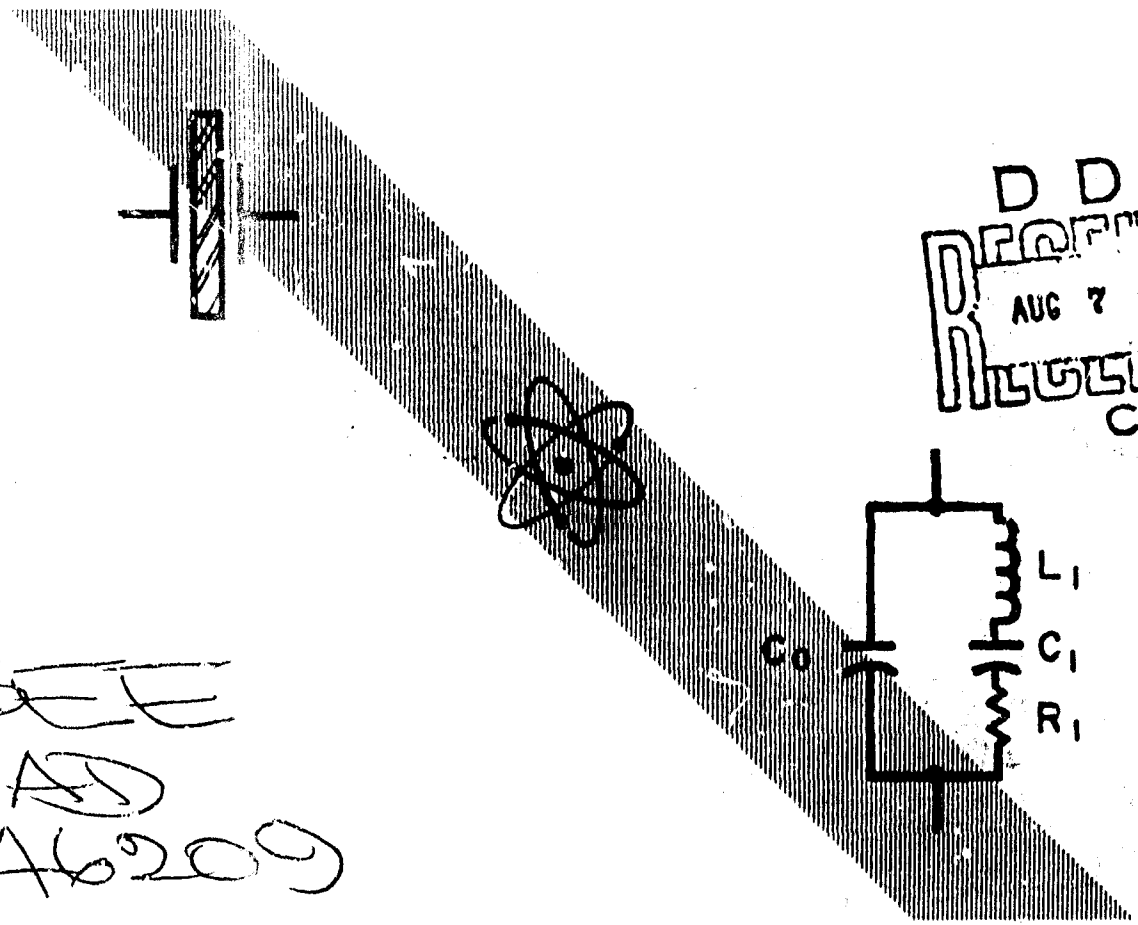
AD 746210

PROCEEDINGS
OF THE
24th ANNUAL SYMPOSIUM ON FREQUENCY CONTROL



DDC
RECEIVED
AUG 7 1972
C

SEE AD 746209

27-28-29 APRIL 1970

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. Department of Commerce
Springfield VA 22151

U.S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

DISTRIBUTION STATEMENT

Approved for public release;
distribution unlimited.

Best
Available
Copy

PROCEEDINGS

OF THE

TWENTY-FOURTH ANNUAL FREQUENCY CONTOL SYMPOSIUM

Sponsored by

SOLID STATE AND FREQUENCY CONTROL
DIVISION
ELECTRONIC COMPONENTS LABORATORY
U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, NEW JERSEY

Details of illustrations in
this document may be better
studied on microfiche.

[redacted]

ELECTRONIC INDUSTRIES ASSOCIATION
2001 Eye Street, N.W.
Washington, D. C. 20006

[redacted]

[redacted]

27 - 29 April 1970

The Shelburne Hotel
Atlantic City, New Jersey

TWENTY-FOURTH ANNUAL FREQUENCY CONTROL SYMPOSIUM

Sponsored by

U. S. ARMY ELECTRONICS COMMAND
Solid State and Frequency Control Division
Electronic Components Laboratory
Fort Monmouth, New Jersey

27-29 April 1970

The Shelburne Hotel
Atlantic City, New Jersey

EXECUTIVE COMMITTEE

General Chairman	Mr. V. J. Kublin
Technical Program Chairman	Dr. G. K. Guttwein
Executive Assistant	Mr. J. M. Stanley
Staff Consultant	Mr. M. F. Timm

TECHNICAL PROGRAM COMMITTEE

Dr. G. K. Guttwein
U. S. Army Electronics Command

Mr. M. Bernstein
U. S. Army Electronics Command

Mr. A. R. Chi
NASA, Goodard Space Flight Center

Mr. J. Dressner
U. S. Army Electronics Command

Mr. J. Giannotto
U. S. Army Electronics Command

Dr. E. Hafner
U. S. Army Electronics Command

Mr. D. L. Hammond
Hewlett-Packard Company

Dr. H. Hellwig
National Bureau of Standards

Mr. A. Machlin
U. S. Army Materiel Command

Mr. J. Sethares
U. S. Air Force

Dr. W. J. Spencer
Bell Telephone Laboratories

Mr. R. R. Stone
U. S. Naval Research Laboratories

Dr. G. M. R. Winkler
U. S. Naval Observatory

HONORARY CRYSTAL INDUSTRY COMMITTEE

Mr. J. J. Colbert
Electronic Crystals Corporation

Mr. D. L. Hammond
Hewlett-Packard Company

Mr. J. D. Holmbeck
Northern Engineering Laboratories

Mr. J. H. Sherman, Jr.
General Electric Company

Mr. R. A. Sykes (Ret.)
Bell Telephone Laboratories

CHAIRMEN FOR TECHNICAL SESSIONS

FUNDAMENTAL STUDIES AND QUARTZ CRYSTAL RESEARCH

Dr. W. J. Spencer, Bell Telephone Laboratories, Inc.

FILTERING TECHNIQUES AND CRYSTAL FILTERS

Mr. M. Dishal, IT&T

CRYSTAL DESIGN AND ENGINEERING

Mr. John Sherman, General Electric Company

OSCILLATORS AND SYNTHESIZERS

Mr. John Sherman, General Electric Company

ATOMIC AND MOLECULAR FREQUENCY CONTROL

Professor Norman Ramsey, Harvard University

TIMEKEEPING AND DISTRIBUTION

Professor William Markowitz, Marquette University

INDEX

PAGE

Introductory Session Honoring Roger Sykes on His Retirement

- E. A. Gerber 1
- W. P. Mason 8
- J. M. Wolfskill 13

Surface Waves and Devices
- H. J. Shaw 16

Fundamental Studies and Quartz Crystal Research

Thickness-Twist Vibrations of a Quartz Strip
- R. D. Mindlin 17

Zero Temperature Coefficient Ultrasonic Delay Lines Utilizing Synthetic Quartz Crystals as Delay Media
- M. Onoe and Y. Mochizuki 21

Frequency-Temperature Dependence of Thickness Vibrations of Piezoelectric Plates
- K. Hruska 33

The Effect of Static Electric Fields on the Elastic Constants of α - Quartz
- J. A. Kusters 46

Selected Topics in Quartz Crystal Research
- C. A. Adams, G. M. Enslow, J. A. Kusters and R. W. Ward 55

Defects and Frequency Mode Patterns in Quartz Plates
- E. W. Hearn and G. H. Schwuttke 64

Filtering Techniques and Crystal Filters

Review of Digital Filtering
- J. D. Heightley 74

Active Filter Capabilities
- P. Geffe 78

* Generalized Filters Using Surface Ultrasonic Waves
- M. G. Holland 83

Monolithic Crystal Filters
- R. J. Byrne 84

Preparation of Quartz Crystal Plates for Monolithic Crystal Filters
- A. J. Miller 93

A Technique for Automatic Monolithic Crystal Filter Frequency Adjustment
- R. P. Grenier 104

* Abstract only - complete paper not received in time for publication

PAGE

Crystal Design and Engineering

Auger Spectroscopy in Studies of the Aging Factors of Quartz Crystal Resonators
- G. W. Simmons, W. H. Hicklin and R. K. Hart..... 111

5 MHz BT Cut Resonators
- J. G. Leach..... 117

Mode Control and Related Studies of VHF Quartz Filter Crystals
- T. J. Lukaszek..... 126

Low Aging Crystal Units for Use in Temperature Compensated Oscillators
- J. F. Silver and L. A. Dick..... 141

Dynamic Temperature Behavior of Quartz Crystal Units
W. H. Hicklin..... 148

The Direct Temperature Control of Quartz Crystals in Evacuated Enclosures
- F. G. Tinta, A. S. Matistic and G. A. Lagasse..... 157

Crystal Measurements

Precision Measurement of Crystal Frequency by Means of "Center Line Method"
- I. Koga..... 168

A Report on IEC Technical Committee TC-49
C. Franx..... 172

Quartz Crystal Measurements
- E. Hafner and A. Ballato..... 177

Oscillators and Synthesizers

Temperature Compensated Crystal Oscillators
- P. Duckett, S. Peduto and G. Chizak..... 191

New Approach to a High Stability Temperature Compensated Crystal Oscillator
- S. Schodowski..... 200

An Ultra Low Noise Direct Frequency Synthesizer
- D. G. Meyer..... 209

Atomic and Molecular Frequency Control

Laser Frequency Stabilization Techniques and its Applications
- H. S. Boyne..... 233

Automatic Frequency Control and Phaselocking of Lasers
- V. J. Corcoran, R. E. Cupp and J. J. Gallagher..... 240

Areas of Promise for the Development of Future Primary Frequency Standards
- H. Hellwig..... 246

Hydrogen Maser with Deformable Storage Bulb
- P. Debely..... 259

Surface Collision Frequency Shifts in the Atomic Hydrogen Maser
- P. W. Zitzewitz..... 263

PAGE

Studies of Hydrogen Maser Wall Shift for High Molecular Weight Polytetrafluorethylene
- R. F. C. Vessot and M. W. Levine..... 270

Frequency Stabilization of CO_2 Lasers with Respect to Passive SF_6 and CO_2 Line Centers
- P. Rabinowitz, R. Keller and J. T. LaTourrette..... 275

Atomic and Molecular Frequency Control

* New Information on the Physics of Rubidium Gas Cells
- P. Bender and V. W. Cohen..... 279

Progress Report on the Rubidium 85 Maser
- J. Vanier, R. Vaillancourt, G. Missout and M. Tetu..... 280

An Optically Pumped Parametric Frequency Converter
- H. Tang and W. Happer..... 285

Cesium Beam Servo System Using Square Wave Frequency Modulation
- H. Daams..... 294

Frequency Comparison of Five Commercial Standards with a NASA Experimental Hydrogen Maser
- A. R. Chi, F. G. Major and J. E. Lavery..... 301

A New Primary Cesium Beam Frequency Standard
- F. Kupersmith, C. Thornburg and J. Ho..... 308

Timekeeping and Distribution

Time/Frequency Technology in System Development
- R. E. Perkinson..... 315

A Survey of Time and Frequency Dissemination Techniques
- J. Jesperson..... 322

Time and Frequency Transfer Via Microwave Link
- D. Phillips, R. Phillips and J. O. Neill..... 325

Diurnal Phase of VLF Signals Near Antipode of a Transmitter
- A. R. Chi 332

A Second Satellite Oscillator Experiment
- R. Easton, C. Bartholomew and J. Bowman..... 339

The Omega Navigation System as a Source of Frequency and Time
- W. Palmer..... 345

The NBS Atomic Time Scale System: AT(NBS), SAT(NBS), and UTC(NBS)
- D. W. Allan, J. E. Gray and H. E. Machlan..... 361

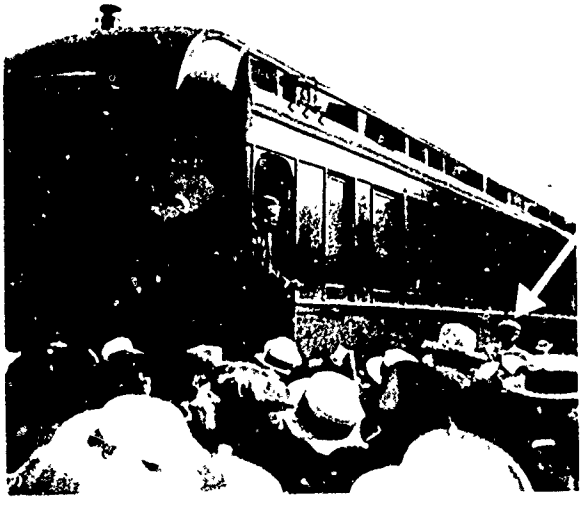
ROGER SYKES AND HIS CRYSTALS

Speech Honoring Roger Sykes on the Occasion of His Retirement

E. A. GERBER
Director, Electronic Components Laboratory
USAECOM

I am indeed happy that the subject of my remarks before this illustrious gathering today is not an inanimate quartz crystal or an atomic clock, but a real person and--even better--a person whom we know and respect, both professionally and personally.

Roger Sykes was born "a long time ago" in Windsor, Vermont, and even as a babe in arms, he came to meet important people--an indication, perhaps, of his range of contacts later in life. This event happened near the end of Teddy Roosevelt's Presidency.



Notice how Teddy Roosevelt is looking right at Roger, no doubt attracted by the fact that Roger was the only one not paying attention. After preparatory school, Roger attended the Massachusetts Institute of Technology while installing telephone apparatus for the Western Electric Company. He received his BS Degree in 1929, and his MS in 1930. Here's how young Roger looked around this time.

After Western Electric and with the New York Telephone Company in New York City for a short while, Roger started his lifetime career in quartz crystals at the Bell Telephone Laboratories in New York City. He began with the development of quartz crystals for filters in 1929; in 1930, he added research on quartz crystals and piezoelectric devices for transducers and networks for frequency control. We see him here at the test bench in Bell adjusting one of the early crystals.

It was indeed fortunate that early in his career Roger became associated with Warren Mason at Bell Laboratories. Their work in the following years resulted in two papers published in Bell Systems Technical Journal. One, published in 1937, was entitled "The Use of Coaxial and Balanced Transmission Lines in Filters and Wide-Band Transformers for High Radio Frequencies,"[1] and the second, published in 1940, was entitled "Electrical Wave Filters Employing Crystals with Normal and Divided Electrodes."[2] Here, Roger and Warren Mason thoroughly discussed all standard types of filters with crystals and the methods for determining their constants and attenuation characteristics.

With the arrival of World War II, Roger's R&D work became mainly concerned with the Armed Services communications systems. One of his important assignments was his work at the Western Electric Production Plant at Hawthorne, Illinois, where, commuting from the East Coast, he was the driving force in setting up the production of DT-cut crystal units. To appreciate the importance of his effort and the effort of many others in this critical war period, may I quote from the introduction to Heising's book, "Quartz Crystals for Electrical Circuits":[3]

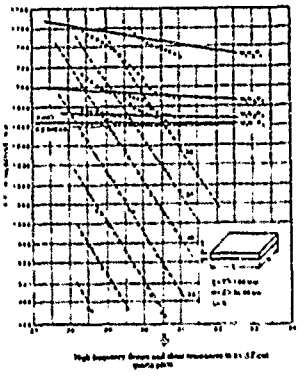
> "... With the thousand-fold expansion in crystal manufacturing facilities that was necessary and the decentralization that is wise from the military standpoint, scores of manufacturers had to be educated in the art and science of crystal manufacture...
>
> "The Western Electric Company in meeting its orders found it necessary to expand its crystal production 750-fold. To accomplish this it was necessary to introduce to the intricacies of crystal manufacture a large number of engineers inexperienced in the art..."

Roger was deeply involved in the national expansion of the U. S. crystal manufacturing capability. Typically, the effect of the strenuous work and strain showed up very soon--he started to have stomach trouble. But, as always with him and in his environment, humor prevailed despite his aches and pains and pressures of the moment. One day when working a long weekend, he sent someone out for a quart of milk to calm his stomach, and someone asked, "You mean to tell me you're now going to wash the crystals in milk?"

In 1943, he moved from New York City to the Bell Telephone Laboratories in Murray Hill, New Jersey, still deeply involved in his war-time activity. In 1944, he published his last joint paper with Warren Mason, "Low Frequency Quartz Crystal Cuts Having Low Temperature Coefficients,"[4] in the Proceedings of the IRE. As you probably all know, it described the famous MT- and NT-cuts which use coupling of the main mode with some other mode to decrease the temperature coefficient of frequency.

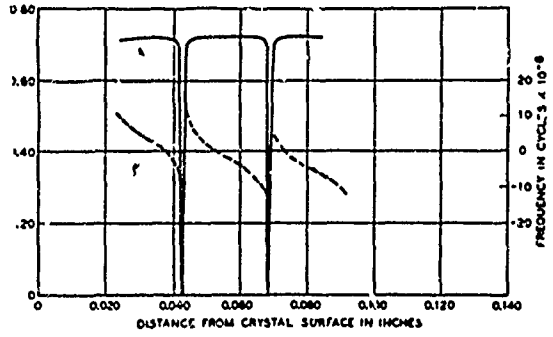
A year later, in 1945, Roger was made Department Head. I have here some memorable comment from Jack Griffith, one of his associates, regarding this event, which I must share with you. Griffith says, "When our Department Head was replaced by Roger A. Sykes, we had an indignation meeting. It took some calling back and forth to get him there. We told him, 'We don't like your way of operating. We are not going to stand for engineering by edict.' We didn't want him to run his engineers in a high-handed manner and dictate details of the job. Roger took it in good spirit and we got along fine." To which I add my observation that it's rather unusual--if not downright dangerous--for a group to tell their boss they don't like the way he operates. It's even more unusual for the boss to accept criticism gracefully from his subordinates. It showed Roger as a sensitive and responsive being, and a man who respected and valued another man's opinion. The problems never surfaced again.

For the next 9 years (until 1954), the main duty of Roger's department was the development of crystal units for network use and frequency control applications in Bell Systems apparatus. Roger published two very important articles in this period, both of lasting credit on his record. The first was an article in Heising's book,[3] published in 1946, on the modes of motion in quartz crystals. Since most of us have frequently used the information in this article, I will not dwell on this important contribution; I will, however, refresh your memory with a graph.



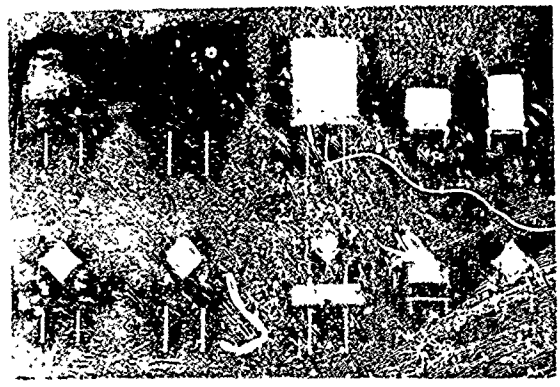
It shows the frequency of flexure and high frequency shear resonances in an AT-cut quartz plate as a function of the length to thickness ratio. By measuring flexure and low frequency shear modes in AT- and BT-cut plates and studying their coupling with the desired high frequency shear, Roger came up with the rules for dimensioning rectangular and square crystal plates in order to avoid activity dips. This article was the Bible to crystal engineers as long as rectangular and square crystals were used.

The second article, also in Heising's book, is the paper on "Principles of Mounting Quartz Plates." This investigation was essential for the development of low frequency CT- and DT-cut resonators. This graph shows an example of how the activity and frequency of a CT-cut crystal are affected by the distance of the clamping from the surface of the crystal.



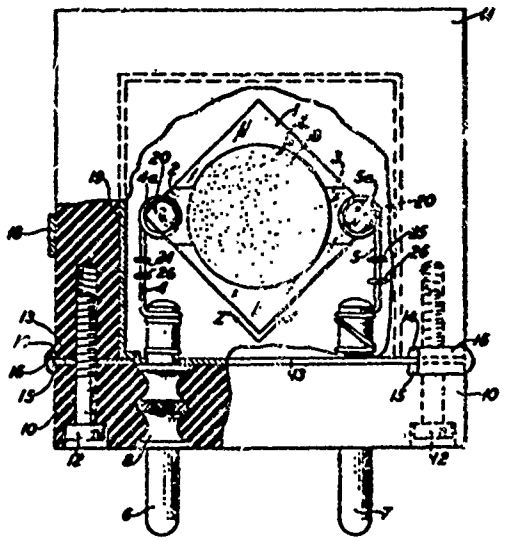
Effect on frequency and activity of clamping one lead of 370 kc. CT-cut crystal.

In January 1948, another paper appeared; this time by Roger alone: "High Frequency Plated Quartz Crystal Units."[5] I understand from his associates that he had quite a time convincing Bell Laboratories' management of the advantage of plated units over clamped type units. How successful he was can be seen from the fact that his design of a wire mounted plated crystal unit is still used today. Shown here are various types of high frequency crystal units that he developed.



The unit in the middle and the type of mounting, I am sure, are familiar to everyone even those who may have recently entered our field.

The drawing below shows a wire mounted unit as described in detail in one of Roger's important patents.



I should mention at this point that he has a total of 17 patents to his credit in the field of crystal units, crystal oscillators, and crystal filters--most of these are described in his publications.

During this highly productive period at Murray Hill (from 1943-1954), he acquired additional duties in 1950; namely, work in ferroelectric crystal units for memory circuits, solid ultrasonic delay lines, sealed terminals, ovens, thermostats, and frequency standards.

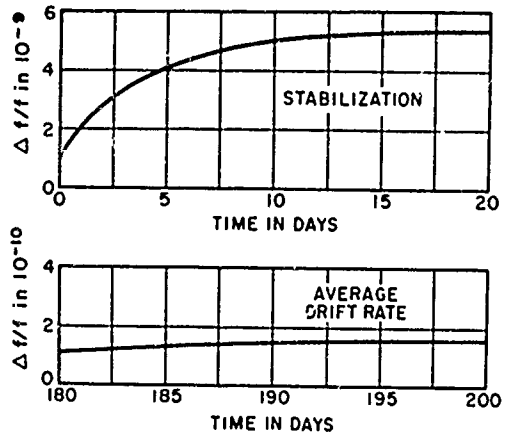
During September and October of 1953, a special event took place in which I had the pleasure of participating, and during which time I learned to know Roger much better. At the suggestion of Mr. W. L. Doxey of the Signal Corps, Major General G. L. Back, the Chief Signal Officer, established a three-man team of consultants to visit each Army contractor working in frequency control, the purpose being to evaluate work going on and to make recommendations as to future work in that field. The team consisted of the unforgettable Karl Van Dyke, Roger Sykes, and myself. I can rightly state that we worked by day and traveled at night for a full month. Among the things that I learned about this man was that he had a warm personality, a wonderful sense of humor, and driving habits that made me wish I had more insurance!

In 1954, the important development of crystals for a submarine cable started at Bell Telephone Laboratories; this must have been quite an interesting challenge and experience for Roger. Then, during that same year, his laboratory was moved to Whippany, New Jersey, where he continued with his assignments from Murray Hill. His work on crystal filters progressed and as a result, he came out with a paper at the 1958 IRE Convention on "A New Approach to the Design of High Frequency Crystal Filters."[6] I will refrain from getting into more details on Roger's filter papers because this will be done much more competently by Warren Mason.

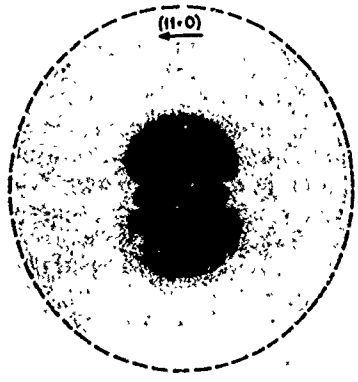
In 1959, his laboratory was again moved; this time from Whippany to North Andover, Massachusetts, where his duties were slightly changed, since, in addition to quartz crystals and circuitry, they now embraced transformers, inductors, thermoelectrics, ovens, and ferrite memory apparatus. I know that he was not too happy with components other than the frequency control devices and so, in 1961, he succeeded in getting rid of transformers, inductors, memories, and so on. At that time, his stomach ulcers kicked up again, and had to resort to surgery in 1961. Asked what the effect of the operation had on his drinking habits, Roger answered that since the operation, he gets more mileage per drink!

Then his laboratory was on the move again, this time to BTL in Allentown, Pennsylvania, where he stayed until his retirement in 1969. His field of work again embraced the entire area of frequency control and selection. In December of 1962, he published a paper in the IRE Transaction on Instrumentation, on "Performance of Precision Quartz Crystal Controlled Frequency Generators,"[7] together with Warren Smith and Bill Spencer. To remind you what big strides have been made by his team, the next photographs of a 2.5 MHz precision crystal unit and a graph on the stabilization and drift rate of a 2.5 MHz precision oscillator show these achievements.

Here, we have plotted the frequency change in parts in 10^9 and parts in 10^{10} as a function of time. Note that the daily drift rate is less than 1 part in 10^{11}/day.

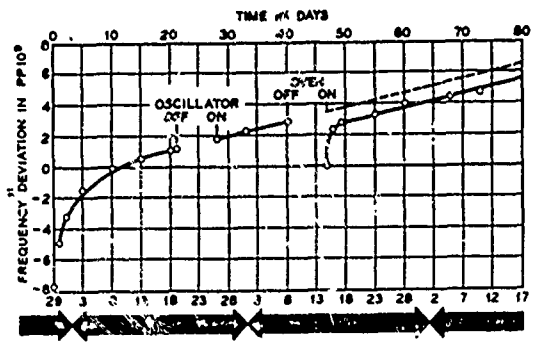


A final paper on the same topic authored again with Smith and Spencer was presented at the 17th Annual Symposium on Frequency Control in May 1963.[8] Again, a very important contribution was made to the art of high precision crystal oscillators. Specifically, the topographic X-ray diffraction method to study imperfections and vibrational modes in crystal units proved to be extremely important. This is an example of an X-ray photograph of the $m_5n_1p_3$ mode of a 5th overtone AT-resonator.



Of equal importance was the study of the effect of stopping the oscillation and of oven shut-down on the stability of the crystal.

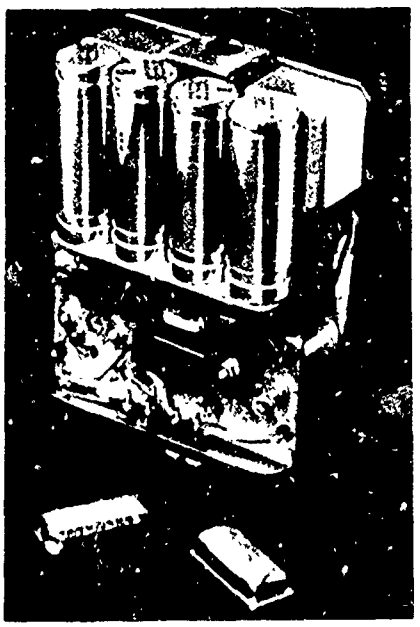
This graph shows how much the influence of those effects on frequency has been reduced due to refinement of evacuation and high temperature baking-out processes of the crystal enclosure.



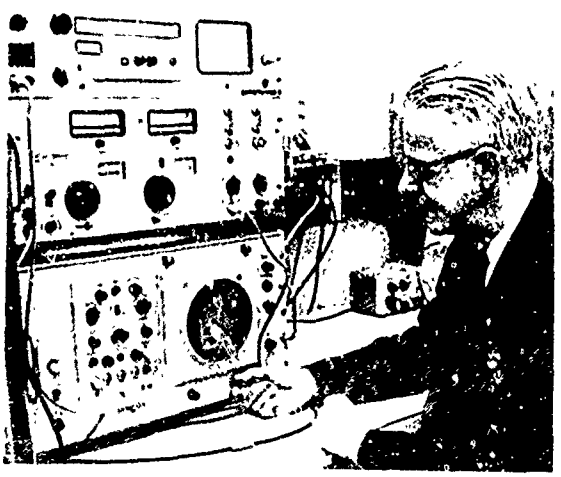
Previously, both effects changed the frequency by a larger amount; namely, in the order of 5 parts in 10^9, and it took some weeks until the old equilibrium was reached again.

The accomplishments reported by the last two papers mentioned are to some extent the result of BTL's several R&D contracts with the Signal Corps (now the Electronics Command). This work was handled in Roger Sykes' Department and involved "Ultra Precise Standards of Frequency," "Quartz Crystal Controlled Oscillators," and "Fundamental Studies of the Properties of Natural and Synthetic Quartz Crystals." I might state that the government profited very much by this contractual relationship, not only by obtaining the valuable results of these competent studies, but also by the close professional association with Roger and his people.

In the following years, our esteemed guest, together with his co-workers, came out with two important papers on monolithic crystal filters.[9, 10] Again, I will leave it to Warren Mason to go into more details. I only want to show what has been accomplished by him and under his leadership.



In the background is an old channel filter which has been widely used by Western Electric. In the foreground, with cover removed, is a modern monolithic filter which has the same performance, is much smaller, and costs less than the old design. This new design requires not only precision adjustment of the properties of each resonator area but of the coupling between the resonator areas as well. The significance of the achievement is evident. Here we have a more mature Roger measuring the transmission characteristic of one of his monolithic filters.



Finally, Roger has to his credit two survey papers entitled "State of the Art - Quartz Crystal Units and Oscillators"[11] and "Quartz Frequency Standards,"[12] which appeared in 1966 and 1967 in the Proceedings of the IEEE. I am indeed happy to say that I had the pleasure of co-authoring these papers with Roger.

In summary, I found that it was the consensus of his people that Roger always demonstrated great technical and administrative ability. He liked to see everything in a simple physical picture and was very able to handle personnel problems.

Roger's professional picture would not be complete if I did not mention the various honors he received and the many committee activities in which he was and is still engaged.

From November 1951 to February 1952, he was technical consultant to Major General John Christmas, Assistant Chief of Staff, Department of the Army, to survey production of the U. S. crystal industry during the Korean emergency. I already mentioned that in 1953 he was a technical consultant to the Chief Signal Officer to evaluate the Department of the Army research and development program in the field of frequency control.

In 1957, he was honored as a Fellow of the Institute of Electrical and Electronics Engineers, and in 1967, he received the Distinguished Engineer Award from the Pennsylvania Section of Professional Engineers. He was and still is very active in various IEEE and EIA committees. Last, but not least, he was extremely successful as the Technical Advisor to Committee TC-49, Crystals, of the International Electrotechnical Commission. I believe that I do not exaggerate if I state that due to Roger's efforts the international standards on crystal units have a high level of perfection. Since they are patterned after the U. S. military and commercial standards and have been accepted internationally by most countries as national standards, it should be very easy for our industry to fill foreign orders for crystal units because they automatically conform with the national standards of most foreign countries.

How busy he was at international committee meetings is clearly shown in this picture (Roger is second from left).

No doubt that this was a fact-finding meeting dealing with 'quartz'---quarts of beer, that is. I suppose that I could also say, "He's in good spirits." Too bad that I missed the occasion! However, I did participate with him at a subsequent IEC meeting in New Delhi, India, and you can see how serious he looked and how strenuously he worked then in behalf of standardization.

In closing, I want to add a couple of personal remarks. I knew of Roger's work in the U. S. when I was still on the other side of the Atlantic involved in quartz crystals, and I already had a profound professional respect for him at that time just through his published works. It would never have occurred to me that fate would bring me into a close professional

association with this man in America, much less that there would develop a bond of friendship between us; but, needless to say, I am extremely happy and grateful that this came to pass. And so, it gives me the greatest pleasure to be able to pay this public tribute to him, not only to a highly competent professional, but also to a wonderful human being.

Roger A. Sykes
1970

REFERENCES

Papers Authored/Co-authored by R. A. Sykes

1. "The Use of Coaxial and Balanced Transmission Lines in Filters and Wide-Band Transformers for High Radio Frequencies," Bell Sys. Tech. J., Vol. XVI, pp. 275-302, July 1937.

2. "Electrical Wave Filters Employing Crystals with Normal and Divided Electrodes," Bell Sys. Tech. J., Vol. XIX, pp. 221-248, April 1940.

3. "Quartz Crystals for Electrical Circuits," R. A. Heising, D. Van Nostrand Co., Inc., 1946.

4. "Low Frequency Quartz Crystal Cuts Having Low Temperature Coefficients," Proc. IRE, Vol. 32, pp. 208-215, April 1944.

5. "High Frequency Plated Quartz Crystal Units," Proc. IRE, Vol. 36, pp. 4-7, January 1948.

6. "A New Approach to the Design of High Frequency Crystal Filters," IRE Convention Record, Part 2, pp. 18-29, 1958.

7. "Performance of Precision Quartz Crystal Controlled Frequency Generators," IRE Trans. on Instrumentation, Vol. I-11, pp. 243-247, December 1962.

8. "Studies on High Precision Resonators," Proc. 17th Annual Symp. on Frequency Control, pp. 4-27, 1963.

9. "High Frequency Monolithic Crystal Filters with Possible Application to Single Frequency and Single Side Band Use," Proc. 20th Annual Symp. on Frequency Control, pp. 288-308, 1966.

10. "Monolithic Crystal Filters," IEEE International Convention Record, Part 11, pp. 78-93, 1967.

11. "State of the Art - Quartz Crystal Units and Oscillators," Proc. IEEE, Vol. 54, pp. 103-116, February 1966.

12. "Quartz Frequency Standards," Proc. IEEE, Vol. 55, pp. 783-791, June 1967.

ROGER SYKES CONTRIBUTION TO QUARTZ CRYSTAL FILTERS

Warren P. Mason
Henry Krumb School of Mines
Columbia University, New York, N.Y. 10027

When Roger Sykes first came to Bell Laboratories as a graduate from MIT, he joined my department which at that time was working on the application of quartz crystals to electrical wave filters. In fact as I recall it, he wouldn't come to Bell Laboratories unless he got in this department. He afterwards paid me the doubtful compliment of saying that I was the best boss he ever had because he could do what he wanted without interference. In his case it was certainly the best policy since with his originality he produced a lot more than if he was directed.

Figure 1 shows the state of the art at that time. The dashed curve shows the best that could be done with a coil and condenser filter at 21.3 KHz whereas the solid curve showed the improvement that could be obtained by using four crystals per section with coils on the two ends to widen the band. The result for 4 KHz spacing was to produce a useful band of 2850 Hz and a distortion of 1 db compared to a band width of 2550 Hz and a distortion of 5 db.

With four crystals in each section this was rather an expensive design. It was soon suggested that two of the crystals could be replaced by a single plate with two sets of platings. Roger developed the first stable holder for this divided plate crystal as shown in Figure 2. In fact, the first quartz crystal filters for the coaxial carrier system were constructed in this form. The subsequent change in the mounting to the headed wire mount shown in Figure 3, was also made by Roger. In this mount, a baked on silver paste is used to provide a base for soldering a headed wire at the nodal point of the vibration. The little motion communicated to the wire, at the nodal point, is reflected back to the vibration by in effect clamping the wire with a solder ball at a flexural nodal point. As a result the low impedance wire resonates at the same frequency as the crystal and simply adds reactance elements to the vibration. This type of mounting is still used for filter and oscillator plates although the solder joint has been replaced by thermo compression bonds to cut down the dissipation.

The original set of channel filters ranged from 60KHz to 108KHz. Since it was impractical to use a divided plate crystal at higher frequencies Roger devised the hybrid coil type of filter shown in Figure 4. In this case, a crystal in each arm of the balanced hybrid takes the place of two divided plate crystals in the balanced lattice on the right. Since these crystals can be made to considerably higher frequencies, they have been used in filters as high as 200KHz.

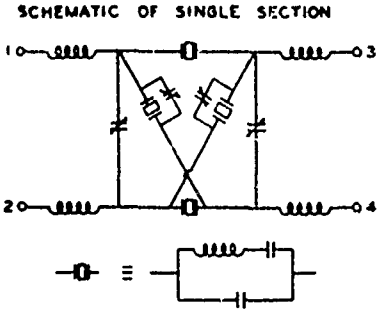
Extension of the coaxial cable system to higher frequencies created a need for narrow band pilotchannel filters for controlling the gain of the system. Some of these filters were as high in frequency as 3MHz. The requirements stipulated that there should be no pass bands of any magnitude below the desired frequency which would affect the transmission of the pilot channel response at 3096KHz. As it was not possible to produce longitudinally vibrating crystals at this high frequency a thickness vibrating crystal of the AT type had to be used. Since the cross-sectional dimensions were much larger than the thickness, there were many other resonances in the neighborhood of the fundamental. In order to relate these resonances to the dimensions of the crystal, Roger made a complete study of the resonances of an AT cut crystal 1 MM in thickness as a function of the dimension along the x or electric axis. The results are shown in Figure 5. As you can see there are strong resonances along the line marked 11,11 which are coupled to another set of resonances crossing them at an angle. These were later shown by R. D. Mindlin to be flexure modes driven by the thickness shear mode.

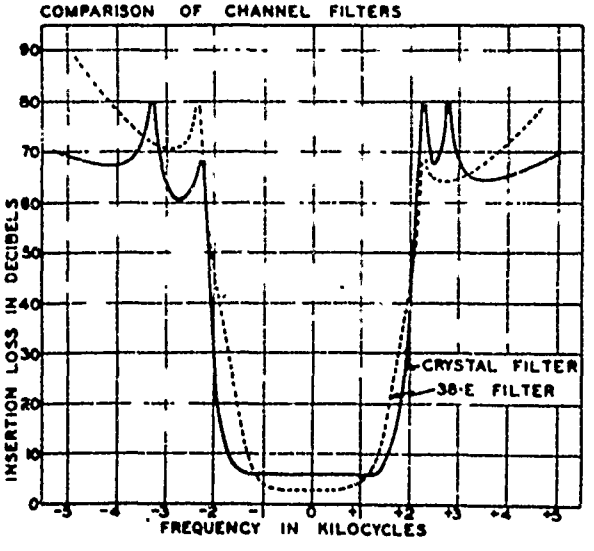
To obtain a pilot channel the dimensions were chosen so that the shear mode was halfway between two flexure modes. The filter structure used by Roger--which I had to abstract from one of his Bell System memorandum dated 1938--is shown in Figure 6. It consists of two sets of plating on an AT cut crystal separated by 1 MM of bare crystal. There were two pass bands one at 3096KHz and the other at 3108KHz. In fact the structure is a forerunner of the monolithic crystal filter which is now being applied to the cable, radio and submarine cable systems of the Bell System. If the separation between the two sets of platings had been increased somewhat to cut down the coupling, a single pass band would have resulted in the manner of the present monolithic crystal filter.

Shortly after this work, World War II broke out, and all the people in crystal research were put to work developing crystals for the military. This phase of Roger's work has been covered by Dr. Gerber. After the war, the company decided that Roger would be more valuable as head of oscillator and crystal development rather than in crystal research.

The last phase of Roger's crystal filter work that I would like to mention is his contributions to the monolithic crystal filter. The monolithic crystal filter is an eight section filter constructed on a single quartz crystal by the use of eight controlled thickness plates separated by regions of bare crystal. This construction shown in Figure 7 produces at high frequencies the equivalent attenuation of the channel filter shown in the background. Figure 8 shows the characteristic of an eight section filter. Because of the high Q of the crystal the mid band loss and distortion at 8MHz is actually less than that of that of the present channel filter at 60KHz.

Roger's contribution to this development was partly inspirational and partly supervisional. The published papers have the names of Sykes and Beaver and Sykes, Smith and Spencer. I recall that Bill Beaver had a thesis at Lehigh and a paper on the properties of the monolithic filter. Smith and Spencer made many contributions to the modes of vibration and the calculation of the filtering properties of such filters. However, it is doubtful if this development would have materialized without the inspiration and direction that Roger supplied it.





	MID-BAND FREQUENCY	BAND WIDTH	DISTORTION
CRYSTAL FILTER	62.3 KC	2850~	1.0 DB
38-E FILTER	21.3	2550	5.0

FIGURE 1 Insertion Loss of Coil and Condenser and Crystal Filters

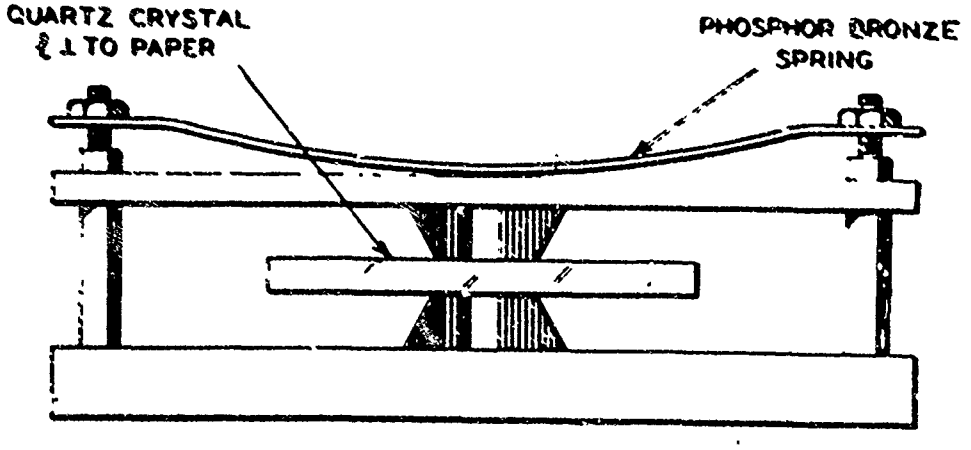


FIGURE 2 Pressure Mounting for Extensional Crystals

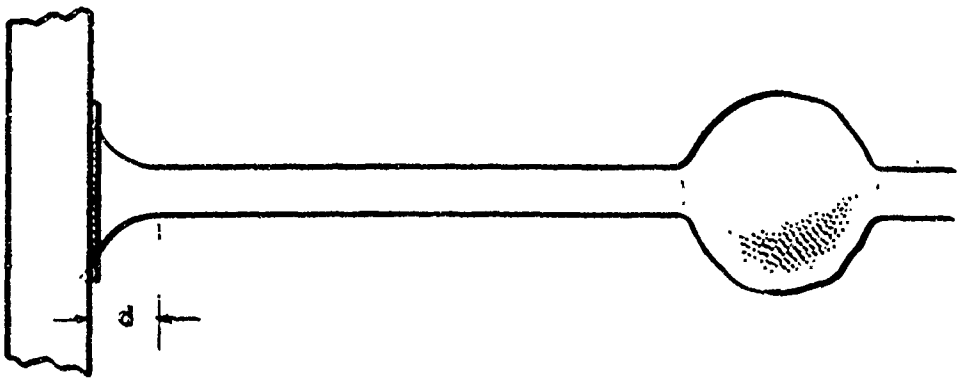


FIGURE 3 - Headed Wire Type Mounting

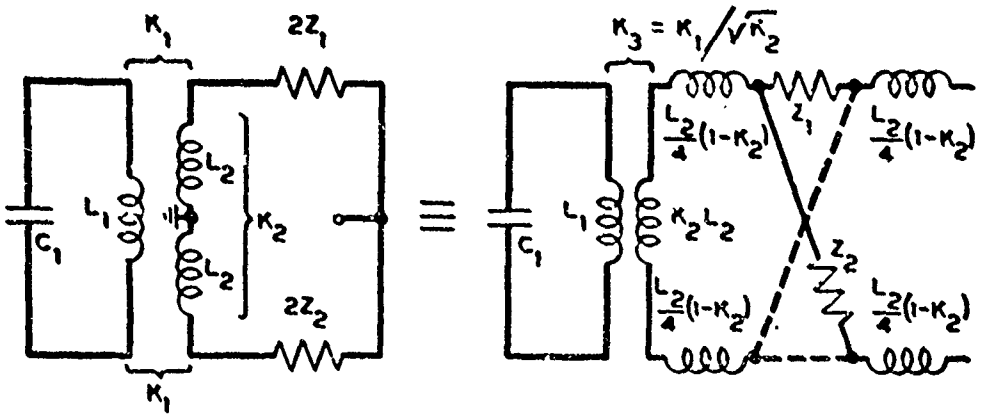


FIGURE 4 - Hybrid Coil Type of Filter

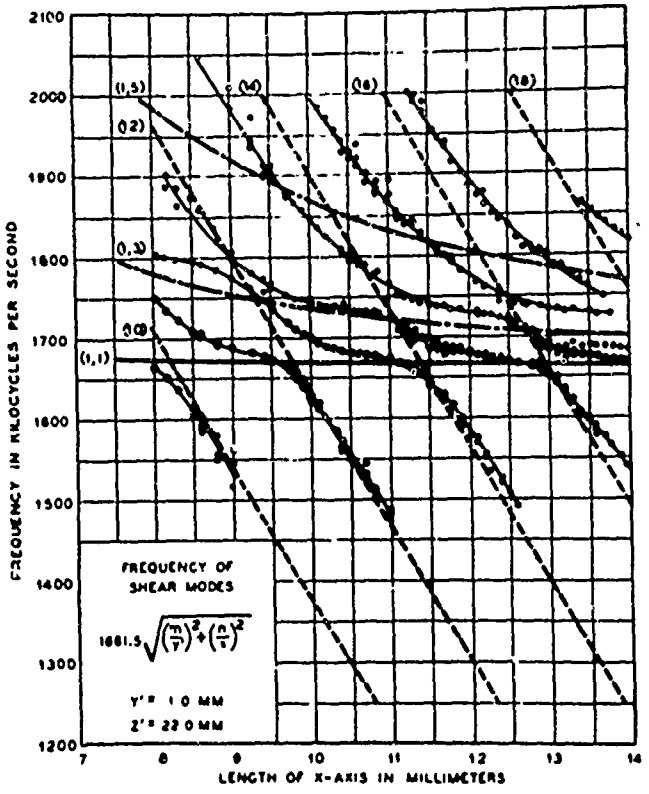


FIGURE 5 - Spectrum of an AT-Cut Crystal as a Function of the Width Dimension

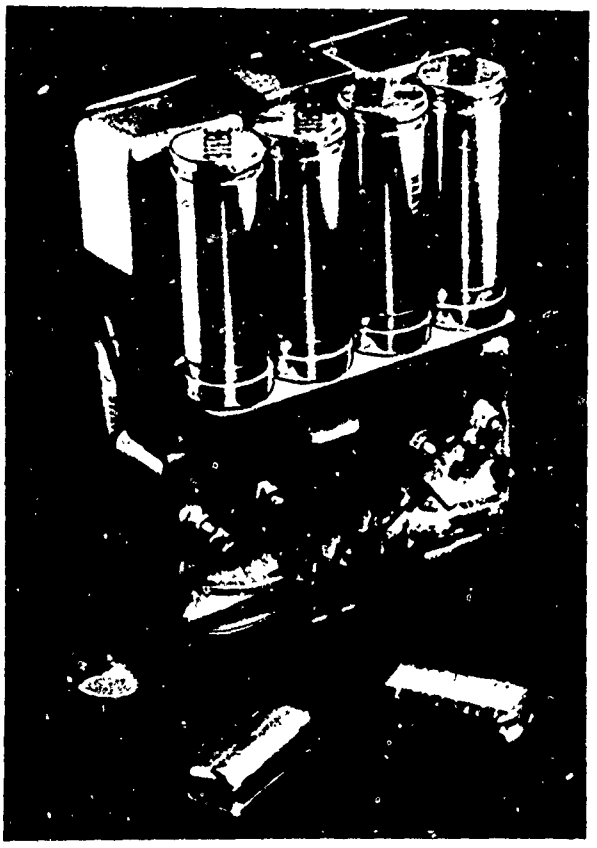


FIGURE 7 - Channel Filter (Back) and Monolithic Counterpart (Front Left)

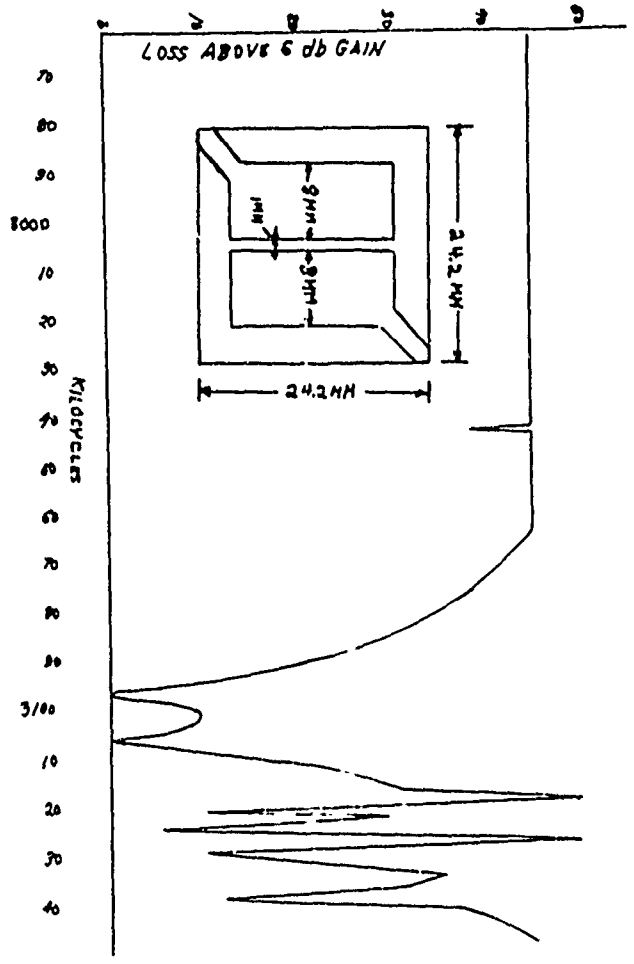


FIGURE 6 - Loss Characteristics of 3090 MHz Pilot Selector Unit - Zero Mode

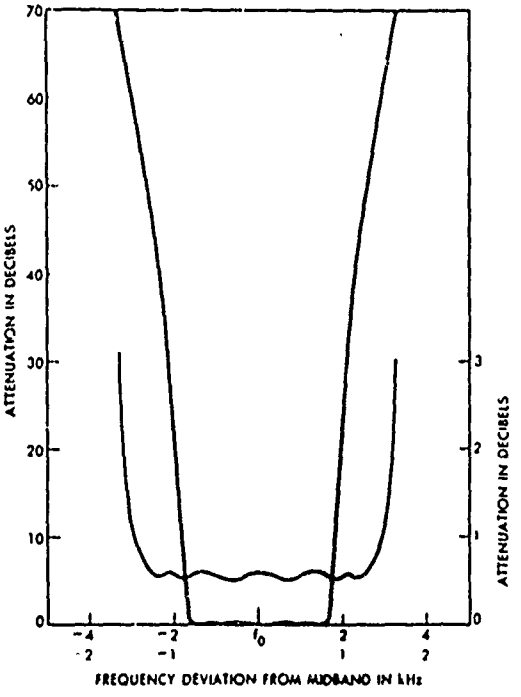


FIGURE 8 - Typical Response Curves of an Eight Resonator Monolithic Filter

TO ROGER A. SYKES ON HIS RETIREMENT

J. M. Wolfskill
Bliley Electric Company
Erie, Pennsylvania

It was forty years ago in 1930 when I first met Roger Sykes. The place was in the crystal cutting shop at Bell Telephone Laboratories, 463 West Street, New York. We both had the same reason for being there: to expedite delivery of some X-cut quartz bars for filters on which we were working. Roger was working for Dr. Mason in the Research Department, and I was working for Mr. Lane in Apparatus Development.

William Broughton was then in charge of the crystal cutting shop, and he was not about to bother much with a couple of "neophyte" engineers wanting a few experimental crystals cut; not when he had a real deal going. To impress us, he then produced a beautiful 1 1/2" square 590 kHz Y-cut crystal unit he had just finished for Station W.O.W. Omaha Nebraska. Boys, he said, this beauty sells for $700.00 and has to be on tonight's mail flight or "Woodman of the World" goes off the air, and we certainly can't let that happen. Properly impressed, we both dickered for best delivery of our individual requirements.

Roger -- as I was to learn later -- always won these rounds. Of course, he was working for Dr. Mason and that alone carried some prestige. We have been "tangling horns" ever since though in many ways -- in the IRE Standardization Committee meetings; EIA, then called RETMA, Standards meetings; in Military Specification Reviews; etc. The real reason Roger always won: "he had the longest horns!"

An astute politician, he always used the right approach. Did you ever notice how he starts off an argument with, "well that's right." He is getting you off guard by agreeing you are right; then comes the "but" and he is off on the trail to accomplish his mission. The old adage "if you can't lick them join them" is not Roger's method. His motto is "if you can't lick them, convince them to join you." You all know what a convincer he can be. But this ability stood him in good stead through the years, and he literally accomplished miracles in some almost hopelessly deadlocked situations.

Roger was using sputter plated crystals on low frequency bar types in the early thirties, and the next big step of course, was to apply this same thinking to high frequency AT crystals and get rid of the old pressure mounted design. Here is where he hit a blank wall. Experts, and there were few then in this field, insisted the heavy electrodes, and close thermal coupling to a radiator were necessary to dissipate the heat; old fashioned maybe, but a necessity. Also, the contention that contiguous electrodes of evaporated or sputtered metal would disrupt and soon disintegrate, was a very strong argument against this approach. Obviously; development work and salesmanship of a high order, were needed for such a radical departure from tried and true methods. Roger Sykes met that challenge, and made extraordinary efforts, by first proving and convincing his own people at Bell Labs, subsequently the Western Electric Company, and later the Military, of the distinct advantages of plated units if properly used. In all these arguments he was careful to underscore, if properly used. This he did through various R & D Panels with Industry and the Military.

Roger was probably the first crystal man to dictate to an oscillator design man how his circuit should be designed, and make it stick. "A crystal oscillator is a frequency control device and not a powerhouse, and must be used accordingly." This was a household phrase in Sykes' department at Bell Telephone Laboratories, and it became one at Western Electric Company.

Old timers like Lack, Willard, Fair, Heising, Llewellyn, Koerner, Thurston, Ziegler, and many others, I am sure can recall the days when the tank circuit of the crystal oscillator was frequently directly coupled to the antenna, and the plate voltage was directly modulated. Special radiators were made and fastened to the crystal holder to dissipate and radiate the heat generated. This type of abuse simply had to be stopped, if the remarkable stability inherent in the quartz crystal were ever to be fully realized. A crusader was sorely needed.

While Roger never made the front page of Time Magazine for this cause, like "Barry Commoner" on "Ecology", he was in the thick of it all, not only at Western Electric and the Military, but wherever there was a clamor for better frequency stability and tighter frequency tolerances. Drive power levels -- and these are Roger's words -- are as much a part of a crystal specification as is the frequency tolerance.

Equipments had to be redesigned; test sets had to be built, which would present the same conditions to the crystal as it would see in the redesigned end use equipment. Obviously, "ball bats" had to be used to make such radical changes in record time. They were accomplished with "ulcers 'a plenty", but no bloodshed.

Probably the largest single impact of these efforts was during World War II in the production of the CR-3 and CR-4 type tank set crystals. What with U-boats sinking boatloads of quartz being shipped from Brazil, and quartz in short supply, and considering the tremendous requirements of the tank set application alone, something had to be done.

The old low frequency AT ice cube 1" square, and anywhere from .050" to 1/4" thick was out of the question from a quartz consumption standpoint. CT and DT cuts were developed to do the job up to 1200 kHz. At 1200 kHz these were only .1" square and approximately .010" to .015" thick, and Roger's eyes gleamed when he said, "Why, one AT ice cube at 500 kHz will yield fifty to one-hundred DT's from the same amount of quartz." Imagine however, the oscillator engineers' "chagrin" when told that this little sliver of quartz would have to do the same job as the so called ice cube!

Needless to say the operation was successful and Western Electric Company Hawthorne Works produced millions of these FT-241 gems, but only after Roger's laboratories staff, and Western Electric engineers developed ulcers and pulled out most of their hair. In fact the operation was so successful that the War Production Board raised serious doubts about what was going on, and thought possibly Western Electric had their own quartz mine because crystal units were being turned out by the hundred thousand, while raw quartz intake was practically nil by comparison to what would have been required had AT cuts been used.

Plating crystals with contiguous electrodes became standard procedure in the Industry, in all varieties, and completely replaced the pressure mounted type, except for a few military types still remaining. Many of the mounting methods, enclosures, and processing techniques, in practically all phases of the crystal industry, whether for oscillators, filters, or transducers, have in some way felt the guiding hand of Roger Sykes.

Introduction to koldwelding enclosures, high temperature bakeout, use of oil free vacuum systems in processing for higher stability and reliability, are just a few of the things which he helped implement and promote. Remember the Loran C 100 kHz GT cut crystal unit; the highest Q crystal ever produced until Warner came along.

Now, we can't give him credit for doing all these things single handed; he had lots of help -- like the farmer who went down to Virginia and bought an old broken down and worn out farm. This fellow had great ambition, and worked hard and long repairing the buildings, fertilizing the soil, and after about five years raised some bumper crops. The local minister passing by one day saw the bountiful crops the farmer was raising, and ringing his hands thought "here is a good and likely heavy contributor to become a member of my church." So he stopped for a visit. The farmer gave him a tour of the farm and the minister marveled at the crops. "My," he said, "what a beautiful crop of corn you and the Lord raised here -- and look at this field of yams! You and the Lord sure did a wonderful job in raising these -- and look at the loaded vines of nice fat cucumbers -- cucumbers everywhere! The Lord and you have raised the most bountiful crops in the valley." The farmer, getting a little tired of receiving only half the credit for all his hard work, snapped his suspenders and finally said, "Pastor, you know, I sure wish you had seen this place when the Lord was taking care of it by himself!"

Art Warner developed his high precision 5 MHz and 2.5 MHz fifth overtone crystal units under an Air Force development contract while working in Sykes' department. This basic design, when properly processed under the most careful conditions, is today still the world's most stable frequency device known, short of Cesium beam or Rubidium standards.

The Bliley Electric Company has been in continuous production of this unit since 1955. Upgrading of vacuum techniques, closure methods, high temperature bakeout, etc, have further improved the stability and retrace characteristics, but the basic design <u>still</u> remains the same.

I will never forget an incident (which I could never repeat till Roger retired) when the Military had called a meeting with the Bell Laboratories engineers and invited members of the Industry, at which the design and processing information of this unit was revealed; the object being to establish an industry capability for the production of this highly stable unit.

Roger spoke eloquently about the design, the great care, and tedious steps that had to be followed in the cleaning procedures prior to encapsulating. Then he called on Art Warner to elaborate further on the cleaning process. Art, with his perpetual grin, said, "oh, I just scrub well with detergent and water and then rinse in methyl alcohol." We were all a bit surprised by this short remark, but understood. Afterwards, Roger told me quietly, "Johnny, if he wasn't such a 'damn good' engineer I'd fire him for that!"

Roger was a "bug" and a "spark plug" on standardization; within his own company, through EIA, for the Industry, and internationally through the IEC, and devoted great time and effort to this cause. He could draw sparks and motivate people to get things done, never shirking hiw own share of the responsibility.

I never heard much about Roger's extra-curricular activities during his travel to other lands, but you know "all work and no play makes Jack a dull boy." Judging from his

accomplishments at those IEC meetings, Roger was not a dull boy.

Incidentally, Industry can pay a tribute to Mr. Sykes by much deeper involvement in the IEC proceedings in the development of International Crystal Standards. Roger was chief U.S. Technical Expert to TC-49 Committee since 1954 and has continued to this day. He needs our wholehearted support.

For his retirement, Roger had a beautiful chalet built on the banks of Lake Sunapee in New Hampshire replete with a good sized workshop. I hope he and his wife Mary will enjoy many hours of fine fishing there, and enjoy the well earned retirement to the utmost.

And so, Roger Sykes, the Engineer, the Inventor, Author, Innovator, Coordinator, Moderator, a great American, and a wonderful person to know, we all salute you.

SURFACE WAVES AND DEVICES

By

H. J. Shaw
Stanford University

The past several years have seen a very strong increase in activity in acoustic surface waves with important contributions to new UHF and microwave devices coming from a large number of laboratories. Propogation characteristics of these waves on solid surfaces have been studied, an example of particular interest at this time being the case of surface waves on strongly piezoelectric lithium crystals. Interdigital transducers for efficient excitation of surface waves on these crystals have been developed and demonstrated in prototype low loss delay lines. Electroacoustic amplifiers in which gain results from the interaction between surface waves on these delay lines and drifting charge carriers in a separate but adjacent semiconductor sample have been demonstrated, which can operate cw and provide stable amplification of surface waves over wide frequency ranges. Frequency and time delay filters using interdigital electrode arrays provide a new degree of flexibility for signal processing and represent an area of applications which is growing rapidly. A new approach to signal processing functions is possible using parametric interactions between surface waves, which can provide signal correlation against an electrically variable reference signal. Looking into the future, attempts are being made to use surface waves to perform many of the functions usually performed by electrical waves in the form of integrated acoustic circuits of very small dimensions. In this talk the basic properties of surface waves on crystal line materials will be discussed and general areas of applications will be considered.

THICKNESS-TWIST VIBRATIONS OF A QUARTZ STRIP

R. D. Mindlin
Department of Civil Engineering, Columbia University
New York, N. Y.

Summary

An exact solution of the three-dimensional equations of elasticity has been found for the thickness-twist modes of vibration of a rotated-Y-cut quartz plate with a pair of parallel, free edges. The solution has a simple form made possible by relaxing the condition that the edge-planes be perpendicular to the main faces of the plate. For the AT-cut, the edges are off perpendicular by about five degrees.

Introduction

In a previous paper[1], the solution was given for the thickness-twist modes of vibration of infinite, rotated-Y-cut, quartz crystal plates. In the present paper[2], the solution is recast in a different form and extended to accommodate similar plates of finite width.

A rotated-Y-cut plate is one whose middle plane contains a digonal axis of elastic symmetry of the quartz crystal and whose normal makes an angle (35° 15' for the widely used AT-cut) with the trigonal axis[3]. In an isotropic plate, the thickness-twist modes of vibration are those in which the displacement direction is parallel to the middle plane of the plate and the wave normal is perpendicular to the displacement. In infinite, rotated-Y-cut quartz plates, analogous modes are possible[1] if the displacement direction is parallel to the digonal axis in the plane of the plate. Again in the isotropic plate, simple reflections of thickness-twist waves occur at free, plane boundaries of the plate parallel to the displacement direction and perpendicular to the middle plane. The same is not true in the quartz plate: with the result that the thickness-twist vibrations of a quartz strip, with a rectangular cross-section perpendicular to the digonal axis, are not expressible in terms of a finite number of elementary functions. The mode shape, if it were found, would comprise an infinite number of elementary shapes; and those with the longer wave lengths would be designated as "unwanted modes" or "spurious modes" which cause difficulties in crystal filter applications in electric circuits.

In the present paper, it is shown that, if the free, plane boundaries of the strip are maintained parallel to the digonal axis but cut slightly off normal to the middle plane (about 5° off normal for the AT-cut) simple reflections occur, a simple, closed solution is obtained and modes with simple shapes and well spaced frequences result.

Solution

If the x-axis is a digonal axis and the z-axis is normal to the middle plane of the plate, the stress-strain relations for the rotated-Y-cuts of quartz are[3]

$$\begin{aligned} T_1 &= c_{11}S_1 + c_{12}S_2 + c_{13}S_3 + c_{14}S_4 \\ T_2 &= c_{21}S_1 + c_{22}S_2 + c_{23}S_3 + c_{24}S_4 \\ T_3 &= c_{31}S_1 + c_{32}S_2 + c_{33}S_3 + c_{34}S_4 \\ T_4 &= c_{41}S_1 + c_{42}S_2 + c_{43}S_3 + c_{44}S_4 \\ T_5 &= c_{55}S_5 + c_{56}S_6 \\ T_6 &= c_{65}S_5 + c_{66}S_6 \end{aligned} \tag{1}$$

where $c_{ij} = c_{ji}$ and the strains, S_i, are related to the displacements, u, v, w, according to

$$\begin{aligned} S_1 &= \frac{\partial u}{\partial x}, & S_4 &= \frac{\partial w}{\partial y} + \frac{\partial v}{\partial z}, \\ S_2 &= \frac{\partial v}{\partial y}, & S_5 &= \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}, \\ S_3 &= \frac{\partial w}{\partial z} & S_6 &= \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}. \end{aligned} \tag{2}$$

The stresses must satisfy the equations of motion:

$$\begin{aligned} \frac{\partial T_1}{\partial x} + \frac{\partial T_6}{\partial y} + \frac{\partial T_5}{\partial z} &= \rho\frac{\partial^2 u}{\partial t^2}, \\ \frac{\partial T_6}{\partial x} + \frac{\partial T_2}{\partial y} + \frac{\partial T_4}{\partial z} &= \rho\frac{\partial^2 v}{\partial t^2}, \\ \frac{\partial T_5}{\partial x} + \frac{\partial T_4}{\partial y} + \frac{\partial T_3}{\partial z} &= \rho\frac{\partial^2 w}{\partial t^2}. \end{aligned} \tag{3}$$

We require a steady state solution, of these equations, of the form

$$u = U(y,z)e^{i\omega t}, \quad v = w = 0, \tag{4}$$

for a strip of finite thickness in the y-direction and of finite width in the z-direction--with all four boundaries free of traction.

Substitution of (4) in (1)-(3) reduces the latter to

$$T_1 = T_2 = T_3 = T_4 = 0,$$
$$T_5 = c_{55}\frac{\partial U}{\partial z} + c_{56}\frac{\partial U}{\partial y}, \qquad (5)$$
$$T_6 = c_{65}\frac{\partial U}{\partial z} + c_{66}\frac{\partial U}{\partial y};$$

$$S_1 = S_2 = S_3 = S_4 = 0, \qquad (6)$$
$$S_5 = \frac{\partial U}{\partial z},\ S_6 = \frac{\partial U}{\partial y};$$

$$c_{66}\frac{\partial^2 U}{\partial y^2} + 2c_{56}\frac{\partial^2 U}{\partial y \partial z} + c_{55}\frac{\partial^2 U}{\partial z^2} = -\rho\omega^2 U; \qquad (7)$$

with the exponential factor omitted.

Now, the four functions[2]

$$U = A \sin\eta y \cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) + B \sin\eta y \sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \qquad (8)$$

$$+C \cos\eta y \cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) + D \cos\eta y \sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right)$$

are solutions of (7) if

$$\rho\omega^2 = c_{66}\eta^2 + \gamma_{55}\zeta^2, \qquad (9)$$

where

$$\gamma_{55} = c_{55} - c_{56}^2/c_{66}. \qquad (10)$$

For a plate of finite thickness, say $2h$, we require

$$T_2\big]_{y=\pm h} = T_4\big]_{y=\pm h} = T_6\big]_{y=\pm h} = 0; \qquad (11)$$

and these boundary conditions are satisfied by (8) if

$$2\eta h = m\pi, \qquad (12)$$

where m is an odd integer for solutions A and B and an even integer (including zero--which corresponds to a face-shear mode) for solutions and D.

