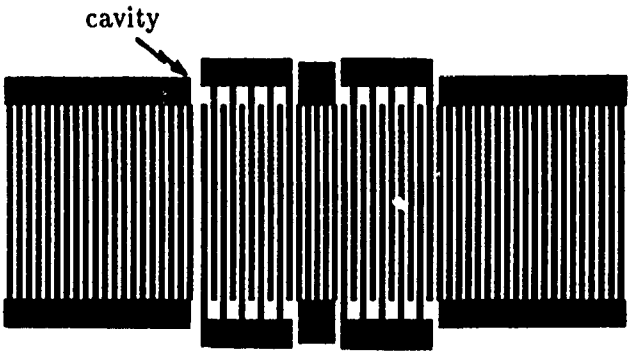


Figure 9: Schematic of a two-port NSPUDT resonator.

in a minimum insertion loss. In addition, the input conductance of a two-electrodes-per-wavelength transducer on an NSPUDT orientation is symmetric about its design frequency $f_0 = v/4w$, where v is the SAW velocity and w is the electrode width [7]. Thus, the spurious responses below resonance are not enhanced as on a conventional crystal orientation by the asymmetric form of the transducer input conductance (Fig. 8). Maximum coupling at resonance and a symmetric transducer input conductance result in a totally symmetric response with very low spurious sidelobes. Fig. 9 shows the schematic of a two-port NSPUDT resonator. Note, that both transducers are on the same side of the $\lambda/2$ resonant cavity. The two transducers are separated by a short non-resonant grating spacer to reduce crosstalk. Both transducers are synchronously placed with respect to the right-hand grating. As a result, the resonant frequency of the device is essentially independent of the localised reflection coefficient $\mathcal{K}(x)$. This is a significant advantage for manufacturability.

Referring to Fig. 9 it is evident that the transducer closest to the resonant cavity will tend to be more closely coupled to the standing wave in the cavity than the second transducer. If the two transducers are identical, therefore, the input characteristics of the resonator will be asymmetric on the two ports. For some applications this can be an advantage. The resonator may be designed with a high input impedance on one port and a much lower impedance on the other port. However, if a more symmetric two-port response is required, this can be achieved in large part by increasing the length of the outer transducer relative to the inner one.

The theoretical response of such an optimized NSPUDT resonator is shown in Fig. 10 This response was generated using the material parameters for alu-

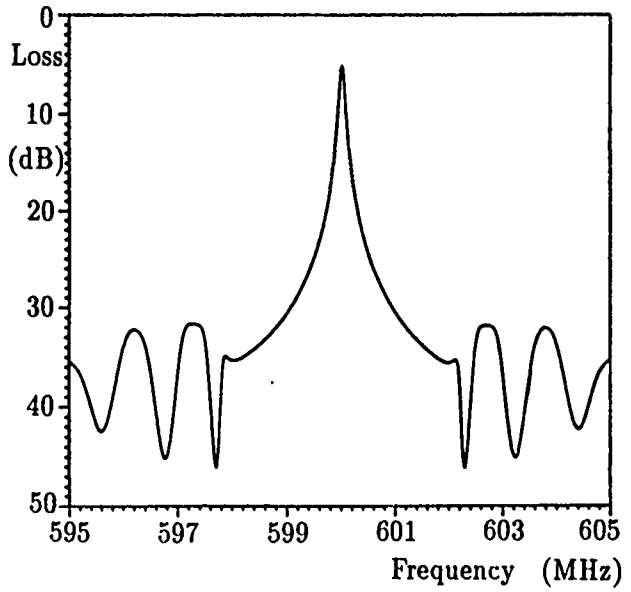


Figure 10: Theoretical response of an optimized two-port NSPUDT resonator.

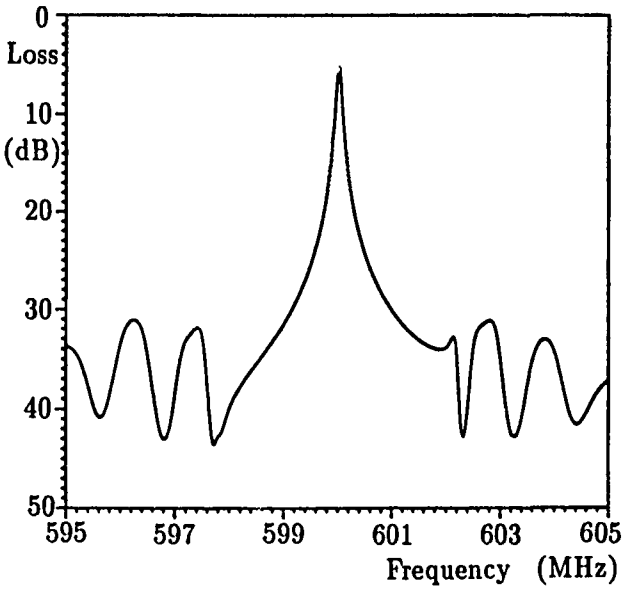


Figure 11: Theoretical response of an optimized two-port NSPUDT resonator on an orientation with 35° phase shift between centres of transduction and reflection.

minum on ST-quartz, but with the assumption of a 45° phase shift between the centres of transduction and reflection. The NSPUDT-resonator response is near ideal and considerably better than any of the conventional design variations considered in the previous section. The insertion loss is lower and all spurious responses are down by more than 27 dB. The predicted response is better even than that of the dual-metal resonator configuration in Fig. 6.

Finding an NSPUDT crystal orientation with exactly 45° phase shift between the centres of transduction and reflection and all the other desirable characteristics for a resonator is a demanding task. Fortunately, it turns out that it is not very critical to the performance of the resonator that the phase angle be exactly 45°. Fig. 11 shows the response of the same NSPUDT resonator on an orientation having only 35° phase shift between the transduction and reflection centres. The response is only a little degraded from that in Fig. 10. The main effect is a small increase in the spurious levels. Thus, it is not critical to find an exact NSPUDT orientation for the resonator. Crystal orientations may therefore be selected that are a compromise between ideal NSPUDT behavior and the other properties desirable for a resonator.

An NSPUDT orientation exists on the ST-cut plane of quartz for propagation in the x+25° direction [11]. The electro-mechanical coupling coefficient K^2 for this orientation is approximately 0.0013. Unfortunately the turnover temperature for devices on this orientation is typically around 125° which is much too high for most resonator applications. It also has approximately 5° beam steering, though this is only a minor complication. Because of its immediate availability, initial verifications of several NSPUDT concepts were performed on this orientation. Since then, other NSPUDT orientations on quartz with much more ideal resonator characteristics have been identified. In particular, a doubly-rotated cut has been identified with good coupling, turnover and reflectivity characteristics, though again with a moderate amount of beam steering. This [111] cut has Euler angles (45°,55°,0°) [12].

Fig. 12 shows the measured transmission response of a prototype NSPUDT resonator constructed on the ST(x+25°)-quartz orientation. The device was constructed from single-level aluminum with a nominal thickness $h/\lambda = 0.02$. Qualitatively, the response is as expected, though the minimum insertion loss of approximately 10 dB is higher than predicted. There are two reasons for this:

1. No compensation was made in the design for beam steering.

2. The value of reflectivity $\mathcal{K}(x)$ achieved by the aluminum stripes was measured by another test structure as 0.012 instead of the design goal of 0.02.

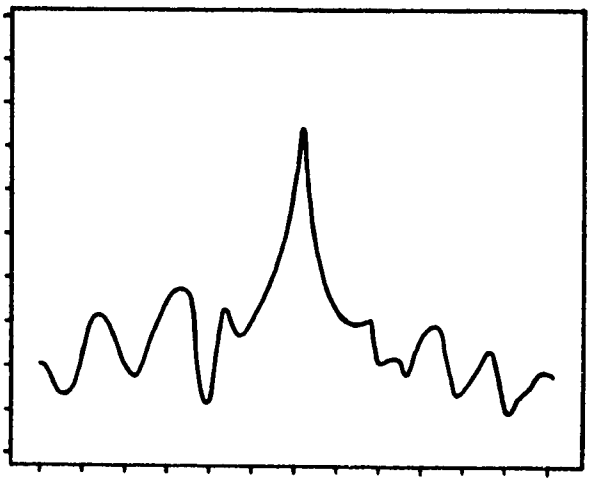


Figure 12: Measured response of an NSPUDT resonator on ST(x+25°)-quartz (vertical 5 dB/div., horizontal 0.5 MHz/div.).

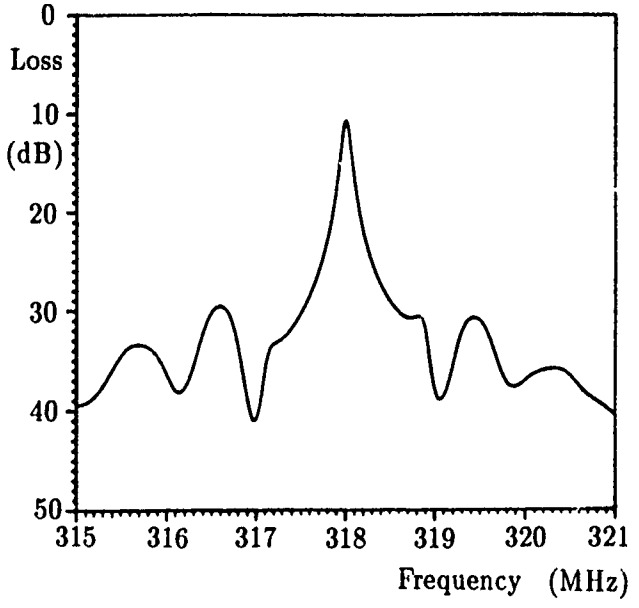


Figure 13: Theoretical response of NSPUDT resonator on ST(x+25°)-quartz.

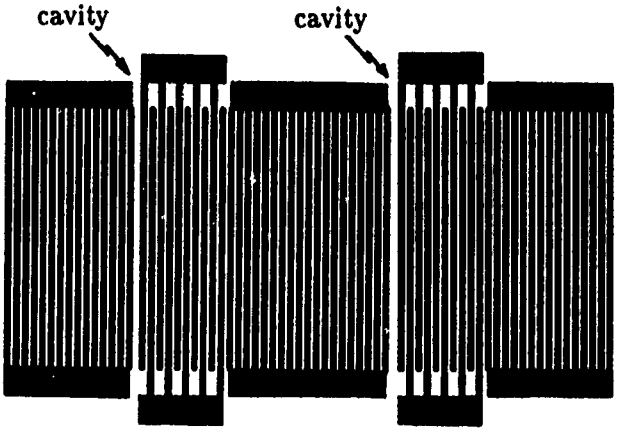


Figure 14: Schematic of a two-port, two-pole NSPUDT coupled resonator.

Further experiments showed that the magnitude of the reflection coefficient of an aluminum stripe on the ST-cut plane actually goes through zero at a propagation angle of approximately x+35°. For the NSPUDT direction x+25°, the reflection coefficient is reduced to approximately one half of its value for x-axis propagation. Taking this reduction in $\mathcal{K}(x)$ into account the theoretically predicted response for the resonator is shown in Fig. 13. The latter is in good agreement with the measured data.

At this point one practical and unusual characteristic of NSPUDT resonators should be pointed out. With conventional resonator structures on typical crystal orientations the devices need only be aligned along the correct propagation direction without respect to the sense. Thus, the mask or crystal may be rotated by 180° without consequence. This is not the case for the NSPUDT resonator. The structure and the crystal properties are both asymmetric which results in a symmetric response. If either the mask or crystal are rotated by 180° the transducers become totally uncoupled from the cavity resonance and thus the device is "anti-resonant" with a null rather than a peak at resonance.

By introducing another cavity into the basic NSPUDT resonator configuration shown in Fig. 9, a two-pole, coupled-resonator response may easily be obtained as shown schematically in Fig. 15. The basic difference between the two-pole and one-pole NSPUDT resonator configurations is simply the addition of a grating coupler between the two transducers and an additional resonant cavity. Both transducers are on the same side of their adjacent resonant cavity and thus both are maximally coupled at resonance. The transducers are also both synchronously placed

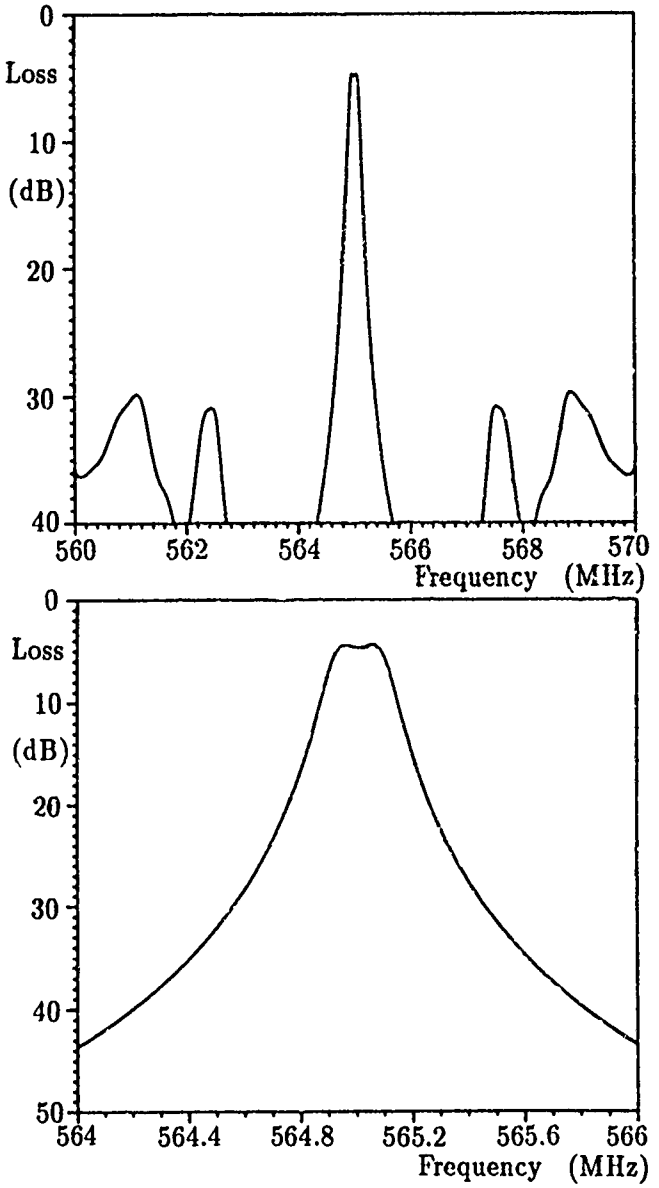


Figure 15: Theoretical response of a two-pole NSPUDT coupled-resonator filter on ST(x+25°)-quartz.

with respect to their neighboring gratings. Thus, all the advantages of the one-pole synchronous design are maintained.

The centre grating coupler controls the coupling between the two cavities and, thereby, the separation of the two poles of the resonator. Theoretical responses of a two-pole NSPUDT coupled resonator with single-element tuning are shown in Fig. 15. Note that the passband is relatively flat and that the device has very low spurious sidelobe levels.

To verify the concept of the two-pole NSPUDT coupled resonator, a prototype device was built on the ST(x+25°)-quartz NSPUDT orientation. The resonator was designed to be used untuned in a coilless oscillator circuit. The measured response of the

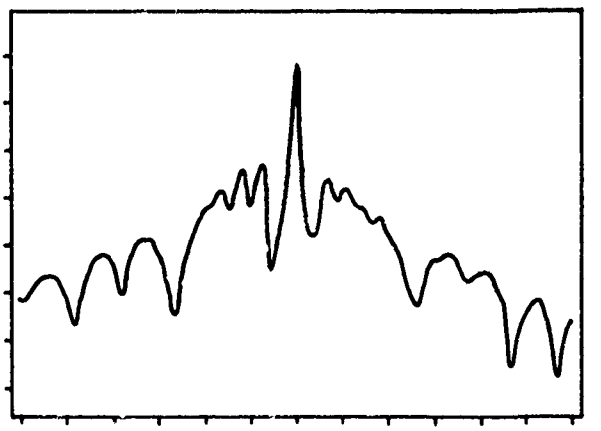


Figure 16: Measured response of a 322 MHz two-pole NSPUDT coupled-resonator filter on ST(x+25°)-quartz (vertical 10 dB/div., horizontal 2 MHz/div.).

resonator shown in Fig. 16 was in excellent agreement with the COM analysis.

4 The "Hiccup" Resonator

The previous section demonstrated one approach for achieving a symmetric one- or two-pole resonant response with low out-of-band sidelobes. The structures described required only a single-level, single-metal, unrecessed two-electrodes-per-wavelength construction. The symmetric responses were achieved by very careful selection of the choice of crystallographic orientation used for the resonators. In this section, we will discuss an alternative approach for obtaining a "near-symmetric" response on conventional crystal orientations.

As was demonstrated in section 2.3, simple topological changes to the basic two-port, one-pole resonator configuration of Fig. 1 do not lead to the desired symmetric response with low spurious sidelobes. Therefore, we now consider how this might be achieved by fundamental changes to the structure.

The synchronous resonator has good high-side rejection as shown in Fig. 3. However, the low-side rejection is inadequate for most filtering applications. There are two principal reasons why the low-side spurious responses are so high:

1. Both transducers become maximally coupled below resonance.

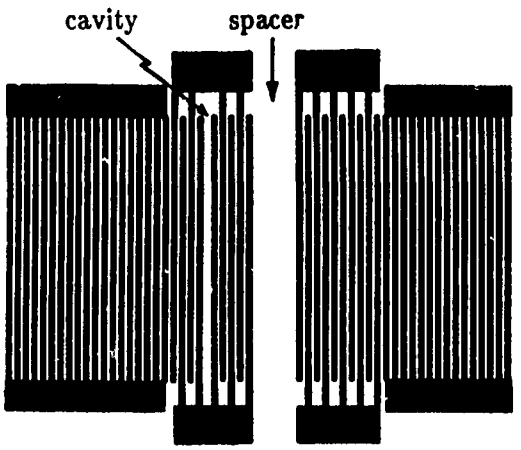


Figure 17: Two-port, one-pole "hiccup" resonator configuration.

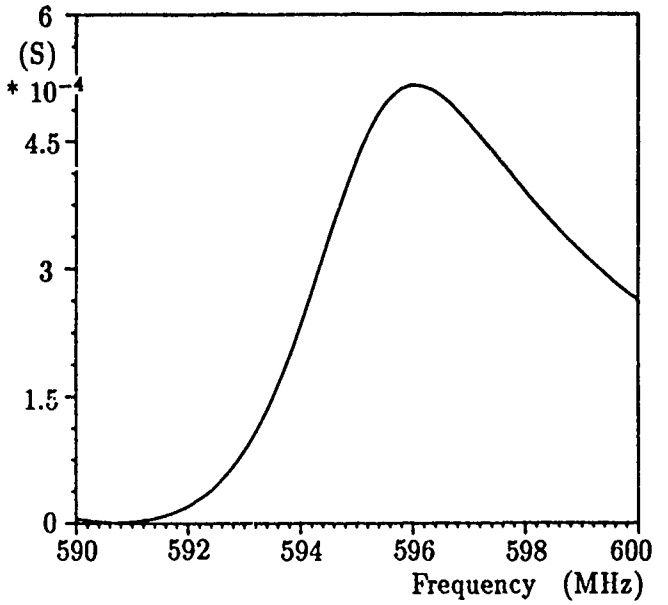


Figure 18: Theoretical input conductance of a "hiccup" transducer on quartz with $\kappa(x) = -0.02$, f-synch. = 595 MHz.

2. The input conductances of the transducers peak below resonance as a result of the electrode reflections (see Fig. 8).

A further consequence of the transducer electrode reflections is that the conductance of both transducers is actually decreased at resonance. This increases the insertion loss at resonance and thus further degrades the out-of-band rejection. To create a more symmetric resonator response, we therefore choose to attack the problem at its root by trying to make the conductance of at least one of the transducers more symmetric.

To accomplish this we shift the resonant cavity from between the two transducers as in Fig. 2, to the centre of one of the transducers as in Fig. 17. Because of the jump in the periodicity in the one transducer, this structure is referred to as "hiccup" resonator. With the exception of the cavity, all stripes are again totally periodic and synchronous. Thus, this structure again has all the desirable features of a synchronous resonator as regards manufacturability. In this structure, the transducer surrounding the cavity will tend to be more tightly coupled to the standing wave than the second transducer. As with the NSPUDT resonator configuration, however, more symmetric input characteristics can be achieved by increasing the length of the less-coupled transducer.

The placement of the cavity in the middle of one of the transducers brings an extra degree of freedom to the design of the resonator. It allows an additional zero to be placed in the transducer input conductance close to the resonance. This additional zero may be placed above, or below the resonance, by choice of the relative phase of the electrical drive to the two halves of the transducer. The aperiodicity in the centre of the transducer also causes a shift in the frequency of maximum conductance of the transducer. In the absence of electrode reflections i.e. $\kappa(x) = 0$, this would occur at the frequency at which the two halves of the transducer radiate constructively. In the latter case, the conductance peak and zero would be symmetrically located about the common synchronous frequency of the two halves of the transducer.

Due to the presence of electrode reflections, the conductance of a uniform two-electrodes-per-wavelength transducer is shifted down in frequency (Fig 8). Consequently, we choose the phase drive to the two halves of the "hiccup" transducer to place the conductance zero below resonance, and the conductance peak above. As the latter is also moved down in frequency by the electrode reflections, the resulting conductance can be designed to be very close to the synchronous frequency of the two halves of the transducer. The precise positioning of the transducer's conductance zero and maximum are controlled by the overall length of the transducer and the size of the cavity.

Fig. 18 shows the theoretical input conductance of a "hiccup" transducer on quartz. This conductance was calculated assuming a reflection coefficient/unit length $\kappa(x) = -0.02$. The latter is typical for aluminum electrodes on quartz. The synchronous frequency for the two halves of the transducr was

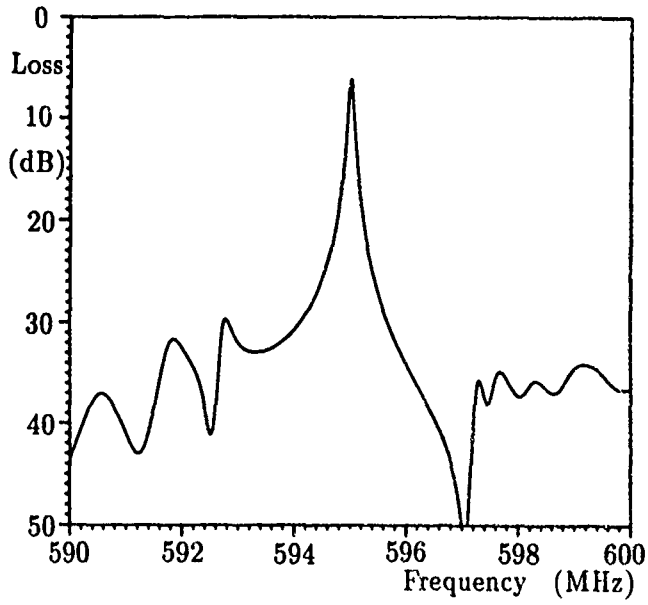


Figure 19: Theoretical response of a two-port, one-pole "hiccup" resonator.

595 MHz. Note, the conductance peak is just slightly above the latter value. There is also a null in the conductance at approximately 591 MHz. This null is very broad and causes a rapid falloff in the conductance of the transducer below resonance. In a resonator configuration the result is a significant reduction in the spurious levels below resonance. This is illustrated by the theoretical response of a prototype "hiccup" resonator shown in Fig. 19. The sidelobe levels below resonance are substantially reduced in comparison to those of a conventional synchronous resonator such as that shown in Fig. 2. In practice, these resonators have been found to be extremely manufacturable and to be very well modeled by the COM analysis. They require only a single-level of metallization, with no recessed electrodes or other critical alignment steps, and can thus be fabricated with resonant frequencies well above 1 GHz.

As with NSPUDT resonators, two-pole coupled-resonator configurations can also be devised that take advantage of the concept of the "hiccup" transducer. Configurations with either one or two "hiccup" transducers can be constructed. The schematic of a coupled resonator with one "hiccup" transducer is shown in Fig. 20

Theoretical responses for two prototype filters, assuming single-element tuning, are shown in Fig. 21. The in-band insertion loss for both devices is very low and the out-of-band spurious responses are all well suppressed. The coupled-resonator with only

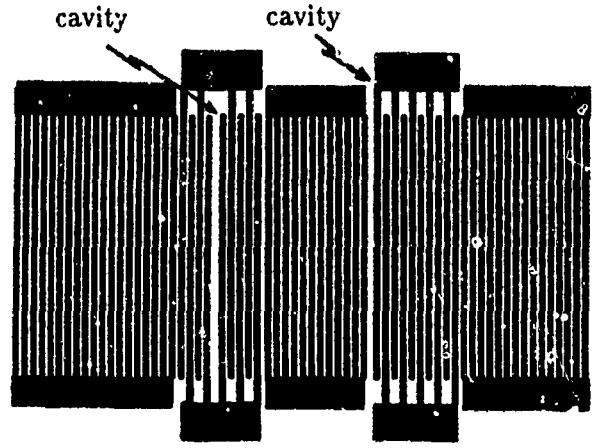


Figure 20: Two-port, two-pole, single "hiccup' coupled-resonator configuration.

one "hiccup" transducer (Fig. 21(a)) has almost symmetric out-of-band rejection. That employing two "hiccup" transducers (Fig. 21(b)) has similar rejection on the high side and substantially greater spurious rejection below the resonance. The former, however, tends to have a slightly flatter in-band response. The excellent low-side rejection of the two "hiccup" structure is a result of the broad low-side null in the conductance of each "hiccup" transducer. Many two-pole "hiccup" coupled-resonator filters have been built and shipped for a variety of applications. They have been found to be extremely easy to construct and very amenable to low-cost, large-volume manufacture.

5 Summary

We have considered the design of single-level two-port resonators and coupled resonators. In particular, we have tried to point out the factors important to high-yield, large-volume manufacturing techniques. The synchronously-placed resonator has most of the desirable characteristics, except for the existence of higher than desired out-of-band spurious responses. These are most troublesome for filtering applications.

Judicious placement of the transducers in a conventional two-port SAW resonator resonant cavity does not significantly reduce the sidelobe levels. In addition, non-synchronous placement of the transducers is highly disadvantageous from a manufacturing standpoint. Two new resonator configurations, the NSPUDT resonator and the "hiccup" resonator, were developed to overcome these problems. They have substantially enhanced out-of-band rejection, com-

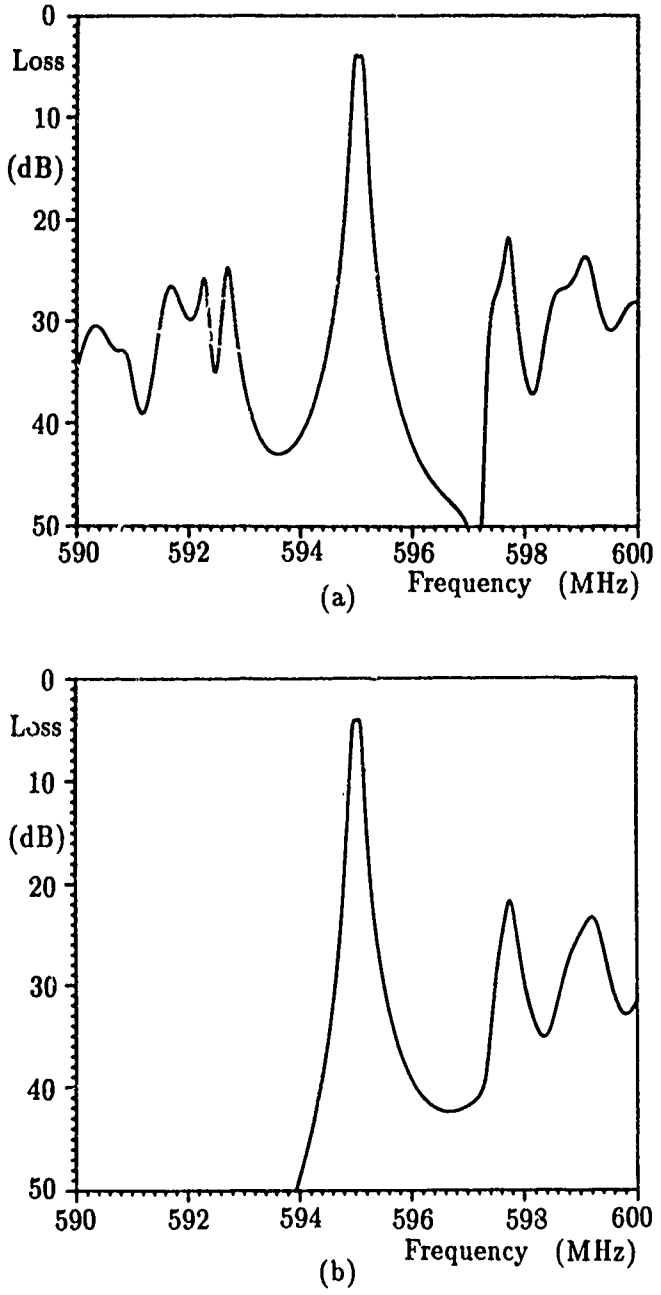


Figure 21: Theoretical responses of single-element tuned, two-pole, coupled resonators, (a) single "hiccup", and (b) dual "hiccup" designs.

bined with the stability of a synchronous design. In addition, both are single-metal, two-electrodes-per-wavelength structures requiring no grooves or critical alignment. Thus they can be readily manufactured for high-frequency applications.

6 Acknowledgement

Frank Perkins motivated and continuously supported this development program. Jeff Andle designed and tested many of the prototype devices and identified the [111]-quartz NSPUDT orientation. I gratefully acknowledge the continuing support of the National Science Foundation for this work.

References

[1] G. K. Montress, T. E. Parker, and J. Callerame, "A miniature hybrid circuit SAW oscillator using an all quartz packaged resonator," *IEEE Ultrason. Symp. Proc.*, pp. 277-282, 1985.

[2] P. V. Wright, "A new generalized modeling of SAW transducers and gratings," *this symposium.*

[3] W. J. Tanski, "Surface acoustic wave resonators on quartz," *IEEE Trans. Sonics Ultrason.*, vol. SU-26, no. 2, pp. 93-104, March 1979.

[4] P. S. Cross, W. R. Shreve and T. S. Tan, "Synchronous IDT SAW resonators with Q above 10,000," *IEEE Ultrason. Symp. Proc.*, pp. 824-829, 1979.

[5] W. R. Shreve, *US patent no. 4,144,507*, March, 1979.

[6] P. V. Wright, "Modeling and experimental measurements of the reflection properties of SAW metallic gratings," *IEEE Ultrason. Symp. Proc.*, pp. 54-63, 1984.

[7] P. V. Wright, "The natural single-phase unidirectional transducer: a new low-loss SAW transducer," *IEEE Ultrason. Symp. Proc.*, pp. 58-63, 1985.

[8] C. S. Hartmann, J. C. Andle, and M. B. King, "SAW notch filters," *IEEE Ultrason. Symp. Proc.*, 1987, pp. 131-138.

[9] C. S. Hartmann and P. V. Wright, *US patent no. 4,599,587*, 1986.

[10] P. V. Wright, *US patent no. 4,731,595*, March, 1988.

[11] P. V. Wright, *US patent no. 4,670,681*, March, 1988.

[12] J. C. Andle, *US patent no. 4,670,680*, March, 1988.

[13] P. V. Wright, *US patent no. 4,616,197*, Oct., 1986.

43rd Annual Symposium on Frequency Control - 1989

LOW NOISE SAW RESONATOR OSCILLATORS

T.E. Parker and G.K. Montress
Raytheon Research Division
131 Spring St.
Lexington, MA 02173

ABSTRACT

The design and performance of surface acoustic wave (SAW) resonator oscillators using techniques for obtaining low phase noise levels are described. These oscillators are very low phase noise engineering prototypes which also meet other vital system requirements, such as temperature stability and vibration sensitivity. A 500 MHz version has a phase noise floor of -177 dBc/Hz and a flicker noise level of -132 dBc/Hz at 1 kHz offset. An L-Band (frequency doubled) version has noise levels of -171 dBc/Hz and -126 dBc/Hz, respectively. Other important performance parameters that are discussed include: temperature dependence of frequency and output power, vibration sensitivity, frequency tuning characteristics, spurious levels, load pulling, voltage pushing and long-term frequency stability.

INTRODUCTION

Techniques for obtaining very low phase noise levels in SAW oscillators have been previously reported in the literature [1]-[3]. These techniques have now been applied to the design and fabrication of engineering prototype low noise oscillators that address real system requirements. The same basic voltage controlled, 500 MHz feedback loop oscillator design which was previously reported has been used for two different oscillator versions. In one case the output of the 500 MHz oscillator loop is amplified and filtered. In the other case, the 500 MHz output is frequency doubled and amplified to provide an L-Band output signal. Though low phase noise is important for these oscillators, other performance parameters that are vital to real system applications are also relevant and they have been taken into consideration. Specifically, the temperature dependence of frequency and output power level have been evaluated, along with vibration sensitivities along all three axes, frequency versus tuning voltage characteristics, spurious levels (harmonic and non-harmonic), load pulling of frequency, bias voltage pushing of frequency, and long-term frequency stability. In addition, a complete thermal analysis and evaluation was performed in order to produce a design with acceptably low transistor junction temperatures under worst case ambient operating conditions. As is usually the case, some degradation in phase noise performance, as compared to laboratory demonstration oscillators [3], has occurred because compromises had to be made in parameters that affect phase noise in order to meet other requirements. Nevertheless, exceptional noise performance has still been maintained.

OSCILLATOR DESIGN

A basic two-port SAW resonator oscillator design with 50 ohm modular components [4] was used for both oscillators. The circuit layout for the UHF version is shown in Fig. 1, while the L-Band version is shown in Fig. 2. The oscillator loop is identical for both versions and contains a varactor diode tuned phase shifter for electronic frequency control [3], and fixed L-C-L or C-L-C phase shifters for loop phase

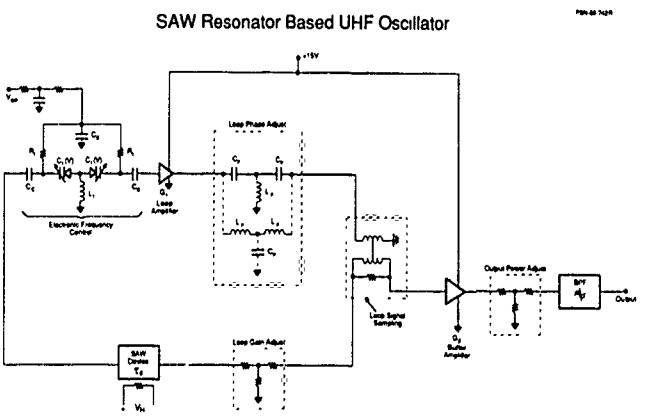


Figure 1. Circuit layout of the UHF Oscillator.

adjustments. The amplifier, power divider and attenuator in the loop are all commercially available components which are essentially the same as those used in Ref. 3. The SAW device is a low noise, nominally 500

CH2690-6/89/0000-588 $1.00 C 1989 IEEE

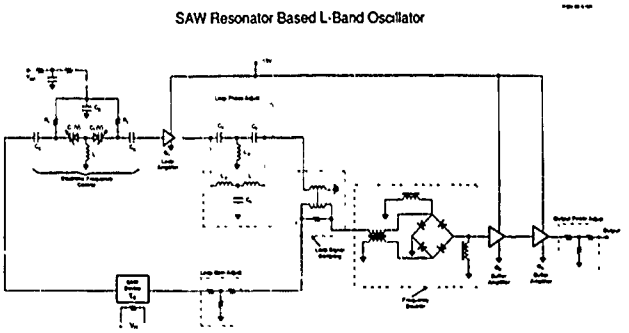


Figure 2. Circuit layout of the L-Band oscillator.

MHz two-port resonator in an "all quartz package (AQP)" [5]. The SAW resonators were fabricated at Raytheon and subjected to a high incident rf power "burn-in" procedure to reduce flicker noise [3]. The residual phase noise of each oscillator component was evaluated individually to insure a low oscillator phase noise level. 100 percent screening of the residual phase noise was required for all SAW devices and amplifiers (including buffer amplifiers).

The SAW resonator was mounted in a standard Kovar hybrid circuit flatpack as shown in Fig. 3. Air core inductors were used at each port of the resonator for series tuning and were also located inside the flatpack. The tuning inductors were used to reduce the insertion loss of the resonators to approximately 5 dB, and to fine tune the resonant frequency of the SAW device. Figure 4 shows the packaged SAW device mounted on an L-Band oscillator circuit board. A commercially available, self contained, DC proportionally controlled heater is mounted on an aluminum block which is placed over the SAW package.

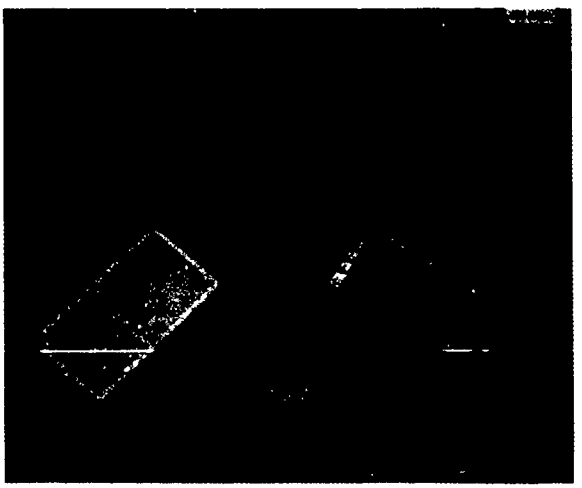


Figure 3. SAW resonator mounted in a flatpack.

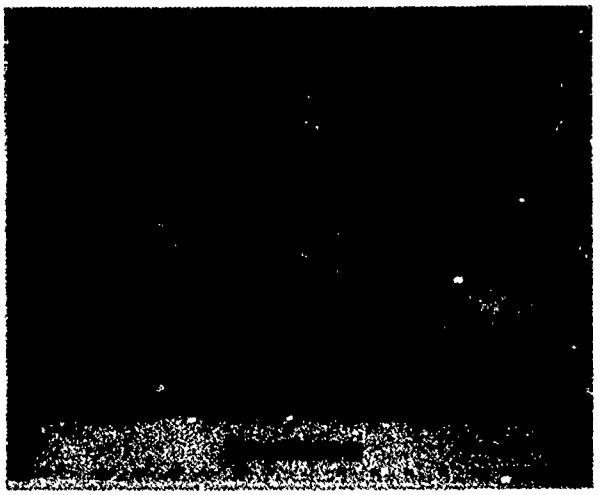


Figure 4. Packaged SAW resonator mounted on an L-Band oscillator circuit board. The heater module is next to the circuit board.

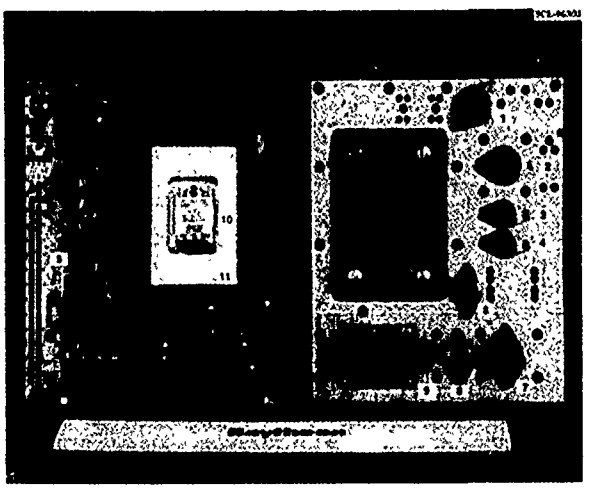


Figure 5. Front and back sides of fully assembled UHF oscillator circuit boards. The components are identified in the text.

The front and back sides of fully assembled UHF oscillator printed circuit boards are shown in Fig. 5. All of the critical oscillator components are housed in TO-8, TO-5 or flatpack packages, except for the filter which is in a rectangular package with PC pins. The location of each component in Fig. 5 is indicated as follows:

1 Electronic phase shifter

2 Loop amplifier

3 Fixed phase shifter

4 Fixed phase shifter or feedthrough (as needed)

5 Power divider

6 Attenuator, loop gain adjust

7 Buffer amplifier

8 Attenuator, output power adjust

9 Low-pass filter

10 Self contained heater module

11 Heater block

All of the components external to the loop for both the UHF and L-Band versions are commercially available, though careful selection for low phase noise is required. A .032 inch thick copper plate was located on one side of the printed circuit board underneath the oscillator components to facilitate the removal of heat from the amplifiers.

During the assembly of the oscillators it is necessary to evaluate the loop gain and phase before the final selection of fixed phase shifters and attenuators can be made. Therefore, provision is made for breaking the loop and evaluating the loop condition with a network analyzer. From these measurements the proper fixed phase shifter(s) and loop attenuator can be chosen. Once this is accomplished the loop can be closed and oscillation will be obtained. The proper output power attenuator can then be chosen.

A completely assembled UHF oscillator is shown in Fig. 6. The overall size of the oscillator is approximately 5" x 4" x 2.5". The L-Band version is essentially the same. Both oscillators operate from +15 VDC and draw less than 300 mA (not including the heater). A wide variety of heater modules is available, but generally a 28 watt version was used that draws up to 1000 mA at 28 VDC. The rf output power of the UHF version is nominally +25 dBm, and for the L-Band version +23 dBm. Small size was not a major factor in the design of these oscillators, but the entire oscillator loop (not including the heater) could potentially be integrated into a single hybrid circuit (less than 2" x 2" x 0.4") for a significant size reduction.

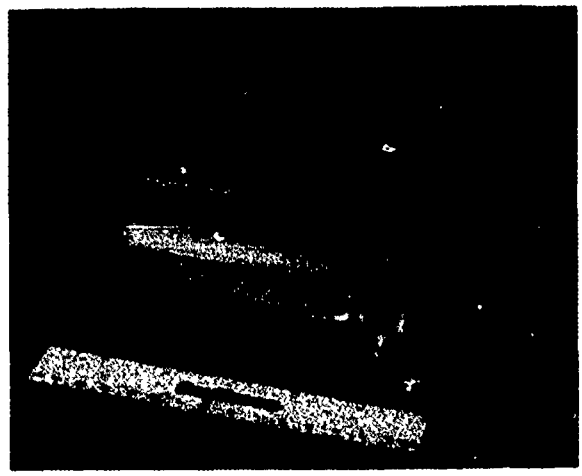


Figure 6. A complete UHF SAW oscillator.

PERFORMANCE

The performance of more than twenty UHF and L-Band oscillators has been evaluated. Figure 7 shows the measured phase noise level of a typical UHF oscillator, and Fig. 8 shows the phase noise of an L-Band version. These measurements were made on pairs of oscillators and 3 dB was subtracted to approximate the noise level of a single oscillator. The L-Band oscillators have a 6 dB higher noise level than the UHF version due to the frequency doubling. These phase noise levels are exceptionally good, but are not quite as low as was reported in Ref. 3. The new 500 MHz oscillators have essentially the same noise floor as reported in Ref. 3, but for frequencies below 100 kHz offset the new oscillators are typically 4 to 7 dB noisier. This is caused primarily by the reduced loaded Q's of the SAW resonators, which are a direct consequence of the need to

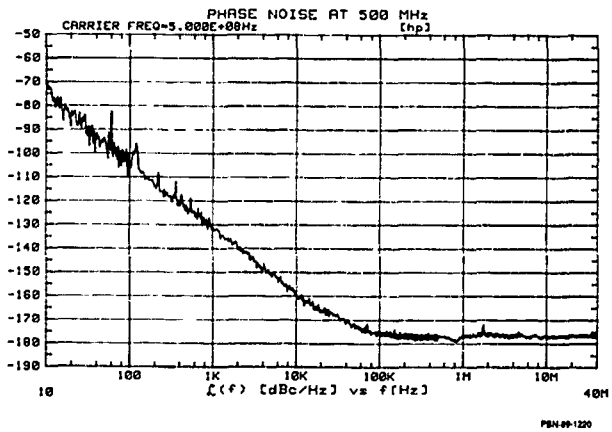


Figure 7. Phase noise level of a typical UHF oscillator.

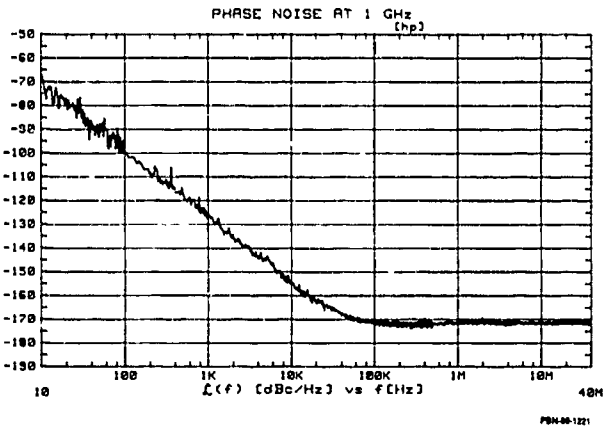


Figure 8. Phase noise level of a typical L-Band oscillator.

series tune the resonators for lower insertion loss. Table 1 summarizes the measured phase noise levels from a sample of twenty oscillators (ten pairs). The noise levels represent a single oscillator at a nominal frequency of 500 MHz (the data from the L-Band oscillators was corrected downward by 6 dB). Since the data was taken on pairs of oscillators the standard deviations and max. to min. values are somewhat smaller than would have been obtained from measurements on individual oscillators. The variations in phase noise levels between oscillators were roughly a factor of two larger in the close-to-carrier flicker noise region than for the noise floor.

TABLE 1

Summary of Phase Noise Levels (500 MHz)

f	Avg. $\mathcal{L}(f)$	Standard Deviation	Max. to Min.
10 Hz	-75.6 dBc/Hz	1.6 dB	4 dB
100 Hz	-105.7	1.3	4
1kHz	-132.6	1.2	3
10kHz	-159.6	1.3	4
100kHz	-174.1	1.6	5
1MHz	-176.3	0.7	2
10MHz	-176.8	0.6	2

The L-Band oscillators were found to have an AM noise floor of less than -170 dBc/Hz (starting at 10 kHz) and an AM noise level of less than -140 dBc/Hz at 10 Hz offset. The AM noise of the UHF oscillators was not accurately measured, but it is expected to be similar to that of the L-Band oscillators.

Figure 9 shows a typical frequency versus voltage tuning curve. All of the oscillators had a useable tuning range of about +/-50 PPM over the voltage range of approximately 2 to 8 volts. Generally the oscillators could be tuned over a larger range but some performance degradation would occur. The center of the useable tuning range was not always at the specified frequency, so some tuning range was used up due to set on errors. However, even with set on errors, all of the oscillators had at least 25 PPM of useable tuning range (either up or down), which was sufficient to allow these oscillators to remain phase locked to a stable reference under all conditions (i.e. frequency variations with temperature, long-term frequency drift, voltage pushing, etc.).

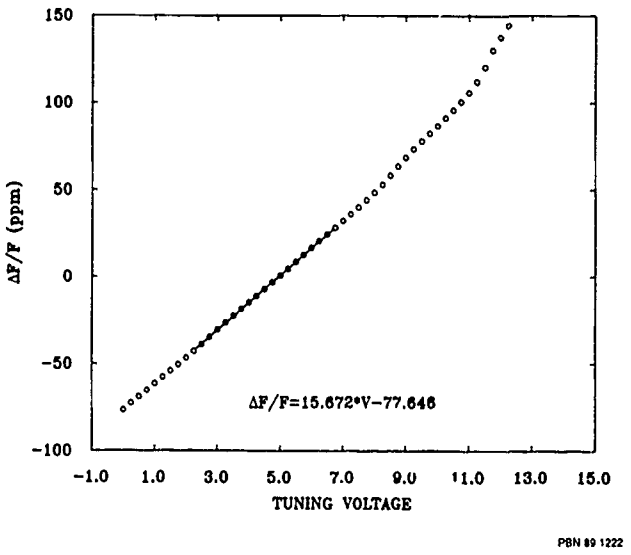


Figure 9. Fractional frequency change versus tuning voltage.

Figure 10 shows the variation in frequency and output power as a function of temperature for one of the UHF oscillators. A complete cycle over the range of -55°C to +85°C is shown. Though only the SAW resonator was temperature controlled, the total fractional frequency variation was less than +/- 4 PPM. The variation in output power was less than 3 dB. The heater module was set to maintain the SAW device at a temperature of approximately 66°C, and therefore at temperatures above 66°C the heater would shut off. The cut of quartz for

the SAW device was chosen so that the turn-over point of the SAW resonator was at about 80°C. Thus at temperatures above 66°C the oscillator frequency would follow the natural frequency-temperature characteristic of the SAW device. The lack of retrace in the frequency curve is a result of a temperature lag between the temperature sensor on the outside of the oscillator and the SAW device on the inside. At room temperature the 28 volt heater draws approximately 250 mA, and . -55°C it draws about 700 mA.

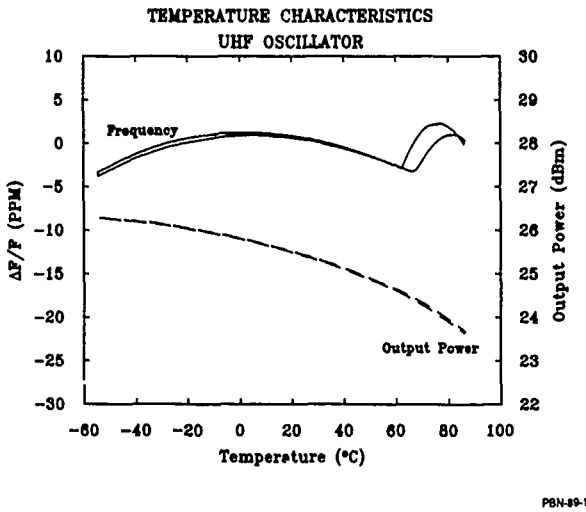


Figure 10. Frequency and power variations as a function of temperature for a UHF oscillator.

Figure 11 shows the temperature dependence of frequency and output power for an L-Band oscillator. The frequency variation is somewhat larger than that of the UHF oscillator but, the output power is significantly more stable. The reason for the increased stability of the output power is that in the L-Band version one of the buffer amplifiers is driven into gain compression. The frequency deviation with temperature varied from oscillator to oscillator, and also depended on the value of the tuning voltage. The difference between the frequency deviations for the oscillators in Fig. 10 and Fig. 11 is well within the normal variation from oscillator to oscillator and does not reflect any fundamental difference between the UHF and L-Band versions. The largest observed frequency deviation with temperature was +/- 15 PPM. Heater modules with less than 28 watts capacity did not control the frequency as well and were not always capable of maintaining a regulated temperature at very low ambient temperatures. At an ambient temperature of 52°C the maximum transistor junction temperature of the rf amplifiers was found to be less than 125°C.

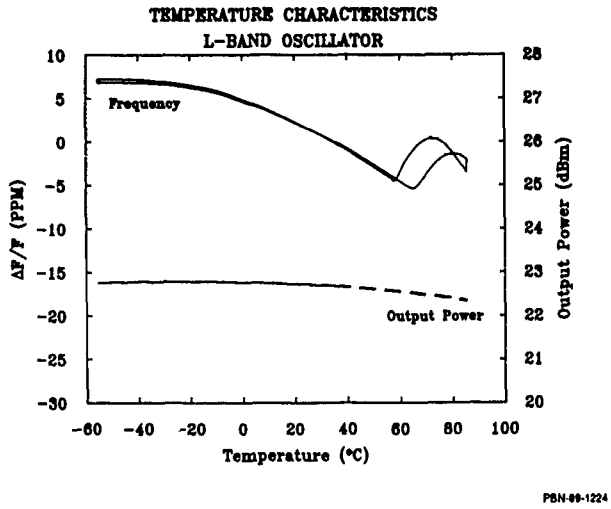


Figure 11. Frequency and power variations as a function of temperature for an L-Band oscillator.

Figures 12 and 13 show the warm-up characteristics of a UHF oscillator for a cold start from 25°C and -54°C, respectively. From 25°C the oscillator is within locking range in a little over 1 minute and from -54°C it takes about 5.5 minutes.

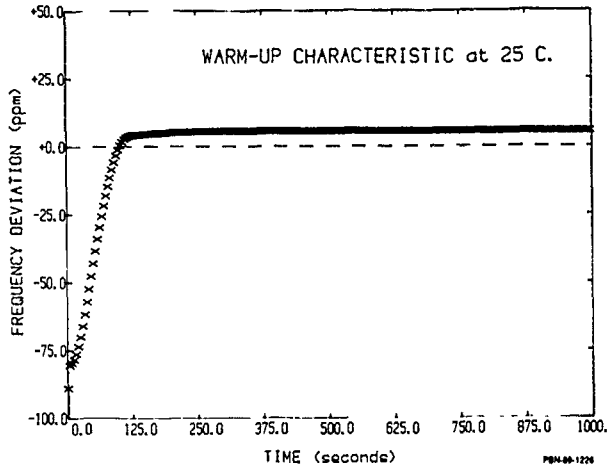


Figure 12. Fractional frequency change versus time for a cold start from 25°C.

Although these oscillators were not intended for use in a high vibration environment, the vibration sensitivity of the oscillators was still of interest, however, because of their very low phase noise levels. Figure 14 shows the measured vibration sensitivity on all three perpendicular axes for one of the UHF oscillators. Generally, γ_1 (normal to the plane of the circuit board) was the most sensitive axis. At frequencies

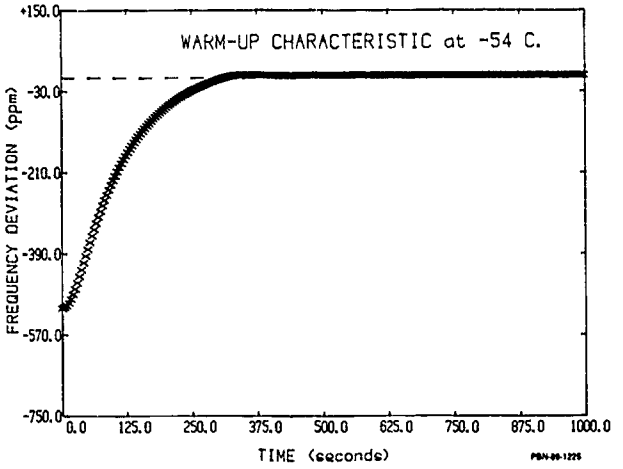


Figure 13. Fractional frequency change versus time for a cold start from -54 °C.

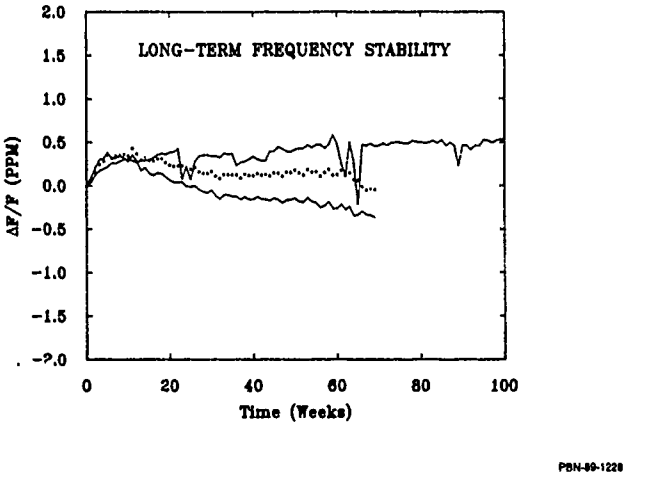


Figure 15. Long-term frequency stability of three oscillators.

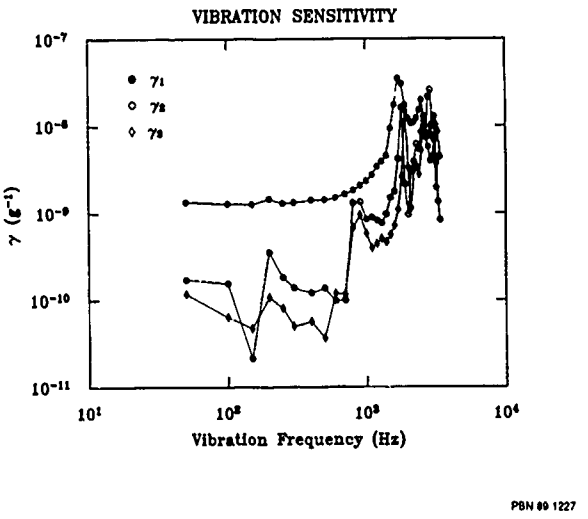


Figure 14. Vibration sensitivity on all three perpendicular axes for a UHF oscillator. γ_1 is normal to the plane of the circuit board.

above 1 kHz resonances in the oscillator chassis and even the vibration test fixture degraded the vibration sensitivity. The average vibration sensitivity (below 1 kHz) of the most sensitive axis for these oscillators was about 2×10^{-9}/g. This was not a problem for our applications.

The long-term frequency stability of these oscillators is also important and therefore aging studies are being carried out. Three oscillators using the same SAW devices, loop amplifiers and power dividers as used in the prototype oscillators, but not the fixed or electronic phase shifters, have been in an aging test for 70 to 100 weeks. The results of these tests are shown in Fig. 15. It does not appear that long-term frequency stability is going to be a problem. Five L-Band oscillators are now being prepared for an aging test which will start in the near future. This test will include both temperature and on-off cycling.

Table 2 provides a statistical summary of many of the performance characteristics for more than twenty of the UHF and L-Band oscillators. This summary includes: tuning sensitivity, output power, harmonic and non-harmonic spur levels, vibration sensitivity, load pulling and voltage pushing and variations of frequency and output power level with temperature. A statistical summary of phase noise levels was presented in Table 1. Since the UHF oscillator design includes a low-pass filter it is not surprising that this version has a much lower level of harmonic spurs than the L-Band version. For the ten UHF oscillators the worst case spur level observed was -78 dBc, whereas for the L-Band version the worst case spur was only -7.5 dBc. More typical values for the highest harmonic spur in individual oscillators were -83 dBc and -14 dBc, respectively. Non-harmonic spurs were all power line related and were always more than 60 dB below the carrier. However, this level is directly dependent on the quality of the power supply being used. Load pulling and voltage pushing were measured on a small

TABLE 2

Statistical Summary of 20 Oscillators

Category	Osc. Type	Avg.	Std. Dev.	Max.	Min.
Frac. Freq. Dev. (-55^{o} to $+85^{o}$)	Both	$\pm$8.4 PPM	2.5 PPM	$\pm$15 PPM	$\pm$3 PPM
Freq. Tuning Slope	Both	15.3 PPM/V	2.0 PPM/V	19 PPM/V	11.1 PPM/V
Vibration Sens. (worst axis)	Both	1.8×10^{-9}/g	0.8×10^{-9}/g	3×10^{-9}/g	0.5×10^{-9}/g
Load Pulling* (2:1 VSWR, any phase)	UHF	~$\pm$3 PPM			
	L-Band	~$\pm$0.2 PPM			
Voltage Pushing*	Both	-3 PPM/V			
Output Power (at 25^{o}C)	UHF	25.0 dBm	0.4 dB	25.6 dBm	24.3 dBm
	L-Band	22.7 dBm	0.4 dB	23.3 dBm	22.0 dBm
Output Power Var. (-55^{o}C to 85^{o}C)	UHF	2.5 dB	0.3 dB	3.1 dB	2.0 dB
	L-Band	0.9 dB	0.4 dB	1.3 dB	0.4 dB
Highest Spur (Harmonic/ Subharmonic)	UHF	-83.4 dBc	2.6 dB	-78 dBc	-87 dBc
	L-Band	-13.9 dBc	4.4 dB	-7.5 dBc	-23 dBc
Highest Spur** (Non-harmonic)	Both	<-60 dBc			

* Measured on only one or two oscillators

** Power supply dependent

number of oscillators. Both the UHF and L-Band versions showed voltage pushing values on the order of 3 PPM/volt. Load pulling on the L-Band oscillators is about +/-0.2 PPM for a 2:1 VSWR at any phase. For the UHF oscillators the value is approximately +/-3 PPM. The L-Band oscillators are less sensitive to load variations because of the greater isolation between the output load and the oscillator loop.

SUMMARY

The oscillators discussed in this paper establish a new state-of-the-art for low noise SAW oscillators. At 500 MHz they typically exhibit a noise floor of -177 dBc/Hz and a flicker noise level of -132 dBc/Hz at 1 kHz offset from the carrier. Furthermore, they also meet all of the other requirements for use in many real system applications.

REFERENCES

[1] M. J. Loboda, T. E. Parker, J. A. Greer, and G. K. Montress, "Reduction of Close-to-Carrier Phase Noise in Surface Acoustic Wave Resonators," in Proceedings of the IEEE Ultrasonics Symposium, 1987, Vol 1, pp.43-46.

[2] G. K. Montress, T. E. Parker, and M. J. Loboda, "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," in Proceedings of the IEEE Ultrasonics Symposium, 1987, Vol 1, pp. 47-52.

[3] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low Phase Noise SAW Resonators and Oscillators: Design and Performance," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol UFFC-35, No. 6, pp. 657-667, November 1988.

[4] T. E. Parker and G. K. Montress, "Precision Surface Acoustic-Wave (SAW) Oscillators, "IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-35, No. 3, pp. 342-364, May 1988.

[5] T. E. Parker, J. Callerame, and G. K. Montress, "A New All Quartz Package for SAW Devices," in Proceedings of the 39th Annual Symposium on Frequency Control, 1985, pp. 519-525.

A New Generalized Modeling of SAW Transducers and Gratings *

P.V. Wright†
RF Monolithics
4441 Sigma Road
Dallas TX 75244

Abstract

A new, general Coupling-Of-Modes (COM) analysis of SAW transducers and gratings is presented. The analysis is applicable to the complete continuum of SAW devices from transducers with or without internal reflections, to grating reflectors comprised of grooves or metallic stripes. The analysis incorporates several important phenomena not previously included in analyses of this type. Both the effects of finite-electrode resistivity and spatially-varying propagation loss are accounted for. In addition, the analysis is sufficiently general to permit the modeling of complex devices on relatively arbitrary crystal orientations. Three coupled, first-order differential equations are obtained relating the forward and reverse propagating acoustic-wave amplitudes in the device and the current on the busbars. These equations lead directly to simple expressions for some very fundamental characteristics of SAW devices. Namely, i) the difference in SAW velocity under isolated- and shorted-stripe gratings, ii) the difference in SAW reflectivity in the latter, and iii) the maximum propagation loss under a SAW metallic grating

1 Introduction

A Single-Phase Unidirectional Transducer (SPUDT) relies on internal reflections to achieve a unidirectional response. The presence of internal reflections makes the impulse response of the device theoretically infinite. Thus, Finite Impulse Response (FIR) techniques, widely employed for the analysis and design of traditional "high-loss" filters, are not applicable to SPUDT's. Traditionally, the analysis of SAW devices with internal reflections, or in which the effects of finite electrode resistivity are important, has been performed using lumped-element analyses. Compared to an FIR analysis the latter are extremely time consuming. In addition, they give little insight into the operation of the device since analytic solutions are not obtained.

By contrast, COM is an ideal candidate for analyzing structures with small distributed reflections and has been employed very successfully for many years for the analysis of microwave and optical devices. An important advantage of a COM analysis over a lumped-element analysis is the analytic nature of the COM equations describing the system. These express the interactions of the various system parameters in a concise fashion that can frequently lead to a much better understanding of the operation of the device. In many cases, relatively simple analytic solutions can be obtained for the scattering parameters of the device. In others, approximate solutions can frequently be obtained which are also very efficient to evaluate.

SPUDTs can be complex structures employing two levels of metallization and/or propagation on special crystal orientations. As a result, electrode resistivity and viscous losses can be very important to an accurate analysis. In addition, the correct relative phase of transduction and reflection within the device is critical to its operation. A very general COM analysis was therefore required to analyze such structures. Though developed primarily for the analysis and design of SPUDT devices, the COM analysis has found most application to-date for resonator and coupled-resonator structures. Agreement between theory and experiment is generally found to be very good and the analysis is extremely efficient. It has been used extensively in the design and development of several new, novel resonator and coupled-resonator structures described in an accompanying paper [1].

*This work was supported by National Science Foundation Grant ECS-8116654

†Present address: Etudes et Productions Schlumberger, 26 rue de la Cavee, Clamart 92140, France

CH2690-6/89/0000-596 $1.00 C 1989 IEEE

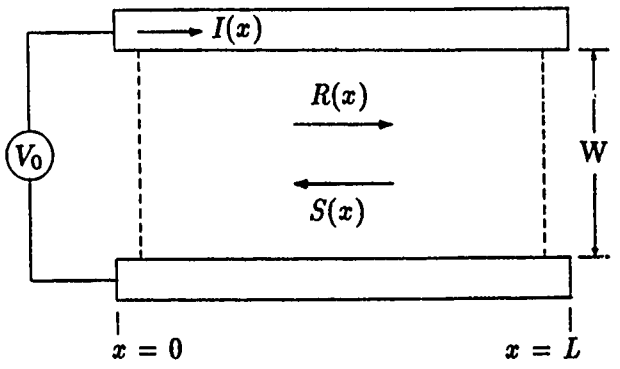


Figure 1: SAW device schematic and wave variables.

2 Theory

The COM equations will be derived for a SAW transducer with internal reflections. A SAW transducer is a three-port device with one electric and two acoustic ports. In contrast, a SAW grating has only two acoustic ports and no electrical port. However, a grating may be regarded as a limiting case of SAW transducer in which the electrical driving-source impedance is either zero or infinity. The COM equations derived for a generalized SAW transducer may therefore be used in limiting form to analyze a SAW grating.

For accurate modeling of a transducer, all device parameters capable of significantly affecting the response must be included. The COM equations will therefore be derived in a form that is sufficiently general to model a device with the following characteristics:

- Spatially-varying transduction of arbitrary magnitude and phase.
- Spatially-varying internal reflections of arbitrary magnitude and phase.
- Spatially-varying separation of the centres of transduction and reflection.
- Spatially-varying SAW velocity.
- Spatially-varying SAW attenuation.
- Spatially-varying electrode resistivity.
- Spatially-varying electrode capacitance.
- Arbitrary termination of the electrical port.

The SAW device will be assumed to occupy the region from $x = 0$ to $x = L$, as shown in Fig. 1. The beamwidth of the device is W. The forward- (i.e. right) propagating acoustic wave is denoted by $R(x)$, while the reverse- (i.e. left) propagating acoustic wave is denoted by $S(x)$. In addition, the spatially-varying electrical current on the busbars is $I(x)$. Busbar resistance will be neglected in the analysis. Thus, a spatially-independent voltage difference V_0 will be assumed between the two busbars. Electrical connection to the busbar occurs at $x = 0$.

A COM analysis is concerned with the perturbation of the freely-propagating wave solutions resulting from a distributed coupling between the waves. For the general case to be considered there are two counter-propagating acoustic waves $R(x)$ and $S(x)$, and the busbar electrical current $I(x)$. In order to be able to concentrate only on the perturbations of the waves and understand their origins, the equations are best derived in terms of normalized wave variables. We define

$$R(x,t) = \tilde{R}(x)e^{j\omega t - j\int_{x_o}^{x} dx_1\, k(x_1)} \tag{1}$$

$$S(x,t) = \tilde{S}(x)e^{j\omega t + j\int_{x_o}^{x} dx_1\, k(x_1)} \tag{2}$$

where x_o is an arbitrary reference plane and the complex, spatially-dependent SAW propagation constant is given by

$$k(x) = k_R(x) - j\gamma(x) \tag{3}$$

with

$$k_R(x) = \omega/v_R(x) = 2\pi/\lambda_R(x) \tag{4}$$

where $\gamma(x)$ is the spatially-dependent SAW attenuation. In (4), ω is the angular frequency, $v_R(x)$ is the SAW velocity and $\lambda_R(x)$ is the SAW wavelength.

$\tilde{R}(x)$ and $\tilde{S}(x)$ as defined by (1) and (2) are the slowly-varying perturbation wave amplitudes. In general they will be complex. In the absence of any device-induced interactions both $\tilde{R}(x)$ and $\tilde{S}(x)$ will be constants. The COM equations will be derived in terms of these perturbation variables. In (1) and (2), the integrals are required for the spatial phase terms to account for a spatially-varying SAW velocity in the device.

In the following section we begin by deriving the COM equations for a SAW grating in the absence of transduction.

2.1 COM Equations for a SAW Grating

A SAW grating consists of an array of periodic, or quasi-periodic, reflective elements on a crystalline substrate. These reflective elements may be grooves,

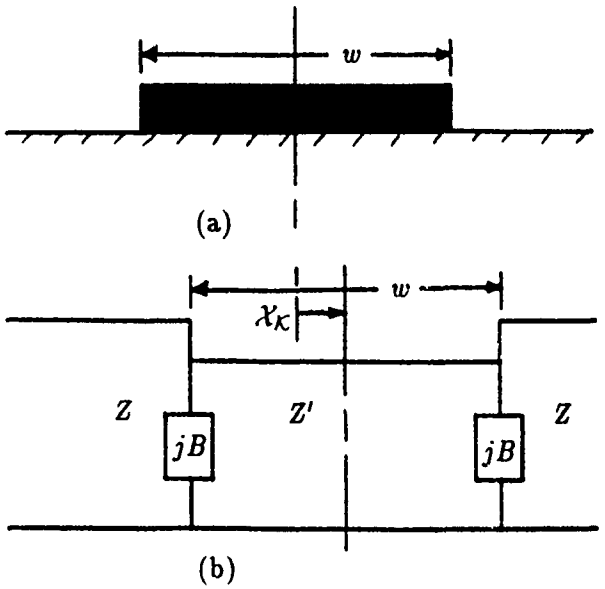


Figure 2: SAW reflection from an isolated stripe, (a) isolated reflective stripe, (b) equivalent-circuit on an arbitrary crystal orientation.

conducting or non-conducting stripe overlays, or indiffused regions etc. The precise nature of the reflective elements need not be considered when deriving the general form of the COM equations. This is one of the major advantages of a COM analysis. The exact details of the reflective elements need only be considered when determining the COM parameters for a particular configuration. It is this generality that makes COM such a powerful technique.