For traction-free boundaries at $z=\pm c$, we should require

$$T_3\big]_{z=\pm c} = T_4\big]_{z=\pm c} = T_5\big]_{z=\pm c} = 0. \qquad (13)$$

The first two of these are satisfied identically, but the third cannot be satisfied by functions as simple as (8). An infinite series of such functions would be required. However, consider a rotation of coordinate axes, through an angle α about the x-axis, according to the scheme of direction consines

	x	y	z
x'	1	0	0
y'	0	$\cos\alpha$	$\sin\alpha$
z'	0	$-\sin\alpha$	$\cos\alpha$

(14)

with (5), the components of stress, referred to the rotated axes, are given in terms of the original components by

$$T_1' = T_2' = T_3' = T_4' = 0,$$
$$T_5' = T_5\cos\alpha - T_6\sin\alpha, \qquad (15)$$
$$T_6' = T_5\sin\alpha + T_6\cos\alpha;$$

and, for traction-free boundaries at $z'=\pm c\cos\alpha$, as shown in Fig. 1, we require

$$T_3'\big]_{z'=\pm c\cos\alpha} = T_4'\big]_{z'=\pm c\cos\alpha} = T_5'\big]_{z'=\pm c\cos\alpha} = 0 \qquad (16)$$

Now, T_3' and T_4' are identically zero whereas

$$\begin{aligned} T_5' = {} & A\eta(c_{56}\cos\alpha - c_{66}\sin\alpha)\cos\eta y\cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & + A\zeta\gamma_{55}\cos\alpha\sin\eta y\sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & + B\eta(c_{56}\cos\alpha - c_{66}\sin\alpha)\cos\eta y\sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & - B\zeta\gamma_{55}\cos\alpha\sin\eta y\cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & - C\eta(c_{56}\cos\alpha - c_{66}\sin\alpha)\sin\eta y\cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & + C\zeta\gamma_{55}\cos\alpha\cos\eta y\sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & - D\eta(c_{56}\cos\alpha - c_{66}\sin\alpha)\sin\eta y\sin\zeta\left(\frac{c_{56}}{c_{66}} y - z\right) \\ & - D\zeta\gamma_{55}\cos\alpha\cos\eta y\cos\zeta\left(\frac{c_{56}}{c_{66}} y - z\right). \end{aligned} \qquad (17)$$

In (17), set

$$\alpha = \arctan(c_{56}/c_{66}) \qquad (18)$$

and note that $(c_{56}/c_{66})y-z=-z'/\cos\alpha$. Then (17) reduces to

$$T_5' = A\zeta\gamma_{55}\cos\alpha\sin\eta y\sin(\zeta z'/\cos\alpha)$$
$$-B\zeta\gamma_{55}\cos\alpha\sin\eta y\cos(\zeta z'/\cos\alpha)$$
$$+C\zeta\gamma_{55}\cos\alpha\cos\eta y\sin(\zeta z'/\cos\alpha) \qquad (19)$$
$$-D\zeta\gamma_{55}\cos\alpha\cos\eta y\cos(\zeta z'/\cos\alpha).$$

Hence, it is apparent that $T_5'=0$ on $z'=\pm c\cos\alpha$ if

$$2\zeta c = n\pi, \qquad (20)$$

where n is an even integer (including zero--which gives simple thickness-shear modes) for solutions A and C, and an odd integer for solutions B and D.

Upon substituting (12) and (20) in (8) and (9), we find the required simple solutions:

$$U = A\sin\frac{m\pi y}{2h}\cos\frac{n\pi}{2c}\left(\frac{c_{56}}{c_{66}}y-z\right) +$$
$$B\sin\frac{m\pi y}{2h}\sin\frac{n\pi}{2c}\left(\frac{c_{56}}{c_{66}}y-z\right) + C\cos\frac{m\pi y}{2h}\times$$
$$\cos\frac{n\pi}{2c}\left(\frac{c_{56}}{c_{66}}y-z\right) + D\cos\frac{m\pi y}{2h}\sin\frac{n\pi}{2c}\left(\frac{c_{56}}{c_{66}}y-z\right). \qquad (21)$$

with circular frequencies

$$\omega = \frac{m\pi}{2h}\left(\frac{c_{66}}{\rho}\right)^{\frac{1}{2}}\left(1+\frac{\gamma_{55}n^2h^2}{c_{66}m^2c^2}\right)^{\frac{1}{2}} \qquad (22)$$

and m and n odd or even integers, for solutions A, B, C, D, according to

A : m odd, n even

B : m odd, n odd

C : m even, n even

D : m even, n odd.

For the AT-cut of quartz $c_{56}=2.53$ and $c_{66}=29.01$, in units of 10^{10} dyn/cm^2, as calculated from Bechmann's values[4] of the principal constants of quartz. Hence, by (18), α is approximately 5° for the AT-cut. Thus, only a slight inclination of the edge cuts is required to reduce the undesirable, complicated mode-shapes to simple ones.

References

1. R. D. Mindlin, Int. J. Solids Structures 1, 141-145 (1965).

2. R. D. Mindlin, J. Acoust. Soc. Am. 41, 969-973 (1967).

3. W. P. Mason, Piezoelectric crystals and their application to ultrasonics, D. Van Nostrand Company, New York (1950).

4. R. Bechmann, Phys. Rev. 110, 1060-1061 (1958).

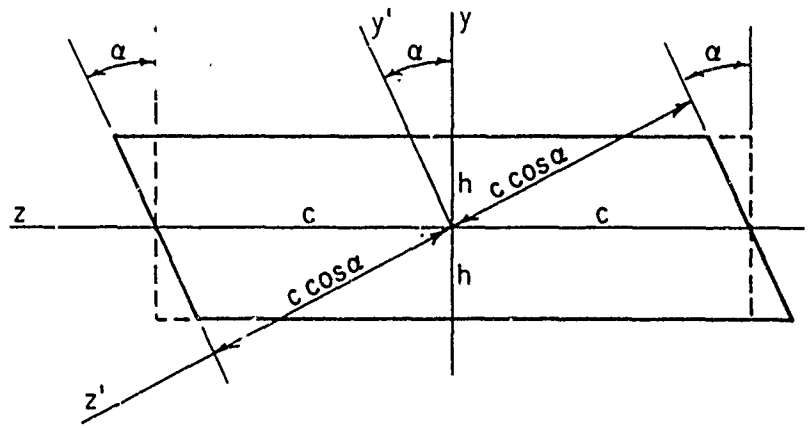


Fig 1. Rotated axes and cross-section of quartz strip.

ZERO TEMPERATURE COEFFICIENT ULTRASONIC DELAY LINES UTILIZING SYNTHETIC QUARTZ CRYSTALS AS DELAY MEDIA

Morio ONOE and Yuzo MOCHIZUKI
Institute of Industrial Science
University of Tokyo
Roppongi, Tokyo, JAPAN

Summary

A study of propagation and reflection of ultrasonic waves in a quartz crystal has made it possible to design multiple reflection ultrasonic delay lines having zero temperature coefficient of delay. Another merits of quartz crystals as delay media are in much higher stability and lower propagation loss in comparison with conventional zero temperature coefficient glass. Since the advent of synthetic quartz crystals grown by hydrothermal methods, raw materials become available in large sizes and at low prices. Hence the use of synthetic quartz crystals as delay media seems promising.

All the delay lines considered in the present paper consist of reflection surfaces which are parallel to the X-axis of quartz. An ultrasonic shear wave with its polarization direction parallel to the X-axis, which is designated SX-wave, can be reflected from any reflection surface without a mode conversion.

Two approaches have been taken in order to realize a zero temperature coefficient of delay at room temperature.

In the first approach, a rectangular shape similar to a line constructed by Merkulov and Yakovlev is used. Two reflection surfaces are perpendicular to the Z-axis. The other two surfaces are perpendicular to the Y-axis and the input and the output shear wave transducers are placed on a part of each surface. A wave left from the input transducer reaches to the output transducer after a number of reflections. An examination of temperature characteristics of each partial path shows that both positive and negative temperature coefficients of delay exist. And it has been found that an cancellation of positive and negative effects can be achieved at certain discrete ratios of the length along the Y-axis to the length along the Z-axis and hence a zero temperature coefficient of total delay can be obtained.

The second approach is based on the fact that there exist two SX-waves with very good temperature characteristics. These are similar to the waves used in AT and BT cut resonators and are designated the AT wave and BT wave, respectively. It has been found that surfaces inclined either -86°5' or -20° from the XY plane reflect the incident AT wave to the BT wave and vice versa. Hence a delay line consisting of only these two kinds of reflection surfaces can have zero temperature coefficient of delay.

Several delay lines have been made and their measured temperature characteristics show a zero temperature coefficient at room temperature as predicted.

1. Introduction

Ultrasonic delay lines have been used in various applications in signal and data processing. The recent trends in requirements for ultrasonic delay lines are a higher operating frequency and a longer delay. This yields also a larger information storage capacity, since the capacity is proportional to the product of the operating frequency and the delay time. A good use of a large storage capacity is only achieved, when the temperature variation of delay time is kept to a minimum. This is because the allowable relative variation of delay time is inversely proportional to the storage capacity. The larger the capacity, the tighter the allowance.

A delay line consists of the input and the output transducers and a delay medium. There are a wide variety of bonded and thin film transducers which are capable of operating at frequencies above say 100 MHz. Whereas the choise of delay media which exhibit low ultrasonic loss at high frequencies and good temperature stability of delay is rather limited.

Fused quartz, which has been one of the most widely used delay media, has a low loss even at a very high frequency. But its large temperature coefficient of delay of -80 PPM requires the use of temperature controlled oven in most applications.

Sapphire is another low loss material which is used at microwave frequencies. Its disadvantages are again the temperature coefficient of 20 PPM and the severe limitation of available size.

If an emphasis is put on temperature stability, there are various isopaustic alloys and glasses which exhibit zero temperature coefficient of delay at room temperature. Their losses, however, increase rapidly with frequency. Hence their use is limited to frequencies below 10 MHz except for a short delay. Their temperature characteristics are sensitive to the composition and the heat treatment, which makes the fabrication difficult. Sometimes a

large aging of delay time exists.

In the present paper, the use of single crystal quartz as delay medium is proposed. It is well known that quartz is physically and chemically stable and has very low loss at high frequencies. A study of propagation and reflection of ultrasonic waves in a quartz crystal has made it possible to design multiple reflection type ultrasonic delay lines having zero temperature coefficient of delay. Since the advent of synthetic quartz crystals grown by hydrothermal methods, raw materials become available in large sizes and at low prices. Hence the use of synthetic quartz crystals as delay media seems promising. Since the loss of quartz has been extensively studied, the emphasis of the present paper is put on temperature characteristics.

The propagation of ultrasonic waves in an anisotropic crystal differs considerably from the propagation in an isotropic material. The velocity of the propagation is determined by the orientation of the propagation direction with respect to the crystallographic axes. Only certain directions yield pure longitudinal or shear modes of propagation. In general, a phase front normal is not collinear with the direction of energy flow. On reflection from a surface plane, the equality of incidence and reflection is not maintained and there may be a mode conversion. All these points must be taken into account in the design of delay lines.

2. Use of SX waves

Mode conversion on reflection should be avoided, because it increases the insertion loss and, more important, the amount of spurious signals. Hence shear waves designated SX waves are used. The polarization of SX waves is parallel to the X-axis and the direction of propagation is within the YZ-plane. It can reflect from surfaces parallel to the X-axis without mode conversion.

Temperature coefficient of delay of SX waves varies with the direction of propagation and becomes zero at two angles corresponding to AT and BT cuts. These waves are designated as AT wave and BT wave, respectively.

A straight path delay line having zero temperature coefficient of delay is easily obtained by the use of AT or BT wave with due regard to the direction of energy flow. Hence the present paper is limited to the multiple reflection type, which can yield a long delay in a small size of crystal. The existence of reflection surface also provides a means of precise adjustment of delay time without removing transducers.

The dimension along the X-axis is not critical if it is sufficiently long. Hence characteristics of a delay line is determined by the contour in the YZ-plane.

Two approaches have been taken in order to realize a zero temperature coefficient of delay. One is based on the overall cancellation of positive and negative temperature variation of delay associated with partial paths. Another is based on the fact that there exist two kinds of surfaces which reflect the incident AT wave to the BT wave and vice versa. Hence a delay line consisting of only these two kinds of reflection surfaces can have zero temperature coefficient of delay.

3. Analysis of SX waves

3.1 Propagation of SX waves

Since electromechanical coupling in quartz is small, a purely elastic analysis suffices for the present purpose. The relationship between stress(T_i) and strain (S_i) refered to crystallographic axes of quartz is as follows:

$$\begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \\ T_6 \end{pmatrix} = \begin{pmatrix} c_{11}, c_{12}, c_{13}, c_{14}, & & \\ c_{12}, c_{11}, c_{13}, -c_{14}, & & \\ c_{13}, c_{13}, c_{33}, & & \\ c_{14}, -c_{14}, & c_{44}, & \\ & & c_{44}, c_{14} \\ & & c_{14}, c_{66} \end{pmatrix} \begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ S_5 \\ S_6 \end{pmatrix} \quad (1)$$

where C_{ij} are stiffness constants. The displacement (u_i) of a SX wave can be written by the following form:

$$u_1 = u_0 \exp j\{k(my + nz) - \omega t\} \quad (2)$$

$$u_2 = u_3 = 0$$

where (0, m, n) is the direction cosine of the phase front normal, and k, ω, u_0 are the wave number, the angular frequency and a constant, respectively. The strain is:

$$S_1 = S_2 = S_3 = S_4 = 0$$

$$S_5 = \frac{\partial u_1}{\partial z} \quad (3)$$

$$S_6 = \frac{\partial u_1}{\partial y}$$

This is substituted in (1) which is then substituted into the equation of motion to obtain the following equation:

$$\rho\omega^2 = k^2(c_{66}m^2 + c_{44}n^2 + 2c_{14}mn) \quad (4)$$

Hence the phase velocity v_p is given by:

$$v_p = \frac{\omega}{k} = \sqrt{\frac{c_{66}m^2 + c_{44}n^2 + 2c_{14}mn}{\rho}} \quad (5)$$

Fig. 1 shows the phase velocity as a function of the direction angle θ, which is measured counterclockwise from the Y-axis.

$$\theta = \tan^{-1}\frac{n}{m} \quad (6)$$

Whereas the direction consine of energy flow is given by

$$0: u_1T_6 : u_1T_5$$
$$= 0 : (c_{14}n + c_{66}m) : (c_{44}n + c_{14}m) \qquad (7)$$

Hence the direction angle ϕ is ;

$$\phi = \tan^{-1}\frac{c_{44}n+c_{14}m}{c_{14}n+c_{66}m} \qquad (8)$$

The difference between two angles, θ and ϕ , is given by the following equation.

$$\phi - \theta = \tan^{-1}\frac{mn(c_{44}-c_{66})+c_{14}(m^2-n^2)}{c_{66}m^2+c_{44}n^2+2c_{14}mn} \qquad (9)$$

Fig. 2 shows the difference as a function of the angle θ.

3.2 Reflection of SX waves

Fig. 3 shows a traction free reflection surface with the angle δ. The incident wave and the reflected wave can be expressed by the following equations, respectively.

$$u_1 = u_0 \exp jk\,(my + nz)$$
$$u_1' = u_0' \exp jk'(m'y + n'z) \qquad (10)$$

The time factor is omitted.
The traction free boundary condition at the surface is:

$$\cos\delta \cdot T_5 - \sin\delta \cdot T_6 = 0 \qquad (11)$$

The substitution of (10) into (11) yields the following two equations

$$u_0 = u_0' \qquad (12)$$

and

$$\frac{c_{66}m^2+c_{44}n^2+2c_{14}mn}{(m\cos\delta+n\sin\delta)^2} = \frac{c_{66}m'^2+c_{44}n'^2+2c_{14}m'n'}{(m'\cos\delta+n'\sin\delta)^2} \qquad (13)$$

The second equation which corresponds to Snell's law may be expressed as follows:

$$\frac{c_{66}+2Bc_{14}+B^2c_{44}}{(1+AB)^2} = \frac{c_{66}+2Cc_{14}+C^2c_{44}}{(1+AC)^2} \qquad (14)$$

where A, B, C are tangent of direction angles.

$$A = \tan\delta$$
$$B = \frac{n}{m} = \tan\theta \qquad (15)$$
$$C = \frac{n'}{m'} = \tan\theta'$$

When the surface and the incident wave are given, the reflected wave is obtained by the following equation,

$$C = \frac{-Bc_{44}+2Ac_{66}+A^2Bc_{66}-2c_{14}}{c_{44}+2ABc_{44}-A^2c_{66}-2A^2Bc_{14}} \qquad (16)$$

Whereas, when the incident and the reflected waves are given, there are usually two surfaces, because Eq. (14) is quadratic in A.

Fig. 4 shows the relationship between incident and reflection angles. The parameter is the angle δ of the reflection surface.

3.3 Temperature characteristics of SX waves

Two factors affect the temperature variation of delay time. One is the variation of propagation velocity and the other is the thermal expansion. The latter is taken into account by changing the contour in the YZ plane in the following manner. A corner point (x, y, z) at the standard temperature T_0 moves to the point ($\ell_X x$, $\ell_Y y$, $\ell_Z z$) at another temperature T, where lx, ly and lz are the length at temperature T of unit vectors along the X the Y and the Z axies, respectively, at the standard temperature.

The total delay time at a certain temperature T is obtained as follows.

(1) Draw the contour.

(2) Trace the path of energy flow from the input transducer to the output transducer and determine reflection points.

(3) Calculate the length, L_i, of the i-th partial path.

(4) Then the delay time, D_i, of the i-th partial path is given by:

$$D_i = \frac{L_i \cdot \cos(\phi_i - \theta_i)}{v_{pi}} \qquad (17)$$

where θi, ϕi, v_{pi} are the directions of phase front normal and energy flow and the phase velocity, respectively, at temperature T.

(5) The total delay time is the sum of D_i.

Repetition of the steps (1), (2) and (3) at each temperature is troublesome and misses intuitive estimates of final characteristics. Hence the following approximation of D_i is used at intermediate stages.

$$D_i' = \frac{L_{i0} \cdot \cos(\phi_{i0} - \theta_{i0})}{v_{pi}'} \qquad (18)$$

where the subscript "0" means that values at the standard temperature are taken. v_{pi}' is the apparent phase velocity, which takes the effect of thermal expansion into account and only variable dependent on temperature in (18). Final temperature characteristics, however, are always calculated

taking the whole steps. Material constants given by Mason are used in all the calculation.[1]

Fig. 5 shows temperature variations of apparent phase velocity at 0 and 50°C with respect to the velocity at the standard temperature of 25°C. It can be seen that there is no variation for the AT and the BT waves.

4. Rectangular delay lines

Merkulov first studied multiple reflection type delay lines made of single quartz crystal.[2] His study was limited to the rectangular contour in the YZ-plane, of which sides are parallel to either the Y- or the Z- axes. No mention was made of temperature characteristics.

Fig. 6 and 7 shows the pattern of energy flow paths. The numbers of partial paths are seven and nine, respectively. The pattern in Fig. 6 was obtained by Merkulov. Input and output transducers are placed at opposite sides parallel to the Z-axis. The phase front at each transducer is also parallel to the Z-axis. Hence the reflected wave at the output traces back the same paths to the input and reflects back again, which yields the triple travel signal as in the case of conventional delay lines.

The pattern of paths is stable against small changes in the contour and the location of input transducer. Also the total delay time is constant regardless of the location of the input transducer, provided that the pattern is preserved.

It can be seen in Fig. 6 and 7 that there are only four kinds of the propagation direction. Table 1 summarizes characteristics of each propagation direction. The apparent phase velocity along the direction I has a large positive temperature coefficient. Whereas the velocity along the direction IV has a large negative coefficient. Hence it is possible to cancel out the temperature variation of the total delay time by adjusting the ratio of side lengths. According to an approximate calculation, zero temperature coefficient of the total delay is to be obtained at the following ratio.

$$\frac{Y_o}{Z_o} = 0.86 \qquad (19)$$

where Y_o and Z_o are the length along the Y and Z axis, respectively. The following total delay is obtained at this ratio.

$$D(\mu s) = 12.5\ Y_o\ (cm) \qquad (20)$$

Fig. 8 shows temperature characteristics calculated according to the whole steps described in 3.3. The curve is nearly parabolic and has a turn over temperature at room temperature. The location of turnover temperature can be shifted to any desired temperature by adjusting the Y_o/Z_o ratio as shown in Fig. 9.

Similary zero temperature coefficient is obtained for the pattern shown in Fig. 7 at the following ratio.

$$\frac{Y_o}{Z_o} = 1.425 \qquad (21)$$

The total delay is:

$$D(\mu s) = 12.8\ Y_o\ (cm) \qquad (22)$$

A comparison of (20) and (22) shows that little advantage is gained by using a more complex pattern.

When the input and the output transducers are placed on surfaces parallel to the Y-axis, two patterns having zero temperature coefficient are obtained as shown in Fig. 10 and 11. The pattern in Fig. 10 has seven partial paths and the Y_o/Z_o is 0.75. The pattern in Fig. 11 has nine partial paths and the Yo/Z_o is 0.44.

Nearly grazing angle of incidence to the vertical surface is not favorable to practical use, because spurious signals are likely to occur when a beam of finite width is used.

5. Polygonal delay lines

In the second approach for obtaining zero temperature coefficient, reflection surfaces has to be found which reflect the incident AT wave to the BT wave and vice versa. The variation of delay with temperature is cubic for the AT wave and parabolic for the BT wave. The shape and the location of the turnover temperature depend somewhat on the direction angle. In the present study, the direction angles are chosen as shown in Table 2. Properties are also included in the table. Substitution of the tangents of both waves into (14) yields the following two tangents for relfection surfaces.

$$A = -14.618 \quad (-86°5')$$
$$A = -0.36377 \quad (-20°) \qquad (23)$$

The design of a contour is to combine these two surfaces together with the input and the output surfaces which are parallel to the phase front of the AT or the BT wave. Fig. 12 and 13 show examples. Although the contour in Fig. 12 is pentagonal, only three surfaces are to be precisely finished. The contour in Fig. 13 is hexagonal and all the surfaces are to be precisely finished. Calculated temperature characteristics will be presented later with experimental results.

6. Experiments

6.1 Fabrication

Several delay lines as shown in Fig. 14 have been made to check the theory. The dimension of the rectangular line is 6 x 21 x 24 mm. The thickness along the X-axis of the pentagonal and the hexagonal lines is 10 mm. The length of the side 1-5 in Fig. 12 is 33.2 mm. The distance between the side 2-3 and the side 4-5 in Fig. 13 is 38.6 mm. Piezoelectric ceramic shear transducers are bonded to the input and the output surfaces. The poling axis is parallel to the X-axis in order to excite

SX waves. The resonant frequency is about 8 MHz, which is too low in comparison with expected operating frequencies of these delay lines. Low loss of quartz at high frequencies, however, has been well confirmed. Hence our present concern is to see whether or not zero temperature coefficient is realized.

6.2 Measurement

Since the delay is rather short and the frequency is low, the variation of delay with temperature is measured as phase change by the circuit shown in Fig. 15. The output of the frequency synthesizer is devided into two paths; one is a delay line in an oven and another is an attenuator. Outputs of both paths are added. The frequency and the attenuator setting are adjusted, so that a phase opposition cancellation is achieved. When temperature varies, a slight change of frequency is necessary to reestablish the cancellation. Then the specific variation of phase delay $\Delta D/D$ is very close to the specific frequency change $\Delta f/f$.

$$\frac{\Delta D}{D} \simeq -\frac{\Delta f}{f} \qquad (24)$$

A gate with a double pulse generator is used for the measurement of total delay time using pulsed RF signals. It is also sometimes used to suppress spurious signals in the phase delay measurement. Impedance of transducers and insertion losses of delay lines are measured in conventional manners.

6.3 Results obtained

Table 3 shows the comparison between calculated and measured delay time. The measurement is done by pulsed RF signals. The agreement is good.

Measured variations of delay of each line with temperature are shown in Fig. 16, 17 and 18. Solid lines are calculated values and dotted lines are measured values. Material constants used in the calculation are not good enough to expect an agreement in details over a wide range of temperature. With this consideration in mind, the agreement in Fig. 17 and 18 for the pentagonal and the hexagonal lines is fairly good and measured characteristics definitely exhibit zero temperature coefficient.

Whereas measured variation in Fig. 16 for the rectangular line exhibits a large positive temperature coefficient. This is probably due to that the accuracy of material constants and dimensions affects more severely in this line, of which partial paths have large temperature coefficients. In other lines partial paths itself have good temperature characteristics.

According to Fig. 9, the length along the Y-axis is shortened in order to shift up the location of turnover temperature. Fig. 19 shows measured characteristics at the Y_0/Z_0 ratio of 0.94. Zero temperature coefficient is obtained this time.

An example of insertion loss and spurious responses is shown in Fig. 20 for the pentagonal line. Fig. 21 is a picture of the input and the output waveforms in the pulsed RF operation. Relatively large spurious responses are due to the tripple travel signal and spurious reflections caused by the beam spread. The latter can be reduced by increasing the operating frequency. Fig. 22 shows the waveforms at 52 MHz, which corresponds to the seventh overtone operation of transducers. An improvement of spurious responses is apparent.

Spurious responses in rectangular and hexagonal lines are worse, because of more reflections at surfaces. Increase of operating frequencies is tested by overtone operations of transducers and found to be effective also in this case.

7. Conclusion

Theory and experiment of multiple reflection type ultrasonic delay lines made of synthetic quarts crystals are presented. Zero temperature coefficient of delay time at room temperature is realized by two different approaches: one is based on the overall cancellation of large positive and negative variation of delay with temperature and another is based on the combination of partial paths having good temperature characteristics.

If good high frequency transducers are introduced in these delay lines, merits of synthetic quartz crystals as delay media with low loss and high temperature stability will be fully utilized.

Acknowledgement

The authors gratefully acknowledge supports in materials and fabrications generouly given by Dr. H. Yoda of Toyo Communication Equipment Co., late Mr. E. Kura of Nippon Electric Co., and Mr. T. Shinada of Kinsekisha Laboratory.

Reference

(1) W. P. Mason, Bell Syst. Tech. J., vol. 30, 4, pp. 366-380, April 1951.

(2) L.G. Merkulov and L.A. Yokovlev, Soviet Physics - Acoustics, vol. 8, 2, pp. 152-155, Oct./Dec. 1962.

TABLE.I CHARACTERISTICS OF FOUR PROPAGATION DIRECTIONS APPEARED IN RECTANGULAR DELAY LINES

Direction	Direction of wave front normal (θ)	Deviation of energy beam ($\varphi-\theta$)	Phase velocity (v_p) m/s	Velocity of energy flow m/s	Temperature coefficient of apparent velocity $\times10^{-5}$/°C
	0°	-24°22'	3890	4270	8.8
II	31°41'	28'	3310	3310	1.6
III	-54°28'	-2°26'	5127	5132	0
IV	63°38'	24°24'	3918	4302	-8.3

TABLE.2 CHARACTERISTICS OF THE AT AND THE BT WAVES

	Direction of wave front normal (θ)	Phase velocity (v_p) m/s	Direction of energy flow (φ)
AT	35°15' B=0.7067	3320.5	40°40'
BT	-49°, C=-1.150	5093.0	-54°36'

TABLE.3
COMPARISON BETWEEN CALCULATED AND MEASURED TOTAL DELAY TIME

Delay lines	Calculated (μs)	Measured (μs)
Rectangular	26.3	26.4
Pentagonal	32.4	32.2
Hexagonal	64.2	64.5

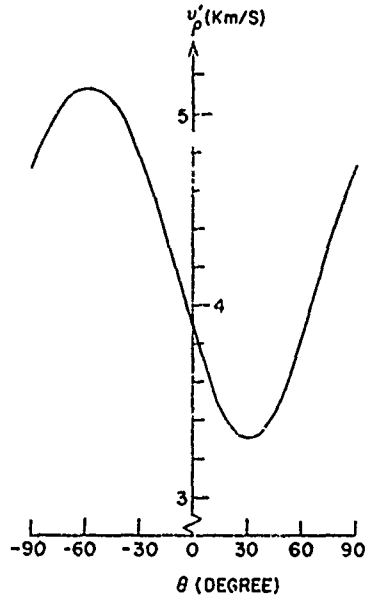


FIG.1. PHASE VELOCITY VERSUS DIRECTION ANGLE OF THE PHASE FRONT NORMAL.

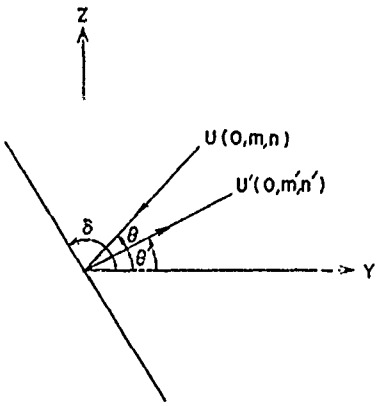


FIG.3 INCIDENT AND REFLECTES WAVES AT A FREE SURFACE.

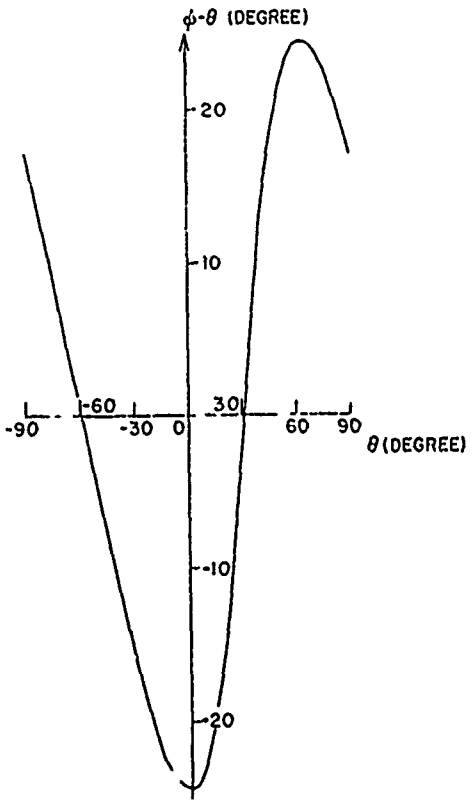


FIG.2 DIFFERENCE BETWEEN THE DIRECTION OF THE PHASE FRONT NORMAL AND THE DIRECTION OF THE ENERGY FLOW.

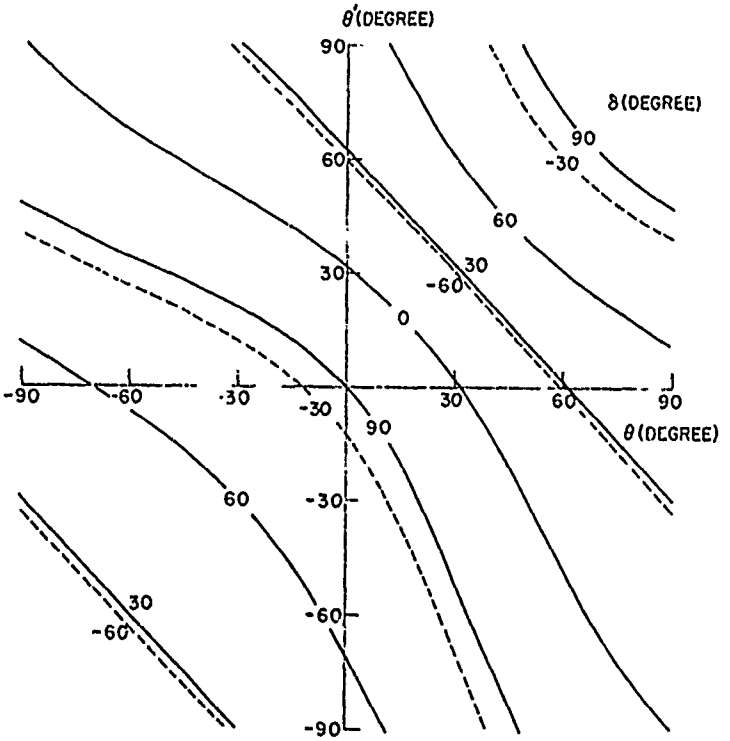


FIG.4
RELATIONSHIP BETWEEN INCIDENT AND REFLECTION ANGLES. THE PARAMETER IS THE ANGLE OF SURFACE.

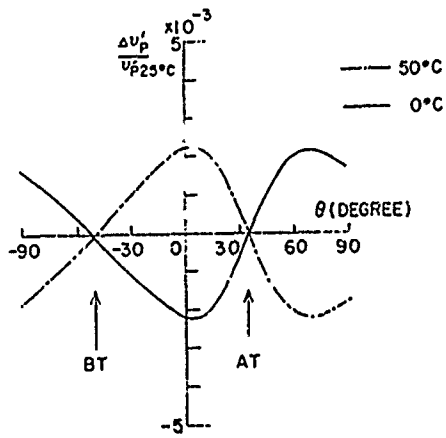


FIG.5 TEMPERATURE VARIATIONS OF APPARENT PHASE VELOCITY.

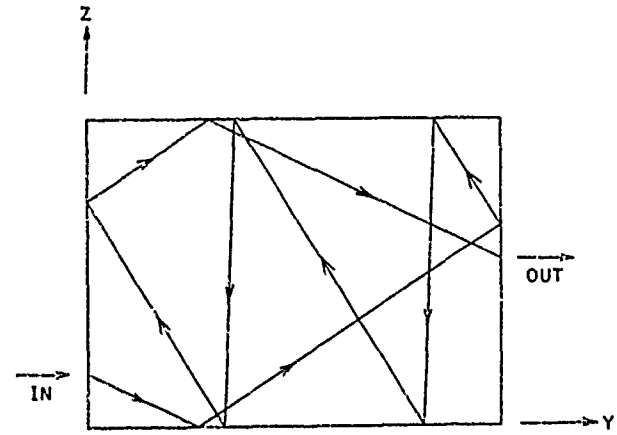


FIG.7
RECTANGULAR DELAY LINE WITH NINE PARTIAL PATHS.

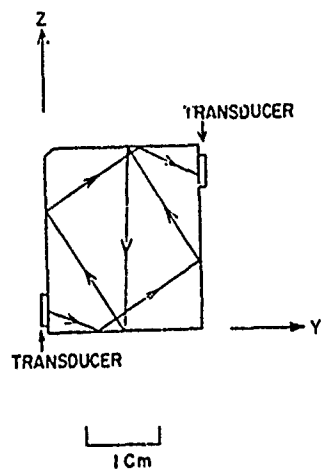


FIG.6
RECTANGULAR DELAY LINE WITH SEVEN PARTIAL PATHS.

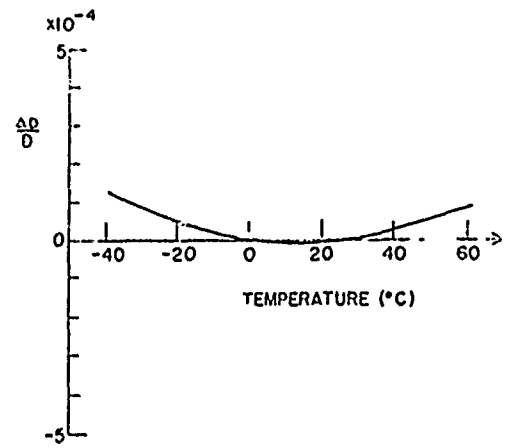


FIG.8 CALCULATED TEMPERATURE CHARACTERISTICS OF A RECTANGULAR DELAY LINE WITH THE YO/ZO RATIO OF 0.86.

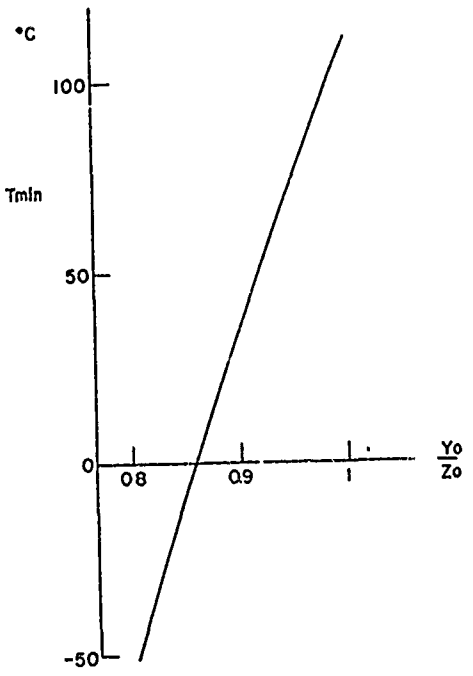


FIG. 9
RELATIONSHIP BETWEEN THE TURNOVER TEMPERATURE AND THE YO/ZO RATIO.

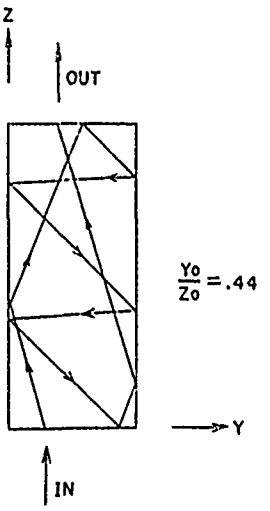


FIG. 11
RECTANGULAR DELAY LINE WITH NINE PARTIAL PATHS AND TRANSDUCERS ON THE Z SURFACES.

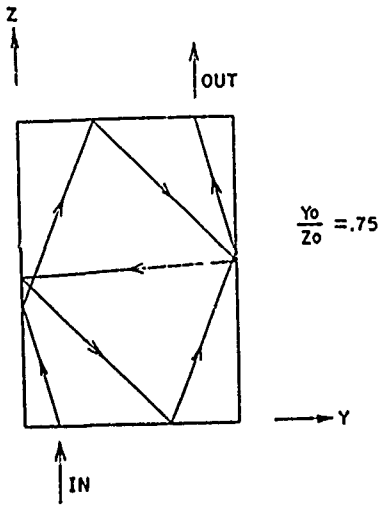


FIG. 10
RECTANGULAR DELAY LINE WITH SEVEN PARTIAL PATHS AND TRANSDUCERS ON THE Z SURFACES.

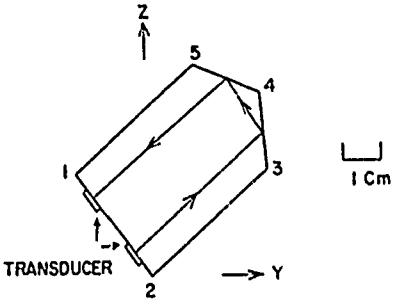


FIG. 12 PENTAGONAL DELAY LINE.

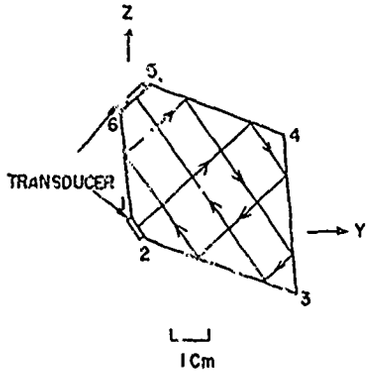


FIG.13 HEXAGONAL DELAY LINE.

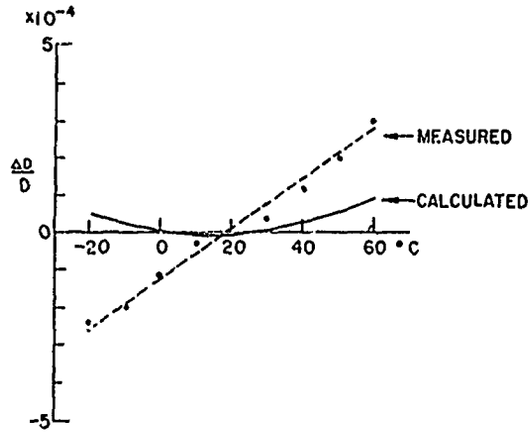


FIG.16 TEMPERATURE CHARACTERISTICS OF THE RECTANGULAR LINE.

Reproduced from best available copy.

FIGURE 14 - Pictures of synthetic quartz crystal delay lines

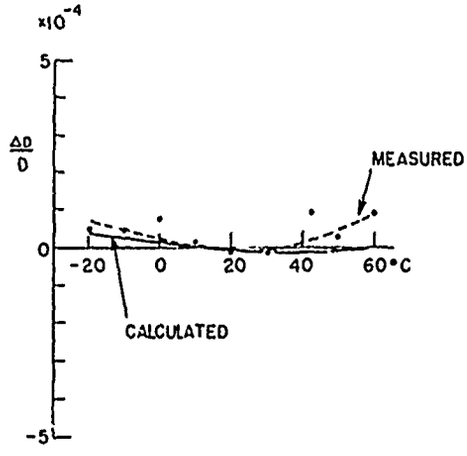


FIG.17 TEMPERATURE CHARACTERISTICS OF THE PENTAGONAL LINE.

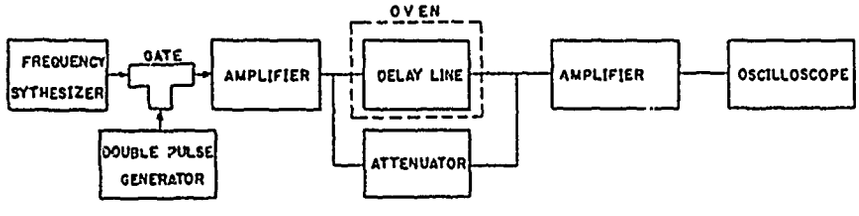


FIG.15 CIRCUIT FOR MEASURING TEMPERATURE VARIATION OF DELAY.

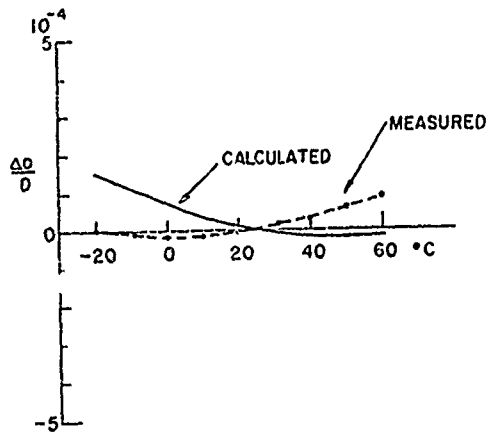


FIG.18 TEMPERATURE CHARACTERISTICS OF THE HEXAGONAL LINE.

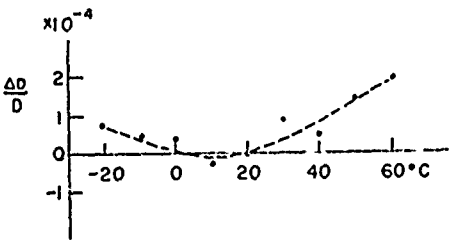


FIG.19
TEMPERATURE CHARACTERISTICS OF THE RECTANGULAR LINE AFTER THE Y LENGTH IS SHORTENED.

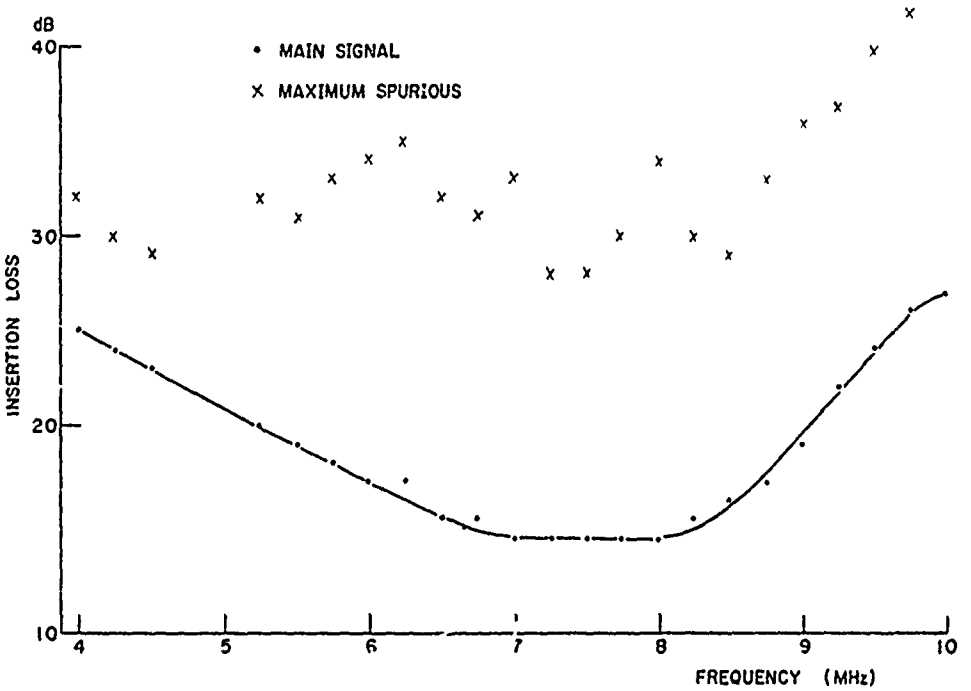


FIG.20 INSERTION LOSS AND SPURIOUS RESPONSES OF THE PENTAGONAL LINE.

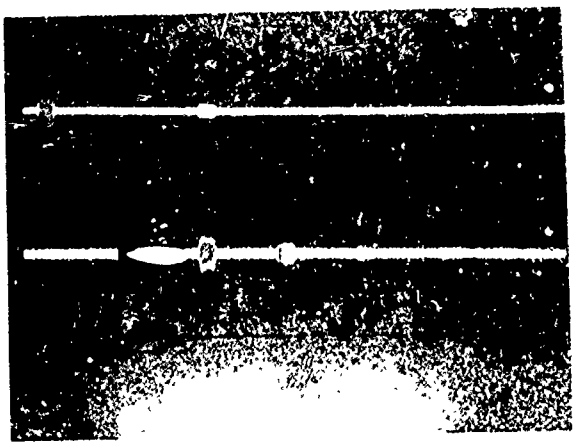


FIGURE 21 - Input and output waveforms of the pentaspiral circuit, low-RF operation

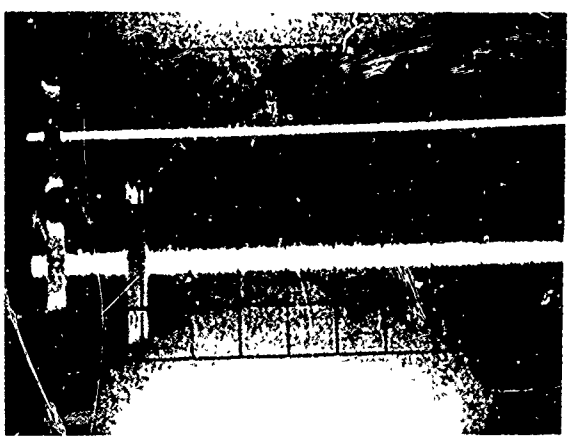


FIGURE [illegible] Input and output waveforms [illegible] several [illegible] operation

FREQUENCY-TEMPERATURE DEPENDENCE OF THICKNESS VIBRATIONS OF PIEZOELECTRIC PLATES

Karel Hruska
University of Khartoum
Khartoum, Sudan

Summary

In this paper anexpression is cbtained for the first-order temperature coefficient of frequency of piezoelectric plates vibrating in thickness between two plane parallel electrodes. The first-order temperature coefficient is given in terms of thermal expansion coefficients and temperature coefficients of elastic stiffnesses, piezoelectric stress constants and dielectric permittivities. It also includes, in a rigorous manner, the effect of the order of harmonic and electrode separation. The boundary conditions are duly respected and the consequent coupling among the three modes of vibrations is taken into account. The resulting expression is presented in a general form applicable to any type of piezoelectric material and plate of any orientation.

The derived expression for the first-order temperature coefficient of frequency was then applied to alpha-quartz. The agreement achieved between the theoretical and experimental results is considerably better than that obtained by other authors thus supporting the new values for the temperature coefficients of elastic stiffnesses and compliances calculated simultaneously. Using the present expression for AT and BT cuts, several frequency-temperature-angle characteristics, effect of overtone and electrode separation were calculated. A comparison between the theoretical and several experimental results yields a good agreement.

The experimental values of the first-order temperature coefficient of frequency on which the application of the derived expression to alpha-quartz plates is based have been known with two to three figures accuracy reaching the order of $10^{-7}(^{o}C)^{-1}$. As a result of this the interpretation was done exclusively in terms of the thermal expansion coefficients and the temperature coefficients of elastic stiffnesses at constant electric field. The contribution of the temperature coefficients of piezoelectric stress constants and dielectric permittivities was found negligible. The presence of these quantities in the general expression for the first-order temperature coefficient of frequency would, however, be appreciated if the existing measurment accuracy were increased or if the general expression were applied to some other piezoelectric material possessing for example a stronger piezoelectric effect that alpha-quartz.

Introduction

The subject of this paper is to give a new interpretation to the temperature dependence of the eigenfrequency of plane parallel piezoelectric plates vibrating in thickness between two plane electrodes parallel with the main plate faces. The frequency-temperature behavior of piezoelectric plates made from a certain crystalline material is knownto be dependent on the direction of the plate thickness, mode of vibration and its order of harmonic, electrode separation, the ratio of the main face dimension to the plate thickness and also, it is maintained, on the contour of the plates. The frequency-temperature dependence was treated theoretically several times in the past while some concessions, be it negligence to the effect of overtone, electrode separation or boundary conditions, had always been made to the requirements of the theory in order that the results might be presented in a relatively simple form.

The aim of this paper is to obtain an expression for the first-order temperature coefficient of frequency including explicitly the possible maximum of the parametres on which it is known to depend. That is why the approach developed by TIERSTEN[1] has been chosen as the starting point for the present work. This ensures, among others positive features, that the boundary conditions at the traction-free surfaces of the plates resulting theoretically in coupling among the three existing modes of the plate vibrations are properly observed.

Basic Relations

A thin rectangular plate of thickness t, width w and length l is surrounded by two plane electrodes equal in dimensions and parallel to the large faces of the crystal plate as shown in Fig.1. The separation of the electrodes is d, where $t,d \ll w,l$. The thickness,

length and width of the plate are parallel respectively to the axes X_1', X_2' and X_3' of an auxiliary orthogonal set of reference axes, while another, basic orthogonal set of reference axes X_1, X_2 and X_3 is fixedly defined in relation to the crystal structure of the plate.[2] The orientation of the plate with respect to the basic set of axes is arbitrary and the two sets of reference axes are related as usual by a matrix of direction cosines

$$\begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ X_1' & \alpha_1 & \alpha_2 & \alpha_3 \\ X_2' & \beta_1 & \beta_2 & \beta_3 \\ X_3' & \gamma_1 & \gamma_2 & \gamma_3 \end{array} \qquad (1)$$

In the anisotropic material of the above plate the following equations of state are valid

$$T_{ij} = c_{ijkl} S_{kl} - e_{nij} E_n \,, \qquad (2)$$

$$D_m = e_{mkl} S_{kl} + \epsilon_{mn} E_n \,, \qquad (3)$$

where further

$$S_{kl} = \tfrac{1}{2} \left(\frac{\partial u_k}{\partial x_l} + \frac{\partial u_l}{\partial x_k} \right) . \qquad (4)$$

Here $\underline{T}$ means the stress tensor, $\underline{S}$ the strain tensor, $\underline{E}$ the electric field intensity, $\underline{D}$ the electric displacement, $\underline{u}$ the displacement vector, $\underline{c}$ the elastic stiffness tensor, e the piezoelectric stress tensor, ϵ the permittivity tensor; the indices i,j,k,l,m,n assume values 1,2,3.

Moreover, we have the fundamental elastic wave equation

$$\frac{\partial T_{ij}}{\partial x_j} = \varrho \frac{\partial^2 u_i}{\partial t^2} \,, \quad i = 1,2,3, \qquad (5)$$

where ϱ means the density and t time. From the electromagnetic theory we have

$$\frac{\partial D_m}{\partial x_m} = 0 \qquad (6)$$

since the space charge density in a dielectric is equal to zero; further we may write

$$\frac{\partial E_m}{\partial x_n} = \frac{\partial E_n}{\partial x_m} \,, \quad m,n=1,2,3 \,, \; m \neq n. \qquad (7)$$

Equation (6) instead of the generally valid curl $E = -\dot{B}$ is here usually considered as correct since it is known that putting $\dot{B} = 0$ has a negligible effect on both the velocity and the amplitude of the plane waves that are the solution of our problem.[3,4]

With reference to the arrangement of the piezoelectric plate, air gaps and plane electrodes in Fig.1., the boundary conditions are formulated in the following way

$$E_{tan}^{o\prime} = E_{tan}' \quad \text{for } x_1' = \pm\tfrac{1}{2}t \,, \qquad (8)$$

$$D_n^{o\prime} = D_n' \quad \text{for } x_1' = \pm\tfrac{1}{2}t \,, \qquad (9)$$

$$T_{1i}' = 0 \quad \text{for } x_1' = \pm\tfrac{1}{2}t \,, \; i=1,2,3 \,, \qquad (10)$$

$$\varphi = \varphi_o e^{-i\omega t} \quad \text{for } x_1' = \tfrac{1}{2}d \,, \qquad (11)$$

$$\varphi = -\varphi_o e^{-i\omega t} \quad \text{for } x_1' = -\tfrac{1}{2}d \,, \qquad (12)$$

Equations (8),(9) and (10) represent the continuity of the electric field component tangential to the plate surfaces, the continuity of the electric displacement component normal to the plate surfaces and the disappearance of the elastic stresses on the plate surfaces respectively. The indices with the individual quantities refer to the primed auxiliary set of coordinate axes with the axis X_1' along the thickness of the plate (Fig.1). The boundary conditions (11) and (12) are in agreement with the presence of a potential across the electrodes exciting the thickness vibrations of the plate.

The relations (2)-(12) form a form a complete set necessary and sufficient to solve the problem of thickness vibrations of the plate in Fig.1. The relations are analogical to those given by TIERSTEN[1] and also by BECHMANN[5] and LAWSON[6] in their earlier approximative solutions.

Strictly speaking the boundary conditions are valid for the case of an infinite plate vibrating between two infinite electrodes. Nevertheless for the problem of thickness vibrations they are regarded as suitable so long as $t,d \ll w,l$.

Following the procedure shown by TIERSTEN[1] in his treatment of thickness vibrations of piezoelectric plates similar relations were obtained here including, however, the case of non-zero air gap between the electrodes and the crystal plate. For our furhter purpose the results were related to the basic set of reference X_1, X_2, X_3 and rewritten into a form valid for plates of arbitrary orientation. For the resonance frequency of thickness vibrations the following transcendental equation was obtained

$$\epsilon_{abc} \epsilon_{def} Q_{ad} Q_{be} Q_{cf} = 0 \,, \qquad (13)$$

$a,b,c,d,e,f = 1,2,3$, where

$$Q_{ik} = A_k^{(2)} \left(\Gamma_{ik} B^{(2)} \operatorname{cotan} B^{(2)} - \frac{\epsilon_o t}{h} \frac{F_i F_k}{\epsilon} \right) \qquad (14)$$

and further

$$\Gamma_{ik} = \alpha_j \alpha_l \left(c_{ijkl} + \frac{\alpha_m e_{mij} \alpha_m e_{mkl}}{\varepsilon} \right) \tag{15}$$

$$B^{(z)} = f \pi t \sqrt{\frac{\varrho}{\Gamma^{(z)}}} \tag{16}$$

$$f_r = \varepsilon_0 t + \varepsilon (d - t) \tag{17}$$

$$F_{ik} = \alpha_i \alpha_j e_{ijk} \tag{18}$$

$$\varepsilon = \alpha_m \alpha_m \varepsilon_{mm} \tag{19}$$

while $\Gamma^{(z)}$ and $A_k^{(z)}$ can be determined from

$$\left| \Gamma_{ik} - \delta_{ik} \Gamma^{(z)} \right| = 0 \tag{20}$$

$$\left(\Gamma_{ik} - \delta_{ik} \Gamma^{(z)} \right) A_k^{(z)} = 0 \tag{21}$$

$$\delta_{ik} A_i^{(z)} A_k^{(z)} = 0 \tag{22}$$

The symbols ϵ_{abc} in (13) and δ_{ik} in (20), (21) and (22) mean the components of the Levi-Civita nad Kronecker tensor respectively. The symbol f appearing in (16) represents the eigenfrequency of the thickness vibrations of the plate, ε_0 means the dielectric permittivity of vacuum. All other quantities in the above equations are sufficiently defined explicitly or can be determined by solving equations (20)-(22).

All indices in the above relations assume values 1,2,3. It should be noted that the symbol z, used as an index or superscript, is not subject to the dummy suffix notation rule replacing the summation symbol throughout this paper. If summation with respect to this symbol is required it is in this paper denoted specifically by the summation symbol Σ.

Solution of the transcendental equation (13) with respect to f yields eigenfrequencies corresponding to all three existing modes of vibration and all odd harmonics. This is sufficient here since it is only odd harmonics that can be excited electrically in the electrode arrangement in Fig.1.

First-order Temperature Coefficient of Frequency

The process of vibration in plates is regarded as adiabatic. This means that the local changes taking place periodically in the elements of a plate are associated with local periodic changes in temperature. It also means that the material constants appearing in the transcendental equation (13) are adiabatic constants. It is further assumed that the changing temperature in the plate has a constant time average value T_0 that is also constant throughout the volume of the plate.

The following treatment requires an assumption that the transcendental equation (13) as well as the auxiliary relations (14)-(22) may be formally differentiated with respect to the implicite parameter of temperature T without their validity being impaired.

In order to obtain the first-order temperature coefficient of frequency the expression (13) is now formally differentiated with respect to the temperature parameter T, the dependence on T being assumed wherever it cannot be ruled out for logical reasons. After the differentiation and rearrangement of terms it is obtained

$$\begin{aligned}
\frac{1}{f}\frac{\partial f}{\partial T} = & \left[-1 + \frac{1}{S} \sum_{z=1}^{3} E_{iz} A_k^{(z)} \left(1 - \frac{\varepsilon_0 t}{f_r} \right) \frac{F_i F_k}{\varepsilon} \right] \frac{1}{t} \frac{\partial t}{\partial T} - \frac{1}{2\varrho} \frac{\partial \varrho}{\partial T} \\
& - \frac{1}{S} \sum_{z=1}^{3} E_{iz} A_k^{(z)} \frac{\varepsilon_0 \varepsilon t}{f_r^2} \frac{F_i F_k}{\varepsilon} \frac{1}{t} \frac{\partial (d-t)}{\partial T} \\
& + \frac{1}{S} \sum_{z=1}^{3} E_{iz} A_k^{(z)} \left[\Gamma_{ik} \left(B^{(z)} \operatorname{cotn} B^{(z)} - \frac{B^{(z)2}}{\sin^2 B^{(z)}} \right) \right] \frac{1}{2\Gamma^{(z)}} \frac{\partial \Gamma^{(z)}}{\partial T} \\
& - \frac{1}{S} \sum_{z=1}^{3} E_{iz} A_k^{(z)} B^{(z)} \operatorname{cotn} B^{(z)} \frac{\partial \Gamma_{ik}}{\partial T} \\
& + \frac{1}{S} \sum_{z=1}^{3} \left(E_{iz} A_k^{(z)} + E_{kz} A_i^{(z)} \right) \frac{F_i}{\varepsilon} \frac{\varepsilon_0 t}{f_r} \frac{\partial F_k}{\partial T} \\
& - \frac{1}{S} \sum_{z=1}^{3} E_{iz} A_k^{(z)} \left(1 + \frac{\varepsilon (d-t)}{f_r} \right) \frac{\varepsilon_0 t}{f_r} \frac{F_i F_k}{\varepsilon} \frac{1}{\varepsilon} \frac{\partial \varepsilon}{\partial T} \\
& - \frac{1}{S} \sum_{z=1}^{3} E_{iz} \left[\Gamma_{ik} B^{(z)} \operatorname{cotn} B^{(z)} - \frac{\varepsilon_0 t}{f_r} \frac{F_i F_k}{\varepsilon} \right] \frac{\partial A_k^{(z)}}{\partial T}
\end{aligned} \tag{23}$$

where

$$S = \sum_{z=1}^{3} E_{iz} A_k^{(z)} \left[\Gamma_{ik} \left(B^{(z)} \operatorname{cotn} B^{(z)} - \frac{B^{(z)2}}{\sin^2 B^{(z)}} \right) \right] \tag{24}$$

$$E_{iz} = \epsilon_{abi} \epsilon_{dez} Q_{ad} Q_{be} \tag{25}$$

All the derivatives, other quantities and values of material constants in (23) and throughout the following calculation are meant at the temperature $T = T_0$ for which the temperature coefficient of frequency $Tf^{(1)}$ is being determined. For the sake of simplicity of writing the formulae the notation stating this fact is completely omitted in this paper.

As can be seen from (23) the first-order temperature coefficient of frequency, $Tf^{(1)}$, is given by eight terms each of them originating from a different part of the transcendental equation (13). The first three terms appearing on the right hand side of (23) represent the contribution of the thermal expansion, the third one being obviously non-zero only if

$(d-t) \neq 0$, i.e. if an airgap exists between the plate and the electrodes. The fourth term in (23) corresponds to the temperature dependence of the effective elastic moduli of the three modes coupled in the vibration. The requirement that the surfaces of the plate be traction-free gives rise to the remaining four terms, the last of them being due to the change in amplitude direction of the waves associated with the vibration. It is only at this stage that the origin of different contributions to the first-order temperature coefficient of frequency can be easily traced up.

The following treatment is aimed at reducing the above expression (23) to simpler terms of the components of the tensor of thermal expansion and the temperature derivatives of elastic stiffnesses, piezoelectric stress constants and dielectric permittivities.

Realising that $\underline{d}$ is the electrode separation and $\underline{t}$ the thickness of the crystal plate we may write

$$\frac{1}{t}\frac{\partial(d-t)}{\partial T} = \left(\delta_{[d-t]} - 1\right)\frac{1}{t}\frac{\partial t}{\partial T} \qquad (26)$$

where

$$\delta_{[d-t]} = \begin{cases} 1 & \text{for } d = t\ , \\ 0 & \text{for } d \neq t\ . \end{cases}$$

Using the known relation between the components of the strain tensor and temperature representing the thermal expansion it may be written

$$\frac{1}{t}\frac{\partial t}{\partial T} = \alpha^{(i)'}_{11} = \alpha_i \alpha_j \alpha^{(i)}_{ij} \qquad (27)$$

Putting further

$$\rho = \frac{m}{V}$$

where $\underline{m}$ is the mass of the plate and $\underline{V}$ its volume, we may write

$$\frac{1}{\rho}\frac{\partial \rho}{\partial T} = -\frac{1}{V}\frac{\partial V}{\partial T} = -(\alpha^{(i)'}_{11} + \alpha^{(i)'}_{22} + \alpha^{(i)'}_{33})$$
$$= -(\alpha_i\alpha_j + \beta_i\beta_j + \gamma_i\gamma_j)\alpha^{(i)}_{ij} \qquad (28)$$

where $\alpha^{(i)'}_{ij}$ and $\alpha^{(i)}_{kl}$, (i,j,k,l=1,2,3), are the first-order isobaric thermal expansion coefficient referred to the primed and the basic sets of coordinate axes respectively. Using (26),(27) and (28) the first three terms of (23) result in

$$\left(A\alpha_i\alpha_j + \tfrac{1}{2}\beta_i\beta_j + \tfrac{1}{2}\gamma_i\gamma_j\right)\alpha^{(i)}_{ij} \qquad (29)$$

where

$$A = -\frac{1}{2} + \frac{1}{S}\sum_{z=1}^{3} E_{iz} A^{(z)}_{ik} \frac{F_i F_k}{\varepsilon}\frac{\varepsilon t}{b}\left(\frac{\varepsilon d}{b} - \frac{\varepsilon t}{b}\delta_{[d-t]}\right) \qquad (30)$$

Differentiating (18) and (19) we have immediately

$$\frac{\partial F_k}{\partial T} = \alpha_i\alpha_j\frac{\partial e_{ijk}}{\partial T} \qquad (31)$$

$$\frac{\partial \varepsilon}{\partial T} = \alpha_m\alpha_n\frac{\partial \varepsilon_{mn}}{\partial T} \qquad (32)$$

that will be substituted in the fifth and sixth term of (23).

In order to find a suitable expression for the components appearing in the last term of (23), auxiliary vectors $\vec{X}^{(z)}$ and $\vec{R}^{(z)}_i$, (i,z=1,2,3), defined by their components

$$X^{(z)}_k = \frac{\partial A^{(z)}_{ik}}{\partial T}\ , \qquad k = 1,2,3\ , \qquad (33)$$

$$R^{(z)}_{ik} = \Gamma_{ik} - \delta_{ik}\Gamma^{(z)}\ , \qquad k = 1,2,3\ , \qquad (34)$$

are introduced. Differentiating (21) and (22) and using the new notation we may write

$$\left(\vec{R}^{(z)}_i, \vec{X}^{(z)}\right) = -\frac{\partial R_{ik}}{\partial T}A^{(z)}_{ik}\ , \qquad z,i = 1,2,3 \qquad (35)$$

$$\delta_{ik}A^{(z)}_i X^{(z)}_k = 0\ , \qquad z = 1,2,3\ . \qquad (36)$$

With respect to the secular equation (20) that takes up, in the new notation, the following form

$$\left|R^{(z)}_{ik}\right| = 0\ , \qquad z = 1,2,3\ , \qquad (37)$$

only two of the three vectors $\vec{R}^{(z)}_i$ may be linearly independent for each value of $\underline{z}$. Assuming now that the value of $\underline{z}$ is fixed let it be vectors $\vec{R}^{(z)}_A$ and $\vec{R}^{(z)}_B$, where A and B are two definite fixed values of the three indices 1,2,3. The two vectors $\vec{R}^{(z)}_A$ and $\vec{R}^{(z)}_B$ form a plane containing also vector $\vec{X}^{(z)}$ as follows from (21) and (36). Thus it may be written

$$\vec{X}^{(z)} = k^{(z)}_\beta \vec{R}^{(z)}_\beta\ , \qquad \beta = A,B\ , \qquad (38)$$

where $k^{(z)}_A$ and $k^{(z)}_B$ are two real numbers. Using the former equations (35) and (38) and forming two scalar products $(\vec{R}^{(z)}_A, \vec{X}^{(z)})$ and $(\vec{R}^{(z)}_B, \vec{X}^{(z)})$ it is obtained

$$-\frac{\partial R^{(z)}_{Ak}}{\partial T}A^{(z)}_k = \left(k^{(z)}_\beta \vec{R}^{(z)}_\beta, \vec{R}^{(z)}_A\right)$$
$$-\frac{\partial R^{(z)}_{Bk}}{\partial T}A^{(z)}_k = \left(k^{(z)}_\beta \vec{R}^{(z)}_\beta, \vec{R}^{(z)}_B\right)$$

Solving these two equations with respect to $k^{(z)}_\beta$, β=A,B , and substituting back in (38) the vector $\vec{X}^{(z)}$ is found

$$\vec{X}^{(z)} = -\frac{\partial R^{(z)}_{\beta k}}{\partial T}A^{(z)}_k \vec{N}^{(z)}_\beta \qquad (39)$$

where components of the vector $\vec{N}_{\beta}^{(z)}$, $\vec{N}_{\beta b}^{(z)}$ b=1,2,3, are

$$N_{\beta b}^{(z)} = \delta_{\beta\alpha} \frac{(\vec{R}_{\beta}^{(z)})^2 R_{ab}^{(z)} - (\vec{R}_{a}^{(z)}, \vec{R}_{\beta}^{(z)}) R_{\beta b}^{(z)}}{(\vec{R}_{\alpha}^{(z)})^2 (\vec{R}_{\beta}^{(z)})^2 - (\vec{R}_{\alpha}^{(z)}, \vec{R}_{\beta}^{(z)})^2} , \qquad (40)$$

α, β, β = A,B , while $\alpha \neq \beta$, and

$$\frac{\partial R_{\beta ik}^{(z)}}{\partial T} = \delta_{\beta i} \frac{\partial \Gamma_{ik}}{\partial T} - \delta_{\beta k} \frac{\partial \Gamma^{(z)}}{\partial T} \qquad (41)$$

Therefore

$$X_{\beta}^{(z)} = - \frac{\partial R_{\beta ik}^{(z)}}{\partial T} A_k^{(z)} N_{\beta b} \qquad (42)$$

In very rare cases it happens that the three vectors $\vec{R}_i^{(z)}$, (i=1,2,3, z fixed), defined in (34) are colinear so that the former procedure to determine vector $\vec{X}^{(z)}$ cannot be used. On the other hand it would be possible to show that in these cases vector $\vec{X}^{(z)}$ has no influence on the magnitude of the first-order temperature coefficient $Tf^{(1)}$. However, the vibration modes where this occurs cannot be excited electrically so that the frequency-temperature behaviour cannot be studied with them experimentally either. Therefore, these cases lie beyond the limits of our interest and will no more be discussed in this paper.

A similar expression to that of (42) can be obtained for any value assumed by the index z .

An expression for $\partial\Gamma^{(z)}/\partial T$, z=1,2,3, will be found now in terms of $\partial\Gamma_{ik}/\partial T$ using the secular equation (20). The calculation is done for a fixed value of z , the result will be similarly valid for arbitrary z .

The equation (20) can be rewritten

$$(\Gamma^{(z)})^3 - \rho_1 (\Gamma^{(z)})^2 + \rho_2 \Gamma^{(z)} + \rho_3 = 0 \qquad (43)$$

where

$$\begin{aligned} \rho_1 &= \delta_{ik} \Gamma_{ik} , \\ \rho_2 &= \tfrac{1}{2} \epsilon_{abc} \epsilon_{def} \delta_{ad} \Gamma_{be} \Gamma_{cf} , \\ \rho_3 &= \tfrac{1}{6} \epsilon_{abc} \epsilon_{def} \Gamma_{ad} \Gamma_{be} \Gamma_{cf} \end{aligned} \qquad (44)$$

and differentiated with respect to T

$$\left[3(\Gamma^{(z)})^2 - 2\rho_1 \Gamma^{(z)} + \rho_2\right] \frac{\partial \Gamma^{(z)}}{\partial T} - \left[(\Gamma^{(z)})^2 \frac{\partial \rho_1}{\partial T} - \Gamma^{(z)} \frac{\partial \rho_2}{\partial T} + \frac{\partial \rho_3}{\partial T}\right] = 0$$

Thus it may be written

$$\frac{\partial \Gamma^{(z)}}{\partial T} = \frac{(\Gamma^{(z)})^2 \frac{\partial \rho_1}{\partial T} - \Gamma^{(z)} \frac{\partial \rho_2}{\partial T} + \frac{\partial \rho_3}{\partial T}}{3(\Gamma^{(z)})^2 - 2\rho_1 \Gamma^{(z)} + \rho_2} \qquad (45)$$

Since $\partial\Gamma^{(z)}/\partial T$ in (45) is to be referred to $T=T_0$ as well as all derivatives in this paper, it is always the solution of (20) with respect to $\Gamma^{(z)}$ that is to be substituted for $\Gamma^{(z)}$ in (45). If the three vectors $\vec{R}_i^{(z)}$ defined in (34) and appearing in the three lines of (20) are not colinear, as is the case under consideration, the three roots $\Gamma^{(z)}$ of the secular equation (20) are different. Therefore the denominator of (45), equal to the partial derivative of (43) with respect to $\Gamma^{(z)}$, is always different from zero. Thus the formula (45) makes always sense.

Substituting from (44) into (45) we obtain finally

$$\frac{\partial \Gamma^{(z)}}{\partial T} = D_{ik}^{(z)} \frac{\partial \Gamma_{ik}}{\partial T} , \qquad (46)$$

where

$$D_{ik}^{(z)} = \frac{(\Gamma^{(z)})^2 \delta_{ik} - \Gamma^{(z)} \epsilon_{abi} \epsilon_{dek} \delta_{ad} \Gamma_{be} + \frac{1}{2} \epsilon_{abi} \epsilon_{dek} \Gamma_{ad} \Gamma_{be}}{3(\Gamma^{(z)})^2 - 2\rho_1 \Gamma^{(z)} + \rho_2} \qquad (47)$$

while, with respect to the definition (15), $\partial\Gamma_{ik}/\partial T$ is given by

$$\frac{\partial \Gamma_{ik}}{\partial T} = \alpha_j \alpha_l \left[\frac{\partial c_{ijkl}}{\partial T} + \frac{e_{mab} \alpha_a \alpha_n e_{mij}}{\varepsilon} \left(2 \frac{\partial e_{nkl}}{\partial T} - \frac{e_{nkl}}{\varepsilon} \frac{\partial \varepsilon}{\partial T}\right)\right] . \qquad (48)$$

Substituting from (24),(25),(29), (30),(31),(32),(40),(41),(42),(46),(47) and (48) into (23) the first-order temperature coefficient of frequency is given by

$$\begin{aligned} Tf^{(1)} = {} & \left(A \alpha_i \alpha_j + \tfrac{1}{2} \beta_i \beta_j + \tfrac{1}{2} \gamma_i \gamma_j\right) \alpha_{ij}^{(1)} \\ & + \frac{1}{2S} \left(G_{ik} + 2H_{ik}\right) \alpha_j \alpha_l \frac{\partial c_{ijkl}}{\partial T} \\ & + \frac{1}{S} \left(G_{ik} + 2H_{ik} + I_{ik}\right) \frac{F_k}{\varepsilon} \beta_j \alpha_m \frac{\partial e_{mij}}{\partial T} \\ & - \frac{1}{2S} \left[G_{ik} + 2H_{ik} + \left(1 + \frac{\varepsilon(d-t)}{b}\right) I_{ik}\right] \frac{F_i F_n}{\varepsilon^2} \alpha_m \alpha_n \frac{\partial \varepsilon_{mn}}{\partial T} , \end{aligned} \qquad (49)$$

where

$$G_{ik} = \sum_{z=1}^{3} \left\{E_{\beta z} A_{\beta}^{(z)} \left[\Gamma_{\beta b} \left(B^{(z)} \operatorname{cotm} B^{(z)} - \frac{(B^{(z)})^2}{\sin^2 B^{(z)}}\right)\right] \frac{1}{\Gamma^{(z)}} D_{ik}^{(z)}\right\} , \qquad (50)$$

$$H_{ik} = - \sum_{z=1}^{3} \left\{E_{iz} A_k^{(z)} B^{(z)} \operatorname{cotm} B^{(z)} + E_{\beta z} \left(\Gamma_{\beta b} B^{(z)} \operatorname{cotm} B^{(z)} - \frac{\varepsilon_0 t}{b} \frac{F_\alpha F_b}{\varepsilon}\right) M_{\beta ik}^{(z)}\right\} , \qquad (51)$$

$$I_{ik} = \frac{\varepsilon_0 t}{b} \sum_{z=1}^{3} \left(E_{iz} A_k^{(z)} + E_{kz} A_i^{(z)}\right) , \qquad (52)$$

$$M_{\beta ik}^{(z)} = N_{\beta b}^{(z)} \delta_{\beta\alpha} A_{\alpha}^{(z)} D_{ik}^{(z)} - N_{\beta b}^{(z)} \delta_{\beta i} A_k^{(z)} , \qquad (53)$$

where β = A,B .

As regards the material constants in (49) it may only be repeated that the values of isagric adiabatic elastic stiffnesses, adiabatic piezoelectric stress constants, adiabatic dielectric permittivities at constant strain and isagric isobaric thermal expansion coefficients are always to be used as it agrees with the formulation of the basic equations in this paper and with the sense in which the thermal expansion coefficients have been used.

Is is essential to stress again that as regards the elastic constants and their temperature coefficients the whole treatment was carried out in terms of elastic stiffnesses as constant electric field-strength while the elastic stiffnesses at constant normal component of electric displacement, $c^{D_n}_{ijkl}$, have always been avoided. When analysed in detail it is found that the elastic stiffnesses $c^{D_n}_{ijkl}$ form one tensor only if all of them are related simultaneously to one direction chosen as the direction of D_n.[7] In general we have as many c^{D_n} tensors as there are possible directions of D_n. Further the symmetry of the c^{D_n} tensor is generally lower than that of the c^E tensor being reduced by the presence of the direction of D_n. For example in the case of alpha-quartz it is valid that $c^E_{1212} = \frac{1}{2}(c^E_{1111} - c^E_{1122})$ while similar equation in terms of $c^{D_n}_{ijkl}$ does not hold at all. Therefore the usage of the isagric stiffnesses c^E_{ijkl} for the interpretation of the frequency-temperature dependence has an indisputable advantage over the use of $c^{D_n}_{ijkl}$. These properties of the tensor $c^{D_n}_{ijkl}$ were not always, in the past, fully appreciated and became the reason for some confusion.

The expression (49) is written in terms of the temperature derivatives of elastic stiffnesses, piezoelectric stress constants and dielectric permittivities rather than the temperature coefficients of the respective quantities. This is done so only because the temperature coefficients do not form tensors. This would make the notation in (49) complicated. However, the way to replace the temperature derivatives by the temperature coefficients is obvious.

Speaking for the moment in terms of temperature coefficients of elastic stiffnesses, piezoelectric stress constants and dielectric permittivities, the relative weight with which each of these three quantities contributes to the temperature coefficient $Tf^{(1)}$ can be estimated from (49) as given by the ratio of

$$c : e^2/\varepsilon : e^2/2\varepsilon$$

respectively, where c, e and ε mean in turn the order of magnitude of the elastic stiffnesses, piezoelectric stress constants and dielectric permittivities.

The resulting expression (49) gives the first-order temperature coefficient of frequency, $Tf^{(1)}$, resolved into four contributions corresponding to the linear components of four qualitatively different effects. These are as arranged on the right hand side of (49): thermal expansion, temperature dependence of elastic stiffnesses, piezoelectric stress constants and dielectric permittivities. These are represented by the material constants α_{ij}, $\partial c_{ijkl}/\partial T$, $\partial e_{ijk}/\partial T$ and $\partial \varepsilon_{ij}/\partial T$, being components of tensors of the second, fourth, third and second rank respectively.

All other quantities appearing in (49) form together coefficients standing in front of the material constants of the above four types. These coefficients depend explicitly on material constants as elastic stiffnesses, piezoelectric stress constants and dielectric permittivities and, also, on the orientation of the plate thickness, electrode separation and amplitudes of the waves coupled in the plate vibrations. Through the quantity $B^{(z)}$ defined by (16) the first-order temperature coefficient $Tf^{(1)}$ depends on the frequency of vibration, i.e. on the mode and order of harmonic.

The real vibrating plates have finite dimensions and it is known experimentally that $Tf^{(1)}$ is affected by this. It is thought that $Tf^{(1)}$ depends on the ratio of the main face dimension (diameter) of the plate to its thickness and on the contour of the plate.[8] This is somewhat contradicted and clarified by a short study on alpha-quartz AT-cuts by GERBER.[9] While the variations in contour do not seem to have any effect on the optimum zero angle of the AT-cuts, the analysis[9] shows that the zero angle depends only on the diameter/thickness ratio multiplied by the order of harmonic. Reformulated, this conclusion means that the temperature coefficient $Tf^{(1)}$ depends only on the diameter/wavelength ratio, the wavelenght corresponding to the wave associated with the plate vibrations. Insofar as the results[9] may be generalized the temperature coefficient $Tf^{(1)}$ depends, as regards the plate dimensions, only on the diameter/wavelength ratio as long as its value is small. When this ratio exceeds certain critical value the temperature coefficient $Tf^{(1)}$ becomes independent of the plate dimensions.

The diameter/wavelength ratio does not appear in the expression (49) as a variable. However, in agreement with the assumption that $t,d \ll w,l$ made before it may be concluded that the expression (49) is suitable rather for greater values of the diameter/thickness ratio.

Thus, by means of the expression (49), the first-order temperature coefficient of frequency is given as a function of the plate thickness, orientation, frequency (i.e. mode of vibration and order of harmonic) and of electrode separation provided that the diameter/thickness ratio exceeds certain critical value. This critical value must be determined experimentally.

Application to Alpha-quartz

The existing experimental material regarding the temperature-frequency behavior of alpha-quartz plates vibrating in thickness is relatively abundant making it possible to verify the new expression for $Tf^{(1)}$ presented in this paper. The paper[10] by BECHMANN, BALLATO and LUKASZEK gives twenty-four experimental values for the first-order temperature coefficient of frequency, $Tf^{(1)}$, of alpha-quartz plates vibrating in thickness as obtained for different fundamental thickness modes of plated blanks at the temperature $T_0=25^{o}C$. Their interpretation was given by means of temperature coefficients of elastic stiffnesses and thermal expansion coefficients. The temperature coefficients of elastic stiffnesses were determined there from the experimental values of $Tf^{(1)}$ and, on their basis, the theoretical values for $Tf^{(1)}$ were calculated back and compared with the experimental ones. The interpretation was carried out neglecting the effect of the order of harmonic as well as the existence of coupling among the modes of vibration. The difficult concept of the tensor of elastic stiffnesses at constant normal component of electric displacement seems to have been used. The degree of agreement between theory and experiment as obtained by the above mentioned authors[10], expressed in terms of the root mean square deviation calculated for the 24 pairs of experimental and theoretical values of $Tf^{(1)}$, is equal to $3.07 \times 10^{-6}\ (^{o}C)^{-1}$. The sum of absolute deviations amounts to $50.8 \times 10^{-6}\ (^{o}C)^{-1}$, the maximum absolute deviation being $6.73 \times 10^{-6}\ (^{o}C)^{-1}$.

The interpretation of the temperature-frequency behavior of plates vibrating in thickness suggested here and the corresponding expression (49) are substantially different from those used by the above mentioned authors[10].

To verify the new expression (49) and to justify the whole relatively complicated approach leading to it it was inevitable to start from the basic experimental data, i.e. from the experimental values of $Tf^{(1)}$. To facilitate the comparison of the relative success of the new expression (49) and the old traditional interpretation 10, the experimental values for $Tf^{(1)}$ were taken over from the paper quoted. All of the data together with the orientations of the corresponding plates are reprinted in Tab. 1.

In this part of the present paper where the expression (49) is actually used the tensor index notation useful in formal calculations was given up and the quantities appearing in (49) were dealt with in the two index matrix notation. From now on the concept of temperature coefficient of material constants was used rather than their temperature derivatives since the letter were not practical here. This means that for example instead of the temperature derivative $\partial c_{1111}/\partial T$ that appears in (49) the first-order temperature coefficient $Tc_{11}^{(1)}=(1/c_{11})\,\partial c_{11}/\partial T$ was used. The whole expression (49) was adjusted accordingly.

Due to the crystal symmetry of alpha-quartz there exist only six independent values of the temperature coefficients of elastic stiffnesses, $Tc_{11}^{(1)}$, $Tc_{12}^{(1)}$, $Tc_{13}^{(1)}$, $Tc_{14}^{(1)}$, $Tc_{33}^{(1)}$ and $Tc_{44}^{(1)}$, two temperature coefficients of piezoelectric stress constants, $Te_{11}^{(1)}$, $Te_{14}^{(1)}$, and only two independent values of the temperature coefficients of dielectric permittivities, $T\varepsilon_{11}^{(1)}$ and $T\varepsilon_{33}^{(1)}$. All remaining temperature coefficients are either dependent on the values of the quantities mentioned before or equal to zero.

Apart from these ten temperature coefficients the temperature coefficient of frequency, $Tf^{(1)}$, also depends on the first-order thermal expansion tensor that possesses, with alpha-quartz, only two independent components $\alpha_{11}^{(1)}$ and $\alpha_{33}^{(1)}$.

Applying the expression (49) and using the method of minimum square deviation an attempt was made to calculate the ten temperature coefficients, $Tc_{\lambda\mu}^{(1)}$, $Te_{i\lambda}^{(1)}$, $T\varepsilon_{mn}^{(1)}$, from the experimental values of $Tf^{(1)}$. The components of the thermal expansion tensor, $\alpha_{ij}^{(1)}$, were regarded as known here as well as further in this paper. Their values and all the values of other material constants as the elastic stiffnesses, piezoelectric stress constants, dielectric permittivities and density were taken over from the paper quoted.[10]

Simultaneously with the values of the ten temperature coefficients their mean quadratic errors were calculated. These were, in the case of $Te_{11}^{(1)}$, $Te_{14}^{(1)}$, $T\varepsilon_{11}^{(1)}$ and $T\varepsilon_{33}^{(1)}$ extremely high and comparable with the values of the temperature coefficients found. As a result of this the values obtained for $Te_{11}^{(1)}$, $Te_{14}^{(1)}$, $T\varepsilon_{11}^{(1)}$ and $T\varepsilon_{33}^{(1)}$ were regarded as not acceptable.

Consequently, the found values of $Tc_{\lambda\mu}^{(1)}$, though having the relative error of the order of 10 per cent, were also rejected being not acceptable separately.

In order to explain this failure the role of $Te_{i\lambda}^{(1)}$ and $T\varepsilon_{mn}^{(1)}$ as contributors to the temperature coefficient $Tf^{(1)}$ was estimated. To this end the published[11] values of $Te_{11}^{(1)}=-166 \times 10^{-6}(^{o}C)^{-1}$, $Te_{14}^{(1)}=-1440 \times 10^{-6}(^{o}C)^{-1}$, $T\varepsilon_{11}^{(1)}=29 \times 10^{-6}(^{o}C)^{-1}$ and $T\varepsilon_{33}^{(1)}=39 \times 10^{-6}(^{o}C)^{-1}$ were used after a slight adjustment to the conditions required in this paper. (The values for $T\varepsilon_{mn}^{(1)}$ were given originally[11] at constant strain.) These values were then substituted in (49) and the magnitudes of the last two terms on the right hand side of (49) for each individual plate and mode listed in Tab.1, were evaluated. This estimation showed that in the majority of cases the contribution from the temperature dependence of piezoelectric stress constants and dielectric permittivities fell behind the limits of accuracy of the experimental values of $Tf^{(1)}$. This was the reason why the quantities $Te_{i\lambda}^{(1)}$ and $T\varepsilon_{mn}^{(1)}$ could not be quantitatively determined with a degree of accuracy warranting their acceptance.

As a result of the above comparison it was concluded that the accuracy of the experimental values of $Tf^{(1)}$ was not sufficient for the last two terms in (49) to play a meaningful role in their interpretation. Consequently, these two terms were dropped from (49) and the following calculations carried out as if the temperature coefficient of frequency depended exclusively on the first two terms in (49) and, as such, was caused only by the thermal expansion and the temperature dependence of elastic stiffnesses.

The simplified version of (49) was then used and suitable values of the temperature coefficients of elastic stiffnesses determined again. On the recommendation of their authors[10] the experimental values $Tf^{(1)}$ corresponding to the X,Y, AT and BT cuts were accepted without reservation as very reliable. The temperature coefficients $Tc_{\lambda\mu}^{(1)}$ were then calculated so as to suit completely these four values and, at the same time, give the best fit over the remaining twenty experimental values $Tf^{(1)}$. The temperature coefficients of elastic stiffnesses obtained in this manner are presented in Tab.2. In several cases they differ considerably from those given by BECHMANN et al.[10] as well as by other authors[12,13]. The reason for this obviously lies in the specific character of the expression (49).

The ultimate justification for the obtained values of $Tc_{\lambda\mu}^{(1)}$ is, however, the degree of agreement between the theoretical and experimental values of $Tf^{(1)}$. To this end the found values of $Tc_{\lambda\mu}^{(1)}$ from Tab.2. were substituted back in (49) and the theoretical values for $Tf^{(1)}$ calculated. These results are also shown in Tab.1. The root mean square deviation obtained for the twenty-four pairs of the experimental and theoretical values of $Tf^{(1)}$ was $1.14 \times 10^{-6}(^{o}C)^{-1}$, the sum of absolute deviations was $19.1 \times 10^{-6}(^{o}C)^{-1}$, the maximum absolute deviation being $2.84 \times 10^{-6}(^{o}C)^{-1}$. The agreement between theory and experiment is therefore considerably better than that obtained by BECHMANN et al.[10] whose theoretical values are also reprinted in Tab.1. for comparison.

On the basis of the newly determined temperature coefficients of elastic stiffnesses the temperature coefficients of elastic compliances, $Ts_{\lambda\mu}^{(1)}$, at constant electric fieldstrength were also determined. They are shown in Tab.3.

With the suitable values of the temperature coefficients of elastic stiffnesses determined the expression (49) can be readily used to calculate the temperature coefficient $Tf^{(1)}$ for alpha-quartz plates of any orientation, mode of vibration and harmonic, any thickness and electrode separation.

As an illustration a number of values characterizing the dependence of $Tf^{(1)}$ on the order of harmonic and electrode separation and the temperature-frequency-angle behavior were calculated for AT and BT cuts. Some of them are shown in Tab.4. and Tab.5.

From Table 4. it can be concluded that the theoretical results obtained for either cut are qualitatively similar.

The temperature coefficient $Tf^{(1)}$ changes with the number of overtone, the increment in absolute value being greatest between the fundamental and the first overtone. Above the 7th harmonic the change in $Tf^{(1)}$ becomes negligible. This agrees qualitatively with the approximate formula given by BECHMANN.[14]

As regards the effect of the electrode separation the temperature coefficient $Tf^{(1)}$ depends, as it was concluded from (49), on the airgap/thickness ratio. An abrupt change in $Tf^{(1)}$ obtained at the point where the airgap/thickness ratio becomes non-zero is due to thethermal expansion term in (49) containing $\delta_{[d-t]}$ whose value changes discontinuously from zero to one with the appearance of a gap between the electrodes and the plate. For a small gap this term becomes relatively very effective and is the reason why the dependence of $Tf^{(1)}$ on the electrode separation is not always monotonous. This seems to be the case with the AT cut so that the angle $\theta_o = 35^{o}15'$ is the zero angle not only for a plated blank but also for a plate with

electrodes separated by a small gap from the plate. This theoretical result is indirectly confirmed by experiment where the zero angles θ_x for different values of the airgap/thickness ratio appear to be on both sides of θ_o.[8]

The temperature-frequency-angle dependence is characterized by the theoretical results in Tab.5. Calculated for the plated blank and fundamental mode, the temperature coefficient $Tf^{(1)}$ depends linearly on the angle of cut at least in an $\pm 2^o$ interval round the zero angle θ_o. The derivative $\partial Tf^{(1)}/\partial\theta$ calculated here for an interval of 6' round the zero angle θ_o agrees in the both cases very well with the experimental values, being in the case of the AT cut only by 12 per cent less.[8] However, for overtones and a non-zero airgap/thickness ratio the same derivative slightly changes which is contradicted by the experimental values published for AT cuts.[8]

For two overtones and two non-zero airgaps the corresponding zero angles θ_x were calculated as an example as shown in Tab.4. Some of the values of θ_x for the AT cuts could be compared with the experimental results.[8] Bearing in mind that the zero angle θ_o in the source quoted is different from that accepted in this paper the agreement is very good because the intervals between θ_o and θ_x in comparable cases are similar.

Conclusion

The improvement in the interpretation of the frequency-temperature behavior obtained in this paper should be basically ascribed to two features that have been introduced here and changed the structure of the expression of the first-order temperature coefficient of frequency.

In the first place it is a fact that the elastic strain does not vanish at the traction-free main faces of the plates since the zero tension is obtained as a result of combination of non-zero strain and electric field. Second, it is the exclusive use of the tensor of elastic stiffnesses as constant electric field in recognition of the difficulties connected with the reduced and changing symmetry of the tensor of elastic stiffnesses at constant normal component of electric displacement.

The dependence of the frequency-temperature coefficient on the order of harmonic and electrode separation and the appearance of the temperature coefficients of piezoelectric stress constants and dielectric permittivities in the final expression (49) are obtained as a direct logical consequence of this approach and would not be obtained otherwise.

The expression (49) for $Tf^{(1)}$ entailing all the above features has the further advantage of being suitable for piezoelectric crystals of any crystalographic symmetry and of being uniform for plates of any orientation and any mode of vibration.

The expression (49) is sensitive to the consistence of the substituted eigenfrequency value with the accepted set of material constants used elsewhere in (49). This is why the complicated transcendental equation (13) had always to be solved to secure the six significant figure accuracy or, better, the degree of consistence of the frequency with the used material constants. In this connection it may be of interest to note that the simpler and only approximate transcendental equation derived long ago[5,6] gives excellent values for the frequency differring, as it was proved on a number of examples, from the"accurate" values of frequency based on the Tiersten approximation[1] not more than by 0.1 per cent. The most frequently used expression for frequency based on the simplified assumption that the plate thickness is equal to an integer multiple of half a wavelength[10] leads, for fundamental modes, to results differring by 1 per cent from the "accurate" ones.

The application of the expression (49) to alpha-quartz leads to the conclusion that the present accuracy of the experimental values for the temperature coefficient of frequency is not sufficient for determination of the temperature coefficients of piezoelectric stress constants and dielectric permittivities.

Further analysis not described in this paper showed that the terms containing the temperature derivatives of piezoelectric stress constants and dielectric permittivities can be,in the case of alpha-quartz, omitted without seriously impairing the quality of interpretation of experimental results.

References

[1] H.F.TIERSTEN,"Thickness Vibrations of Piezoelectric Plates," J.Acoust.Soc. Am., vol.35,pp.53-58,January 1963.

[2] "Standarts on Piezoelectric Crystals" Proc.IRE, vol.37,pp.1378-1395,December 1949.

[3] V.ALDA,K.HRUSKA,J.TICHY,"Propagation of Waves Through Infinite Piezoelectric Medium," Czech.J.Phys., vol. A 13,pp.345-366,1963.

[4] K.HRUSKA,"Relation Between the General and the Simplified Condition for the Velocity of Waves in a Piezoelectric Medium," Czech.J.Phys., vol. B 18,pp.214-221,1968.

[5] R.BECHMANN,"Dickenschwingungen piezoelektrisch erregter Kristallplatten,"

Hochfrequenz. und Elektroak., vol.56, pp.14-21,1940.
[6] A.W.LAWSON,"The Vibration of Piezoelectric Plates," Phys.Rev., vol.62, pp.71-76,July 1942.
[7] W.G.CADY,"Piezoelectricity," McGraw-Hill Book Co., New York,N.Y.,London, England,pp.271-272,1946.
[8] R.BECHMANN,"Frequency-Temperature-Angle Characteristics of AT-Type Resonators Made of Natural and Synthetic Quartz," Proc.IRE, vol.44,pp. 1600-1607,November 1956.
[9] E.A.GERBER,"Temperature Coefficient of AT Cut Quartz Crystal Vibrators," Proc.IRE, vol.43,p.1529,October 1955.
[10] R.BECHMANN,A.D.BALLATO,T.J.LUKASZEK, "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz," Proc.IRE, vol. 50,pp.1812-1822,August 1962.
[11] R.BECHMANN,"The Elastic,Piezoelectric, and Dielectric Constants of Piezoelectric Crystals," Numerical Data and Relationships in Science and Technology, Group III, Crystal and Solid State Physics," Springer, Berlin, Heidelberg, New York, p.83 and 94, 1969.
[12] W.P.MASON,"Zero Temperature Coefficient Quartz Crystal for Very High Temperatures," Bell Sys.Tech.J., vol.30,pp.366-380,April 1951.
[13] I.KOGA,M.ARUGA,Y.YOSHINAKA,"Theory of Plane Elastic Waves in a Piezoelectric Crystalline Medium and Determination of Elastic and Piezoelectric Constants of Quartz," Phys.Rev., vol.109, pp.1467-1473,March 1958.
[14] R.BECHMANN,"Influence of the Order of Overtone on the Temperature Coefficient of Frequency of AT-Type Quartz Resonators," Proc.IRE, vol.43,p.1667, November 1955.

Acknowledgement

The author wishes to acknowledge his indebtedness to Mr. Julian Rees, Computer Manager of the University of Khartoum, and his Staff, especially Dr. Otakar Blahnik and Dr. Jaromir Janko, for their support facilitating fast progress in this work.

ORIENTATION (yxwl)Φθ		MODE	FIRST-ORDER TEMPERATURE COEFFICIENT OF FREQUENCY $Tf^{(1)}$ 10^{-6} $(^{o}C)^{-1}$			CONTRIBUTION TO $Tf^{(1)}$ NEWLY CALCULATED 10^{-6} $(^{o}C)^{-1}$	
Φ	θ		OBSERVED [10]	CALCULATED [10]	CALCULATED NEWLY	THERMAL EXPANSION	NEGLECTED CONTRIBUTION OF $Te^{(1)}_{m\lambda}$ AND $T^{(1)}_{mn}$
30°	0°	A	-20.5	-20.5	-20.50	3.74	-0.47
30°	10°	A	-29.3	-29.8	-30.2	3.98	-0.34
30°	20°	A	-42.0	-41.4	-42.81	4.47	-0.23
30°	30°	A	-54.8	-51.7	-55.51	5.30	-0.15
0°	-49° 13'	B	0	0	0.00	7.31	-0.34
5°	-47°	B	- 0.9	1.7	0.34	7.07	-0.37
5°	-48°	B	- 1.5	- 0.2	- 1.20	7.18	-0.35
10°	-38°	B	1.1	6.8	2.06	6.10	-0.51
10°	-40°	B	- 2.4	4.3	- 0.11	6.31	-0.47
15°	-34° 30'	B	- 7.4	- 0.67	- 6.70	5.74	-0.47
15°	-35°	B	- 7.45	- 0.98	- 7.06	5.79	-0.46
30°	20°	B	- 9.3	- 9.3	- 6.47	4.47	-0.13
30°	30°	B	-19.1	-18.2	-19.35	5.30	-0.19
0°	0°	C	92.5	92.5	92.50	3.74	-1.47
0°	35° 15'	C	0	0	0.00	5.82	0.17
10°	-32°	C	0.26	- 1.5	0.22	5.49	0.00
10°	-33°	C	- 0.87	- 1.9	- 0.38	5.59	0.00
12.5°	-33°	C	[illegible]	- 1.62	0.17	5.59	0.01
12.5°	-33° 30'	C	- 0.4	- 2.07	- 0.29	5.64	0.01
15°	-34° 30'	C	1.3	- 2.9	- 0.85	5.74	0.01
15°	-35°	C	- 0.2	- 3.5	- 1.48	5.79	0.01
20°	34° 20'	C	- 0.06	- 0.5	- 0.22	5.72	0.15
30°	34°	C	0.75	0.26	1.75	5.69	0.09
30°	36°	C	- 4.55	- 5.4	- 2.34	5.89	0.09

Tab.1. Measured and calculated first-order temperature coefficient of frequency $Tf^{(1)}$ and the contribution to it from the thermal expansion and from the temperature coefficients of piezoelectric stress constants and dielectric permittivities. The data is given for the thickness modes A,B,and C of quartz plates selected by BECHMANN et al.[10]

λμ	$Tc^{(1),E}_{\lambda\mu}$ 10^{-6} $(^{o}C)^{-1}$	
	NEW	[10]
11	- 43.4	- 48.5
12	-2159	-3000
13	-765	-550
14	78.1	101
33	-262	-160
44	-143	-177
66	142	178

Tab.2. Values for the first-order temperature coefficients of the stiffnesses of alpha-quartz at 25°C.

λμ	$Ts^{(1),E}_{\lambda\mu}$ 10^{-6} $(^{o}C)^{-1}$	
	NEW	[10]
11	6.7	15.5
12	-996	-1370
13	-333	-166
14	105	134
33	231	140
44	168	210
66	-117	-145

Tab.3. Values for the first-order temperature coefficients of the compliances of alpha-quartz at 25°C.

MODE ORDER	AIRGAP THICKNESS	AT CUT $\theta_o = 35°15'$		BT CUT $\theta_o = -49°13'$	
		$Tf^{(1)}_{\theta_o}$	$Tf^{(1)}_{\theta_x} = 0$	$Tf^{(1)}_{\theta_o}$	$Tf^{(1)}_{\theta_x} = 0$
		$10^{-6}(°C)^{-1}$	θ_x	$10^{-6}(°C)^{-1}$	θ_x
1	0	0.00		0.00	
3	0	0.53	35°22' 42"	0.29	-49°23'17"
5	0	0.57	35°23' 22"	0.31	-49°24'02"
7	0	0.58		0.32	
9	0	0.59		0.32	
99	0	0.61		0.32	
1	10^{-6}	-0.16		0.05	
1	0.05	0.00		0.10	
1	0.10	0.10		0.13	
1	0.25	0.27		0.19	
1	0.50	0.38		0.24	
1	1.00	0.48	35°21'52"	0.27	-49°22'35"
1	2.00	0.53	35°22'43"	0.29	-49°23'25"
1	100.00	0.59		0.32	

Tab.4. Calculated values for the first-order temperature coefficient of frequency $Tf^{(1)}$ for AT and BT cuts under different conditions. The symbol θ_x means the new zero angle calculated for the conditions stated.

MODE ORDER	AIRGAP THICKNESS	$(\partial Tf^{(1)}/\partial\theta)_{\theta_o}$	
		$10^{-6}(°C.°\theta)^{-1}$	
		AT CUT $\theta_o = 35°15'$	BT CUT $\theta_o = -49°13'$
1	0	-4.52	1.75
3	0	-4.14	1.70
5	0	-4.10	1.68
7	0	-4.12	1.69
9	0	-4.09	1.69
99	0	-4.34	1.30
1	10^{-6}	-4.52	1.75
1	0.05	-4.44	1.74
1	0.10	-4.38	1.71
1	0.25	-4.29	1.73
1	0.50	-4.21	1.73
1	1.00	-4.17	1.70
1	2.00	-4.14	1.70
1	100.00	-4.09	1.69

Tab.5. Calculated values for the derivative of the temperature coefficient of frequency with respect to the angle of orientation for AT and BT cuts under different conditions.

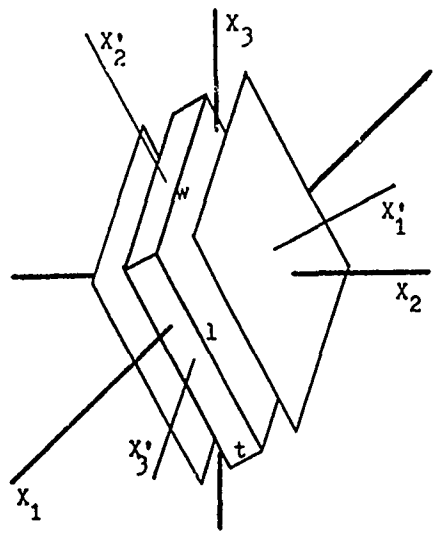


Fig.1. Arrangement of the crystal plate, air gaps, electrodes, and the basic and auxiliary sets of reference axes.

List of Captions

Tab.1. Measured and calculated first-order temperature coefficient of frequency $Tf^{(1)}$ and the contribution to it from the thermal expansion and from the temperature coefficients of piezoelectric stress constants and dielectric permittivities. The data is given for the thickness modes A,B,and C of quartz plates selected by BECHMANN et al.[10]

Tab.2. Values for the first-order temperature coefficients of the stiffnesses of alpha-quartz at 25°C.

Tab.3. Values for the first-order temperature coefficients of the compliances of alpha-quartz at 25°C.

Tab.4. Calculated values for the first-order temperature coefficient of frequency $Tf^{(1)}$ for AT and BT cuts under different conditions. The symbol θ_x means the new zero angle calculated for the conditions stated.

Tab.5. Calculated values for the derivative of the temperature coefficient of frequency with respect to the angle of orientation for AT and BT cuts under different conditions.

Fig.1. Arrangement of the crystal plate, air gaps, electrodes, and the basic and auxiliary sets of reference axes.

THE EFFECT OF STATIC ELECTRIC FIELDS ON THE ELASTIC CONSTANTS OF α-QUARTZ

John A. Kusters

Hewlett-Packard Laboratories
Palo Alto, California

Summary

The elastic constants of quartz are generally assumed to be independent of an applied electric field. However, recent experimental evidence indicates that the change in resonant frequency which occurs when a DC voltage is applied to the electrodes of a crystal cannot be explained by a simple change in the thickness and density of that crystal through the piezoelectric effect. For certain orientations, the measured frequency change is an order of magnitude larger than that predicted by the simple theories. The effect in quartz can best be explained as a modification of the elastic constants when intense electric fields are applied to the material. This effect can be adequately represented by a fifth-rank polarizing tensor, which describes the effect of a unidirectional electric field on the elastic constants of quartz.

Introduction

In the course of several experiments on resonating crystals, a rather large frequency change was observed on doubly rotated cuts of quartz when a constant potential was applied across the electrodes. In another case, unexpected frequency shifts in a resonating crystal were traced to static charge buildup on the electrodes. (Ref. 16) Minor frequency changes were detected even on rotated Y cuts, such as the AT, with fields applied in the thickness direction, (Ref. 17) a result not consistent with the accepted crystalline symmetry of quartz. Substantial changes are expected even on these cuts if the field applied has large components in the X-direction. As this effect can be a possible source of problems in high stability, high precision crystals, an attempt was made to explain these effects in terms of the fundamental properties of quartz. The results of this study indicate that one must in general consider the effect of electrostatic fields on all of the properties of quartz, although the largest effect is on the effective elastic constants.

Theory

To a first approximation, the resonant frequency of the thickness mode of a quartz resonator is given by:

$$f_n = \frac{n}{2t}\sqrt{\frac{\bar{c}}{\rho}} \qquad (1)$$

where n is the overtone number, t is the crystal thickness, ρ is the density, and $\bar{c}$, the effective elastic constant for the acoustic mode of interest. If one differentiates (1) above with respect to some field variable and approximates the resultant partial differentials with their finite differences, one gets:

$$\frac{\Delta f}{f_0} = \frac{\Delta C}{2C_0} - \frac{\Delta \rho}{2\rho_0} - \frac{\Delta t}{t_0} \qquad (2)$$

where:

$$\Delta f \equiv f(E) - f(0) \quad \text{etc.} \qquad (3)$$

If the field variable in question is a uni-directional static electric field, applied to the electrodes of a thickness mode resonator, it can be shown that the strain in the direction of the applied field, and hence $\Delta t/t$, is given by: (Ref. 1)

$$\frac{\Delta t}{t} = E\left(l^3 - 3\,lm^2\right) d_{11} \qquad (4)$$

where l and m are the directional cosines of the applied field with respect to the X and Y axes, E is the magnitude of the applied field, and d_{11} is the piezoelectric constant.

More properly, one should also consider higher order effects which are dependent on higher powers of the applied field. Strongest of these should be the electrostriction effect which is proportional to the square of the applied field. (Ref. 2) However, all experimental runs to date show negligible departure from linearity with field strengths as high as 5×10^7 volts/meter. The conclusion is that second-order and higher effects may be ignored in any further calculations.

Since an electric field generally causes a change in the resonator thickness, one must also consider the possibility of a change in density due to an applied field. The change in density is equivalent to the change in volume of a unit mass. This change in volume can be related to a first approximation to the strain as shown:

$$\frac{\Delta V}{V} \simeq \left(S_{xx}' + S_{yy}' + S_{zz}'\right) \tag{5}$$

where each strain element represents a longitudinal strain in some orthogonal axis system. If one expresses the strain matrix as:

$$S = d_t E \tag{6}$$

it can be shown that (5) leads to the conclusion that no change in density occurs for an electrostatic field.

At this point, one can ignore any possible change in the effective elastic constant and examine how well a simple change in thickness accounts for the observed frequency shifts. From the above, the change due to thickness alone is independent of the type of acoustic mode. The fractional frequency change, and the direction of the frequency shift, are the same for all modes. Both of these are contradicted by experimental evidence as will be shown shortly.

It is easy to visualize that, for electric field strengths approaching the dielectric breakdown of quartz, quite high internal stresses may be generated. An attempt was made to see if, through the third-order elastic constants, these internal stresses could account for the observed frequency shifts. From the work of Thurston, Brugger, McSkimin, and Andreatch, (Ref. 3-6) the expected fractional frequency change expressed as a function of the third-order elastic constants is given by:

$$\frac{\Delta f}{f} = \frac{P}{2\omega} f_{ab}'(0)\left[N_a N_b + V_j V_k \left(2\omega S^T_{jkab} + N_r N_s S^T_{ipab} C_{jrksip}\right)\right] \tag{7}$$

where:

ω is the natural frequency;

N_a is the directional cosine with respect to the a^{th} axis of the natural direction of acoustic propagation;

U_j is the directional cosine of the natural particle displacements;

s^T_{jkab} is a component of the elastic compliance tensor;

C_{jrksip} is a third-order elastic constant;

t_{ab} is the second Piola-Kirchhoff stress tensor; and,

p is a scalar variable.

The expression t'_{ab} (0) indicates differentiation with respect to the scalar variable p, evaluated at p=0.

Ignoring possible rotational effects of the electric field, the t_{ab} tensor is, to first order, equivalent to the stress tensor given by:

$$T_{ab} = -e_{abq} E N_q \tag{8}$$

where E and N_q are as defined previously, and e_{abq} is an element of the piezoelectric tensor. Differentiating (8) with

respect to E and evaluating at E=0 gives:

$$T'_{ab}(0) \simeq T_{ab}'\big|_{E=0} = -e_{abq}N_q \qquad (9)$$

With the substitution of (9) into (7), one has a means of evaluating the change in frequency which includes in one expression changes in the thickness, density, and the elastic constants, and which also shows the necessary dependence on the type of acoustic mode. Unfortunately, using the third-order elastic constants of quartz yields results that are an order of magnitude lower than those observed experimentally. The conclusion is that the model used must be incorrect. The value of the elastic constant must have a stronger direct dependence on the electrostatic field strength than on the magnitude of the strain developed through the linear piezoelectric and elastic constants.

One can at this point examine all of the possible properties of quartz and attempt to determine the influence of an electric field upon them. Such a study leads to the conclusion that, to be complete, one should include at least one new third-rank tensor, which is related to the electro-optic effect, at least two new fourth-rank tensors which are related to the elasto-optic and electrostrictive effects, and at least two new fifth-rank tensors which may be related to the electro-acoustic effect. From the symmetry of quartz, this would involve the determination of at least thirty new independent constants. Rather than attempt this monumental task, we have adopted a phenomenological approach which was suggested and examined in a series of papers by Hruška. (Ref 7-11) This approach is valid only when the contributions from the piezoelectric constants are small enough to be neglected.

Recalling (2), (4), and the discussion following (6), one has:

$$\frac{\Delta f}{f_o} = \frac{\Delta \bar{C}}{2\bar{C}_o} - E\left(l^3 - 3lm^2\right) d_{11} \qquad (10)$$

Expanding the effective elastic constant in a MacLaurin's series using as the independent variable the field strength E, and keeping only the constant and linear terms in E, one has:

$$C(E) = C(0) + E\left(\frac{\partial C}{\partial E}\right)_{E=0} \qquad (11)$$

The partial derivative of the elastic constant with respect to the applied field forms a tensor of rank five. The change in a particular elastic constant may be expressed as:

$$\Delta C^E_{ijkl} = \delta_{ijklm} E_m \qquad (12)$$

where:

$$\delta_{ijklm} \equiv \left(\frac{\partial C^E_{ijkl}}{\partial E_m}\right)_{\substack{E=0 \\ T=0}} \qquad (12a)$$

All of the effects mentioned previously, with the exception of the thickness change, are assumed to be lumped together in the elements of this tensor. The actual form of the tensor is shown in Figure 1.

Using the above, and the values for the tensor elements as given by Hruška,[11] we attempted to predict our experimental results. As shown in Figure 2, agreement was fair in some cases, rather poor in others.

Experimental

Measurements were made on a variety of doubly-rotated crystal cuts at various temperatures and field strengths. All measurements were taken using the two thickness shear modes. The experimental setup used the test network shown in Figure 3. The network differs from the conventional 'pi' network in that the crystal is in a shunt arm instead of the series arm. This was done to simplify measurements in a temperature bath as it only requires one lead per crystal instead of the two required for the conventional series network.

Although this does not allow precise measurements of the anti-resonant frequency, measurements of the resonant frequency of zero phase may be made to a precision of better than one part in 10^8 The values of the DC blocking capacitor and the shunt inductor were chosen to

minimize their effects on the crystal being measured and yet provide protection for the input probes of the detector.

The rest of the test setup is identical to that described at previous symposia (Ref. 12-14) where phase changes detected by a vector voltmeter are used in a closed loop system to adjust the operating frequency so that the crystal is always operated at the resonant point. With a step function of voltate applied to a crystal, the total system rise time is less than 150 msec.

As seen in Figure 4, a 100 volt potential, approximately 500,000 volts/meter for this crystal, causes a frequency change of less than 3 parts in 10^6.

A further complication is due to a quasi-exponential decay of the observed frequency shift. In this case, less than twenty seconds after the field was applied, the net frequency shift decreased to less than 30 per cent of that observed when the field was first applied. The 1/e time constant of the decay was about 7 sec. for this crystal at 80° C. The combination of a small effect and a strong decay phenomenon affect the accuracy of ..ay measurements. Serious measurement errors can occur if measurements are made after the steady state condition occurs.

The total amount and the rate of decay have been observed experimentally to be strongly dependent on the temperature, material, and crystallographic orientation of the crystal.

A suitable explanation of the decay phenomenon, which seems to explain the experimental observations to date, lies in the ionic concentrations present in crystalline material. When the electrostatic field is applied, ions apparently migrate in the crystal toward the electroded surfaces setting up a static charge which effectively reduces the total field within the crystal.

The rate of migration depends on the mob..lity of the ions. Ionic mobility has a high temperature dependence. The experimental data reflects this as shown in Figure 5. With the exception of those points labelled as synthetic ('Syn #'), all of the data was taken using natural quartz thickness shear mode resonators. All of the natural crystals measured have a decay rate which is roughly proportional to:

$$e^{(0.57eV)/kT} \tag{13}$$

where T is in degrees Kelvin, and k is Boltzmann's constant. This indicates that the decay mechanism is essentially the same for all natural quartz crystals. The experimental activation energy of 0.57 electron volts is interesting because it lies within the range of measured activation energies for motions of positive ion vacancies. (Ref. 15)

Crystals made from optical grade synthetic quartz behave quite differently. As seen in Figure 5, a synthetic crystal must be at a temperature substantially higher than a natural crystal to exhibit the same decay rate. Although not shown, the synthetic crystals have one other property which indicates that the mechanism is slightly different, the net amount of decay for synthetics is substantially less for the same decay time constant than that of a natural crystal at the same orientation. As synthetic quartz is known to have a substantially lower impurity concentration than natural quartz the above supports our hypothesis as to the basic mechanism.

One can express or define a relationship between the observed decay times and the temperature as:

$$t = t_0 e^{(0.57eV)/kT} \tag{14}$$

where t_0 is a normalizing factor which is different for each cut. Plotting the value of t_0 versus the polar angle of the crystal orientation gives the results shown on Figure 6. 0° in this case is the X-Y plane; 90° is the Z-axis. As seen, as the crystal orientation approaches the X-Y plane, the decay time constant increases almost exponentially. t_0, which is inversely related to the mobility of the ions, and hence to the electrical conductivity of quartz, is a strong

function of the crystal polar angle. As the electrical conductivity along the Z-axis in quartz is known to be some 300 times greater than that along any path in the X-Y plane, this also supports the ionic impurity hypothesis.

Finally, at any given temperature, the amount of decay varies as much as fifty per cent from crystal to crystal for cuts of the same orientation. This also seems directly related to the impurity concentrations in the individual crystals.

Data Reduction

A complete set of data was obtained for twelve distinct cases which involved both thickness shear modes in seven different crystallographic orientations. Among the various cuts used were the RT and IT cuts. The crystals were plano-convex units with diameters of 6.35 mm. and thicknesses ranging from 0.18 to 0.38 mm. All measurements were made on the fundamental over a temperature range of -50 to +100°. The data obtained included the temperature, frequency, and amount and direction of frequency shift measured when the potential was first applied. Data was taken for both positive and negative potentials, and the final results averaged. All experimental data was expressed in terms of right-hand quartz. This is required since it can be shown that, for the model chosen, the fifth-rank tensor elements have the same magnitudes but opposite signs for left and right hand quartz. Although the original crystals were an unknown mixture of left and right hand quartz, suitable adjustments were made to the sign of the experimental data depending on the polarity of the potential measured when the crystals used were subjected to a longitudinal compressive stress.

In all cases, more than one crystal of each orientation was used. For the final data reduction, because of scatter in the data points and a spread of about 5 percent in the fractional frequency change from crystal to crystal in the same group at the same temperature, some data smoothing was done by fitting the experimental data to a third order polynomial. The resultant smoothed data was used as input for a non-linear determination of the eight independent tensor elements.

The final values determined are shown as a function of temperature in Figure 7. The influence of the third-order data smoothing is evident in this figure. However, the fantastically large temperature dependence of the tensor elements is not a function of data smoothing. An examination of the original raw data indicates that the linear temperature coefficient of the electrostatic field sensitivity approached values of 10,000 ppm. per °C. in one group of crystals, and averaged better than 1000 ppm per °C. for the rest of the crystals measured. For this reason, no attempt was made to compensate for the known variations with temperature of the elastic constants, the thickness, or the density. The temperature dependence observed is at least an order of magnitude greater than anything previously observed in quartz. Considering all measurement errors and the ability of the data presented in Figure 7 to reproduce experimental data, a tolerance of ± 20 per cent must be placed on the values shown.

Conclusion

The large temperature dependence of the fifth-rank polarizing tensor indicates that this is a property of quartz which cannot be explained in terms of the traditional properties of the material. Although the fractional frequency change per unit change in the field magnitude is in the parts in 10^{12} range, consideration must be given to this effect if one is interested in high precision, high stability crystals.

Acknowledgements

I wish to acknowledge the influence of the published work of Dr. Harel Hruska on the model and terminology used. I wish also to thank Mr. Don Hammond and Mr. Al Benjaminson of the Hewlett-Packard Laboratories for many discussions that led to the concepts expressed in this paper, and also Mr. Benjaminson for designing the test fixture used in all measurements.

References

1. J. F. Nye, "Physical Properties of Crystals", Oxford University Press, 1964, pp. 127-8.

2. W. G. Cady, "Piezoelectricity", McGraw-Hill, 1946, pp. 198-9.

3. K. Brugger, "Thermodynamic Definition of Higher Order Elastic Coefficients" Phys. Rev., Vol. 133, No. 6A, March 1964, pp. A1611-2.

4. R. N. Thurston and K. Brugger, "Third-Order Elastic Constants and the Velocity of Small Amplitude Elastic Waves in Homogeneously Stressed Media," Phys. Rev., Vol. 133, No. 6A, March 1964, pp. A1604-10.

5. K. Brugger, "Pure Modes for Elastic Waves in Crystals," J. Appl. Phys., Vol. 36, March 1965, pp. 759-68.

6. R. N. Thurston, H. J. McSkimin, and P. Andreatch, Jr., "Third-Order Elastic Coefficients of Quarts", J. Appl. Phys., Vol. 37, Jan. 1966, pp. 267-75.

7. K. Hruška, "The Influence of an ElectricField on the Frequency of Piezoelectric Cuts," Czech. J. Phys., B11, 1961, pp. 150-2.

8. K. Hruška, "An Attempt at a Phenomenological Interpretation of the Influence of a Polarizing Field on Piezoelectric Resonators," Czech. J. Phys., B12, 1962, pp. 338-53.

9. K. Hruška, "Tensor of Polarizing Correction Terms of Quartz Elastic Coefficients," Czech. J. Phys., B13, 1963, pp. 307-8.

10. K. Hruška and L. Janík, "Change in Elastic Coefficients and Moduli of -Quartz in an Electric Field," Czech. J. Phys., B18, 1968, pp. 112-6.

11. K. Hruška and V. Kazda, "The Polarizing Tensor of the Elastic Coefficients and Moduli for -Quartz," Czech. J. Phys., B18, 1968, pp. 500-3.

12. R. P. Grenier, "Technique for Crystal Resonance Measurements Based on Phase Detection in a Transmission Type Measurement System," Proc. 22nd Annual Symposium on Frequency Control, 1968.

13. C. Adams, J. A. Kusters, and A. Benjaminson, "Measurement Techniques for Quartz Crystals," Frequency Technology, Vol. 6, Aug. 1968, pp. 22-5.

14. M. E. Frerking, "Vector Voltmeter Crystal Measurement System," Proc. 23rd Annual Symposium on Frequency Control, 1969.

15. C. Kittel, "Introduction to Solid State Physics," Wiley, 1956, pg. 487.

16. A. Benjaminson, private communication

17. D. Hammond, C. Adams, and L. Cutler, "Precision Crystal Units," Proc. 17th Annual Symposium on Frequency Control, 1963.

δ_{111}	$(-1/2)(\delta_{111}+\delta_{221})$	δ_{131}	δ_{141}	0	0
	δ_{221}	$-\delta_{131}$	δ_{241}	0	0
		0	δ_{341}	0	0
			δ_{441}	0	0
				$-\delta_{441}$	$(1/2)(\delta_{241}-\delta_{141})$
					$(-1/2)(\delta_{111}+\delta_{221})$
0	0	0	0	$-\delta_{241}$	$(1/4)(\delta_{111}+3\delta_{221})$
	0	0	0	$-\delta_{141}$	$(-1/4)(3\delta_{111}+\delta_{221})$
		0	0	$-\delta_{341}$	$-\delta_{131}$
			0	δ_{441}	$(1/2)(\delta_{141}-\delta_{241})$
				0	0
					0
0	0	0	0	δ_{153}	0
	0	0	0	$-\delta_{153}$	0
		0	0	0	0
			0	0	$-\delta_{153}$
				0	0
					0

Figure 1 - Form of the Polarizing Tensor

$\frac{\Delta f}{fE} \times 10^{-12}$ METERS/VOLT

CUT	ORIENTATION		Δt/t	3rd ORDER ELASTICITY	HRUSKA CONSTANTS	EXPERIMENTAL VALUES
	PHI	THETA				
IT	20.0	34.2	-1.1	-.3 -2.7	21.8 -8.8	-4.0 2.5
LC	11.3	9.4	-1.2	1.6 -2.0	26.8 -19.9	-5.7 16.7

meters/volt

Figure 2 - Comparison of Theoretical Values Derived from Three Different Models with Experimental Data for Thickness Shear Modes

TEST FIXTURE

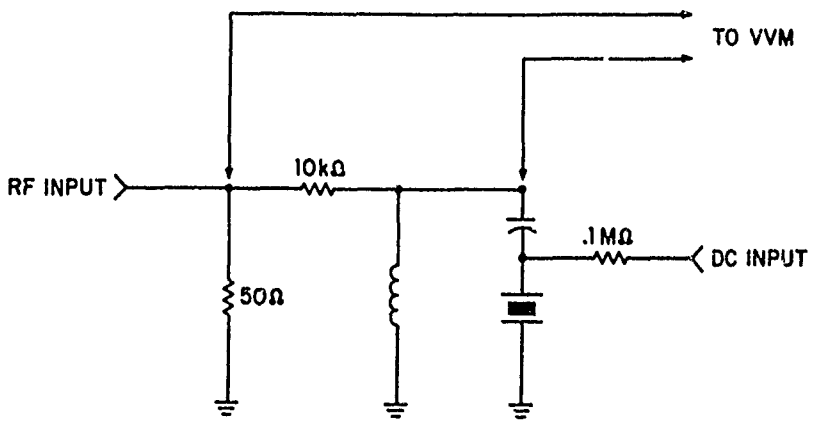


Figure - 3 Experimental Test Fixture

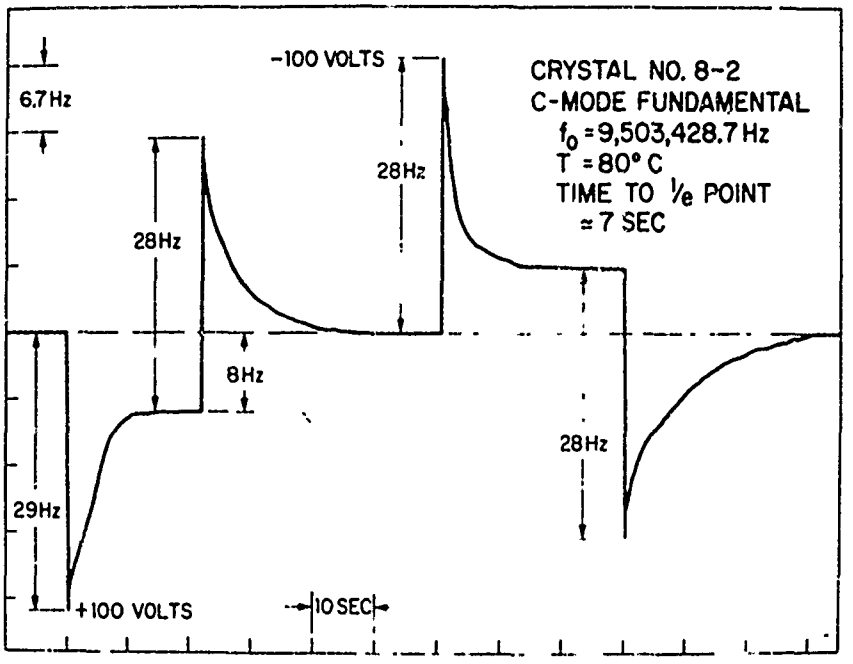


Figure 4 - Experimental Run showing Exponential Decay-Back Phenomena

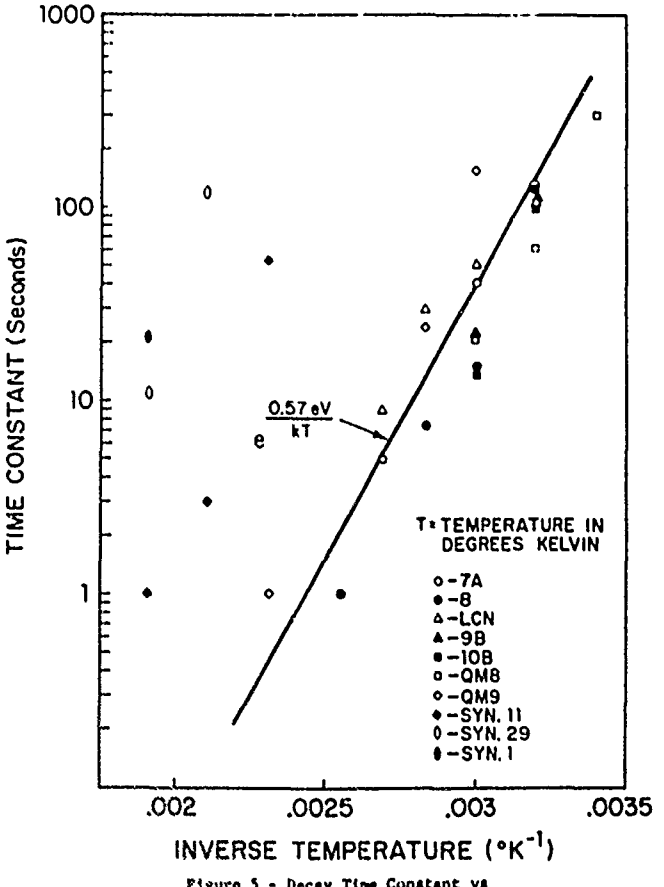


Figure 5 - Decay Time Constant vs Inverse Temperature

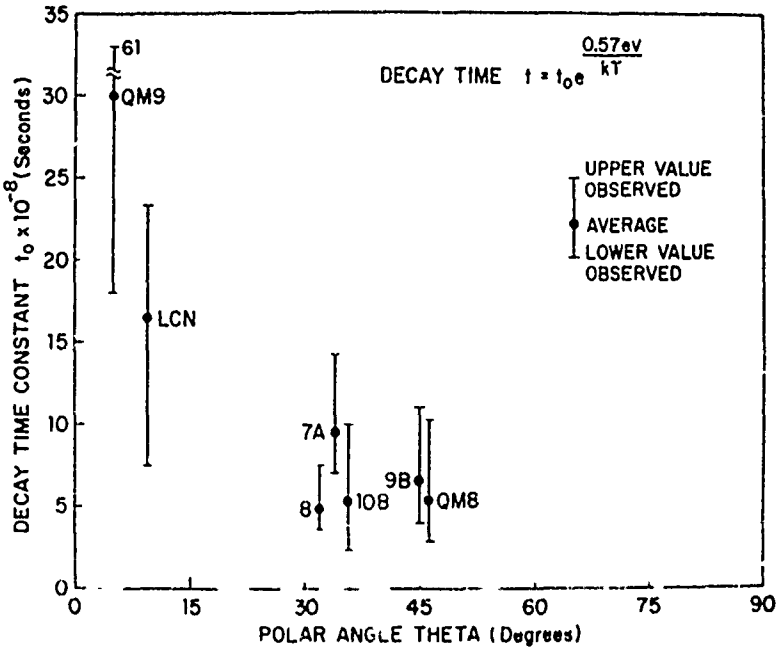


Figure 6 - Decay Time Normalizing Factor vs Polar Orientation Angle

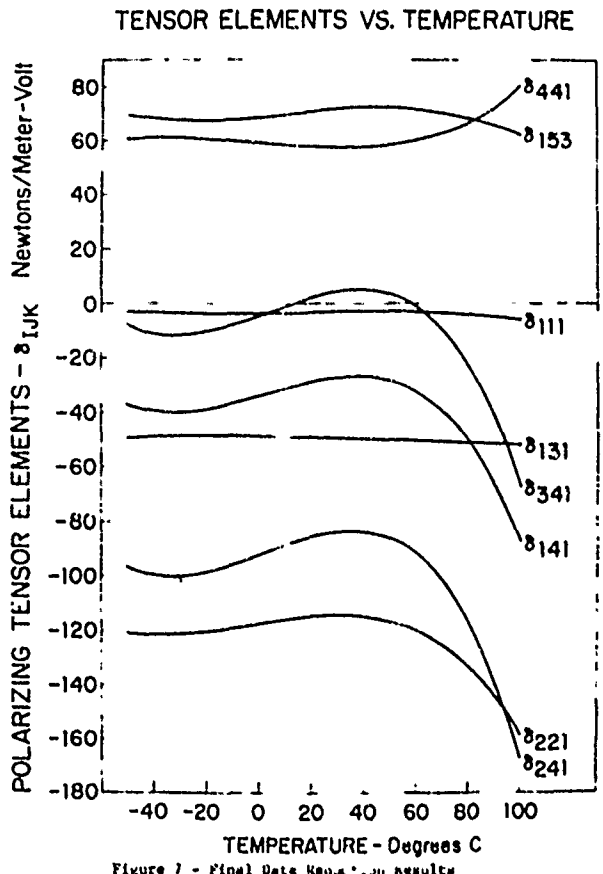


Figure 7 - Final Data Reduction Results
Values of the Polarizing Tensor Elements

SELECTED TOPICS IN QUARTZ CRYSTAL RESEARCH

Charles A. Adams, George M. Enslow,
John A. Kusters, and Roger W. Ward
Hewlett-Packard Laboratories
Palo Alto, California

Summary

Two recent areas of quartz crystal research have included the measurement of the temperature coefficients of the elastic stiffnesses of quartz and a study of the angular dependence of the temperature-frequency behavior of AT cuts.

The measurement of the temperature coefficients of the stiffnesses of quartz resulted in excellent agreement with Bechmann's (Ref. 1) value for c_{11}. Good agreement was seen for c_{14}, c_{44}, and c_{66}. Rather poor agreement was obtained for c_{13} and c_{33}. Locus lines of zero temperature coefficients of frequency are shown.

The study of the angular dependence of the temperature-frequency behavior of AT cuts resulted in experimental frequency-temperature data which deviated from a third-order curve by an RMS deviation of better than 1 part in 10^7. A determination was made of turn-over temperatures and inflection temperatures versus angle over the temperature range -35 to +100°C. Empherical equations were derived from which the turn-over and inflection temperature of an AT resonator may be calculated.

Part I: Measurement of the Temperature Coefficients of the Elastic Stiffnesses of Quartz

Introduction

The purpose of this study was to accurately determine the temperature coefficients of the elastic stiffnesses of quartz. The elastic stiffnesses are defined in (1) by

$$f_n = \frac{n}{2t}\sqrt{\frac{\bar{c}}{\rho}} \qquad n = 1,3,5,\cdots \qquad (1)$$

to first approximation, where t is the thickness of the resonator, ρ is the density of quartz, and $\bar{c}$ is a function of the six independent elastic stiffnesses, c_{ij}. The importance of this study lies in the fact that accurate values of the temperature coefficients of the elastic stiffnesses are necessary in order to predict important orientations in quartz; i.e., the RT and LC cuts; and to predict the frequency-temperature behavior of any cut.

Bechmann and associates[1] made significant progress in the specification of the temperature coefficients in 1962. Their values resulted from a third-order fit to data taken over the temperature range -196 to +170°C, on a large number of crystals whose accuracy of orientation angles was reported to be "better than 10' for the angle ϕ and better than 5' for the angle θ according to the manufacturer".

Experimental frequency-temperature curves often deviate significantly from a third-order curve, especially over large temperature ranges, thereby restricting the accuracy of any determination of the temperature coefficients of frequency, as defined in (2)

$$\frac{f - f_o}{f_o} = \sum_n T_f^{(n)} (T - T_o)^n \qquad (2)$$

$$T_f^{(n)} \equiv \frac{1}{n! f_o}\left(\frac{\partial^n f}{\partial T^n}\right)_{T=T_o}$$

where f_o is the frequency of the resonator at the reference temperature T_o, and

where $T_f^{(n)}$ is the n^{th} order temperature coefficient of frequency for a given resonator. For this reason a more accurate third-order fit to experimental data should result from taking frequency-temperature data over limited temperature ranges. Additional accuracy is obtained by knowing accurately the orientation angles of the crystals used.

Experimental

In order to determine the temperature coefficients of the elastic stiffnesses of quartz accurately, several crystal parameters must be measured to high precision. Among these are: crystal orientation angles, resonant frequencies, and temperatures.

Crystals

Angular orientation of a number of crystals to be studied was performed, using an optical dividing head and double-crystal refractometer techniques, to an accuracy of better than 15". The crystals were high-frequency, plano-convex units operating in thickness-shear mode. Diameter-thickness ratios of the blanks ranged from 20 to 40, thereby yielding quite thin units.

Temperature-Frequency Measurements

Temperature-frequency measurements were carried out with the resonators immersed in a temperature controlled oil bath. Temperature measurements were made using a quartz thermometer which was calibrated against a platinum resistance thermometer. Temperature accuracy was better than 0.01^oC over the temperature range -50 to $+100^oC$. Frequency measurements were made on the units to an accuracy of better than 1 part in 10^8.

Data Evaluation

The raw data, consisting of a number of temperature-frequency data pairs per crystal, was fit to a third-order curve over the temperature range -20 to $+70^oC$. The RMS deviation of the data from the calculated curves ranged from a few parts in 10^7 to 1 part in 10^8, with respect to the resonant frequency of the crystals. The coefficients of the calculated curve for each crystal are the experimental values of the n^{th} order temperature coefficients of frequency for that crystal.

Determination of the Temperature Coefficients

Knowing the values of the orientation angles for the crystals of this study, the components of the effective stiffness in terms of the elastic stiffnesses, c_{ij}, were determined. The theoretical value of the effective elastic stiffness of a particular orientation, however, is dependent upon an accurate knowledge of the thickness of the crystal. Due to the thinness of the units used, their thickness could not be measured accurately enough to be able to predice new values of c_{ij}. For this reason the thicknesses of our crystals were normalized to the thicknesses calculated from Bechmann's stiffnesses at 20^oC. The effects this had upon the determination of the new temperature coefficients is discussed later.

Using the relationships between the n^{th} order temperature coefficients of frequency and elastic stiffness, the new temperature coefficients were calculated. For brevity, only the first-order equation is shown in (3)

$$2T_f^{(1)} = T_c^{(1)} - T_\rho^{(1)} - 2T_\theta^{(1)} \qquad (3)$$

where $T_\rho^{(1)}$ is the first-order temperature coefficient of density, $T_c^{(1)}$ is the first order temperature coefficient of the elastic stiffness, and

$T_e^{(1)}$ is the first-order temperature coefficient of thickness for quartz. Higher-order relationships are given in Appendix I.