As in any grating structure, the total grating response is the product of an "element factor" and an "array factor" [2]. The element factor characterizes the properties of the individual reflective elements, whereas the array factor depends upon the periodicity of the grating. We consider first, the element factor of a SAW reflective grating on an arbitrary crystal orientation.

In Fig. 2 an isolated SAW stripe of width w, is shown together with an equivalent-circuit model. At the edges of the stripe, lumped susceptances jB are included to account for energy storage [3], [4], [5]. Note that in Fig. 2, the impedance discontinuities in the lumped-element model are not shown spatially aligned with the physical discontinuities of the stripe. This is an important generality of the model. For propagation on conventional, highly-symmetric crystal orientations the physical and model discontinuities should generally be aligned. However, on asymmetric or "Natural" SPUDT (NSPUDT) orientations

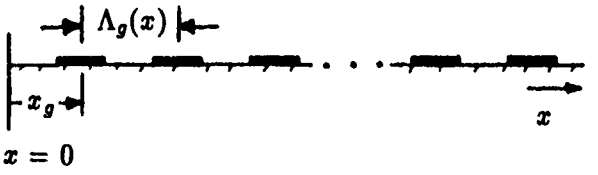


Figure 3: SAW grating array of reflective stripes.

this is not the case.

On a symmetric crystal orientation, if the reflection coefficient from the front edge of a stripe discontinuity is Γ_e, by reciprocity, the reflection from the back edge of the stripe is $-\Gamma_e^*$, where "$*$" denotes the complex conjugate. The net reflection coefficient of the stripe, referenced to the centre of the stripe, for low attenuation, is thus

$$\Gamma_s \approx 2j\,|\Re(\Gamma_e)\sin\theta + \Im(\Gamma_e)\cos\theta| \qquad (5)$$

where $\theta = k_R(x)w$, w is the stripe width, and $\Re$ and $\Im$ denote real and imaginary parts respectively. Hence, on a symmetric crystal orientation, the reflection coefficient is pure imaginary at the centre of the stripe, regardless of the stripe width. The locations in a grating at which the local reflection coefficient is pure imaginary will be defined as "centres of reflection."

On asymmetric crystal orientations, by contrast, the stripe reflection coefficient is not generally pure imaginary at the centre of the stripe. However, such a location can always be found at a displaced location. The equivalent-circuit model can therefore be generalized to model asymmetric crystal propagation, simply by introducing a spatial offset between the physical stripe location and the corresponding impedance discontinuities in the model, as in Fig. 2. If this displacement is $\mathcal{X}_K(x)$ in the $+x$ direction, the reflection coefficients of the forward and reverse propagating waves, referenced to the centre of the stripe, are respectively

$$\Gamma_R = e^{+2jk(x)\mathcal{X}_K(x)}\,\Gamma_s \qquad (6)$$

$$\Gamma_S = e^{-2jk(x)\mathcal{X}_K(x)}\,\Gamma_s \qquad (7)$$

In COM, the discrete nature of the stripes in a grating is ignored. Instead, the coupling between counter-propagating waves is viewed as the result of a continuous, distributed process. The element factor for the coupling is also usually approximated by its value at synchronism. As a result, in place of the discrete stripe reflection coefficient Γ_s, an equivalent,

distributed, coupling coefficient $\mathcal{K}(x)$ is defined for the COM equations as

$$\mathcal{K}(x) = 2\,|\Re(\Gamma_e)\sin\theta + \Im(\Gamma_e)\cos\theta|\,/\Lambda_g(x) \qquad (8)$$

evaluated at synchronism, where $\Lambda_g(x)$ is the local grating period (Fig. 3). Note, that $\mathcal{K}(x)$ as defined here represents only the spatially-dependent element factor of the grating. We must now consider the grating array factor.

The coupling between counter-propagating acoustic waves in a SAW grating results from the periodic perturbation of the waves by the grating. For significant scattering to occur, the frequency must be close to synchronism. At synchronism, the phase matching condition

$$k_R(x) - k_{grating}(x) = -k_R(x) \qquad (9)$$

is satisfied, where

$$k_{grating}(x) = 2\pi/\Lambda_g(x) \qquad (10)$$

Energy is then coherently scattered between the counter-propagating acoustic waves. Only the Fourier component of the grating that results in synchronous, or near synchronous, scattering is important when determining the array factor. From (3), (4), (9) and (10), at synchronism

$$\omega_0/v_R(x) = \pi/\Lambda_g(x) \qquad (11)$$

The location of the first grating stripe will be taken at $x = x_g$, (Fig. 3), rather than at $x = 0$. This generality is useful when considering grating/transducer combinations and adds flexibility for varying the phase origin of the equations. The full COM equations for a grating are derived by considering scattering from an incremental section. In the equations, there are two dephasing terms that arise

$$\begin{aligned}\varphi_{\mathcal{K}}(x) &= k(x)|_{synch.}\mathcal{X}_{\mathcal{K}}(x) + \int_0^{x_g} dx\,\pi/\Lambda_g(x) \\ &\approx [\pi/\Lambda_g(x)]\,\mathcal{X}_{\mathcal{K}}(x) + [\pi/\Lambda_g(x_g)]\,x_g \qquad (12)\end{aligned}$$

and

$$\phi_g(x) = \int_0^x dx_1\,[k(x_1) - \pi/\Lambda_g(x_1)] \qquad (13)$$

The first term $\varphi_{\mathcal{K}}(x)$ is frequency independent and simply expresses the fact that the phase of the grating Fourier component is not necessarily zero at $x = 0$. The second term $\phi_g(x)$ is the phase mismatch that arises in the scattered waves away from synchronism.

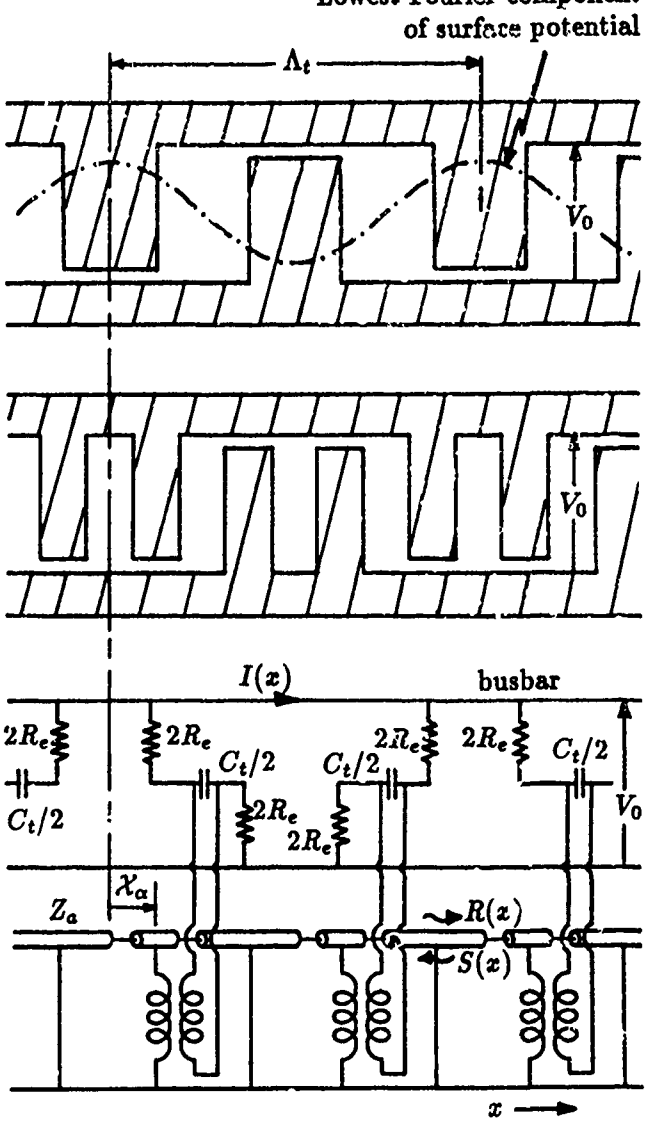


Figure 4: Schematics of two- and four-electrodes-per-wavelength transducers, and associated equivalent-circuit model ignoring electrode reflections.

This term is strongly frequency dependent. With these phase terms, the COM equations for a SAW grating on an arbitrary crystal orientation are

$$\frac{d\tilde{R}(x)}{dx} = j\mathcal{K}(x)\,e^{+2j\varphi_{\mathcal{K}}(x)}\,e^{+2j\phi_g(x)}\,\tilde{S}(x) \qquad (14)$$

$$\frac{d\tilde{S}(x)}{dx} = -j\mathcal{K}(x)\,e^{-2j\varphi_{\mathcal{K}}(x)}\,e^{-2j\phi_g(x)}\,\tilde{R}(x) \qquad (15)$$

2.2 COM Equations for a Generalized Transducer

In the previous section, the effects of transduction were completely ignored. Transduction will now be

introduced into the analysis to obtain the full COM equations for a generalized transducer with internal reflections. Again, propagation will be assumed on an arbitrary asymmetric crystal orientation. The effects of both substrate propagation loss and finite electrode resistivity will be considered.

Fig. 4 shows schematics of a two- and a four-electrodes-per-wavelength transducer for the purposes of the analysis. (Note, the analysis is valid for all SAW transducers with an arbitrary number of electrodes per wavelength). At the bottom of Fig. 4, a one-dimensional equivalent-circuit model is shown ignoring electrode reflections. This model, together with the grating analysis of the previous section, form the basis of the COM equations.

In the equivalent-circuit model, the capacitance between adjacent electrodes of opposite polarity is denoted by $C_t/2$. The resistors of value $2R_e$ are included for two purposes:

1. They permit the effects of finite electrode resistivity to be incorporated in an approximate fashion.

2. In the limit $R_e \rightarrow \infty$, they permit the properties of isolated-stripe gratings to be modeled.

The resistors only allow the effects of finite electrode resistivity to be modeled in an approximate fashion, as the model is only one-dimensional. However, a detailed electrostatic analysis indicates that this simple model is probably adequate for typical electrode resistivity values [6]. Referring to Fig. 4, R_e is the "effective" resistance of an electrode or split-finger pair in the transducer. From the electrostatic modeling, and from power dissipation arguments [7], the optimum value of R_e for modeling the effects of finite electrode resistivity is

$$R_e = (1/3)\,\rho_e W \tag{16}$$

where ρ_e is the resistance of an electrode, or split-finger pair, per unit width, and W is the beamwidth of the transducer.

In the equivalent-circuit model of Fig. 4, the coupling between the electrical drive and the counter-propagating acoustic waves occurs via transformers periodically located on the acoustic transmission line. Each transformer radiates and couples symmetrically to waves propagating in the $+x$ and $-x$ directions. The lowest Fourier component of the surface potential on the electrodes is shown at the top of Fig. 4. It is this component of the surface potential that results in the transduction of acoustic energy. Note, that the transformers are not shown spatially aligned with either the zeros or the extremes of the potential, as is customarily done. Instead a relative displacement of $\mathcal{X}_\alpha(x)$ is assumed between the two. The reason, as in the SAW grating analysis, is that on asymmetric crystal orientations a more complex situation can exist. The centres of transduction will be defined as those locations in the transducer at which counter-propagating acoustic waves are launched in-phase (i.e. the transformer locations in Fig. 4). On arbitrary crystal orientations, the centres of transduction are not restricted to being coincident with the centres of the electrodes, or gaps. Hence, the need for the offset $\mathcal{X}_\alpha(x)$ in the model.

As with the grating analysis, the COM equations describing transduction require two dephasing terms to be defined. These are

$$\begin{aligned}\varphi_\alpha(x) &= k(x)|_{synch.}\mathcal{X}_\alpha(x) + \int_0^{x_t} dx\, 2\pi/\Lambda_t(x) \\ &\approx [2\pi/\Lambda_t(x)]\,\mathcal{X}_\alpha(x) + [2\pi/\Lambda_t(x_t)]\,x_t \end{aligned} \tag{17}$$

where x_t is the location of the centre of the first positive electrode, or electrode pair, and

$$\phi_t(x) = \int_0^x dx_1\,[k(x_1) - 2\pi/\Lambda_t(x_1)] \tag{18}$$

These are completely analogous with (12) and (13) for reflective coupling in a grating. As in the latter case, $\varphi_\alpha(x)$ is frequency independent and $\phi_t(x)$ is strongly frequency dependent.

The COM equations, in the absence of electrode reflections, are derived by considering the electro-acoustic coupling of an incremental equivalent-circuit model exactly analogous to that shown in Fig. 4. Again, an incremental model must be used for deriving the COM equations, since they approximate the coupling as being continuous rather than discrete. The equations governing an incremental section are found to be

$$\begin{aligned}\frac{d\tilde{R}(x)}{dx} &= j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\phi_t(x)} \\ &\quad \cdot \left[V_0 + 2R_e(x)\Lambda_t(x)\frac{dI(x)}{dx}\right] \end{aligned} \tag{19}$$

$$\begin{aligned}\frac{d\tilde{S}(x)}{dx} &= -j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\phi_t(x)} \\ &\quad \cdot \left[V_0 + 2R_e(x)\Lambda_t(x)\frac{dI(x)}{dx}\right] \end{aligned} \tag{20}$$

$$\frac{dI(x)}{dx} = -\frac{1}{[2R_e(x)+1/j\omega C_t(x)]\,\Lambda_t(x)}V_0 + \beta_R(x)\tilde{R}(x) + \beta_S(x)\tilde{S}(x) \qquad (21)$$

where V_0 is the voltage on the busbars and $\alpha(x)$, $\beta_R(x)$ and $\beta_S(x)$ are acousto-electric coupling coefficients. (Note, in (21), $I(x)$ is also a perturbation variable. It is the current on the busbars normalized by the implicit time dependence $\exp(j\omega t)$). The two coupling coefficients $\beta_R(x)$ and $\beta_S(x)$ might be thought to be independent variables. However, a reciprocity analysis reveals that this is not the case. For the transducer to satisfy reciprocity, we must have

$$\beta_R(x) = \frac{2j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)} \qquad (22)$$

$$\beta_S(x) = \frac{2j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)} \qquad (23)$$

If we introduce (22) and (23) into (19) - (21), and also include the coupling terms from (14) and (15) to include the possibility of internal reflections in the transducer, the COM equations become

$$\frac{d\tilde{R}(x)}{dx} + \frac{4\alpha^2(x)R_e(x)\Lambda_t(x)}{1+2j\omega C_t(x)R_e(x)}\tilde{R}(x) = jK(x)\,e^{+2j\varphi_K(x)}\,e^{+2j\phi_g(x)}\,\tilde{S}(x) + \frac{j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}V_0 - \frac{4\alpha^2(x)R_e(x)\Lambda_t(x)\,e^{+2j\varphi_\alpha}\,e^{+2j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\tilde{S}(x) \qquad (24)$$

$$\frac{d\tilde{S}(x)}{dx} - \frac{4\alpha^2(x)R_e(x)\Lambda_t(x)}{1+2j\omega C_t(x)R_e(x)}\tilde{S}(x) = -jK(x)\,e^{-2j\varphi_K(x)}\,e^{-2j\phi_g(x)}\,\tilde{R}(x) - \frac{j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}V_0 + \frac{4\alpha^2(x)R_e(x)\Lambda_t(x)\,e^{-2j\varphi_\alpha}\,e^{-2j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\tilde{R}(x) \qquad (25)$$

$$\frac{dI(x)}{dx} = -\frac{j\omega C_t(x)/\Lambda_t(x)}{1+2j\omega C_t(x)R_e(x)}V_0 + \frac{2j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\tilde{R}(x) + \frac{2j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\phi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\tilde{S}(x) \qquad (26)$$

The form of these equations is very revealing. We observe from (24) and (25) that even in the absence of a voltage on the busbars (i.e. $V_0 = 0$), and in the absence of counter-propagating waves, the derivatives of $\tilde{R}(x)$ and $\tilde{S}(x)$ are both non-zero. This is a result of the existence of the second term on the left-hand side of the equations. The implication is that $\tilde{R}(x)$ and $\tilde{S}(x)$ are not the "natural" perturbation variables for the transducer. The assumption in their definition in (1) and (2) was that the uncoupled surface waves in the transducer propagate with the same propagation constant $k(x)$ as on the free surface. The COM equations (24) and (25) now reveal that this assumption for the general case considered is incorrect. The propagation constant of the uncoupled-wave solutions under the transducer is in fact modified by the presence of the transducer. To remedy this oversight, we now rewrite the equations in terms of the new acoustic perturbation wave amplitudes $\hat{R}(x)$ and $\hat{S}(x)$, where

$$R(x,t) = \hat{R}(x)e^{j\omega t - j\int_{x_o}^{x} dx_1\, k_e(x_1)} \qquad (27)$$

$$S(x,t) = \hat{S}(x)e^{j\omega t + j\int_{x_o}^{x} dx_1\, k_e(x_1)} \qquad (28)$$

and $k_e(x)$ is the modified, complex, effective SAW propagation constant under the transducer/grating structure defined by

$$k_e(x) = k_{Re}(x) - j\gamma_e(x) \qquad (29)$$

To eliminate the second term on the left-hand sides of (24) and (25), requires

$$k_e(x) = k(x) - \frac{4j\alpha^2(x)R_e\Lambda_t(x)}{1+2j\omega C_t(x)R_e(x)} \qquad (30)$$

Hence, from (29) and (30)

$$k_{Re}(x) = k_R(x) - \frac{8\alpha^2(x)\omega C_t(x)R_e^2(x)\Lambda_t(x)}{1+(2\omega C_t(x)R_e(x))^2} \qquad (31)$$

and

$$\gamma_e(x) = \gamma(x) + \frac{4\alpha^2(x)R_e(x)\Lambda_t(x)}{1+(2\omega C_t(x)R_e(x))^2} \qquad (32)$$

New grating and transducer dephasing terms analogous to those defined in (13) and (18) respectively, must also be defined in terms of this effective propagation constant. The definitions are

$$\varphi_g(x) = \int_0^x dx_1 \left[k_e(x_1) - \pi/\Lambda_g(x_1)\right] \qquad (33)$$

$$\varphi_t(x) = \int_0^x dx_1 \left[k_e(x_1) - 2\pi/\Lambda_t(x_1)\right] \qquad (34)$$

The COM equations (24) - (26) may then be written in the much simpler form

$$\begin{aligned}\frac{d\hat{R}(x)}{dx} &= j\Bigg[\mathcal{K}(x)\,e^{+2j\varphi_{\mathcal{K}}(x)}\,e^{+2j\varphi_g(x)} \\ &\quad + \frac{4j\alpha^2(x)R_e(x)\Lambda_t(x)\,e^{+2j\varphi_\alpha}\,e^{+2j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\Bigg]\hat{S}(x) \\ &\quad + \frac{j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}V_0\end{aligned} \qquad (35)$$

$$\begin{aligned}\frac{d\hat{S}(x)}{dx} &= -j\Bigg[\mathcal{K}(x)\,e^{-2j\varphi_{\mathcal{K}}(x)}\,e^{-2j\varphi_g(x)} \\ &\quad + \frac{4j\alpha^2(x)R_e(x)\Lambda_t(x)\,e^{-2j\varphi_\alpha}\,e^{-2j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\Bigg]\hat{R}(x) \\ &\quad - \frac{j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}V_0\end{aligned} \qquad (36)$$

$$\begin{aligned}\frac{dI(x)}{dx} &= -\frac{j\omega C_t(x)/\Lambda_t(x)}{1+2j\omega C_t(x)R_e(x)}V_0 \\ &\quad + \frac{2j\alpha(x)\,e^{-j\varphi_\alpha}\,e^{-j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\hat{R}(x) \\ &\quad + \frac{2j\alpha(x)\,e^{+j\varphi_\alpha}\,e^{+j\varphi_t(x)}}{1+2j\omega C_t(x)R_e(x)}\hat{S}(x)\end{aligned} \qquad (37)$$

These are the COM equations for a generalized SAW transducer with internal reflections. The effects of propagation loss and finite electrode resistance are included. Furthermore, they contain no assumptions about crystalline symmetry implicit in many other transducer analyses. In many cases of practical interest, significant simplifications can be made to the equations.

3 Important Results

In this section we show that (30) - (32), together with the COM equations (35) - (37), lead directly to some very simple expressions for some fundamental characteristics of a SAW device. The analytic nature of the equations is of great help in understanding or predicting device properties. As mentioned at the outset, this is a principal advantage of a COM analysis over a lumped-element analysis.

It has long been observed that there can be a significant difference between the velocities and reflection coefficients in isolated- and connected-stripe gratings [8], [9]. On strong coupling piezoelectrics the reflection coefficients can even be of the opposite sign. This can have important consequences for device design, especially for SPUDTs. In addition, it is important to understand how the propagation loss in a SAW transducer or grating is related to the electrode resistivity. Simple algebraic expressions will now be obtained for all these device parameters.

In the limit $R_e(x) \to 0$, from (31) and (32)

$$k_{Re}(x) \to k_R(x) = \omega/v_R(x) \qquad (38)$$

and

$$\gamma_e(x) \to \gamma(x) \qquad (39)$$

This is the case of an interconnected-stripe grating.

In the limit $R_e(x) \to \infty$, the generalized transducer analyzed becomes an open-stripe reflective grating array (see Fig. 4). In this limit, from (31) and (32)

$$k_{Re}(x) \to k_R(x) - \frac{2\alpha^2(x)\Lambda_t(x)}{\omega C_t(x)} \qquad (40)$$

and

$$\gamma_e(x) \to \gamma(x) \qquad (41)$$

Thus, the attenuation coefficient in an interconnected-stripe grating is the same as that in an isolated-stripe grating. This is as expected since the only attenuation mechanism is substrate propagation loss. However, there is a difference in the propagation phase constants in (38) and (40). Defining the velocity under an isolated-stripe grating array as

$$v_{R_0}(x) = v_R(x) + \Delta v(x) \qquad (42)$$

from (38) and (40)

$$\frac{\Delta v(x)}{v_R(x)} = \frac{2\alpha^2(x)v_R(x)\Lambda_t(x)}{\omega^2 C_t(x)} \qquad (43)$$

This is, thus, the fractional change in SAW velocity in an isolated-stripe grating compared to that in a connected-stripe grating.

The maximum propagation loss in a transducer or grating can also be predicted from the equations. Taking the derivative of $\gamma_e(x)$ in (32) with respect to the "effective" electrode resistance $R_e(x)$, and setting it equal to zero, gives

$$2R_e(x) = 1/(\omega C_t(x)) \tag{44}$$

This is the value of electrode resistance which will result in maximum SAW propagation loss in the device. From (32) and (44) the maximum attenuation coefficient is

$$\gamma_e(x)|_{max.} = \gamma(x) + \frac{\alpha^2(x)\Lambda_t(x)}{\omega C_t(x)} \tag{45}$$

Expressions for the difference in reflectivity between isolated- and connected-stripe gratings, may be obtained by considering the limiting forms of the COM equations. In the limit $R_e(x) \to 0$, with $V_0 = 0$, (35) and (36) become

$$\frac{d\hat{R}(x)}{dx} = jK(x)\,e^{+2j\varphi_K(x)}\,e^{+2j\varphi_g(x)}\,\hat{S}(x) \tag{46}$$

$$\frac{d\hat{S}(x)}{dx} = -jK(x)\,e^{-2j\varphi_K(x)}\,e^{-2j\varphi_g(x)}\,\hat{R}(x) \tag{47}$$

However, the limiting form as $R_e(x) \to \infty$, is

$$\frac{d\hat{R}(x)}{dx} = j\left[K(x)\,e^{+2j\varphi_K(x)}\,e^{+2j\varphi_g(x)} + \frac{2\alpha^2(x)\Lambda_t(x)\,e^{+2j\varphi_\alpha}\,e^{+2j\varphi_t(x)}}{\omega C_t(x)}\right]\hat{S}(x) \tag{48}$$

$$\frac{d\hat{S}(x)}{dx} = -j\left[K(x)\,e^{-2j\varphi_K(x)}\,e^{-2j\varphi_g(x)} + \frac{2\alpha^2(x)\Lambda_t(x)\,e^{-2j\varphi_\alpha}\,e^{-2j\varphi_t(x)}}{\omega C_t(x)}\right]\hat{R}(x) \tag{49}$$

In a metallic grating

$$\Lambda_t(x) = 2\Lambda_g(x) \tag{50}$$

hence, from (33) and (34)

$$\varphi_t(x) = \varphi_g(x) \tag{51}$$

Also, for most SAW devices on symmetric crystal orientations

$$\varphi_\alpha = \varphi_K(x) = \text{const.} \tag{52}$$

Thus, the COM equations (48) and (49) for a metallic, isolated-stripe grating on a conventional crystal orientation reduce to

$$\frac{d\hat{R}(x)}{dx} = j\left[K(x) + \frac{2\alpha^2(x)\Lambda_t(x)}{\omega C_t(x)}\right] \cdot e^{+2j\varphi_K}\,e^{+2j\varphi_g(x)}\,\hat{S}(x) \tag{53}$$

$$\frac{d\hat{S}(x)}{dx} = -j\left[K(x) + \frac{2\alpha^2(x)\Lambda_t(x)}{\omega C_t(x)}\right] \cdot e^{-2j\varphi_K}\,e^{-2j\varphi_g(x)}\,\hat{R}(x) \tag{54}$$

Defining the effective, distributed coupling coefficient between counter-propagating acoustic waves in an isolated-stripe grating as

$$K_0(x) = K(x) + \Delta K(x) \tag{55}$$

and comparing (46) and (47) with (53) and (54), gives

$$\Delta K(x) = \frac{2\alpha^2(x)\Lambda_t(x)}{\omega C_t(x)} \tag{56}$$

This is the difference in the acoustic reflection coefficient/wavelength of an isolated- and a connected-stripe grating. Note, that it is entirely attributable to electrical regeneration and is independent of metal thickness.

The above expressions can be evaluated more easily if the transduction coefficient $\alpha(x)$ is expressed in terms of the more familiar electro-mechanical coupling coefficient K^2. In a uniform two-electrodes-per-wavelength transducer, with a metallization ratio of 0.5

$$\alpha^2 = (2/\pi)K^2\omega C_t/\Lambda_t^2 \tag{57}$$

Hence, the fractional change in SAW velocity in an isolated-stripe, two-electrodes-per-wavelength grating, compared to that in a connected-stripe grating, from (43) and (57) is

$$\Delta v/v_R = (2/\pi^2)K^2 \approx 0.41(\Delta v/v)_\infty \tag{58}$$

where $(\Delta v/v)_\infty$ is the fractional change in SAW velocity on a shorted surface, compared to the free surface. Thus, the expected difference in velocity between the two alternate grating configurations is approximately 0.41 of the difference between the open and shorted SAW velocities on the substrate.

From (45) and (57) the maximum attenuation coefficient in a two-electrodes-per-wavelength transducer or grating is

$$\gamma_e(x)|_{max.} = \gamma(x) + (2/\pi)K^2/\Lambda_t(x) \qquad (59)$$

This is for the worse case electrode resistance, as defined by (44).

Finally, the difference in reflectivity between isolated- and connected-stripe two-electrodes-per-wavelength SAW devices from (56) and (57) is

$$\Delta\mathcal{K}(x) = (4/\pi)K^2/\Lambda_t(x) \qquad (60)$$

Thus, for 34°-quartz, with $K^2 = 0.00114$

$$|\Delta\mathcal{K}| = 0.145/\lambda \ \% \qquad (61)$$

where λ is the synchronous wavelength.
For X-$LiTaO_3$-Y112°, with $K^2 = 0.008$

$$|\Delta\mathcal{K}| = 1.02/\lambda \ \% \qquad (62)$$

Experimental measurements of the reflectivity of shorted- and isolated-stripe, two-electrodes-per-wavelength gratings are shown in Figs 5 and 6. The least-squares, straight-line fits to the data yielded the following difference between the two grating types, essentially independent of the metal thickness as predicted by (56).

For 34°-quartz

$$|\Delta\mathcal{K}| = 0.145/\lambda \ \% \qquad (63)$$

and for X-$LiTaO_3$-Y112°

$$|\Delta\mathcal{K}| = 1.09/\lambda \ \% \qquad (64)$$

which are in very good agreement with the theoretical values in (61) and (62). Thus, the experimental data appears to validate the model.

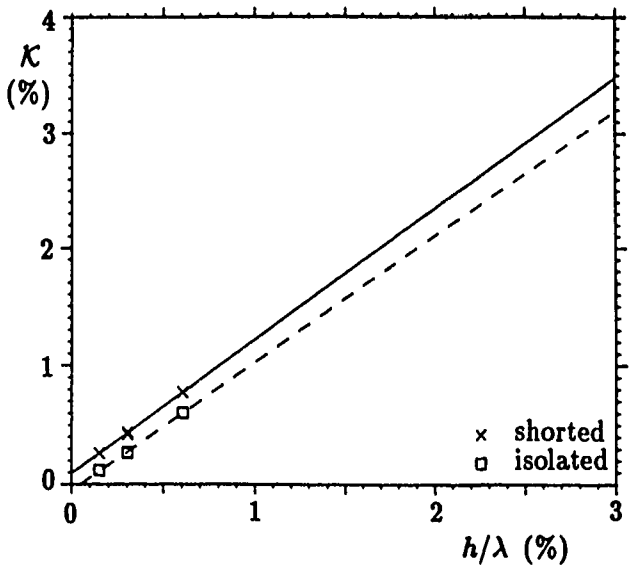


Figure 5: Reflectivity versus metal thickness for isolated- and shorted-stripe two-electrodes-per-wavelength aluminum gratings on 34°-quartz.

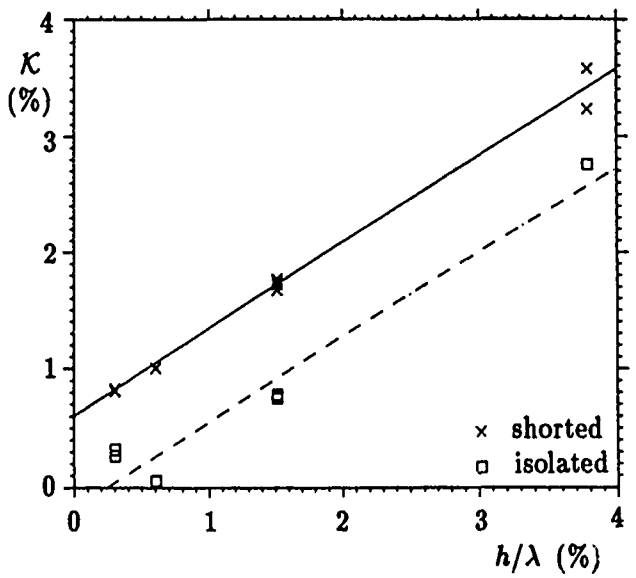


Figure 6: Reflectivity versus metal thickness for isolated- and shorted-stripe two-electrodes-per-wavelength aluminum gratings on X-$LiTaO_3$-Y112°.

4 Conclusion

A new, very general coupled-mode analysis of a SAW transducer with internal reflections has been achieved. In this analysis, all device variables are assumed to be spatially varying and the analysis is valid even for propagation on arbitrary asymmetric crystal orientations. Both electrode resistivity and substrate propagation loss are included in the model. The COM equations relating the amplitude of the counter-propagating acoustic waves in the transducer and the electrical current on the busbar are relatively

compact and lead to simple algebraic expressions for some fundamental SAW device parameters.

For special cases, the COM equations can be solved exactly. In others, approximate, though very exact solutions can be obtained. Finally, for the most general case the equations must be solved numerically. One technique, for example, is to approximate the transducer by discrete uniform sections. In each of the latter, exact solutions to the COM equations may be used. The overall device response is then obtained by parallel/cascading the individual sections. Such a numerical approach is still usually relatively efficient.

The COM equations are particularly useful for the analysis of complex SAW devices such as SPUDTs. However, they have also been used extensively to analyze and design much simpler resonator-type structures [1]. In the latter case, they have the dual advantages of computational efficiency and excellent agreement with experiment. It is hoped that these equations will find increasing application in the design and analysis of SAW devices.

References

[1] P. V. Wright, "Low-cost high-performance resonator and coupled-resonator design: NSPUDT and other innovative structures," *this symposium.*

[2] J. D. Kraus, "Antennas," McGraw-Hill, New York, pp. 66-67, 1950.

[3] R. C. Li. and J. Melngailis, "The influence of stored-energy at step discontinuities on the behavior of surface-wave gratings," *IEEE Trans. Sonics Ultrason.*, vol. SU-22, no. 3, pp. 189-198, May 1975.

[4] P. V. Wright and H. A. Haus, "Theoretical analysis of second-order effects in surface-wave gratings," *Proc. 34th Ann. Symp. Frequency Control*, pp. 262-268, May 1980.

[5] P. V. Wright, "Modeling and experimental measurements of the reflection properties of SAW metallic gratings," *IEEE Ultrason. Symp. Proc.*, pp. 54-63, 1984.

[6] P. V. Wright, "Experimental and theoretical research on an innovative unidirectional surface-acoustic-wave transducer," *Phase II Final Report, National Science Foundation Grant ECS-8116654*, July, 1986.

[7] K. M. Lakin, "Electrode resistance effects in interdigital transducers," *IEEE Trans. Microwave Theory tech.*, vol. MTT-22, pp. 418-424, 1974.

[8] W. H. Haydl, P. Hiesinger, R. S. Smith, B. Dischler and K. Heber, "Design of quartz and lithium niobate SAW resonators using aluminum metallization," *Proc. 30th Ann. Symp. Frequency Control*, pp. 346-357, June 1976.

[9] P. S. Cross, "Properties of reflective arrays for surface acoustic resonators," *IEEE Trans. Sonics Ultrason.*, vol. SU-23, no. 4, pp. 255-262, July 1976.

43rd Annual Symposium on Frequency Control - 1989

SHEAR MODE GRINDING

Norman J. Brown and Baruch A. Fuchs,
Lawrence Livermore National Laboratory*

ABSTRACT

The thesis of this paper is that shear mode grinding of glass 1) occurs with abrasive particle sizes less than 1μm, 2) that it is the mechanical limit of the more common mechanical-chemical glass polishing, and 3) that the debris is insufficient in size to perform the function of eroding the binder in the grinding wheel and thus necessitates the addition of an abrasive and/or chemical additions to the coolant to effect wheel-dressing.