Table I gives a comparison between the newly determined temperature coefficients and Bechmann's values. Note the excellent agreement in the temperature coefficients for c_{11}, and the generally good agreement in c_{14}, c_{44}, and c_{66}. The temperature coefficients for c_{13} and c_{33} show quite poor agreement. This type of behavior is to be expected. For example, c_{11} is the only elastic stiffness which influences x-cut crystals. As such, it is a well understood and widely studied cut. c_{14}, c_{44}, and c_{66} are the elastic stiffnesses which relate to all rotated y-cuts, and in particular the AT and BT cuts. These, again, are widely used and rather well understood cuts. On the other hand, relatively few common crystal cuts show strong dependence on c_{13} and c_{33}, making these stiffnesses less studied and more difficult to measure. For these reasons, the overall agreement with Bechmann is actually quite good.

Due to the accuracy of the orientation of the crystals studied, the temperature stability and measurement resolution of the crystals' environment, the frequency measuring accuracy, and the excellence of the third-order fits to the raw data, these new values for the temperature coefficients of the elastic stiffnesses should be substantially more accurate than previous determinations.

Locus Lines of Zero Temperature Coefficients of Frequency

Figures 1-7 show orthographic projections of the primitative region of quartz with the locus lines of zero temperature coefficient of frequency marked for the fast-shear, slow shear, and extentional modes of resonance, as calculated from the new values for the temperature coefficients, and from Bechmann's values. The left-hand side of the projection represents the y-z plane and the bottom is the x-y plane. This provides a graphic comparison of the two sets of values. Note that there is excellent agreement in the prediction of the general vicinity of the zeros of the temperature coefficient of frequency, and only small disagreement between them as to the exact location of the zero. Orientations for some of the crystals used in this study are indicated.

One point is of interest--every temperature coefficient of frequency exhibits a zero value at $\phi = 60^\circ$, $\theta = 23\text{-}24^\circ$. This point represents a singularity in quartz, as the gradients of the temperature coefficients of frequency are very steep in this region. This is physically related to the extentional, fast-shear, and slow-shear modes being degenerate at that orientation.

Effects of Assuming Bechmann's Values for the c_{ij}'s

In order to determine the dependence of the new temperature coefficients of stiffness upon Bechmann's values for c_{ij}, c_{44} was intentionally increased by varying amounts. These "new" values of c_{44} were then used to recalculate the temperature coefficients. The result of this study showed that the change in c_{ij} with respect to c_{44} was linear at least out to c_{44} 3% larger than Bechmann's value. From this, it is evident that the determination of the temperature coefficients of the stiffnesses is linearly dependent upon the accuracy to which the stiffnesses are known.

A complete theory of the temperature behavior or quartz is more complicated than the theory considered in this study. A more complete theory considers also the temperature coefficients of the dielectric constants and the piezoelectric stress constants, among other things. Such a consideration, however, increases the complexity of the theory many-fold, to a point far beyond the scope of this study.

Part II: The Frequency-Temperature Behavior of 5MHz AT Resonators

Introduction

The purpose of this study was to make very high precision measurements on a group of AT resonators. High precision measurements on three parameters of the resonators were basic to this study; angular orientation, frequency, and temperature.

Preparation of the Resonators

The resonators under study were 5MHz, 5th overtone, plano-convex, natural quartz plates. Their diameter was 15mm and the radius of curvature of the convex surface was 50.8mm. Angular orientation of the resonators was performed on an optical dividing head, using double-crystal refractometer techniques, to an accuracy of 5". It was decided not to polish the crystals in order to prevent any deviations in the orientation angle from occuring. Instead, the crystals were lightly etched before final processing.

Testing of the Resonators

The resonators were placed in a temperature controlled oil bath whose temperature was programmed to increase at a constant rate of either 3.3 or $4.2^{o}C$/hr, depending upon the particular resonators under test. This small rate of change of temperature provided a quasi-adiabatic environment for the crystals. The temperature was monitored with a quartz thermometer which was calibrated against a platinum resistance thermometer. Resolution was $0.1X10^{-3}\ ^{o}C$ with a maximum error of $5X10^{-3}\ ^{o}C$ over the range -25 to $+100^{o}C$. The frequency of the resonators was measured to 0.005Hz using a computing counter.

Frequency-Temperature Data

Figure 8 shows a raw data plot for one of the crystals used in this study. This plot has had no smoothing--it is actual raw data. The scaled frequency represents the difference between the crystal's frequency and 5MHz. When this raw data was fit by a third-order curve, the resulting calculated curve had an RMS deviation from the raw data of 6 parts in 10^{8} with respect to the actual frequency of the crystal. Similar data was taken on a number of other crystals with the worst RMS deviation being 1 part in 10^{7}. This degree of accuracy in fitting frequency-temperature data with a third-order equation gives added credence to the idea that AT cuts are best represented by a third-order theory.

Angle of Cut Versus First-Order Temperature Coefficient of Frequency

The first-order temperature coefficient of frequency, as defined in (2) for n = 1, determines the turn-over temperatures of AT cuts. As such, it is a sensitive indicator of the precision of preparation of a group of crystals, to the extent that quartz is homogeneous. Since the angular orientation was performed on the crystals of this study to 5", a measure of the RMS deviation of the first-order temperature coefficients of frequency versus angle from a third-order curve (Ref. 1) should be a measure of how well the processing of the crystals was done and of the uniformity of natural quartz. Such a measurement was made, with the result being an RMS deviation of 8" of angle. Hence, one must conclude that, to within the limits of our ability to prepare quartz resonators, natural quartz is exceptionally uniform and homogeneous.

Turn-Over Temperatures

Figure 9 shows a plot of the upper and lower turn-over temperatures of the crystals versus their orientation angles. The turn-over temperatures were calculated from the third-order equations mentioned earlier. Note the scatter in the experimental points (actually some points are 2 points on top of one another). When one takes into consideration the slope of the curve at each data point, the scatter is seen to lie within a 5" band about the calculated curve, indicative of the angular orientation error, and more importantly, of the homogeneity of natural quartz. The equation under the figure is the em-

pherical equation of the calculated curve. This equation predicts a value of θ, the polar angle, for any temperature between -35 and +100°C, which is accurate to better than 30". The minimum of the curve, as calculated from the equation, occurs at 35° 19' 9", at a temperature of 32.5°C. Hence, the data predicts that for AT crystals of the type used in this study, oriented at an angle below 35° 19', there will be no turn-over in the frequency-temperature behavior of the crystal.

Inflection Temperatures

The inflection temperatures of the curves are plotted against angle in Figure 10. The slope of the line is 18.6°C/degree of angle, which agrees well with similar slopes calculated from graphs published by Bechmann (Ref. 1) and Chi, (Ref. 3) and with the slope as calculated from the angular dependence of the temperature coefficients of frequency of AT cuts.

Conclusions

In conclusion, the results of this study have 1) verified the third-order behavior of AT resonators to within an RMS value of better than 1 part in 10^7, 2) provided a means of predicting the turn-over temperature for any precisely oriented AT 5MHz resonator, and 3) determined an analytical expression for the inflection temperature of the crystals, while at the same time providing further evidence of the homogeneity of natural quartz.

Note:

It should be pointed out that the information presented here is only exactly applicable to crystals having the same physical characteristics as the ones in this study; i.e., same diameter, thickness, contour, etc. However, the information should be of use in predicting the general behavior of any AT resonator.

Appendix I

Relationships between the temperature coefficients of frequency and the temperature coefficients of thickness, stiffness, and density:

$$2T_f^{(1)} = T_c^{(1)} - T_\rho^{(1)} - 2T_\theta^{(1)}$$

$$2\left[T_f^{(2)} - \frac{1}{2}\left(T_f^{(1)}\right)^2\right] = T_c^{(2)} - T_\rho^{(2)} - 2T_\theta^{(2)} - \frac{1}{2}\left[\left(T_c^{(1)}\right)^2 - \left(T_\rho^{(1)}\right)^2 - 2\left(T_\theta^{(1)}\right)^2\right]$$

$$2\left[T_f^{(3)} - 2T_f^{(2)}T_f^{(1)} + \frac{1}{3}\left(T_f^{(1)}\right)^3\right] = T_c^{(3)} - T_\rho^{(3)} - 2T_\theta^{(3)} - \left[T_c^{(2)}T_c^{(1)} - T_\rho^{(2)}T_\rho^{(1)} - 2T_\theta^{(2)}T_\theta^{(1)}\right] + \frac{1}{3}\left[\left(T_c^{(1)}\right)^3 - \left(T_\rho^{(1)}\right)^3 - 2\left(T_\theta^{(1)}\right)^3\right]$$

Where $T_\rho \equiv -(2\alpha_x + \alpha_z)$, α_x and α_z are the expansion coefficients.

Appendix II

To determine the temperature coefficients of the stiffnesses, it is necessary to assign values to the density, dielectric constant, etc., of quartz. In this study, the following values, taken from Bechmann (Ref. 1) were used:

Piezoelectric stress constants:

$$e_{11} = 0.171 \text{ C/m}^2$$

$$e_{14} = 0.0403 \text{ C/m}^2$$

Dielectric constants:

$$\epsilon_{11}^T = \epsilon_{22}^T = 39.97 \times 10^{-12} \text{F/m}$$

$$\epsilon_{11}^S - \epsilon_{11}^T = -0.76$$

$$\epsilon_{33}^T = 41.03$$

$$\epsilon_{33}^S - \epsilon_{33}^T = 0$$

Coefficients of thermal expansion:

$$\alpha_{11}^{(1)} = \alpha_{22}^{(1)} = 13.71 \times 10^{-6}/°C$$

$$\alpha_{33}^{(1)} = 7.48$$

$$\alpha_{11}^{(2)} = \alpha_{22}^{(2)} = 6.5 \times 10^{-9}/(°C)^2$$

$$\alpha_{33}^{(2)} = 2.9$$

$$\alpha_{11}^{(3)} = \alpha_{22}^{(3)} = -1.9 \times 10^{-12}/(°C)^3$$

$$\alpha_{33}^{(3)} = -1.5$$

Density:

$$\rho = 2.65 \times 10^3 Nm^{-4}S^2$$

Temperature coefficients of density:

$$T_\rho^{(1)} = -34.92 \times 10^{-6}/°C$$

$$T_\rho^{(2)} = -15.9 \times 10^{-9}/(°C)^2$$

$$T_\rho^{(3)} = 5.30 \times 10^{-12}/(°C)^3$$

Bibliography and References

(1) R. Bechmann, A.D. Ballato, and T. J. Lukaszek, "Higher-order temperature coefficients of the elastic stiffnesses and compliances of alpha-quartz," Proc. IRE, Vol. 50, pp 1812-1822; August, 1962.

(2) A. W. Warner, "Design and performance of ultraprecise 2.5-mc quartz crystal units," Bell Sys. Tech. J., Vol. 36, pp. 1193-1216; Sept., 1960.

(3) A. R. Chi, "Some resonator properties of synthetic and doped snythetic quartz," 1959 IRE Convention Record. pp. 70-75.

(4) V. E. Bottom, The Theory and Design of Quartz Crystal Units, The McMurry Press, Abilene, Texas; 1968.

TEMPERATURE COEFFICIENTS OF THE STIFFNESSES OF QUARTZ

		$C^D(10^9 N/M^2)$	$A(10^{-6}/°C)$	$B(10^{-9}/°C^2)$	$C(10^{-12}/°C^3)$
C_{11}	BECHMANN	87.49	-48.5	-107	-70
	NEW		-49.6	-107	-74
C_{13}	BECHMANN	11.91	-550	-1150	-750
	NEW		-651	-1021	-240
C_{33}	BECHMANN	107.2	-160	-275	-250
	NEW		-192	-162	67
C_{14}	BECHMANN	-18.09	101	-48	-590
	NEW		89	-19	-521
C_{44}	BECHMANN	57.98	-177	-216	-216
	NEW		-172	-261	-194
C_{66}	BECHMANN	40.63	178	118	21
	NEW		167	164	29

Table I: A comparison between Bechmann's temperature coefficients of the elastic stiffnesses of quartz and the newly determined values.

Figure 1

LOCUS OF ZERO FIRST ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

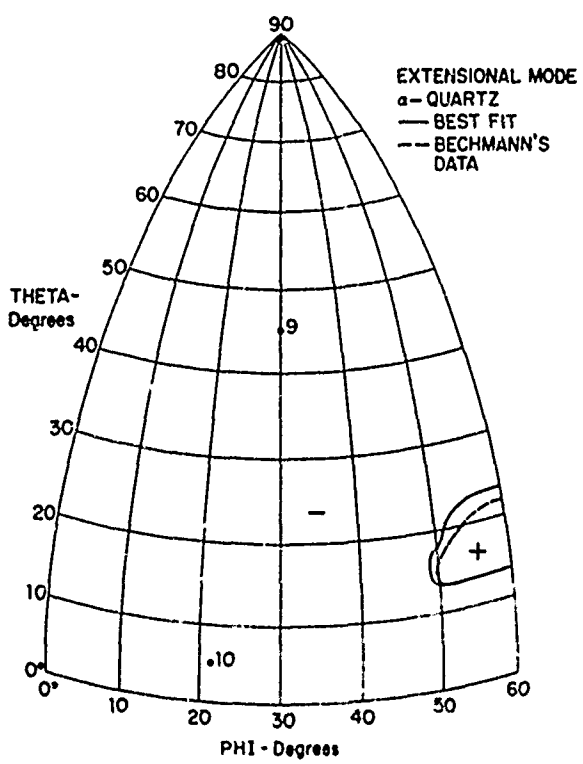


Figure 2

LOCUS OF ZERO FIRST ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

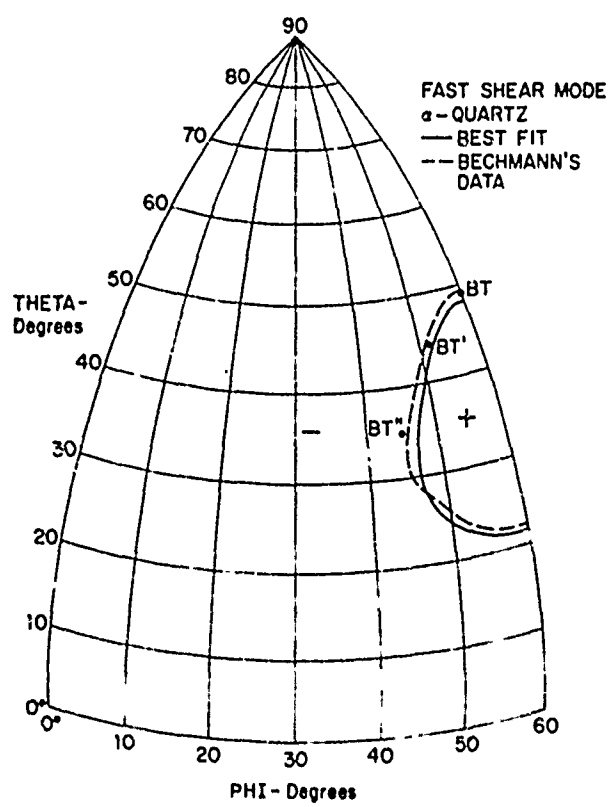


Figure 3

LOCUS OF ZERO SECOND ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

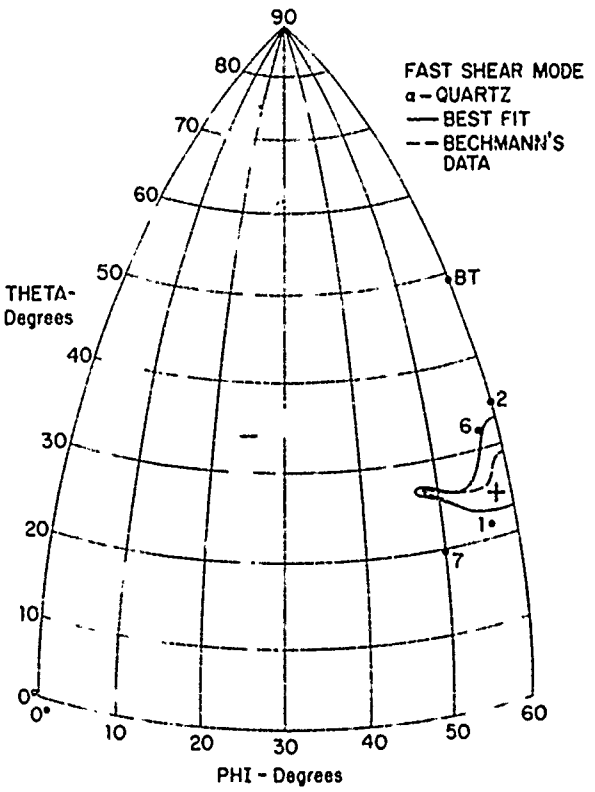


Figure 4

LOCUS OF ZERO THIRD ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

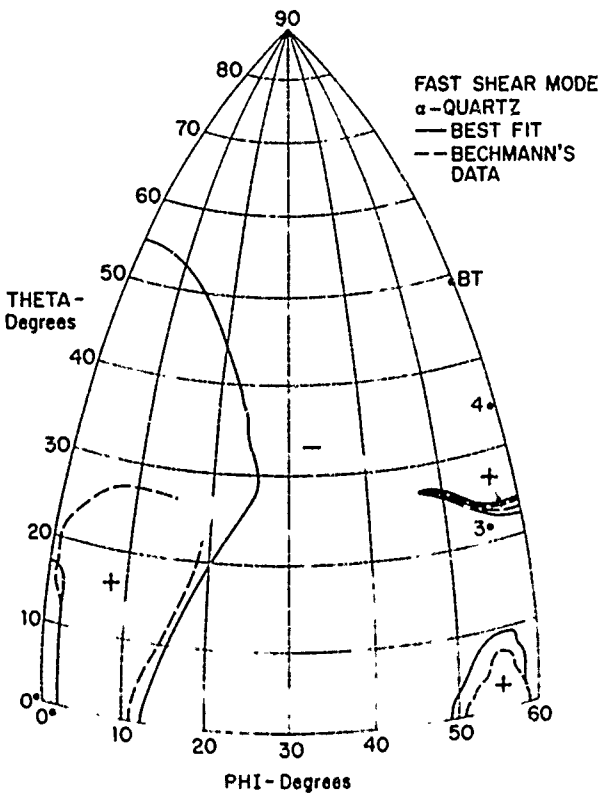


Figure 5

LOCUS OF ZERO FIRST ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

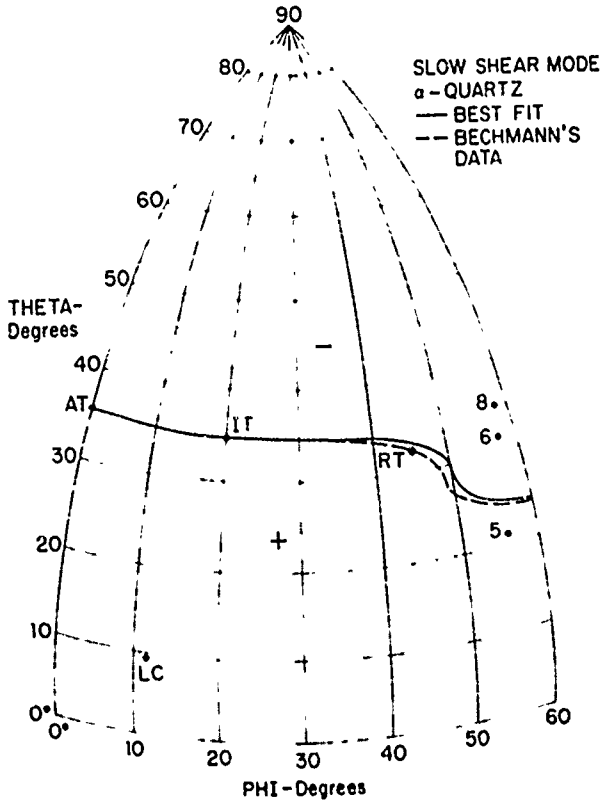


Figure 6

LOCUS OF ZERO SECOND ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

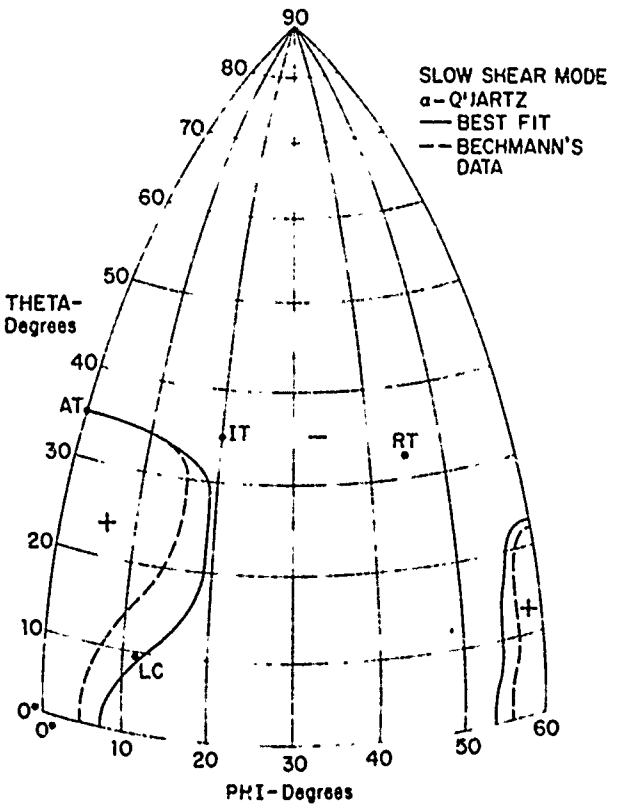


Figure 7

LOCUS OF ZERO THIRD ORDER FREQUENCY-TEMPERATURE COEFFICIENTS

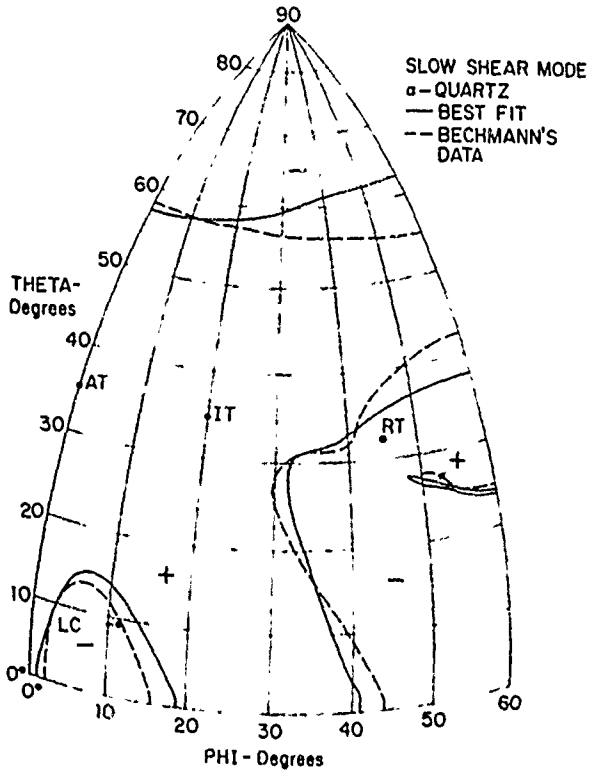


Figure 9

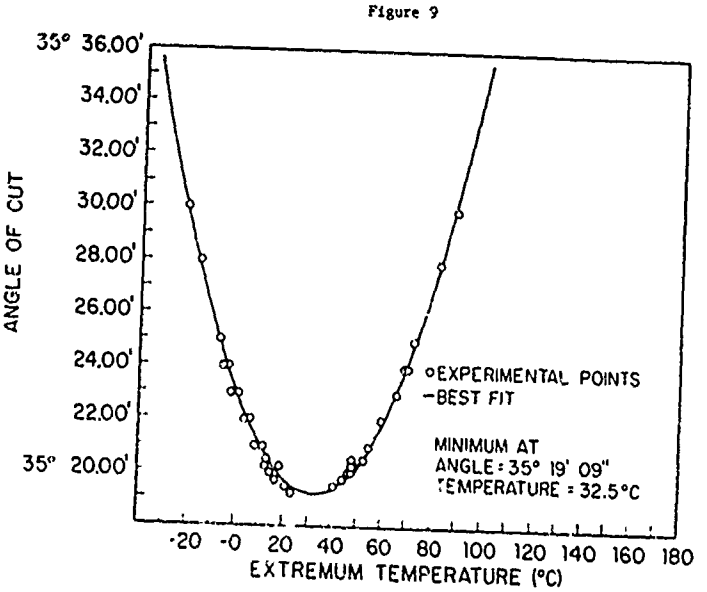


FIG.2: LOCUS OF TURN-OVER TEMPERATURE FOR AT CUTS
$\theta = 35\,3831\,(1\text{-}1.10881 \times 10^{-4}T + 1.7049 \times 10^{-6}T^2)$ Degrees, T IN °C.

Figure 8

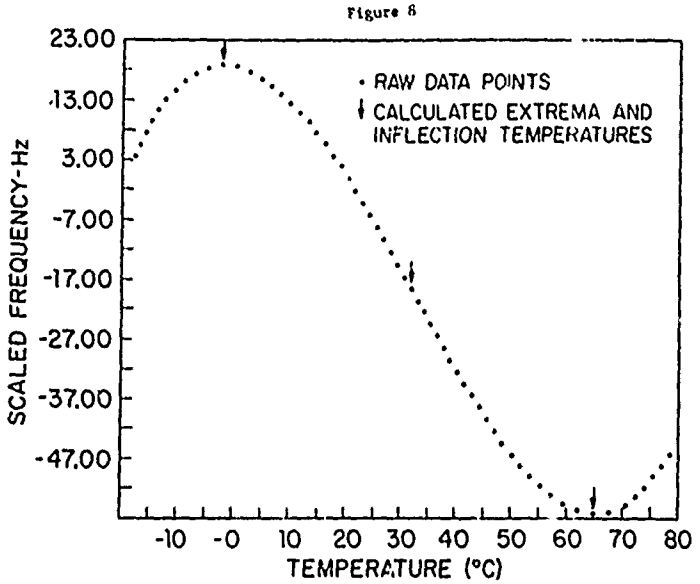


FIG.1: FREQUENCY-TEMPERATURE BEHAVIOR OF AN AT CRYSTAL

Figure 10

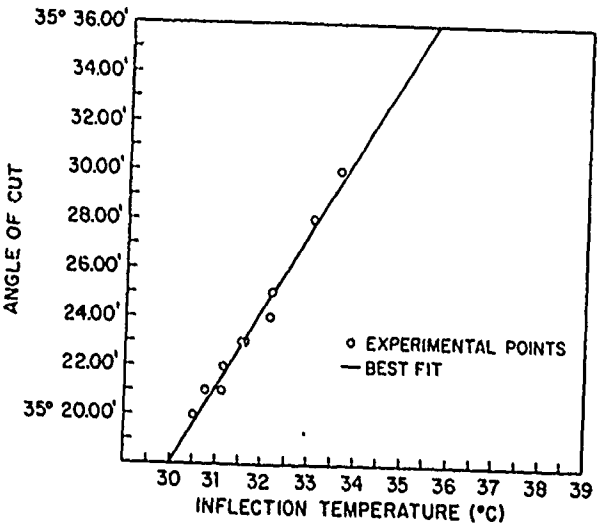


FIG. 3: $T_j = -635.664\,(1\text{-}2.9665 \times 10^{-2}\theta)$ °C
θ IN DEGREES

DEFECTS AND FREQUENCY MODE PATTERNS IN QUARTZ PLATES

E. W. Hearn & G. H. Schwuttke
International Business Machines Corporation
East Fishkill Laboratories
Hopewell Junction, New York 12533

Summary

Internal defects in quartz plates and their influence on the resonance vibrations of A. T. cuts are investigated. Crystal defects and frequency mode patterns are displayed through x-ray topography while the 5 MC frequency domain of the A. T. cut plates is explored through frequency spectrum analysis. Investigation of 12 different oscillators showing varying amounts of crystal defects are correlated with the frequency spectrum of the plates. A definite relation between the number of peaks in the spectrum of the oscillator and the defect concentration in the plate is established. The highest number of peaks is observed for the "perfect" quartz plate.

Introduction

The purpose of this paper is to relate quartz crystal perfection to oscillator efficiency. This is done by observing the effects of crystallographic defects in quartz on the frequency spectra. X-ray diffraction topography displays first the perfection of the quartz plate and then records the modes of oscillation of the plate at various frequencies in the spectra. Through this sequence a correlation is established between defects and oscillator efficiency.

The crystal structure of quartz is more complicated than the simple diamond structure of silicon. Consequently, a much larger variety of crystal defects is expected in quartz compared to silicon. This expectation has been confirmed by the x-ray topography work of Lang and Miuscov.[1] They have shown that defects in natural quartz, such as dislocations and two dimensional "sheet defects" or 'fault surfaces" are quite common. A typical example of such a fault is shown in Fig. 1.

A recent study of defects in synthetic quartz by Spencer and Haruta[2] is also based on x-ray topographic display of quartz defects. They observe "line defects" as the predominant faults in synthetic quartz through the use of high resolution topography. Lang and Miuscov[1] identified these "line defects" as dislocations. An example of dislocations in synthetic quartz is shown in Fig. 2.

Spencer and co-workers have measured the acoustic loss in natural[3] and synthetic[2] A. T. cut plates and have attempted a correlation with the crystallographic defect density in the plate. The Q of the 5 MC oscillator was measured by the logarithmic decrement technique and the defect density was determined by the x-ray topography technique. For natural quartz plates they arrived at the conclusion that no correlation exists,[3] and for synthetic plates they found a somewhat qualitative correspondence between defect density and Q measurements.[2] Published work on this subject indicates[2,3] that there seems to be some kind of influence of crystal defects on the oscillating efficiency of a resonator, but this influence is not clear cut and a precise correlation between defect density and oscillator efficiency cannot be established.

In our investigations we approached the problem in the following way: 12 commercially manufactured doubly beveled natural quartz oscillators, A. T. cuts, at 5 MC were selected. Examples of the crystal perfection of these 12 oscillators are displayed in Fig. 3. A variety of standard defects found commonly in natural quartz, such as the two dimensional fault planes--stacking fault type--are characteristic for all 12 plates. The topographs are recorded at room temperature. The crystals are mounted on the top graphic camera as received from the vendor; however, the caps are removed for topography. The topographs display also the metallurgy and the holder as seen at two positions around the periphery as a shadow.

For every quartz plate we explored the 5 MC frequency domain through the use of frequency spectrum analysis. Typical examples of such frequency spectra showing the A. C. rectified impedance vs. frequency are shown in Figs. 4 A and 4 B. We observe the fundamental mode at 5 MC and in addition a number of anharmonic modes. Note that the oscillators shown in Figs. 4 A and 4 B have a different number of modes. The number of modes of the sample shown in Fig. 4 A is 36 and in Fig. 4 B is 34.

For every single mode vibration visible in the frequency spectrum, we recorded the corresponding x-ray topograph, even the small peaks

result in good topographs. In other words, the oscillating quartz crystal was topographed at every single frequency displayed in the spectrum, thus the actual vibrational pattern of the quartz plate was recorded and correlated with the frequency spectrum. An example is shown in Fig. 5. This figure shows the topograph of the non-oscillating crystal at the top and all the possible vibrational states of this A.T. cut crystal in its frequency domain. The electrical vibrational states are evidently characterized by a characteristic number of mechanically vibrational plate modes. For instance, the fundamental mode is singular, while all the other satellite frequencies display a characteristic pattern of their own. For example, the 131 mode is indicated as number 4 in Fig. 5 and the 113 mode is number 6 in Fig. 5. While in the 131 mode the first number "1" refers to the n^{th} odd harmonic number of the crystal, the second number "3" refers to the number of anti nodes in the X_2 direction and the last number "1" refers to the number of anti nodes in the X_1 direction. X_1 is the vertical direction through the crystal and X_2 is the horizontal direction through the crystal. Figure 6 shows an enlargement of the 131 mode.

According to the number of vibrational states visible in the topograph, the frequency domain can be organized into a matrix. This is shown in Fig. 7. The numbers on the horizontal axis of this matrix correspond to the number of anti nodes that appear in the x-ray topograph in the X_1 direction of the crystal while the ordinate of the matrix is formed by the number of anti nodes in the X_2 direction of the crystal. The even row marked "rare" occurred only one time in the 12 crystals investigated. Each matrix element designates the number of plate-modes characteristic for a particular frequency domain of the A.T. cut. It follows that the number of matrix elements is identical with the number of frequency peaks in the frequency domain of the oscillator. Therefore the matrix describes uniquely all possible frequency levels of the oscillator.[4]

We have measured the number of matrix elements for each single A.T. cut using frequency analysis and x-ray topography. Two examples of a matrix for individual crystals are shown in Figs. 8 A and 8 B. In the V.F.-7 matrix each element box having a number corresponds to a certain frequency of the spectrum and each number in the element box is the number of dark areas, or anti nodes, in the topograph in all directions. Figure 9 is an enlargement of pattern number 27 from Fig. 5. Although it is complicated by mode coupling, it follows from its matrix position that it is the 177 mode.

Based on these measurements, we have ordered the 12 oscillators into a group shown in Fig. 10. It is seen that the A.T. cut with the highest number of matrix elements is the first plate in the group and the plate with the lowest number of matrix elements is the last plate in the group. The number at the left side of each topograph shown in Fig. 10, is the number of peaks in the corresponding spectrum.

The quality as determined by the crystallographic defect density of the oscillators goes from best to worst and, correspondingly, the number of frequency peaks in each spectrum goes from maximum to minimum. This then would indicate that the perfection of the crystalline oscillator can be correlated with the number of peaks in its spectrum.

Spencer et al have reported a correlation in "Q" measured by the logarithmic decrement technique as a function of defect density in synthetic quartz.[2] However, they report no noticeable correlation in natural quartz.[3] We have ordered the oscillator plates according to their "Q" as measured by the logarithmic decrement technique and again related this to the perfection of the plates as seen in the topographs. This group is shown in Fig. 11. The corresponding "Q" values are given in Fig. 12. This group is ordered by decreasing relative "Q" values as measured by the log decrement method going from best to worst. Generally, this now follows a similar pattern as shown in Fig. 10. However, we observe some interesting exceptions. The most notable one is that of V.F.-12. V.F.-12 is the most perfect of all the crystals topographed. It also has the highest number of peaks in its spectrum; however, the "Q" value is quite low. One can also notice in the topograph that the edge of the crystal is cracked off. In addition, the number of thickness fringes due to beveling is greater than in other crystals. On the other hand, the spectrum of V.F.-5 has the lowest number of frequency peaks and, in good agreement, an extremely low relative Q, and the highest defect density. Similar observations can be reported for other "misfits" in comparing the frequency spectra and the logarithmic decrement measurements. This suggests then that the frequency peak measurement is a very sensitive measure of the crystal perfection, while the logarithmic decrement values are also influenced by variables due to the manufacturing process.

In synthetic crystals--specifically the ones investigated by Spencer et al[2]--the dislocation density is the overriding factor; therefore, Spencer et al[2] find a correlation between the logarithmic decrement technique and the crystal perfection.

Contrary, Spencer's natural crystals[3] are much more perfect than his synthetic crystals, therefore, the logarithmic decrement Q measurements are influenced by the manufacturing conditions and no correlation is found.[5]

One interesting aspect of the frequency peak measurements is that it can be used to monitor or control manufacturing variables. Once the frequency peak numbers are put into matrix form, the manufacturing parameters can be monitored to see if they preferentially affect the matrix in any way.

It is also possible to predict the mode to be observed on the topograph as well as the generating frequency by merely filling in the missing elements of the matrix.

Summary

The spectra technique as presented in this paper seems to be capable of serving as a guide in evaluating the performance of A. T. cut oscillators as determined by their crystallographic perfection. Good correlation between "oscillator efficiency" and crystal perfection has been achieved for 12 A. T. cut natural quartz plates. However, more detailed measurements of quartz plates are required to further substantiate these findings.

Acknowledgement

This work is sponsored by the Air Force Cambridge Research Laboratories, Bedford, Massachusetts, U.S.A., under Air Force Contract F19-(628)-68-C-0196.

References

(1) A. R. Lang and V. F. Miuscov, J.A.P. 38, 2477, (1967).

(2) W. J. Spencer and K. Haruta, J.A.P. 37, 549, (1966).

(3) W. J. Spencer and W. L. Smith, J.A.P. 37, 2557, (1966).

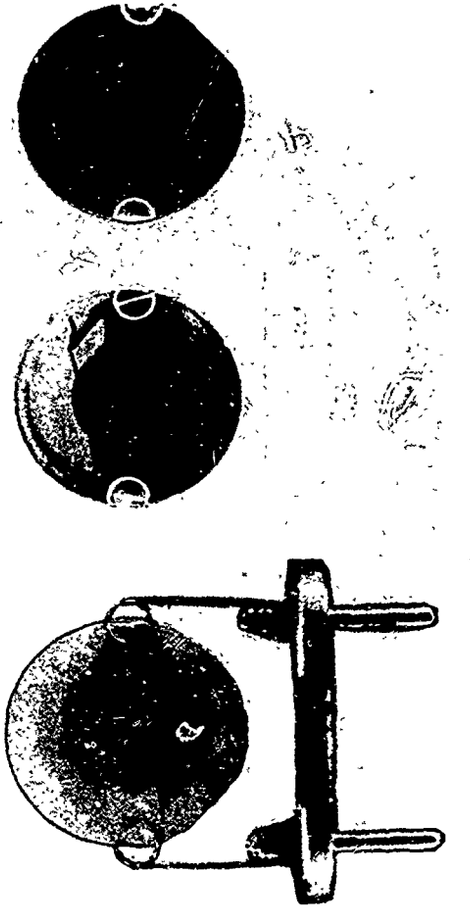
(4) Note that in order to obtain the matrix, it is necessary to overdrive the crystals. A drive level + 13 db is used.

(5) Compare Fig. 2 in reference 2 and Figs. 9, 10, 11 and reference 3.

Fig. 1. Typical example of defects in natural quartz.

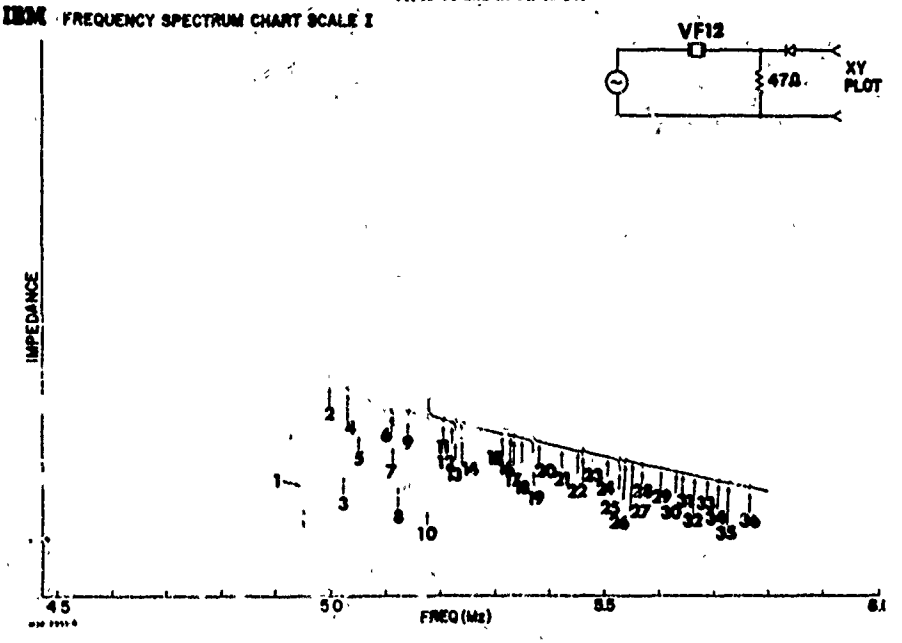
Fig. 2. Typical example of defects in synthetic quartz.

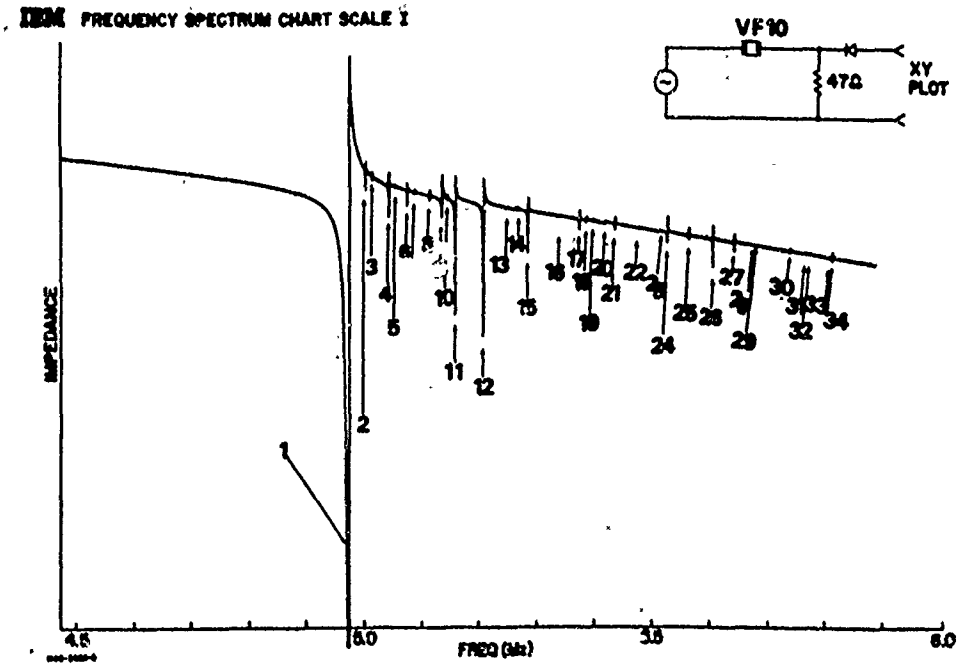
Fig. 3. Photograph (a) and topographs (b, c) of typical units investigated



Reproduced from best available copy.

Fig. 4. A & B Frequency spectra of V.F.-12 (A) and V.F.-10 (B) of harmonic and anharmonic modes around 5 MHz. The mode number in 4 A is 36 and in 4 B is 34.





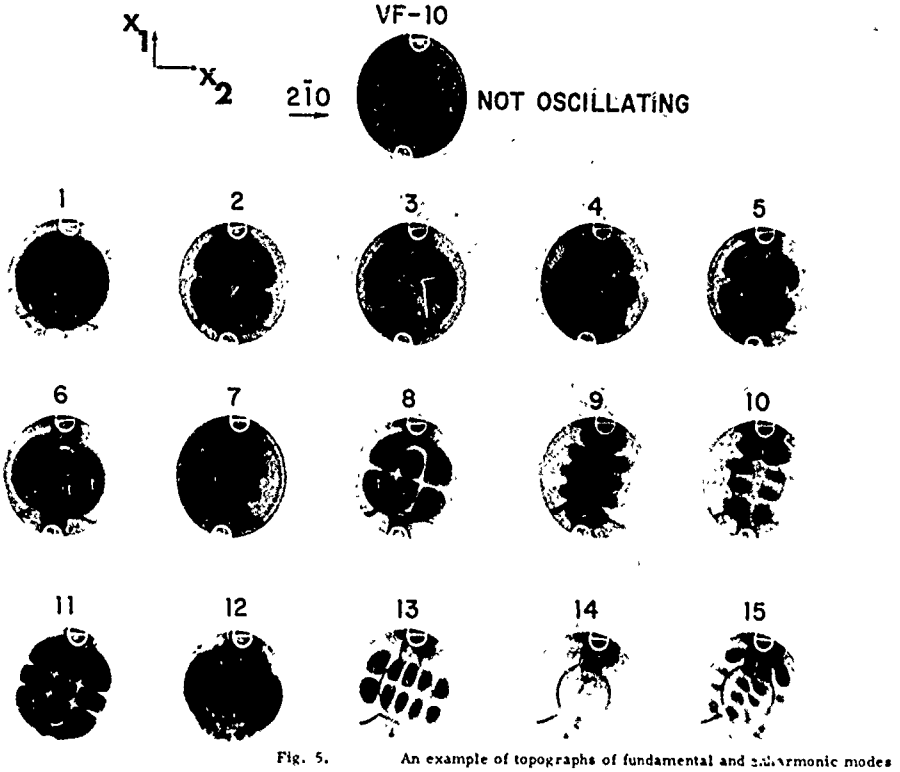
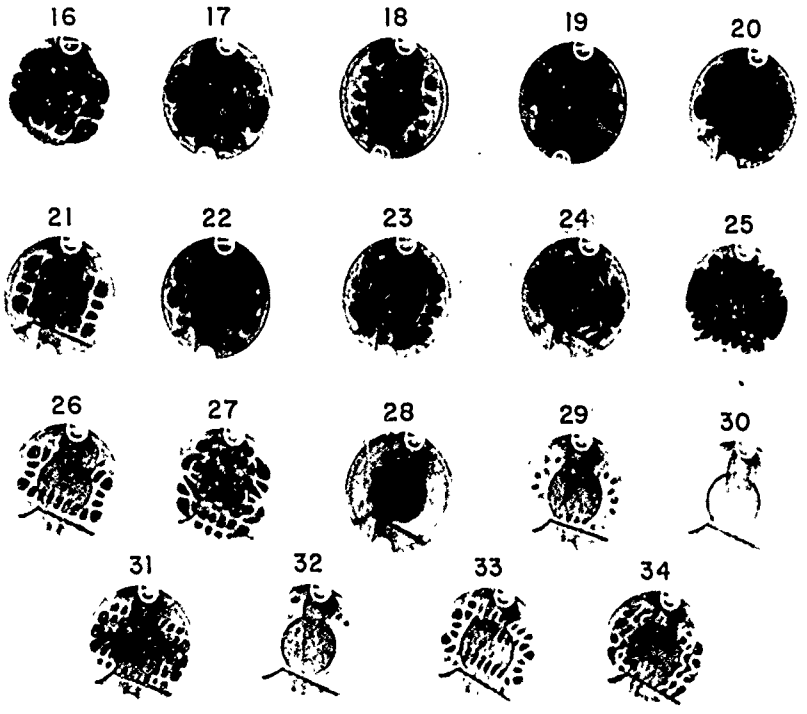


Fig. 5. An example of topographs of fundamental and subharmonic modes in V. F -10.



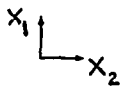
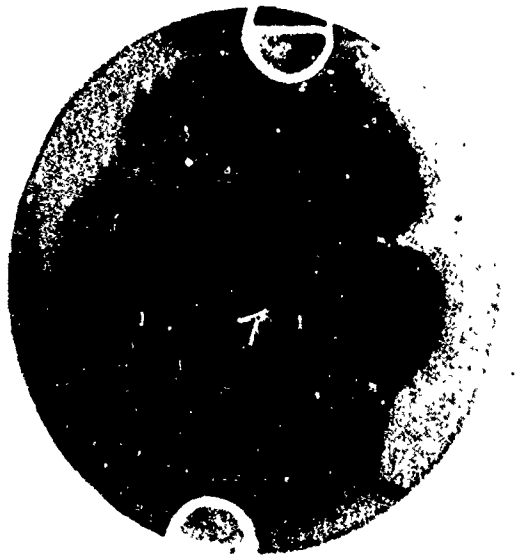


Fig. 6. Enlargement of topograph of the 131 mode.



Rare

χ_2 \ χ_1	1	2	3	4	5	6	7
1	1	2	3	4	5	6	7
2	2	4	6	8	10	12	14
3	3	6	9	12	15	18	21
5	5	10	15	20	25	30	35
7	7	14	21	28	35	42	49
9	9	18	27	36	45	54	63
11	11	22	33	44	55	66	77

Fig. 7. Matrix derived from topographs and frequency spectra.

VF 7

χ_2 \ χ_1	1	2	3	4	5	6	7
1	1	2	3		5		
2							
3	3		9		15		
5	5		15				
7	7				35		
9	9	18					
11	11						

Fig. 8. A Matrix of V.F.-7

VF 10

χ_2 \ χ_1	1	2	3	4	5	6	7
1	1	2	3	4	5		
2	2	4		8	10		
3	3	6	9		15		
5							
7	7	14	21	28	35		
9	9	18		36			
11			33	44	55		

Fig. 8. B Matrix of V.F.-10

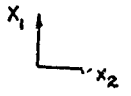


Fig. 9. Enlargement of topograph of the 177 mode.

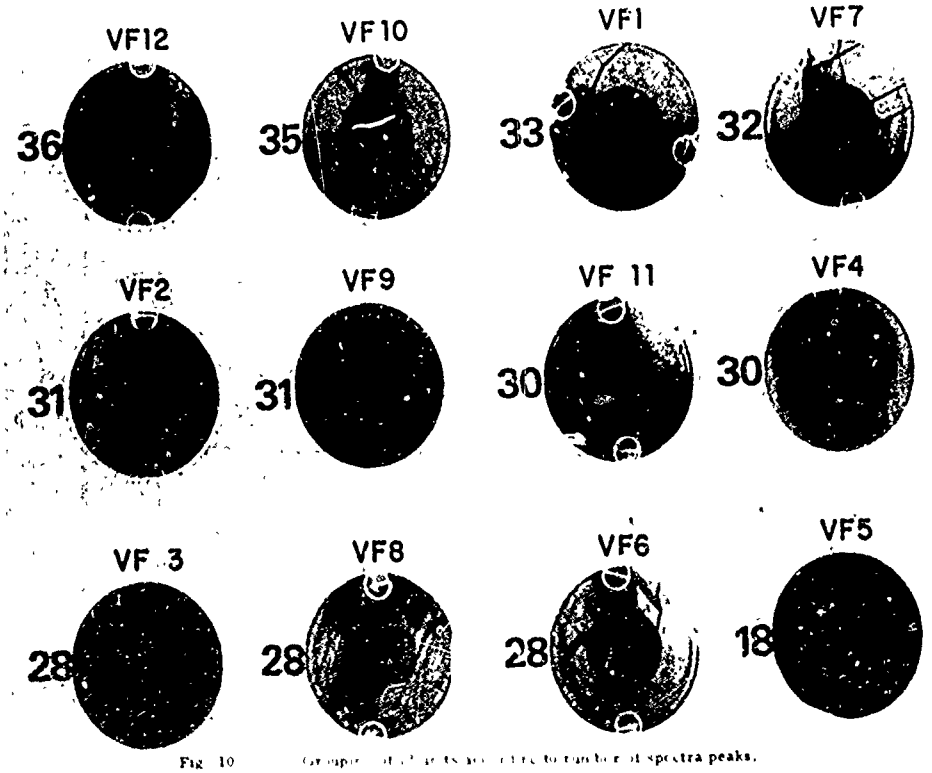


Fig. 10 Groups of [illegible] to number of spectra peaks.

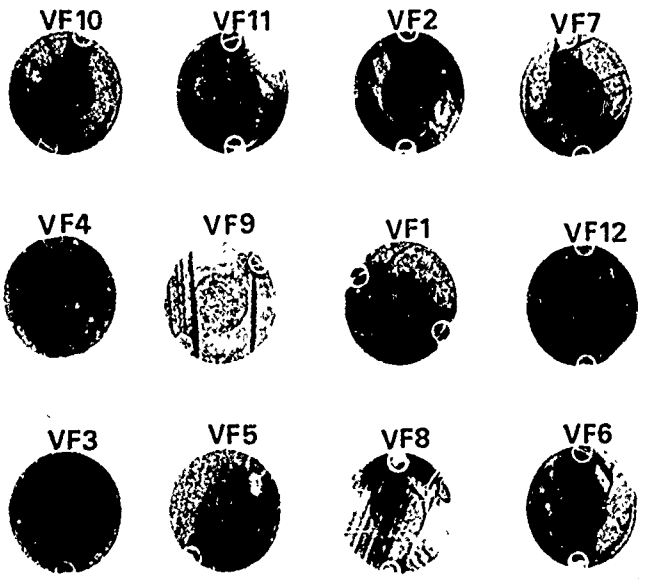


Fig. 11. Grouping of 12 units according to Q value measured by the log decrement method in decreasing order of relative Q

Reproduced from best available copy.

Fig. 12. Q values as determined by the log decrement method. The measurements were made in air.

V.F. No.	Frequency (MHz.)	Decay Time τ - ms	Q × 4
10	4.999987	5.4	8.4[illegible]
11	5.000083	5.3	8.[illegible]
2	4.998731	4.9	7.6
7	4.997805	4.9	7.6
4	4.998979	4.6	7.1
9	4.999948	4.4	6.9[illegible]
1	4.999651	3.9	6.1
12	4.999572	3.7	5.7[illegible]
3	4.999925	3.2	4.9[illegible]
5	4.999124	3.2	4.9[illegible]
8	4.999939	1.7	2.6[illegible]
[illegible]	4.999629	1.6	2.5[illegible]

REVIEW OF DIGITAL FILTERING

John D. Heightley
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey

Summary

In general, digital filtering embodies a wide range of signal processing and signal analyzing applications but this paper is concerned with the more conventional filtering requirements of frequency selection and rejection. Some of the principles involved in digital filtering are discussed from an intuitive point of view rather than in terms of the well-established mathematical treatments.[1,2,3] The form of a general second order filter section is established using this intuitive approach and the extension to higher order filters is discussed.

Introduction

Conventional analog filters employ inductors, capacitors, resistors and crystals in their realization while digital filters employ multipliers, adders and storage (delay) in their realization.[4] The coefficients associated with the transfer function of a digital filter are numerical inputs to the multipliers and are thus not sensitive to component variations. Since the transfer function of a digital filter is determined numerically rather than through the use of physical elements, certain restrictions on realizability are removed and the characteristics of the filter are easily modified.

How Is Filtering Accomplished Digitally?

This question is essentially equivalent to asking how the three operations of multiplication (scaling), addition (linearly combining) and storage (delay) may be combined to perform frequency selection and rejection. Consider the case where these operations are performed on a continuous analog input signal, $e_i(t)$, to obtain an output signal, $e_o(t)$, in the following manner:

$$e_o(t) = e_i(t) + \alpha_1 e_i(t-T) + \alpha_2 e_i(t-2T). \quad (1)$$

We have restricted ourselves to delays of either one unit (T) or two units (2T), since a larger number of delayed inputs may be treated by repeatedly applying the same process. The Laplace transform of the above output signal may be written by inspection:

$$E_o(s) = E_i(s)\left[1 + \alpha_1 e^{-sT} + \alpha_2 e^{-2sT}\right] \quad (2)$$

where s is the complex variable $\sigma + j\omega$. The operations performed here obviously have some spectral shaping effect on the input signal and the transfer function of this filter will be:

$$H(s) = \frac{E_o(s)}{E_i(s)} = 1 + \alpha_1 e^{-sT} + \alpha_2 e^{-2sT}. \quad (3)$$

If we wish to have this transfer function reject an input signal of frequency, ω_z, then we simply set the magnitude of (3) equal to zero with s replaced by $j\omega_z$. Solving for α_1 and α_2 we find

$$\alpha_1 = -2 \cos \omega_z T \quad (4)$$

$$\alpha_2 = 1. \quad (5)$$

It is this concept of linearly combining a number of appropriately scaled and delayed replicas of an input waveform which is at the heart of the nonrecursive digital filter. In order to convert the above analog filtering operation to a digital filtering operation, the input waveform is sampled periodically, at a rate at least twice as high as the highest frequency present in that waveform, and the sample values converted to a linear digital code (number). The indicated operations may then be done numerically.

In digital filters, the sample interval and the delay interval are the same so that in a digital realization, (1) becomes:

$$e_o^*(nT) = e_i^*(nT) + \alpha_1 e_i^*(nT-T) + \alpha_2 e_i^*(nT-2T) \quad (6)$$

where nT is a particular sample time. The asterisk is intended to indicate that these are digitally coded amplitude samples. The coded samples are easily scaled by digitally multiplying the delayed samples by the coefficients and the linear combination is performed by simply adding the scaled values. The delay is realized by storing the coded samples. A digital filter which will realize (6) is shown in Fig. 1.

Normally we are not only interested in having zeros in the transfer function but also poles. If we had a transfer function which had the form of the reciprocal of (3), we would achieve a pole and thus resonant behavior, i.e.,

$$H(s) = \frac{E_o(s)}{E_i(s)} = \frac{1}{1 + \beta_1 e^{-sT} + \beta_2 e^{-2sT}}. \quad (7)$$

This function will have infinite gain at ω_p if

$$\beta_1 = -2 \cos \omega_p T$$

$$\beta_2 = 1.$$

We may rewrite (7)

$$E_o(s) = E_i(s) - E_o(s)\left[\beta_1 e^{-sT} + \beta_2 e^{-2sT}\right]. \tag{8}$$

The inverse Laplace transform of (8) is by inspection:

$$e_o(t) = e_i(t) - \beta_1 e_o(t-T) - \beta_2 e_o(t-2T). \tag{9}$$

The output is therefore a linear combination of the input and scaled versions of the output which have been delayed by one unit of time (T) and two units of time (2T). A filter with infinite gain at its resonant frequency is an oscillator and if this is to be avoided, we must insert some loss in the feedback paths of the delayed output. Let us assume that the feedback waveforms are scaled further by a factor $e^{-\sigma T}$ for each unit of delay where $\sigma T > 0$, i.e.,

$$\beta_1 \rightarrow -2e^{-\sigma T} \cos \omega_o T$$

$$\beta_2 \rightarrow e^{-2\sigma T}.$$

This is equivalent to moving the poles into the left half of the s-plane.

We may rewrite (9) when the voltages are represented by a digital code at each sample instant as was done with (6) to give:

$$e_o^*(nT) = e_i^*(nT) - \beta_1^* e_o(nT-T) - \beta_2^* e_o(nT-2T). \tag{10}$$

This filter function may be realized as shown in Fig. 2.

Consider next the case where we wish to realize a general second order section having two zeros and two poles, i.e.,

$$H(s) = H_D(s)H_N(s) = \frac{1 + \alpha_1 e^{-sT} + \alpha_2 e^{-2sT}}{1 + \beta_1 e^{-sT} + \beta_2 e^{-2sT}}. \tag{11}$$

The total number of storage locations may be reduced if we generate an intermediate result which is due only to the denominator of (11), i.e.,

$$E_x(s) = E_i(s)H_D(s)$$

$$E_o(s) = E_x(s)H_N(s).$$

From the results in (6) and (10), it follows that

$$e_x^*(nT) = e_i^*(nT) - \beta_1 e_x^*(nT-T) - \beta_2 e_x^*(nT-2T) \tag{12}$$

$$e_o^*(nT) = e_x^*(nT) + \alpha_1 e_x^*(nT-T) + \alpha_2 e_x^*(nT-2T). \tag{13}$$

Only two storage locations are required to implement these equations rather than the four which would result from combining the filters of Fig. 1 and Fig. 2. A digital realization of the second order filter described by (12) and (13) is shown in Fig. 3. Higher order filters may be realized by cascading or paralleling sections of the form shown in Fig. 3. One additional multiplier is required for scaling the output of the filter in order to control the gain of the filter independent of the pole and zero locations.

An important point to note is that the processing portion of the filter is only in use when an input sample is present. If the input signals are low frequency and thus the sampling rate low, there may be idle periods between samples during which the processing equipment may be used to process a number of additional input sample streams which are time division multiplexed. Two words of storage are required for each sample stream being processed. Note also that the coefficients may be changed for the processing of each sample stream so that many different filters may be realized simultaneously with a single processor. A cascade of filter sections may be realized in the same fashion by storing the output associated with one section as it is generated and feeding that stored output back as input in the next sample interval with the coefficients changed to represent the second section. This process may proceed through as many sections as required. In both of these applications, if variable coefficients are required, the coefficients may be stored in a read-only memory and read out at the required time. This multiplexing of sample streams through a single set of processing elements is illustrated in Fig. 4.

Impairments

There are three primary sources of impairment associated with digital filtering which are: (1) signal quantizing noise, (2) filter roundoff noise and (3) coefficient quantizing. Signal quantizing noise arises because the continuous range of sampled signal amplitudes must be coded digitally using a finite number of quantized levels. This means the level represented by the code may differ slightly from the

actual level thus generating noise. Roundoff noise results when an N-bit sample is scaled by an M-bit coefficient if the storage word length is only N-bits. The M+N bit product cannot be stored and must be rounded off so that the stored value is slightly in error, thus causing noise to be generated. Both quantizing noise and roundoff noise are held to acceptable levels by providing enough bits to do adequate coding and to absorb the roundoff noise in the filter. The third impairment results from the fact that the coefficients in a digital filter are also quantized. Coefficient quantizing leads to a quantization of the pole and zero locations. This means that if a limited number of bits are used to represent the coefficients, we can only approximate the desired locations of the singularities. The number of bits used to represent the coefficients must be adequate to meet the requirements on the transfer function. Once the coefficients are fixed for the filter, however, the transfer function will remain fixed.

Applications

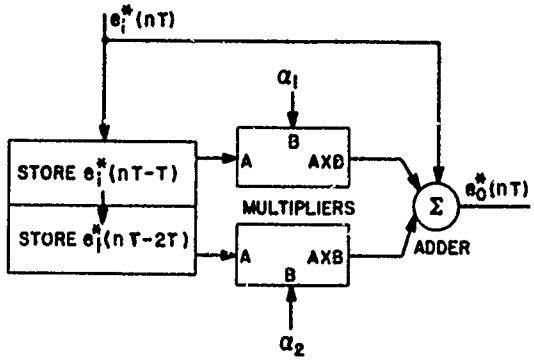
With the continuing advance in complexity, density, and performance of digital integrated circuits and, at the same time, decreasing cost, it appears inevitable that digital filtering will take over some of the functions now performed by analog filters. In applications where high order filters are required or where a number of individual channels are to be filtered and the frequency is not excessive, a single processor may serve many sample streams and thus be shared by many filters. In these cases, only the storage elements are provided for each sample stream and these elements stand to gain the most from the integrated circuit technology because they are realized as orderly arrays.

There are also applications such as very low frequency filtering which were not feasible with analog filters that can be performed quite readily with digital filters.

In conclusion then, the stability and flexibility of digital filtering techniques, combined with the rapidly moving integrated circuit technology, will make it possible to not only take over many analog filtering applications but to extend the meaning of the term filtering to processing applications not possible before.

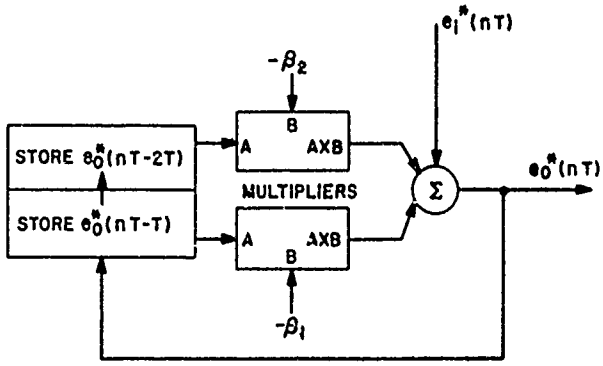
References

1. J. F. Kaiser, "Digital Filters," in System Analysis by Digital Computer, J. F. Kaiser and F. F. Kuo, Eds., New York; Wiley, 1966, pp. 218-85.
2. C. M. Rader and B. Gold, "Digital Filter Design Techniques in the Frequency Domain," Proc. IEEE, Vol. 55, No. 2, pp. 149-171.
3. C. M. Rader and B. Gold, Digital Processing of Signals, New York: McGraw-Hill, 1969.
4. L. B. Jackson, J. F. Kaiser, and H. S. McDonald, "An Approach to the Implementation of Digital Filters," IEEE Trans. on Audio and Electroacoustics, Vol. AU-16, No. 3 (September, 1968), pp. 413-21.



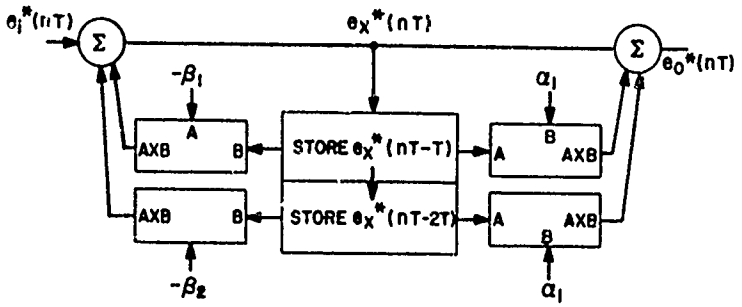
$$e_0^*(nT) = e_i^*(nT) + \alpha_1 e_i^*(nT-T) + \alpha_2 e_i^*(nT-2T)$$

TWO-ZERO DIGITAL FILTER



$$e_0^*(nT) = e_i^*(nT) - \beta_1 e_0^*(nT-T) - \beta_2 e_0^*(nT-2T)$$

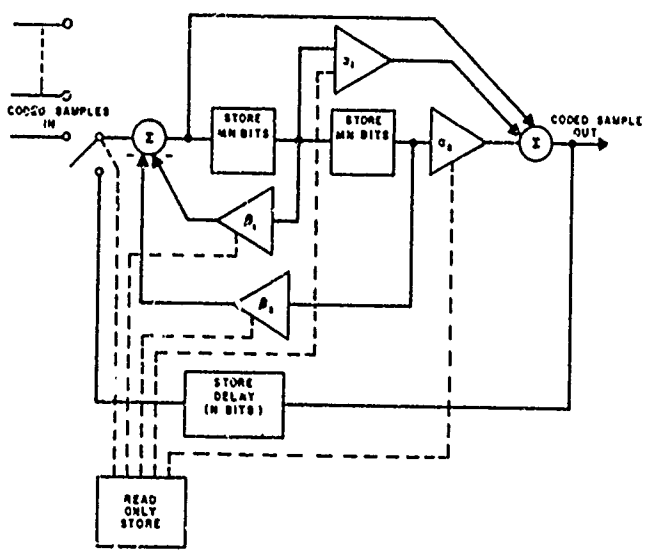
TWO-POLE DIGITAL FILTER



$$e_x^*(nT) = e_i^*(nT) - \beta_1 e_x^*(nT-T) - \beta_2 e_x^*(nT-2T)$$

$$e_0^*(nT) = e_x^*(nT) + \alpha_1 e_x^*(nT-T) + \alpha_2 e_x^*(nT-2T)$$

GENERAL SECOND ORDER FILTER SECTION



M SECTION MULTIPLEXED FILTER

ACTIVE FILTER CAPABILITIES

Philip R. Geffe
Westinghouse Electric Corporation
Baltimore, Maryland 21203

Summary

Active filters are best made as a cascade of simple sections. Each section has a specified center frequency, Q, and two transfer zeros.

This paper discusses the practical circuit design of the most useful sections. Sections using unity-gain amplifiers can realize LP and HP filters having pole Q's of 10 at low frequencies to a limit of Q=1 at about 1 MHz. A Sallen and Key section gives BP filters over the same range of Q and frequency. The dual integrator resonator can realize high Q filters of any type with Q=200 to above 100 KHz.

Also discussed are frequency stability, thermal coefficients, and certain elements of size and cost.

Introduction

Design formulas are given for normalized active resonators. The gain function of each resonator has one pair of complex conjugate poles:

$$S = -\sigma \pm j\omega \qquad (1)$$

The design formulas are expressed in terms of the pole Q and the pole center frequency:

$$\omega_o = \sqrt{\sigma^2+\omega^2} = \text{"pole center frequency"} \qquad (2)$$

$$Q = \frac{\omega_o}{2\sigma} = \text{"pole Q"} \qquad (3)$$

If the Q is 5 or more, these definitions correspond quite accurately to the frequency of maximum gain, and to the ratio of center frequency to half-power bandwidth, respectively.

The normalized resonators all have $\omega_o = 1$, and the RC element values are given in ohms and farads. A practical design is then obtained by denormalizing the resonator with frequency and impedance scaling. This is done by multiplying all resistances in the network by K_z, and dividing all capacitances by $2\pi f_o K_z$, where f_o is the desired center frequency, and K_z is a freely chosen impedance factor.