INTRODUCTION

This paper will group together a series of observations and work by ourselves and others which, when taken together, forms a picture of shear mode grinding. While we have been working in this field for a long time we are now talking to firms such as Matsushita[1] who unquestionably have had the process in production for sometime. We do not know in detail how they view shear mode grinding or the steps by which they arrived at their goal. This is simply our own understanding and the odyssey by which we arrived at this position.

Before we begin, we must alert the reader to the fact that we shall be linking physics and chemistry rather indiscriminately as we ramble back and forth between polishing and grinding of materials with both ductile and brittle behavior. We do wish to assure the reader that we are not consciously aiming to frustrate him by this seeming indirection. We firmly believe each point made will ultimately prove important in the field of shear mode grinding of brittle materials and glass in particular. We will address fracture behavior because we must be aware of what we seek to avoid. We will address material removal rate and surface roughness also because damage is bound to them. We will address abrasive size, not just as a parametric descriptor of force and fracture, but because abrasive size may be a critical parameter for thermal and chemical diffusion. We will address chemistry because even in the cases where pure mechanical gripping is sufficient to produce surface shear, it does so on a scale where the coolant (water) is chemically reacting with the near surface which has three dimensions and is a region rather than a geometric boundary. We will discuss polishing because shear mode grinding is really nearing the mechanical extreme of the more common mechano-chemical polishing. Machinists like to use the word grinding as denoting a more deterministic process but physical phenomena are not so easily partitioned by convenient semantics.

Most of the data and derivations in this paper have been previously published. This is not intended as a paper of justification so much as an essay toward understanding.

We will develop our picture as the jigsaw puzzler pieces together patches of the whole, and we will not apologize from this point on that we do not have every connection in place.

The Structure of Glass

Glass is basically a polymeric network of silicate tetrahedrons, a silicon atom at the center and an oxygen at each corner usually shared by two tetrahedrons in which case it is termed a "bridging" oxygen. A two dimensional slice appears as a series of loops usually involving four to seven tetragonal units. In crystal quartz these loops form a repeating hexagonal pattern. If an oxygen does not complete a bridge, there is normally an electron donor, usually an alkali, in the vicinity. In three dimensions a glassy material has a microscopically irregular form. Under strain the irregular structure insures that links are unequally loaded. The void space permits extraordinary hydrostatic compression as overloaded links buckle and permanent compressive densification occurs. The majority of plastic deformation is of this type since the irregular structure permits little shear relocation except in regions of rapid reorganization under thermal or chemical influence, such as near the surface. Surfaces brought together are almost certainly poorly coordinated on the atomic scale unless they have just parted, so they will not rejoin as metals can. Glass can support little tensile stress because overloaded links rupture and various chemicals attack and weaken the structure in the vicinity of a crack tip. Little ductility is evidenced in glass so we shall use the term "shear" in place of the alternative "ductile".

A Size Limitation

This section deals with the fact that there is a critical size limitation (a threshold size) for cracks. Knowledge of this goes back to the latter part of the eighteenth century when it was found that there is a minimum particle size that can be produced by compressive crushing. For glass and other oxides this is in the vicinity of 1 μm, as evidenced by the size distribution of many of our polishing abrasives. This, by hindsight, can be construed as an argument for a minimum fracture size on the order of 1 μm.

The next observation occurred with the use of indentation testing in this century. As testers went from wedges producing cracks to a more blunt shape indentor[2], they noticed crack propagation occurred primarily on unloading as the surrounding material took up the function of the indentor in restraining the elastic release of the densification zone. This compacted material, reorganized with respect to the non-compacted surrounding material, now put the latter in tension. Cracks, some initiated in tensile zones created during loading, are now free to propagate. These are not the only conditions from which cracks arise, but since the mechanism begins with a compressive state that is dominant in grinding, this is one of the more significant fracture sources.

In the latter half of this century, workers such as Lawn, Swain, Marshal[3], et al., have described and analyzed the size effect. Note that there are several elements here. First, there is as T. Bifano points out, a volume distributed elastic energy whose ratio to the crack surface energy, into which some energy will transform, leads to a dimension almost certainly dictating a critical size effect.

Surprisingly, Brian Lawn[4] and his co-workers had pointed this out quantitatively in a most elegant letter to the Journal of Materials Science in 1976 and drew a remarkable quantitative conclusion largely overlooked by the grinding community. Lawn noted that separate equations describe the radius of the densified or compacted zone and the radius of the crack originating from it. By equating these two radii so that any smaller crack would fall within the compression region (which is almost a conceptual contradiction), Lawn showed a way to determine the smallest crack size in a most elegant way. For glass this also turns out to be very close to one μm. We also have the well known fact that near this size the strength of glass fibers increases dramatically as it becomes less and less probable that the fiber can hide such a "least flaw".

There we have it, several different but very convergent lines of thought that point out there is a minimum flaw size. The obvious conclusion to be drawn for grinding is that something different must be expected for any sequence of phenomena that tend to produce smaller and smaller cracks as the size of something is reduced. Lawn in fact did draw this conclusion most explicitly at the end of his letter contrasting a possible "ploughing" or polishing mode with a "chipping" mode of removal. Before we were aware of Lawn's predictions of a decade

CH2690-6/89/0000-606 $1.00 C 1989 IEEE

earlier, we began looking for the same thing for a different reason and found it right where he had predicted.

Of Lichens and Chemistry

The bible of silicate chemistry for years has been "The Chemistry of Silica" by Ralph Iler, J. Wiley. In Table 1.1 of this standard work there are several interesting entries. At room temperature a little catechol in water is seen to raise the dissolution rate of quartz by a factor of a million over that in pure water, and saturated catechol in ammonium hydroxide can add another five orders of magnitude beyond that. Catechol is usually described as a breakdown product of lignin, wood structure, and much of our clays are broken down from rocks by decaying lichen. This would be just an out of place tidbit if it were not for the fact that opticians have lapped with wood pitches for nearly five millenia. Never underestimate the effectiveness of a little chemistry.

Tetsuro Izumitani[5] laid to rest the theories of simple abrasion and thermally induced flow in the early 70's by showing there was no correlation between the polishing susceptibility of glasses when plotted as either a function of their softening point or their hardness, but there was a great deal of correlation with both the weight change of powdered samples in acids and even in water leaching baths. Conventional polishing had to involve chemistry and water would be one of the chemicals.

In 1982 Walter Silvernail[6] showed that ceric oxide (Ce^{IV}) would polish but cerous oxide (Ce^{III}) would not. Lee Cook of Schott and Brown[7] began studying other polishing agents following this lead, and found that only one ion from each metal group involved in a polishing oxide was active, usually a IV valence or III valence. Closer examination of these showed a number of interesting but only partially understood correlations. The ions occupied a narrow range of ionic charge to radius ratios, tended to form amphoteric compounds, tended to form Lewis acids, and precipitated monosilisic acid from aqueous solutions. This last turned out to be crucial. The hydrolyized debris from conventional polishing is near monomeric monosilisic acid plus normally soluble modifiers (alkali electron donors etc.). The debris from the polishing compound causes much of this monosilisic acid to drop a couple of waters and repolymerize as pure silica on the glass surface. This is what puts the smooooth in polishing. Furthermore, this property of precipitating the hydrolized silica is a litmus test as it were for polishing ability. A polishing agent does this and a nonpolishing ion doesn't.

Now I am going to jump back to Izumitani and what appears to be another litmus test. One doesn't just go to a chemist's shelf and pull out a polishing compound. T. Izumitani, the Sage of Hoya, had shown that there is an optimum firing temperature for a polishing abrasive. Its density increases nearly linearly with firing temperature until full crystallinity is reached, but the polishing rate at first increases with firing temperature and then drops off. In effect, the compound needs to get fired to a strength sufficient to shear the glass but no farther. It needs to keep a maximum of site defects, of unsatisfied chemical potential ready to glom onto the glass in a chemical embrace. How does this connect?

Well, Lee Cook[8] has continued to investigate this Lewis acid question. When the surface silica is hydrolized and a tetrahedron is broken off, the bridging oxygen is left with the tetrahedron on the glass gaining an extra pair of electrons in the process. Now a Lewis acid is one which wants to gobble up an excess electron presented in just that fashion. Lee is about to publish this as the reaction we shall call "chemical tooth" and that the condition for the optimum firing is that one which will make the fired oxide gain maximum Lewis acidity in an aqueous solution. The importance of this discovery is that it gives us a test for firing temperatures other than the hit or miss polishing test which usually never gets done anyway. Lee says it is easy for any good ceramist to run the tests.

Under the heading of chemistry we still have the question of water. Izumitani had shown that water was a chemical agent in polishing and had shown that abrasion did not work nearly as well in petroleum products. This is important because machinists still like to use oil as a coolant on their machines and the high heat capacitance of water and its large heat of vaporization lead to awkward thermal control problems as well as rust. Water is essential.

Brown and Cook published an early version of this work in April of 1984, still puzzled by the effects of water and observations of surface plasticity. The next month Minoru Tomozowa[9] and his student Nogami published the latter's thesis and all was light. Basically they state that water diffuses into glass at a rate that increases exponentially with tensile strain and reacts internally with the glass by hydrolysis at a rate exponential with compressive strain (or PdV mechanical energy input). What this means is that water can travel easily in a glass structure by several mechanisms. Now assume a water molecule is sitting next to a bridging oxygen when the pressure pulse preceding any abrasive grain passes. The oxygen drops one bond to silicon, and picks up the H^+ from the water. The silicon grabs the OH^- from the water at the same time. Thus, where we had a solid link blocking the way of the approaching molecule, we now have an open gate since the formerly linked silicons are now each terminated with a hydroxyl group. Furthermore, in the tensile region just behind the particle, the diffusion coefficient rises to suck in water for the following particle, and as the first particle passes the hydroxyl gate, the change of compression to tension sucks out energy and reverses at least some of the hydrolytic reactions. We have just seen a mechanism for pumping in plasticity.

Since 1984, Tomozowa has been apparently accelerating his research into the aqueous chemistry effects on glass and Lee Cook who has seen quite a bit of it in prepublication form says wonderful things are coming. The grinding polishing community needs to keep tabs on this. It is chemistry geared directly for them!

Now we are going to leave chemistry for a bit to look at metals, the Soviet Union, and Japan.

Shear Mode Grinding

In the late 70's Brown, Baker, and Maney[10] were polishing the early diamond turned optics and working informally with Bob Parks on some Air Force Optics that were leading to LODTM. In the process Brown found that a spherical particle wear model gave nearly exact results if a Hertzian elastic penetration of the surface were followed by a plowing such that the resulting trough has a cross section area proportional to the cross section area of the penetration. Note that there is no chip science detail. We jump from the penetration to the trough skipping the intervening details. The simple analysis is given on a single page at the end of this report. The three key equations may be summed up as saying that the Preston equation holds, the wear rate coefficient is independent of abrasive particle size and concentration, and the roughness is proportional to the product of abrasive size and the two thirds power of the ratio of pressure to the product of concentration and Young's modulus. That's an awful mouthful, but to wake everybody up we should say that every single factor in those equations has been experimentally verified for shear mode (ductile grinding). The equations were developed for work

with ductile metals. They should work with any isotropic material that responds to abrasion with a plowing type response.

Now we jump up to 1985, a very good year. It started out with a rush. Miyashita[11] brought word that the group at Citizen Watch, (i.e., professors from Tokyo Metropolitan University, Tokyo Metropolitan College of Aeronautical Engineering, etc., as well as the engineering staff of Citizen Watch Co.) had succeeded in producing a 2 nm Rz (peak to valley max) surface on a 1 inch specimen of crystal quartz with a specially dressed diamond wheel on the newest version of their ultra stiff machine. There was no question of their achievement. Miyashita was loaded with his own stylus traces and others from Albert Franks and Kevin Lindsay at NPL, Teddington. The specimen was also examined with the Nomarski Microscope at the LLNL Optics Shop. It was, in a word, gorgeously unmarked by any sign of a wheel. Note that the glass was natural quartz, presumably formed in water and pretty well water saturated.

Miyashita also brought word that things were stirring in the Soviet Union. This became evident as the year unfolded and paper after paper poured out of Khodakov[12]. Khodakov made no secret of his goal. He was after shear mode grinding. In one of his earlier papers of that year, he did a variant on the analysis included here for spheres, showing that both wedges and pyramidal grains should provide size and concentration independence. After an excellent analysis he was forced to admit to what appeared to be defeat in his grinding studies which were done with a wheel. He surmised that his analysis had overlooked some important effect since his results obviously showed wear rate coefficients linear with abrasive size instead of size independence. As we shall see, he was mistaken and not in the way he expected. We believe he overlooked the fact that in going from fracture to shear mode, he left a region where grinding debris was of a size on the order of the least flaw (near 1 μm) to one where it was the order of a scraping (near a few nanometers). This new debris was too fine to dress his wheels and they glazed. We went to laps where we could study parameters and found the region he sought right beside the region he selected as his limit.

Late in 1985 Peter Takacs from Brookhaven asked us to begin looking at SiC for supersmooth synchrotron mirrors. As we worked this material on our laps we realized we were studying shear mode grinding at a slow pace, but shear mode grinding nevertheless. When you consider the acoustic velocity in glass is near 4 km/s and the highest grinding wheel peripheral velocities are 100 m/s or less, it is obvious that all speeds are insignificant as far as acoustic effects are concerned. (There are thermal and chemical diffusion rates to consider however).

While we were thinking in this vein it occurred to us that all shear mode grinding was in some degree 2-body wear whether the abrasive was loosely applied or tightly bound in a wheel matrix since a shearing particle does not roll or move with respect to the lap during shearing. This is not to say that a lap and a wheel will behave identically over this region. Particles held in place in reaction to the shearing forces are free to move after the part has passed, and such particles on a lap could move differently when dull as they could not when grown dull in a wheel. Yet at the same time there should be some similarity between the results. As we indicated we verified every single factorial in the three equations. This does not say the equations are exact. In fact, there is an order of magnitude variation wandering around for the constant in some cases, but the parameter dependence appears correct. In this time frame, Brown[13] also derived, on geometric grounds, relationships for certain other factors as shown in an Appendix.

To Grab or Not to Grab

Now we come to the discussion of tooth. To abrade a surface in a shearing mode we must somehow grip the surface we want to wrench or shear off. We know of two methods for mechanical shearing. We can penetrate with a sharp discontinuity where sharpness is an edge radius of curvature somewhat small with respect to the depth of cut. This is a mechanical gripping and note that there is a distinction between sharpness and keenness. We have seen sharp diamond tools with included angles of 120 degrees. The other method of gripping is by some form of a chemical bond, a gluing of the abrasive particle to the surface.

Now consider the scale of things. We know from polishing that depths of shearing will range from about one tetrahedron depth to a couple or tens of these. A very small chemical reaction and removal will seriously damage any edge trying to mechanically shear such a surface. Mechanical-chemical polishing agents are worn to well rounded surfaces where they are presented to the work. Now it takes very little wear to knock the edges off a diamond which can react chemically with both silicon and oxygen, although we believed it is oxygen that is doing most of the damage. Very fine submicron CBN might be a better choice than diamond if edge shearing is the method of choice. On the other hand, with a chemical bond, we wear away the abrasive layer by layer, "good to the last drop." It is not clear whether edged or chemical abrasives are preferable for this task. An abrasive wearing through very fast compared to diamond may still outlast it because we are comparing the total wear of a grain to the wear of an edge.

Now if we are contemplating edged wheels we must face the fact that in shifting from the transition region to complete shear region we have gone from at least a few near μm particles ready to abrade the binder for wheel renewal to shavings a few nanometers thick utterly incapable of abrading it due to a two order of magnitude size discontinuity. We must make provision for a chemical or abrasive additive to the coolant to provide for binder wear to provide renewed edges. Colloidal abrasives exist in a size range that should be applicable here.

We want small abrasive sizes, even submicron, which should lead to ultrafine wheels outperforming coarser wheels in the lifetime category. Consider a factor of 2 difference in size at the same concentration. The smaller size has eight particles where the larger had one. Each two of these have the same length of cutting edges as the larger so we have increased our cutting edge length by four. We must cut shavings half the thickness or make twice the cut per volume of material removal. Still we are a factor of two ahead. We may lose the last factor of two from a requirement that the narrower depth of cut requires twice the sharpness but we may gain in lower loads and greater surface for heat dissipation.

The chemical bonding abrasive may have the advantage of not having to be removed by binder control, or needing significant binder abrasion.

Both types of wheels may need to be guarded against silica buildup by abrasive or chemical means. High pH especially when accomplished by LiOH remains an attractive possibility, with a pH near 10.5 a good place to start. See Figure 1.6 of Iler.

Acknowledgements

This work has been performed under an Air Force grant monitored and supervised by W. K. Stowell. The work has also been frequently discussed with I. F. Stowers, T. T. Saito and P. P. Hed all of whose many contributions are greatly appreciated.

Appendix I

The text does not contain any calculations. The results of an analysis of Brown (SPIE, Vol. 306 and Precision Engineering, July 1987), are provided in the slide copies that follow. Assuming spherical particles and a Hertzian elastic penetration we can derive the fact that abrasive wear rate should be independent of particle size D and area concentration K_A if we assume the wear trough area is proportional to the cross section of penetration a. The factor M_f is a material coefficient containing among other things this constant of proportionality and probability factor. This simple derivation provides an analysis of the depth of penetration δ and the fact that wear rate follows the Preston equation. This last is of enormous importance since it provides the missing link to machine stiffness requirements for a given process.

For example, assume we are using a thin wheel with axis tangential to a rotating blank whose aspheric shape is determined by the wheel edge as it traces out the aspheric meridian. The wheel edge is similar to an oversize circular edged diamond tool. Assume we know the rotation rate of the blank and the wheel and the depth of cut and the rate of feed or traverse. We can form a volume increment of the depth of cut by the wheel thickness by the increment of traverse per blank rotation. This is the volume removed by the wheel in the time interval it takes for a point in the blank to pass the thickness of the wheel, a time interval we know. In that time interval we know the length of rotating wheel circumference that has been in contact with the point as it makes its passage across the wheel face. Now by integrating with respect to time and multiplying the Preston equation by the contact area, we obtain a form that states the volume of material removed is equal to the constant in the Preston equation multiplied by the total load and the relative displacement of the wheel across the part. We have just calculated this displacement and volume so a knowledge of the constant from a simple lapping experiment allows us to compute the load and machine stiffness for a given process.

The second formula is a parametric relationship determining the maximum chip depth Dmax in terms of abrasive size, feed to peripheral velocity ratio, and depth of cut to wheel radius ratio. The maximum chip depth is critical because it is in fact part of a chemical and thermal diffusion length necessary to assist in the shear mode grinding. Note that the concentrations K in the two analyses are different, one being an area concentration which is the two thirds power of the other which is a volume concentration. The concentration is not a significant parameter in reducing chip depth since the need to keep a bit of binder cushion between abrasive grains for shock reduction limits concentration to about 50.

In the question and answer session following the presentation of this paper, I was asked to include an estimate of process parameters for a quartz disk about the size of a quarter. This is strictly guesstimate. I do not have a coefficient of removal for quartz, but I estimate it to be about 10^{-13} cm^2/dyne. I will assume a 1.5 cm radius workpiece worked by a 2 cm radius cup wheel with a 3 mm wide abrasive area of 50 concentration, half micron resinoid bonded diamond. I will assume a maximum pressure of 1 atm at the edge of the workpiece yielding a 2/3 atm average pressure. The actual contact area for the cup wheel which passes over the workpiece center is 0.3 cm by 4 R $\sin^{-1}$ 3/8 = 0.923 cm^2 yielding a load of 6.15×10^5 dynes. Assuming a peripheral velocity of 30 m/sec or 3000cm/sec yields a volumetric removal rate of ($10^{-13} \times 6.15 \times 10^5 \times 3 \times 10^3$ =) 1.85×10^{-4} cm^3/sec which on a 7.07 cm^2 surface translates into 0.26 micrometers/sec. The grinding wheel speed would translate into about 29,000 rpm. A shallow removal of about 13 A per rotation would be consistent with a workpiece speed of about 12,000 rpm. The colloidal abrasive to achieve binder wear would be about a quarter to a fifth of the wheel abrasive size. Coolant would be water with a few percent organics to lubricate the diamond. Please note: We have not run this. This is a wild extrapolation from lapping experience.

* This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7405-ENG-48 and in accordance with Air Force Agreement AFWL-88-162.

References

1. Ueda, S., Mockida, S., Mahino, M., Nakata, K., Precision Grinding of Aspheric Glass Lenses, (Presentation at LLNL Aug. 1988).

2. Lawn, B., Swain, M., Microfracture Beneath Point Indentations in Brittle Solids, J. Mat. Sci. 10, 113, (1975).

3. Bifano, T., Ductile Regime Grinding of Brittle Materials, PhD Dissertation, North Carolina State University (1988).

4. Lawn, B., Jensen, T., Arora, A., Brittleness as an Indentation Size Effect, J. Mat. Sci., 11, 573, (1976).

5. Izumitani, T., Polishing, Lapping, and Diamond Grinding of Optical Glass, pp 115-172, Treatise on Material Science & Technology, Vol. 17 Edited by Tomozawa and Doremus, Academic Press, 1979.

6. Silvernail, W., Observations on Glass Polishing, OF&T Digest, OSA, Dec 1982.

7. Brown, N., and Cook, L., The Role of Abrasion in the Polishing of Metals and Glass, Tech. Digest, Science of Polishing, OSA, 1984.

8. Cook, L., Chemical Processes in Glass Polishing, 10th Glass Conference, Penn. State Univ. June 1989.

9. Nogami, M., Tomozawa, M., Effect of Stress on Water Diffusion into SiO_2 Glass at Low Temperature, J. Amer. Cer. Soc. 67(2) 1984.

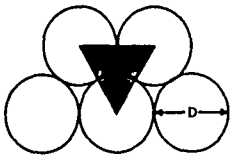
10. Brown, N., Baker, P., Maney, R., Optical Polishing of Metals, Proc. SPIE, Vol. 306 (1981).

11. Yoshioka, Hashimoto, Miyashita, Kanai, Abo and Daito, Ultraprecision Grinding Technology for Brittle Materials: Application to Surface and Centerless Grinding Processes, ASME Milton Shaw Grinding Symposium, PED, Vol. 16, 1986.

12. Khodakov, G., Characteristics of Grinding of Optical Materials by a Tool Having a Multilayer Distribution of Abrasive Granules, Soviet J. of Opt. Tech. 52(8) 1985.

13. Brown, N., Some Speculations on the Mechanisms of Abrasive Grinding and Polishing, Precision Engineering, Vol 9, 5, 1987.

Hertzian plowing model of abrasion



$$\frac{N}{A} = \frac{0.5K}{A_t} = \frac{0.5K_A}{\frac{\sqrt{3}}{4}D^2}$$

$$F = \frac{L}{N} = \frac{L}{A}/\frac{N}{A} = P\frac{\sqrt{3}D^2}{2K_A}$$

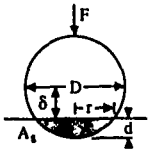
$$\frac{D^2}{4} - \delta^2 = \left(\frac{D}{2}+\delta\right)\left(\frac{D}{2}-\delta\right) = r^2 = Dd \quad \textbf{since} \quad \delta = \frac{D}{2} \text{ and } d = \frac{D}{2} - \delta$$

$$r^2 = \frac{3}{8}\frac{DF}{E} = \frac{3\sqrt{3}PD^3}{16EK_A} = (r^2)^{3/2} = (Dd)^{3/2}$$

$$A_s = H\frac{2}{3}(2r)d = H\frac{4}{3}\sqrt{Dd}\,d = H\frac{4}{3}\frac{(Dd)^{3/2}}{D} = H\frac{\sqrt{3}PD^2}{4EK_A}$$

$$\frac{\Delta V}{A} = \Delta H = \frac{N}{A}A_s\Delta S = \left(\frac{2KM}{\sqrt{3}D^2}\right)\left(\frac{\sqrt{3}PD^2}{4EK_A}\right)\Delta S = \left(\frac{M}{2E}\right)P\Delta S$$

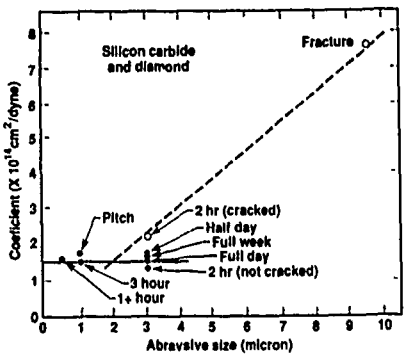
$$\boxed{\frac{\Delta H}{\Delta t} = \frac{M}{2E}P\frac{\Delta S}{\Delta t}}$$



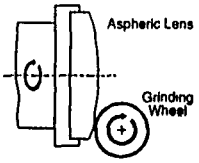
$$(Dd^{3/2}) = \frac{3\sqrt{3}PD^3}{16EK} = \left(\frac{3D^2}{4}\right)^{3/2}\frac{P}{2EK} \qquad \boxed{d = \frac{3}{4}D\left(\frac{P}{2EK_A}\right)^{2/3}}$$

$$\epsilon = \frac{F}{\pi r^2 E} = \frac{\sqrt{3}}{2\pi}\frac{PD^2}{KEDd} = \boxed{\frac{4}{\sqrt{3}\pi}\left(\frac{P}{2K_AE}\right)^{1/3}}$$

Rate independent of size



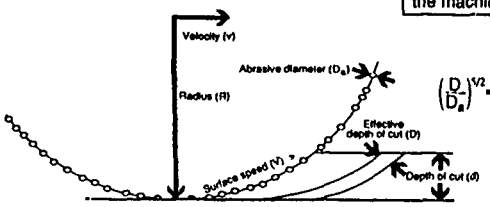
The Preston equation can be used inversly to help specify the grinding machine characteristics



$$\Delta H = C_P P \Delta S$$
$$\Delta H A = \Delta V = C_P L \Delta S$$

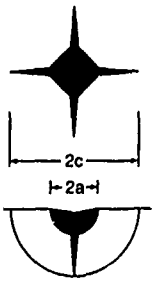
Using the Preston relation, the removal rate ΔV, and the displacement ΔS the load L can be ascertained.

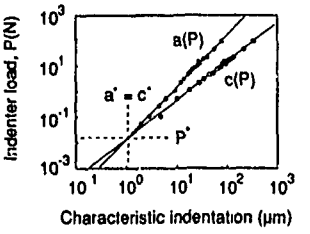
L can be used to determine the machine stiffness.



$$\left(\frac{D}{D_a}\right)^{V2} = \frac{3.42}{K^{2/3}}\left(\frac{v}{V}\right)\sqrt{\frac{r}{R}}$$

Prediction and demonstration of minimum crack size





al, J. Mat. Sci 1976

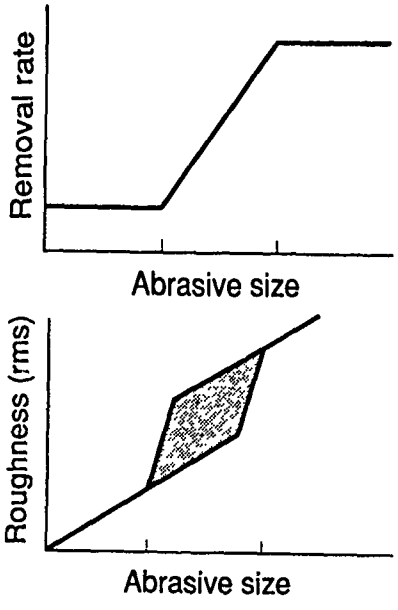
43rd Annual Symposium on Frequency Control - 1989

THE RESPONSE OF ISOTROPIC BRITTLE MATERIALS TO ABRASIVE PROCESSES

Norman J. Brown, Baruch A. Fuchs, P. Paul Hed and Irving F. Stowers
Lawrence Livermore National Laboratory*

ABSTRACT

This paper is an overview of brittle material response to abrasion. It proposes a 3 region model for both 2-body and 3-body wear, two regions being where removal rate is nearly independent of abrasive size at otherwise identical conditions and a transition region where rate is proportional to abrasive size. The mechanics will be explored through these 3 regions in 2-body and 3-body wear and some comparisons will be made to the wear of nonbrittle materials.



INTRODUCTION

It is the purpose of this paper to attempt to paint a picture of the response of brittle materials to abrasive processes in very broad brush strokes, paying particular attention to abrasive induced damage. For this paper we will restrict our attention to isotropic materials with particular attention to glass and a few other materials of interest for optical applications. We shall adopt for the moment the tribological nomenclature of 2-body (wheel and part) wear and 3-body wear (lap, loose or free abrasive, part) although we shall point out the degree to which this division has become oversimplified and at times overburdened with long associated concepts.

We shall address a central theme that wear in each of the two dimensions above can be divided into at least three regions, in two of which wear rates are independent of abrasive size, (all other parameters being constant), and connected by a third region in which wear rates are linear with abrasive size when all other conditions are fixed. We shall point out that to some extent the same type of phenomena can be found for more ductile materials, but we shall provide no data for this. We shall also throughout this paper use the terms shear mode to describe what is sometimes called ductile grinding, or nanogrinding, or non-fracture grinding, and we shall use the term densification to refer to a form of permanent deformation in response to near hydrostatic compressive loading. The reason for these distinctions will become clear throughout the paper.

We shall appear at first to be wandering at random and talking about rates, roughnesses etc, rather than about subsurface damage. This is because there is a relationship between the roughness and the depth of subsurface damage, and between the roughness and the size of the chunks whose removal resulted in the roughness, and between the size of the chunks and the rate of removal. Removal rate, roughness, and subsurface damage are different aspects of the same phenomena.

We shall throughout this paper emphasize size and dimension. For a given concentration of abrasive, the number of points transmitting the load will vary as the inverse square of size, so size has an obvious effect on load per impacting or abrading particle. However, dimension is also tied to diffusion lengths, both thermal and chemical and to atomic and molecular structure sizes.

We shall be speaking of physical abrasion, but at both the finest and coarsest particle sizes, the coolant (water) becomes a chemical whose effects may dominate the phenomena. We shall see that on the finest scale, chemical interactions between substrate and abrasive are crucial to behavior and process strategies.

We are restricting ourselves to isotropic materials, since variations in bond strength and lattice structure introduce orientation effects. This should be interpreted neither too strictly nor too loosely. Few crystals can be described as isotropic, but some of the simpler ones such as quartz and sapphire are well behaved in most isotropic analyses, but even in these cases, anisotropy can be significant. Namba[1] for example has made effective use of the chemical anisotropy of sapphire while others would probably consider it isotropic. Vinokurov[2] points to rhombododecahedral crystals as being particularly anisotropic, pointing out a fracture depth to surface roughness ratio of 7 as compared to a factor of 4 for most materials. At the present time we do not know where to draw this line but we are aware of cautionary flags when dealing with crystals.

Strain Fields

Glass and brittle materials tend to be more restricted to an elastic behavior than do ductile or malleable materials. They yield in a generally linear manner and then fracture quite abruptly with little of the nonlinear (plastic) strain and displacement that generally precedes rupture in ductile materials. There is one exception to this generalization, however, and at times it can dominate both our observations and our thinking. Under extreme compression, regions of the glass structure that are in near hydrostatic compression can collapse and take on a more or less permanently compacted geometry which we shall describe as densification to distinguish the phenomena from the more general plasticity.

In saying that elastic behavior dominates glass response to loading, we are singling out elastic descriptions of this response. We shall be concerned here only with the most general features of these descriptions and will look at Boussinesq type of solutions. Displacements in response to compressive loads (displacement per unit of load) may be described as a function of location and orientation beneath the surface and Poisson's ratio, all divided by or inverse to Young's modulus. Similarly, stress per unit load may be described as a function of location, orientation, and Poisson's ratio. Poisson's ratio is linked to the dimensional function and in effect is a measure of the strain or displacement cross linking. Young's modulus however shows up as a constant factor just like the load. It is a scaling factor and the dependence of strain or displacement on this parameter is an inverse relationship. This will turn out to be very important for us in our scaling of the rates, roughness and damage.

CH2690-6/89/0000-611 $1.00 C 1989 IEEE

Glass is a very irregular structure of small tetrahedral units made up of silicon cores surrounded by four oxygen atoms on the corners, each oxygen obtaining one of the valence electrons from each of the silicon atoms it links. A silica network is a linkage similar to a foam with the silica tetrahedrons surrounding spaces that are quite large with respect to the tetrahedron themselves. A plane slice through one of these will show four to eight of these units surrounding a typical space. Thus, there is plenty of room for compression and permanent displacements to take place as a highly overloaded link buckles or permanently twists. This produces compaction, but by and large each tetrahedron remains linked to the neighbors it had prior to collapse and densification.

Now if an indentor permanently deforms a glass surface, it forms a near hydrostatic compression region. Sometimes Poisson effects produce tensile cracks on loading. However, as the indentor is removed and its load is reduced, the surrounding material must take up the load of confining the compacted region. The tensile stresses thus generated propagate any cracks already initiated as well as initiate new ones. These tend to be of two types, a lateral or petal type crack that propagates parallel to the surface and produces chips and vents, and a type which penetrates orthogonal to the surface either as a near half circle centered on the center of the indentation, or as a quarter circle extending from a normal axis through the center up to the surface. There are other types of cracks and fissures, but these types appear to dominate in much of what we shall examine.

Now let us assume we are operating in 3-body wear in the transition region. Here we will find wear rates, surface roughness and damage depth proportional to abrasive diameter. Let us run three or four specimens of different materials simultaneously on the same lap, at least one specimen being BK7, a more or less universal standard for both grinding and polishing. The specimens if properly mounted will see the same velocities, pressures, etc. The only thing that will be different to first order will be the materials. We will assume Poisson's ratio is similar for all. Thus, to a first approximation the only thing different will be Young's modulus. We have just seen that the displacement field scales with the inverse of Young's modulus E. Strains are derivatives of this so they will also scale as inverse with E. Thus we should not be too surprised when we observe that "roughness and damage tend to scale with 1/E". Note that word "tends". The dependence is not exact and Poisson's ratio seems to have a little larger part than usual to play in this type of loading.

In one of the first papers we published[3] we pointed out that removal rates seemed to be varying as the cube of the roughness for a given size abrasive. Now if you think about it, the fracture impacts are produced by abrasive particles rolling between the part and the lap in the 3-body wear mode. There should be about the same number of impacts per unit area under each of the specimens. The simplest assumption is that the chunks removed are proportional in size to the roughness. In an isotropic field, similar but for scaling, we should expect the same maximum strain geometry and the same shape of particle debris. Since the hollows are where a chunk or chunks come out, we should expect a proportionality between chunk and hollow size and a geometric similarity between the chunks from different materials. If we have the same number of chunks removed, since we have the same number of impacts, we would expect the rate of volume removal to vary as the cube of the chunk size or the cube of the roughness. Note that we are not postulating this and looking for an experimental verification but rather holding experimental data and looking for an explanation. The experience of a decade and a half of grinding and polishing research has taught us that this is one of the more reliable rules of grinding. If all other things are held equal, "roughness will vary as the cube root of the grinding rate". Since it is a cube root it is a fairly insensitive parameter, rather steady and stable, the kind of bellweather we need to check grinding results. If the roughness does not check out consistent with this model, it can indicate a need for caution in assessing subsurface damage. Thus it is fairly important to put this cubic relationship on a firm basis.

In addition to our observations, we have a completely different form of analysis by a leading Soviet researcher, L. S. Tsesnek[4]. In 1980 he made an analysis on probabalistic grounds that indicated there should be a cubic dependence. Note that he was not primarily interested in wear rate but in the size distribution of wear particles. Consequently, he did not look at a rate analysis for confirmation but toward a size distribution experiment. This appears to confirm his analysis and his equation for rate confirms an equation we published in 1976. Note that we are not claiming priority of discovering this fact. We have recently received a copy of a 1957 paper by Aleinikov[5] that simply states the same cubic dependence. What we are claiming is a long awareness of this fact and extensive experience with its use.

In summary, we have two observations; that roughness scales as 1/E and also scales as the cube root of the grinding rate.

Now consider how we would use these rules. One of the easiest measured properties is E. Suppose we get a new material in the shop. Assume it's SiC. We look at E and find it's 5 times that for BK7. That means it may grind as much as 125 times slower than BK7, our standard, or require 125 times the force for the same rate. We know we're in some kind of trouble. If we pick a Blanchard grinding feed rate that we would ordinarily use for BK7, we know something's going to give and it's probably either our Blanchard or at the very least our wheel. Next we are going to run a grinding rate experiment. We know the E is very high for SiC so the abrasive particles that provide chips won't penetrate quite as far. The part will ride a bit higher on the slurry and only the bigger particles will provide support. Thus the chips will be a tad bit larger than we expect and it will be ground a bit faster than we expect. We expect it to go about 125 times slower than BK7 and we measure only 80 times slower. That's in the right direction and we know enough to plan on a long run. Remember, we were predicting a roughness and cubing this to obtain a rate. A cubing operation tends to triple any errors in our assumptions regarding things like Poisson's ratio, particle size etc., by the time we test these with a rate measurement. On the other hand, assume we get our rate measurement. We can now ratio this to the rate of BK7, take the cube root of this ratio and we have the ratio of the roughness of our new material to that of BK7. If we have a stylus instrument we have a computed check on it. If we don't have a surface roughness instrument but know the roughness of BK7 for this process from published data, we probably have a pretty good value for both surface roughness and by inference subsurface damage just from this computation.

We should not overlook the backup value mentioned. There have been cases with very hard materials such as SiC, where the computed values were probably more representative of the surface than the stylus trace values due to the very small size of the roughness. Note here that since we are taking a cube root we are reducing our errors by about a factor of three, whereas going in the other direction we increased our errors by three. Furthermore, here we are making no assumptions about Young's modulus or Poisson's ratio or anything else. We are measuring a removal rate, an integration or summation of chip volumes, where the volume shape already has the real values factored in.

3-Body Wear - Regions of Behavior

We started out as all good literature should somewhere "in medias res", into the middle of things. Now let's survey the terrain.

We mentioned that we were in the transition region where both rate and roughness were proportional to particle size. Let's proceed to smaller particle sizes. We should realize we have a problem. It has long been known that there is a minimum fracture size. Particles produced by crushing have a minimum size, evidence of a minimum fracture size of 1 μm. Further crushing of these debris particles will lead to densification and other forms of permanent deformation. Well, with this in the offing we certainly can't reduce size indefinitely without seeing new phenomena. The fracture experiment outlined above indicates a minimum size near 1 μm. This same mechanism is why our polishing compounds are centered in this size region.

Brian Lawn[6] and coworkers by equating fracture radius from the standard fracture mechanics formulation to the compaction radius from the definition of hardness came up with a condition that postulates that crack radius should penetrate the densified volume and disappear at a particle size near 1 μm and a load per abrasive near 1 gram. Few in the grinding community seem aware of Lawn's prediction which was published in a letter to the Journal of Materials Science in 1976, but it agrees remarkably well with what we are seeing and what some Japanese firms now have in production. Recently, R. Mouginot[7] working from another more sophisticated analysis has proposed a factor of E/H should multiply Lawn's criterion but this is not a major change.

In any event, at about 1 μm abrasive size we should see a change in behavior and Lawn predicts we will go from a chipping mode to a plowing mode. We have found that and the parametric dependence is one Brown[8] published for the plowing abrasion of metals in 1981. Since we have another paper in this conference on the shear mode grinding we shall not describe this beyond a few notes necessary to complete our overall grinding picture. We mentioned that the analysis was developed to describe metal abrasion in an optical regime, a very gentle plowing mode of abrasion. Thus, the fact that these two plowing regimes, one in ductile and one in brittle materials agree except for a constant of proportionality leads us to believe we have connected the ductile and brittle response in the small size abrasive region.

We also have coupled 3-body and 2-body wear to a degree. 3-body wear is almost a contradiction in terms in shear mode grinding since the abrading particles are apparently locked to the lap by grinding loads if not by actual physical socketing during the plowing. The profile of the scratches do not change during shear mode grinding and in more than a few cases, identifiable characteristics persist along the path, confirming a constant grit orientation, a lack of independent action by the particle. This means that even particles loosely applied are acting in 2-body wear. There are however, a few differences and we wish to sound a note of caution. First, for abrasive loosely applied, the particle may be free to change orientation at the end of a plowing stroke, i.e. when it passes from between particle and lap. Secondly, it is certainly not as firmly socketed as when bound in a conventional metal, ceramic, or resinoid bonded wheel. Thus the additional forces generated as it becomes dull may elicit some different response. We were aware of the possibility of differences and experimented with particles deeply imbedded in pitch. We found behavior in these cases quite consistent with the cases of abrasive loosely applied, in general, but we also detected what may be termed minor differences with time.

In shear mode we did not see any rate dependence on concentration or size of abrasive, but we did see roughness dependent on both parameters and of just the order predicted for each parameter. We found evidence in the literature and in personal communications that this mode of abrasion has been largely overlooked for years because, we believe, people encountering it in 2-body wear encountered wheel glazing due to the fact that a step change of two orders of magnitude in debris size precluded standard continuous wheel dressing by the grinding debris.

Finally we wish to point out that shear mode grinding appears to be the mechanical extreme of mechano-chemical polishing. This does not mean chemistry is absent from the process since the chemistry of the coolant and workpiece is definitely part of the process (hydrolysis). It simply means there is less chemical reaction between abrasive and workpiece, but what little there is strongly influences the interaction (i.e. the effect of glass or silicon, a carbide former, on diamond wear).

At this point we wish to return to our original position on the rate versus abrasive size curve in the middle of the transition region for 3-body wear and direct our attention in the increasing size direction.

In 1976 we assumed the Preston wear equation adequately described this regime. We determined a coefficient at a number of points along the curve and found it was indeed linear with abrasive size. We found several pieces of Soviet data[9,10] that were in excellent agreement. Thus, we had a set of Preston coefficients that could be expressed as a constant with dimensions of inverse pressure per length or size of the abrasive particle. When multiplied by the abrasive size this would become a conventional Preston coefficient for that size of abrasive.

Then we encountered a puzzling set of data from Schott Glass Co. It is still being published as a set of grinding data for many of their glasses for loose abrasive grinding. This data gave a Preston coefficient per unit of grinding abrasive diameter almost exactly a factor of 4 below our data. Something appeared to be wrong.

We did not uncover the reason until we began working with SiC. This had a surface roughness about five times smaller than that of BK7 for a comparable size abrasive, and the SiC wear coefficient was about 80 times lower than than for BK7. We needed a higher abrasion rate so we went to a larger abrasive size than we traditionally used in the shop. We found as we expected that the roughness and by inference the subsurface damage increased with abrasive size in just the linear fashion we expected, not only for the SiC, but for the BK7 and fused silica that went along on the same lap at the same pressure as comparative standards. The roughness ratio and rate ratio between the materials at each abrasive size remained constant and consistent when the size was increased. The trouble was, the rate did not increase with size beyond a certain point. We appeared to be working in a region where rate is independent of abrasive size again. With this as a clue we returned to the Schott data and interpreted it in the same terms. It had been ground with 115 μm abrasive and it would become consistent with the other data if we assumed the rate linearity with size stopped at a size somewhere near 30 μm. We have found several different pieces of data and they tell the same consistent story. At about 30 μm abrasive size, 3-body wear coefficients cease being linear with abrasive size. However, here we encountered a real puzzle. Roughness and damage in the transition region are also linear with size. They continue to increase linearly with size past the abrupt rate transition point. They cannot of course continue to increase indefinitely. In a 1986 paper by O. Podzimek[11] we may be seeing a transition to size independence for roughness in the vicinity of 90 μm abrasive size for BK7 and

SF6 and near 100 μm in size for Herasil I, a form of fused silica. This last behavior is consistent to a degree for fused silica. On the other end of the scale, fused silica makes the transition to shear mode behavior at a smaller size (1 μm) than BK7 does (near 2.5 μm).

The existance of regions in which the wear rate is independent of abrasive size while the roughness continues to be dependent on abrasive size has significant economic consequences. If an increase in abrasive size does not increase the removal rate but increases the abrasive damage, it is a very poor economic trade-off. For this reason, identifying the transition point or the "critical size" as Rabinowicz[12] termed it for metals two decades ago is extremely important.

This same logic does not necessarily hold for 2-body wear. Here we force the abrasive deeply into the work. There are relationships that must hold between the size of the abrasive and the depth of abrasive penetration into the workpiece if binder abrasion for self sharpening of the abrasive is to occur. This relationship is to our knowledge not completely explored in the literature and we will not pursue it further in this paper.

We should also point out some interesting comparisons between our observations in the transition region and those of the tribological community. By and large they have tended to work at the upper abrasive size range of the transition region and with metal. We also worked with metal, polishing out grooves of diamond turned optics over a decade and a half ago. In the mid 50's the tribologists developed a form of the Preston wear equation involving a coefficient with an inverse hardness dependence. Brown is developing a similar equation with inverse Young's modulus dependence at the lower abrasive size extreme. This transition from Young's modulus to hardness dependence is quite rational. At the smaller abrasive sizes the forces tend to be quite gentle and tend to simply rupture bonds whose strength is in some way measured by elastic properties. As we increase the abrasive size the particle loads increase dramatically, pushing in the case of metals into the plastic region. The reader will note that the definition of hardness in terms of load per unit of indentation area has the same units as Young's modulus. The latter describes displacements where structures remain connected to neighbors. The former describes plastic strain where neighbors are exchanged. Now in a sense we see the same thing happening in glass. Instead of general plasticity we have a more restricted form, compaction or densification, but still permanent deformation increasing with increasing particle size (all other things held equal) and associated with it, cracking. Thus, similar phenomena are going on in glass as in metals, but there are differences. Glasses cannot stand tension to any serious degree, and are not only homogeneous but without grain boundaries. Thus, metal tribology must be considerably more complex in some respects than that of glass. Nam Suh's elegant theories on spallation and delamination are usually inconceivable in glass.

Finally, we wish to point out that this region of rate independent of particle size is not a new discovery. Rabinowicz had studied it rather extensively in metals and some other materials in the 60's but somehow that information had never gotten back to the optical community and apparently little was done in the realm of brittle materials and glass in particular. Similarly this 3-body region concept is not thoroughly new. Nam Suh and his coworkers have been exploring a similar demarcation in metals terming the lower size regime sliding wear.

Linking 2-Body to 3-Body Wear

Since 3-body wear is somewhat more predictable and uniform, it has traditionally been interposed between 2-body wear processes known as generation and the fine 2-body wear known as polishing. Thus, since very slow polishing must remove the damage of the 3-body transition region wear, that damage is fairly well known and will be described in an appendix. However, as we begin to enter a period of more deterministic manufacturing we expect to see far more 2-body wear cases, both due to the precision and variety of processes available, as well as the speed dictated by modern economics. It is thus increasingly imperative that we begin to understand 2-body processing the same way we understand 3-body processes.

There have been a number of problems in studying 2-body wear. One of the more serious has been the tendency to allow 2-body wear processes to terminate in a spark out. This usually superimposes a fine ground upper structure on a coarser lower structure clouding the surface damage issue. However, we have begun to make some inroads.

Recently we used a 9 μm diamond abrasive wheel from our Blanchard Grinder as a lap and obtained five wear rate measurements. We computed the Preston coefficient for these experiments and found they were quite comparable to results we were obtaining with loose abrasives and with diamond. We are going to explore some of the similarities and differences between two and 3-body wear by examining the transition region behavior, using BK7 as our principal references. In the Soviet Union BK7 is known as K8 glass. In a 1969 paper, L. S. Tsesnek produced 4 data points for K8 glass with 0.083, 0.100, 0.127 and 0.107 mg/s removed using M28 alumina abrasive on cast iron laps at a load of 4 x 10^4 dynes and a velocity of 110 cm/s. There is a question on his pressure which on a 4 x 4 x 40 mm specimen could have been either 25 g/cm^2 or 250. From the similarity of his results to our own half a decade later we tend to believe it was the lower value. He claims to have removed both the fine and course fractions from his abrasive to achieve a 26 μm mean size, but there is a tendency of the parts to discriminate against the largest particles by pushing them into grooves and riding on the next larger particles so he may still have had a near 28 μm size. The reason we point this out is that his values are almost exactly identical to our own if this is the case. At any rate dividing the mean removal (0.10425 mg/s ± .01821) by load, velocity and density we obtain a Preston coefficient of 9.40x10 $^{-12}$ cm^2/dyne and normalizing to abrasive size we obtain either (3.62 ± .63)x10^{-13} or (3.36 ± .59)x10^{-13} cm^2/dyne-μm. Our own results published in 1976 for 9, 15, 20 and 30 μm alumina on cast iron at about 25 g/cm^2 at 114 cm/s gave 3.41x10^{-13} cm^2/dyne-μm for BK7. We also have a similar piece of data from Sergeev and Golovanova where they report 0.6 mg of K8 removed in 15 minutes at a load of 460 g at a velocity of 125 cm/s on a brass tool. The brass is a problem. Our measurements on dead soft brass gave a factor of 0.52 compared to cast iron but other data in the same chapter where we found the above data yielded 0.83. The above data yields a cast ion equivalent coefficient either 3.23x10^{-13} cm^2/dyne-μm or 2.02x10^{-13} cm^2/dyne-μm, depending on the type of brass.

Now let us compare the results for bound abrasive. We had a 10 inch OD by 9 inch ID, 9 μm diamond 100-concentration resinoid bond Blanchard wheel in good shape. We inverted it and loaded a 1.9 inch BK7 specimen with 1393 g. We ran 5 runs of 100 turns each with an average mass removal of 0.1004 ± 0.008 g to yield a coefficient of 4.27x10^{-13} cm^2/dyne-μm. If we consider the load distributed across a 1/2 x 1.9 inch area, the pressure is about 230 g/cm^2. These values are not drastically different from the 3-body wear values indicating a somewhat similar process is involved. We should be able to estimate loads and removal rate combinations even in the absence of a force dynamometer on the Blanchard Grinder.

Next we shall examine another set of 3-body wear data. There is a rather broad range for reasons that we do not fully understand. The Tsesnek data cited had a 17.5% standard deviation and it is not by any means an extreme case. This has masked much fine structure. For example, few tests have closely compared the results of pressure variations on the same lap run. A wide variation in pressure tends to make a free-floating ring stop with the highest load in the drag position. We set up a lap with a gear drive on rings and varied pressures from 10 to 140 g/cm^2 on samples in the same ring. We were quite surprised to see our rate data take on an unmistakable inverse cube root dependence on pressure when we simply normalized to pressure and velocity on the assumption of Preston equation behavior. It is not much in most cases but it is clearly there. This sheds a whole new light on 3-body wear in the transition region since it indicates the Preston equation holds only approximately. In our best judgement we see rates in this region varying as the diameter of the abrasive and the two thirds value of the pressure. Now this is suddenly beginning to show up as a familiar expression. The Hertzian elastic penetration of a plane by a sphere has a penetration depth equal to the diameter of the sphere multiplied by the two thirds power of the ratio of pressure to the product of concentration and Young's modulus. We have not tested for concentration and Young's modulus effects in detail but do know that rate is not a simple inverse two thirds power of Young's modulus. There is close to an inverse cubic dependence in fact. Thus while rate may well be linear with penetration depth, there will have to be other factors involved.

Currently we cannot measure loads on our Blanchard, but we can measure roughness as a function of rate. On one Blanchard we had a 100 mesh resinoid bonded wheel, 100-concentration. We measured the roughness of the surfaces as a function of removal rates. The BK7 specimens were placed 5 inches from the center of the table which was rotating at 50 rpm. A removal schedule was set so that there would be no residual damage from previous grinds. Six 0.5 inch stylus traces were taken at six equispaced circumferential positions between the 0.5 and 0.7 radius regions in a tangential direction. The average of the Ra traces were determined from 36 equispaced positions on each trace to yield the following values in μm; 1.628 ± .312, 1.148 ± .178, 0.936 ± .108, 0.843 ± .143 and 0.606 ± .050 at removal rates of 0.050, 0.010, 0.005, 0.001, 0.0005 and 0.0001 in/min respectively. The ratio of the first two roughnesses is 0.22 and the first and third yield 0.24, and the ratio of the first to the fourth yield 0.17. We hoped to find a cube root dependence on rate but the values at the higher rates which are the ones we trust seem to follow a bit lower power, about two ninths, about two thirds of the value we anticipated. You see we would have the same number of impacts but the depth would change with feed rate. We have had a situation similar to one we experience in 3-body wear where parts with different properties are done simultaneously on the same lap. With 3-body where the difference in debris size and hence roughness and hence subsurface damage varies as the cube root of the rate. Here we don't quite have that and we are dealing with a fairly large size abrasive.

Now the 9 μm resinoid wheel is too fine for most Blanchard work. Used as a lap it gave an Ra of about 0.09 μm. On the Blanchard, at 0.0001 in/min it gave 0.035 to 0.04 μm Ra and at 0.0005 in/min downfeed it gave 0.09 μm Ra. It is information like this that lead us to look for the cube root of rate behavior since nine fourths is very close to the cube root of 5. When we went to 0.001 in/min we overloaded the wheel and measured 0.046 to 0.05 μm Ra. At two intermediate rates 0.00067 and 0.0008 in/min downfeed we obtained 0.09 and 0.08 μm Ra. The onset of overload is rather sudden.

While we are on the topic of Ra we would like to pass on a few tidbits. We usually use what we term peak-to-valley readings on stylus charts, picking out the larger or extreme average values. This is formalized in metrology as Rz, the ten point spread where the difference of the average of the five lowest values in an interval is subtracted from the average of the five highest peaks, a reasonable length of rather level traverse being selected. The Ra or arithmetic average is four fifths of the rms for near random distributions and we should expect the Ra to be about one fifth of the average peak-to-valley. It may be, but we seem to find it is about one seventh of Rz, the ten point spread. The major damage fraction is confined to a depth about 4 times Rz and virtually all the damage of the crack variety that can open up on an etch will be found within about 7 Rz. Thus, about 50 Ra will turn out to be the maximum damage depth and 30 Ra the major damage depth for free abrasive grinding. Now 25 to 30 times the Ra on 3-body wear usually yields the abrasive diameter for our standard BK7 worked with alumina on a well designed lap. Our Blanchard results are at best about 3 to 5 times lower than we would predict.

Conclusions

Most of the conclusions have already been drawn in the appropriate places. We have shown we see three regions of grinding in 3-body wear in glass, and presumably three in 2-body wear in glass. We certainly see all three regions in both three and 2-body wear in metals. We have also shown that there is no true 3-body case in the lowest region we have described as shear mode grinding.

Acknowledgements

No one works in a vacuum. Most of this work was supported by Air Force funds on a contract monitored and supervised by a W. K. Stowell. Many discussions have been held with T. T. Saito and many of the important experiments could not have been accomplished without the skills of Tony Demiris and Felicia Ochoa.

Appendix

The ratio of subsurface damage to surface roughness is about 4 for most of the damage and about 7 for the last damage trace as determined by Hed and Brown using the taper polished method. The factor of 4 had been observed many times beginning with Preston in 1921 and continuing through Aleinikov. It seems to be on a very firm footing.

The problem that occurs with inferring subsurface damage from surface roughness is that some other process such as a sparkout may superimpose a lesser surface roughness on a larger subsurface damage. Note that we are also seeing variations in roughness with pressure or rate for material ground with the same size material, small effects on the order of cube root or less but departures from constancy. Using the cubic relationship between rate and roughness for materials ground on the same abrasive as a check is a simple procedure and will certainly point out gross errors. A factor of two or three difference in roughness would imply a factor of 8 to 27 in rate, no small difference.

Fairly wide ranges in rates are experienced in different loose abrasive grinding runs under nominally identical grinding conditions, 10 to 20% being rather common. However, comparative differences between materials run simultaneously on the same lap show a much smaller range, so we have followed the lead of Schott and Hoya in making most of our runs with a BK7 monitor and normalizing our data to that for BK7.

* This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract No. W-7404-ENG-48 and in accordance with Air Force Agreement AFWL-88-162.

References

1. Namba, Y., and Tsuwa, H., Ultra Fine Finishing of Sapphire Single Crystal, Annals of CIRP, V28, 1, 1977.

2. Vinokurov, V., Ardamatskii, A., Popov, V., The Structure of the Disrupted Layer, Generation of Optical Surfaces, Kumanin (Edit), Focal Library, 1962.

3. Brown, N., Prochnow, E., Blackman, R., Whistler, W., Free Abrasive Grinding In Precision Optics, UCRL 77844, 1976.

4. Tsesnck, L., Kogan, Y., Sorokin, V., Electon-microscopic Investigation of the Products of Glass-surface Dispersal by Abrasive Working, Sov. J. of Opt. Tech. 47(12) 1980.

5. Aleinikov, F., The Effect of Certain Physical and Mechanical Properties on the Grinding of Brittle Materials, Sov. Physics, Tech. Physics, 27, 2529, 1957.

6. Lawn, B., Jensen, T., Arora, A., Brittleness as an Indentation Size Effect., J. Mat. Sci. 11, 573, 1976.

7. Mouginot, R., Blunt or Sharp Indenters: A Size Transition Analysis, J. Amer. Cer. Soc. 761 98) 1988.

8. Brown, N., Baker, P., Maney, R., Optical Polishing of Metals, SPIE Vol. 506, 1981.

9. Tsesnck, L., Experimental Determination of the Surface Energy of Solids by a Grinding Method.

10. Serguv, O., and Golovanova, M., Rate of Wear and Frictional Force in Abrasive Disruption, 87 (Generation of Optical Surfaces, Kumanin (Edit), Focal Press 1962.

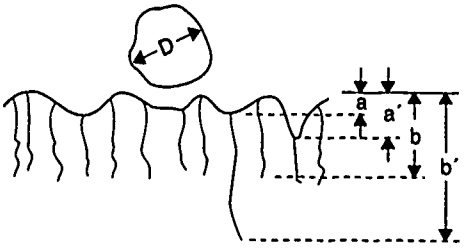
11. Podzimek, O., Residual Stress and Deformation Energy Under Ground Surfaces of Brittle Solids (Fig. 4) Annals of CIRP, V35, 1, 1986.

12. Rabinowicz, E., Mutis, A., Effect of Abrasive Size on Wear, Wear V8, 1965.

13. Nam Suh, Reclamination Theory of Wear, Elsevier, 1977.

14. Nam Suh, Tribophysics, Fig. 7.13, Prentice Hall, 1986.

Surface roughness may be related to subsurface damage in the brittle removal mode



a = peak-to-valley roughness

b = cluster criteria estimate - 95%

b' = last fracture trace

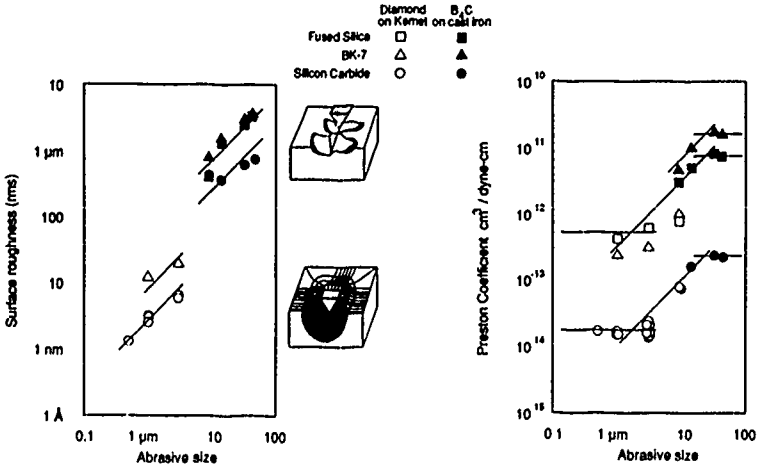
b = D abrasive diameter

$\frac{b}{a} = 4$ Preston 1921 $\quad \therefore a = R_{p\text{-}v} \approx \frac{D}{4}$

$\frac{b'}{a} \approx 7$ Hed - Brown 1985

$\frac{b'}{a'} \approx 4$ $\quad \therefore \frac{a'}{a} \approx \frac{7}{4} \approx 1.75$

All of the observed phenomenon associated with ductile grinding are illustrated in these figures



43rd Annual Symposium on Frequency Control - 1989

EFFECTS OF SURFACE ABRASION AND IMPURITY LEVELS ON STRESS-INDUCED DAUPHINE TWINNING IN ALPHA QUARTZ

DAVID A. COCUZZI
Frequency Electronics, Inc.
55 Charles Lindbergh Blvd.
Mitchel Field, NY 11553

and

JAMES W. LAUGHNER
New York State College of Ceramics
at Alfred University
Alfred, NY 14802

Abstract

Two crystals of synthetic quartz were suitably prepared to study the effects of abrasion flaws and impurity content on the levels of stress required to induce Dauphine twinning.

Four sample lots representing three surface treatments and two impurity levels were studied. One lot, designated as a control for purposes of comparison, was prepared from a crystal of "low" purity. Two lots prepared from the same crystal were abraded on one face with SiC grinding media, one with 15 micron particles and the other with 60 micron particles. These flaws were expected to act as stress concentrators and to facilitate the onset of twinning. A fourth group, prepared in the same manner as the control group but from a crystal of higher purity, was expected to be more resistant to twinning.

Two stages of stress induced Dauphine twinning are observed; nucleation and switchover. Nucleation refers to the formation of the characteristic strip shaped twins. Twins nucleate at the surface of a specimen, often at an edge or corner which is in contact with the rams. Switchover is defined as the complete transformation of the specimen, from its initial twin state to the alternate twin state.

Raw data was collected in the form of applied load as read from a strip chart recorder attached to the Instron stressing device. Contact areas of the specimens were determined and stress levels were calculated. These were ranked from lowest to highest and the standard form of the Weibull model was used to plot the data. A Newman-Keuls range test was used to estimate the significance of the difference between mean twinning stresses for all groups for nucleation and switchover.

Initial results indicate that surface flaws and impurity concentration affect the stress level required for twin nucleation and are factors in determining switchover stresses.

Introduction

This study investigates possible relations between surface conditions and impurity levels on the two phases of stress induced twinning described previously.

It is well known that [1, 2, 3] surface flaws and abrasions reduce the ultimate strength of materials, in particular glasses and ceramics. The mechanism by which this occurs is generically referred to as stress concentration. Stress is concentrated at the tip of the flaw under an applied load. At some critical stress the crack lengthens and the material could fail. This critical stress is related to the length of the original flaw and certain material properties and is given by the Griffith Equation

$$\sigma_c = \sqrt{\frac{2E\gamma}{\pi c}} \quad . \qquad (1)$$

E is Youngs modulus for the material, γ is the surface energy per unit area, c is the length of the flaw.

Equation 1 illustrates the expected relation between nucleation stress and flaw size: surface abrasions will concentrate stress and cause localized twin nucleation at lower applied loads than would have been observed had the abrasions not been present or had been smaller in size.

The effects of impurities and related defects on stress levels required for twinning were also considered. The strain energy associated with interstitial and substitutional impurities is thought to contribute to the amount of energy required to nucleate twins or advance twin walls. Therefore, it was expected that lower stress levels would cause twinning events to occur in the samples with higher levels of impurities.

Dislocation density, which has been associated with hydrogen content [4] may also be a factor influencing stress levels required for twinning. Dislocations may act in much the same manner as surface flaws or abrasions. Since dislocation lines generally lie normal to the growth face [4] and Z growth material was used in this study, dislocations in the twinning specimens lie primarily along the Z axis. Figure 1 shows the orientation of the twinning specimens. Dislocations which are present lie parallel to the Z axis and are well positioned to act as stress concentrators.

A study involving the effects of radiation on twinning in quartz by Shaiu [5] indicates that neither neutron nor gamma irradiation produced the expected effect of increasing switchover stress. Electrical and optical observation of gamma-irradiated crystals did seem to indicate suppression of twin formation prior to switchover. This suggests that fairly severe lattice flaws such as Brazil twins [6] or extremely high dislocation densities are required to pin domain walls.

Sample Preparation

Two crystals of synthetic quartz grown by Motorola were used for all experiments. Crystal 1 was grown in the early to mid seventies (run #336-14 T5) and was used for the control group and the two abraded groups.

CH2690-6/89/0000-617 $1.00 C 1989 IEEE

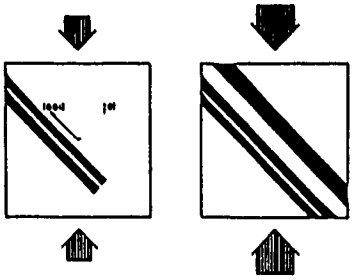


Figure 1. Orientation of twinning specimens used in the experiments. The Y and Z axes correspond to the [120] and [001] directions respectively. The X axis lies perpendicular to the page. Comparison of the two specimens shows the effects of increasing stress on domain wall expansion. (After 11)

Crystal 2 was grown in the early to mid eighties (run #X173) and was used to fabricate the high purity specimens. All specimens had physical dimensions of 2.5mm (± .2mm) x 2.5mm (± .2mm) x 2.5mm (-.1mm). Crystals will often be referred to as the low purity crystal or control group and crystal 2 will often be called the high purity or low hydrogen crystal.

Figure 2 shows the room temperature infrared absorption spectrum of one sample from each crystal. The infra-red samples were approximately 1 cm x 1 cm x .3 cm in the X, Y, and Z dimensions respectively. The spectrometer beam was oriented parallel to the z axis. Both runs were started at the same baseline but the spectra have been displaced vertically for clarity. Several of the primary absorptions for quartz have been indicated. Brice and Cole [7] indicate that absorptions at 3585, 3500 and 3410 cm^{-1} may be linearly related to hydroxyl content. The absorptions at 3410, 3300 and 3200 cm^{-1} have been associated with hydrogen with possible components due to hydrogen plus sodium, silver, lithium and copper [7]. Brice [4] also suggests that there is a correlation between hydrogen content and dislocation density. It can be seen from Figure 2 that absorptions between 3585 and 3410 cm^{-1} are noticeably stronger in crystal 1 than crystal 2. Additionally the absorptions at 3300 and 3200 cm^{-1} are slightly stronger in crystal 1 than crystal 2. This indicates that crystal 2 has a lower impurity content than crystal 1, and in particular, a lower hydrogen and hydroxyl content.

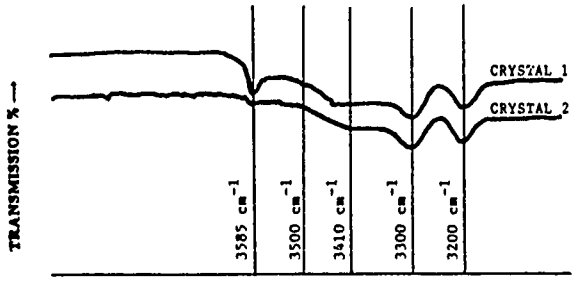


Figure 2. Infrared spectra for the two crystals used in the experiments. Crystal 1 showed noticeably stronger absorption at 3585, 3500 and 3410 cm^{-1}. Absorptions at these wavenumbers have been linearly related to hydroxyl content [9].

Table 1 contains impurity data compiled by Brice and Cole [7] for quartz grown during the early to mid seventies and the impurity data for crystal 2 used in this study.

TABLE 1. IMPURITY LEVELS

Element	Ref. 7	Crystal 2
Al	3	1.62
Fe	0 .1	.21
Li	20	.51
Na	3	.05
K	0.3	.15

*Impurity levels from the study in reference 7 are median values based on several samples from each of twelve crystal producers. These values are taken to be representative of crystal 1, which was grown in the same time period. All values are in atomic parts per million relative to silicon.

There is no impurity data available for crystal 1, but the data compiled by Brice and Cole is taken to be representative of the impurity levels attainable at that time.

The twinning specimens were cut using a low speed diamond abrasion saw. Rough grinding was done with silicon carbide backed pads on a grinding wheel. Fine grinding was done using diamond paste and the samples were polished with aluminum oxide and cesium oxide.

Four sample lots were prepared. One control group and two abrasion groups were fabricated from crystal 1. The abrasion groups were abraded following the polishing procedure, one with 15 micron and the other with 60 micron silicon carbide.

Experimental Procedures

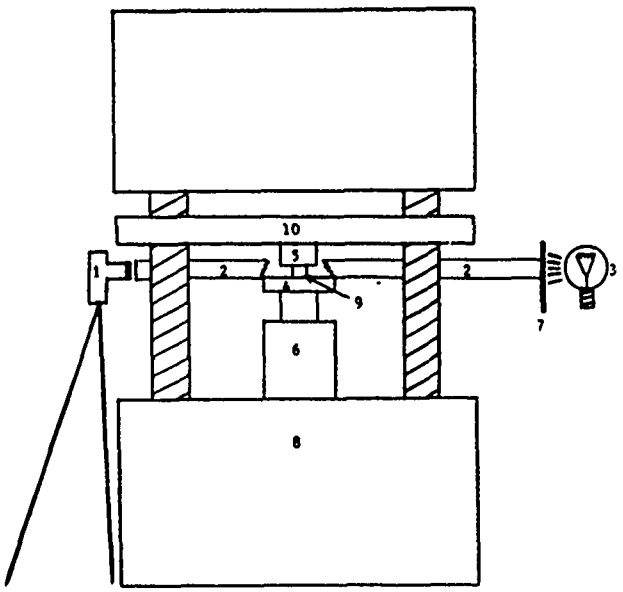
The specimens were twinned using an Instron loading device (model 1123) with a 2500 kg load cell (Figure 3). A stainless steel platen and ram were used. Polished alumina disks were cemented to the center of both the ram and platen to prevent permanent indentation and deformation of the ram and platen.

Specimens were observed through a 35 mm camera mounted on a tripod. To observe Dauphine twinning under stress, it is necessary to view the sample between polarizer and analyzer [5,8,9] and to keep the sample dark relative to the background illumination. The polarizer was mounted on the light source and the analyzer was mounted on the camera lens. A black cardboard tube shielded the specimen from incidental lighting.