Thermal coefficients are calculated from a knowledge of the sensitivities.[1] As an example of this, suppose a resonator has a sensitivity of Q to R_1 of 0.2, and let the thermal coefficient of R_1 be -100 ppm/°C:

$$S^Q_{R_1} = 0.2$$

$$(TC)_{R_1} = -100 \text{ ppm/°C}$$

Then R_1 contributes to the thermal coefficient of Q as follows:

$$(TC)^Q_{R_1} = (0.2)(-100 \text{ ppm/°C}) = -20 \text{ ppm/°C}$$

The algebraic sum of all such contributions will give the net thermal coefficient of Q.

Throughout this paper, A will represent the closed-loop amplifier gain, and A_o the open-loop gain.

Resonators Made with Voltage Followers

Figure 1 shows two resonators that are very useful in simple LP and HP filters, such as Butterworth and Chebyshev types. In both circuits,

$$\left.\begin{aligned} S^{\omega_o}_A &= 0 \\ S^{\omega_o}_x &= -\tfrac{1}{2} \end{aligned}\right\} \qquad (4)$$

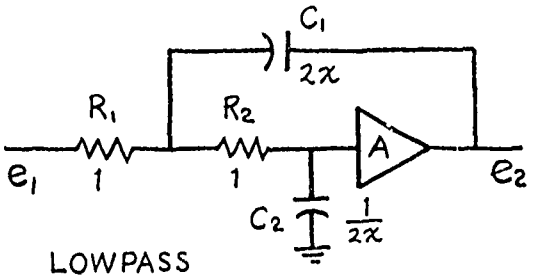
where $\mathcal{Z}$ may be any of the four passive elements.

In the lowpass filter,

$$\left.\begin{aligned} S^{Q}_{R_1} &= S^{Q}_{R_2} = 0 \\ S^{Q}_{C_1} &= -S^{Q}_{C_2} = \tfrac{1}{2} \end{aligned}\right\} \qquad (5)$$

In the highpass filter,

$$\left.\begin{aligned} S^{Q}_{R_2} &= -S^{Q}_{R_1} = \tfrac{1}{2} \\ S^{Q}_{C_1} &= S^{Q}_{C_2} = 0 \end{aligned}\right\} \qquad (6)$$



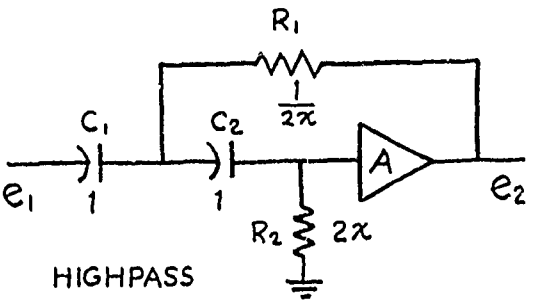


FIGURE 1: Voltage Follower Resonators

The numbers shown in the figures give the correct design on the assumption that $A = +1$. In this case, $Q = \mathcal{Z}$. On the other hand, A will always be slightly less than unity, and this requires us to modify the design formulas. Instead of identifying Q with $\mathcal{Z}$, we will therefore calculate $\mathcal{Z}$ from a knowledge of the gain:

Case I: A is given

$$\mathcal{Z} = \frac{1 \pm \sqrt{1 - 8(1-A)Q^2}}{4(1-A)Q} \qquad (7)$$

and the Q is not realizable unless

$$A \geq 1 - \frac{1}{8Q^2} \qquad (8)$$

Case II: A_o is given

Then $$A = \frac{A_o}{A_o + 1} \qquad (9)$$

$$\mathcal{Z} = \frac{A_o+1}{4Q} \pm \sqrt{\left(\frac{A_o+1}{4Q}\right)^2 - \left(\frac{A_o+1}{2}\right)} \qquad (10)$$

and the Q is not realizable unless

$$A_o \geq 8Q^2 - 1 \qquad (11)$$

In Case I, the gain information may be given as in Figure 2. This is the

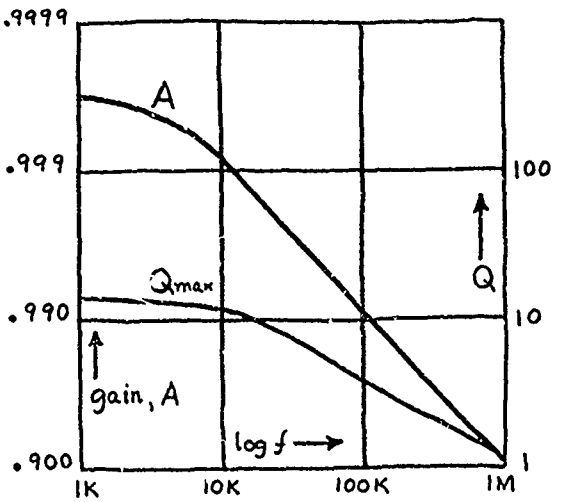


FIGURE 2: LM-102 Frequency Response

frequency response of a voltage follower made with an operational amplifier which was especially designed for this type of service. The maximum value of realizable Q is also shown on the graph as a function of frequency.

If some other amplifier is used, then the open-loop gain (with phase compensation networks in place) can be obtained directly from the manufacturer's catalog, or measured. The formulas of Case II are then used, and the available Q is shown in Figure 3.

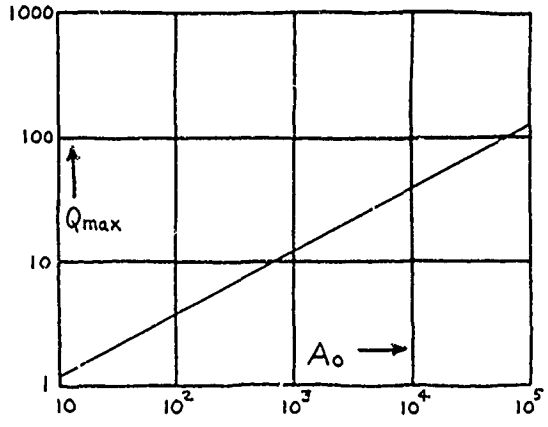


FIGURE 3: Available Q

The sensitivity of Q to the nominal gain is $2Q^2$, which seems very high. But, if we calculate the sensitivity to the open-loop gain, we obtain

$$S_{A_o}^{Q} = \frac{2Q^2}{A_o+1} \tag{12}$$

The worst figure obtainable here occurs when Eq. (11) applies. The sensitivity is then equal to 1/4, and will be better than this if excess gain is available.

A Bandpass Resonator

Figure 4 shows a useful BP resonator, originally given by Sallen and Key. The element values may be chosen as follows:

$$\left.\begin{aligned} A &= -5Q^2 \\ R_1 &= 1 \\ R_2 &= 4.235 \\ C_1 &= \frac{0.5002}{Q} \\ C_2 &= \frac{0.1181}{Q} \end{aligned}\right\} \tag{13}$$

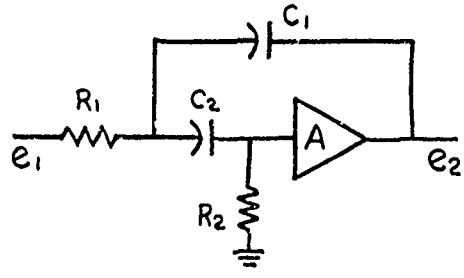


FIGURE 4: A BP Resonator

The sensitivities are

$$\left.\begin{aligned} S_z^{\omega_o} &= -\tfrac{1}{2} \quad \text{for all } z \\ S_A^{\omega_o} &\cong -\tfrac{1}{2} \\ S_{R_1}^{Q} &= S_{C_2}^{Q} = -0.053 \\ S_{R_2}^{Q} &= S_{C_1}^{Q} = +0.053 \\ S_A^{Q} &\cong \tfrac{1}{2} \end{aligned}\right\} \tag{14}$$

The Q of this resonator is also severely gain limited, and the available Q is given, approximately, by Figure 3. One must keep in mind, however, that the center frequency will drift with about half the percentage change of the gain, so that additional open-loop gain may be needed for enough stability.

The Dual Integrator Resonator

Figure 5 shows one form of the dual integrator circuit. Complete design information on this circuit has been given by several authors,[1-5] and the reader is referred to them for details.

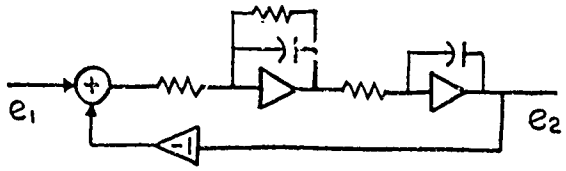


FIGURE 5: The Dual Integrator Resonator

The remarkable success of this circuit is due to the way in which the integrators contribute to the total gain. So far as is presently known, all low sensitivity circuits require gain proportional to Q^2, and this circuit is no exception. What is exceptional here is that the gain is factored into two amplifiers without, at the same time, causing any problems with feedback stability. This means that low sensitivities are obtained when each amplifier has an open loop gain of about $5Q$. The available Q is shown in Figure 6, which shows that values greater than 100 are readily obtainable. At present this is the only known method of obtaining high Q with high stability.

The degree of stability that can be obtained is very high indeed. At Westinghouse, we recently made an experiment with a resonator of this type which had a Q of 160 at 7.5 KHz. The center frequency was compensated by using components with inverse thermal coefficients. After compensation, a temperature test showed that the total drift (not the thermal coefficient) was within ± 100 ppm of the mean value from -35C to +105C. This is better than can be obtained with a quartz crystal resonator at this frequency.

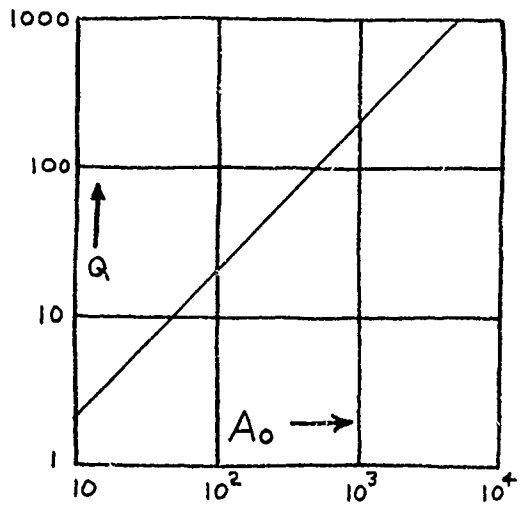
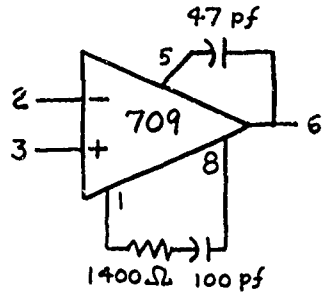


FIGURE 6: Available Q

Recently, the upper frequency limit for high Q in this circuit was moved from 2 KHz to 100 KHz. In order to see how this enormous improvement came about, consider the problem of designing phase compensating networks for the integrator amplifiers. Because a capacitor connects the output to the input, the textbook reasoning went as follows: Since the feedback approaches 100% at high frequencies, we must use the network that is needed for voltage follower service. With this network in place, the open loop gain of a 709 or 702 amplifier crosses the 60 dB point around 1 or 2 KHz, so this is the upper frequency limit for a Q of 200.

Last year at the Bell Telephone Laboratories in Holmdel, New Jersey, it was discovered that this reasoning is entirely wrong. In fact, the integrator is feedback stable with light lag compensation that would be totally inadequate for a voltage follower.[3] While space forbids giving the complete rationale of this discovery here, we show the compensation networks, for the μA709 and μA702 amplifiers, in Figure 7. These permit the 709 to be used without problems to 20 KHz, and the 702 to 100 KHz. Pin numbers are for the TO-99 case.



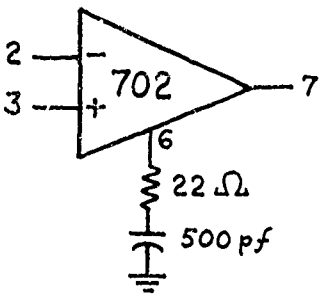


FIG. 7: Compensation for an Active Integrator

Conclusions

Practical design data has been given which, together with the information in the references, enables design of high Q active filters to 100 KHz. For moderate Q values, the resonators using only one amplifier can effect considerable cost savings.

References

1. P.R. Geffe, "RC-amplifier resonators for active filters," IEEE Trans. Circ. Theory, Dec. 1968, pp. 415-419.

2. W.J. Kerwin etal, "State-variable synthesis for insensitive integrated circuit transfer functions," IEEE J. Solid-State Circuit, Sept. 1967, pp. 87-92.

3. J. Tow, "Design formulas for active RC filters using operational amplifier biquad," Electronics Letters, 24 July 1969, pp. 339-341.

4. J. Tow, "A step-by-step active filter design," Spectrum, Dec. 1969, pp. 64-68.

5. L.C. Thomas, "The biquad: A multi purpose active filtering system," to be published.

GENERALIZED FILTERS USING ULSTRASONIC WAVES

By

M. G. Holland
Raytheon Research Division

Interdigital electrodes on piezoelectric substrates are used to generate surface ultrasonic waves. The equivalent acoustic circuit of these structures can be trated as a series combination of resonant elements which are electrically connected in parallel. By adjusting the spacing and size of the electrode elements, the frequency response (phase and amplitude) of the units can be controlled. It is possible to combine electrodes in such a way that many filter functions can be synthesized. Matched chirp filters have been previously reported, and units such as bandpass filters, transversal equalizers and Taylor weighting filters have been made. Computer derived responce curves will be compared to experimental results on these devices, and the Fourier transform property of the structures will be demonstrated. Bandwidth and frequency limitations, mask design and fabrication, photolithograph problems, properties of various substrate materials, and compatibility with other integrated circuit elements will be discussed. The future development potential of this field will be assessed.

MONOLITHIC CRYSTAL FILTERS

by

R. J. Byrne
Bell Telephone Laboratories, Inc.
Allentown, Pennsylvania

Introduction

Since the introduction of monolithic crystal filters (MCF's) in 1966,[1] a number of such devices have been used in Bell System applications. Most of these are two-pole devices used as single frequency band pass filters whose frequency, temperature and aging tolerances are well within the present state-of-the-art for crystal units. The advantages of MCF's in terms of size, cost, high reliability and reduced quartz usage led to the development of a channel bank system at 8 MHz using monolithic crystal filters. Although channel filters represent a much larger application, the requirements are much more stringent and are more closely related to precision crystal units. The mechanical design and the processing operations necessary to produce high frequency channel filters, in very large quantities and at a reasonable cost, were reviewed and attempts were made to take full advantage of batch processing techniques widely used in other modern planar device designs.

Telephone channel banks are used to assemble individual voice channels into groups for subsequent transmission over various wire and radio systems. Presently twelve individual voice channels, each 4 kHz wide, are assembled into an Internationally agreed group between 60 and 108 kHz as shown in Fig. 1. In the Bell System the channel filters use four "E" element crystal units in a lattice structure with additional capacitors and inductors. The new MCF system was designed so that the basic channel group would be located between 8.140 and 8.188 MHz, Fig. 1. This frequency was selected because it is high enough to avoid the problems of the lower end of the AT range, uses a plate of reasonable size and the resonator Q necessary to realize the desired filter characteristic is readily attainable. After the group is formed it is translated to the 60-108 kHz range in a second modulation step.

Compatability of the new channel bank with the present one requires that the same absolute tolerances be held at 8 MHz as at 80 kHz, a change from parts in 10^4 at the low frequency to parts in 10^6 at the new channel frequencies but with the advantage of the 12 filters being nearly identical.

Figure 2 shows the size comparison of the present filter with its four low frequency crystal units and associated circuitry and the MCF. The generally "planar" nature of the MCF is apparent as well as the lower number of components and interconnections. Unlike the low frequency crystal filters where the quartz plate dimensions vary by two to one over the twelve channel range of 60 to 108 kHz, the plate and array dimension of each of the twelve MCF channel filters are identical. This lends itself to automatically controlled batch operations compared to the one at a time methods used in existing devices.

Filter Design

The general filter design process used at Bell Laboratories is illustrated in Fig. 3. From the system requirements, the number of poles necessary for adequate band edge slope are determined and the resonator Q's attainable at the desired frequency are estimated. Using an insertion loss synthesis technique[2] the critical frequencies of the filter are determined and related through the equivalent circuit to resonator tuning frequencies and inter-resonator couplings. The plate dimensions and electrode sizes are selected by the designer to provide the required termination impedance and anharmonic suppression. The resonator spacings are determined using an analysis of the piezoelectric equations[3] and interactive computer programming. The resonator tuning frequencies that are computed take into account the effect of adjacent resonators on the measured frequency of an individual resonator. This effect can be relatively large, running as much as 40-50 parts per million in the 8 MHz design. With these array dimensions, the necessary masks are obtained and filters are fabricated.

The desired filter characteristic for the channel bank application is illustrated in Figure 4. It is realized using an 8-pole Chebyshev design with a .1 db ripple and a 3 db bandwidth of 3260 Hz. f_c is the channel carrier frequency. The tolerances shown at various frequencies across the band are the allowable variations. These variations must include those of initial adjustment, temperature variation from 10-50°C, and 20 year aging. The most critical requirements are those at the lower

end of the passband, between 200 and 600 Hz above the carrier frequency where this filter must be compatible with equalizers installed in existing plant. Translating the loss requirements to a frequency scale, the lower 3 db point must be held within ± 25 Hz or ± 3 parts per million for all causes. Out of band, there must be no responses within 250 kHz of the passband stronger than -60 db. For this filter design, resonator Q's of approximately 200,000 are required.

The 8-pole MCF has a monotonic loss characteristic while the present filter is a peak section design. The loss shapes are shown in Fig. 5. While the MCF characteristic is not quite as steep as the peak section design, interchannel noise suppression of the MCF is more than adequate to achieve the current objectives. For this application, the plate is cut in two after mounting and a capacitor is used to couple the two middle resonators. This guarantees adequate suppression of unwanted modes and eliminates some variation in the loss in the skirt region that otherwise occurs in high order MCF's on a single plate due most probably to unwanted propagating modes. The mechanical design is such that processing is the same whether or not the plate is cut.

Assembly Process

The manufacturing process and the tolerances allowed at each step are governed principly by the inter-resonator couplings. The objective is to achieve the calculated couplings at the same time that the resonators are tuned to the proper frequencies so that the correct set of natural frequencies are obtained when the unit is connected in a filter configuration. The couplings depend mainly on the resonator spacings and the plateback and to a lesser extent on the manner in which the electrode mass is distributed. The accumulated tolerances on each of those must be small enough so that the couplings will be within adjustment range which is approximately 10%. For our present geometry using the equivalent circuit, it can be shown that for a worst case, such as all couplings being either high or low, the couplings must be held to ± 2%.

The assembly process, Fig. 6, begins with the quartz plate. It is 1.35" long by .440" wide and is cut from r-face grown synthetic quartz. The plate is oriented for lowest frequency variation between 10 and 50°C and this orientation is held to ± 1' over all the plate. The parallelism of the plate is a critical point, since variations of thickness along the coupling direction cause variations in the inter-resonator coupling. At the present time, this is held to 8 μ inches or less, resulting in an uncertainty in the inter-resonator coupling of about 6% depending on the location of the resonators on the plate. The end etch frequency is held to ± 1 kHz, adding another 1/2 to 1% uncertainty in the coupling. The process used to produce these plates is described in the next paper at this meeting.[4]

The electrode array is vapor deposited on the quartz plate using masks held to tolerances of .0001" on the resonator spacings, the most critical dimensions. The film is either chromium-copper-gold or chromium-gold depending on the mounting method to be used. 96% to 98% of the total mass needed to bring the resonators to frequency is applied in this operation using quartz crystal monitors for thickness control. A very even distribution of mass is insured by using rotating crystal holders. The total plateback of this design is approximately 250 kHz or 3.1%.

Until very recently, the plated quartz wafers were solder mounted to leads riveted to a ceramic plate, Fig. 7. The ceramic is an inexpensive mounting platform that permits easy handling through subsequent operations and prevents any strain resulting from the final cold weld sealing operation being transmitted to the quartz plate. The quartz plate is accurately located with respect to the ceramic platform so that all of the resonators on the plate can be properly positioned in subsequent operation by positioning the ceramic platform.

With sixteen leads being attached to the quartz plate, it is difficult to attain the "strain-free" mounting condition and low contamination levels necessary for very low long term aging. A thermocompression bonded design is now being introduced, Fig. 8, that will permit high temperature cleaning and strain relief operations.[5] The leads are first bonded to the quartz plate, eight leads at a time, then formed as shown, and finally bonded to a ceramic substrate in one operation. These operations are relatively free of operator control and appear to be suitable to a higher degree of mechanization than could be achieved with soldering techniques.

The resonator frequencies are adjusted by evaporated gold. A collimated beam of evaporant is directed by suitable masking to each individual resonator while the frequency is measured. The adjustment spot is centered and approximately 80% of the area of the individual resonators to keep the effect on inter-resonator coupling as small as possible. The resonators are adjusted to frequencies slightly higher (about 1.5 kHz) than those required when the filter is finished. This leaves room for the frequency changes that will occur when the inter-resonator couplings are adjusted.

Until recently, the coupling adjustment has been made by adding mass between the resonators. Mass removal works just as effectively and a laser machining facility to perform this operation is now in use.[6] Since a vacuum is not required, mechanization and automatic control of the facility is simpler as well as less costly. An additional stripe of electrode material is deposited during base plating between the resonators and designed as part of the resonant structure of the initial array. Removing a bit of this stripe with a laser raises the cut-off frequency in the inter-resonator area, thereby reducing the coupling. With the present array dimensions, the adjustment range is 10% and the accuracy is on the order of 1/2%.

The individual resonators are readjusted, again using vapor deposition. Since our frequency location requirement for this filter is not on the mid-band frequency, but on the location of a specific loss level on the skirt, the lower 3 db point, the accumulated error in couplings, resonator tunings and resonator Q's make it unlikely that acceptable filters could be made using resonator frequency tunings along. The resonators are left slightly high in frequency so that after this operation the filter characteristic has the proper bandwidth and ripple characteristics, but is located 50-100 Hz high in frequency.

At this point, a coupling capacitor is installed, the required internal connections are made, and the assembly is mounted on a suitable header, Figure 9. Final cleaning operations are done, limited to standard solvents and low temperatures (125°C) for soldered units and higher temperatures (400°C) stress relief and cleaning processes for thermocompression bonded assemblies.

The final adjustment operation involves adjustment of all resonators at once by vapor deposition of gold to shift the entire filter characteristic to the desired frequency. The adjustment areas amount to about 1/2 the individual resonator areas, and is limited to less than 150 Hz. Another paper at this meeting describes the measurement technique used to control this operation.[7] The accuracy of this adjustment is better than 0.5 ppm. A cover is then cold weld sealed to the header and the device is complete.

Results

A few hundred of these filters have now been fabricated with excellent results. The skirt regions are extremely uniform as shown in Fig. 10. The average compares very closely with the calculated characteristic and the standard deviation is small. The skewness of the band edges is felt to be due to stray capacitance in the measurement fixtures and in any event is not important in this case.

The inband results are equally good, as seen in Fig. 11. The agreement between the measured values and that predicted by the equivalent circuit are on the order of .05 db. The standard deviations are approximately .1 db at either end of the band and less than .05 db in the center.

Aging is a problem with solder mounted units, being on the order of 2-4 parts per million in the first three months. It is believed that this is mainly a contamination problem that can be reduced somewhat by more careful attention to the assembly process and virtually eliminated by utilizing thermocompression bonds and high temperature processes.

In summary, successful 8-pole MCF's at 8 MHz meeting channel filter requirements have been made. These units are compatible with planar batch fabrication techniques which should result in low cost in high level production. Thermocompression bonding has been used to reduce aging and frequency shifts due to contamination. When additional out-of-band suppression is required the plate may be cut without any change in process steps. The use of thin film processes and TC bonding permits full advantage to be taken of the inherent stability of the quartz to provide high Q, precision filters.

References

1. R. A. Sykes and W. D. Beaver, Proceedings of the 20th Annual Symposium on Frequency Control, p. 288, (1966).
2. R. C. Rennick, private communication.
3. K. Haruta and P. Lloyd, private communication.
4. A. J. Miller, Proceedings of the 24th Annual Symposium on Frequency Control, (1970).
5. R. J. Byrne and J. L. Hokanson, IEEE Transactions on Instrumentation and Measurement, IM-17, 1, p. 76 (March 1968).
6. J. L. Hokanson, Proceedings of the 23rd Annual Symposium on Frequency Control, p. 163 (1969).
7. R. P. Grenier, Proceedings of the 24th Annual Symposium on Frequency Control, (1970).

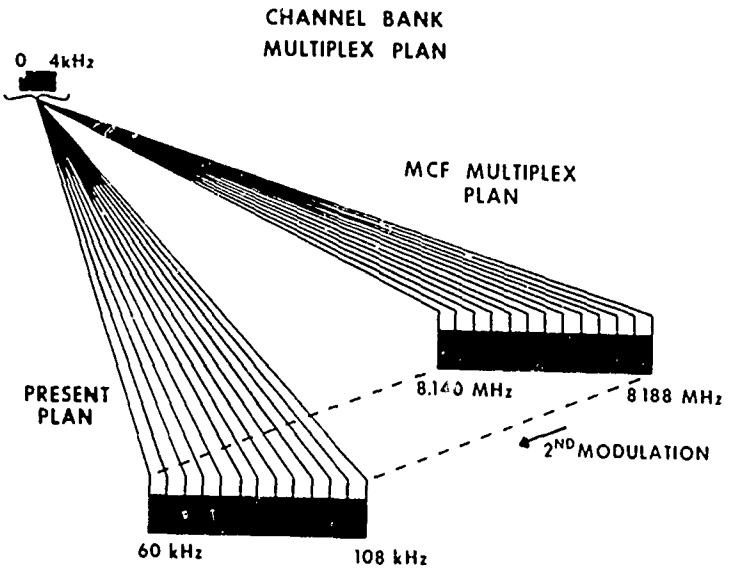
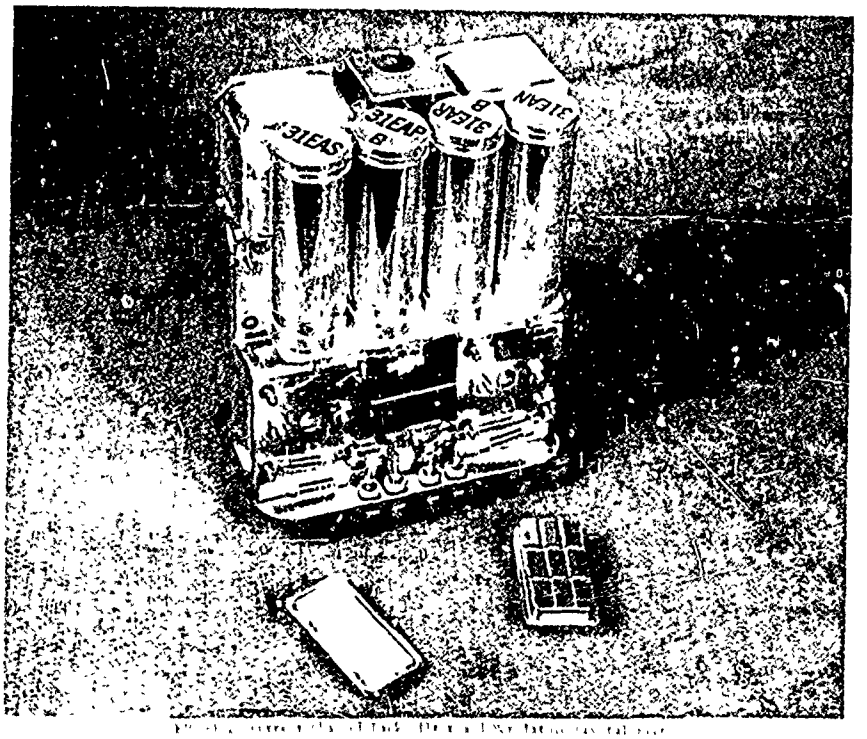


FIGURE 1 Channel Bank Multiplex Plans

Reproduced from
best available copy.



DESIGN PROCEDURE FOR MCF'S

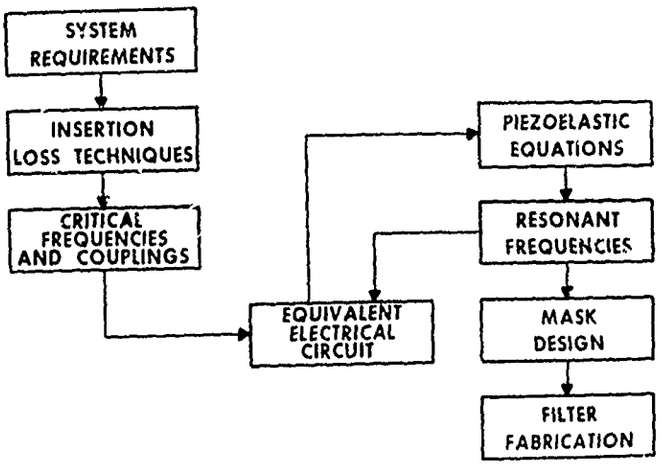


FIGURE 3 MCF Design Procedures

8 POLE CHARACTERISTIC

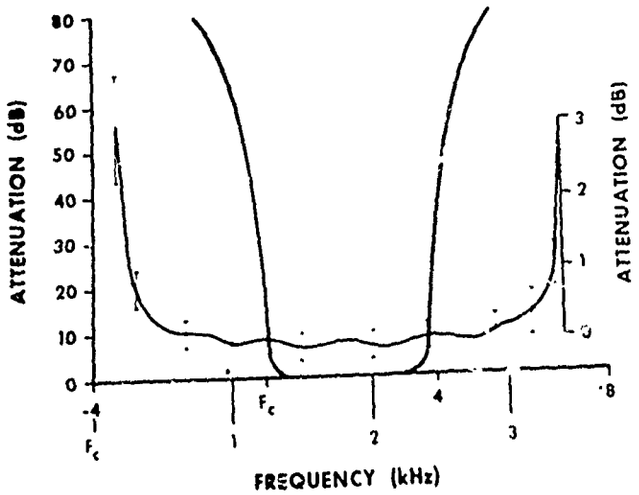


FIGURE 4 Channel Rank Filter Requirements

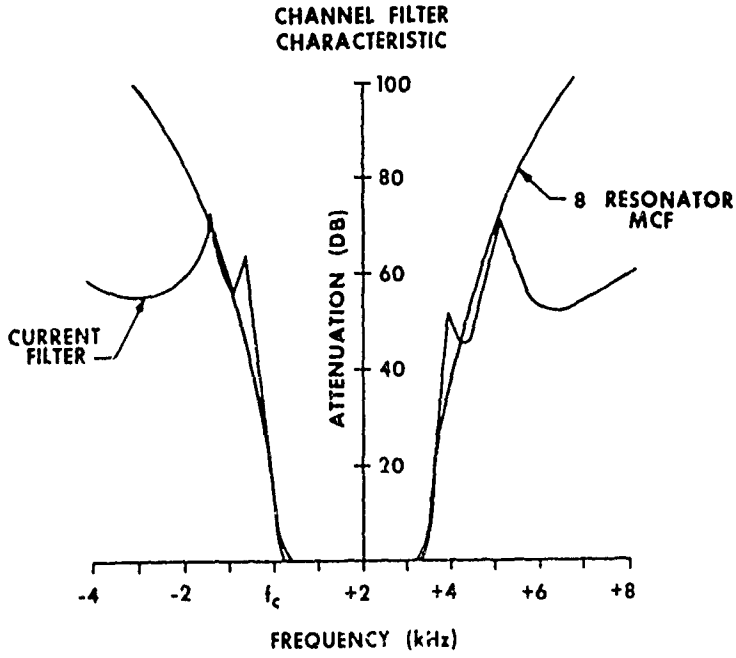


FIGURE 5 Filter Characteristics of [illegible] MCF Channel Bank Filters

PROCESSING STEPS FOR MONOLITHIC CRYSTAL FILTERS

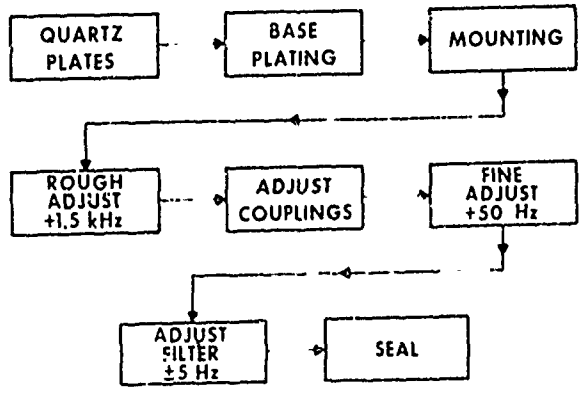


FIGURE 6 MCF Processing Steps

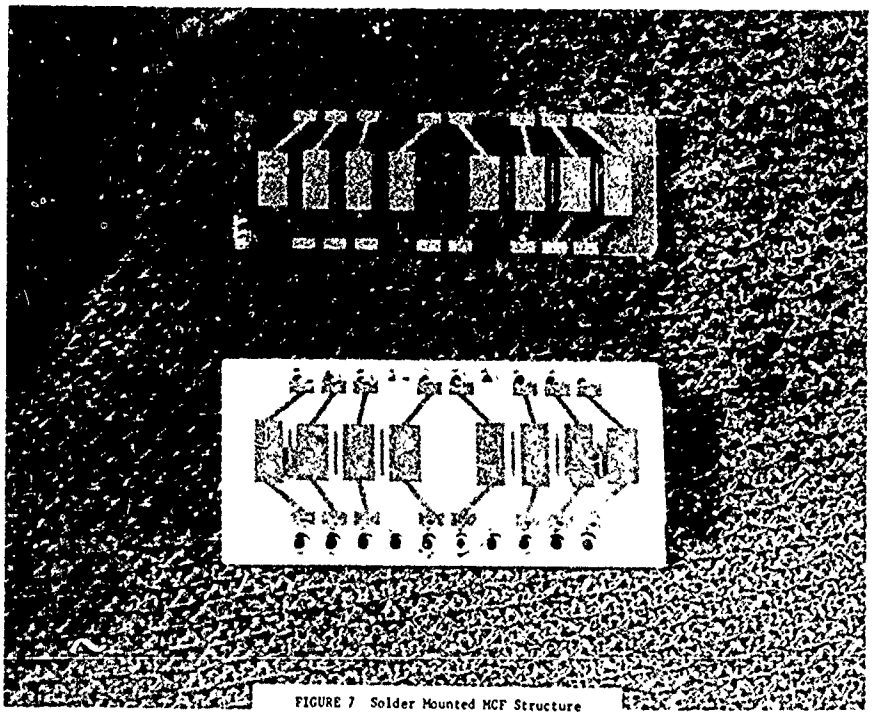


FIGURE 7 Solder Mounted MCF Structure

Reproduced from best available copy.

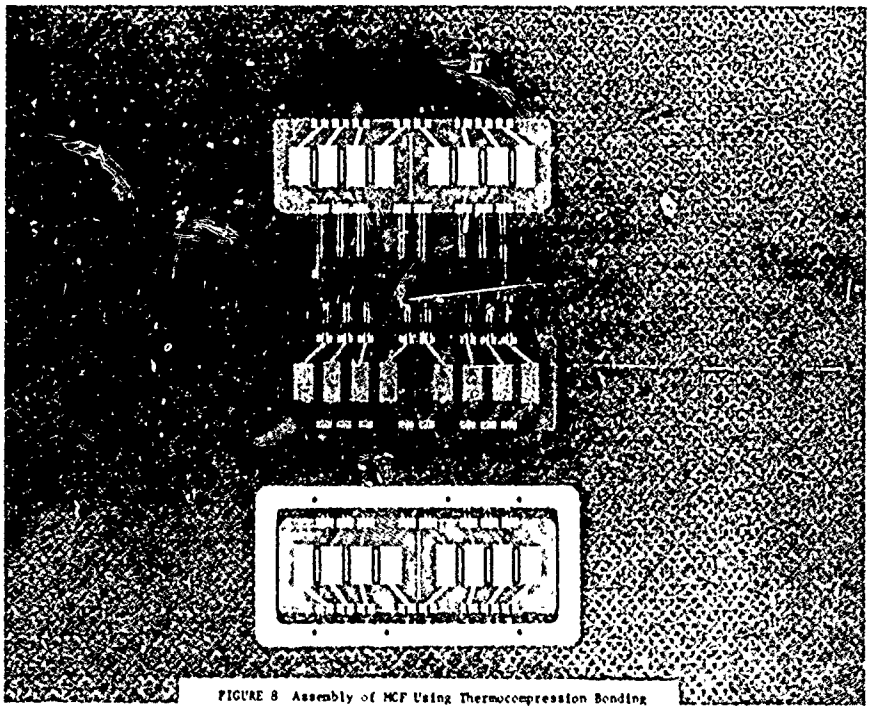


FIGURE 8 Assembly of MCF Using Thermocompression Bonding

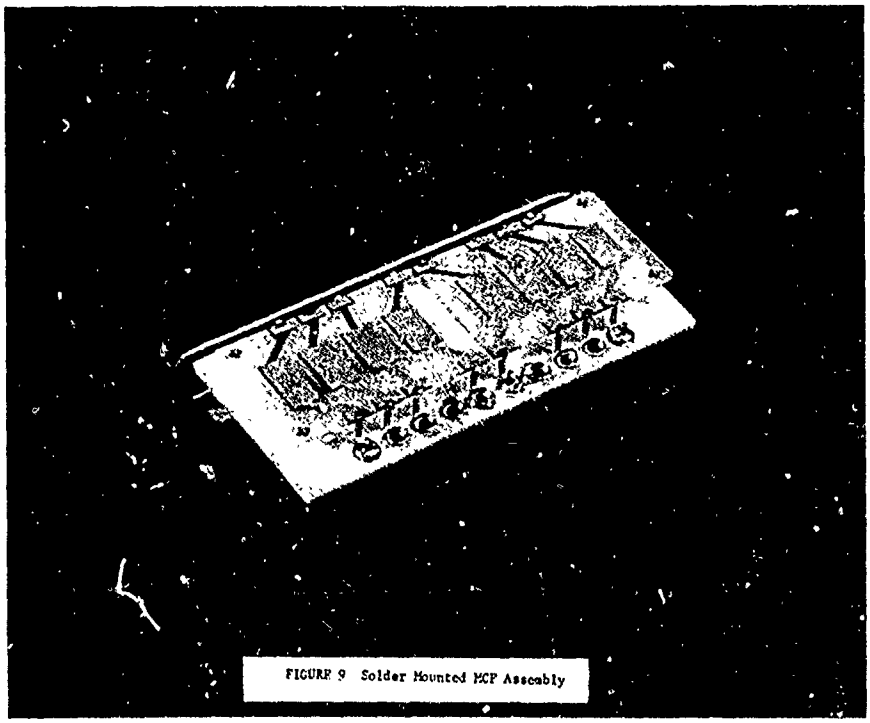


FIGURE 9 Solder Mounted MCF Assembly

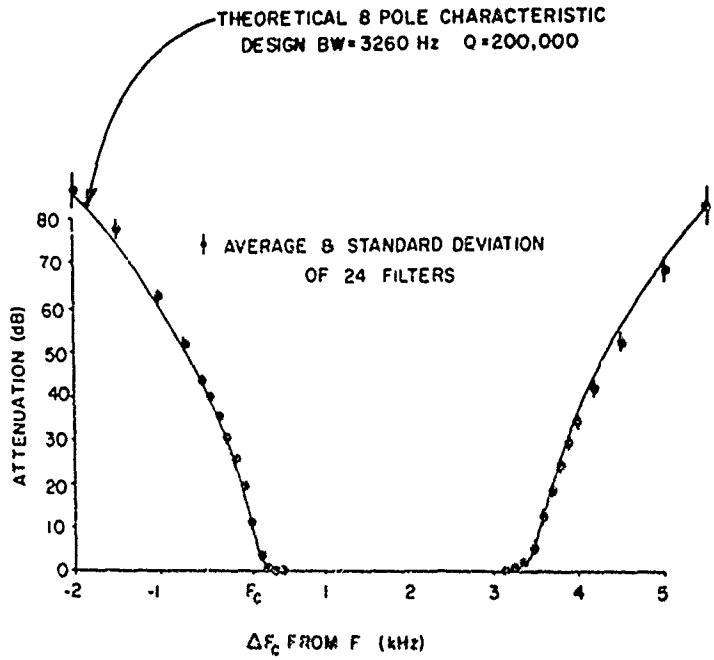


FIGURE 10 Average and Standard Deviation of 24 MCF Filters Out-of-Band Characteristics

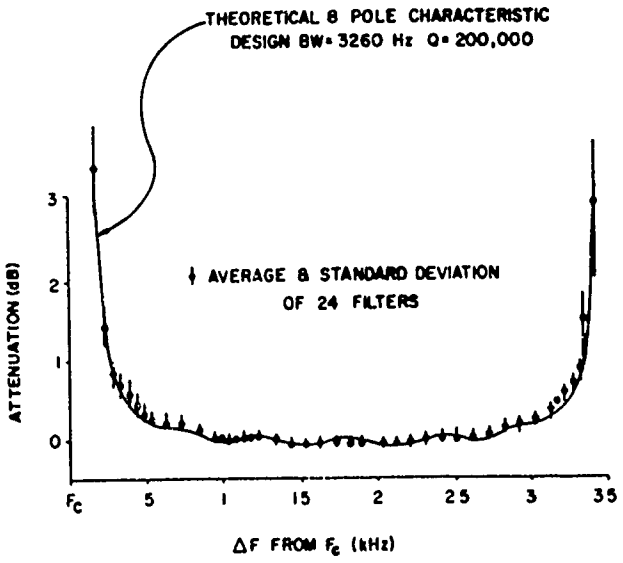


FIGURE 11 Average and Standard Deviation of 24 MCF Filters Inband Characteristics

PREPARATION OF QUARTZ CRYSTAL PLATES FOR MONOLITHIC CRYSTAL FILTERS

by

A. J. Miller
Bell Telephone Laboratories, Inc.
Allentown, Pennsylvania

Introduction

The previous paper[1] has shown the design and performance of Monolithic Channel Bank Filters (Fig. 1). In my presentation I will concentrate on important details pertinent to processing crystal plates for these filters.

It is apparent that the unusually strict requirements the filter has to meet will in turn place unusually strict requirements on the crystal plate. The principle requirements put on the crystal plate can be summarized as follows:

Of the eight resonators on one plate
A. The frequencies must be almost identical.
B. The frequency variations over the temperature range must be small and uniform.
C. The Q-values must be high and rather identical.
D. The Q-variations over the temperature range must be avoided or minimized.

Quite obviously these requirements can be met only if crystal plates are very parallel, the crystallographic orientation angle is optimum and deviations are kept to minimum, the major surfaces are smooth and uniform in texture, and the effects of unwanted flexure modes are adequately controlled.

Specific Requirements and Process Sequences

Fig. 2 shows the important dimensions of the AT-cut plate and specific requirements. The length along ZZ' is determined by the number of resonators, and the dimensions of electrodes and spacings. For economical reasons the width is kept as small as practical, guided by basic rules to avoid couplings to flexure type modes[2].

The nominal frequency is at 8.14 MHz. The tolerance of the AT-cut orientation is ± 1 minute of arc. Note at this point that the English measurement system is preferred throughout this paper simply because a frequency change of 1 kHz at 8 MHz is identical to a thickness change of 1 μ inch. The frequency variations, or the error of parallelism over the entire plate must be less than 8 kHz or 8 μ inches. This is measured at three points as indicated in the figure.

Figure 3 shows two process sequences, and approximate thicknesses and frequencies at each respective process step. Mainly because of difficulties in preserving the AT-cut orientation within the acoustically active area (roughly the area provided for electrodes) the existing standard process sequence was not satisfactory. In this sequence the thickness was reduced entirely by double face lapping, that is from both sides at the same time. This led to the modified process sequence. The main difference is that the thickness is now reduced almost entirely by single face grinding, that is from one side only, while crystal plates are mounted firmly onto a backing plate minimizing any angle variation.

By changing from double face lapping to single face grinding we avoid, or reduce, angle variations, but, unfortunately, we pay with parallelism. This means that with single face grinding alone, the yield in obtaining very parallel plates would be rather low. Therefore, we could not neglect double face lapping. In fact it is still the only economical technique capable of producing very parallel crystal plates. Two steps of double face lapping are still needed. The basic step before single face grinding removes irregularities in thickness and surface damage from the cutting step. A minimum of .00[illegible] inches was found to be sufficient for that. The final step is to equalize errors in parallelism from the single face grinding, and, of course, to be able to apply radio receiver control during lapping.

For this final lapping step the thickness reduction of ≈.001 inch may be considered a compromise because it is a minimum to equalize errors in parallelism from single face grinding, and a maximum to avoid AT-cut angle variations. With this modified process sequence under control, the only X-ray angle test necessary comes after the basic double face lapping step with the reading made at the center of the plate only.

Orienting-Cutting-Trimming

The R-face cultured quartz (See Fig. 4) recently developed by Western Electric Co.[3] has proven to be satisfactory in both economy and quality. Filters made of natural quartz and R-face quartz were fully identical, within the limits of the present experiment. These two drawings show

briefly the principles in orienting, cutting and trimming. Cutting the oriented blocks into crystal plates is done be a multiblade gangsaw. Because abrasives do the actual cutting, the crystal plates have little surface damage. The present yield in meeting the required tolerance of ± 1 minute of arc about the nominal AT-cut angle is greater than 80%. Trimming crystal plates to the final length and width dimension is done after basic double face lapping because the subsequent process steps produce no significant edge damage.

Single Face Grinding

Single face grinding is subdivided into cementing, pressing, calibrating, diamond and abrasive grinding (See Fig. 5). Up to 80 plates may be cemented onto one backing plate. The backing plate is serrated to allow excessive cement and unwanted particles to flow off. Figure 6 shows an alignment fixture placed over a backing plate. This fixture simplifies and speeds up proper locating of crystal plates during cementing. After cementing, the alignment fixture is removed and crystal plates are pressed onto the backing plate during cooling. The minimum pressure per crystal plate is approximately two pounds. Six boron carbide tipped pins screwed into the backing plate have to be adjusted to extend above the top surface so that the grinding action is stopped whenever the thickness of the crystal plates during grinding reaches this preset height. The height is determined by the frequency requirements after single face grinding. Despite the hardness of boron carbide, naturally the grinding action is not stopped at this point, but rather retarded. The retardation, however, is sufficient to allow good thickness control.

The results of one grinding load after single face grinding are shown on Fig. 7. This load consists of 72 crystal plates. The numbers with a plus sign represent frequency differences in kHz from plate to plate as measured at the center of each plate. Differences are recorded in reference to the thickest plate, the one with the lowest frequency which is here 7426 kHz. It can be seen that only a slight wedge, amounting to about 40 kHz is apparent. These results are plotted in the histogram at the lower left hand corner. Note, that the requirement at this point is ± 200 kHz about 7.3 MHz. The other numbers represent the error of parallelism in kHz as measured at three points along the length of each crystal plate. This error may be attributed to uneveness of the cement layer between crystal plate and backing plate. These results are statistically shown in the right hand corner. The requirement here calls for 80 kHz maximum.

Parallelism and Double Face Lapping

The basic requirement that the frequencies of the eight resonators be nearly identical requires that the major surfaces of the plate be very parallel. This is achieved by double face lapping.

Planetary lapping machines can be divided into those having two and those having four motions (See Fig. 8). The two motions are, of course, the rotation of each carrier around its own axis and the counterclockwise revolution of all carriers between two ring-shaped laps. The two additional motions are the clockwise rotation of the upper lap and the counterclockwise rotation of the lower lap. Both types of machines are capable of producing the required parallelism provided the two laps remain very flat and the crystal plates are transposed frequently.

After experimenting with several machines, we chose a four motion machine for our work because (1) the laps maintain their flatness for a longer time, (2) more crystal plates could be loaded into each blue steel carrier (the layout of such a carrier is shown in the lower right figure), (3) there was less breakage and edge damage, and (4) there were significantly less orientation angle changes during lapping. The main reason for these advantages is obviously based to a large extent on the opposing lap rotation which reduces or even equalizes unwanted forces applied to crystal plates and carriers, forces which are in turn applied to the laps. Because of their uniform particle size distribution micrograded aluminum oxide abrasives were found to add to the high degree of parallelism achieved. The histogram in the lower left corner of the figures shows characteristic results of this process. From several machine loads (about 200 plates) over 90% meet the 8 kHz or 8 μ inch parallelism requirement.

AT-Cut Angle Variations

The effect of AT-cut angle variations on frequency-temperature characteristic of filters is shown on results of five crystal plates (See Fig. 9). The frequency of each one of the eight resonators on each plate was individually measured over the temperature range. Subsequently the AT-cut angle was measured at each one of the areas covered by a pair of electrodes. It can be seen that the frequency deviations agree quite well with what one would expect the AT-cut angle to be for such a respective deviation. The parallelism on these plates is better than 6 kHz, and only 2 kHz on plate "E". This plate shows the largest angle deviations of all and represents a worst case condition. In spite of almost perfect parallelism, the AT-cut angle varies 7 minutes of arc from one plate end to the other. On the other side plate "A" may be considered a characteristic example of AT-cut variations obtained from the modified process sequence. The angle varies only 22 seconds of arc. Plate "D" is shown because the angles and frequency slopes measured at areas occupied by the four resonators on one half of the plate were identical, indicating no angle variations. The angles and slopes on the other hand, however, varied with progressively increasing deviation toward the end of the plate. "B" and "C" are of interest because the orientation angles of the resonators do not vary uniformly along the length of the plate but are random. Such variations are generated by deformations of the crystal plate during thickness reduction from double face lapping . A discussion of models which explain why

such variation occur is too lengthy a matter to be covered at this time. Suffice to say, the new modified process sequency greatly reduces or even eliminates such variation, as illustrated by the results of plate "A".

The Effect of Mass Loading on the Frequency-Temperature Characteristic

The effect of mass loading (or plate back as it sometimes is called) on the frequency temperature characteristic was experimently evaluated on resonators provided for these filters. The base-electrodes used consist of a combination of metals. A thin chromium film is followed by a layer of copper and a layer of gold. These metals are plated by evaporation. The mass of the base-electrodes was such as to lower the frequency about 1.5% below the frequency of the unplated plate. After evaporation the mass was increased in increments by electroplating nickel or evaporating gold, until a maximum frequency lowering of about 9% had been achieved. After each increment the frequency deviation between -50°C and +110°C was measured. The results for one crystal plate are shown in Fig. 10. The measurement at "zero" mass loading was made while the unplated crystal plate was sandwiched between two metal plates which had the same dimensions as the electrodes subsequently applied by evaporation. This family of frequency-temperature curves is the same as one would obtain by changing the crystal plate's orientation angle in increments up to a maximum of about 12 minutes of arc; keeping the mass loading constant, of course.

The frequency deviations between the lower and upper turnover-points (i.e., points where the tangent has zero slope) have been plotted as a function of mass loading, and are shown in Fig. 11. Six different resonators on six different crystal plates have been studied. On three of these resonators the mass was increased by nickel electroplating only. These results are drawn in solid lines. On the other three resonators the mass was increased entirely by evaporating gold. These results are shown in broken lines. The readings taken between the highest and lowest values have been omitted unintentionally. They do however fall on, or very near to the broken lines shown. Although the frequency deviations show some irregularities as mass increases, especially in the case of nickel electroplating, the relation of mass loading versus frequency-temperature characteristic can clearly be seen. Irregularities are caused probably by slight differences in the mass distribution over the plated area, and by differences in the mass balance between two electrodes of one resonator. No special attention was given to distribution and balancing of mass during this study. In spite of these disturbances, a coarse but useful rule can be established. It can be said that each one percent of mass loading causes the same change in the frequency-temperature characteristic of an AT-cut quartz-resonator as that resulting from a change in the crystallographic orientation angle of about 1-1/3 minutes of arc.

It is interesting to note that the Q-values of nickel electroplated resonators with heavy mass loads are very low and erratic, while the Q-values of resonators with heavy evaporated loads (8% for example) are almost identical to those of moderate mass loads.

Etching

Because of high requirements on Q and aging, much attention was paid to the condition of the two major surfaces. The amount of etching proved to be an important parameter. The minimum amount could not be determined by mechanical surface tests. For example, no differences in surface texture were revealed by stylus type measurements. Using a Talysurf instrument the stylus recordings of lapped, unetched and etched surfaces at a variety of etching stages were rather identical. The minimum etch was finally determined by microscopic tests. Samples were etched sequentially in intervals to increase the frequency a maximum of 1000 kHz. After each etching interval, the surface topology was observed with a scanning electron microscope.[4] Some of these surface samples at several etching stages are shown in Figs. 12, 13, and 14. The viewing angle is 45°, and we look along the Y -axis, where minus X is toward the top of the pictures. Magnifications at 2000 and 5000 appeared best to uncover important details. For realistic size comparison the white dust particle at zero etching has about the size of a 3 μm abrasive grain, the abrasive which has been used to lap these surfaces. If one takes complete removal of the layer formed by loose or disturbed matter as measure, he will observe that little has changed after mass was etched away equivalent to 20 kHz. 10 kHz, not included here, indicated hardly any change. Surfaces at 30 kHz (and even at 40 kHz) still show sections which may be unstable with time and hence may contribute to aging of filters. The first indication that the disturbed layer is removed -- that is when the surface really appears crystalline -- may be detected to some degree at 40 kHz, but certainly at 50 and 60 kHz. Thus the minimum etch was put at 50 kHz. At this point I would like to point out that the minimum amount of etching specified in MIL-C-3098 is ≈ 2.5 kHz at 8 MHz for plates lapped with 3.75 μm abrasive.

Because Q does not decrease with longer etching, a maximum amount of etching is not so important. For example, resonators etched about 150 kHz still were at the same Q level as resonators etched about 50 to 60 kHz. On the other hand, resonators etched amounts below 40 kHz, especially between 10 to 30 kHz showed nonuniform, sometimes erratic Q's, and because of that many filters did not perform as required. After the etching and position of electrodes were fully under control, generally no Q differences were observed between polished plates and those lapped with 3 μm, 5 μm or even slightly larger grain sized abrasives. It is strongly believed that by removing the layer of disturbed matter not only the Q but also the aging of filters was advantageously affected.

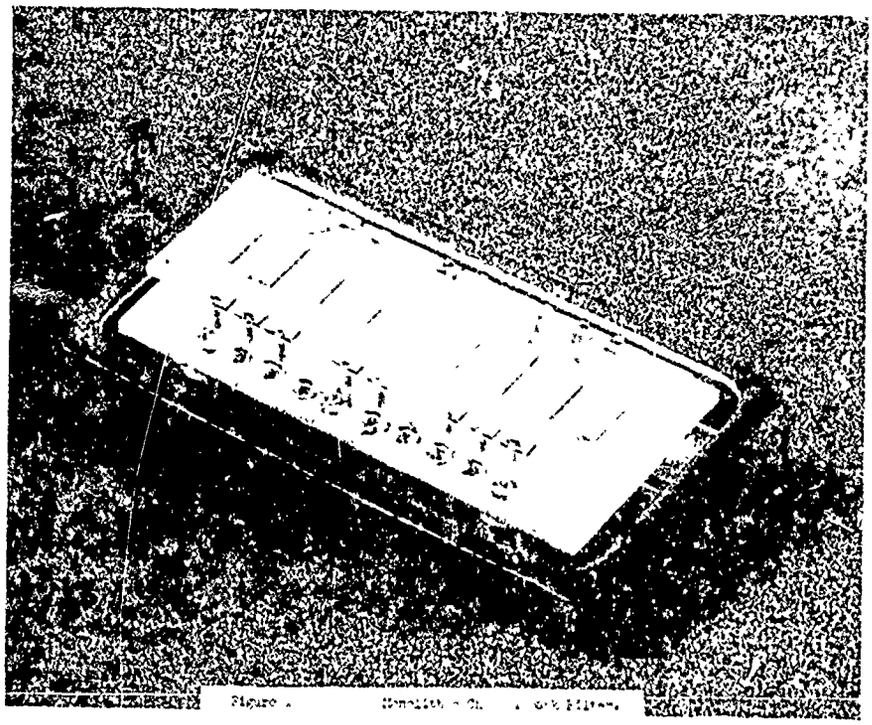
References

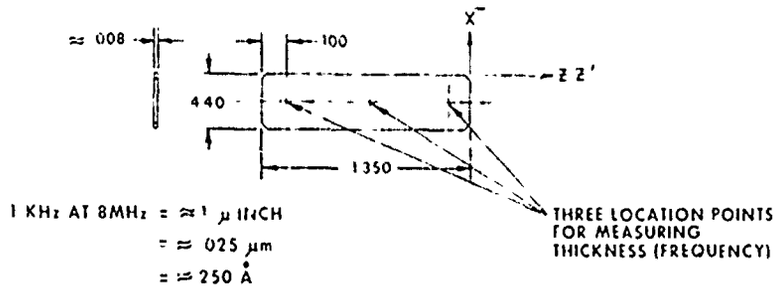
1. R. J. Byrne, Proceedings of the 24th Annual Symposium on Frequency Control, (1970).

2. P. C. Y Lee and W. J. Spencer, The Journal of the Acoustical Society of America, Vol. 45, No. 3, 637-645, March 1969.

3. N. C. Lias and D. W. Rudd, The Western Electric Engineer, Vol. XIII, No. 2, April 1969.

4. P. R. Thornton, Scanning Electron Microscopy, Chapman and Hall Ltd., London, 1968.





1 KHz AT 8MHz = ≈ 1 μ INCH
= ≈ 025 μm
= ≈ 250 Å

NOMINAL FREQUENCY : 8140 kH.
AT-CUT ORIENTATION : NOMINAL ± 1'
PARALLELISM : 8 kHz MAXIMUM
MEASURED AT THREE POINTS
AS INDICATED
MATERIAL . R-FACE CULTURED QUARTZ
SURFACE CONDITION . LAPPED AND ETCHED
END ABRASIVE MICRO GRADED 3 μm
ALUMINUM OXIDE

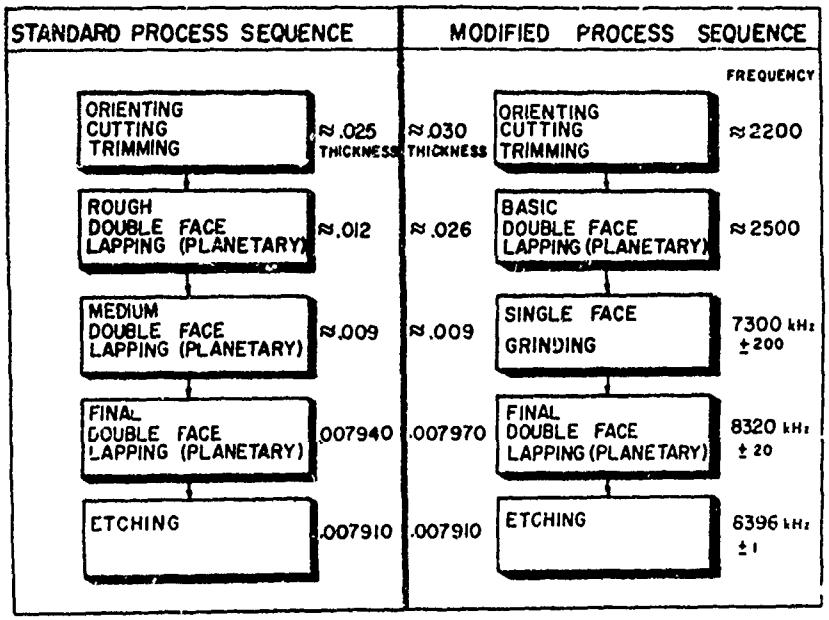


Figure 3 Process Sequences.

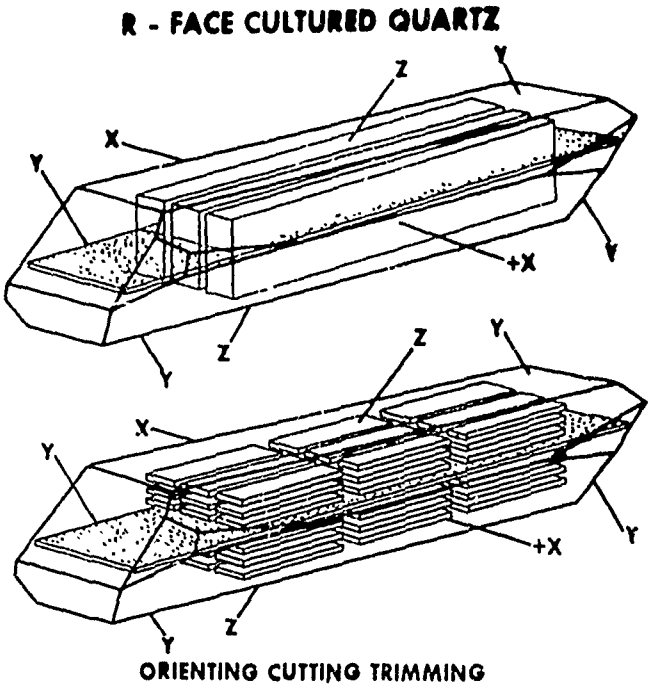


Figure 4 Orienting-Cutting-Trimming.

SINGLE FACE GRINDING

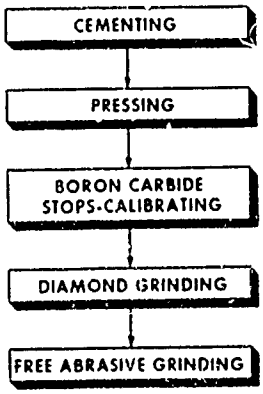


Figure 5 Single Face Grinding - Process Steps.

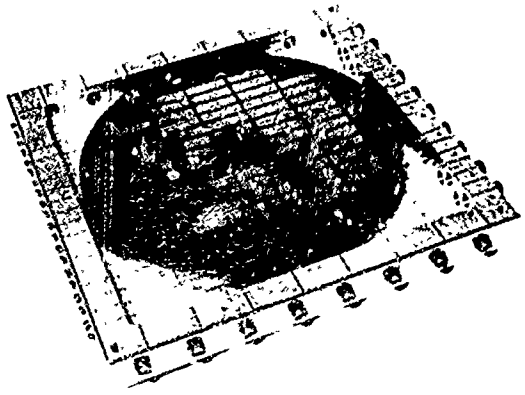


Figure 6 Single Face Grinding - Backing Plate and Alignment Fixture.

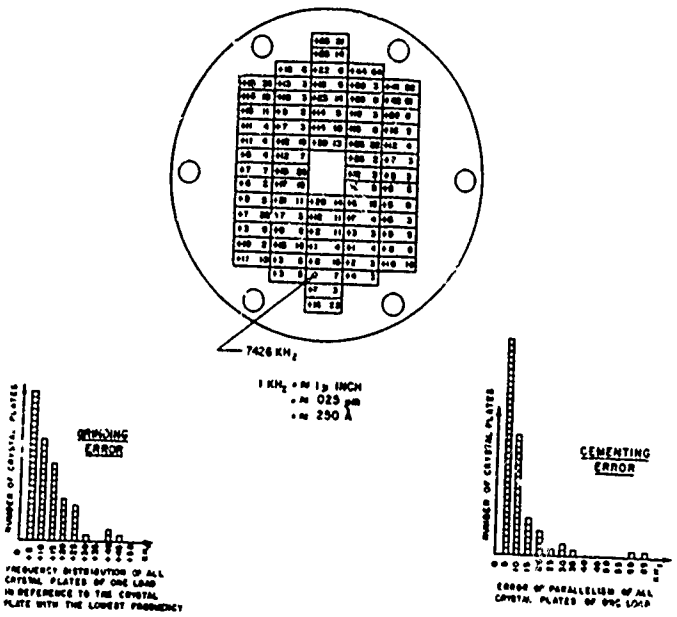


Fig. 7 SINGLE FACE GRINDING (CHARACTERISTIC RESULTS)

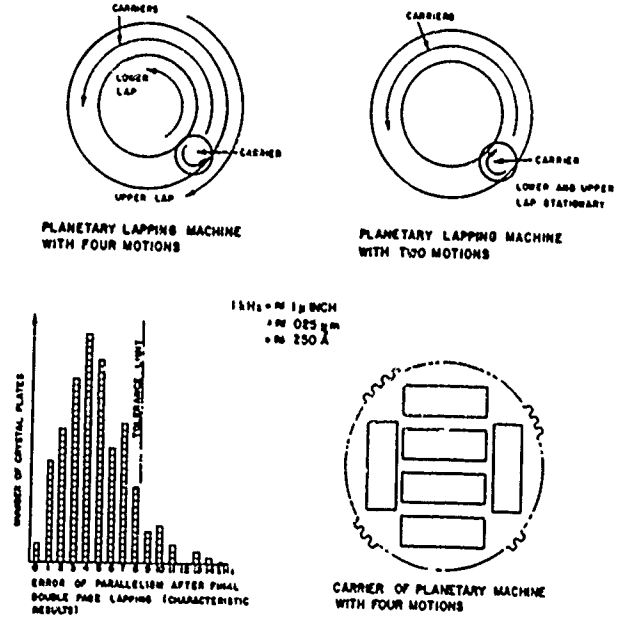


Figure 8 DOUBLE FACE LAPPING

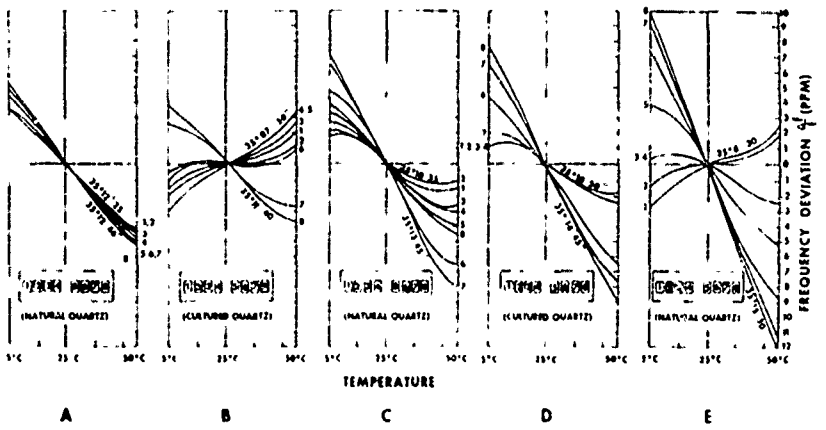


Figure 9

EFFECT OF AT-CUT ANGLE VARIATIONS ON FREQUENCY TEMPERATURE CHARACTERISTICS

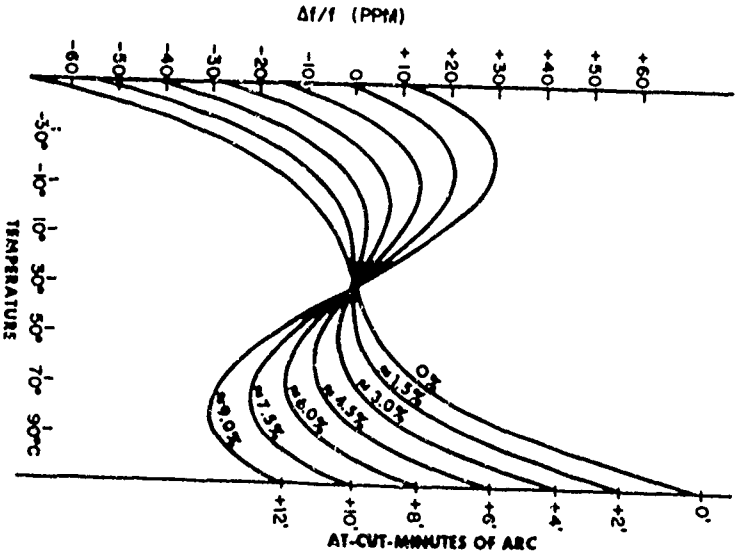


Figure 10 Effect of Mass Loading on Frequency Temperature Characteristics - Results of One Resonator.