The loading rate used for all samples was:

$$8.3 \times 10^{-4} \text{ mm/s.}$$

When twin nucleation was observed, loading was momentarily halted and the load at which nucleation was observed was recorded. Loading was resumed until switchover was observed. At switchover there is an observable drop in load [10]. The load indicated after switchover represents the difference in strain between the two twin states due to the change in sign of the elastic compliance S_{1123}. The maximum load recorded was used in all calculations for switchover.



1 = Camera and analyzer mounted on the tripod.
2 = Cardboard tube used to shield crystal from incident light.
3 = Light source.
4 = Platen used to hold crystal.
5 = Ram used to transmit load to crystal.
6 = Load cell used to detect magnitude of load and strain rate.
7 = Polarizer.
8 = Base of Instron loading device. (Model 1123)
9 = Crystal specimen resting on platen.
10 = Crosshead of Instron.

Figure 3. Schematic of the experimental equipment.

Results and Discussion

Raw data was collected in the form of applied load as read from a strip chart recorder connected to the Instron. The contact areas of the specimens were determined and stress levels were calculated. These were ranked from highest to lowest and a standard form of the Weibull model [12] was used to plot the data. This data is plotted in figures 4 through 9. All data sets were compared to the control group which was an unabraded group of specimens from crystal 1. The Weibull modulus, m [12,13] is shown in the upper left corner of each figure for the data in that figure. The Weibull modulus, m is the slope of the best fit straight line through the plotted data. This number is often used an indicator of the degree of scatter in data. The greater the value of m the less scatter is observed in the data. This type of analysis is often used to characterize the strength of brittle materials such as ceramics and glass. Weibull moduli for these materials are often between 10 and 20 [12]. Figures 4,6 and 8 are the data for nucleation and figures 5,6 and 9 are the data for switchover. It is clear that there is a great deal of scatter in the nucleation data and somewhat less scatter in the switchover data. Tables 2 and 3 contain the mean average, the median average, and the maximum and minimum stress values for nucleation and switchover for all groups. Also tabulated are the variance (standard deviation squared), the Weibull moduli with the calculated standard error range, and the number of samples N.

Nucleation

The Weibull moduli for nucleation are quite low. This suggests that there are several types of flaws present in all samples which can concentrate stress and nucleate twins over a broad range of stress values. Two types of flaws which were considered in this study are surface abrasions, which were purposefully introduced, and growth dislocations whose volume density is related to impurity levels.

The Weibull plot for the 15 micron abrasion group is shown in figure 4. The maximum stress required (see table 2) for nucleation for this group was 464 MPa and the maximum for the control group was over 800 MPa. The maximum for the high purity group was also over 800 MPa. This data indicates that surface abrasions may be concentrating stresses in such a manner that twin nucleation occurs at a relatively low overall stress. Minimum values of stress required for twin nucleation in the 15 micron abrasion group were very similar to the values recorded for the control group.

TABLE 2. SUMMARY OF NUCLEATION DATA

Parameter*	Control	15 Micron	60 Micron	Crystal 2
Average	3.85	3.22	2.87	5.29
Median	3.51	3.22	2.89	5.34
Maximum	8.16	4.64	5.53	8.07
Minimum	1.22	1.08	1.18	2.96
Variance	2.70	0.92	1.81	1.26
$m \pm m(2n)^{-1/2}$	2.55 ± .21	3.24 ± .34	2.15 ± .34	5.01 ± .62
n	73	47	20	33

*Units are Pa x 10^8 for all stress and $(Pa \times 10^8)^2$ for variance.

The maximum stress required for twin nucleation in the 60 micron abrasion group (see figure 6) was also lower than the maximum for the control group and the high purity group. This data suggests that surface flaws act to concentrate stress and lower the overall stress required for twin nucleation. The observation of twin nucleation for the samples with 60 micron abrasions was difficult due to the severity of the abrasions. Therefore it is suspected that the maximum stress value recorded for nucleation is in error and might be too high. This also applies for the minimum value recorded for nucleation for this group.

Figure 8 contains the data for the control group (crystal 1) and the low hydrogen group (crystal 2). It is immediately apparent that the minimum stresses observed for nucleation are quite dissimilar. This is attributed to a difference in dislocation density. The stress distribution is narrower for the low hydrogen group because there are fewer dislocations and growth defects present to cause sufficient stress concentration and induce twinning at the lower stress levels. This suggests that dislocations and growth defects are the dominant nucleation mechanism at the lower stress levels.

Since the control group and the low hydrogen group were prepared similarly it would be expected that they have the same type of surface flaw distributions. This is indicated by the fact that the maximum stresses required for nucleation are nearly identical.

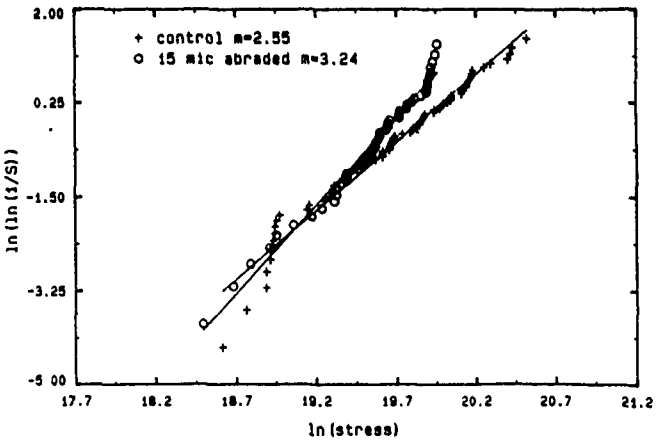


Figure 4. Weibull plot for nucleation of (crystal 1) control and 15 micron abrasion groups.

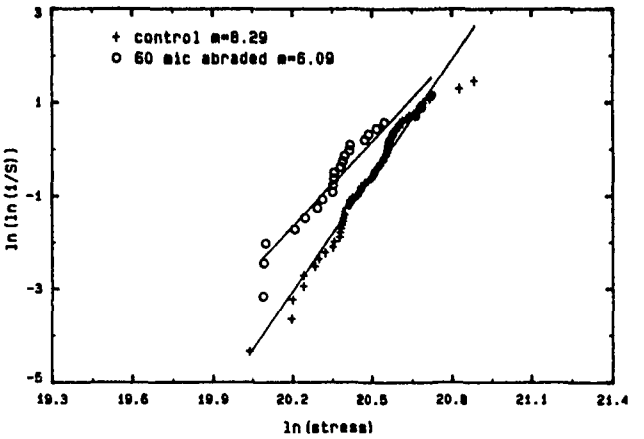


Figure 7. Weibull plot for switchover of (crystal 1) control and 60 micron abrasion groups.

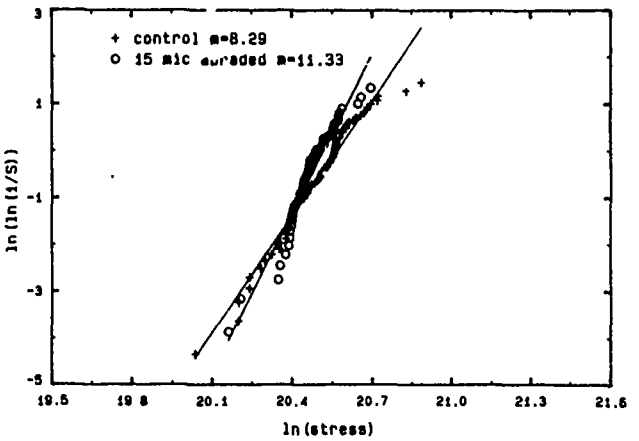


Figure 5. Weibull plot for switchover of (crystal 1) control and 15 micron abrasion groups.

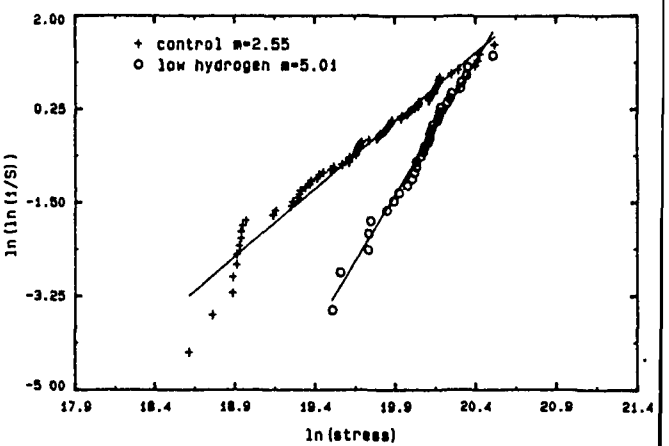


Figure 8. Weibull plot for nucleation of (crystal 1) control and (crystal 2) high purity groups.

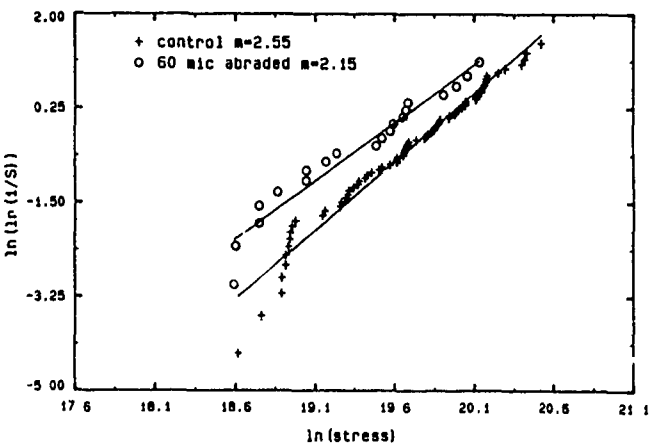


Figure 6. Weibull plot for nucleation of (crystal 1) control and 60 micron abrasion groups.

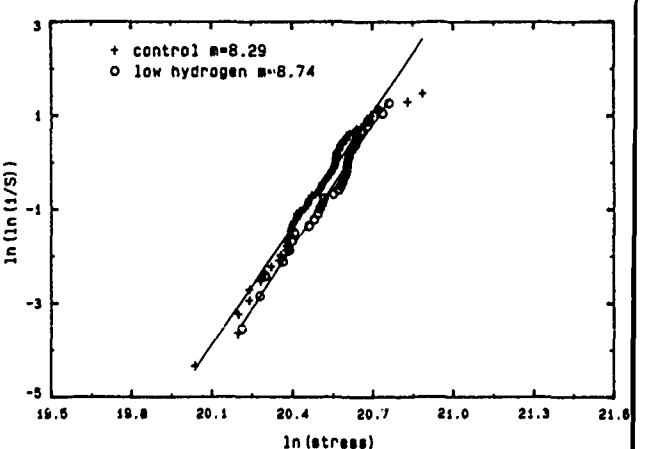


Figure 9. Weibull plot for switchover of (crystal 1) control and (crystal 2) high purity groups.

Switchover

Weibull moduli calculated from switchover data are higher than those calculated from nucleation data because the stress values for switchover tended to group within a narrower range around a central value with fewer data at the extremes of the distribution. This is true for all sample groups (see Table 3).

TABLE 3. SUMMARY OF SWITCHOVER DATA

Parameter*	Control	15 Micron	60 Micron	Crystal 2
Average	8.09	7.79	7.23	8.43
Median	8.18	7.72	7.31	8.70
Maximum	11.77	9.70	10.01	10.37
Minimum	5.05	5.71	5.30	6.00
Variance	1.27	0.58	1.64	1.08
$m \pm m(2n)^{-1/2}$	8.29 ± .67	11.33 ± 1.18	6.09 ± .9	8.74 ± 1.62
n	76	47	23	34

*Units are Pa x 10^8 for all stress and (Pa x 10^8)^2 for variance.

Table 4 contains the results of a Newman-Keuls range test performed on all combinations of sample pairs for nucleation and switchover. These comparisons were made at the α = .05 (95% confidence) level. Sample pairs whose mean switchover (or nucleation) stresses were not significantly different could have come from a common population. These will be discussed in the order they appear in table 4.

TABLE 4. RESULTS OF NEWMAN-KEULS RANGE TEST

Sample Pairs	Nucleation	Switchover
Control and high purity	**	
Control and 15μ abraded	**	
Control and 60μ abraded	**	**
High purity and 15μ abraded	**	**
High purity and 60μ abraded	**	**
15μ and 60μ abraded		**

A double asterisk (**) indicates that the difference between means of the sample pair was found to be significant at the 95% confidence level.

Table 4 indicates that there was no difference between mean switchover stress for the control and low hydrogen pair and the control and the 15 micron abrasion pair. The remainder of possible combinations of sample pairs have significantly different mean switchover stresses.

The mean switchover stress for the 60 micron abrasion group was about 10.4% (86 MPa) less than that observed for the control group. Again, as stated earlier, these figures might be in error because of the difficulty encountered in observation. This may also be the cause of the scatter in this data which is indicated by the relatively low Weibull modulus. It appears that abrasions of this size are effective in influencing switchover behavior. It should be noted that these abrasions were quite deep (.06 mm), approximately 2.4% of the smallest dimension of the twinning specimens.

The combination of purity level and surface abrasion caused a significant difference in mean switchover stress between the low hydrogen group and the two abrasion groups. It is not clear why surface abrasion should influence switchover stress. It is known [9, 10] that twins tend to nucleate on the +X or -X face for the orientation used in these experiments. Surface abrasions were expected to play a role during twin nucleation and purity and dislocation density were expected to have an influence on switchover. It was seen earlier that surface abrasions tend to dominate twin nucleation stresses at the upper end of the stress distribution. That is, if surface abrasions are present, all samples will nucleate twins before a certain level of stress is attained. The presence of the nucleated twin and stress-concentrating flaws might facilitate the movement of domain walls through the volume of the specimen. If the specimen is small enough, the stress fieldassociated with the abrasion might extend through a large portion of its volume. If there is a nucleated twin present, its expansion throughout the volume is facilitated by a properly oriented stress field.

Conclusions

1) Abrasions reduce the stress required for twin nucleation. They appear to be the dominating stress concentration mechanism at the upper end of the stress distribution.

2) Sample groups prepared from the lower purity crystal 1 showed a minimum nucleation stress that was nearly one third that observed for the higher purity crystal 2. This suggests that growth dislocations resulting from impurities could be an influence in lowering nucleation stresses at the lower end of the stress distribution.

3) Relatively severe abrasions (60 micron) appear to influence switchover stress.

4) There was no statistical difference between the mean switchover stress for the control group (low purity crystal 1) and the low hydrogen (high purity crystal 2). When the low purity samples are abraded there is a statistical difference in mean switchover stress between the abraded group and the high purity group.

References

[1] C.E. Inglis, "Stress in a Plate due to the Presence of Cracks and Sharp Corners", Transactions of the Institution of Naval Architects, Volume 55, pp. 22-1230 (1913).

[2] A.A. Griffith, "The Phenomena of Rupture and Flow in Solids", Philosophical Transactions of the Royal Society of London, Volume 221, No. 5, pp. 163-198 (1920).

[3] K.J. Pascoe, "Physical Basis of Fracture", Journal of Materials Education, Volume 3, No. 4, pp. 735-751 (1981).

[4] J.C. Brice, "Crystals for Quartz Resonators", Reviews of Modern Physics, Volume 57, No. 1, pp. 105-146 (1985).

[5] S. Shiau, "The Effect of Radiation on Twinning in Quartz", PhD thesis Solid State Science, Pennsylvania State University (1986).

[6] T.W. Cline, J.W. Laughner, R.E. Newnham and L.E. Cross, "Electrical and Acoustic Emission during Ferrobielastic Twinning in Quartz", Proceedings of 32nd Symposium on Frequency Control, pp. 43-49 (1978).

[7] J.C. Brice, A.M. Cole, "The Characterization of Synthetic Quartz Using Infra-Red Absorption", Proc. 32nd Symposium on Frequency Control, pp. 1-10 (1978).

[8] K. Aizu, "Second Order Ferroic State Shifts", Journal of the Physical Society of Japan, Volume 34, No. 1, pp. 121-128 (1973).

[9] J. Laughner, "Induced Dauphine Twinning in Quartz", PhD thesis Solid State Science, Pennsylvania State University (1982).

[10] J. Markgraaff, "Elastic Behavior of Quartz During Stress Induced Dauphine Twinning", Physics and Chemistry of Minerals, Volume 13, pp. 102-112 (1986).

[11] R.E. Newnham, "Domains in Minerals", American Minerologist, Volume 59, pp. 906-918 (1974).

[12] G. Weaver, "Engineering Ceramics Part 1. the Weibull Model", Journal of Materials Education, Volume 5, pp. 773-804 (1983).

RAPID,NON-CONTACT,DAMAGE FREE SHAPING OF OPTICAL & OTHER SURFACES WITH PLASMA ASSISTED CHEMICAL ETCHING

C.B.Zarowin & L.D.Bollinger
Research Dept., AST Sector
Perkin Elmer Corporation
100 Wooster Heights Road
Danbury Connecticut 06810

INTRODUCTION

Experimental results are presented demonstrating that Plasma Assisted Chemical Etching (PACE) can rapidly and controllably figure (shape) and smooth optical surfaces without mechanical contact. This process, for example, significantly reduces the constraints on the design of quartz optical elements imposed by mechanical processes, such as the requirement of substrate rigidity and the limitation of sphericity of the resulting surface. It offers similar possibilities for other applications with material removal rates of the order of 1 mm^3/min. It employs a plasma etch process which takes place at gas pressures of about 10 Torr (~0.01 atm) originally developed to pattern microelectronic circuits by etching through (openings in) photolithographically patterned masks(1)(2). After describing our progress to date in figuring and polishing quartz and other materials with the PACE process, the pertinent chemistry and physics of this process will be summarized to provide a basis for understanding its characteristics. In contrast to the mechanical process, a theory of surface evolution applicable to PACE (3), permits us to show the range of surface topographies predicted for PACE, showing, among other things, that isotropic and anisotropic etch processes both intrinsically micro-smooth features smaller than 0.1 mm and that removal smooths while deposition roughens.

The shaping of a quartz surface with PACE, for example, takes place by removing material from a region under a confined reactive gas plasma moved over this surface (see Fig(1)), the dwell time and removal rate determining the amount of removed quartz. We have obtained rates of removal for figuring fused silica as high as 10 microns per minute with open loop control to about 1%. The static removal "footprint" can be varied during the process, as shown in Fig(2). PACE inherently polishes or smooths while etching material, as shown in Fig(3). The process is benign, exposing virgin surfaces free of process generated contamination and subsurface damage. It is a ("dry" or gaseous) chemical process, free from any high ion energy species that can cause sputter induced subsurface damage because it employs low energy (<10eV) plasma species that only enhance the chemical rates(Fig(4).

PACE has two operating modes, shown in Fig(1): (i) (right) in "contact" with plasma generated short lived species such as ions, which "deposit" super-thermal kinetic energy on striking the reactive surfaces; (ii) (left) downstream of the plasma, where longer lived species, such as excited metastable neutrals ("free radicals"), which release their super-thermal stored energy at the surface.

Although other materials (see Table(1)) can also be figured by a PACE process (frequently, at much higher rates) for this study, an apparatus was developed to explore the figuring of fused silica. Results are shown demonstrating: repeatability and control of removal rate and "footprint"; repeatable and predictable removal rate dependence on "tool" slew rate(Fig(5); smoothing of initially rough surfaces; freedom from optical edge figuring distortion; and evidence that the process does not introduce subsurface damage or contamination. Finally, we show the results of figuring fused quartz substrates to wedged (Fig(6)), and spherical (Fig(7)) surfaces, in addition to demonstrating PACE's figure correcting ability (Fig(8)), by neutral removal or removing mapped figure errors. Although PACE is highly accurate and repeatable without rate controlling feedback (open loop), it is also compatible with in situ interferometric figure monitoring, and, in addition, from the intensities of spectroscopic lines of the etched species in the plasma, which are an accurate (<<0.1%) measure of the amount of material removed, as shown in the experimental apparatus used to perform the figuring (Fig(9)). The PACE removal chemistry used for fused quartz is shown in Table (1), where the reaction converts the solid quartz and gaseous carbon tetrafluoride to the vapor silicon tetrafluoride and gaseous carbon dioxide.

OUTLINE OF CHEMISTRY & PHYSICS OF THE PACE PROCESS

The PACE process uses plasma assisted chemical vapor transport (PACVT) to remove (etch) a solid by converting it to a volatile compound in a reaction with a gas. The plasma assistance of this heterogeneous chemistry speeds up the rates without significantly raising the surface temperature. Control of this process then requires control of the flux and energy of species striking the reacting surface. This is achieved by controlling the parameters of the plasma in which they are generated. Thus, there are three components to PACE : 1-chemical vapor transport (CVT); 2-plasma assistance of the CVT (PACVT); 3-plasma generation and control of the number and energy of species suitable for such assistance.

1-CHEMICAL VAPOR TRANSPORT (1)(2)

For the moment, postponing consideration of the plasma assistance of this chemistry to the next section, we outline the properties of the CVT process: a solid (S, e.g., quartz) reacts with a reactive gas (A, e.g., CF_4) to produce (a) volatile compound(s) (B, e.g., SiF_4 & CO_2). At a single surface (at surface temperature, T), such a generic heterogeneous chemical reaction, as shown in Fig (10), is:

$$S + A \underset{b(T)}{\overset{a(T)}{<\!\!-\!\!-\!\!-\!\!-\!\!-\!\!>}} B$$

with a forward rate, a(T), and a back rate, b(T). The rate of generation of B (or, equivalently, consumption or etch rate of S) implied by the above reaction is: $d[B]/dt = a[A] - b[B]$. From this rate relation, the limiting transport rate (for "removing" S) is easily obtained for a total number of molecules in the reactor, $[N] = [A] + [B]$, at a given total pressure, $p = [N]kT$:

$$d[B]/dt = a[N]\{1 - ((a+b)/a)[B]/[N]\}$$

where [A] and [B] are the number of molecules of A & B, respectively. For a small concentration of the volatile reaction product, the initial fractional rate of removal is equal to the forward rate constant,

$$(1/[N])d[B]/dt \sim a \qquad (\text{when } [B]/[N] << a/(a+b))$$

Chemical equilibrium occurs when $[B]_0/[N] = a/(a+b) = k/(k+1)$, where the equilibrium concentration of $[B] = [B]_0$ and the equilibrium constant $k = a/b$.

Between the two or more such reactive surfaces shown in Fig (11), with reaction equilibrium constants, $k_1 = a_1/b_1$ and $k_2 = a_2/b_2$ (i.e., at surface temperatures, T_1 & T_2), with no flow in/out of the reactor, there is transport between the surfaces (e.g., etching at surface "1" and deposition at "2") at a rate proportional to the difference (k_1-k_2). In terms of the reaction Gibbs Free Energy and surface temperatures, the equilibrium constants of these surface reactions are:

$$k_{1,2} = \exp(-DG/kT_{1,2})$$

Thus, for small temperature differences, the transport rate is proportional to $DG(T_1-T_2)$, so that transport is from surface "1" (etch) to surface "2" (deposition) when $DG > 0$ (endoenergetic), but from "2" to "1" if $DG < 0$ (exo-energetic). Put another way, transport is from hot to cold for endo-energetic reactions and from cold to hot for exo-energetic reactions. The latter is exemplified by the reactive gas filled incandescent lamps in which exo-energetic CVT of tungsten reacting with halogens &/or oxygen compensates for the higher evaporation rate of the hotter (and brighter) filament by etching at the cool envelope and depositing on the hot filament. When super-thermal ions provide an energy source to "drive" the transport reaction, the surface bombarded by more energetic ions acts as the "hot" surface, providing control in directing the transport

CH2690-6/89/0000-623 $1.00 C 1989 IEEE

Finally, flow (of B out of the reactor) adds its own transport component algebraically (and non-linearly) to differences in surface "temperatures", this is shown in Fig(12), in which the transport rate at the etch surface ("1") is increased by the flow of A into and B out of the reactor while the rate of deposition on the other surface ("2") is reduced to zero and then starts to etch at higher flows. Thus, flow provides another useful control parameter for such CVT processes.

2-PLASMA ASSISTANCE OF THE TRANSPORT CHEMISTRY

Experience (1)(2) shows that the rate of the CVT reactions increases when the surface on which the chemistry occurs is subject to super-thermal species having either kinetic (as for ions made energetic in a plasma sheath field) or potential energy (as for metastable excited neutrals generated by inelastic collisions with electrons). The details of this enhancement remains argumentative, but by following conventional chemical rate theory, these can non-the-less be represented as due to a reduction of the chemical activation energy. Such rate theory requires that:

$$a(T) = Z^{*}\exp(-E_a/kT) \text{ and } b(T) = Z^{*}\exp(-E_b/kT),$$

where Z is a frequency, kT is the surface thermal energy and $E_{a,b}$ are the forward/back activation energies. When reaction rates, a & b, increase for fixed Z and kT, $E_{a,b}$ must have decreased. Letting $E_{a,b}(W_+) = E_{a,b}(0) - U_{a,b}$, where $U_{a,b}$ represents the reduction of the activation energy and is an increasing function of the energy (kinetic or stored) of the plasma species striking the surface (W_+). This allows us to summarize plasma enhancement of the transport reaction as a reduction of the reaction activation energies. We may speak of a "hot" surface as one for which the reaction rate increase occurs whether or not it is thermally hot. It should be noted that purely thermal reactions (no activation energy reduction) coexist (occur in parallel) with those of reduced activation energy, say, due to release of kinetic or potential energy at the reacting surface.

3-GENERATION AND CONTROL OF PLASMA SPECIES

Electrical generation of a plasma, i.e., ionizing the gas, involves making the gas conductive by causing a voltage break down. This frees electrons from a small fraction (10^{-7} to 10^{-3}) of the gas atoms to generate positive ions. Equal numbers of ions and electrons are generated. On the other hand, the electron diffusivity and mobility is many orders of magnitude greater than that of the ion. Because of their smaller mass the electrons are "heated" more (or extract more energy) by the electric fields than the heavier ionic species. Consequently, the electrons diffuse more rapidly to walls or electrodes where they become neutralized. This causes the central part of the plasma, the plasma body, to become slightly charged to a plasma potential of +30 to +300 volts (2), allowing the electron and ion loss rates to equilibrate. The energetic electrons in the plasma body collisionally excite strong spectral emission, mostly from the predominant atoms. Near the electrodes, in a region called the plasma sheath, there lower density of electrons than ions and across which these ions are driven by the sheath fields caused by the voltage between the plasma body and the electrodes. This region has less spectral emission both because there are fewer electrons to collide inelastically with the subsequently emitting atoms and because they are less energetic, having been reduced in energy by this sheath field.

Although the range of electrical excitation frequencies can vary widely, the time average effects of delivering energy to the charged species in the plasma is the similar (but not identical) at all frequencies Typical excitation frequencies are 10KHz to 13.56 MHz, and less commonly, microwave frequencies. The time average schematic of such a plasma between electrodes, one or more of which is the reactive solid, with two plasma sheaths and a plasma body, is shown in Fig(13).

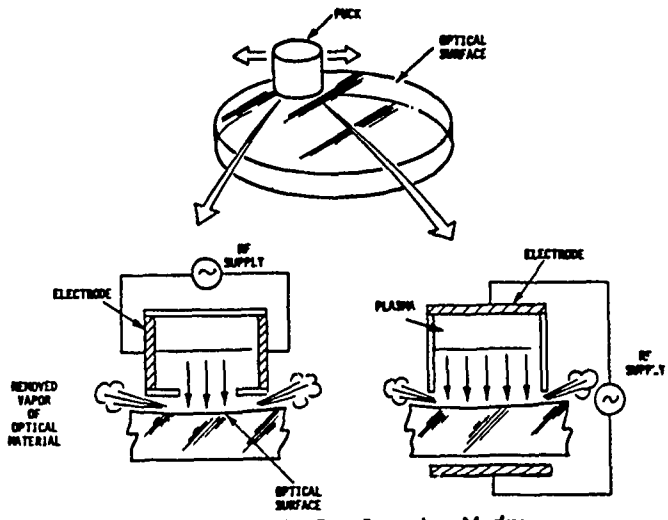
In the plasma sheaths, the ions with a positive charge, e, on the average, collide with the predominantly neutral species with a mean free path, l_+, so that between collisions they "see" a mean electric field, E_s, and gain an energy, $W_+ = eE_s l_+$. Since $(E_s)^2$ is proportional to rf power density and the mean free path is inversely proportional to the pressure, the ion energies can be shown to be proportional to the square root of the rf power density, P_{rf}/V divided by the gas pressure, p. Similar behavior occurs for the electrons in the plasma body ($W_- = eE_p L$, where Ep is the plasma electric field), so the electron and ion energies are both proportional to $(P_{rf}/V)^{1/2}/p$, showing that the geometry of the plasma is also important in controlling the PACVT rates

CONCLUSIONS

We showed that the PACE figuring and polishing process is able achieve (at least) comparable results the conventional mechanical process for optical surfaces, while offering significant advantages that also invite reappraisal of our present commitment to mechanical processes for other precision shaping processes. PACE is not applicable to all materials because there are sometimes no known CVT reactions for a particular material. None-the-less, PACE does offer intriguing possibilities for other shaping and machining applications that are inaccessible to conventional processes or for which such processes are difficult to control, particularly in the shaping of unusual non-spherical surfaces and to high precision.

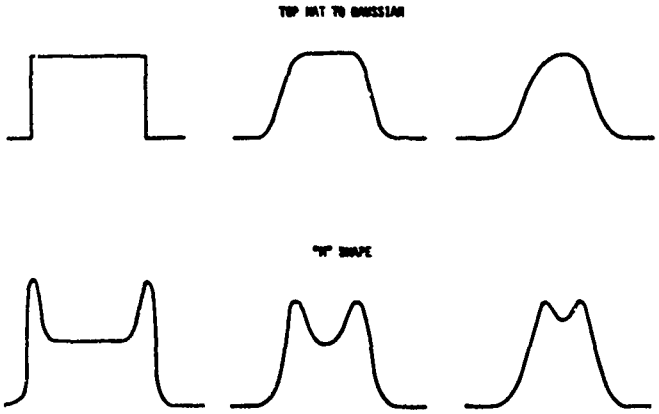
REFERENCES

(1) C.B.Zarowin., J.Appl.Phys., 57(3), 929, 1985.
(2) C.B.Zarowin., J.Vac.Sci.Tech.(A) , 2(4), 1537, 1984.
(3) G.Gallatin & C.B.Zarowin, J.Appl Phys. (1989).

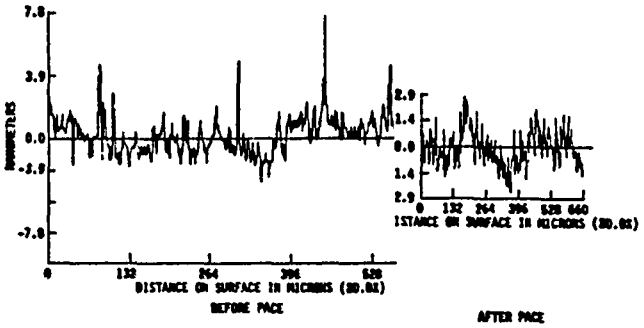


Fig(1) PACE Schematic: Two Operating Modes

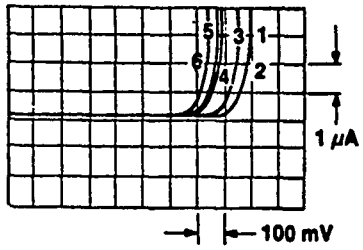
MATERIAL	(SOME) POSSIBLE GASES	EXAMPLE
FUSED SILICA (SiO_2)	CF_4, SF_6, NF_3	$SiO_2 + CF_4 \rightarrow SiF_4 \uparrow + CO_2 \uparrow$
SILICON CARBIDE (SiC)	CF_4, SF_6, NF_3	$SiC + NF_3 \rightarrow SiF_x \uparrow + C_y \uparrow$
BERYLLIUM (Be)	Cl_2	$Be + Cl_2 \rightleftharpoons BeCl_2$



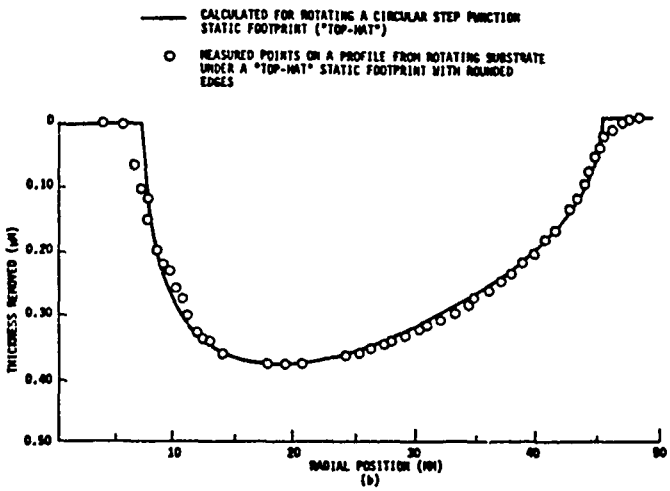
Fig(2) Various Static PACE "Footprints"



Fig(3) Smoothing/Polishing w/PACE

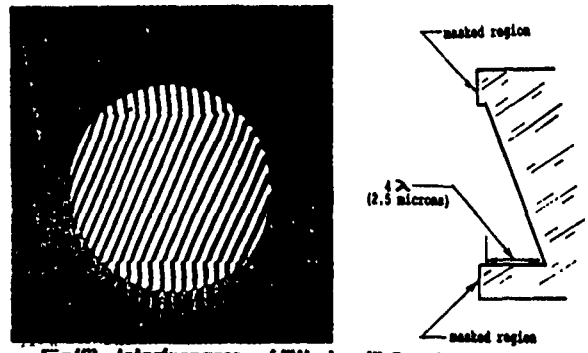


Fig(4). No Subsurface Ion Damage at Low Ion Energy
Current-Voltage Plot for a Surface Subjected to Ion Bombardment:1-none; 2- 30 eV; 3- 60 eV; 4- 100 eV; 5- 150 eV; 6- 210 eV; Ashok, Chow & Baliga, Appl.Phys.Lett.42,(8)687(1983)

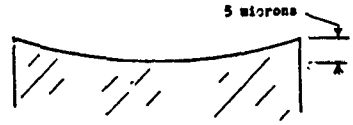


Fig(5) Calculated & Experimental Dynamic "FootPrint"

THIS 4-WAVE WEDGE WAS FIGURED BY PACE. SPATIAL CONTROL OF THE DEPTH OF REMOVAL WAS OBTAINED BY THE DWELL TIME OF THE PLASMA REACTOR AS IT WAS MOVED OVER THE FUSED SILICA SURFACE.
(4" diameter x 1/2" thick fused silica)



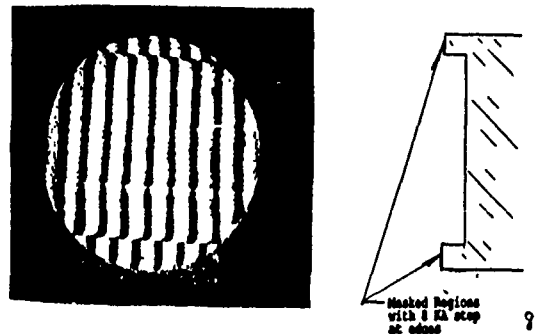
Fig(6) Interferogram of "Wedged" Quartz Figure



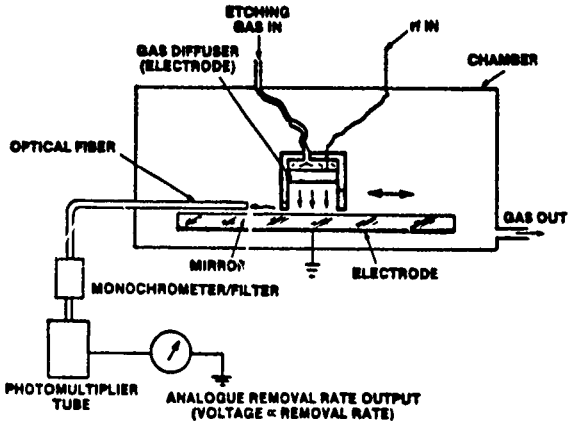
SPHERICAL SURFACE FIGURED BY PACE
(4" Diameter x 1/2" thick fused silica)

Fig(7) Interferogram of "Spherical" Quartz Figure

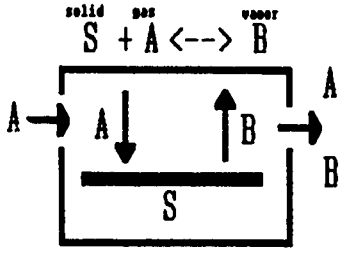
PACE PRODUCED THIS OPTICALLY FLAT SURFACE WITHOUT INTRODUCING SURFACE ERROR BY REMOVING MATERIAL FROM AN INITIALLY FLAT SURFACE (NEUTRAL REMOVAL).
4" Dia. x 1/2" thick fused silica, plasma tool size 1" dia.