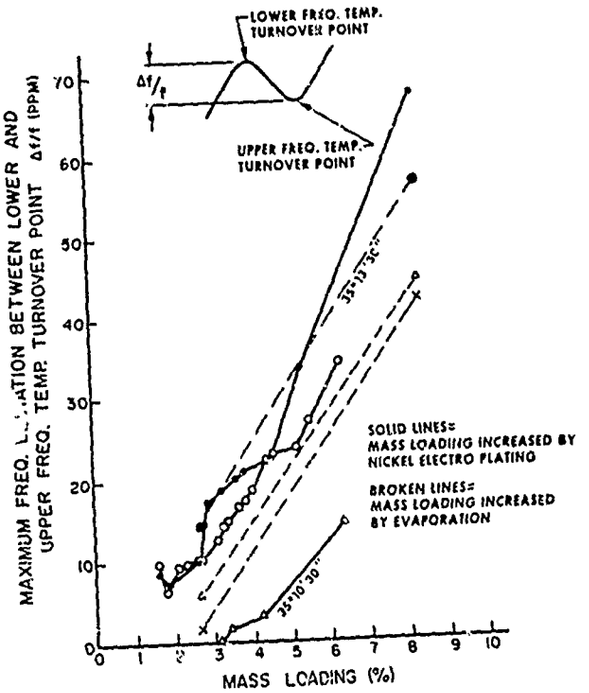
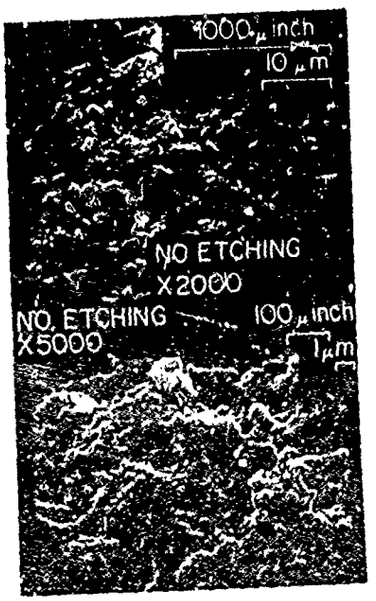
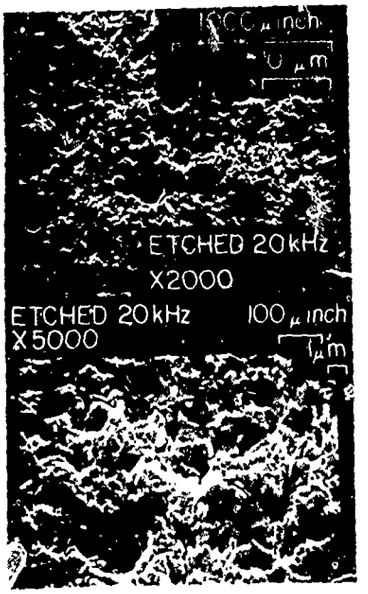
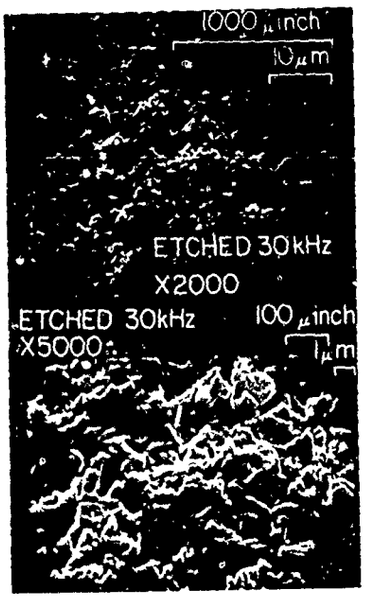


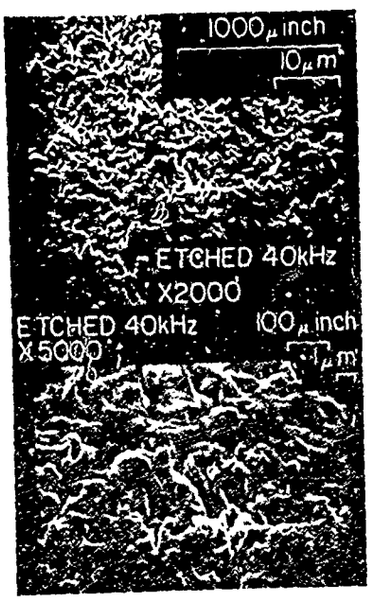
Figure 11 Effect of Mass Loading on Frequency Temperature Characteristics - Frequency Deviation Between Lower and Upper Turnover Point.

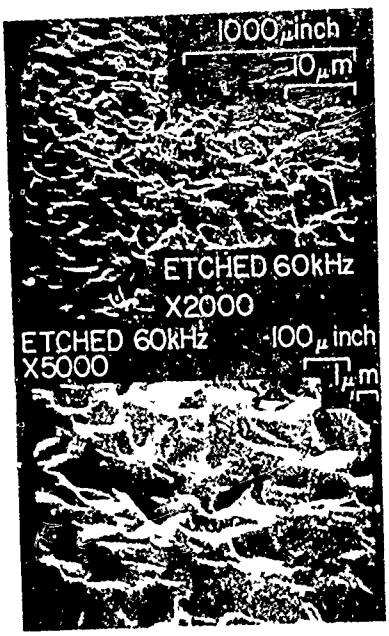
Figures 12a, 12b, 12c, 13a, 13b, 13c, 14a, 14b, 14c Etching Sequence - Surface Topology.

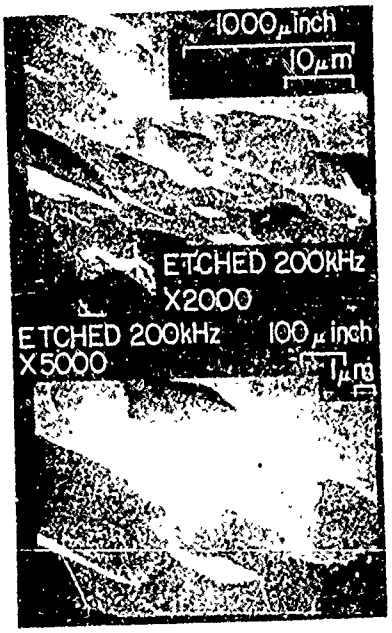


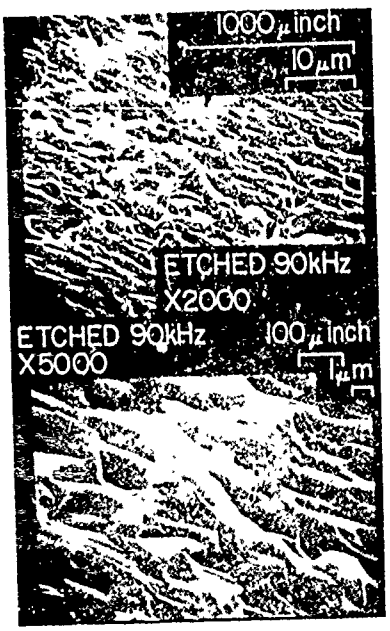




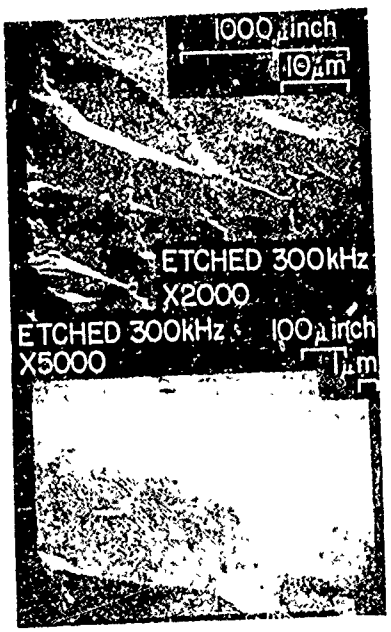


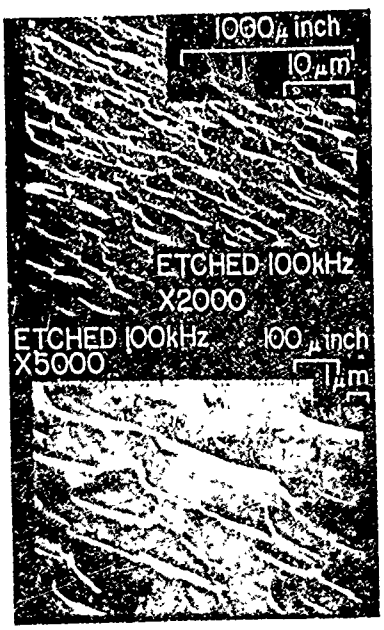






Reproduced from best available copy.





A TECHNIQUE FOR AUTOMATIC MONOLITHIC CRYSTAL FILTER FREQUENCY ADJUSTMENT

R. P. Grenier
Western Electric Co., Inc.
Merrimack Valley Works
North Andover, Massachusetts

Abstract

This paper describes a differential insertion loss measurement technique presently used by the Western Electric Co. for the automatic frequency adjustment of monolithic crystal filters. The information given places emphasis on the technique's practical as well as theoretical aspects.

Introduction

The technique described was developed as a result of a manufacturing shop requirement for a differential insertion loss measurement system suitable for controlling a deposition process acting on a monolithic crystal filter frequency adjustment. The shop requirement was such that the system had to automatically determine whether or not a filter needed adjustment or could be adjusted when it was inserted into the measuring circuit. If the filter could be adjusted, the system also had to automatically stop the deposition process when the filter was adjusted to a specific differential insertion loss frequency requirement.

Figure 1 shows the monolithic crystal filter involved and figure 2 shows this unit's desired loss vs frequency characteristics. This paper concerns the filter's adjustment to the requirement that the insertion loss at the fc + 200 Hz frequency exceed that at the fc + 1000 Hz frequency by precisely 3 db. Figure 3 shows the filter's loss vs frequency characteristics prior to adjustment. In this state the filter is adjusted to have all the desired characteristics except that it is approximately 50 Hz high in frequency. The lowering of frequency is accomplished by a simultaneous deposition of an approximately equal amount of gold on each of the filter's eight accessible electrodes.

The equipment used for the deposition process is a commercially available turbomolecular plating facility. The deposition process can be stopped with this facility by operation of a remotely located switch.

It was obvious at the beginning of this project that conventional differential insertion loss measurement techniques were not ideally suited for the machine control application required and that a new technique had to be developed. The following discussion describes the technique developed and how it was implemented by the design and construction of a system that has been in use since March of 1969.

Discussion

Figure 4 shows a block diagram of the basic circuit used to automatically determine the relationship of two rf signal levels. To simplify circuit description, it is assumed that relay switching occurs instantaneously and that the phase detector reacts only to the phase relationship of its input signals. It is also assumed that this phase detector has a positive dc voltage output when its input signals are in phase and a negative dc voltage output when its input signals are 180° out of phase. This circuit alternately applies the two rf signals to an amplitude detector through operation of the square wave generator driven relay. The amplitude detector output signal, which is a dc voltage with an amplitude proportional to the input rf signal level, is applied to the phase detector for phase comparison with the square wave generator signal. The result of this comparison is indicated by the dc voltmeter polarity reading. A positive voltmeter reading indicates the rf signal "2" voltage level exceeds that of rf signal "1" and a negative voltmeter reading indicates that the rf signal "1" voltage level exceeds that of rf signal "2". A signal is adjusted to have approximately the same voltage level as the other signal simply by adjusting its voltage level to the closest setting that causes a shift in the voltmeter polarity reading. Figure 5 shows the signal waveforms associated with this circuit.

Figure 6 shows a block diagram of the basic circuit used to automatically determine the magnitude of difference in rf signal voltage levels. To simplify circuit description, it will again be assumed that relay switching occurs instantaneously. The operation of this circuit is similar to the one previously described and the waveforms shown by figure 5 also apply. The exception in operation is that this circuit measures the magnitude of the ac voltage component contained in the amplitude detector output signal. This ac component voltage magnitude is proportional to the difference in rf signal voltage levels and is indicated by the ac voltmeter reading. A signal is adjusted to have the same voltage level as the other signal simply by adjusting its voltage level to the setting that produces a zero ac voltmeter reading.

Figure 7 shows a block diagram of a basic differential insertion loss measuring system using the two basic circuits previously described. This system is calibrated by setting the signal sources to the two test frequencies of interest and then adjusting the signal source output voltages to the same level. The latter calibration is accomplished by setting the calibrated adjustable attenuators to their 0 db positions, placing a short circuit across the A & B test terminals, adjusting one signal source voltage to the appropriate drive level and then adjusting the other signal source voltage level to produce a zero ac voltmeter reading. The ac voltmeter is calibrated in terms of db, after the rf signal voltage calibration, by noting its pointer deflection for various settings of one of the calibrated adjustable attenuators. After calibration, the short circuit across the test terminals is removed and the calibrated adjustable attenuator used for the ac voltmeter calibration is reset to the 0 db position. A measurement is made by connecting the device under test to the test terminals, adjusting the appropriate calibrated adjustable attenuator to obtain a zero ac voltmeter reading and then noting the attenuator setting which is the differential insertion loss of the device. The appropriate attenuator is determined when the device is connected to the test terminals by noting the dc voltmeter polarity reading. For example, a positive reading indicates that the device's insertion loss at the rf signal "1" frequency exceeds that at the rf signal "2" frequency and that the rf signal "2" attenuator must be operated to obtain a zero ac voltmeter reading. A go no-go type measurement is made by pre-setting the appropriate attenuator to the desired differential insertion loss value, connecting the device under measurement to the test terminals and then observing the meters. The ac voltmeter indicates the degree of differential insertion loss deviation from the attenuator setting and the dc voltmeter indicates the deviation's polarity.

The basic circuits described assume ideal conditions and are practical for many applications. However, rf signal switching does not occur instantaneously, switching transients do exist and settling time is required when testing highly reactive networks. Figure 8 shows a typical amplitude detector output signal waveform when a monolithic crystal filter is connected into the measuring circuit. It is evident that if a very precise measurement is to be made, the effects of switching and filter ringing have to be removed from the measurement. This can be accomplished with a circuit to sample a portion of the detected signal at a time free from the unwanted effects and hold this sample until a new one is taken. Figure 8 shows the signal waveforms associated with this circuit. The sample and hold circuit output signal is treated in the same fashion as described for the basic circuits except that phase comparison is made relative to a signal delayed with respect to the square wave generator relay energizing signal. The delayed signal's leading and trailing edges are coincident with the sampling pulse's leading edge and are therefore coincident with the sample and hold circuit output signal's leading and trailing edges. The sample and hold circuit output signal is either in phase or 180° out of phase with the delayed signal as determined by the voltage level relationship of the two rf signals applied to the amplitude detector.

Figure 9 shows the basic block diagram of the system that has been constructed and successfully employed for the application previously described. This system is calibrated by performing the following operations in the sequence given.

(1) Placement of a short circuit across the A & B test terminals.
(2) Adjustment of the fc + 200 Hz signal voltage level to a predetermined value.
(3) Operation of the "CAL-OPERATE" switch to the "CAL" position.
(4) Operation of the "METER GAIN" control to the maximum sensitivity position.
(5) Adjustment of the fc + 1000 Hz signal voltage level to obtain an ac voltmeter null reading. This adjustment sets the voltage levels of the two signal sources to within .01 db of each other.
(6) Operation of the "CAL-OPERATE" switch to the "OPERATE" position. This operation causes the fc + 1000 Hz signal voltage level, applied to the square wave driven relay, to be 3 db below that of the fc + 200 Hz signal voltage level.
(7) Adjustment of the "METER GAIN" control to obtain a full scale ac voltmeter indication. This adjustment provides an ac voltmeter calibration that is employed to determine the degree of filter differential insertion loss deviation when this deviation is within 3 db of the desired differential insertion loss.
(8) Removal of the short circuit across the A & B test terminals.

The calibrated system produces a zero ac voltmeter reading and is at the point of a phase reversal between the sample and hold circuit output signal and delayed signal when a filter with a differential insertion loss of 3 db is connected across the test terminals. This is so because the filter attenuates the fc + 200 Hz signal voltage level by 3 db more than that of the fc + 1000 Hz signal, the same amount by which the fc + 200 Hz signal voltage level exceeds that of the fc + 1000 Hz signal as applied to the filter. The machine control relay, which is controlled by the phase detector output signal, is in a de-energized condition when a filter differential insertion loss exceeds 3 db and is in an energized condition when a filter differential insertion loss is less than 3 db. The machine control relay illuminates a "PLATING ON" indicator lamp and allows the gold deposition process to occur when it is in the de-energized condition. This relay illuminates a "PLATING OFF" indicator lamp and does not allow the gold deposition process to occur when it is in an energized condition. The calibrated system's normal operating procedure is as follows:

(1) The unit to be adjusted is connected to the test terminals.

(2) The deposition chamber is closed and evacuated to obtain the desired pressure level.

(3) The system's indicator lamps and ac voltmeter are observed to determine whether or not the unit can or need be adjusted. A "PLATING OFF" lamp illumination indicates that the unit cannot be adjusted. A low ac voltmeter deflection within a prescribed limit indicates that the unit does not require adjustment. A "PLATING ON" lamp illumination in conjunction with an ac voltmeter deflection above a prescribed limit indicates that the unit can and needs to be adjusted.

(4) When a unit can and needs to be adjusted, which is the normal case, a plating facility "START" switch is operated to the "ON" position. This operation applies heat to the gold source in the deposition chamber and opens a solenoid operated shutter placed between the gold source and unit under adjustment so that the gold deposition process is started.

(5) The deposition process lowers the unit's frequency characteristics and consequently its differential insertion loss. When the unit's differential insertion loss reaches a value that is .025 db less than 3 db, the machine control relay energizes. This relay energization stops the plating process by turning off the plating facility gold heat source and operating the solenoid operated shutter so that it is placed between the gold source and the unit under adjustment. The relay energization also extinguishes the "PLATING ON" lamp and illuminates the "PLATING OFF" lamp so that the operator knows when the adjustment is completed.

(6) The ac voltmeter deflection is observed to assure that the unit has been adjusted to within the prescribed limits.

(7) The deposition chamber pressure is released, the unit in the chamber is replaced with the next unit to be adjusted and the operating procedure is repeated.

Figure 10 shows a detailed block diagram of the operating system and figure 11 shows this system's physical appearance. The system's capabilities are such that monolithic crystal filters are frequency adjusted for the desired 3 db differential insertion loss to within 1 part in 10^7. This frequency accuracy represents a differential insertion loss deviation of .05 db.

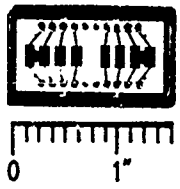
Conclusion

We believe that a differential insertion loss measurement technique suitable for machine control application has been developed. The technique's feasibility has been proven by the construction and employment of a system for automatic monolithic crystal filter frequency adjustments. We also believe that the technique has many practical applications and with slight inovations in equipment design it may be employed for other transmission network measurements such as transmission and return loss.

REFERENCES

A. J. Miller, Proceedings of the 24th Annual Symposium on Frequency Control (1970)

R. J. Byrne, Proceedings of the 24th Annual Symposium on Frequency Control (1970)



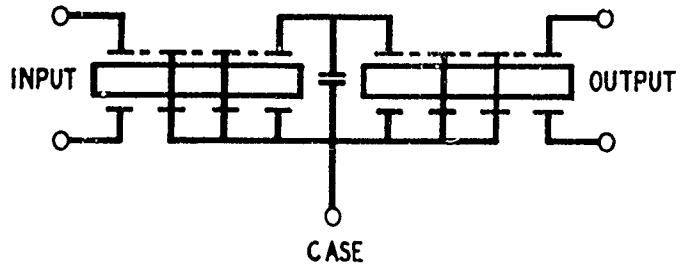
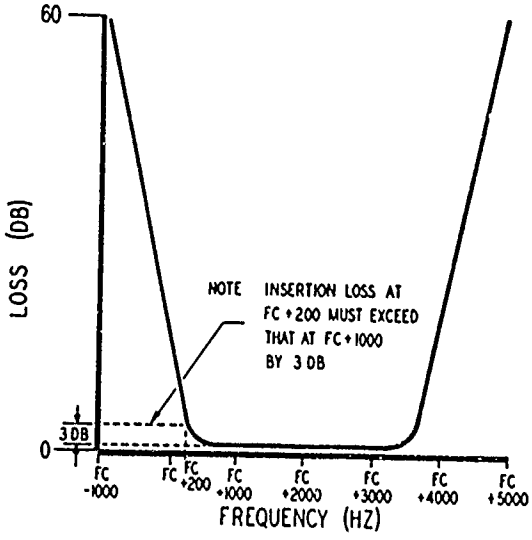
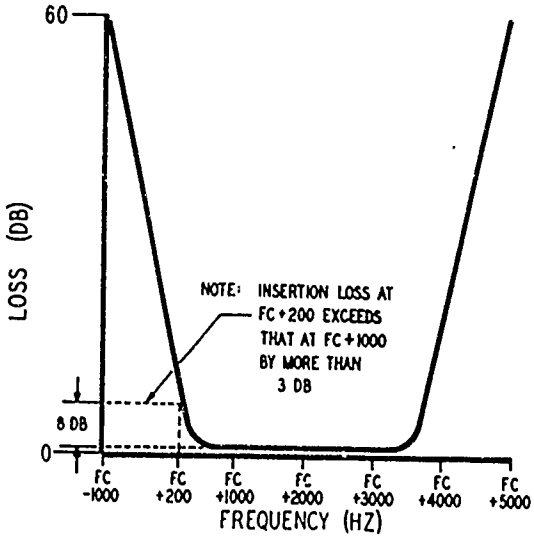


FIGURE 1

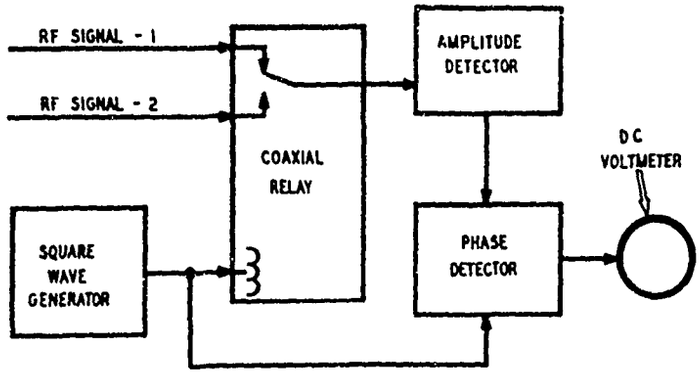


CHANNEL	FC (KHz)
1	8140
2	8144
3	8148
4	8152
5	8156
6	8160
7	8164
8	8168
9	8172
10	8176
11	8180
12	8184

DESIRED A6 CHANNEL FILTER CHARACTERISTICS
FIGURE 2

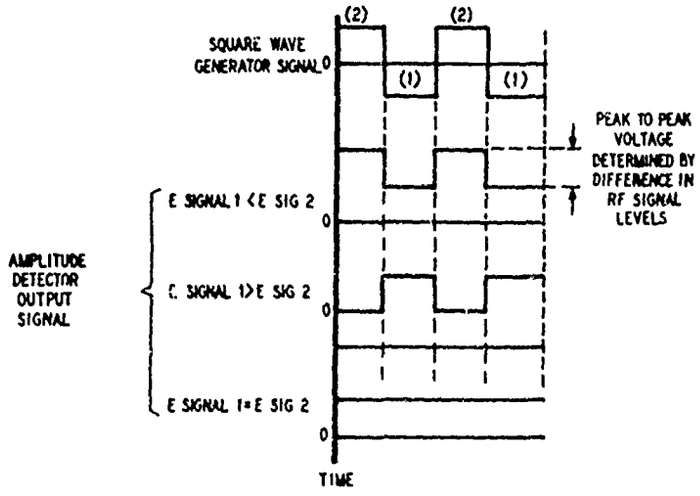


A6 CHANNEL FILTER CHARACTERISTICS
PRIOR TO FINAL FREQUENCY ADJUSTMENT
FIGURE 3



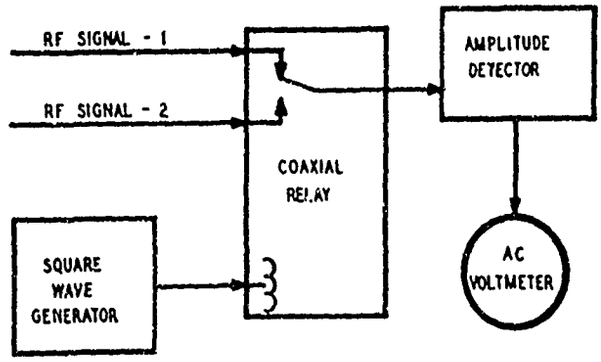
BASIC CIRCUIT FOR DETERMINING RELATIONSHIP
OF TWO RF SIGNAL VOLTAGE LEVELS

FIGURE 4



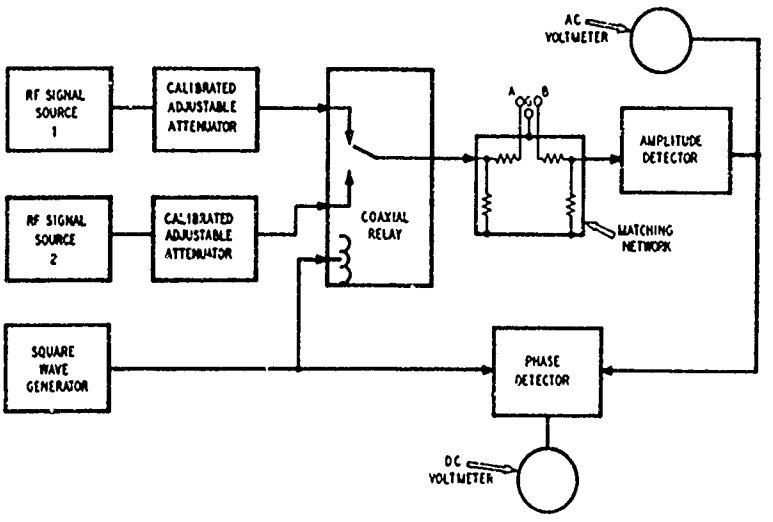
AMPLITUDE DETECTOR OUTPUT SIGNAL RELATIONSHIP
TO SQUARE WAVE GENERATOR OUTPUT SIGNAL

FIGURE 5



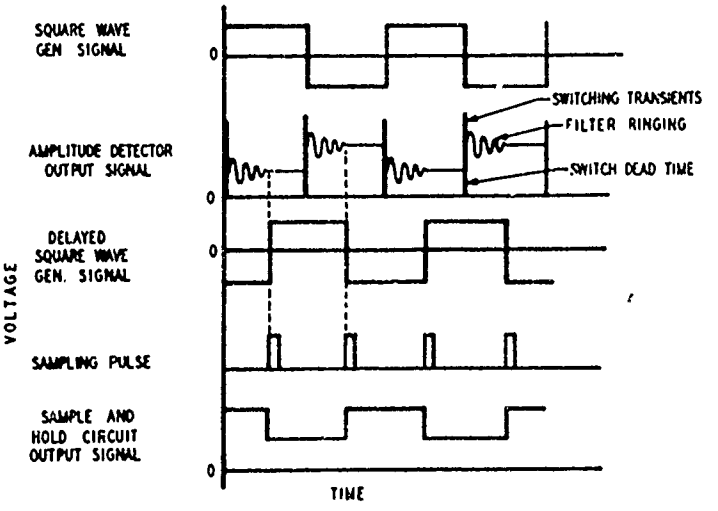
BASIC CIRCUIT FOR DETERMINING MAGNITUDE
OF DIFFERENCE IN RF SIGNAL LEVELS

FIGURE 6



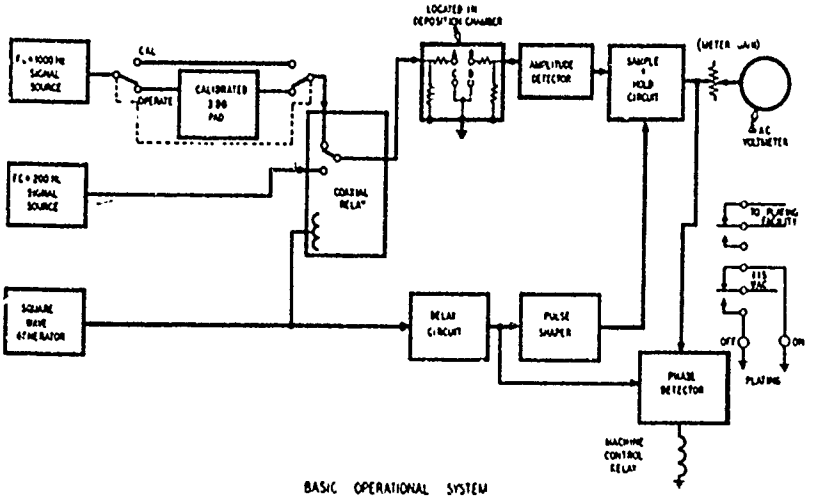
BASIC DIFFERENTIAL INSERTION LOSS MEASURING SYSTEM

FIGURE 7



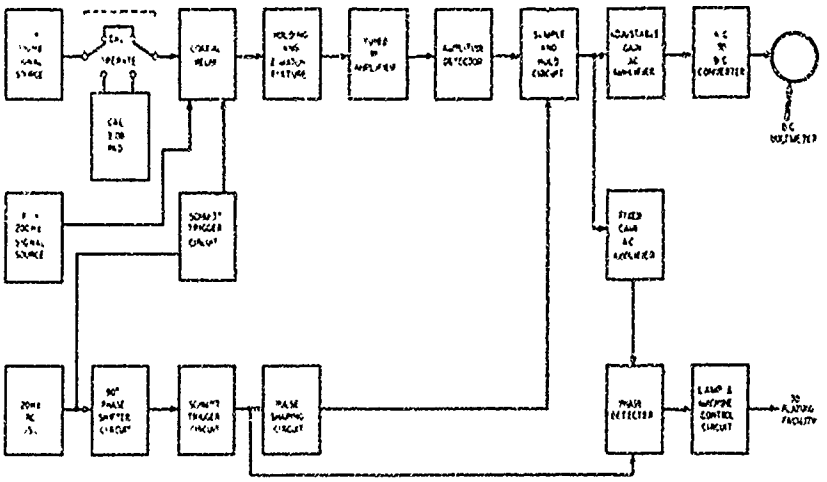
SIGNAL RELATIONSHIPS

FIGURE 8



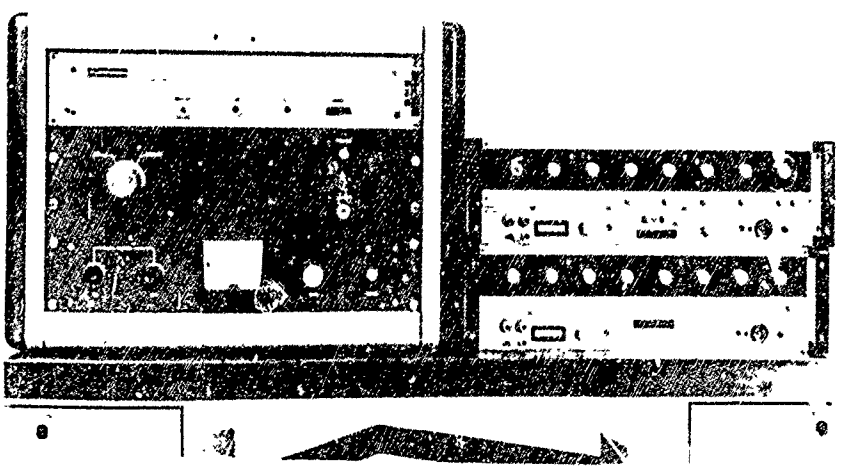
BASIC OPERATIONAL SYSTEM BLOCK DIAGRAM

FIGURE 9



OPERATIONAL SYSTEM BLOCK DIAGRAM
FIGURE 10

FIGURE 11



AUGER SPECTROSCOPY FOR STUDIES OF THE AGING FACTORS OF QUARTZ CRYSTAL RESONATORS

Gary W. Simmons, Walter H. Hicklin, and R. K. Hart
Georgia Institute of Technology
Atlanta, Georgia

Summary

The recent development of Auger spectroscopy has made it possible to detect and to identify surface impurities in concentrations as low as 10^{13} atoms/cm^2 (0.01 atomic layers). This technique can be readily used to monitor the purity of the interfaces in each step of resonator fabrication. We have initially investigated the surfaces of two quartz samples that were prepared and processed according to procedures used for resonator fabrication. The first sample was cleaned in an ultrasonic chromic acid bath. Several microns were etched from the second sample with amonium hydrogen fluoride. Auger spectra were measured for the samples after the vacuum system bakeout of 250°C and an initial outgassing of the specimens at 350°C in ultrahigh vacuum. Auger analysis showed the presence of carbon contamination for both samples. Similar results were observed for magnesium oxide samples that were used in another study. Since the treatment of the quartz samples in this preliminary investigation did not differ appreciably from the actual fabrication procedures, it is suggested that a layer of some form of carbon exists between the quartz and metal electrode. It is possible at present only to speculate about the effects of this impurity on the aging characteristics of resonators. Since rf-sputtering of magnesium oxide in argon was successful in removing carbon contamination, it is suggested that quartz surfaces can be cleaned *in situ* by this procedure.

Introduction

The changes in the resonance frequency of quartz crystal units due to aging has been a major problem in their use for frequency control applications. Improvements in stability have been made in the past by continued step-wise refinements in processing and fabrication techniques. Progress has now reached the stage where there is considerable difficulty in the identification of the factors that are responsible for crystal aging. The frequency stability should ultimately be restricted by the intrinsic properties of quartz. This ultimate, however, can not be realized until all of the extrinsic contributions to aging are firstly identified and then minimized.

Steps toward a fundamental approach have been initiated by the application of ultrahigh vacuum techniques in the work of Hafner and Blewer[1] and in the studies by Warner, Fraser and Stockbridge[2]. The recent development of Auger spectroscopy provides an analytical tool which can be readily used to complement the approach taken by these investigators. The present paper is an introduction to the Auger effect, and we describe the experimental techniques used in Auger spectroscopy of surfaces. Some preliminary results of surface analysis of quartz and magnesium oxide are included to illustrate the potential for using Auger spectroscopy in aging studies.

Auger Effect

When a material is bombarded with electrons, it is possible to ionize the inner energy levels of its atoms. The vacancy created by this ionization is then filled by an electron from one of the higher levels. The energy emitted in such a transition is equal to the difference in the ionization energy of the two levels. This transition energy results either in the emission of a photon (characteristic X-ray) or the ejection of an electron from another shell. The latter radiationless transition is called the Auger effect, and the emitted electron is called an Auger electron. For example, if a vacancy in the L_{II} level of an atom is filled by an electron from the M_I shell, it is possible that a second electron will be emitted from the M_{II} level. The shorthand notation for this process is written $L_{II}M_IM_{II}$. Figure 1 illustrates the Auger effect for a $L_{II}M_IM_{II}$ transition in a solid and a KL_IL_I transition from a surface impurity. It should be noted that there are six possible transitions of the type $L_{II}MM$ for the example shown in Fig. 1, and there are three possible KLL transitions for the impurity. Additional transitions are allowed when all of the possible initial vacancies are considered.

Energies for the possible Auger transitions can be calculated from ground state binding energies. The Auger electron, however, is emitted from an excited atom due to the initial ionization. Burhop[3] has suggested that the shift in energy levels for the atom of the excited state could be obtained by using energy levels

for the atom of the next higher atomic number for the ejected electron. This empirical rule is given by $E_{WXY} = E_W(Z) - E_X(Z) - E_Y(Z + 1)$, where E_{WXY} is the energy of the emitted Auger electron, $E_W(Z)$ is the ionization energy for the initial vacancy in level W for an atomic number Z, $E_X(Z)$ is the energy level from which the vacancy in level W is filled, and the Auger electron is emitted from the energy level E_Y as determined from the next highest atomic number (Z + 1). The calculated Auger energies considered in this paper have been determined with this relationship.

Surface Analysis by Auger Spectroscopy

Although the Auger effect was discovered in 1925[4], it has only been recently applied to surface analysis. Lander[5] suggested in 1953 that the excitation of Auger transitions by low energy electrons could be used as a method for identifying surface impurities. The instrumental developments of Tharp and Scheibner[6] and of Harris[7] have provided the techniques suitable for Auger spectroscopy of surfaces.

Bishop and Riviere[8] have estimated the efficiency of production of electrons emitted by the Auger process for oxygen adsorbed on copper. These determinations were based on the approximate relationships that have been derived for calculating the efficiency of characteristic X-ray production in the electron microprobe. In a typical LEED system with a primary beam of 1 μA at normal incidence, the total Auger current from chemisorbed oxygen was estimated to be about $2 \times 10^{-10}\ \alpha$A, where α is the fractional surface coverage. Although this estimated Auger yield is low, Weber and Johnson[9] have demonstrated that a surface coverage of 0.02 monolayer Cs can be detected on a Si surface.

A number of factors contribute to the energies and shape of the Auger peaks. The line width depends upon the breadth of the participating energy levels and the inelastic losses which the Auger electron may suffer in escaping from the sample. The type of binding between the element of interest and the other atoms in the solid modifies both the energy levels and the density of states in the valence band. These latter two effects also contribute to the line shape and transition energies. It is possible for the emitted electron to excite bulk and surface plasmons. These energy loss peaks are observed on the low energy side of the Auger peak, and discretion is required to avoid confusing these peaks with Auger peaks.

Sulfur, carbon and oxygen are the most common impurities encountered in our studies. Figure 2 shows the Auger spectra observed for these elements. The sulfur peak was obtained from the diffusion of this element to the surface of a tantalum sample during heating at 700°-800°C. The Auger electron distribution from carbon was observed as a contaminant on a titanium surface. The characteristic oxygen spectrum shown in Fig. 2 was obtained from a clean magnesium oxide surface. These elements are characterized by their line shape and by the energies at which they appear along the retarding field plot. The asymmetric shape of the carbon peak is a characteristic feature of this element. Since chemical effects would not be expected to drastically alter the energies and line shapes, the Auger spectra shown in Fig. 2 should be generally representative of these elements.

Experimental Techniques

The essential features of the apparatus[6,7,10] used for measuring energy distributions and Auger spectra are described below. Experiments are performed in a titanium pumped ultrahigh vacuum system with a base pressure of 2×10^{-10} torr. The optics and electronic arrangement used for energy analysis are shown schematically in Fig. 3. Electrons that are scattered and emitted from the sample pass through three wire mesh grids before reaching the collector. The closest grid to the sample is held at ground potential and serves to maintain field free conditions between the sample and analyzer. The grid adjacent to the collector is also at ground potential and acts as an electrostatic capacitive shield between the collector and the center, suppressor, grid. Improved energy resolution is obtained by connecting the latter two grids, if capacitive nulling is provided to reduce the capacitive coupling between the collector and these analyzer grids.

Energy distributions are measured by applying a retarding field voltage to the middle grid along with a small modulating voltage. With the lock-in amplifier tuned to the frequency of the modulating voltage, the retarding voltage and the amplifier output are fed to the X and Y channels of a recorder. The resulting energy distribution represents the derivative of the current reaching the collector as a function of the retarding field. In order to observe Auger electrons, considerable amplification of the energy distribution is required. Since the Auger peaks are present in a high background signal, it is convenient to take the derivative of the energy distribution. In effect, the second derivative of the current reaching the collector is measured as a function of the suppressor grid voltage. In order to obtain the second derivative, the signal channel of a phase sensitive detector is tuned to a frequency double that applied to the retarding field grid. This method of energy analysis has the advantage that for a gradually varying background, the signal is essentially constant except in the region of Auger electrons. The instrumental factors that effect sensitivity and energy resolution have been considered recently by Taylor[11].

Results and Discussion

Quartz

Surface analysis was made for two separate quartz samples. The first sample was cleaned in an ultrasonic chromic acid bath, and several microns were etched from the second sample with ammonium hydrogen fluoride. The samples were mounted on a tantalum heater and placed in the ultrahigh vacuum system for observation. Auger spectra were measured for the samples after a vacuum system bakeout of 250°C and an initial outgassing of the specimens at 350°C in ultrahigh vacuum. Although there were considerable differences in surface preparation, the surface analysis for the two samples was essentially the same. Figure 4 shows a typical Auger spectrum that was observed for the two quartz samples described above. Peaks from silicon, carbon and oxygen are readily identified.

At present only the elemental constituents of the surface can be determined, since it is not possible to specify the chemical composition from this data. Consequently, the carbon peak may result from either the presence of elemental carbon or a surface carbonate. Similar difficulties are likewise involved in determining the chemical nature of oxygen on this surface. The oxygen peak may be attributed to the quartz substrate or to a radical containing oxygen such as OH^- or $CO_3^=$. The relatively low intensity of the silicon peak is probably due to scattering of the silicon Auger electrons by the layer of contamination. The intensity ratio of the silicon peak to the oxygen peak could be used as an indication of the presence of OH^- or H_2O, since one would expect this ratio to increase as these contaminants were removed from the quartz surface.

There are two possible sources for the contamination of the quartz by carbon. First, atmospheric adsorption of carbon dioxide is likely to occur during transfer of the samples from the cleaning bath to the vacuum system. Second, it is possible that the specimens adsorbed residual gases in the vacuum chamber. Undoubtedly, the quartz surfaces adsorbed H_2O during sample preparation, and the residual hydroxyl groups present after outgassing may have contributed to the oxygen Auger peak.

The contamination as described above for quartz was found to be stable at temperatures as high as 700°C. Since the treatment of the quartz samples in this preliminary investigation did not differ appreciably from the actual fabrication procedures, it is suggested that a layer of some form of carbon exists between the quartz and the metal electrode. It is possible at present only to speculate about the effects, if any, of this impurity on the aging characteristics of resonators.

Magnesium Oxide

Studies of MgO surfaces were conducted in conjunction with a separate program. It is of interest to present data on MgO for comparison with quartz, since these two oxides show a similar susceptibility to carbon contamination. Surface analysis was performed on samples that were cleaved in air and for specimens that were mechanically cut and polished. A typical spectrum is shown in Fig. 5 for MgO following vacuum system bakeout and sample outgassing at 750°C. Although both the cleaved and mechanically prepared samples showed the presence of carbon, the intensity ratio of carbon to oxygen peaks for the cleaved samples was less than the carbon to oxygen peak ratio for the polished specimens. Since attempts to remove the carbon from the MgO surfaces by extensive heating at 750°C failed, the MgO samples were rf-sputtered in argon. A source of rf power of approximately 350 volts (rms) with a frequency of 20 MHz was applied to the specimen holder in 5×10^{-3} torr of argon. A self sustained plasma was established after initiation by electron emission from a heated filament. An Auger spectrum of MgO cleaned by this technique is shown in Fig. 6. Clean magnesium oxide surfaces were found to be relatively insensitive to background gases at 10^{-10} torr for several days, which suggests that the primary source of carbon is the adsorption of atmospheric CO_2 during sample preparation.

Conclusions

1. Auger spectroscopy when combined with ultrahigh vacuum techniques can be effectively applied to surface related studies of quartz resonators.

2. The oxides of silicon and magnesium are highly susceptible to some form of carbon contamination.

3. Low power rf-sputtering by argon can be readily used to remove contaminated layers from oxides such as MgO and quartz.

4. The results reported here indicate only that carbon is present on quartz during resonator fabrication. We have made no attempts to establish the effect, if any, of this contamination on the aging characteristics of quartz resonators.

References

1. E. Hafner and R. S. Blewer, Proc. IEEE (letters) 56, 366 (1968).

2. A. W. Warner, D. B. Fraser, and C. D. Stockbridge, IEEE Trans. Sonics and Ultrasonics 12, 52 (1965).

3. E. Burhop, "The Auger Effect and Other Radiationless Transitions", Cambridge University Press, New York, 1952.

4. P. Auger, Comptes Rendus 180, 65 (1925).

5. J. J. Lander, Phys. Rev. 91, 1382 (1953).

6. L. N. Tharp and E. J. Scheibner, J. Appl. Phys, 38, 3320 (1967).

7. L. A. Harris, J. Appl. Phys. 39, 1419 (1968).

8. H. E. Bishop and J. C. Riviere, J. Appl. Phys. 40, 1740 (1969).

9. R. E. Weber and A. L. Johnson, J. Appl. Phys. 40, 314 (1969).

10. P. W. Palmberg, Appl. Phys. Letters 13, 183 (1968).

11. N. J. Taylor, Rev. Sci. Instrum. 40, 792 (1969)

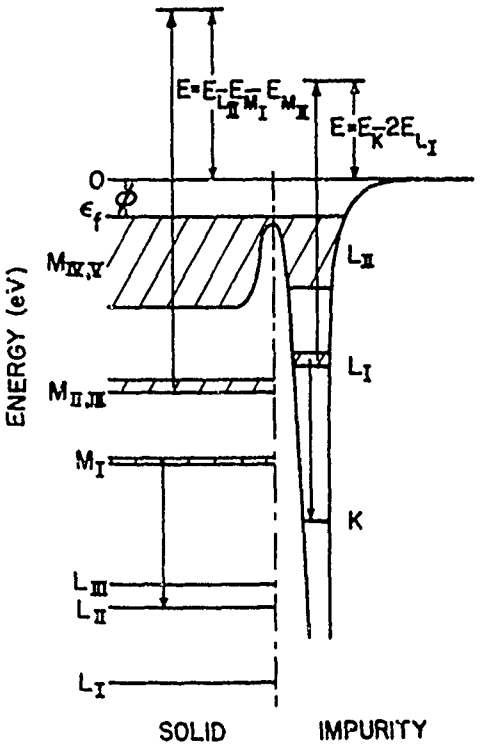


FIGURE 1

Diagram showing the origin of Auger electrons that result from the ionization of atoms in a solid and from a surface impurity.

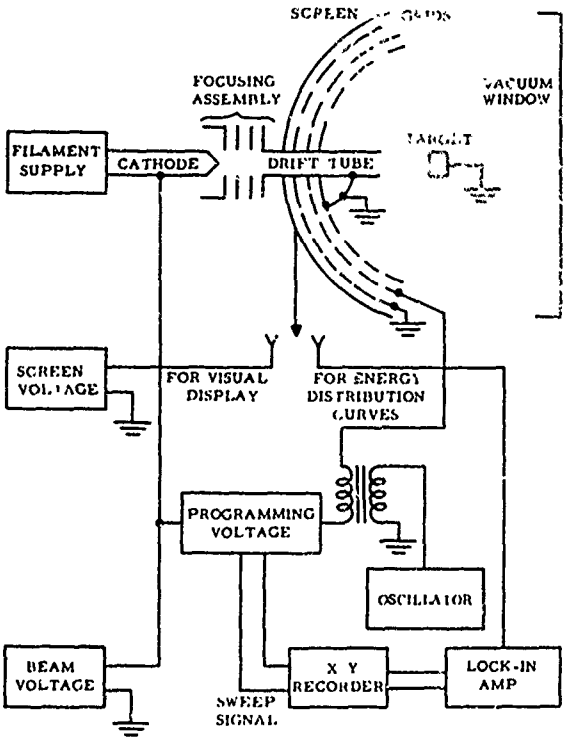


FIGURE 3

Schematic of experimental arrangement used for Auger spectroscopy of surfaces.

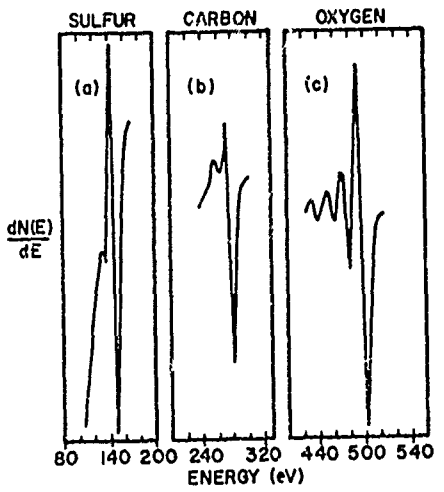


FIGURE 2

Auger spectrum for (a) sulfur, (b) carbon and (c) oxygen.

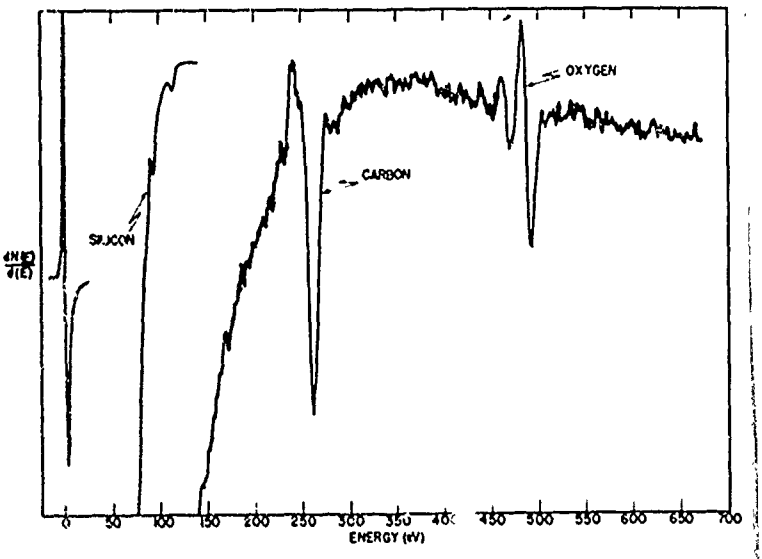


FIGURE 4

Auger spectrum of quartz showing carbon contamination.

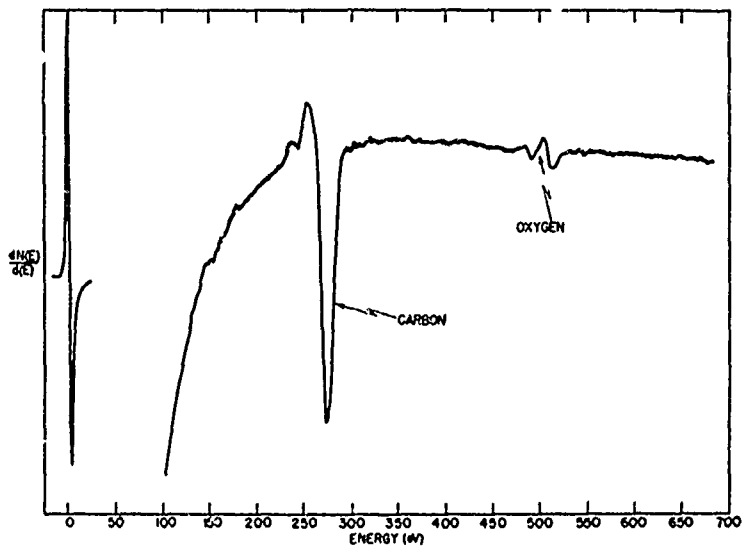


FIGURE 5

Auger spectrum of magnesium oxide showing carbon contamination.

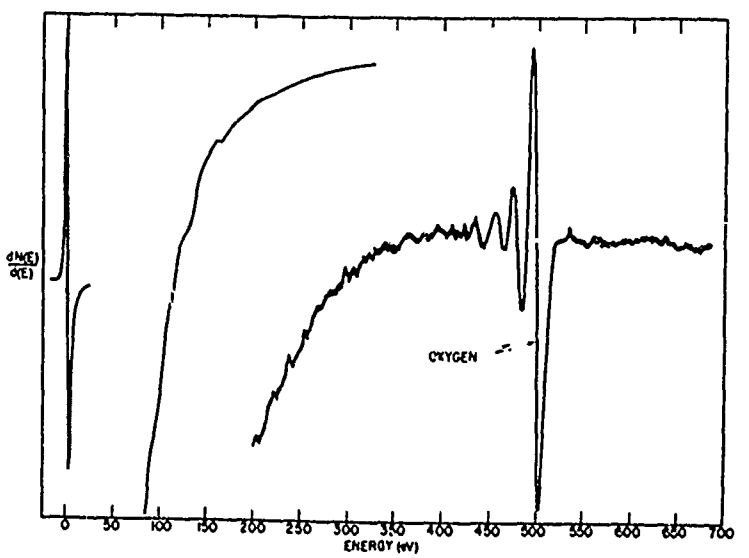


FIGURE 6

Auger spectrum of magnesium oxide cleaned by rf-sputtering.

5MHZ BT CUT RESONATORS

Jerry Leach
Hewlett-Packard Laboratories
Palo Alto, California

Summary

Design parameters for circular BT cut resonators are sparse. Some data by Seed, Spencer, and others has been published but in comparison to the information published on AT cuts, the BT cut has been somewhat neglected.

In a program to develop a quartz resonator pressure transducer, it became evident that the BT cut has a number of features that make it superior to other orientations. Of primary importance is the fact that the temperature coefficient of pressure sensitivity for BT cuts is very nearly zero.

To make the transducer useful over a wide temperature range, it is necessary to eliminate the effects of temperature on the resonant frequency of the device. This can be done by using another BT cut as a reference. The orientation angle vs. the turn-over point temperature for BT cuts is such that one can easily achieve a good match of frequency-temperature characteristics.

A study was undertaken to design a 5 MHz, third overtone, BT cut resonator having a diameter of 15mm and mounted in a cold-weld enclosure. A number of designs were fabricated and evaluated over a temperature range from 0 to +125° Celsius with particular attention being given to the Q of the resonator and to the extent to which the third overtone thickness shear mode of the resonator was coupled to other modes of vibration. As have previous investigators, we found that the orientation angle, radius of curvature on the resonator surface, dimensional ratio, and the overtone of operation are reflected in the value of the temperature coefficients of frequency.

A plot of the observed temperature of zero temperature coefficient of frequency vs. the polar angle for the BT cut was made for a number of different resonator design parameters. Using this plot, we were able to predict turn-over points as the resonator design was optimized.

During the study, measurements of the amplitude and frequency of the anharmonic overtones were made as a function of the radius of curvature.

The above data allowed selection of a final design which has essentially a constant impedance over the temperature range from 0 to +125° Celsius. The crystal parameters, Q, drive level sensitivity, and short term stability for some designs will be presented.

Introduction

Seed[1] has shown that the BT cut resonator exhibits a very high value of Q but that the resonator is very sensitive to finite plate effects. This is evident in the fact that the crystal turn-over point is greatly altered when the mode of operation is changed from the fundamental to the third overtone and when the surface of the resonator is contoured.

Although these effects can increase the difficulties normally associated with the design and fabrication of high precision resonators, the BT orientation offers some distinct advantages which are utilized in the quartz resonator pressure transducer[2]. The transducer is shown in Fig. 1. A 5 MHz BT cut resonator is constructed as an integral part of a crystalline quartz cylinder so that application of an external pressure will uniformly and elastically stress the resonator periphery. This stress causes a nominal change in the resonator frequency of -1.5Hz/psi and at 10,000 psi the frequency has been linearly decreased by 0.3 percent. The BT orientation is used in this device primarily because the temperature coefficient of pressure sensitivity is nearly two orders of magnitude smaller than that of the AT cut. Quartz end caps are used to seal the cylinder bore and thus isolate the resonator surfaces from the pressure media. With the transducer in a laboratory instrument and operating at its turn-over point under temperature controlled conditions, a resolution of ±0.001 psi and an absolute accuracy of ±0.01 psi can be obtained. To extend the usefulness of the device beyond the laboratory, a scheme to eliminate thermal effects is required. One method to achieve temperature independent operation of the transducer for applications where high resolution pressure measurements must be made and where the ambient temperature can vary as much as 100°C, is to incorporate it into a sonde along with another BT cut which serves as a reference crystal. If the reference can be crystallographically oriented to have the same temperature coefficients of frequency as the

transducer and then by some means it is thermally coupled to the transducer, the frequency difference between the two resonators is then only a function of ambient pressure on the transducer. Using this method requires that the reference have as high a Q as possible and that it be free from the effects of coupled modes which cause anomilies in the frequency-temperature characteristics. It was desirable that these requirements be met over a temperature range of 0 to +125°C.

Using Seed's data as a starting place and with the information already gathered on the transducer, a study of BT cuts was initiated to determine whether or not a suitable reference crystal design could be found and to determine to what extent one could reasonably hope to match the temperature coefficients of the reference and transducer. All resonators evaluated in this study were 5 MHz third overtone units 15mm in diameter. All blanks were angle corrected to ±15 seconds of arc. Both plano-convex and bi-convex designs were fabricated and tested. All units were optically polished and processed using state of the art precision crystal finishing techniques.

Test Equipment

Fig. 2 shows the equipment used in measuring the temperature coefficients of each resonator design. The system is a variation of the crystal parameter recording system described by Adams[3] at a previous frequency control symposium. An oil bath oven accommodates a fixture holding twelve resonators with each resonator having its own oscillator. The bath is programmed to cover the temperature range of 0 to +125°C in thirty hours. Through an input scanner and a stepping switch, the resonators are sequentially sampled by a frequency counter and the output is stored on magnetic tape along with temperature information provided by a quartz thermometer which continuously monitors the oil bath. Using computer methods, the data was reduced and either printed or plotted out.

To determine the presence of coupled modes, measurements of resonator impedance over the temperature range were made using a vector voltmeter system. In addition to the vector voltmeter system, a computing counter, with its programable arithmetic capability and having an absolute accuracy of 1×10^{9} for one-second periods, made it possible to quickly and easily determine the Q, amplitude frequency effect, and fractional frequency deviation of the resonators.

Experimental Results

Coupled Modes

When the influence of contour upon the reference crystal impedance had been explored, it became evident that not all contours would result in acceptable resonator designs. The criteria in the sonde application was that the reference has as high a Q as possible with a constant impedance over the temperature range. Plotting impedance-temperature curves for resonators having various contours indicated that numerous coupled modes become quite predominant on bi-convex resonators having contours steeper than about 4.25 meters^{-1}. In the case of the plano-convex blank, a region between 2.83 and 7.55 meters^{-1} was found which was relatively free of these modes.

Two types of coupled modes manifest themselves. The first is altered in amplitude and in the temperature at which it occurs by contouring. These modes cause large changes in the main mode impedance and their influence may last over temperature intervals as large as ten or fifteen degrees C and can cause perturbations in the frequency-temperature curve as large as 2×10^{-7} This is an undesirable condition in sonde since small frequency changes appear to be pressure fluctuations. The second type is one which can be altered in amplitude by contouring but not in the temperature at which it occurs. Fig. 3 shows the impedance-temperature plot of a resonator in which both types of modes are present with the second type of coupled mode occurring at about 30°C.

Of the designs evaluated, plano-convex resonators having a contour of 6.13 meters^{-1} have a very high Q and are relatively free from coupled modes except at about 30°C. To determine the effect of this mode on the frequency-temperature curve, frequency measurements were made at temperature intervals of < 0.1°C from 0 to +125°C. With no smoothing, the raw data was fitted to a third order polynomial. The RMS value of frequency deviations was in the order of 5×10^{-8} even in the region of the coupled mode.

Turning Point

A plot of turnover temperature for various contours on resonators having a constant polar orientation angle θ is shown in Fig. 4. It will be noted that in order to obtain equal turnover points, the curvature on a plano-convex blank must be a little greater than twice that on the bi-convex. Taking the average slope of these curves indicates that the dependence of turning point on contour is -2.2°C/meter^{-1} on plano blanks and twice this on bi-convex.

Conversely, by holding constant curvature and plotting turnover temperature as the orientation angle θ is changed, a plot such as shown in Fig. 5 is obtained. The dependence of turnover point on orientation angle is -1°C/minute at θ = 45 degrees and -2°C/minute at θ = 46 degrees.

Using the data from Figs. 4 and 5, one finds that in order to control the turnover point of a

plano-convex resonator to 65.0 ± 1°C requires a control of ± 0.3 meters^{-1} on contour and ± 1 minute on orientation angle. Also plotted in Fig. 5 is the distribution of turnover points of resonators having various contours for θ = 45 degrees. It is evident that in order to accurately predict the orientation angle and contour required to obtain a specified turnover point for each design, more data points would be required so that families of curves could be drawn covering the regions of interest. In order to be completely general, the diameter-to-thickness ratio should also be considered since it also influences the turnover point. See Fig. 6.

With regard to the design of reference crystals, which will be used to track the pressure transducer, it must be pointed out that not only must the turnover temperatures be matched but that the total frequency change away from the turnover points be the same. As it turns out, the second order temperature coefficients of plano-convex and bi-convex resonators are not equal even when the turning points are. In Fig. 7 the value of these coefficients for various contours is plotted. At a 65° turnover temperature, the plano-convex design has a second order coefficient which is about 4.5 percent larger than a bi-convex unit. This is unfortunate since the transducer utilizes a bi-convex resonator and a close match of frequency temperature characteristics could not be obtained. In this case, the mismatch is in the order of 36 Hz at 60° from turnover points. Therefore, it became necessary to use a bi-convex reference crystal even though it is more difficult to control orientation during the manufacturing process. Some loss of orientation angle almost always results when the second convex surface is generated. In order to match a bi-convex reference to the transducer, such that the frequency difference between the two is no more than ± 1 x 10^{-6} at a temperature ± 60°C away from the turnover point, the second order temperature coefficients of frequency must agree to within about ± 1 percent which is equivalent to a turnover point of < ± 0.2°C. Fabrication of a number of bi-convex resonators and transducers has shown that this matching condition can be met without too much difficulty.

Anharmonic Overtones

A plot of the main mode response and the first four anharmonic modes above 5 MHz for a contoured BT resonator is shown in Fig. 8. Making this type of plot for a number of resonators having various contours shows that the frequency difference between the main mode and the anharmonic modes increases as shown in Fig. 9. While this increase is only about 30 percent as large as what one would expect from an AT cut having similar dimensions, the modes are sufficiently attenuated that no difficulties with the resonator jumping modes has been experienced. With the exception of the second anharmonic which is attenuated only 10 db, the remaining anharmonic modes are attenuated by 20 to 30 db.

Resonator Parameters

A tabulation of electrical parameters for the various designs evaluated in this study is shown in Fig. 10. Although a number of designs exhibited high Q values, bi-convex resonators having a curvature of 6.30 meters^{-1} were the only ones which provided both high Q and the desired matching characteristics. For other applications, a number of these designs could prove to be useful. As a matter of interest, the plano-convex resonator having a curvature of 6.13 meters^{-1} was observed to have Q values approaching 4 x 10^6. The drive stability on one such unit was found to be in the order of 2 x 10^{-10}/db change in driving voltage amplitude and had a short term stability of ± 1 x 10^{-11} RMS.

Conclusions

While the BT cut resonator has temperature coefficients which are a function not only of orientation angle but also of contour, diameter-to-thickness ratio, and harmonic order, careful preparation of the resonator geometry and control of the crystallographic orientation angle has resulted in a reference crystal design which can be used to make the quartz resonator pressure transducer a useful device for field applications. A frequency match of ± 1 x 10^6 at a temperature 60°C away from the turnover point of the transducer has been accomplished with a bi-convex resonator and with improvements in fabrication techniques, matching in the order of ± 2 x 10^{-7} appears feasible.

Although plano-convex resonators could not be used as reference crystals for the pressure transducer device, it was found that one particular design offered very high values of Q, low sensitivity to drive level, and very good short term stability. These factors make this design attractive for frequency standard applications.

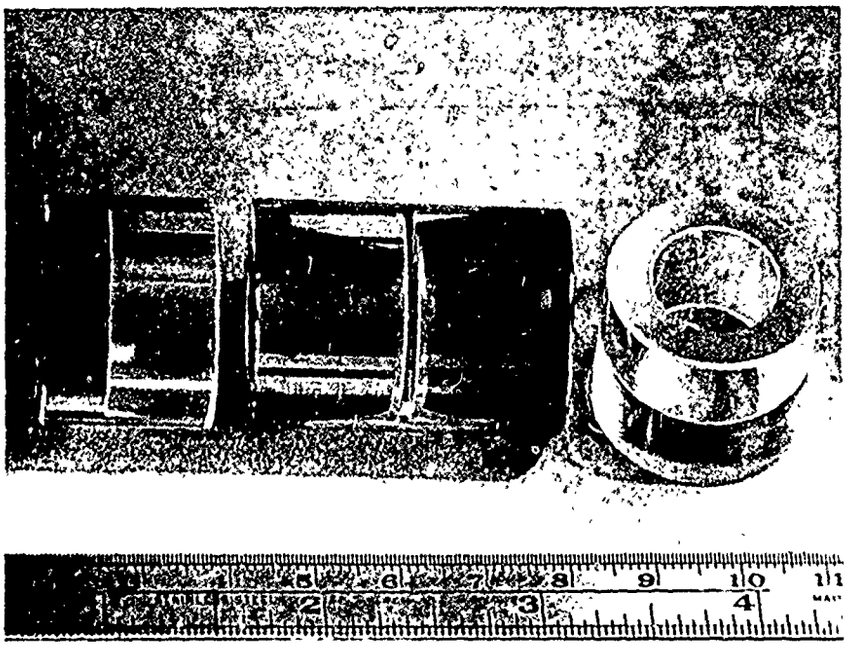
Acknowledgments

The author would like to acknowledge the valuable discussions and suggestions offered by Mr. Donald Hammond during the course of this study and the assistance of Mr. John Kusters in evaluating large amounts of data.

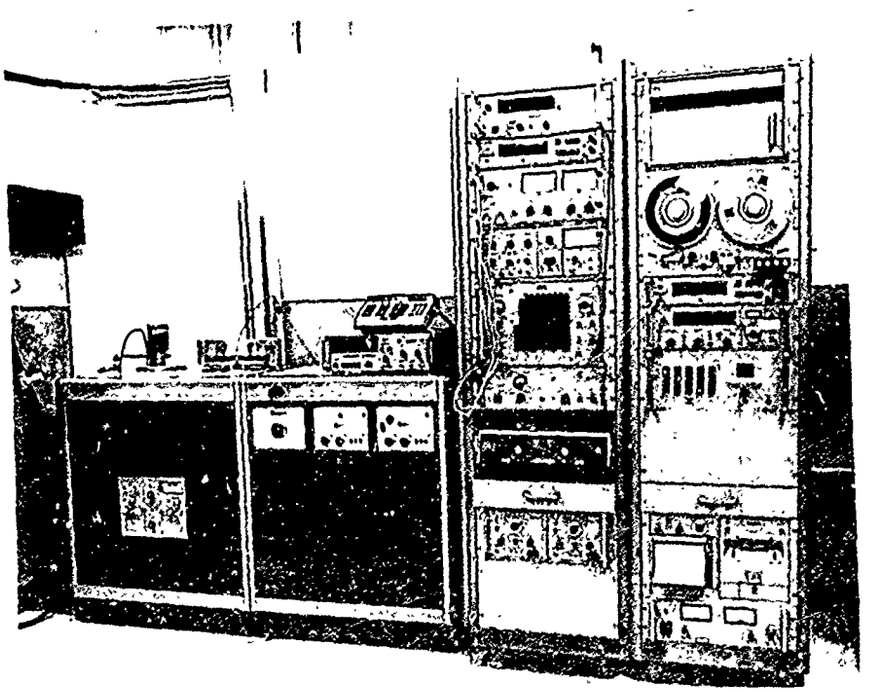
References

1. Seed, A., "Development of a High Q, BT-Cut Quartz Resonator," *Brit. J. Appl. Phys.*, 1965, Vol. 16, pp 1341-6
2. Karrer, H. E., J. Leach, "A Quartz Resonator Pressure Transducer," *IEEE Transactions*, Vol. IECI-16, Number 1, July 1969, pp 44-50
3. Adams, C., J. Kusters, A. Benjaminson,

"Measurement Techniques for Quartz Crystals,"
Proceedings of 22nd Annual Symposium on Frequency Control, April 1968, pp 248-51



Reproduced from
best available copy.



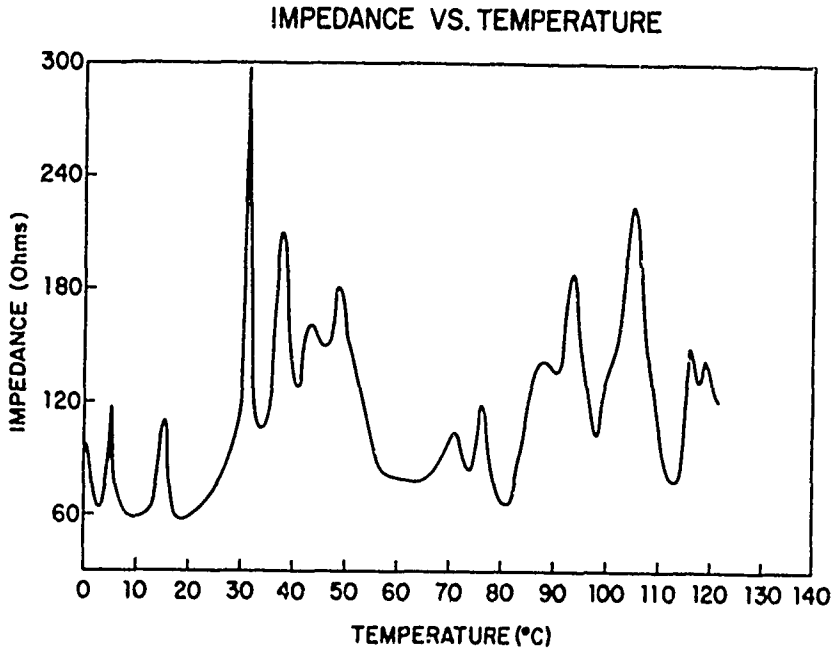


Fig. 3 Coupled Modes Present in a 5 MHz BT Cut Resonator

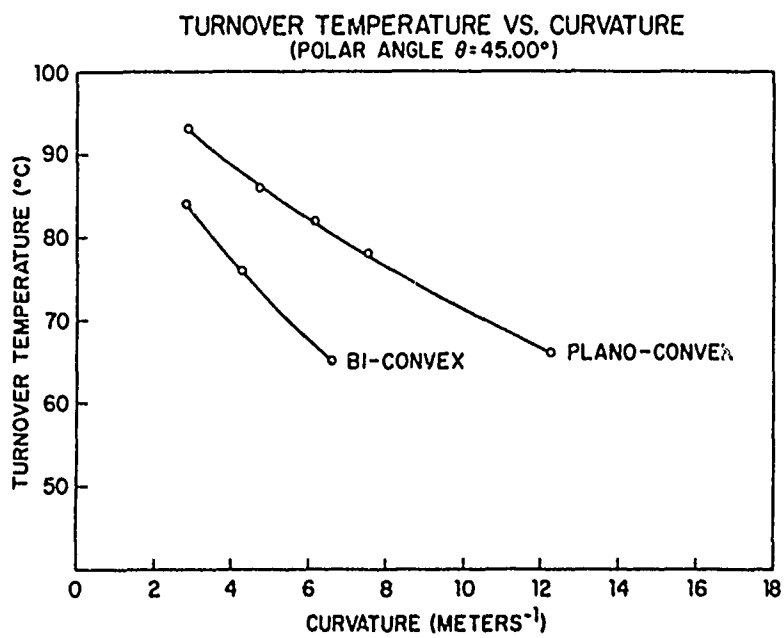


Fig. 4 Turnover Temperature vs. Curvature for BT Resonators

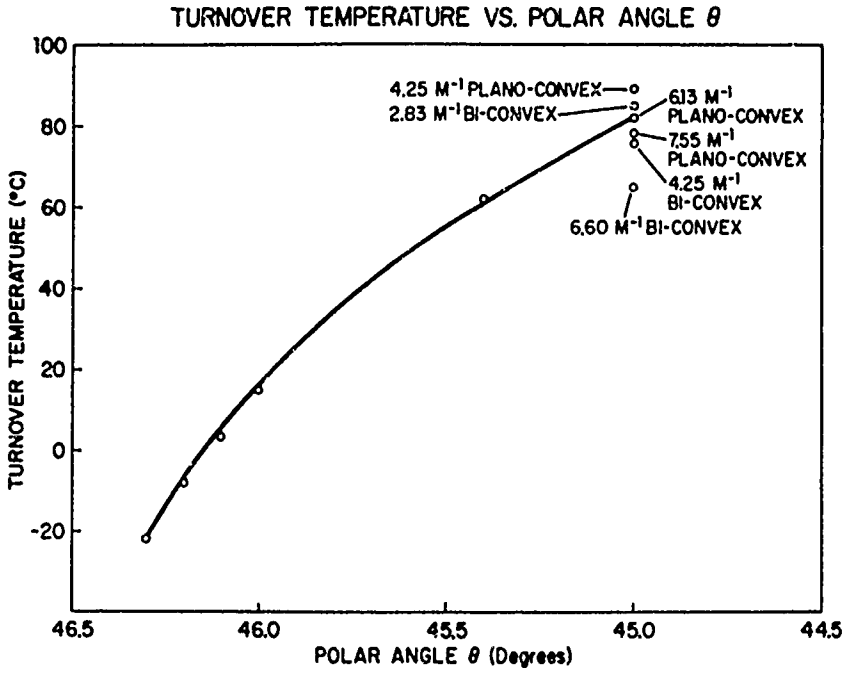


Fig. 5 Turnover Temperature vs. Orientation Angle θ

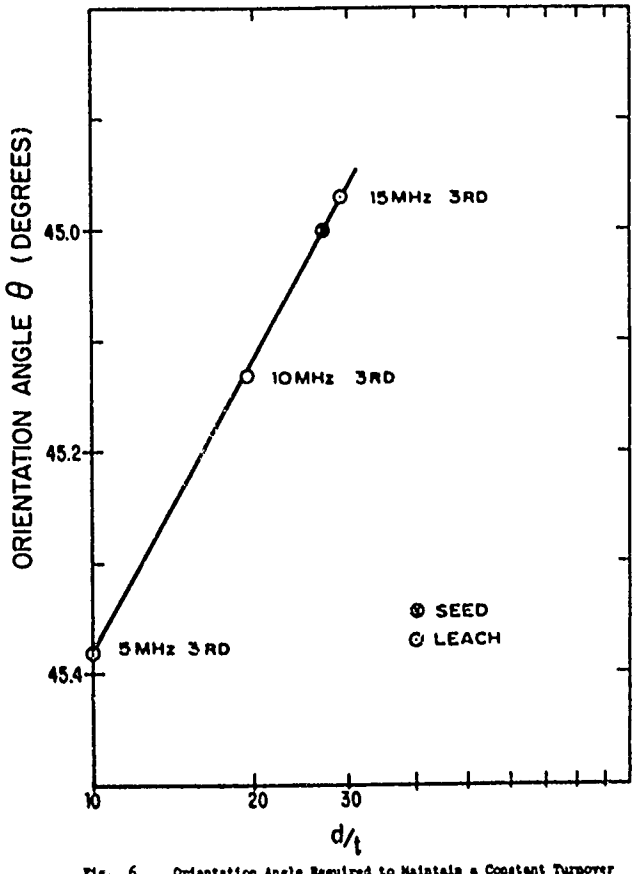


Fig. 6 Orientation Angle Required to Maintain a Constant Turnover Temperature vs. Diameter/Thickness Ratio

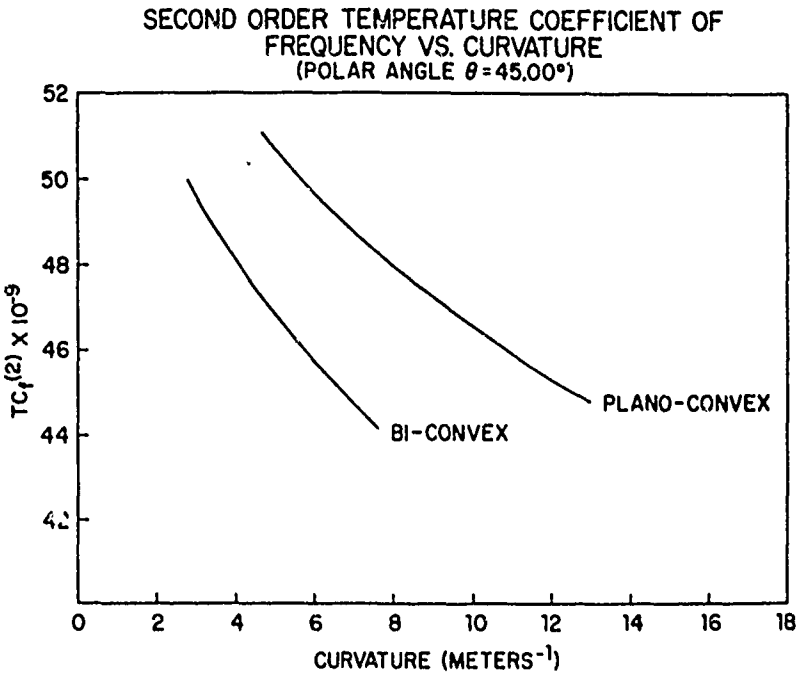


Fig. 7 Second Order Temperature Coefficients of Frequency vs. Contour

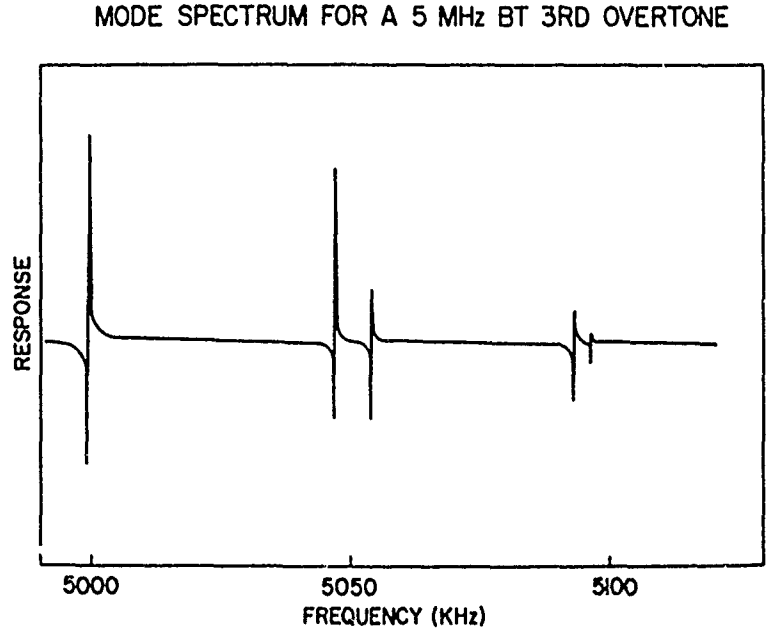


Fig. 8 Mode Spectrum for a 5 MHz Plano-Convex BT Cut

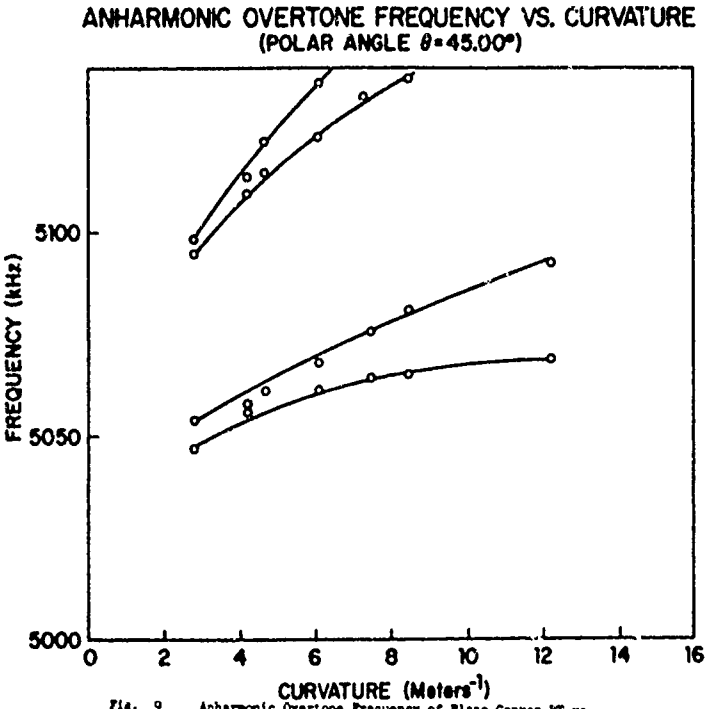


Fig. 9 Anharmonic Overtone Frequency of Plano-Convex BT vs. Curvature

ELECTRICAL PARAMETERS

5 MHz 3rd OVERTONE BT CUT RESONATORS

PLANO-CONVEX

CURVATURE (M^{-1})	Rs (OHMS)	Q×10⁶	INDUCTANCE (HENRIES)
2.83	235	0 32	1.20
4.25	60	1.60	3.00
6.13	36	3.20	3.80
8.49	73	1.90	4.40
12.26	155	0.76	5.10

BI-CONVEX

CURVATURE (M^{-1})	Rs (OHMS)	Q×10⁶	INDUCTANCE (HENRIES)
2.83	40	2 54	3.24
4.25	135	1.10	4.05
6.13	90	1.64	4.85
7.55	143	1.43	5.40

Fig. 10 Electrical Parameter of 5 MHz BT Resonators

MODE CONTROL AND RELATED STUDIES OF VHF QUARTZ FILTER CRYSTALS

Theodore J. Lukaszek
Electronic Components Laboratory, USAECOM
Fort Monmouth, New Jersey

Summary

Criteria for designing filter crystals with minimum disk diameter and maximum electrode diameter for the achievement of high quality, unwanted mode free AT-cut VHF filter crystals below 150 MHz has been reported previously. The technique used was to consider certain inherent boundaries on resonator design and performance arising from the intrinsic properties of quartz ($\mathcal{T}_1$) and on the diameter and thickness of the electrode (ϕ_e and t_e) in accordance with trapped-energy concepts. At frequencies above 150 MHz, however, such conventionally used crystal designs result in units possessing either an increasingly higher content of unwanted modes and/or resistance values normally greater than acceptable for most equipment applications.

The purpose of this paper, is to present new information which alleviates previously encountered restrictions and permits fabrication of filter crystals to frequencies as high as 200 MHz. This has been made possible on the basis of first a recently assembled crystal plating and monitoring system and secondly the introduction of a new processing method employing ion-etch techniques.

Introduction

Filter resonators operating in the upper VHF frequency range, are particularly applicable for increasing front-end selectivity of receivers in the parametric up-conversion scheme. In this approach gain is proportional to frequency so that the upper frequency is chosen as high as possible, compatible of course with crystal filter capabilities.

As part of an overall program aimed at meeting such increased demands, new design data have been developed for the fabrication of filter type quartz crystal resonators at frequencies up to 200 MHz. This represents a considerable increase over the previously attainable frequency of 150 MHz.[1] To achieve this goal, the investigations pursued were divided into two areas. In the first, the problem was approached on a semi-theoretical basis utilizing a new experimental technique, so that energy trapping requirements which relate proper frequency plateback to electrode dimensions could be precisely determined. The investigations were conducted for various frequencies within the 150 - 200 MHz range, using a unique plating-monitoring system recently assembled in the Electronic Components Laboratory. This new setup provides means for direct observation of a crystal's mode spectrum during and after the plating process and allows for direct measurement of the magnitude of each mode.

The straightforward use of conventional fabrication procedures at these VHF frequencies will not, in general, satisfy the required design specifications. The primary problem encountered, is that the electrical resistance of the required thin film electrodes and leads, necessary to achieve adequate mode control, becomes significantly higher as the electrode thickness is continually decreased in going to higher frequencies. It was necessary, therefore, that these filter crystal investigations be supplemented by a second area of work in which the electrical resistance of thin film electrodes and leads is taken into account and means for reducing it be developed. The approach introduced to minimize this film resistance problem while still allowing proper energy trapping conditions to be fulfilled is ion bombardment of a depression in the quartz wafer, such that the electrode may be partially embedded in the wafer. Earlier attempts chemically to etch selectively such a depression with hydrofluoric acid were unsuccessful because of preferential etching and virtually no depth control by this process. Recently advanced ion etching techniques have now been uniquely applied to quartz crystal development with the result that etch rates can easily be controlled, preferential etching and undercutting is not observed and in general the technique is capable of producing etch patterns with vertical walls to a definition limited only by the mask. Data on the interrelationships between essential design parameters, mode spectrographs, and related experimental evidence pertaining to this new approach are presented.

Resonator Design Criteria

The functional relationship between crystal resistance, electrode diameter and frequency for units operated between 150 and 200 MHz at the 5th and 7th harmonics of a fundamental mode is given in Fig. 1. As seen from this figure, for a fixed frequency, the crystal resistance depends primarily on the electrode diameter. As an example, to obtain a crystal resistance below 100 ohms at 210 MHz with a 7th harmonic unit, an electrode diameter of approximately 2.0 mm must be utilized. For applications where a 5th harmonic, 150 MHz

unit is required to have 40 ohms crystal resistance, an electrode diameter somewhat larger than 2.0 mm is required. The curves given in this figure were generated from calculations based on considerations concerning the motional time constant τ_1, a capacitance constant f_1 and a resistance constant P_1 according to the relations $P_1 = f_1/\tau_1$ and $R_1 = P_1 t/A_e$. Agreement of these calculated resistance values with experimental units has been good; however, it should be pointed out that use of subsequent experimental data could result in more refined values.

The second design consideration of importance concerns the unwanted mode content. To obtain a strong, clean, response in bound VHF AT-cut quartz resonators, it is necessary to lower the resonant frequency of some portion of the blank below that of the remainder so that this portion can act as an energy storage tank. This can be achieved if electrodes of finite thickness are applied to a limited portion of an AT-cut wafer, then separate cut-off frequencies will exist for the electroded and surrounding regions. Optimum conditions will occur when the electrode diameter and the plateback are coordinated by energy trapping considerations.[2]

For resonators below 150 MHz, criteria of this type, together with certain initial conditions on blank parallelism and surface finish, were sufficient to design units having a range of motional parameters while still maintaining suppressed unwanted modes by a trade off process between electrode dimensions and frequency lowering.[3,4] It has been found, however, during this investigation, that in trying simultaneously to satisfy these conditions at higher frequencies, very definite limitations do exist.

In the design of filter crystals in this upper frequency range with usable resistance values, the electrode diameter requirement is such that in satisfying the corresponding plateback requirement, the resulting aluminum electrode film thickness results in appreciable film resistance values. Consequently, one must observe and is bound to a minimum usable plateback figure since the metal film placed on the resonator must remain of sufficient thickness as not to cause excessive resistance in this film, which would add to the total crystal resistance. Experiments in this VHF range have shown that in order to reduce this film resistivity consideration to negligibility using aluminum, platebacks in the order of 0.2% are required.

In an attempt to overcome this problem and obtain data which will lead to the fabrication of suitable VHF resonators, investigations have been performed with a structure originally suggested by the Clevite Corp., under an ECOM-sponsored contract.[2] The technique is to deposit a sufficient thickness of aluminum without regard to spurious mode considerations. One thus approaches the bulk conductivity properties of the aluminum film and negligible film resistance is obtained. Because the mass loading of such a unit is excessive in terms of energy trapping, pronounced spurious modes appear but the cutoff frequency of the electrode and surrounding regions are then adjusted by depositing a nonconductive material layer on the surrounding area of the quartz structure. The effect of this layer is to diminish gradually the unwanted modes as proper conditions are finally approached. Fig. 2 schematically compares this configuration with one of a conventional design. In the conventional design (top illustration), the only means of adjusting the difference frequency Δf between the region marked f_e and f_s is by an appropriate change in the plating thickness t_e. As mentioned previously, for designs requiring minimal frequency depression, the film thickness at times is reduced below 300Å, resulting in films which no longer exhibit bulk properties with the consequence of poor film conductivity. In the lower drawing it is seen that t_e may be adjusted to any suitable thickness, normally in the order to 1000Å, and the proper cutoff frequency between the two regions may now be adjusted by an application of a dielectric film on the surrounding regions of the quartz wafer.

Crystal Plating & Monitoring System

Investigations of this type require experimentation with a large number of crystal units. In the past it had been necessary to remove the unit from the vacuum system to perform the required measurements after each step variation and, as a result, the collection of data was very slow. To permit investigations of this type to be conducted, not only in a more efficient manner, but with a precise determination of frequency lowering data, a system has been developed which permits direct observation of a crystal mode spectrum during and after the plating process without breaking the vacuum. The plating and monitoring system set up for this purpose in the Electronic Components Laboratory is shown in Figs. 3 and 4. Figure 3 is a closeup of the bell jar assembly that was constructed, showing a vacuum tight coaxial connector mounted in the wall of a stainless steel bell jar. One side of this connector will be attached to the unwanted mode monitoring instrumentation (Fig. 4) and the other, through an ell and rotary joint, to the masking arrangement shown in the lower right hand corner. Essentially, this is a G. R. component mount which was specially fitted to accomodate a variety of mask sizes and configurations. The crystal to be investigated is mounted in a conventional crystal holder, set in the component mount and the appropriate mask then aligned to this crystal. Figure 4 provides a photograph of the complete system. In this approach a half electrical wavelength (or integral multiple thereof) section of coaxial air transmission line is used between the plating apparatus and the unwanted mode monitoring system, thus allowing remote measurements to be performed quickly and in

a more efficient manner than was required by the old process of removing the unit from the vacuum system after each step variation of plateback to perform required electrical measurements. The new procedure is relatively straightforward; the coaxial feedthrough is connected to one arm of a hybrid bridge. The bridge is then excited from a frequency source that is swept by a function generator. The voltage at the bridge output is the response to this excitation which can then be monitored by an appropriate detector and recorded on an oscilloscope. The bridge is initially balanced for system capacitance at a frequency removed from the expected crystal resonance and then continually swept through the intended resonance region. At this point little or no output is observed on the scope. The crystal is then gradually plated until sufficient electrode material has been deposited such that as the signal source passes through the resonance region, the bridge balance is upset and the scope records the unbalance in the form of an initial resonance curve. The bridge is then rebalanced for the additional capacitance due to Co and plating gradually continued until the desired response is obtained on the oscilloscope. Having calibrated the grid lines on the oscilloscope for both frequency and amplitude in terms of resistance and dB, the frequency shift of the mode responses (Δ f) and magnitude of each, may be observed directly. A bridge arrangement of this type may also be used to evaluate the actual crystal parameters for both the main and unwanted modes in a very rapid fashion as discussed by Priebe and Ballato.[5]

Dielectric-Layered Resonator Considerations

As an example of the information than can be realized in the 150 - 200 MHz range by using this technique, Fig. 5 gives a seven-step sequence at various stages in the fabrication of a crystal at 150 MHz. The procedure utilized for this unit, began by selecting a 30 MHz fundamental quartz blank that had been processed to a parallelism greater than 1/5 wavelength of mercury light[3] (approximately 0.1 μm) and mounted in a suite holder. The unit was then plated in the manner just described with the sole intent of depositing a sufficient thickness of aluminum, to assure good crystal activity. The mode spectrum characteristics of the crystal that was visible at this point is shown in step (a) of this figure and is seen to be quite poor. The reason for this, is that in obtaining adequate film conductivity in the electrode, the resulting frequency depression was approximately 300 kHz. To fulfill energy trapping conditions at this frequency with the 2.4 mm electrode diameter that was employed requires a frequency plateback of only 19 kHz. The unit was then masked in such a way as to permit the deposition of a dielectric layer on the surrounding portions of the quartz wafer, to adjust the cutoff frequency to the conditions required for adequate mode control.

The improvements in the mode spectrum that can be observed as this layer is put on is shown in a step wise fashion in sequence (b) thru (g). Starting with step (b) we see that as a coating of zinc oxide, the dielectric used in this experiment, was initially applied, one first observes a slight decrease in the first unwanted mode when compared to its magnitude shown in sequence (a). Additional deposition of zinc oxide, beyond step (b), resulted in a more notable change as shown in steps (c) (d) (e) until finally, the condition shown in (f) was reached. At this point we have a complete suppression of unwanted modes on the high frequency side (right of the main mode), however, in the process of compensating for these unwanted modes some spurious mode activity is noted to have appeared on the low frequency side (left of the main mode). In fact in sequence (f) we see one fairly strong response. It was reasoned that this was due to over-compensation with the dielectric layer thus resulting in a case where a cavity was actually formed over the electroded region of the plate. Under these conditions, the electroded area is at a higher frequency then the surrounding regions, thereby giving rise to unwanted modes having frequencies lower than that of the main response. At this point, additional aluminum was evaporated into what is essentially a cavity formed by the dielectric layer and the spectrograph shown in step (g) resulted. Here we see a well controlled mode spectrum. It is pointed out that there was a slight decrease in main mode activity as we went thru steps (b) thru (f). However, upon deposition of the additional aluminum between steps (f) and (g) the main mode is seen to have improved and a crystal resistance of only 80 ohms was measured. For comparison, it had been previously reported that filter crystals fabricated using the conventional process at 150 MHz had resistance values in excess of 160 ohms.[1] The two-for-one resistance reduction is due to the combination of achieving negligible film resistance and to the practicability of employing somewhat larger electrode areas when the mode spectrum can be compensated for by some such technique as using a dielectric layer.

In another series of experiments utilizing techniques of this type, three optically parallel quartz blanks had ZnS layers of approximately 1000Å deposited on each side. The inverted masks used for these dielectric depositions were made so that three different cavity diameters (ϕ c) were formed. The cavity diameters chosen were ϕ c = 3.6, 3.0 and 2.4 mm. A 2.4 mm mask was then selected to be used in depositing 2.4 mm aluminum electrodes into these cavities. By using the same electrode diameter in each of the above-cited cavities, information was obtained on the effects of having various separations between the edge of the electrode and the edge of the cavity. In the first experiment, the crystal with the dielectric cavity hole of 3.6 mm and an electrode mask diameter of 2.4 mm provides for an annulus between electrode and

dielectric of 0.6 mm width. This unit was inserted into the monitor plating system and the modes observed while aluminum electrodes were gradually deposited in the setup as previously discussed. When sufficient material was deposited to just barely excite the plate, one was able to observe at this early stage the unwanted modes as well as the main mode. Continued plating gradually increased the magnitude of all modes until the condition shown on the left of Fig. 6 resulted. With this electrode cavity arrangement, at no time was there a condition in which the unwanted modes appeared suppressed to any satisfactory extent. Apparently the large separation between the electrode edge and ZnS dielectric edge is detrimental in the sense that the required Δf between the two regions was not achieved. Most likely, a third frequency region had been established between these two coated areas. The experiment was then repeated using the 3.0 mm dielectric cavity and again a 2.4 mm diameter electrode deposited within the cavity thus giving a 0.3 mm edge separation. The mode spectrum which resulted, center illustration, appears much improved over the previous combination and the crystal resistance was satisfactory. In the third and final experiment of this series a crystal was fabricated in which the separation between electrode edge and dielectric edge was completely eliminated. Both cavity size and electrode diameter were 2.4 mm. Alignment of the electrode within the cavity was very good and measurement of the crystal resistance under these conditions gave a value of only 70 ohms. The curve on the right contains the mode spectrum for this condition and appears far superior to cases 1 and 2 above. The main contributing factor is considered to be the elimination of the separation between the electrode to dielectric edge, thus providing attainment of proper frequency lowering conditions.

On this basis, several groups of crystals in the 150 - 200 MHz frequency range have been processed and direct observations of mode patterns, motional parameter values and a correlation between mode spectrographs and introduced discontinuities made. Figure 7 is given, summarizing the results of this investigation, i. e., data of plateback vs. electrode diameter conditions which was found to provide optimum suppression of unwanted modes for frequencies in the 150 - 200 MHz range.

Ion-Etch Approach

In view of the results obtained in the manner just discussed, it is apparent that mode control and resistance requirements can be met simultaneously even at upper VHF frequencies, providing proper conditions are met. A search for a method which would essentially provide a similar structure without, however, the deposition of dielectric layers, has led to the utilization of ion-etch techniques to produce an inverted mesa structure. Figure 8 provides a schematic of the structure. With this approach an electrode may be partially embedded in the wafer rather than the previous requirement of building up the surrounding area with a dielectric. Doing away with this layer, has the effect of eliminating the instabilities which may occur with respect to its aging and frequency-temperature behavior. Several earlier attempts to chemically etch a selective depression of this type with hydrofluoride acid were only partially successful, because the inverted mesa region had a poorly defined slope rather than the sharp edges required.[2] What is being introduced at this time, is the application to crystal resonator development the recently advanced ion etching, or ion bombardment technology to achieve this configuration. The phenomenon involved in this erosion technique is the disintegration of a solid surface by striking it with high energy ions. Here an inert gas such as argon is used as an ionizable medium to transfer potential energy built up by an electric field into kinetic energy of the ion that, in turn, knocks out neutral target atoms or molecules. The basic mechanism of ejected surface material being momentum transfer. A good deal of the necessary technology for the realization of ion erosion is currently being published in the literature in allied disciplines, though in a somewhat different form.[6,7] Since only a relatively modest effort is required to bend this technology toward the realization of the device to be discussed here, the ion-etch applications and results not operational procedures will be presented.

To obtain an appreciation for the extensive capabilities provided with the utilization of this approach, several figures relating information on a number of basic studies that have been completed, are first presented. Starting with Fig. 9, an interferogram is shown comparing the interference fringe pattern of a quartz plate surface with and without ion erosion. The portion labeled exposed is the area that was ion-etched and the non-exposed area was protected by a mask. The interpretation of this fringe pattern is that the bands across the crystal plate are bent according to the degree of thickness variation across the unit. Each band is equal to one-half the wavelength of the green mecury light used in this experiment 5462/2 Å, thus, it is seen that the amount of material removed due to ion erosion is a little more than two half wavelengths or approximately 6000A. It is pointed out that the definition between the exposed and not exposed portions is quite poor because in this first preliminary experiment the masking technique used was just a crude metallic mask. The primary information that was sought in this trial was the determination as to the uniformity with which surface removal could be obtained and whether the heating effects introduced, would cause significant thermal stresses within the crystal leading to possible fractures. From an examination of this photograph it appears that the removal of material by ion-etch

techniques even to the extent of 6000Å, much more than will be normally required for resonator fabrication, results in a surface that should be quite suitable for filter crystal fabrication. This is affirmed by the observation that the interference fringes in the exposed portion are not deviated in shy peculiar manner, as would be the case if the surface had an irregular finish, if anything the fringes appear to show improved parallelism. An additional investigation on uniformity of surface removal is shown in Fig. 10. Here we have an interferogram of two quartz wafers, a portion of each which again has been exposed to ion bombardment only this time, the unit on the right has a convex surface. In this trial, approximately one-half wavelength (2730Å) was removed from each surface. One should note the markedly improved edge definition not only for the unit with the flat surface but a near perfect boundary division for the unit having a convex curvature as well. In these examples, a glass mask had been used to obtain the required division for the flat surface, and photo resist masking techniques were employed for the convex unit. An extremely interesting conclusion drawn from this experiment, in addition to the edge improvements, is the fact that even though the unit on the right has a convex surface, the technique of surface removal by ion-etching does not show any apparent preferential erosion. The interference fringes are as evenly spaced in the exposed surface as the portion that was not exposed.

In the third experiment of this type, a quartz crystal surface was examined before and after ion etch, by scanning the surface with an electron microscope. Figure 11 provides this comparison, in which the photo on the left represents the surface before ion etching and the photo on the right the same crystal, after a 5000Å layer was removed. As can be seen, the crystal on the left, taken out of a group of crystals processed using conventional lapping and polishing techniques possessed an area with an appreciable number of scratches. This unit was then exposed to ion bombardment, removing a layer of approximately 5000Å. The complete surface was re-examined under the electron microscope and was found to be free of all scratch marks previously observed. With the exception of a small surface defect indicating a build up of some sort shown in the photo on the right, the surface was completely smooth. The fact that a surface finish may be improved by a technique of this type may have far reaching implications not only for filter crystal applications but for crystal resonators in general. It appears that this ion etch technique may lend itself to final finishing of quartz crystal surfaces.

Encouraged by the physical improvements observed by optical means during the three preceding experiments, two high quality 5 MHz plano convex resonators were fabricated to obtain information as to the effects ion etched surfaces would have on the electrical properties of such devices. The two units selected were processed in an identical manner, the only exception being one of the two units was etched by ion-bombardment techniques removing a half wavelength Hg layer from each surface. Figure 12 is presented, providing a comparison of the essential parameters. In comparing each individual parameter, it is apparent that the process of ion etching does not result in a degradation of crystal performance. To the contrary, one might conclude a slight improvement had been obtained. The fact that satisfactory parameter values are achievable in such high quality resonators, probably reflects the contention that no sub-surface damage to the crystal structure, such as twining etc., is introduced during the erosion process.

Ion Etch Results

The application of this erosion technique to filter crystal development has enabled the usable frequency range to be extended to frequencies as high as 210 MHz. Examples of this achievement have been chosen and will now be discussed as representative of the possibilities now attainable. Figure 13 shows a magnified photograph and interferogram of a 210 MHz, 7th harmonic, quartz blank into which a keyhole pattern of $\phi e = 2.0$ mm and a tab width of 1 mm was eroded. The notable achievements obtained here are the very sharply defined keyhole edge, uniformity in surface removal within the keyhole cavity and the scratch free surface finish. The well defined boundary visable was accomplished using photo-resist techniques and the erosion, as mentioned previously, by ion etching. The depth of erosion for this particular case is seen to be approximately $\lambda/4$ Hg. The unit was then mounted in a conventional HC-18/U holder and inserted into the crystal plating and monitoring facility for electrode evaporation into this cavity.

A three step sequence at various stages in the plating process of this unit is provided in Fig. 14. Below each response curve, is illustrated schematically, cross sectional diagrams of the various plating conditions within the cavity that give rise to the respective mode spectrum. In the first sequence, designated (a), one observes a strong main mode response together with several weaker unwanted responses, appearing on the low frequency side of the primary mode. The configuration illustrated below this spectrum show a condition such that the plated region, f_e, is higher in frequency than the surrounding region f_s. If the unwanted modes are though of as being associated with the plate as a whole, this would account for the fact that these modes occur at a frequency somewhat below the main response.

Evaporation of additional aluminum material into the ion-etched cavity gradually diminished the low frequency unwanted responses and finally produced the desirable condition illustrated in

(b). At this point, the frequency difference between the plated and unplated portions of the resonator, Δf, are such that proper energy trapping conditions have apparently been achieved. Measurement of the crystals resistance parameter under these conditions resulted in a value of less than 160 ohms. If the frequency of the electroded region is decreased beyond this point, by a further increase in electrode thickness, energy trapping requirements are upset and the mode spectrum shown in (c) of Fig. 14 results. Since a spectrograph of the type shown in sequence (c) is quite typical of crystal units fabricated at this frequency by conventional techniques, an excellent comparison is provided showing the significant improvement that is achieved by incorporating this proposed method of fabrication.

The mode spectra responses of a resonator having an even larger area electroded structure, in which part of the electrode thickness was embedded into the wafer thereby reducing electrical resistance without contributing to overall mass loading is shown in Fig. 15. In this case we have a 150 MHz, 5th harmonic unit into which an ion etched keyhole pattern having a cavity diameter of 2.4 mm, a tab width of 1 mm and an erosion depth of approximately 1500Å on each side of the plate was prepared. Upon completion of electrode mask to cavity alignment to assure exact registration, the units performance was then monitored as plating commenced, by the procedure outlined previously. Illustration (a) of this figure shows the crystal response in the initial stage of excitation, i. e., the result as soon as a minimum amount of plating required to begin electrical conduction has been achieved. The bridge is then rebalanced for the additional capacitance due to C_o and the plating gradually continued until the desired response is observed on the oscilloscope. This condition is shown in sequence (b). The crystal resistance at this point measured slightly over 100 ohms again a considerable reduction compared to the best previously reported value at this frequency, 150 ohms.[1] As the unit was plated with additional aluminum beyond this point, an interesting observation made was that unwanted modes gradually appeared on the high frequency side, before the undesired modes on the lower frequency side were completely diminished. This means that a degree of compromise must be made between the mode structure on either side of the desired response. By the time the lower frequency unwanted mode structure is completely eliminated, several high frequency spurious mode are already visible as shown in sequence (c). Intentional frequency lowering beyond this point, to the extent that energy trapping requirements were unbalanced, again resulted in the introduction of a family of high frequency unwanted modes as shown in sequence (d) and (e).

On the basis of the work completed during this investigation it appears that the combination of photo resist masking and ion etching techniques should be capable in producting controlled accuracy down to several microns. In general, the attainable definition is limited primarily to ones masking abilities. An example of a notable achievement in this area is shown in Fig. 16. Here we see a quartz blank into which a grid network pattern was ion etched. The grid spacings are 0.75 mm (0.0295"), the diameter of the dots within the grids 0.43 mm (0.0169") and the ion etched grid slots only 0.107 mm (0.0042") in width. Approximately 1/3 λ_{Hg} of quartz material had been removed (1810Å), and even with this closely spaced pattern, again we see very sharp edge definition and uniform material removal within the cavities.

Conclusion

This paper has presented a new design aspect for crystal resonator fabrication. Proper mode levels and resonator impedance control for low insertion loss filter application can be achieved at upper VHF frequencies if the techniques outlined are utilized. Examples verifying the technique together with optimum plate-back design data for controlling the mode spectrum level were provided.

In general, the techniques of photo resist masking and ion etching offers an economical and practical means of performing difficult erosion tasks. Only one particular application was presented in this paper i. e., the selective removal of material in the fabrication of quartz filter crystals, but techniques of this type should also be applicable to a variety of other devices, since it has been shown that selective uniform removal of material can be precisely controlled with good definition. An additional application that comes to mind is the task of optical polishing, i. e. complete surface removal to a prescribed depth.[8] A technique of this type can assure absolute cleanliness of a quartz wafer immediately prior to film deposition thus having application to resonator technology in general. Other possibilities might be the selective etching of channels or grooves in piezoelectric materials for guiding acoustic surface waves and possibly mode control in acoustically coupled resonators.

Acknowledgments

I would like to acknowledge the assistance and advice received from Mr. Mike Tarasevich formally with Kollsman Instrument Corporation, the many stimulating discussions held with Dr. Guttwein and Mr. Arthur Ballato, and the enthusiastic support of Mr. Stanley Emmons in the course of these investigations.

References

1. G. K. Guttwein, T. J. Lukaszek and A. D. Ballato: "Practical Consequences of Modal Parameter Control in Crystal Resonators," Proc. 21st Annual Symposium on Frequency Control, pp. 115-137, April 1967.

2. W. Shockley, D. Curran and D. Koneval: "Energy Trapping and Related Studies of Multiple Electrode Filter Crystals," Proc. 17th Annual Symposium on Frequency Control, AD-423381, pp. 81-126, May 1963.

   D. R. Curran and D. J. Koneval: "Research Into New Approaches for VHF Filter Crystals," Final Report, Contract No. 36-039 AMC-02245 (E), USAECOM, Fort Monmouth, New Jersey, June 1967, 211 p.

3. A. Ballato, T. Lukaszek, H. Wasshausen and E. Chabak: "Design and Fabrication of Modern Filter Crystals," Proc. 20th Annual Symposium on Frequency Control, AD-800523, pp. 136-160, April 1966.

   T. J. Lukaszek, A. D. Ballato and H. P. Wasshausen: "Additional Considerations for VHF Filter Crystal Design," Proc. IEEE, Vol. 54, #10, October 1966, pp. 1466-1467.

4. G. K. Guttwein, T. J. Lukaszek and A. D. Ballato: "Design Considerations for Crystal Resonators with Prescribed Mode Suppression," ECOM Technical Report 3018, September 1968.

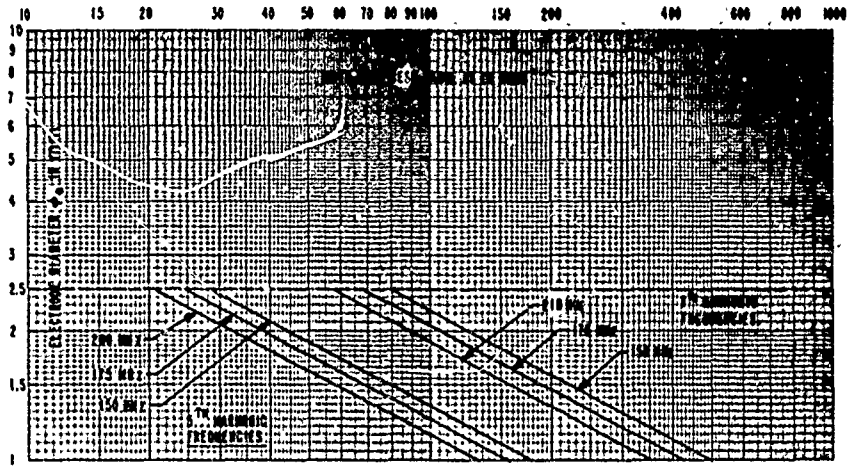
5. F. K. Priebe and A. D. Ballato: "Measurement of Mode Parameters by Sweep Frequency Methods in the Frequency Range From 20 to 250 MHz," Proc. 20th Annual Symposium on Frequency Control, AD-800523, pp. 465-499, April 1966.

6. M. Tarasvich, "Experimental Verification of Ion Beam Process," Interim Technical Report, September 69, Contr. No. F33615-69-C-1744, U. S. Air Force, Wright-Patterson AFB, Ohio.

7. J. L. Vossen and J. J. O'Neill, Jr.: "RF Sputtering Processes," RCA Review, Vol. 29, p. 149, June 1968.

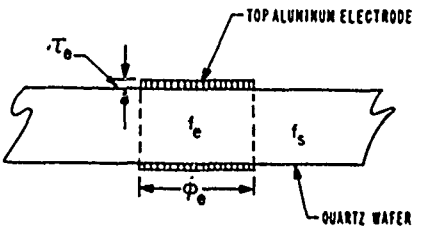
8. M. Tarasevich: "Ion Beam Erosion of Rough Glass Surfaces," Applied Optics, Vol. 9, p. 173, Jan. 1970.

9. M. Tarasvich, Kollsman Instrument Corp., Private Correspondence.

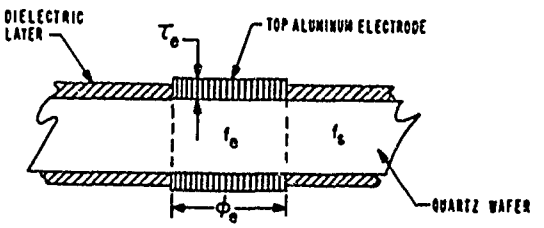


RELATION BETWEEN ELECTRODE DIAMETER, CRYSTAL RESISTANCE AND FREQUENCY FOR 5^{TH} AND 7^{TH} HARMONIC UNITS

Figure 1

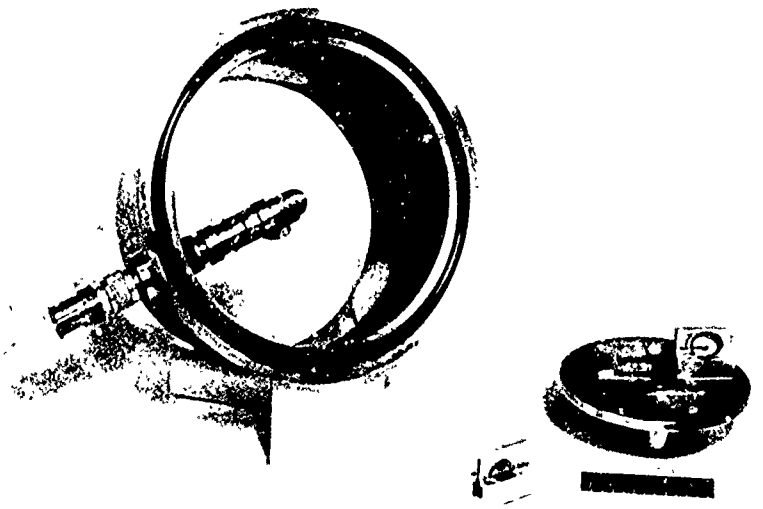


CONVENTIONAL RESONATOR



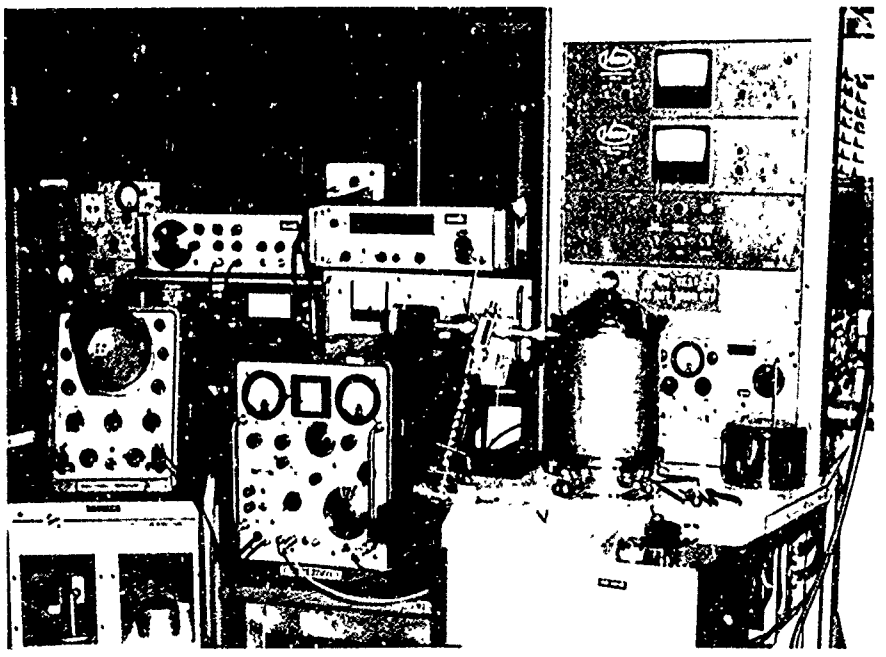
DIELECTRIC LAYERED RESONATOR

Figure 2



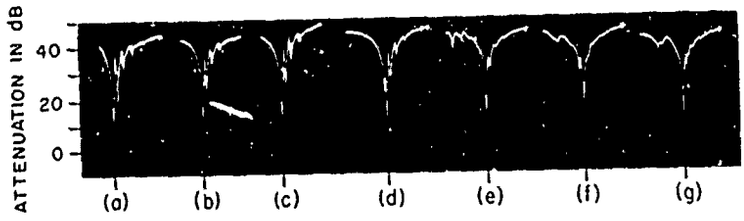
BELL JAR ASSEMBLY: COAXIAL VACUUM FEED-THROUGH AND PLATING MASK

Figure 3



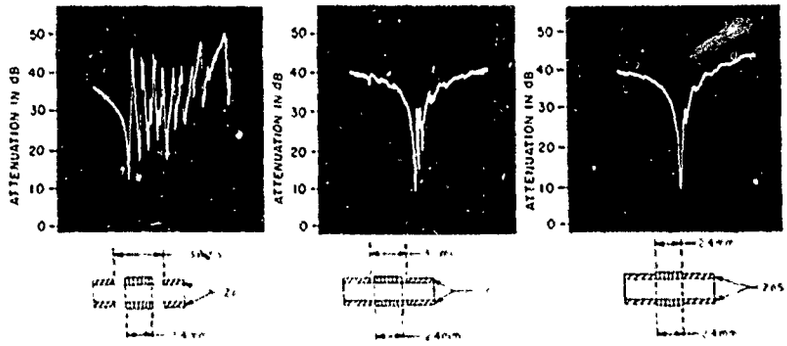
FILTER CRYSTAL PLATING & MONITORING SYSTEM

Figure 4



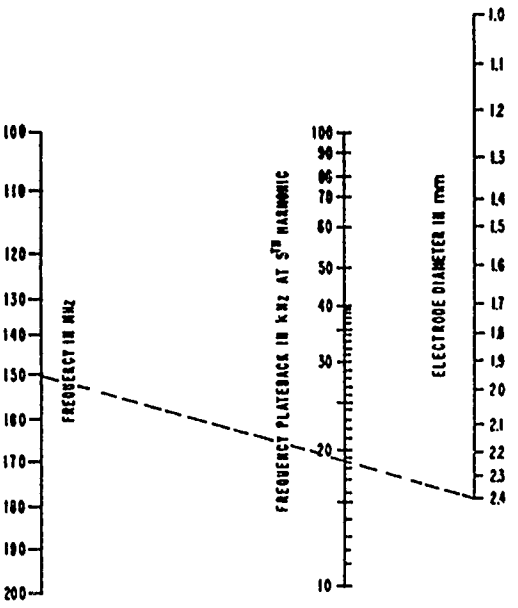
SEQUENCE OF MODE SPECTRA IMPROVEMENT EMPLOYING DIELECTRIC COATING TECHNIQUES

Figure 5



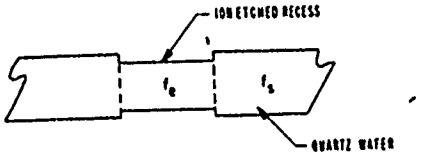
MODE SPECTROGRAPH EFFECTS OF SEPARATION BETWEEN ELECTRODE TO DIELECTRIC EDGE

Figure 6

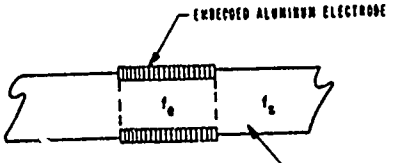


FILTER CRYSTAL DESIGN CRITERIA

Figure 7



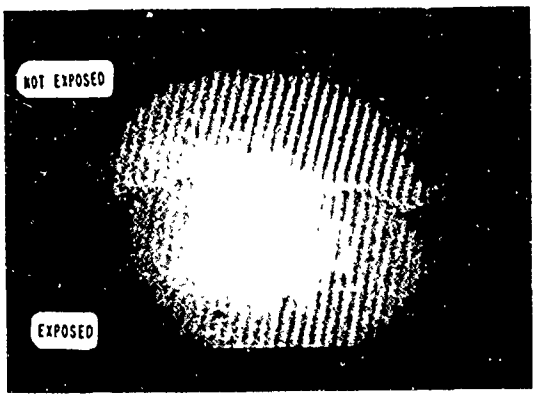
BEFORE PLATING



AFTER PLATING

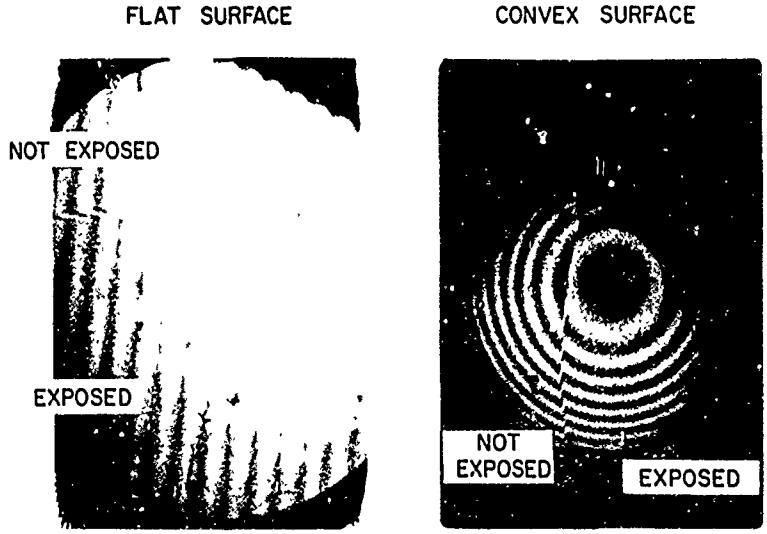
PROPOSED ION ETCHED STRUCTURE

Figure 8



INTERFERENCE FRINGE PATTERN
OF A QUARTZ PLATE COMPARING
THE SURFACE BEFORE
AND AFTER ION EROSION

Figure 9



INTERFERENCE FRINGE PATTERNS OF TWO QUARTZ WAFERS COMPARING
THICKNESS DEVIATIONS BEFORE AND AFTER ION EXPOSURE

Figure 10

BEFORE ION EXPOSURE

AFTER ION ETCH

SCALE 0 40μm

SCALE 0 70μm

COMPARISON OF A QUARTZ CRYSTAL SURFACE BEFORE AND AFTER REMOVAL OF A 5000 Å LAYER BY ION ETCH TECHNIQUE

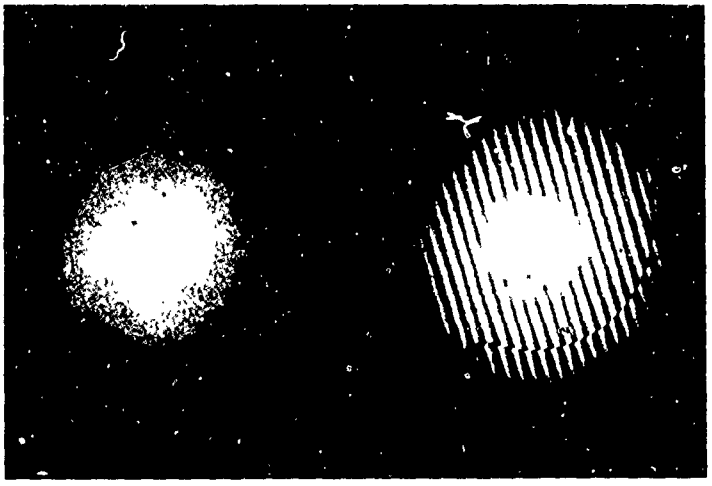
Figure 11

PLANO-CONVEX CRYSTAL

PARAMETER	CONVENTIONAL UNIT	ION ETCHED UNIT
FREQUENCY	5.0 MHz	5.02 MHz
ELECTRODE DIAMETER	5.0 mm	5.0 mm
MOTIONAL RESISTANCE	2.2 OHMS	1.95 OHMS
MOTIONAL CAPACITANCE	11.0 fF	11.2 fF
STATIC CAPACITANCE	5.2 pF	4.95 pF
CAPACITANCE RATIO	472	442
TIME CONSTANT	24.2 f SEC	21.9 f SEC
QUALITY FACTOR	1,316,000	1,454,200

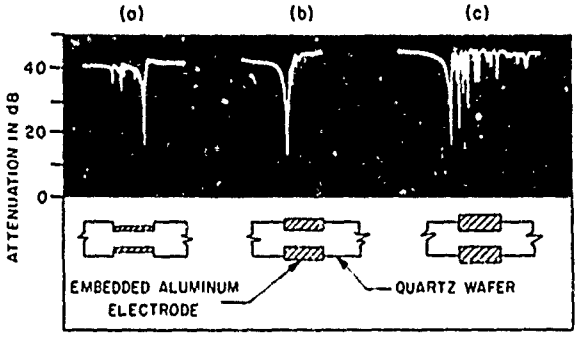
COMPARISON OF RESONATOR'S MOTIONAL PARAMETERS WITH AND WITHOUT ION ETCH

Figure 12



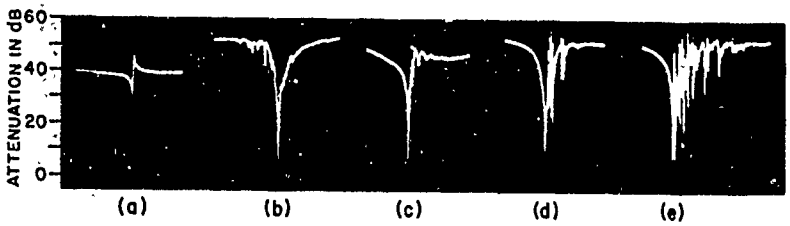
KEYHOLE PATTERN ERODED INTO A QUARTZ PLATE
BY ION ETCH TECHNIQUES

Figure 13



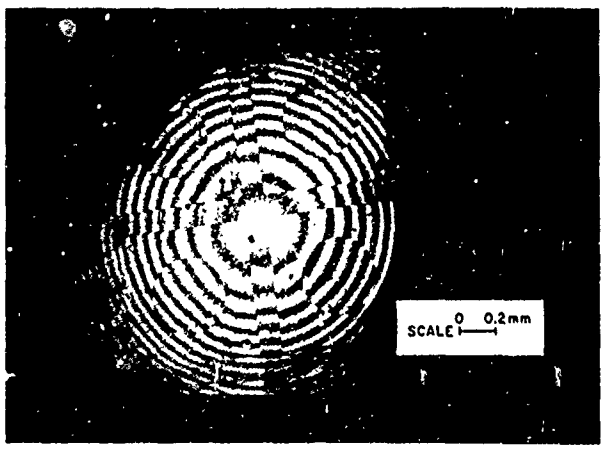
MODE SPECTRA VS ELECTRODE THICKNESS 210 MHz ION ETCHED CRYSTAL

Figure 14



MODE SPECTRA SEQUENCE OF AN ION ETCHED 150 MHz CRYSTAL
AS A FUNCTION OF ELECTRODE THICKNESS

Figure 15



INTERFEROGRAM OF TEST PATTERNS ION ETCHED INTO QUARTZ
DEMONSTRATING DEFINITIONS ACHIEVABLE WITH THIS TECHNIQUE

Figure 16

LOW AGING CRYSTAL UNITS FOR USE IN TEMPERATURE COMPENSATED OSCILLATORS

L. A. Dick and J. F. Silver
CTS Knights, Inc.
Sandwich, Illinois

Introduction

Claude Shannon of the Bell Telephone Laboratories first published his classical treatise on information theory some twenty years ago. In this interval of time the demands for interference free communications channels has multiplied many fold. The communications industry has tried to meet these increased demands by extending the useful range of frequencies available for channel assignment and by changing the mode of modulation to decrease the band of frequencies required per channel and/or to improve the effective signal to noise performance, and increase the effective range per channel.

The rapid advance of the communication equipment suppliers has imposed continuous demands upon the quartz crystal industry to provide resonators having the prerequisite stability, particularly over the broad and severe environmental conditions associated with military communications.

In a paper presented at the 1969 Annual Symposium on Frequency Control, Owen P. Layden outlined the requirements for "secured voice" communications for tactical military use. [1] If successful, Mr. Shannon's "bits" of information theory would be utilized in a most effective way. But this could only be made possible by meeting very stringent stability requirements for both transmitter and receivers in the network. For example, at 30 MHz the transmitter and receiver can have a difference frequency of no more than 5 Hz which allows 2.5 Hz for each equipment. In addition, the system must operate for 120 weeks without calibration. Figure 1, borrowed from Mr. Layden's paper, recalls these specifications. It shows the aging per week vs. the frequency-temperature stability required. It can be readily seen that there is a trade-off between aging and frequency stability. Current state of the art aging for a 30 MHz system using a 5 MHz fundamental crystal unit is 5×10^{-10}/week. From Figure 1, this would require a frequency stability over the temperature ambient of 2×10^{-8}. Actually, aging is approximately an exponential function of time, hence a crystal which reached an aging rate of 5×10^{-10}/week after two month operation would allow the system engineer to design to 5×10^{-8} ambient stability.

Performance

Under Contract DA-28-043-AMC-02183, with the U. S. Army Electronics Command, the CTS Knights Company undertook the task of developing quartz crystal units to meet these requirements. After a couple of less successful attempts a group of crystal units was submitted that were 5 MHz fundamental mode housed in coldweld equivalents of HC-6/U type holders. These units met all of the parameter specifications, and typical aging behavior is exhibited in Figure 2. The aging data shown here was supplied by the USAECOM. We see that for the best units the aging per week is less than 5×10^{-10}. These units had been aged a minimum of two weeks prior to shipping.

The aging just discussed was obtained with the crystal resonators in what might be called non-operating modes. This is to say that the units were operating during the test periods only long enough to take readings. To further investigate the behavior of a crystal unit under more or less continuous operation, a precision ovenized oscillator was designed so that the crystal current could be changed or the oscillator turned off without interrupting the oven operation. We believe it is of special interest to study the retrace characteristics of a resonator that result from turning off the excitation and maintaining the resonator and circuitry at constant temperature. The assumption here is that any transferable mass due to residual contamination will settle on the quiescent crystal with its relatively high "gettering" properties (when the oscillator is turned off) only to be driven off by the combination of a slight increase in temperature and the motion of the resonator when the oscillator is reactivated. The frequency change resulting is one of the more positive indicators of transferable mass residue retained in the enclosure.

Figures 3 and 4 show the behavior of one of these crystals in the special oscillator previously described. The important points to be considered in Figure 3 are the

resonator's behavior as a result of change in crystal current and of oscillator power off, and of oscillator and oven power off. First, we see that the data starts 24 days after turn on and with the crystal operating at 6 ma current. Then there was a one day period of power off followed by a return to operation at a lower crystal current of 0.7 ma. This change in drive level of 37 db produced an off set in frequency of -1.75×10^{-6} which is approximately 5×10^{-9}/db. The data was normalized to show the slight, if any, change in aging rate under the two drive conditions. Second, on the 54th day, the oscillator power was turned off for 24 hours. When returning to the power on condition, the aging was again unaffected. This time, since there was no drive change, there was no need to normalize the data. The retrace, if any, was less than 1×10^{-9}. The insert on the upper left hand of the chart shows the performance on a minute by minute basis. The frequency excursion occurring during the first few minutes of operation is small and might well be the result of thermal stress which is produced by the return to oscillation. However, since the frequency returned in twenty minutes to within 2×10^{-10}, it is assumed that the retained contamination would not prevent predictable retrace behavior of less than 1×10^{-8}. This behavior is indicative of a quality resonator free from reversible aging components. Third, after 58 days of evaluation, the power to both the oscillator and oven were turned off for a period of 4 days. We see the retrace is 4×10^{-9} and the aging rate recovered within 24 hours. Figure 4 is the hour by hour retrace just discussed and shows a recovery to 8×10^{-9} of the turn off frequency within 2 hours. After 26 hours, we are within 4×10^{-9} of the turn off frequency.

Summing up Figures 3 and 4, of extended operation, we have a positive aging occurring over a time period of 2 months after which we reach an aging behavior of approximately 1×10^{-10}/day and retrace behavior of less than 5×10^{-9}. We believe the relatively long stabilization period shown here is due to circuit components other than the crystal, because previously this crystal unit had an initial aging of approximately 5×10^{-10}/week under intermittent operating conditions.

Having a resonator capable of the performance shown on the previous figures challenges the oscillator designer to make optimum use of the device by providing a compact temperature-compensated oscillator which will meet the service requirements for "secured" single side band voice communications as previously set forth.

Figure 5 is a curve of frequency vs. temperature for such a compensated oscillator unit designed by Dr. D. E. Newell. The temperature in the thermal chamber was varied by 10°C steps and allowed to stabilize for 30 minutes between readings. The solid line was obtained by moving up in temperature and the dotted line by moving down in temperature. The corresponding data-points on both curves were made at the same temperature within 0.5°C. We see that the frequency stability of $\pm 5 \times 10^{-8}$ (or $\pm$ 2.5 Hz) is met over the range of -40 to +80°C as required. It should be stated that Dr. Newell used the segmented compensation approach in the above design.

Crystal Design

The design and processes to produce these crystal units was reported on in detail in the Final Report on Contract DA-28-043-AMC-02183 for the USAECOM. In general this work was done under the hypothesis that crystal resonators could be rendered stable with regard to the three major design and fabrication areas outlined as follows:

1. Crystal units were purged of mass sorption contaminants by high temperature vacuum baking (250°C) in an ultra clean system.

2. No major reversible temperature induced stress was generated in the fabrication which could not be relieved to a linear retraceable phenomena prior to being subjected to actual field use.

3. That the purging and stress relieving techniques used must not in any way deteriorate the viability of the materials involved.

To accomodate the relatively high edge motion of the fundamental mode resonator, viable conducting cements capable of processing temperatures approaching 300°C were used, and bake-out temperatures and operation sequences were critically selected. To collect the expurgated contaminants, the base mounted resonators were pumped down and baked out in glass enclosures and then removed from these enclosures just prior to introduction to the high temperature, high vacuum sealing machine. Completed sealed units were cycled over temperature extremes while being intermittently activated by a sweep oscillator to strain relieve the assemblies.

Figure 6 is a picture of the crystal units in the interim stage after being enclosed in glass and baked out under high vacuum for several hours.

Figure 7 is a picture of the special high temperature high vacuum machine designed to vacuum bake and coldweld the units under ultra clean conditions The crystal units travel through the device on a chain conveyor and the cans and units are manipulated in and out of the coldweld die by means of the bellows-sealed high vacuum manipulator shown protruding from the machine.

Conclusion

It would seem important to consider the matter of expected life or very long stability. It must be recognized that the cement is the one organic material remaining in the enclosure. Since it is reasonable to assume that the consistent performance of the crystal unit will depend upon the viability of the cement, additional tests should embrace continuously working as well as non working conditions at various temperatures. Microscopic and chemical studies of the physical and or chemical changes in the cement and its interface relations would be desirable during the investigation.

Although the data presented is admittedly limited, the possibility of providing the temperature-compensated crystal oscillator designer with crystal units which will meet the "secured voice" single side band requirements for operation at 30 MHz has been established.

References

1. Layden, Owen P., "Frequency Control For Tactical Net SSB Equipment", Proceedings Of The 23rd Annual Symposium On Frequency Control, May 1969.