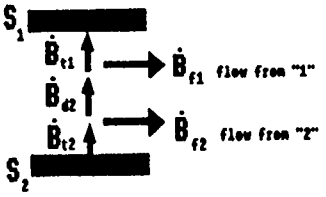
Fig(8) Interferogram of "Flat/Neutral Removal" Quartz Figure



Fig(9) Schematic Experimental PACE Figuring Reactor



Fig(10) One Surface Transport

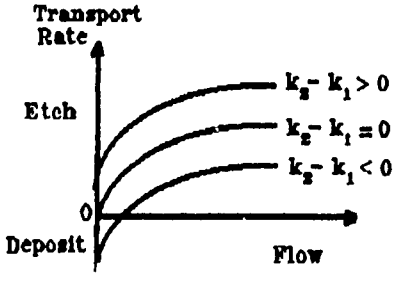


$\dot{B}_{t1}$ - transport to "1" (deposit)

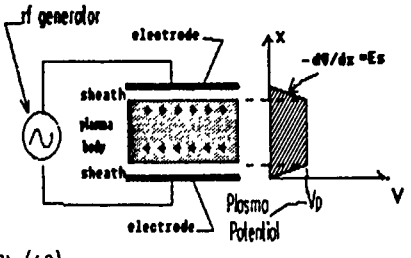
$\dot{B}_{d2}$ - diffusion "2" to "1"

$\dot{B}_{t2}$ - transport from "2" (etch)

Fig(11)Two Surface Transport



Fig(12) Transport Rate vs Flow



Fig(13) TIME AVERAGE RF GAS DISCHARGE

CRYSTAL ADJUSTMENT BY LASER CHEMICAL DEPOSITION

R. J. Williamson
STC Technology Ltd, London Road,
Harlow, Essex, England, CM17 9NA

Abstract

Maskless adjustment of quartz crystals has been developed using laser induced deposition of material. This is achieved using an excimer laser operating in the UV which dissociates volatile precursors leaving a stable material. This material is deposited where the laser strikes the surface of the crystal and hence can be used to adjust it. A simple means of beam steering has been developed and the necessary process parameters such as laser fluence and precursor flow rate have been investigated. Problems such as window fouling and gas phase reactions have been solved. Up to 2000 ppm of adjustment to crystal resonators up to 125 MHz fundamental has been demonstrated. Initial ageing of canned crystals looks very promising.

Much of the present work on laser deposition has been undertaken using trimethyl aluminium (TMA) as the precursor. This has proved to have a number of problems which have largely been solved. The other precursor used has been chromyl chloride (CrO_2Cl_2) which deposits out as chromium dioxide, and which is now used as the final process.

Introduction

The usual method of adjusting a quartz crystal to frequency is by adding extra mass to the electrode surface evaporated through a complex machined mask from an incandescent filament loaded with a metal. Since the crystal has to be active during this process, the mask must be out of contact. Much above 20 MHz, the reduced electrode dimensions, together with the tolerances from the crystal mount and filament position limit the accuracy of placement. Furthermore, with high frequency fundamental devices, any external strain induced by the mask can cause undesirable frequency shifts. Adjustment of dual resonator devices, with the requirement to adjust in specific areas, becomes very problematical.

Laser adjustment of crystals is regularly used in production. Typically this uses a Q switched YAG laser operating at 1.06 micron which ablates a small area of the electrode through to the quartz and hence reduces the mass loading. However, this method is limited in the amount of adjustment that can be achieved before low Q and drive level effects dominate. This appears to be caused by damage to the electrode/quartz interface at the adjustment spot together with fall out of the metal back onto the electrode.

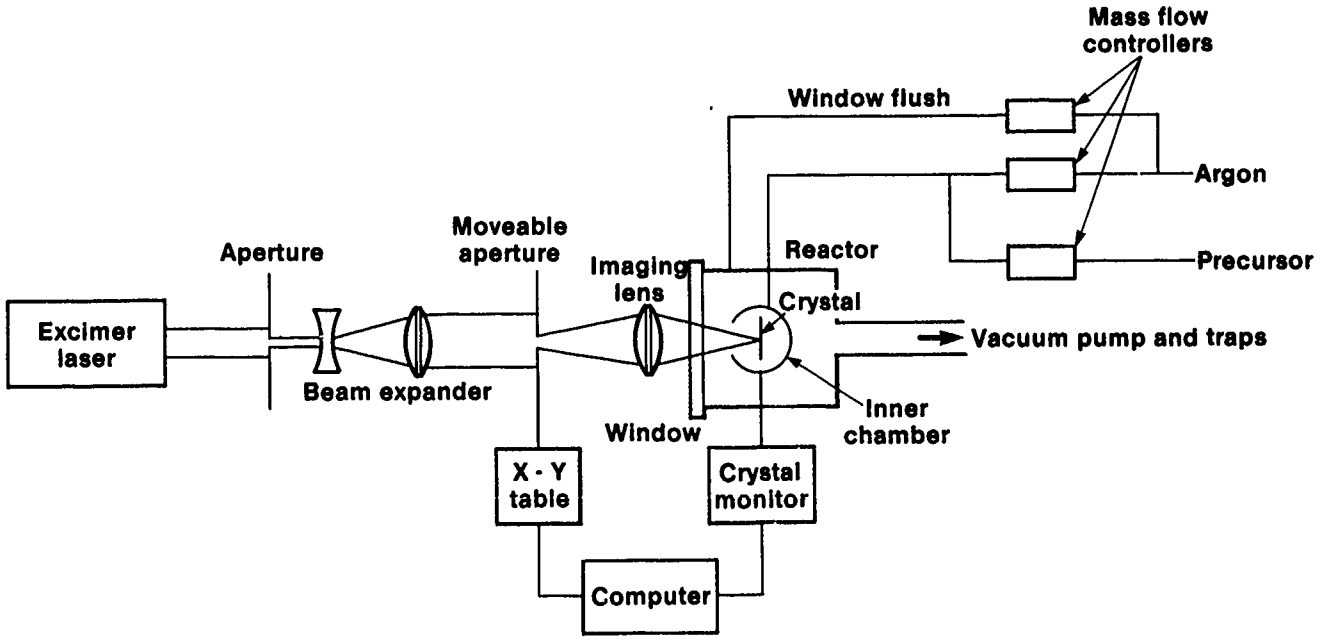


Fig.1 General layout

CH2690-6/89/0000-627 $1.00 C 1989 IEEE

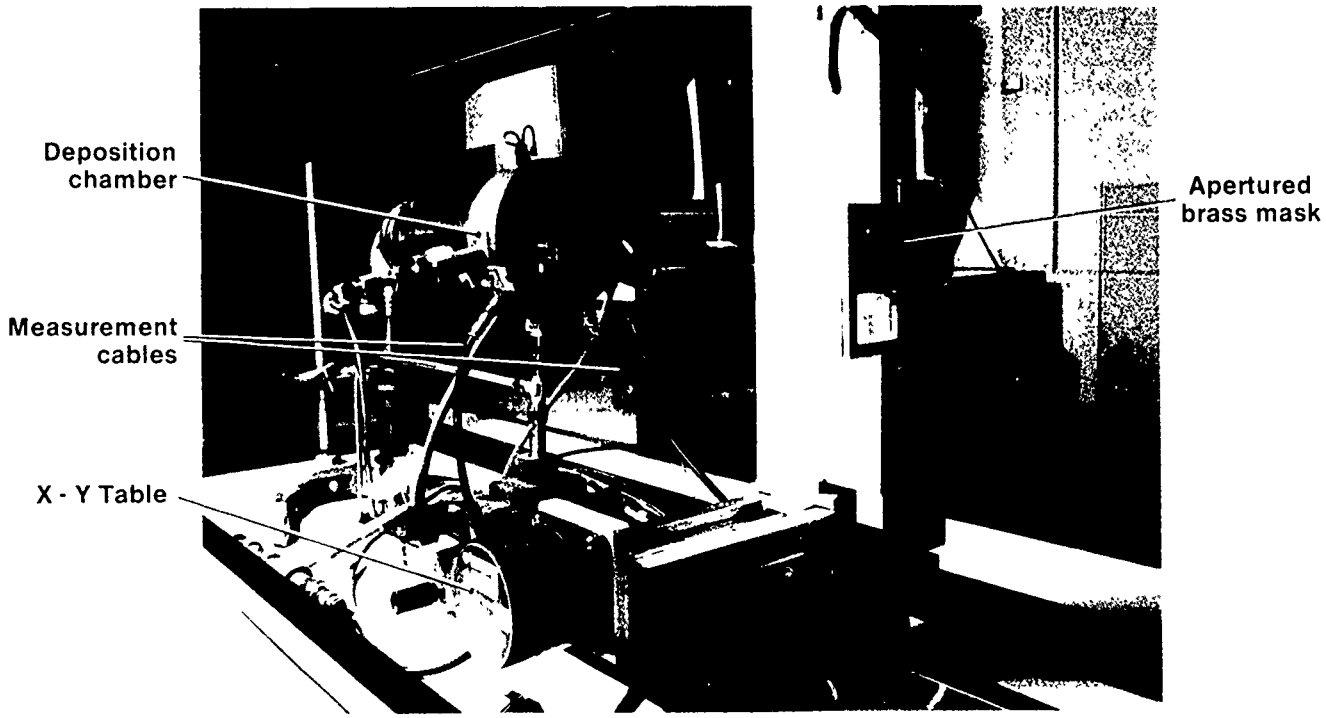


Fig.2 Optical bench with deposition chamber and driver

In chemical vapour deposition (CVD process), precursors such as organometallics are used to plate out materials onto a substrate. There are a number of organometallics available, and they often have sufficiently high vapour pressures which enable their transport into a vacuum system. In traditional CVD, the heat in the substrate causes the bonds between the metal and the rest of the molecule to break, resulting in the deposition of the metal. With laser deposition, the energy is supplied by photons, and the area of deposit is defined by the size of the laser beam. The need for a mask is therefore eliminated.

Equipment

The overall layout of the experimental equipment is shown in Figures 1 and 2 and comprises six sections, namely:

1. The deposition reactor with the quartz crystal device to be adjusted.
2. The frequency monitoring system.
3. The gas handling system.
4. The laser beam source.
5. The beam definition and placement facility.
6. Location of resonator area.

A detailed description of the various sections is given below:

A. The Deposition Chamber

The device is held in a holder inside an (inner) reaction chamber and connected to the measurement system (Figure 3). The reaction chamber is of the chimney type with the precurser gas admitted at the bottom. A hole in the side of this chamber allows the laser light to impinge on the electrode of the crystal.

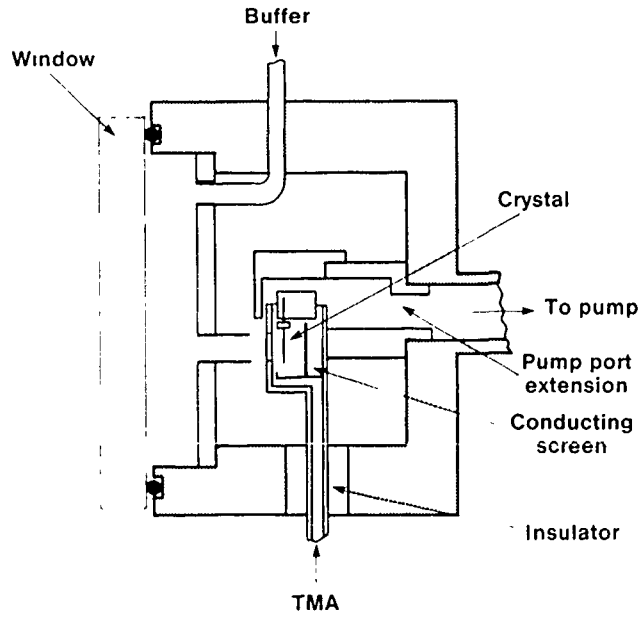


Fig.3 Deposition chamber

The inner chamber has been electrically isolated from the outer chamber and an extra socket installed allowing electrical connection to the inner chamber. This has enabled bias voltages to be applied. The exit to the pump has been extended into the chamber so that the inner chamber is more directly pumped. In this way the spent precursor is pumped away rapidly and hence prevented from being recirculated within the outer chamber.

The main chamber is pumped from the rear end using a double stage roughing pump and a liquid nitrogen trap. The chamber can be heated using three strip heaters clamped to the back, and is normally held to within 0.2°C at an elevated temperature (preferably at the minimum of the f(T) curve).

The main vacuum chamber has a window made from 6 mm thick Spectrasil (silica) which transmits UV radiation to below 200 nm. An inner window is placed between this and the device, to minimise deposits at the window surface. A buffer gas is fed into the space between the two windows and excited through a hole in the centre which also passes the laser beam.

For loading, the windows are removed, the inner chamber moved down, and the device inserted into its holder. After loading, the device is kept clean by flushing the system with zero grade argon through all operations. This chamber-in-a-chamber design maintains 125 MHz devices, for instance, stable to within 1 ppm over many hours, which is perfectly adequate for the present investigations, and makes fast cycle times possible without involved heating and pumping cycles.

For deposition, the flow of the buffer gas is partitioned, the smaller fraction being mixed with the reactant gases and passed through the inner chamber, and the larger fraction directed to flush the front window in the beam area to avoid deposition on the inner surface.

B. Frequency Monitoring

A flexible frequency monitoring system capable of operating at any frequency up to the maximum frequency envisaged (1 GHz) has been devised. The crystal is measured in transmission using the cables entering the lower part of the chamber through coaxial feed-throughs.

The method is based on the IEC 444 measurement system but with the d.c. frequency modulation input of the signal generator driven by the phase output of the vector voltmeter, as shown in Figure 4, in order to lock onto the crystal.

C. Gas Handling

Conventional gas handling components are used. The gas manifold is fed by three mass-flowmeters for the precursor and for the buffer gas (Argon). A separate branch of buffer gas is used for the window flush and the proportion of flow between this and the manifold is set by a pair of needle valves. The precursor cylinder is located in a vented steel cabinet.

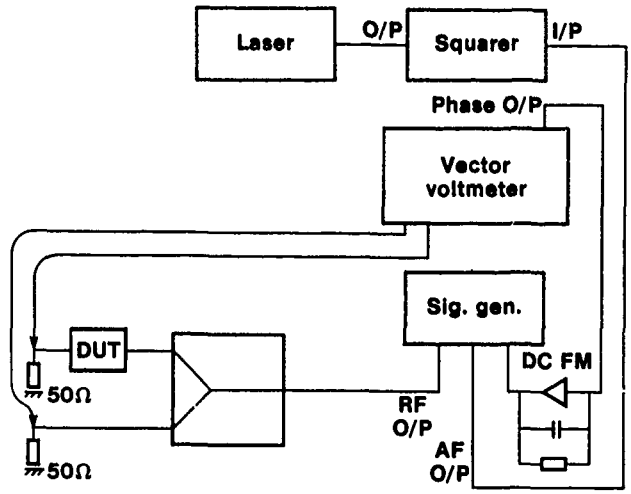


Fig.4 Measurement layout

D. The Laser Source

The laser used for the current work was a Questek 2040 excimer laser. Excimer lasers have the advantage over other available laser sources in that they operate in the UV where the chemical reactant gases for this work absorb most strongly, and that the operating wavelength may be varied by changing the filling gas. In operation, excited dimers are formed between a noble gas and halogen, which radiate at the following wavelengths: 193 nm (ArF*), 222 nm (KrCl*), 248 nm (KrF*), 30-8 nm (XeCl*) and 351 nm (XwF*).

The output energy of the laser can be varied between 20 and 200 mJ per 15 ns pulse with a repetition rate up to 100 Hz. The output beam is about 10 * 20 mm^2 so that the fluence is between 10 and 100 mJ/cm^2. The large beam size enables a very simple means of beam placement as shown in Figure 1. A computer controlled X-Y table moves a brass foil containing a number of apertures thus enabling different areas to be irradiated with the same fluence. In front of the foil is a beam expander which allows lower fluences to be used and evens out the intensity distribution across the beam. A simple lens system is positioned at twice the focal length from both the foil and the crystal, to minimise the diffraction effects from the apertures.

The Deposition Process

A number of papers have reported on laser deposition [1, 2], where high temperature, partial pressures and fluences can be used. The restriction of deposition on an active, temperature sensitive device is central to this investigation. For instance, the effects of laser fluence need investigation since excess energy will lead to heating effects and hence the uncertainty in the crystal frequency. In like manner the excess pressure of the precursor will lead to an uncertainty as to the mass loading and hence frequency of the device. However, too low a value of both these parameters will result in too slow a deposition. Clearly there is an optimum window of operating conditions. A description of those conditions is given below.

A. Photodeposition

There exist a number of organometallics where a metal (for instance aluminium) is combined with organic groups such as methyl. Trimethaluminium (TMA) is one such chemical with a vapour pressure of about 12 torr at room temperature. With sufficient energy (hard UV of 193 nm produced by an argon fluoride excimer laser) the bonds between the metal and the organic radical can be broken leaving the metal to condense out. In practice TMA is a dimer and one photon is required to split this bond with another photon for each organic group, thus requiring four photons for complete dissociation.

TMA is pyrophoric and reacts with any absorbed surface water. All the surfaces of a crystal are typically of this type and on first exposure to the TMA the initial chemisorbed layer can shift frequency almost as much as the required adjustment. This initial shift can be reduced to negligible levels by a preclean of the crystal by exposing the whole crystal to the laser.

It is also necessary to remove any absorbed precursor layers prior to bringing an adjusted crystal up to air, otherwise a milky deposit forms. The laser 'fixes' the TMA wherever it reaches. However, this may not be reliable on the reverse side of the crystal. Furthermore, all the surfaces of the inner chamber are covered with TMA and will recontaminate the crystal surfaces.

A solution to this problem was found by using a plasma. By applying a bias voltage to the inner chamber, a plasma may be produced. The optimum conditions of voltage and gas pressure are given by the Paschen curve. For setting it up, the novel technique of 'listening' to the plasma has proved useful [Figure 5].

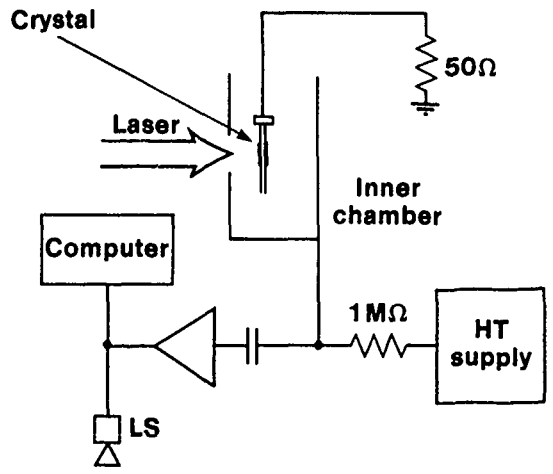


Fig.5 Bias circuit

The voltage drop across a resistor in series between the power supply and the inner chamber has been used to monitor the current. By capacitively coupling this into an audio amplifier it is possible to hear the onset of break-down as the voltage is raised. In this way, the Townsend discharge region before breakdown can be found which has a characteristic white noise signature. This discharge, which can only just be seen, surrounds the crystal and is contained within the inner chamber. At higher voltages, the gas starts to break down, the emission becomes much brighter and the discharge becomes more localised, causing damage to the crystal. Using the Townsend discharge, however, the TMA was satisfactorily desorbed without damage to the crystal.

In view of the complexities involved in the use of TMA, it was decided to investigate alternative precursors. Iron carbonyl was considered but was rejected on health and safety grounds. Chromyl chloride (CrO_2Cl_2) is safer and has been used successfully. The deposition rates dramatically improved because of its higher molecular weight and greater photodissociation efficiency. This precursor is a liquid with a higher vapour pressure than TMA and has been successfully used before with an Excimer laser[2]. The deposit so produced is a mixture of Cr_2O_3 and CrO_2 and has excellent adherence.

The energy of 193 nm ArF is about 6 eV and hence produces photoelectrons from surfaces. By positively biasing the inner chamber, these photoelectrons can be accelerated away from the surface of the crystal. Their subsequent collision with the precursor has been shown to approximately double the deposition rate.

Thermal Effects

The thermal diffusion length for a 15 ns heat pulse in aluminium is 1.5 micron, thus it will have uniformly heated all 1000 A of a typical electrode plating. If it is assumed that this volume is uniformly heated, then for every mJ/cm^2 adsorbed the film will be heated by 40^0C. The peak temperature experienced by the surface of the electrode will be higher. It is clearly possible, therefore, to damage the electrode. In practice the threshold for damage inferred from observed damage of an aluminium film subjected to 100,000 shots at 50 Hz is at a fluence of 3 mJ/cm^2.

Plotting the Electrode Area

In order to accurately place material onto the electrode of a crystal, it is necessary to determine the electrode position. Normally, laser adjustment methods use a beamsplitter, microscope and a visible laser such as a HeNe so that the operator can align the device with the beam manually. As noted earlier, the 6 eV photons of the laser produce photoelectrons from most surfaces. These photoelectrons can be used to devise a method of plotting the electrode area. Furthermore, by applying a positive bias between the surface and the collector, and using a background pressure of about a torr, enhancements of several orders of magnitude can be achieved by

avalanche breakdown of the gas. The optimum pressure for this to occur depends on the Paschen minimum [Figure 6].

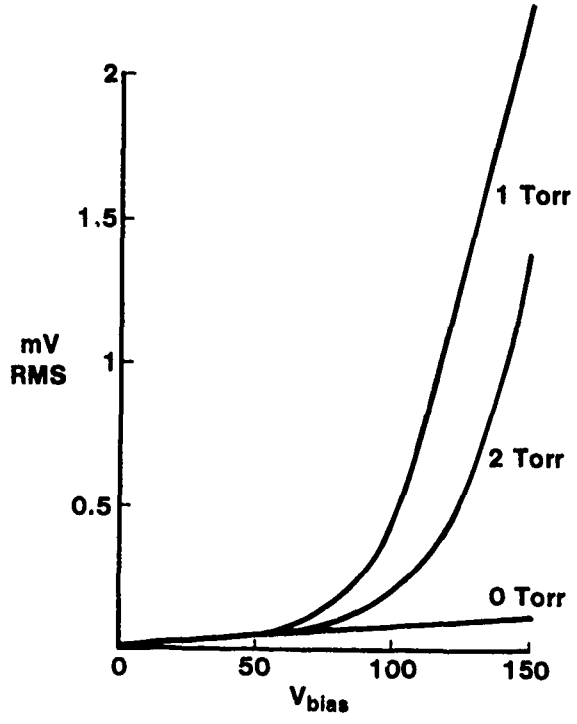


Fig.6a Photoelectrons with bias

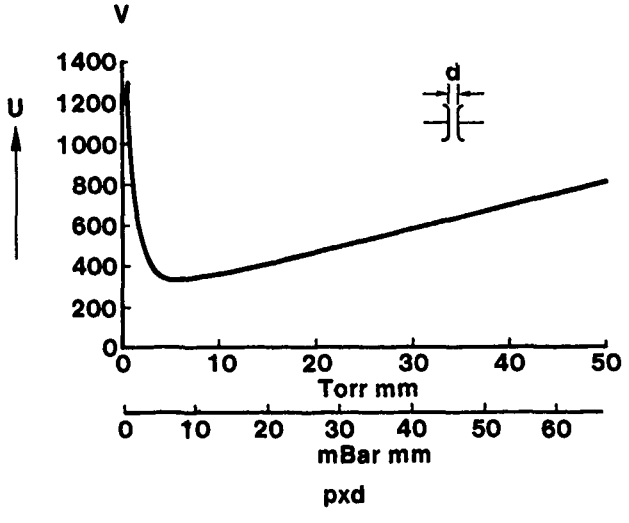


Fig.6b Paschen curve

By positively biasing the inner chamber, electrons produced from the surface of the electrode are accelerated away from the surface and collected by the inner chamber. Rastering the laser across the area of the crystal enables an image of the electrode closest to the laser to be plotted. However, in order to obtain a picture of both electrodes together, the transmitted image rather than the reflective image is needed. The second mode of biasing supplies this. Here the inner chamber is biased negatively and the rear of the chamber is made into a plane parallel to the plane of the crystal. Since quartz is transparent to the UV light, where there is no electrode, it passes straight through and produces photoelectrons from the rear of the chamber. Hence, outside the shadow of the electrodes, electrons accelerate away from the negatively biassed inner chamber to be collected by the electrodes at earth potential.

The data from the raster across the crystal is used to form a contour corresponding to the edge of the shadow of the electrode configuration. The problem is then to find the centre of the electrode overlap from this contour. It is not a closed form (because of the resonator tabs) and is not symmetrical since the electrode is not necessarily centred on the laser raster. However, there should always be two orthogonal lines of reflective symmetry. Humans find it easy to find the centre of such a shape but mathematically this is not a simple task. The solution found to the problem is an interesting example of applying a physical process to mathematics rather than the other way round. The algorithm consists of expanding a circle inside the contour until one or more points of the contour intersect the circle. For each of these points a vector is formed pointing towards the centre of the circle and with a magnitude corresponding to the overlap. The vectors are then summed and the resulting vector used to move the centre of the circle. This procedure is repeated until the circle centre can no longer move. In this way, irrespective of the shape of the contour, the largest circle that can fit into the space is found. The algorithm reflects the physical process of blowing up a two dimensional balloon with the forces on the surface of the balloon moving it into place. Figure 7 shows a plot of the electrode edge with the best fit circle inside it.

Fig.7 Plot of electrode edge with best fit curve

Experimental Results

Deposition tests on silicon samples

Maximisation of deposition rate requires high precursor partial pressures and laser fluence. This also leads to increased window fouling and background deposits from gas phase reactions. In order to investigate these effects, 5 mm square silicon chips were used in a modified crystal holder. These have a surface of silica not unlike quartz. Silicon has the advantage that the deposits can easily be measured with a surface profilometer (a Dektak was used), and that interference fringes show up well under a microscope.

The condition required for minimal window fouling is that the flow of buffer gas away from the window must be sufficient to prevent the back-diffusion of precursor to the window. In practice this means that the precursor flow must be kept down, and hence the deposition rate will be low. A typical Dektak result is shown in Figure 8. Here a fluence of 3 mJ/cm^2 was used with a 0.3 mm diameter slot. The system pressure was 1 torr with 12 sccm of argon on the window and 4 sccm each of TMA and argon across the crystal. The number of laser pulses was 50,000 at 50 Hz repetition rate so that the deposition rate was 0.06 Å per pulse or about 1/75 monolayer, which is about the physisorbed layer for that flow.

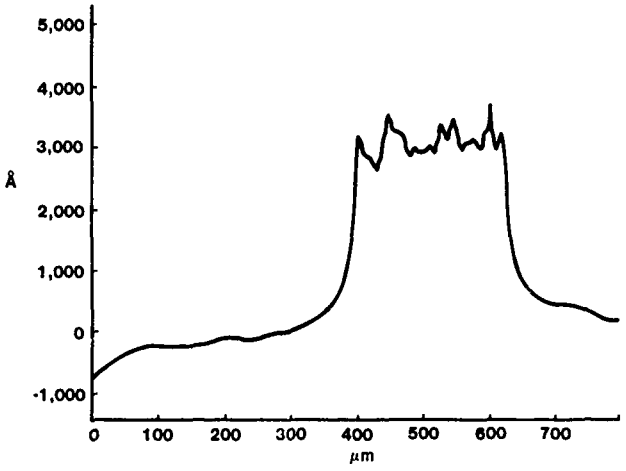


Fig.8 Surface profilometer (DEKTAK) trace of silicon slice

Adjustment of Quartz Devices

The present method is primarily intended for use on the high frequency fundamental devices currently manufactured by STC Quartz Crystal Division. Adjustment trials were performed on a number of crystals from 50 MHz to 125 MHz fundamentals.

The adjustment conditions that were finalised upon were a flow of chromyl chloride of 4 sccm (calibrated for nitrogen) across the crystal, with a flow of argon across the window to bring the system up to 400 mTorr. The laser was used at a repetition rate of 40 Hz with a fluence at the crystal surface of about 3 mJ/cm^2, with a bias voltage of +150 volts between the inner chamber and the crystal electrode.

There are a number of ways of presenting the deposition rates. Typically for 125 MHz fundamental crystals the average frequency shift per laser pulse was 300 Hz or 12.4 kHz/sec or 100 ppm/sec at 40 Hz. In terms of f^2 units this corresponds to 2×10^{-5} f^2 per pulse. Alternatively a typical 1000 ppm adjustment could be achieved in 10 seconds at 125 MHz.

The areal density is given by df/f^2 4.4×10^5 gm/cm^2. This gives per pulse a deposition rate of 8.4×10^{-9} gm/cm^2.

Assuming the deposit to be primarily chromium dioxide with a molecular weight of 84 this gives a deposition rate of 6×10^{13} molecules/cm^2. The laser fluence used was 3 mJ/cm^2 which corresponds to 3×10^{15} photon/cm^2 so that the ratio of photons to deposited molecules is 50:1.

Figure 9 shows some preliminary ageing results of a canned 125 MHz device adjusted by 1000 ppm together with another crystal with no adjustment. The crystals were given a 120°C bake before canning. This demonstrates that even at 125 MHz, acceptable ageing results can be achieved with this method so that the films must be stable.

Conclusion

A method of adjusting quartz crystals to frequency has been developed using the deposition of stable films onto the surface of the electrode. Those films are deposited by the dissociation of volatile precursors with a UV laser. The deposit occurs only at the laser spot so that the adjustment process is maskless.

The necessary process parameters have been established. These include the laser fluence, pre- and post-cleaning techniques and the flow rates necessary to establish reasonable deposition rates with minimum window fouling. The preliminary ageing results of adjusted units looks promising.

Acknowledgements

The author wishes to thank R.A. Heinecke and D.J. Carter both of STC Technology for invaluable contributions. This work has been carried out with the support of the Procurement Executive, Ministry of Defence (Royal Signals and Radar Establishment) and sponsored by the Admiralty Research Establishment (Portsdown), and also by STC Quartz Crystal Division.

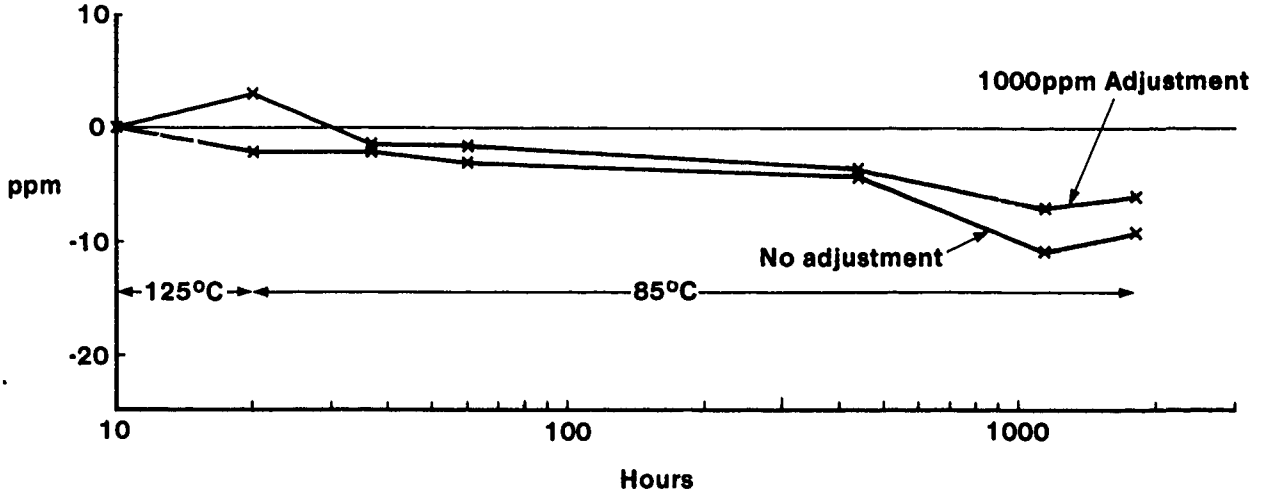


Fig.9 Ageing of 125 MHz fundamental crystal

References

[1]Photodeposition of metal films with ultraviolet light. D. Ehrlich, R. Osgood, T. Deutsch. J.Vac.Sci.Tech. 21(1) May 1982.

[2]Visible-laser photodeposition of chromium oxide films and single crystals. C. Arnone, M. Rothschild, J. Black, D. Ehrlich. Appl.Phys.Lett. 48 (15) 14 April 1986.

43rd Annual Symposium on Frequency Control - 1989

THE FABRICATION OF HIGH FREQUENCY FUNDAMENTAL CRYSTALS BY PLASMA ETCHING

F.M. Stern and J. Dowsett
STC Components, Quartz Crystal Division,
Harlow, Essex, CM20 2DE, England

D.J. Carter and R.J. Williamson
STC Technology Ltd, Harlow, Essex, CM17 9NA, England

Abstract

This paper describes a method of manufacturing AT-cut high frequency fundamental crystals up to 150 MHz. Samples at fundamental frequencies up to 500 MHz have been fabricated, demonstrating that batch production of these very high fundamental frequency crystals is also feasible. Plasma etching is used for recessing blanks to the desired thickness. High quality quartz is essential as the starting material and careful preparation of the blanks prior to plasma etching is required. Examples of the electrical parameter measurements on production batches of crystals at fundamental frequencies between 60 MHz and 150 MHz are given, together with results from sample crystals at frequencies up to 500 MHz

Background

High frequency fundamental (HFF) crystals, 30-150 MHz (and above) are increasingly in demand for a wide range of commercial and military applications. Conventional crystal processing, by lapping and polishing to thickness, has a mass production thickness limit of about 50 µm, corresponding to a fundamental frequency of 33 MHz. In order to make robust high frequency resonator devices operating in the fundamental mode, it has been necessary to develop processing techniques for the selective thinning of quartz.

Berte [1] has reported using ion beam milling to achieve a frequency of 270 MHz. Wang et al. [2] have used reactive ion beam milling to selectively etch quartz and produce a resonator with a fundamental frequency of 200 MHz. Hunt et al. [3] have combined traditional chemical etching in saturated ammonium bifluoride solution with use of resist masking to produce fundamental mode resonators at frequencies as high as 1655 MHz.

Introduction

STC has been manufacturing AT-cut high frequency fundamental crystals of up to 75 MHz since 1984, initially using saturated ammonium bifluoride solution to etch the blanks to the desired frequency [4]. Crystals of up to 150 MHz were produced, but as a production process it proved impracticable above 75 MHz. Without the necessary masking to maintain a thicker supporting structure, the blanks become increasingly fragile and difficult to handle as etching progresses.

This phase of fabricating high frequency fundamental crystals by overall wet etching was, however, only an interim activity. Various etching and masking techniques had previously been evaluated with the objective of identifying a more practical process for producing high frequency fundamental crystals. These included wet etching, ion beam milling and plasma etching. Plasma etching was favoured as a dry process, offering the possibility of in situ measurement, better yields than those with wet etching and good control of the blank geometry. Etch rates were lower than with wet etching methods, but this was outweighed by the potential batch capacity possible with plasma etching. Ion beam milling gave good etch rates but had limited batch size potential as it was necessary to rotate the blank continuously during etching to ensure good uniformity. Wet etching is an operator-intensive process and a loss of blank parallelism is observed as etching progressed. Plasma etching techniques were therefore developed further and samples at frequencies up to 150 MHz were produced. Production equipment was designed and constructed, and finally commissioned in 1985.

Blank Geometry

The high frequency fundamental blank produced in the plasma etching process is a single-side recessed structure as shown in Fig. 1. The recess profile is controlled by masking developed specifically for the process. The outer support rim, 1 mm wide and 65 µm thick, reduces gradually in thickness to the central membrane where, depending on the exact masking used, the active area may be parallel or contoured as required.

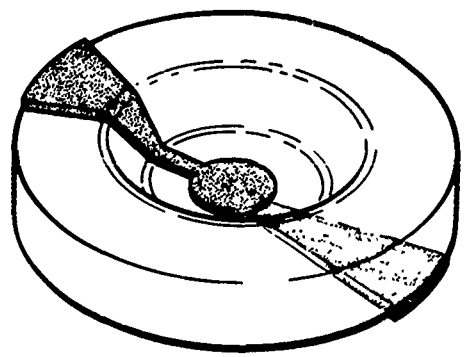


FIG. 1 Schematic of blank structure

CH2690-6/89/0000-634 $1.00 C 1989 IEEE

Quartz Quality

For best Q, C_1 and R_1, the surface finish of the blank must be perfect, therefore swept quartz is essential as the starting material. The preparation prior to plasma etch involves etching in saturated ammonium bifluoride solution. If standard synthetic quartz is used, tunnelling occurs at dislocation sites [5] with the undesirable and characteristic pit formation at emergence points of a tunnel on each side of a blank (as shown in Fig.2).

FIG. 2 Tunnelling in an unswept quartz blank after etching in ammonium bifluoride.

A programme to improve the quality of the quartz grown by STC was initiated and material with Q of higher than 2.6×10^6 was grown. Although a reduction in the tunnel density from >250/sq.cm. with standard synthetic quartz ($Q = 1.8 \times 10^6$) to <100/sq.cm was observed, further reduction was still required. Quartz grown on natural quartz seed was found to give a very low tunnel count, less than 25/sq.cm, but a shortage of material of this type made this impractical. Commercially available swept quartz was found to give excellent results with <10/sq.cm.

Due to the cost of this swept material a sweeping facility was set up. Pairs of quartz blocks ($Q = 2.5 \times 10^6$) are sandwiched between stainless steel plates with graphite electrode foils interspersed. Sweeping takes place in a convection oven at 500° C over a total period of 155 hours, with a field strength of 1000 V/cm applied throughout including heat-up and cool-down. Etch tunnel densities of 100-200/sq.cm. are reduced to 0-10/sq.cm.

Blank Preparation

For most HFF applications, 5 mm diameter blanks are processed; there is an occasional requirement for 7.5 mm. The swept quartz bars are cut, trepanned and the blanks lapped to 14 MHz by conventional processing methods. The blanks are then edged down to 5 mm and lapped up to 20 MHz using a calcined alumina abrasive with a nominal particle size of 5 μm. At this stage it is important that the blanks are flat and parallel. Therefore, the flatness of the plates used for lapping blanks for the HFF process is rigorously monitored.

The lapped blanks are etched in saturated ammonium bifluoride to remove surface stress and then polished up to 25 MHz, a critical stage of the processing. The polished blanks must be free of all visible surface damage at x10 magnification and must be parallel to better than 1/2 a light fringe across their diameter. For most production work, parallelism is inspected using a sodium lamp. This is generally adequate for inspecting blanks up to 20 MHz. Above 20 MHz the birefringent property of quartz becomes increasingly dominant as the frequency increases and at certain thicknesses the two fringe patterns arising from this birefringence cancel out, creating an erroneous impression of flatness. To overcome this problem, equipment for inspecting parallelism was devised at STL. The fringe viewer, as shown in Fig. 3, utilizes a He/Ne laser (wavelength 630 nm) to produce fringes which may be viewed on a TV monitor, and incorporates a rotatable polarizer so that both fringe patterns may be observed.

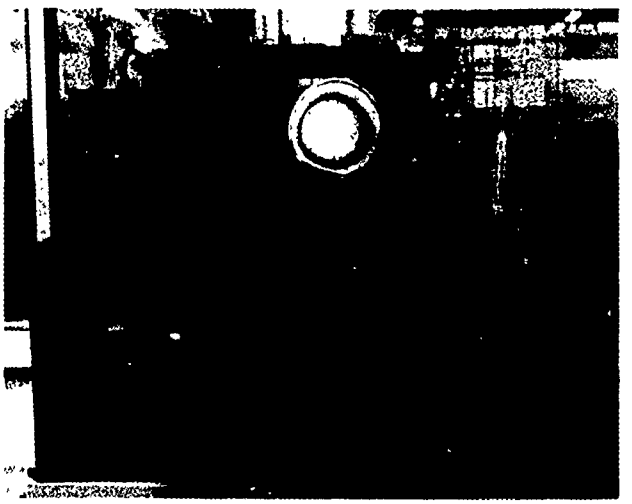


FIG. 3 Parallelism inspection equipment

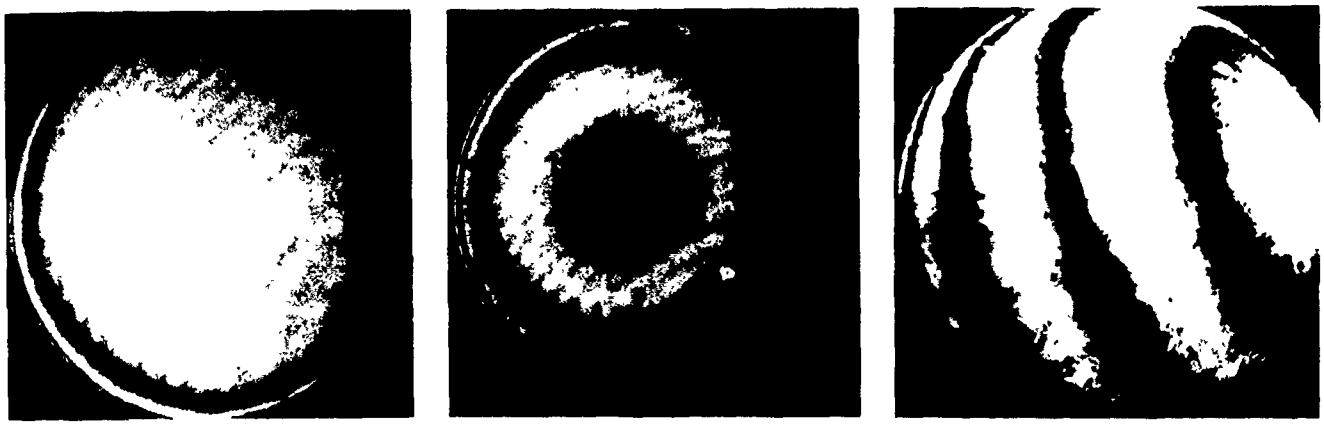
A further refinement on this system has been to replace the He/Ne laser with a sodium light and replace the sample support table with a tilting version, so now the degree of convexity/concavity of the sample may be determined by tilting the table.

Fig. 4 shows some typical fringe patterns. (a) demonstrates excellent parallelism and concentricity, (b) is convex and eccentric and (c) is a wedge.

From this stage onwards all processing takes place in a class 10,000 clean room. The polished blanks are cleaned in detergent, then in an ammoniacal peroxide solution and then rinsed in deionised water. The jigs containing the blanks are then immersed in hot deionised water and spin dried.

Plasma Process and Equipment

The next stage is plasma etching. Extreme care is essential to avoid any trace of particulate matter on the surface of the blanks as the plasma will etch around them leaving a protrusion (as shown in Fig. 5). Therefore facemasks and gloves are worn by the operators in addition to the usual clean room overalls. The clean, dry blanks are transferred, using vacuum tweezers, from the cleaning jigs directly into trays, and the masks positioned so that only the centre (3.5 mm diameter) of each blank will be exposed to the plasma. The trays are



(a) Parallel, concentric

(b) Convex, slightly eccentric

(c) Wedge

FIG. 4 Examples of polished blank fringe patterns

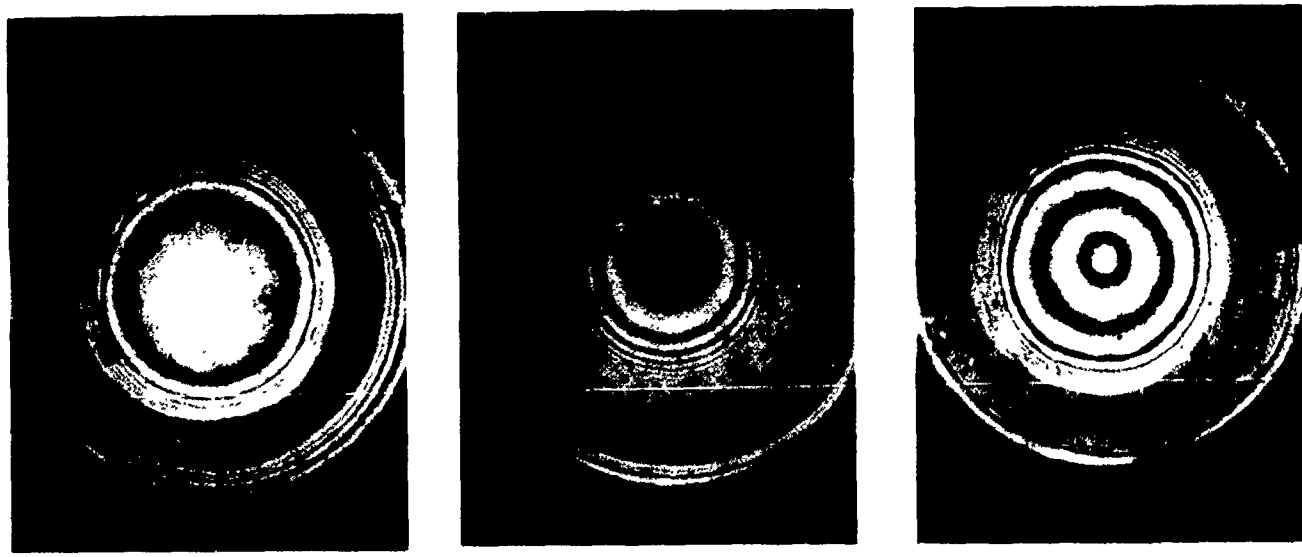
removed from the plasma etcher at intervals during the processing to inspect the blanks for concentricity and parallelism of the recesses, and alternative masking deployed as necessary. The etch rate may be varied by altering the gas flow and power level. For optimum control of the recess profile in conjunction with the masking, the equipment is set up to give an etch rate of 0.1 μm/min.

Fig. 6 shows the plasma etching equipment. This is a two chamber reactive ion etching plant powered by a 3 kW 13.56 MHz generator. Each chamber can accommodate five trays each of which holds up to 100 blanks. The potential batch capacity of this plant is 1000 blanks, although this has not yet been utilized.

The equipment is run on a semi-automatic basis; running conditions and etch time are set up manually. On completion of the etch time the plasma automatically switches off.

Frequency Trimming

As with wet etching, a frequency spread is observed as the frequency increases. A frequency spread of 400 kHz is allowed at the polishing stage, corresponding to a thickness spread of 1.2 μm. At 100 MHz this corresponds to a frequency spread of 8 MHz. In fact, at 100 MHz a frequency spread of 10 MHz or more may result. A trimming etch into frequency is therefore necessary. After plasma etching the blanks are 100% checked for frequency, sorted and etched into the desired preplate frequency band. Tightly specified at 0.02 f^2, this corresponds to a thickness spread of only 0.03 μm. Until recently this final etch into frequency was carried out in saturated ammonium bifluoride solution; an operator intensive operation. This had poor yields, 95% at 50 MHz, decreasing to 60% at 110 MHz, due to blank breakages and over-etching.



(a) 50 MHz

(b) 100 MHz

(c) 125 MHz

FIG. 7 Typical fringe patterns for recessed blanks

FIG. 5 Vertical pillar caused by a permanent etch stop (x2000)

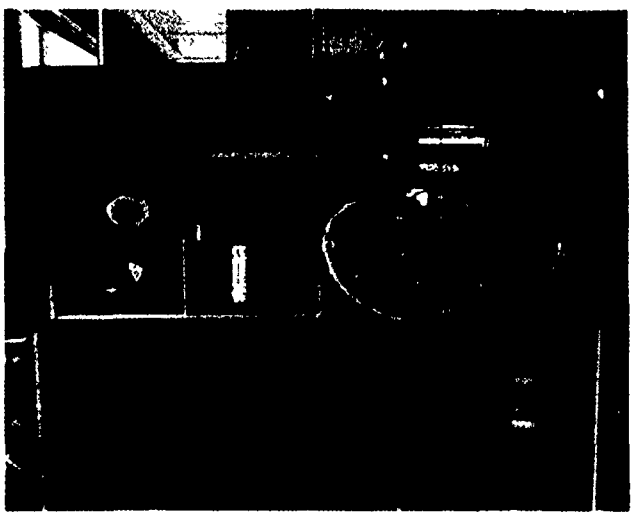


FIG. 6 Plasma etching equipment

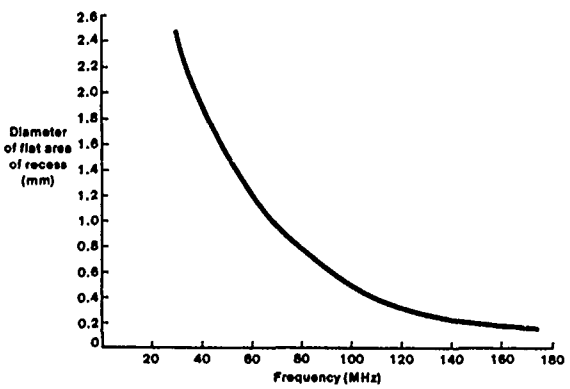


FIG. 8 Frequency vs. diameter of flat area of recess

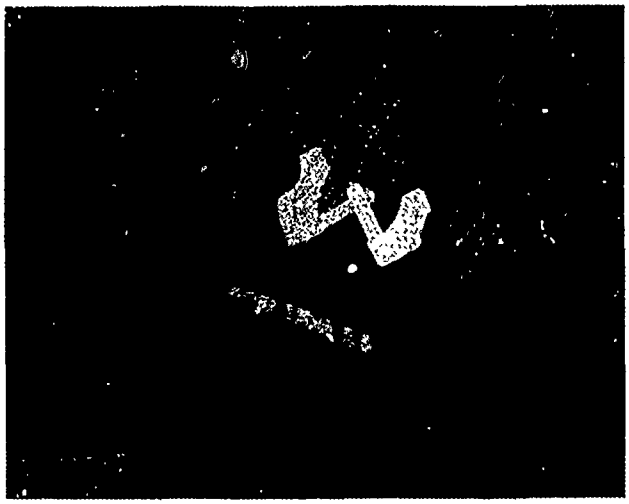


FIG. 9 100 MHz crystal before sealing

A new plasma etcher that will automatically trim the blank frequency into the desired band has been devised. This equipment automatically measures each blank frequency and individually plasma etches it into the preset frequency specification, reducing the overall time for the final trimming operation.

Resonator Design and Final Processing

As a consequence of controlling the blank profile to provide a smooth slope from rim to recess, the central active area of the recess decreases with increasing frequency. Fig. 8 shows frequency versus the maximum diameter over which it is practical to deposit electrodes.

Round electrodes are used for most crystal designs up to 70 MHz, while above 70 MHz a crossover configuration has been adopted as shown in Fig. 9. Silver metallization with a 30 angstrom nichrome underlayer is usual for oscillator crystals up to 150 MHz. To improve the frequency response and Q and reduce spurious modes for filter crystals and 3rd overtones, aluminium is increasingly used above 70 MHz.

The crystals are mounted, pasted and cured, adjusted for frequency, and sealed.

Results

Crystal parameter measurements are made on a HP4195A network analyzer using the admittance circle method of measurement as described by Williamson [6] and Morley et al. [7]. In Table 1 are shown examples of the actual measurements on production batches of 63 MHz filter crystals manufactured over a period of 18 months demonstrating the consistency of the process. The motional capacitance is specified at 2.5fF +/-10%. Frequency tolerance at room temperature is +/-10ppm.

Table 2 shows some typical results from production batches of 101 MHz, 139.264 MHz, and 150.045 MHz HFF crystals.

Measurement results for third overtone crystals at 150 MHz, 300 MHz and 450 MHz are given in Table 3.

Table 4 shows measurements on some very high frequency fundamental crystals which have been fabricated as part of the development programme to increase the production frequency limit from 150 MHz to 500 Mhz.

Fig. 10 compares some best Q results for plasma etched units with the theoretical values and with those reported for wet etched units by Hunt et al. (3).

Conclusion

Selective thinning of quartz blanks by plasma etching has been found suitable for producing large numbers of high frequency fundamental crystals up to 150 MHz. Swept quartz is essential as the starting material as well as careful preparation of the blanks prior to the plasma etching.

Records of measurements made on 63 MHz filter crystals over the past 18 months show that the process is consistent.

Samples at fundamental frequencies above 150 MHz have been fabricated demonstrating that with further development of the resonator design and metallization, batch production at these higher frequencies is entirely feasible.

Acknowledgements

The authors wish to thank all those who have contributed to this work and, in particular, R.A. Heinecke of STL.

This work has been partly supported by the Procurement Executive, Ministry of Defence (Royal Signals and Research Establishment) and sponsored by the Admiralty Research Establishment (Portsdown).

Table 1. Parameter measurements on 63 MHz fundamental mode filter crystals manufactured over a period of 18 months.

(a) May 1987

FREQ.(KHz)	RES.(Ohms)	DRIVE(uW)	C1(fF)	Q
62943.026	52.64	88.01	2.712	17709
62944.552	24.38	100.78	2.614	39658
62943.911	20.28	99.70	2.480	50251
62944.681	25.28	100.79	2.441	40961
62944.161	21.21	100.11	2.470	48243
62943.981	20.32	99.72	2.563	48534
62944.024	22.37	100.48	2.517	44895
62944.325	22.68	100.56	2.461	45269
62944.059	19.67	99.36	2.672	48079
62943.793	18.13	98.23	2.951	47244
62944.657	20.58	99.84	2.605	47156
62944.323	27.19	100.61	2.079	44716
62945.045	74.24	75.98	2.461	13838
62944.119	20.44	99.78	2.499	49468
62943.576	26.49	100.71	2.203	43317
62943.776	21.17	100.10	2.451	48699
62943.733	20.94	100.00	2.366	51025
62944.672	21.21	100.11	2.416	49337
62943.667	20.83	99.96	2.484	48839
62944.227	21.38	100.18	2.423	48794

(b) December 1987

FREQ.(MHz)	RES.(Ohms)	DRIVE(uW)	C1(fF)	Q
62943.507	24.58	20.99	2.548	40362
62943.697	24.61	20.99	2.507	40978
62943.435	20.92	20.83	2.497	48385
62943.268	21.20	20.85	2.542	46905
62943.951	21.46	20.87	2.473	47642
62943.784	20.95	20.83	2.716	44408
62943.578	21.03	20.84	2.531	47488
62943.455	22.10	20.92	2.691	42506
62943.505	21.20	20.85	2.602	45820
62943.272	23.20	20.97	2.685	40570
62943.442	20.86	20.83	2.637	45953
62943.463	20.70	20.81	2.665	45802
62943.063	23.45	20.97	2.500	43103
62943.697	21.76	20.90	2.576	45084
62943.733	21.47	20.88	2.741	42949
62943.320	20.03	20.74	2.537	49728
62943.606	20.05	20.74	2.649	47593

(c) December 1988

Crystal code	fs (ppm)	drive (uW)	RI (ohms)	CI (%)	Q
Nom :	62943.880 kHz			Nom :	2.50 fF
1	-2.8	99.10	24.07	-3.03	43339
2	.5	99.03	23.35	-.47	43518
3	-1.9	98.85	22.43	-2.02	46026
4	-3.0	99.00	23.19	-1.73	44373
5	1.1	98.94	22.85	-4.35	46274
6	-.1	99.09	23.93	-1.90	43092
7	1.6	98.62	21.63	4.57	44709
8	5.7	98.37	20.96	7.02	45097
9	-.5	99.11	24.15	-4.34	43774
10	1.0	98.96	22.96	-3.75	45764
11	-.6	98.49	21.24	.87	47205
12	-1.8	98.34	20.87	5.86	45781
13	2.1	98.94	22.83	-.43	44495
14	-1.7	98.65	21.71	9.42	42579
15	0.0	99.13	25.54	-5.52	41920
16	-3.6	99.14	25.29	.90	39635
17	3.8	98.82	22.32	-3.41	46922
18	-1.7	98.85	22.42	-.50	45337
19	-2.0	99.09	23 94	-5.88	44887
20	-3.4	99.14	24.81	-5.55	43161

Table 2. Parameter measurements taken from production records for (a) 101 MHz , (b) 139.264 MHz and (c) 150.045 MHz fundamental mode oscillator crystals.

Crystal code	fs (ppm)	drive (uW)	R1 (ohms)	C1 (fF)	Q	C0 (pF)
Nom : 101000.000 kHz						
2	11.0	95.99	21.16	2.94	25351	1.81
3	14.0	96.09	21.45	2.83	25982	1.80
4	8.3	96.45	22.77	2.84	24399	1.76
5	13.4	96.64	25.74	3.06	20022	1.83
6	7.1	96.61	23.91	2.77	23760	1.82
7	11.6	95.51	20.09	2.96	26504	1.82
8	6.6	96.64	24.25	2.78	23387	1.80
12	12.3	95.37	19.82	2.84	27967	1.82
13	9.5	96.63	24.23	2.87	22649	1.81
14	5.3	96.66	25.00	2.71	23251	1.77
Nom : 139264.000 kHz						
1	5.0	102.68	36.24	2.21	14280	1.09
3	-7.9	105.72	28.85	2.55	15517	1.16
4	5.1	102.36	36.86	2.02	15353	1.05
5	4.7	105.69	28.95	2.50	15812	1.16
6	-7.7	98.13	44.11	2.08	12430	1.06
7	-.9	102.49	36.60	2.32	13476	1.11
9	-4.1	104.31	32.85	2.54	13709	1.14
11	5.8	105.53	27.52	2.33	16588	1.13
Nom : 150045.000 kHz						
1	-7.0	93.81	18.46	2.99	19204	1.74
2	-1.5	95.95	25.89	2.68	15269	1.68
3	1.7	95.25	20.99	2.73	18539	1.67
4	1.5	95.19	20.85	2.87	17742	1.74
5	8.2	93.93	18.62	3.03	18800	1.75
6	-4.9	95.91	23.70	2.56	17489	1.66

Table 3. Parameter measurements on third overtone crystals, (a) 150 MHz, (b) 300 MHz, (c) 450 MHz.

Crystal code	fs (ppm)	drive (uW)	R1 (ohms)	C1 (fF)	Q	C0 (pF)
Nom : 150000.000 kHz						
1	-6.1	93.83	49.19	.42	50776	1.88
2	-5.3	94.45	48.20	.41	54226	1.88
3	-5.8	94.13	48.70	.43	50566	1.92
4	-8.2	95.95	45.76	.40	58562	1.94
5	-7.9	92.72	50.97	.41	50229	1.93
6	-9.0	92.19	51.82	.42	48542	1.87
7	-7.2	94.33	48.38	.44	49890	1.89
8	-7.5	95.81	46.00	.43	53185	1.92
9	-6.7	95.65	46.26	.43	53695	1.94
10	-4.0	95.22	46.95	.40	56989	1.87
17	300034.096	61.87	93.63	.26	21396	1.72
19	299833.852	46.90	143.96	.19	19918	1.70
20	300166.430	63.41	89.69	.31	19064	1.74
21	300170.460	66.49	82.28	.33	19593	1.73
22	300013.998	46.50	145.69	.30	11984	1.77
24	300090.650	61.24	95.29	.26	21730	1.69
26	299940.868	23.90	337.21	.36	4362	1.70
27	300023.409	51.15	126.88	.23	18069	1.73
28	300004.906	65.93	83.58	.24	26443	1.74
30	299959.401	62.65	91.63	.21	27584	1.73
1	450497.129	36.58	115.39	.31	9825	1.59
2	450376.456	36.51	115.74	.21	14208	1.53
3	450424.433	40.90	96.28	.26	14031	1.52
4	450516.041	44.97	81.23	.27	15947	1.60
5	450565.536	47.48	73.03	.31	15649	1.60
6	450295.136	35.16	122.59	.26	11277	1.51
7	450460.024	36.73	114.68	.33	9297	1.57
8	450361.867	38.73	105.41	.23	14714	1.52

Table 4. Selected results from very high frequency fundamental crystals from 150 MHz to 500 MHz.

fs (kHz)	drive (uW)	R1 (ohms)	C1 (fF)	Q	C0 (pF)
91628.820	50.79	25.84	1.70	39498	1.55
164110.525	50.54	24.24	1.55	25761	1.45
225406.647	50.80	47.77	.91	16230	1.27
353391.234	50.55	48.55	1.37	6766	1.29
449551.303	45.41	57.33	1.01	6119	1.85
498977.468	29.78	56.46	.97	5853	1.08

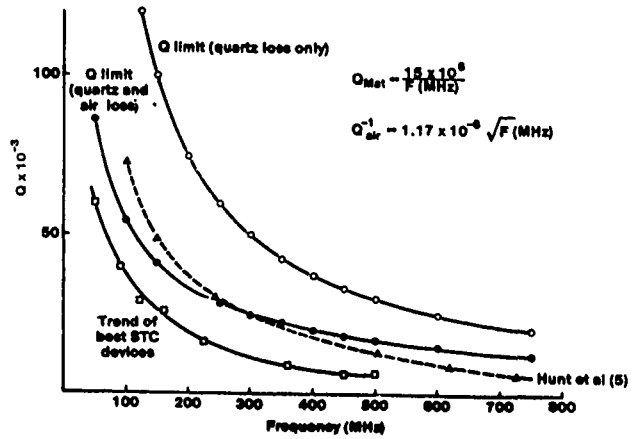


Fig. 10. Variation of Q with frequency for fundamental AT devices.

References

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

[2] J.S. Wang, S.K. Watson and K.F. Lau, "Reactive ion beam etching for VHF crystal resonators", Proc. 34th Ann. Freq. Control Symp., 1980, pp. 101-104.

[3] J.R. Hunt and R.C. Smythe, "Chemically milled VHF and UHF AT-cut resonators", Proc. 39th Ann. Freq. Control Symp., 1985, pp. 292-300.

[4] J. Dowsett, D.F.G. Dwyer, F. Stern, R.A. Heinecke and A.H. Truelove, "Etch processing of bulk and surface wave devices", Proc. 39th Ann. Freq. Control Symp., 1985, pp. 301-309.

[5] J.R. Hunt, "Etch pits and channels in swept AT- and SC-cut quartz", Proc. 41st Ann. Freq. Control Symp., 1987, pp. 183- 191.

[6] R.J. Williamson, "An improved method for measuring quartz crystal parameters", IEEE Trans. UFFC-34, no. 6, 1987, pp 681-689

[7] P.E.Morley, R.J.T. Marshall and R.J. Williamson, "Crystal parameter measurement using a Hewlett Packard HP4195A network/spectrum analyzer", presented at the Third European Frequency and Time Forum, Besancon, March 1989.

AUTHOR INDEX

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201)981-0060

176-1987 (ANSI/IEEE) Standard on Piezoelectricity (SH11270)
177-1966 Standard Definitions and Methods of Measurements for Piezoelectric Vibrators
180-1986 (ANSI/IEEE) Definitions of Primary Ferroelectric Crystal Terms (SH10553)
319-1971 (Reaff 1978) Piezomagnetic Nomenclature (SH02360)
1139-1988 Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology (SH12526)

DEPARTMENT OF DEFENSE

Order from: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120
(215)697-2000

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for
MIL-C-24523 (SHIPS), Chronometer, Quartz Crystal
MIL-F-15733 Filters and Capacitors, Radio Interference, General Specification for
MIL-F-18327 Filters, High Pass, Low Pass, Band Pass Suppression and Dual Functioning, Gen Spec for
MIL-F-28861 Filters and Capacitors, Radio Frequency/ Electromagnetic Interference Suppression, Gen Spec for
MIL-F-28811 Frequency Standard, Cesium Beam Tube
MIL-H-10056 Holders (Encl), Crystal, Gen Spec for
MIL-O-55310 Oscillators, Crystal, Gen Spec for
MIL-O-39021 Oven, Crystal, Gen Spec for
MIL-S-49433 Surface Acoustic Wave Devices, Bandpass Filter
MIL-S-49433(ER) Surface Acoustic Wave Devices, Gen Spec for
MIL-STD-683 Crystal Units, Quartz/Holders, Crystal
MIL-STD-188-115 Interoperability and Performance Standards for Communications, Timing and Synchronization Subsystems
MIL-STD-1395 Filters and Networks, Selection and Use of
MIL-W-46374D Watch wrist: General Purpose
MIL-W-87967 Watch wrist: Digital

GENERAL SERVICES ADMINISTRATION

Order from: Naval Publication & Form Center or
General Services Administration Business Service Centers in major U.S. cities

FED-STD-1002 Time and Frequency Reference Information in Telecommunication Systems

ELECTRONIC INDUSTRIES ASSOCIATION

Order from: 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

EIA 512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

EIA/IS-17-A, Assessment of Outgoing Nonconforming Levels in Parts Per Million (PPM).

EIA/IS-18, Lot Acceptance Procedure for Verifying Compliance with the Specified Quality Level (SQL) in PPM

(c) Application Information

EIA Components Bulletin No. CB6-A, Guide for the Use of Quartz Crystal Units for Frequency Control, Oct 1987

(d) EIA-477, Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985)

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC)

Order from: American National Standards Inst. (ANSI)
1430 Broadway
New York, NY 10018
(212)354-3300

IEC PUBLICATIONS

122: - Quartz crystal units for frequency control and selection.
122-1 (1976) Part 1: Standard values and test conditions Amendment No. 1 (1983).
122-2 (1983) Part 2: Guide to the use of quartz crystal units for frequency control and selection.
122-3 (1977) Part 3: Standard outlines and pin connections.
Amendment No. 1 (1984).
122-3A(1979) First supplement.
122-3B(1980) Second supplement.
122-3C(1981) Third supplement.
283 (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 (1 '1) First supplement.
368: - Piezoelectric filters.
368-1 (1982) Part 1: General information, standard values and test conditions.
368A (1973) First supplement.
Amendment No. 1 (1977).
Amendment No. 2 (1982).
368B (1975) Second supplement.
368-2: - Part 2: Guide to the use of piezoelectric filters
368-2-1(1988) Section One - Quartz crystal filters.
368-3 (1979) Part 3: Standard outlines.
368-3A (1981) First supplement.
444: - Measurement of quartz crystal unit parameters by zero phase technique in a π-network.
444-1 (1986) Part 1: Basic method for the measurement

of resonance frequency and resonance resistance of quartz crystal units by zero phase technique in a π-network.

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

444-3 (1986) Part 3: Basic method for the measurement of two-terminal parameters of quartz crystal units up to 200 MHz by phase technique in a π-network with compensation of the parallel capacitance C_0.

444-4 (1988) Part 4: Method for the measurement of the load resonance frequency f_L, load resonance resistance R_L and the calculation of other derived values of quartz crystal units, up to 30 MHz.

483 (1976) Guide to dynamic measurements of piezoelectric ceramics with high electromechanical coupling.

642 (1979) Piezoelectric ceramic resonators and resonator units for frequency control and selection. Chapter I: Standard values and conditions. Chapter II: Measuring and test conditions.

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.
Amendment No. 1 (1985).

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watches and standard values.

758 (1983) Synthetic quartz crystal.
Chapter I: Specification for synthetic quartz crystal.
Chapter II: Guide to the use of synthetic quartz crystal.
Amendment No. 1 (1984).

862: - Surface acoustic wave (SAW) filters.

862-1-1(1985) Part 1: General information, test conditions and methods.

862-3 (1986) Part 3: Standard outlines.

PROCEEDINGS OF THE ANNUAL SYMPOSIA ON FREQUENCY CONTROL

A complete index to the Proceedings for 1956 (10th) through 1988 (42nd) is available as part of the Proceedings of the 42nd Annual Symposium on Frequency Control.

NO.	YEAR	DOCUMENT NO.	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	AD746210	"	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	AD046089	"	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	64.00
38	1984	84CH2062-8	"	61.00
39	1985	85CH2186-5	"	70.00
40	1986	86CH2330-9	"	70.00
41	1987	87CH2427-3	"	64.00
42	1988	88CH2588-2	"	70.00
43	1989	89CH2690-6	"	70.00

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161

* EIA - Electronic Industries Association
c/o Frequency Control Symposium
2001 Eye Street, Washington, DC 20006

*IEEE - Institute of Electrical & Electronics Engineers
445 Hoes Lane, Piscataway, NJ 08854