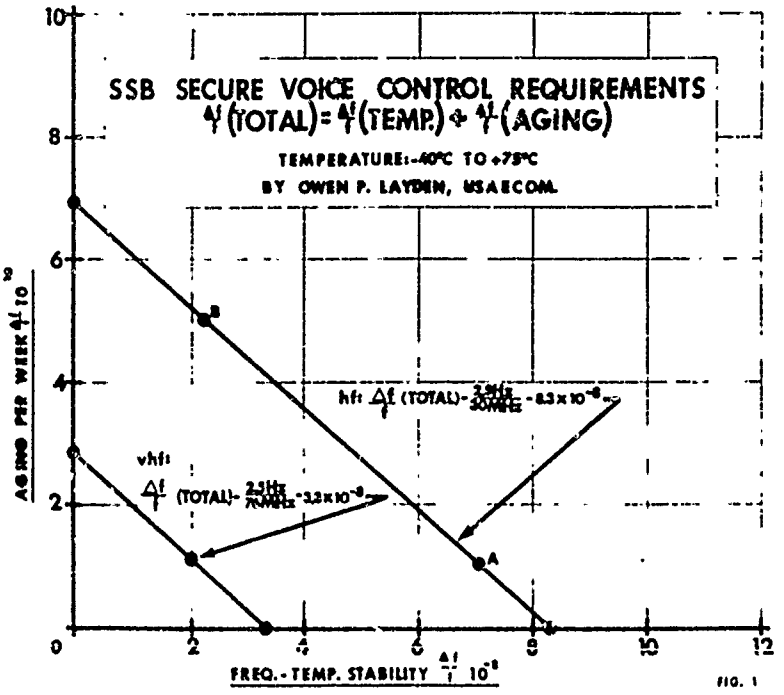

SSB SECURE VOICE CONTROL REQUIREMENTS
Δf/f (TOTAL) = Δf/f (TEMP.) + Δf/f (AGING)
TEMPERATURE: -40°C TO +75°C
BY OWEN P. LAYDEN, USAECOM.
AGING PER WEEK Δf/f x 10^8
hf: Δf/f (TOTAL) = 2.5Hz/30MHz = 8.3 x 10^-8
vhf:
Δf/f (TOTAL) = 2.5Hz/70MHz = 3.2 x 10^-8
A
B
FREQ. - TEMP. STABILITY Δf/f 10^-8

FIG. 1

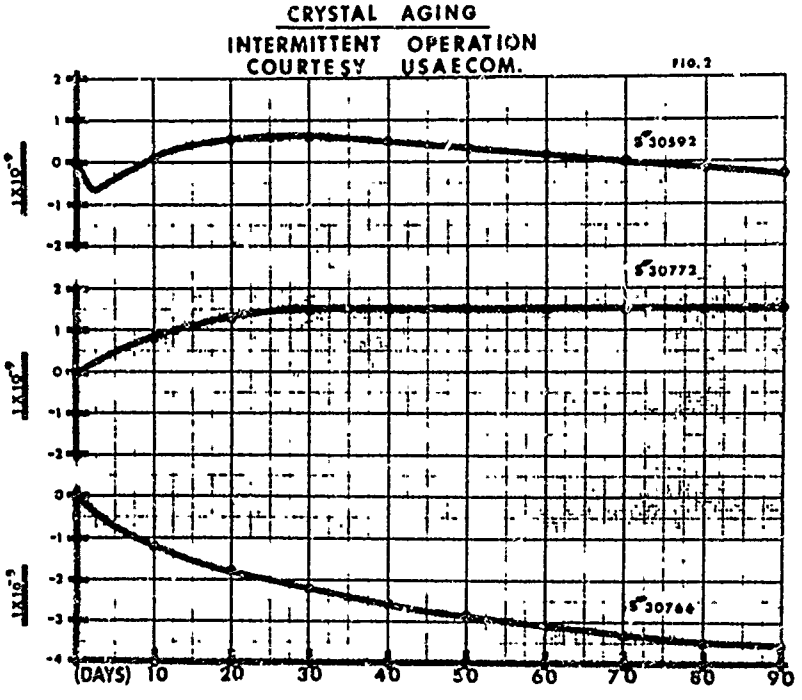

CRYSTAL AGING
INTERMITTENT OPERATION
COURTESY USAECOM.
FIG. 2
S/N 30592
S/N 30772
S/N 30766
(DAYS) 10 20 30 40 50 60 70 80 90

FIGURE 3

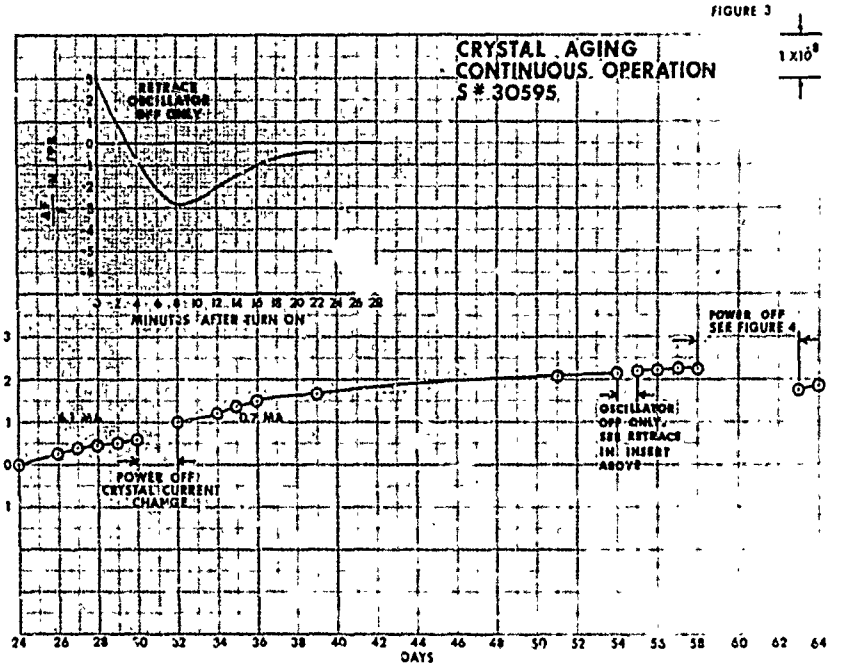
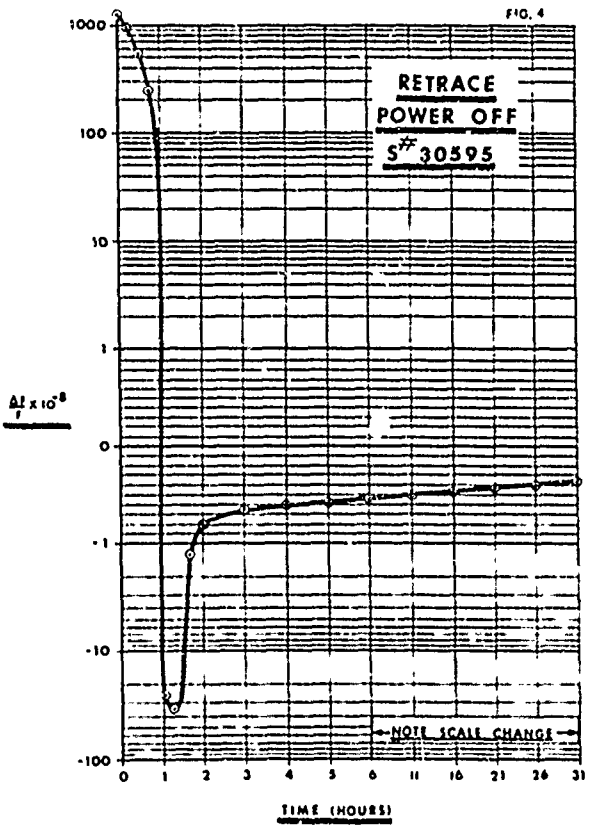


FIG. 4



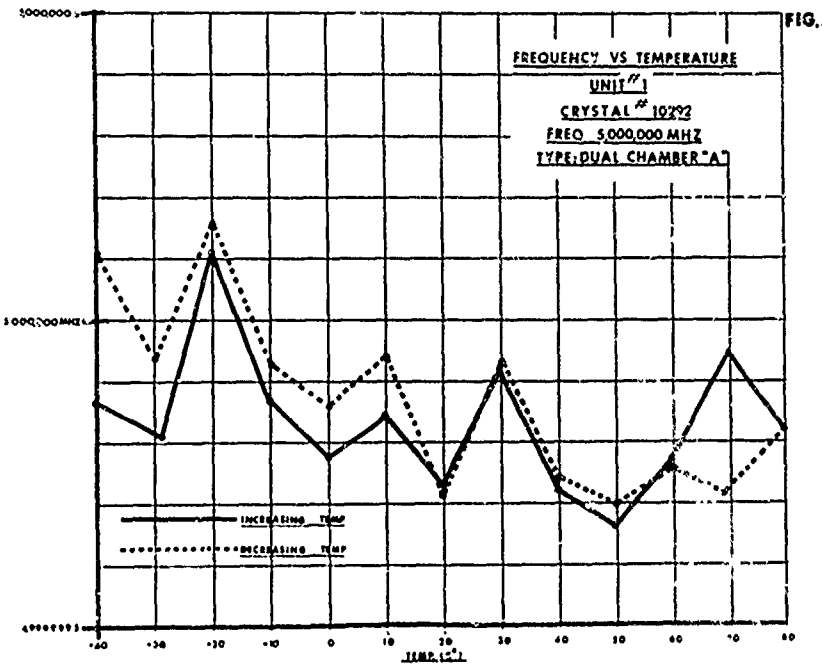
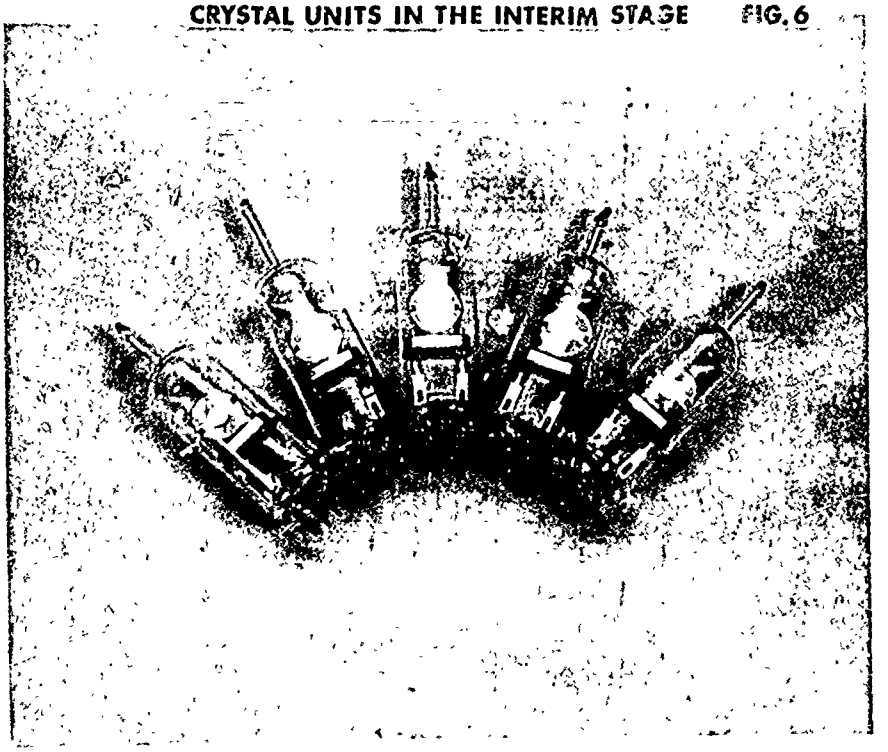
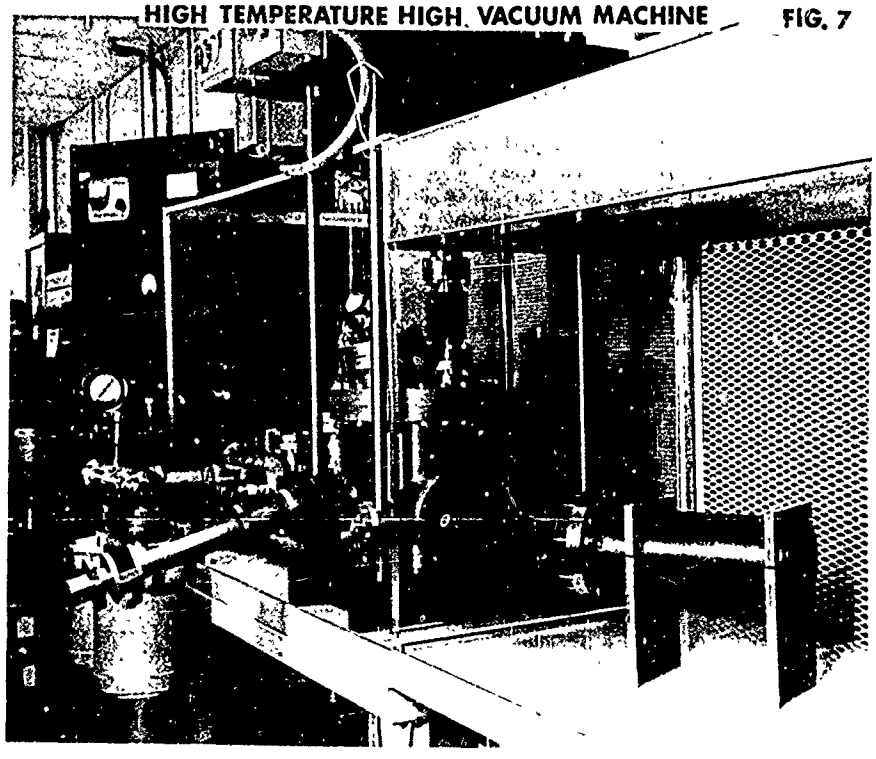


FIG. 5

CRYSTAL UNITS IN THE INTERIM STAGE FIG. 6



HIGH TEMPERATURE HIGH VACUUM MACHINE FIG. 7



DYNAMIC TEMPERATURE BEHAVIOR OF QUARTZ CRYSTAL UNITS

Walter H. Hicklin
Georgia Institute of Technology
Atlanta, Georgia

Summary

Quartz resonators of 5 MHz fundamental frequency have been fabricated for meeting the requirements of fast warm-up time and aging rate for the SSB, secure-voice communications system planned for the future. These requirements were discussed by Owen P. Layden of the U. S. Army Electronics Command during the 23rd Annual Symposium on Frequency Control, April 1969.

Measurements of resonator series resistance (R_s) vs gas pressure (P_{gas}) in atmospheres of argon, air, and helium have shown that air and argon at 50 torr cause the R_s value to increase by a factor of about two. With helium, however, a pressure of about 50 torr does not appreciably increase the value of R_s from that obtained in vacuum.

Studies of methods of heating the resonators have shown that convection ovens preset to 85°C require about 10 minutes to heat a resonator to the operating temperature regardless of the design of the crystal unit. An oil bath preset to 85°C heats the resonator more rapidly provided the holder and the leads are submerged in the oil. Oil bath heating was used for the fast warm-up studies and also for the thermal repeatability studies.

The effect of helium pressure in the holder on resonator warm-up time was found to be small for helium pressures between 740 and about 10 torr.

Aging rates of less than 5×10^{-9}/week at 85°C have been obtained for gold-plated units in HC-6 cold-weld holders after evacuating, baking, and then sealing in commercial grade helium at 10 to 50 torr. Copper-plated units sealed in helium have given less consistent results, although some were better than the gold-plated units after about 1000 hours of aging at 85°C.

Thermal repeatability measurements were made on aged units stored between measurements at -10°C. The frequency of the unit measured after three days in the oil bath at 85°C was the reference for these tests. Prior to warm-up, the unit was stabilized at 0°C and then transferred promptly to the oil bath. The frequency was recorded each minute to 10 minutes, then at 10-minute intervals to one hour. This cycle was repeated daily for four days, then once a week. No significant uncertainty in frequency measurement was noted other than that of $\pm 2 \times 10^{-8}$ caused by the periodic variation of the oil bath temperature.

The USARTL technical requirements of frequency repeatability of $\pm 1 \times 10^{-7}$ at one minute, $\pm 5 \times 10^{-8}$ at two minutes, $\pm 1 \times 10^{-8}$ at one hour were approximated. With both gold and copper-plated 5 MHz crystals, a small change from the reference frequency was found during the first few days. Thereafter, the repeatability appears to be within $\pm 5 \times 10^{-9}$.

Selection of Holder Filling Gas

Introduction

Consideration of the problem and the performance of a few experiments showed that the crystal holders must be gas filled, rather than evacuated, to meet fast warm-up requirements. The first questions to be answered were then: (1) which gas should be used, and (2) to what pressure should the holder be backfilled.

Thermal Conductivity of Various Gases

The thermal conductivity and density data used to construct the bar graph of Figure 1 were obtained from the literature. Note that as the density of the gas decreases, its thermal conductivity increases. The fact that the least dense gases are also the best heat conductors is a factor of considerable value as will be seen later.

R_s vs P_{gas}

Data in Figure 2 relate R_s to P_{gas} in the holder for argon, air, and helium; the densities of which are 1.78, 1.29, and 0.18 g/l respectively at STP. At 1 torr and below R_s is at its lowest value and is independent of the residual gas. The pressure of helium could be increased to 50 torr before a measurable increase in R_s was noted. However, for air and argon R_s doubled at 50 torr and, of course, the resonator Q decreased by a factor of two.

Hydrogen should give similar results to those obtained with helium, but we preferred the use of an inert gas, so helium was selected as

the most promising gas for additional work.

Warm-Up Measurement Procedure

Heating Methods

At this point we investigated methods of heating the resonators and recording the data. Heating of the holder by an 85°C "slug" placed over the holder cap was not used because of the uncertain thermal contact. Figure 3 shows the warm-up times obtained by placing crystal units in an air oven preset to 85°C. The warm-up time was about 10 minutes regardless of the atmosphere in the holders. Figure 3 also shows the results obtained when the same crystals were immersed in an 85°C oil bath. The oil used was mechanical vacuum pump oil. The oil bath heating method gave data consistent with the experimental variables and was used for all of the warm-up experiments including the thermal repeatability measurements to be reported later. It is important that the crystal holder, pins, socket and lead wires be immersed in the oil as will be shown subsequently. The oil causes the resonator frequency to change about - 4 parts in 10^7 but there is no measurable change in R_s and thus no loss of Q.

Data Readout

Digital warm-up time measurements were obtained as shown in Figure 4 for the slow-heating evacuated units or for ones heated in the air oven. The crystal impedance bridge was balanced and the frequency read from the synthesizer for each data point. Although this method is laborious it is accurate within a few parts in 10^9.

Analog warm-up time measurements were obtained for the fast-heating, gas-filled units in the oil bath. The Y axis of a T-Y recorder was used as the null detector of a synthesizer driven, crystal-impedance bridge. After the base line was established and the reference frequency measured (at 85°C), the synthesizer was offset 0.5 Hz and the error signal representing $\pm 1 \times 10^{-7}$ was established and labeled. Similarly, the error signal at a frequency offset of 5 Hz was established. The synthesizer frequency was then returned to the reference frequency and the crystal unit removed from the oil bath. After a cooling period (the temperature to which the units were cooled was varied from room temperature to -40°C) the unit was returned to the oil bath and simultaneously the recorder was started. The T axis of the recorder was then used to time when the resonator frequency crossed the pre-established "off-set" frequency levels, as shown in Figure 5. The crystal being measured was a 5 MHz fundamental unit sealed in an HC-6 holder at 100 torr of helium. The 10^7 crossover was reached in about 0.6 minutes, which is typical for such units; the precision of measurement was about ± 2 parts in 10^8.

Additional Measurements

As stated previously, it is necessary that the entire resonator including the pins and electrical connection be immersed in the oil bath in order to obtain reproducible data. An example of warm-up time measurements with only the cap in the oil is shown in Figure 6. The crystal unit frequency did not return to the reference frequency in nine minutes because heat was being lost by the wafer through the leads at almost the same rate it was being received through the walls of the cap. The indicated null point at 0.8 minutes is an artifact caused by a low turning point in this particular crystal.

Aging rates and Q values of 5th-overtone 5 MHz units are usually superior to fundamental resonators, but such units probably can not be used since they will not meet the fast warm-up specifications. Figure 7 shows the warm-up data for an overtone unit sealed in air at 740 torr. The 10^{-7} crossover occurred about seven minutes after total immersion in the 85°C oil bath; a similar unit using a fundamental quartz plate reached the 10^{-7} crossover in 1.5 minutes. Thus, the overtone unit containing five times as much quartz as the fundamental unit required almost five times longer to heat.

Determination of Optimum Helium Pressure

The final step before proceeding with the next phase of the project was to determine the optimum helium pressure in the sealed holder for meeting the fast warm-up specifications and maintaining maximum resonator Q. The experimental results are shown in Figure 8; oil-bath heating and analog readout were the conditions used. The time plotted is that required to reach the 10^{-7} crossover starting at room temperature. Several of the points were rechecked using the total dynamic temperature range of -40 to +85°C and warm-up times only 0.1 minute longer than those shown were obtained. Note that for a considerable pressure range below 740 torr the warm-up time was relatively independent of the gas pressure. This observation is consistent with the literature relating the thermal conductivity and pressure of the gases used in these experiments.

Based on the data shown in Figures 2 and 8, the optimum helium pressure range in the holder should be between 10 and 50 torr. The 10 torr value was determined by the warm-up measurements and gave a warm-up time to the 10^{-7} crossover of 0.8 minutes for a temperature range of -40 to +85°C. The 50-torr value had been obtained previously and is the maximum helium pressure that can be used without appreciable loss of Q.

Summary of Phase I

The method used to meet the fast warm-up requirements is summarized in the following paragraphs.

(1) Five MHz fundamental quartz wafers should be used.
(2) Particular attention should be paid to angle control so as to place the upper turning point at exactly 85°C. Activity dips over the temperature range of -40 to +85°C should be eliminated if possible.
(3) The metal holder should have the HC-6 configuration with the quartz centrally located so as to allow uniform heating or cooling of each major surface.
(4) The holder must be evacuated, then refilled with 10 to 50 torr of helium before cold-weld sealing.
(5) Electrical connection to the plating can either be made by ultrasonic bonds or conducting cement - the results were similar.

Aging of Resonators in Helium-Filled Holders

Introduction

When a method for fabricating crystal units capable of meeting the warm-up specifications was completed, the next step was to determine the long-term aging effect of helium filling at 85°C and under conditions of thermal cycling. Twenty resonators plated with gold and twenty with copper were fabricated for this study. All were given a low-pressure bakeout before back-filling with 10 to 50 torr of commercial-grade helium.

Aging of Gold-Plated Units at 85°C

The aging characteristics of evacuated gold-plated 5 MHz resonators was presented at last year's symposium by Belser and Hicklin. The aging of similar units in helium-filled holders was less in most cases than the evacuated crystals. A typical aging curve is shown in Figure 9. There was no apparent difference in the aging of units sealed at low and high helium pressures.

Aging of Copper-Plated Units at 85°C

The aging of copper-plated resonators in helium-filled holders was less consistent than similar gold-plated ones. Units helium-filled in the low torr range frequently had low aging rates as shown in Figure 10. Helium-filling in the high torr range frequently caused high negative aging rates as shown in Figure 11. The reason for this anomalous behavior was not determined.

Thermal Repeatability Measurements

Resonators having several hundred hours of aging at 85°C were transferred to an 85°C oil bath for a reference frequency measurement and then stored at -10°C. Each day for one week and weekly thereafter they were placed in the oil bath for 60 minutes during which time the frequency was measured using digital readout. This procedure is called a "cycle" for convenience. The data obtained are shown in Figure 12.

Gold-plated resonators typically aged positive at a rate of about 1×10^{-8}/day during the first few days, after which they became very stable. This effect is displayed in the upper graph section shown in Figure 12. Here the ninth cycle data are given relative to the original reference frequency. In the lower section, the ninth cycle data are re-plotted relative to a new reference frequency obtained during the fourth cycle when the aging was about complete. Frequency stabilization by temperature cycling is not a new concept, but seems to be especially important for units intended for fast warm-up operations.

Copper-plated resonators behaved similarly to gold-plated ones except the shift of frequency was negative. After this initial aging process the thermal repeatability was as good as that obtained with gold-plated units.

Summary of Phase II

The aging of gold-plated units in helium-filled holders at 85°C is at least as good and frequently better than those in evacuated holders. Copper-plated resonators have given less consistent performance but the problem is probably one of "quality control."

During the first few days of thermal repeatability measurements, aging rates of about $\pm 1 \times 10^{-8}$/day are obtained for gold-plated and copper-plated units. Frequency stability is usually obtained in one week or less in either case after which the aging is as good or better than that obtained at 85°C.

Recommendations for Future Research

Crystals meeting the warm-up and aging requirements can be fabricated using state-of-the-art laboratory techniques. The design of ovens and electronic circuits to complete the frequency-control package should be started. The fast warm-up oven may present a special problem, the solution of which may require a design "break-through."

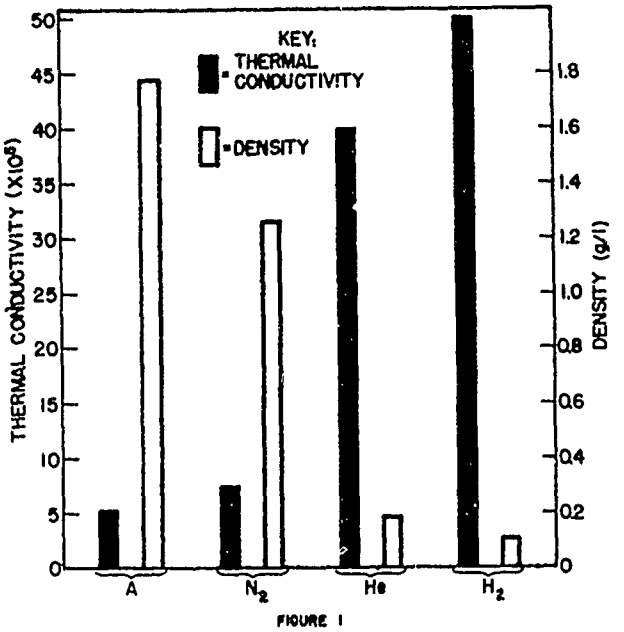


FIGURE 1

Bar Graph of Thermal Conductivities and Densities of Argon, Nitrogen, Helium and Hydrogen.

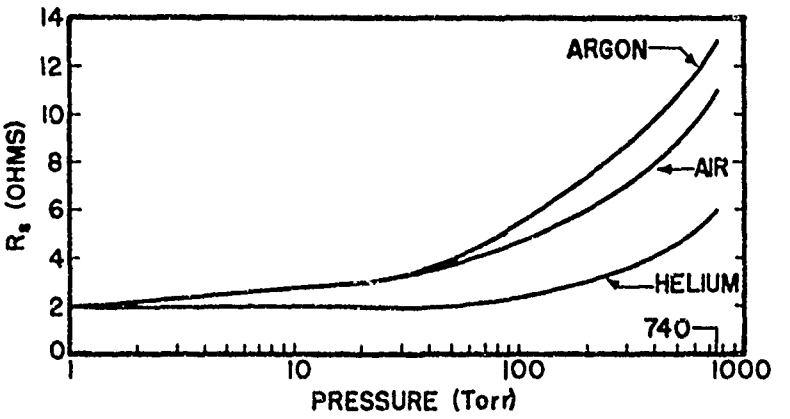


FIGURE 2

Series Resistance Plotted as a Function of Gas Pressure for Argon, Air and Helium.

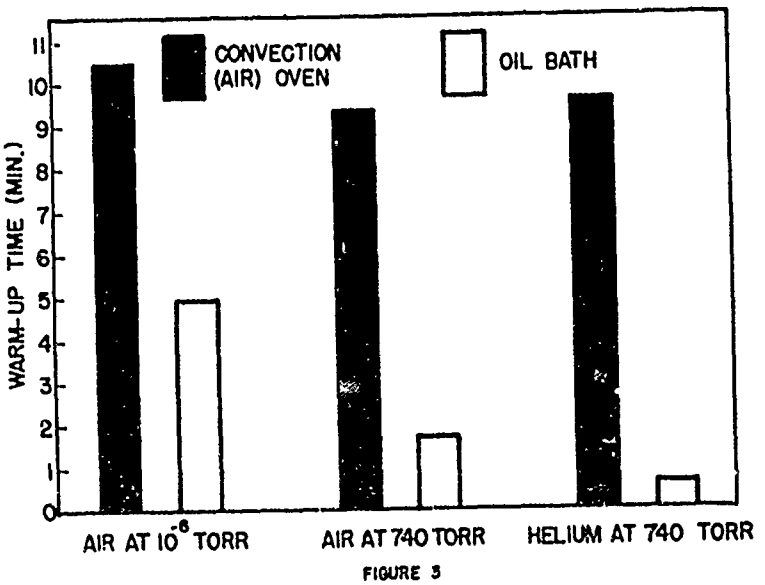


FIGURE 3

Bar Graph Relating Heating Methods and Warm-Up Time.

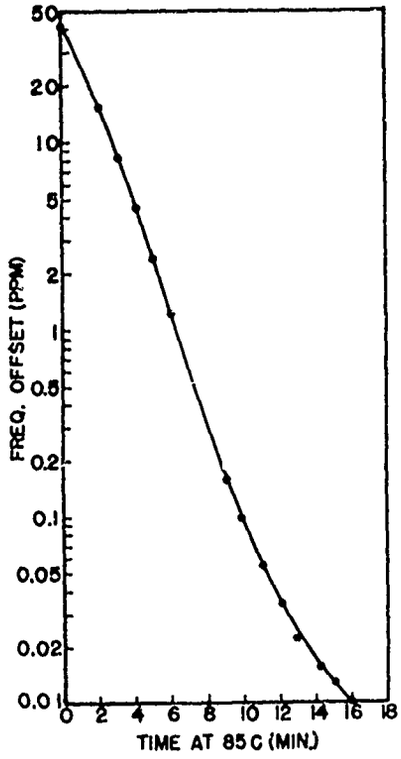


FIGURE 4

Digital Warm-Up Data for an Evacuated Unit Measured in an 85 C Convection Oven.

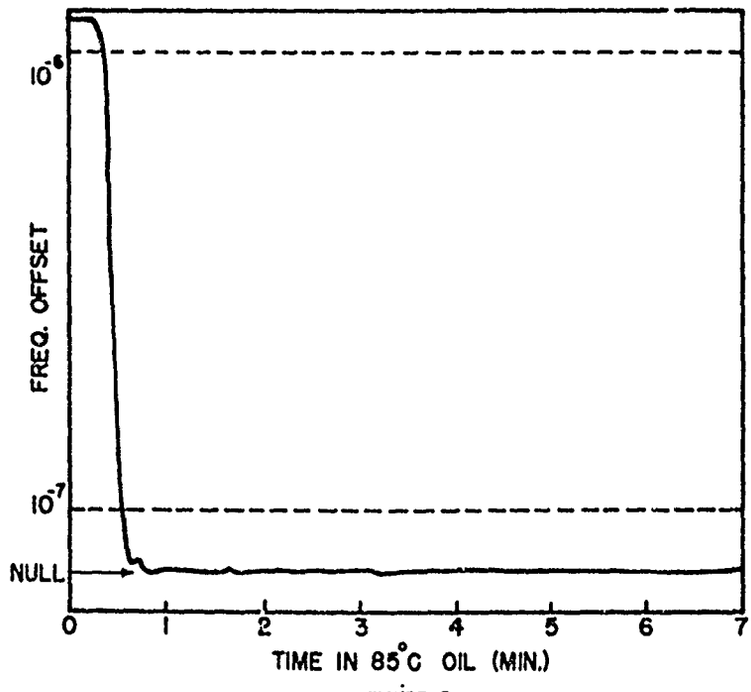


FIGURE 5

Analog Warm-Up Data for a Gas-Filled Unit (100 torr He) Measured in an 85°C Oil Bath.

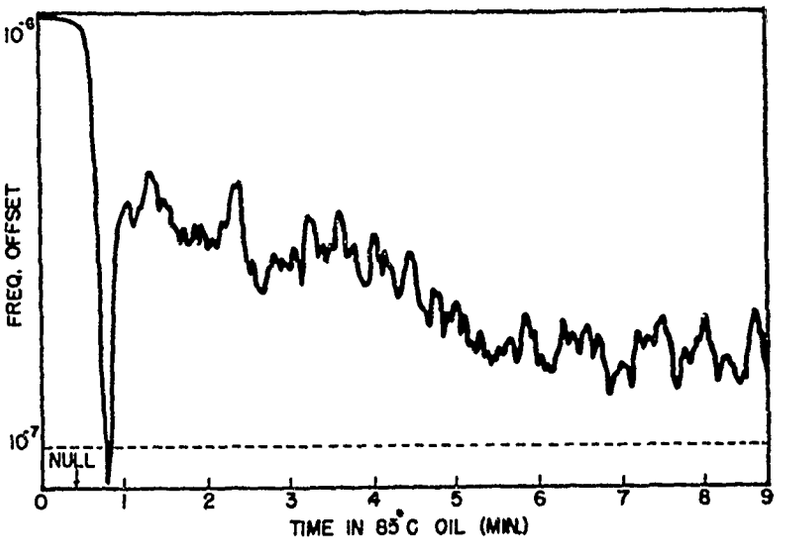


FIGURE 6

Analog Warm-Up Data for a Gas-Filled Unit Measured in an 85°C Oil Bath With the Holder Pins Above the Oil Level.

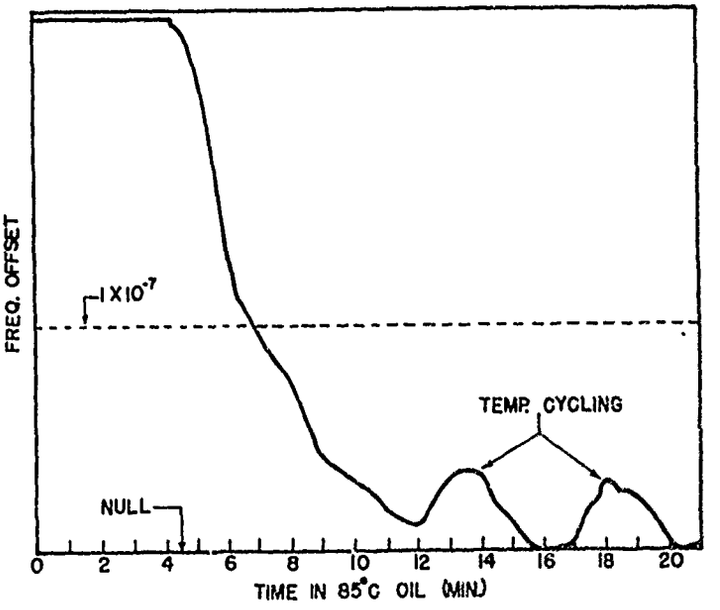


FIGURE 7

Analog Warm-Up Data for a Gas-Filled Fifth Overtone Unit Measured in an 85°C Oil Bath.

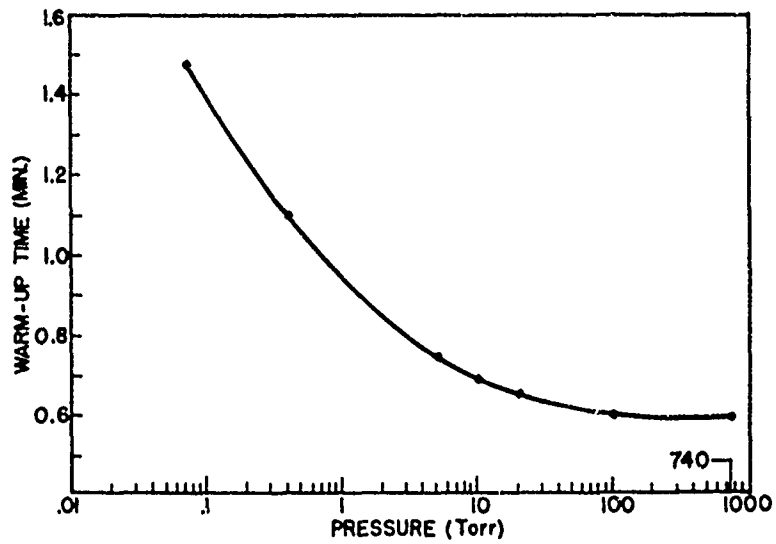


FIGURE 8

Warm-Up Time Plotted as a Function of Helium Pressure in the Crystal Unit Holder.

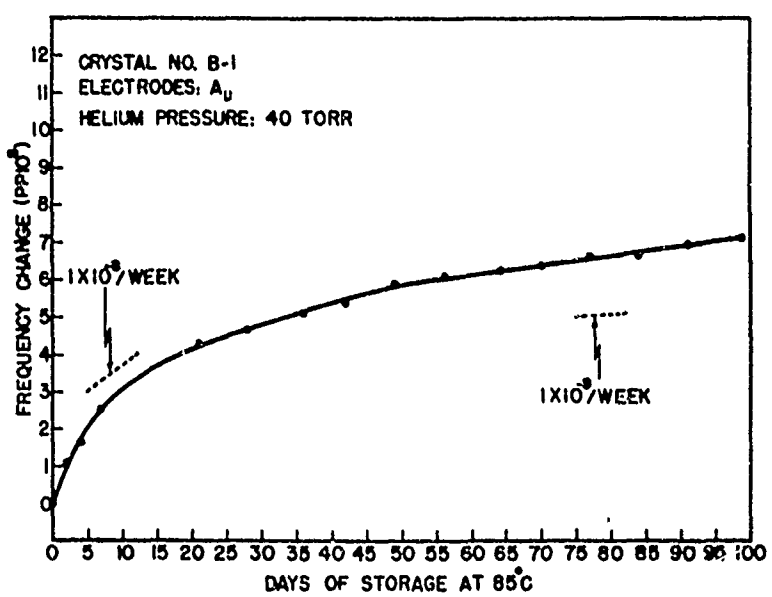


FIGURE 9

Aging Data at 85°C for a Cold-Plated Unit Backfilled With 40 Torr of Helium.

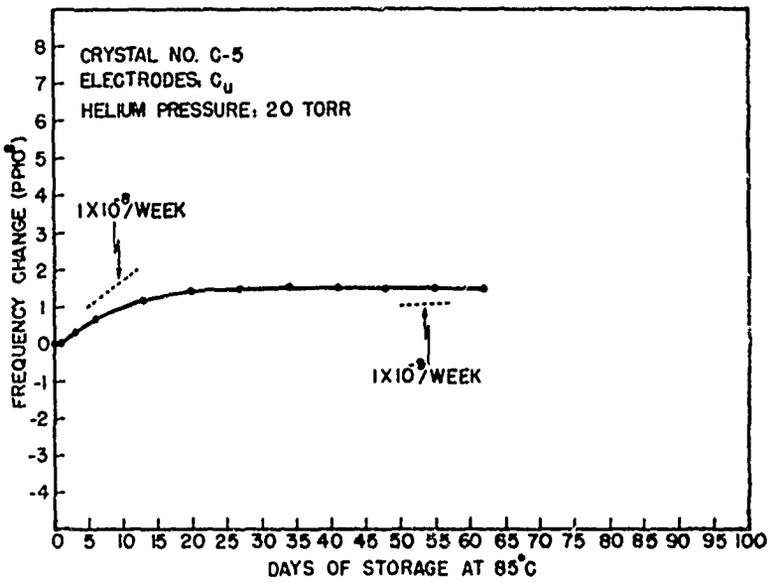


FIGURE 10

Aging Data at 85°C for a Copper-Plated Unit Backfilled With 20 Torr of Helium.

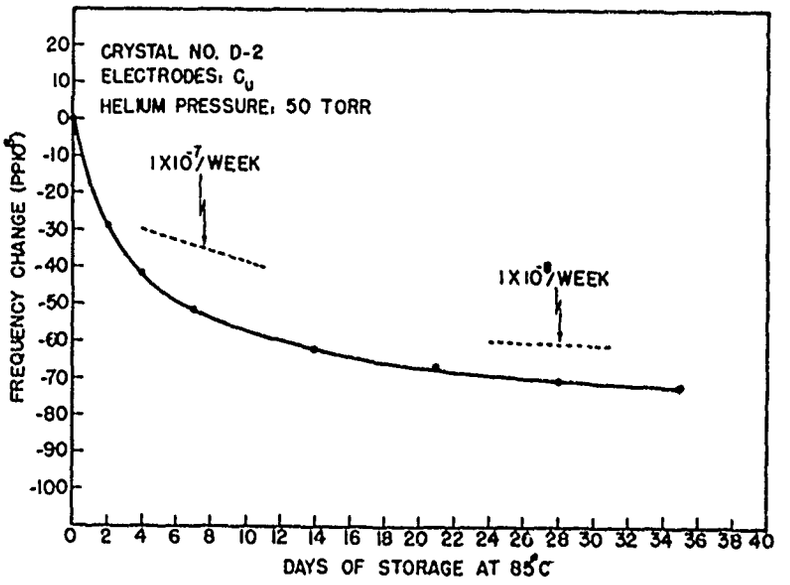


FIGURE 11

Aging Data at 85°C for a Copper-Plated Unit Backfilled With 50 Torr of Helium.

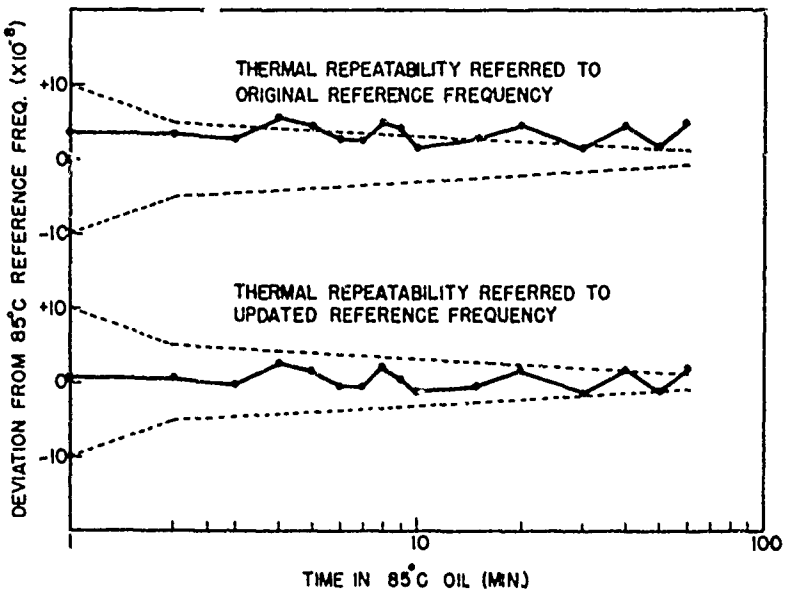


FIGURE 12

Thermal Repeatability Measuring Showing the Effect of Frequency Aging.

THE DIRECT TEMPERATURE CONTROL OF QUARTZ CRYSTALS IN EVACUATED ENCLOSURES

F. G. Tinta, A. S. Matistic, and G. A. Lagasse
Frequency Electronics, Inc.
New Hyde Park, New York 11040

Summary

This paper presents a method for direct temperature control of a quartz crystal in an evacuated enclosure. The crystal enclosure acts as a Dewar Flask, and the heat exchange with the ambient, when compared with conventional crystal ovens, is greatly reduced. Thin metallic films deposited on the crystal surface and traversed by electric currents were used as contact heaters. Thus the time required for temperature stabilization is very short.

The thermal and piezoelectric behavior of quartz crystals, temperature controlled by contact heaters, is analyzed, and the experimental results are reported. The acquired information could be useful for the design of micro-oven integrated systems. The feasibility of integrated micro-oven crystal oscillators is demonstrated.

This method shows promise in building temperature controlled crystal oscillators of very small size and power requirements and perhaps opens a new approach for achieving Peltier-effect cooled units.

Key words (for information retrieval) Quartz Crystal, Contact Heater, Integrated Micro-Oven, Temperature Control, Analysis, Experimental Results.

Introduction

Previous experiments, in which quartz crystals were heated through a direct contact with a warmer body placed inside the crystal holder case, were performed in the course of a general study concerning the thermal behavior of the crystals used for temperature compensated oscillators.

The problems under study were related to the frequency instabilities occurring in TCXO's at abrupt temperature changes as a result of different time constants of quartz crystals and compensation networks.

For accurate temperature measurements, precision thermistors were bonded to the quartz crystals, in positions in which the oscillations were not disturbed. The resulting dissipation constant inside the evacuated enclosure of the crystal was measured using the self-heating characteristic of these thermistors. The results obtained indicated that it is possible to heat quartz crystals to high enough temperatures with relatively small amounts of power. Third-overtone, AC-cut crystals with essentially linear frequency versus temperature variation (about +20 PPM/^{o}C coefficient) were used in these tests.

Heat Transfer

The heat transfer between the heater and the crystal is essentially, in our case, a conduction process. The heat energy is supplied from an external power source in the form of an electric power, which is transformed into heat through the heater itself.

Roughly, if i_s is the supplied power (in watts), and ΔT the difference in temperature between the heater and the crystal in the contact area, then:

$$i_s = \frac{1}{R_s}\Delta T \qquad (1)$$

where R_s could be considered a heat transfer resistance, measured in ^{o}C/watt (or ^{o}K/watt) units.

The heat transfer from the inside of the crystal holder case to the external surfaces is more complex, involving:

1. Conduction through all the leads in (thermal) contact with the crystal body, i.e., through the crystal electrodes connected to the holder base pins, through the heater leads and pins and eventually through the control thermistor leads, if used. Conduction is also involved in the heat transfer between

different points of the crystal holder case. This process could be described with the same type of general equation as (1), in integral form:

$$i_c = \frac{1}{R_m}\Delta T \qquad (2)$$

There is, in our case, a physical difference between R_s and R_m; R_s being practically a contact resistance, while R_m applies rather to longitudinal structures. Here, ΔT corresponds to the temperature difference between the extremities of the heat conductor.

2. Radiation exchanges between the crystal body (including the heater) and the crystal holder case. For discussion purposes and assuming uniform temperatures for the crystal (T_x) and for the case (T_c); the heat flow i_r (watts) could be considered as being represented by:

$$i_r = k_1 T_x^4 - k_2 T_c^4 \qquad (3)$$

where T_x and T_c are measured in ^oK

k_1 and k_2 are global factors, including areas, radiation, reflection and absorption constants of the crystal body and of the case.

For a given T_x, an equivalent thermal resistance R_r, as in equations (1) and (2) will be a nonlinear function of T_c.

3. Convection through residual gases in the crystal case becomes negligible when the enclosure is properly evacuated.

The heat transfer from the external surfaces of the crystal case and pins to the ambient is also complex but could be considered, as in the Newton cooling law, as a linear process. In this way, a new thermal resistance, R_n, is introduced.

The thermal lumped constant equivalent circuit of an internally heated crystal could be represented, using electrical symbols, as in Figure 1.

In this schematic,

R_1: represents the thermal resistance between the crystal body and the ambient medium. It is advantageous to increase this resistance as much as possible, therefore using thin, long leads inside the case and thin, short pins outside, ($^oC/W$).

R_3: The resultant thermal resistance between the crystal body and the crystal case, corresponding thus to radiation exchanges and conduction between base pins and case. The polished, smooth surfaces of the cold-welded holder reduce radiation loss. The experimental results have indicated that this resistance may practically be considered as linear.

R_2: the thermal insulation resistance of case.

R_c: the equivalent thermal resistance for the temperature control system (if used).

C_1: the thermal capacitance of the crystal body ($J/^oC$).

C_2: the thermal capacitance of the case.

T_x: the crystal temperature

T_c: the case temperature

E_1: the ambient temperature

E_2: the reference temperature of the control system (if used).

Thermal Analysis

Steady Conditions

It is possible to determine the values of R_1, R_2, and R_3 by measurement of temperatures (using appropriate sensors, for example thermistors) and of heating power.

If the crystal is heated with a constant power i, we have

$$V_1 = T_x - E_1 = \frac{R_1(R_2 + R_3)}{R_1 + R_2 + R_3}\, i \qquad (4)$$

$$V_2 = T_c - E_1 = \frac{R_1 R_2}{R_1 + R_2 + R_3}\, i \qquad (5)$$

Immersing the crystal case in an oil bath maintained at the constant temperature E_1, a thermal shortcircuit of R_2 is obtained. In this case, for the same T_x is necessary to use a larger power, i'.

$$V_1 = T_x - E_1 = \frac{R_1 R_3}{R_1 + R_3}\, i' \qquad (6)$$

From (4), (5), and (6), one obtains

$$R_1 = \frac{V_2}{i - i'(1 - V_2/V_1)} \quad (7)$$

$$R_2 = \frac{V_2}{(i'-i)(1-V_2/V_1)} \quad (8)$$

$$R = \frac{V_2}{i' - i} \quad (9)$$

Transient Response

The thermal transient response to a step function of the heating power can be obtained using the classical heat diffusion equations or applying the Laplace transform.

The simplest case corresponds to a thermally shortcircuited R_2; we have

$$i' U(t) = C_1 \frac{dV_1}{dt} + \frac{V_1}{R_1} + \frac{V_1}{R_3} \quad (10)$$

and correspondingly

$$V_1(s) = \frac{i'}{C_1} \frac{1}{s(s + \frac{1}{\tau'})} \quad (11)$$

where $\tau' = R_3C_1R_1/(R_3 + R_1)$

Finally we obtain

$$V_1(t) = i' R_3(1 - e^{-\frac{t}{\tau'}}) \quad (12)$$

which is the well-known exponential form.

If R_2 is not shortcircuited, it results (13)

$$V_1(s) = \frac{i}{s} \frac{1}{sC_1 + \frac{1}{R_1} + \frac{1}{R_3 + \frac{R_2}{R_3 + R_2/(sC_2R_2+1)}}}$$

The shape of $V_1(t)$ is not very much different from (12); it resembles the transient response of a RC transmission line at a step function of electric current.[1]

Temperature Uniformity

In order to study the problem of the temperature distribution in the crystal body, it is necessary to consider a distributed constant equivalent circuit.

One of the simplest forms of such a circuit is given in Figure 2. The crystal body is deemed as being split into two sections. It is possible to draw some practical conclusions even without going into a detailed thermal analysis.

T_x' is close to T_x'' if the heat is brought in a proportional measure with the heat loss at each point.

Too, it is possible to obtain better temperature uniformity if the heat loss resistances are higher.

These findings were adhered to in the construction of the contact heaters.

Contact Heaters

Material

The contact heaters were built by vacuum depositing thin resistive films on the crystal surface. Various metals were used for this purpose, namely aluminum, nickel, and platinum. Aluminum was selected because of its low specific weight, high electric conductivity. The contact heater would be very light, thus practically eliminating some motional influence on the operation of the quartz crystal.

This material being also used for the crystal electrodes, the manufacturing process could also be simplified. However, aluminum gave unsatisfactory results with respect to the heater resistance stability.

Nickel and platinum gave, in this regard, very satisfactory results. Besides, the temperature coefficient of the resistivity of these metals allows a reliable temperature sensing by the heater itself and to use the same metallic film as heater and as temperature sensor.

It was possible to build convenient heaters with these metals, without disturbing the operation of the crystal. By depositing the heater, the corresponding frequency change was in the 100 Hz range. The Q factor ranged within the same limits as for the crystals without integrated heater, 1-1.6 million.

Geometry

In order to study the problem of direct heating, various heater geometries were used. These geometries provided a large contact area with the crystal surface and a certain degree of symmetry in heating. (Figure 3)

The heater leads are connected at the periphery of the crystal, on both front-end surfaces, and at points

disposed at 90°, in relation to the crystal electrode leads.

At (a) the heater is a narrow ring, deposited at the periphery of the crystal. It is composed of two half rings, each of them disposed symmetrically on the crystal opposite surfaces.

At (b) the heater is also a narrow ring, with the difference that in this case the metallic films on the opposite surfaces are overlapping each other about 240°. They are in electrical contact at the ends of the overlapping region. In this region the crystal is heated from both surfaces.

At (c) the heater is built in the same way as in (b), with the difference that the ring is wider.

At (d) one of the crystal electrodes also acts as a heater. Heat is dissipated in the central region of the oscillating quartz crystal and along the connecting bands (strips).

At (e) one of the crystal surfaces is completely plated, thus covering the lower electrode area. The heater supply voltage is applied at the same points as before. The distribution of the dissipated heat depends on the distribution of the electric current in the heater, does not reproduce directly as its geometry.

Geometries (b) and (c) were found, after experimental tests, as being the most suitable for the application. It should be noted that in these geometries we have more specific dissipated heat in the vicinity of the crystal electrode leads. This results from the fact that in this zone (non overlapping layers which are electrically in parallel for the remainder of the crystal), the heater displays a larger resistance per unit length. This is convenient for temperature uniformity because the heat loss in this zone is also larger.

Electrical Resistance

The DC resistance measured between the heater leads was in the range of 100 to 1000 ohms. It is possible to obtain the required value of this resistance with sufficient reproducibility. The most often used value was about 300 ohms.

Integrated Oven Crystals

Third overtone, AT-cut, 5 MHz quartz crystals were used for these experiments. In some cases, independent temperature sensors (precision thermistors) were bonded to the crystal electrodes and the heater, as shown in Figure 4. This made it possible to check the temperature distribution and make measurements on the thermal transients. Photographs of such crystals with integrated heaters are shown in Figure 5.

The crystals were properly cleaned and enclosed in cold welded type C holders. The base of these holders has been provided with five pins. The metallic holder case was processed in order to obtain smooth, reflecting surfaces (by electro-polishing). The vacuum was in the 10^{-8} mm Hg range.

Experimental Results

Thermal Quiescent State

Temperature Distribution. Precision thermistors (Model 51TF25, Gulton) were simultaneously placed in various positions on the crystal surfaces. In the geometries of (b) and (c) these positions are shown in Figure 4. A few seconds after turn-on of a constant voltage power supply, the thermistors indicated a uniform temperature within the calibration error (±0.2°C).

Thermal Dissipation of the Crystal. Several tests were performed, using a constant voltage power supply. Operating at ambient temperature (25°C), the crystal and case temperature dependence on the heating power was examined. Some typical results are displayed in a graphic form in Figure 6.

Curve (1) shows the temperature dependence of a non-encapsulated crystal.

Curve (2) shows the temperature dependence of an encapsulated crystal whose case was immersed in an oil bath maintained at 25°C. The temperature of the case and of the pins was practically identical to that of the oil bath.

Curve (3) shows the temperature dependence of the same crystal, whose case was wrapped in a thin glass wool bed, in order to avoid the disturbing effects of the air currents.

Curve (3') corresponds to the temperature dependence of the case, under these conditions.

Curve (4) shows the temperature dependence of the crystal for the best realized device. No thermistor was placed on the crystal, thus eliminating a part of the heat loss by radiation and

conduction through the corresponding leads.

Another series of tests were performed in which the environmental temperature was changed in the range from -40°C to 75°C. The temperature of the crystal was maintained at 75°C, supplying the heater with the required amount of power. Reference results are shown in Figure 7.

The performed measurements allowed the computation of thermal loss resistances, according to the formulas of (7), (8), and (9).

It was generally found that in the final integrated oven crystals R_1 were very large as compared with R_2 and R_3 and could thus be neglected in some computations. The ratio between R_3 and R_2

$$R_{3/R_2} = 1 - V_{2/V_1} \qquad (14)$$

was in the range between 2 (worst case) and 6 (best case), thus giving an estimate of the achieved degree of the thermal insulation of the crystal body in the evacuated enclosure.

The worst case involved crystals with heavy leads and pins, with one of the ends of the heater connected electrically (and thermally) to the case, and with bonded thermistors on the crystal surfaces.

A figure of practical interest is the overall dissipation constant G, in our situation:

$$G = \frac{1}{R_2+R_3} \qquad (15)$$

It was found that the value of G was in the range of 8-3mW/°C.

Good Dewar-flask enclosed crystal ovens have a dissipation constant of about 12mW/°C. In very carefully produced units, dissipation constants of 8mW/°C are found, approaching the more pessimistic values. Nevertheless, practical advantage of the integrated oven remains in its space saving and rugged construction.

Further improvements of G are still possible, using thinner leads and pins and mirror-like finished internal surfaces. A closer examination of the heating power versus ambient temperature graphs shows that for the best integrated oven crystals, the radiation has a certain contribution to the heat loss. In this respect, nickel film heaters displayed very convenient features.

Thermal and piezoelectric transients

Turn-on and turn-off characteristics were studied, and temperature and frequency recordings were made at one second time intervals during the stabilization time. The crystals were heated at constant power.

For the study of thermal transients, thermistors in conjunction with a digital printing ohmmeter were used. The measuring current for the thermistor was kept to a sufficiently low value to avoid self-heating effects. The graph of Figure 8 was obtained for a type b geometry (see figure 3 and 4), the measuring thermistor being placed in position 2). The heating power was 200 mW. The results shown are typical. The curves have the expected features, similar, as previously stated, to the transient response of a noninductive electric cable (RC transmission line). About five seconds after turn-on or turn-off (time related with a heat propagation delay), we have an inflexion point. Later, the temperature changes in an exponentiallike manner, asymptotically approaching the stabilization temperature. After three minutes the temperature difference is generally less than 2% of the whole temperature shift.

In general, the frequency variation does not reproduce the thermal response, although the correlation between starting and stabilization values was very good (within the experimental errors, corresponding to temperature differences in the .1°C). Solid state, three stage, Pierce configuration oscillators were used for the measurement of crystal frequency transients. These oscillators were placed in temperature controlled chambers, allowing the possibility of having various starting temperatures. Printing counters were used for frequency recordings. The most significant measurements for secondary piezoelectric effects were made for temperature shifts outside frequency turning points.

Generally the frequency change versus time, after both turn-on and turn-off of the heater, displays an overshoot or an undershoot, followed by a slower transient. In figure 9, a unit on the Y scale represents the total change in frequency, corresponding to about 8ppm. The overshoot was noted for crystals provided with heaters in (a), (b), and (c) geometries, whereas for the geometries (d) and (e), an undershoot took place.

This phenomenon is probably related to thermo-mechanical action. A possible, rough explanation could consist in the fact that the central oscillating area would experience mechanical stresses due to thermal expansion of the first heated parts of the crystal, at power turn-on. For the geometries (a), (b), and (c), the pheriphery is heated first, and thus the central area is subjected to pulling mechanical efforts. For the geometries (d) and (e) the central area is heated first, but its thermal expansion is hampered by the remaining zone of the crystal. It is known that such mechanical efforts could affect the oscillating frequency of a quartz crystal. But, in fact, the actual situation is more complex. It is to be noted that for geometries (a), (b), and (c), the heat is distributed from both front end surfaces at the same time, but not in a truly symmetric way with respect to the X axis of the crystal (vertical position in Figure 4); whereas for (d) and (e) only one of the front end surfaces is heated. The crystal bodies would experience different kind of deformations at turn-on.

Another fact to be noted is that the turn-on and turn-off frequency characteristics are more or less symmetrical. Without delving into further explanation of these facts, this not being the purpose of this paper, we conclude with the remark that the overshoot can be used efficiently to shorten the overall frequency stabilization time.

The slower transient which follows is very likely related with the equalization of the temperature in the crystal body.

If the stabilization temperature corresponds to a turning point of the crystal, the frequency will be approached, within a given frequency error, in a shorter time. An integrated oven crystal oscillator with proportional oven control in operation reached the nominal frequency within a relative error of 2×10^{-7} in about 20 seconds and of 4×10^{-8} in about 90 seconds, starting from room temperature. The total frequency change was about 9×10^{-6}. (Low turning point crystal)

The output of integrated oven crystal oscillators was also studied with a spectrum analyzer (HP Model 8553L). No extra spurious signals were noted during the operation of such oscillators under either steady or turn-on conditions, as compared to conventional oscillators.

Operation of the Integrated Oven Crystal Oscillators

As stated earlier, integrated oven crystals were used to build proportional controlled oven oscillators. The electronic circuitry for both the oscillator and the control amplifier used hybrid integrated semiconductor chips. The temperature sensor was a 50 kΩ thermistor bonded to the crystal electrodes as in Figure 4 (position 1 or 2).

The operation proved to be very satisfactory, and the construction simple, rugged, lightweight, and small in size. An experimental breadboard model was displayed during the 24th Annual Symposium on Frequency Control.

Further simplification of such oscillators is contemplated, using the heater circuit as a sensor as well. A schematic diagram is shown in Figure 10. The heater is mounted in a four arm bridge. Small temperature coefficient resistors will be used for R_1, R_2, and R_3. For instance, Corning Glass C style resistors have a temperature coefficient of $.1\times10^{-3}/^{o}C$. The temperature coefficients for Platinum is about $3.9\times10^{-3}/^{o}C$ and $6.17\times10^{-3}/^{o}C$ for Nickel.[2]

The differential amplifier will sense the difference voltage between the centers of the two branches. A gain of 80 dB will be more than enough to control the temperature of the crystal to a fraction of a degree. The dissipated power in R_1 will represent an extra amount of about 10%. R_2 and R_3 could be made large enough to achieve negligible power consumption. The fact that the voltage supply of the bridge varies with the temperature difference between the nominal and actual value will contribute to a fast thermal transient response.

Till now we do not have conclusive data regarding the aging of such oscillators. Further tests are still necessary. The temperature changes of the crystals at turn-on are fast. The temperature uniformity would not be perfect, and relaxation processes could play a more important role in the aging. However, such processes will probably not be much more important than in the case of TCXO's, and for continuous operation the situation will not differ too much from those of the conventional ovens.

For a thermistor controlled integrated oven oscillator, the total frequency change (excluding the initial five minutes) was about 2×10^{-8} after the first three days of continuous operation.

Future Prospects

The use of hybrid integrated circuits could allow a very compact form of construction, including all electronic circuitry and the quartz crystal in the same evacuated case.

The oscillating quartz crystals could also be cooled (or heated), using the reversible Peltier effect. A stable, low operating temperature contributes to a very low aging rate.[3] Including the cooling junction in the evacuated crystal case, the needed amount of power, which is still large, could be reduced.

The method of direct temperature control (inside the crystal holder case) could be of use not only for oscillators, but also for other piezoelectric devices, such as crystal filters or delay lines.

References

1) Ernst Weber: Linear Transient Analysis, volume II, p. 318, 331 (J. Wiley & Sons 1956)

2) W. Hunsinger:Temperaturmessung. Encyclopedia of Physics, editor S. Flügge, volume XXIII p. 393 (Springer Verlag 1967)

3) P. Gheorghiu: The Peltier Effect in Frequency Stability Management. Frequency Technology, vol. 7, p. 35 (July 1969)

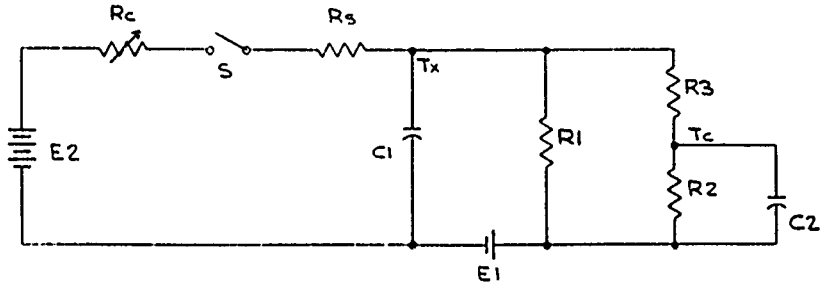


FIGURE 1. A THERMAL LUMPED CONSTANT EQUIVALENT CIRCUIT FOR AN INTERNALLY HEATED CRYSTAL

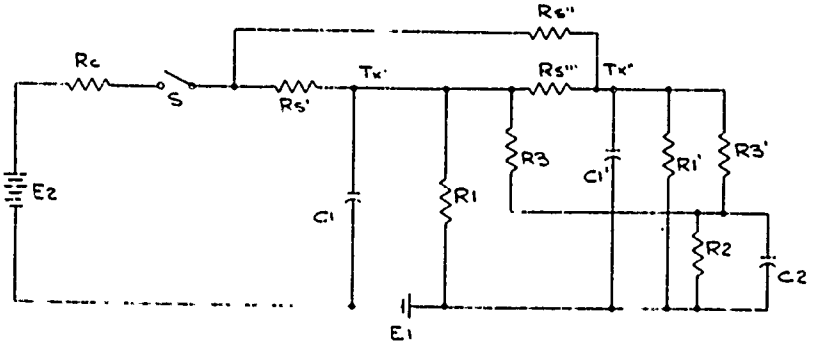


FIGURE 2. A THERMAL DE [illegible] CONSTANT EQUIVALENT [illegible] FOR AN INTERNALLY HEATED CRYSTAL

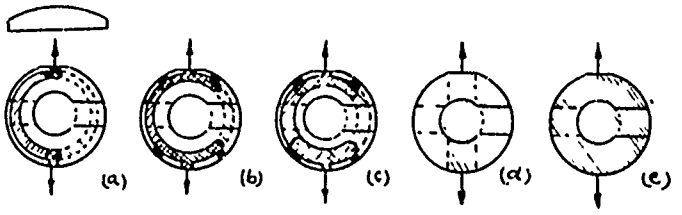


FIGURE 3. GEOMETRIES OF CONTACT HEATERS

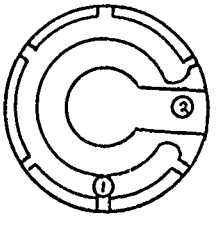
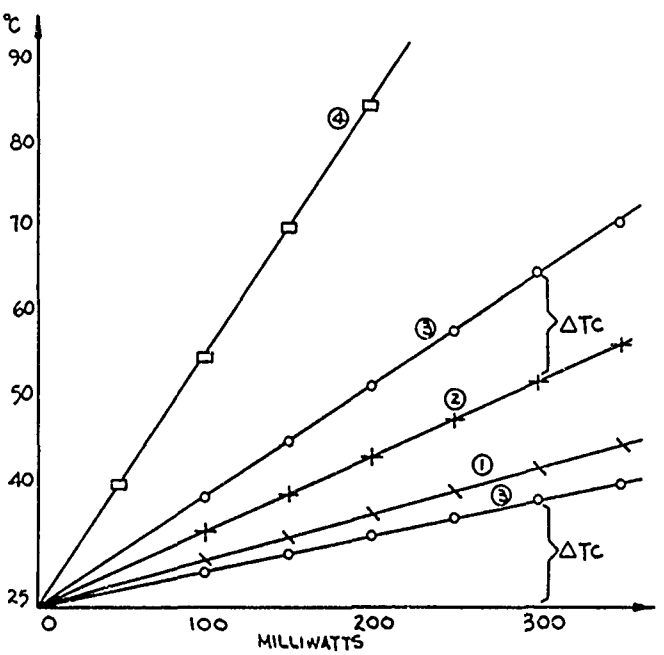


FIGURE 4. SENSOR LOCATIONS ON QUARTZ CRYSTALS WITH HEATERS

FIGURE 5. QUARTZ CRYSTALS WITH CONTACT HEATERS

FIGURE 6. CRYSTAL AND CASE TEMPERATURE VS. HEATING POWER



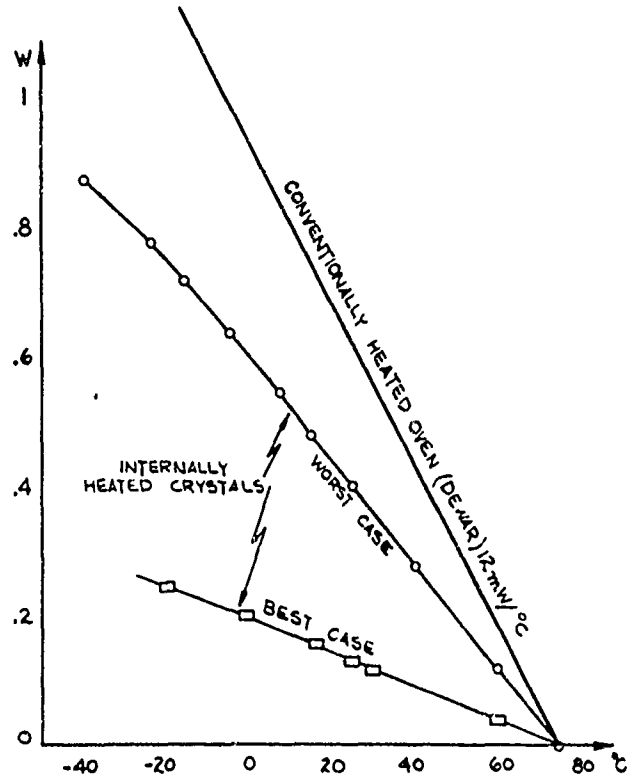


FIGURE 7. HEATING POWER VS. AMBIENT TEMPERATURE

FIGURE 8. THERMAL TRANSIENT RESPONSE

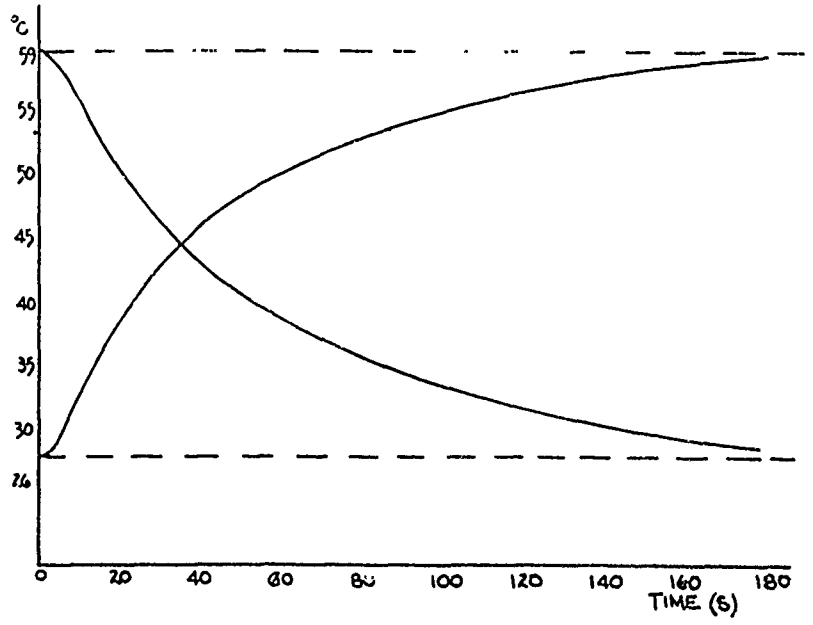
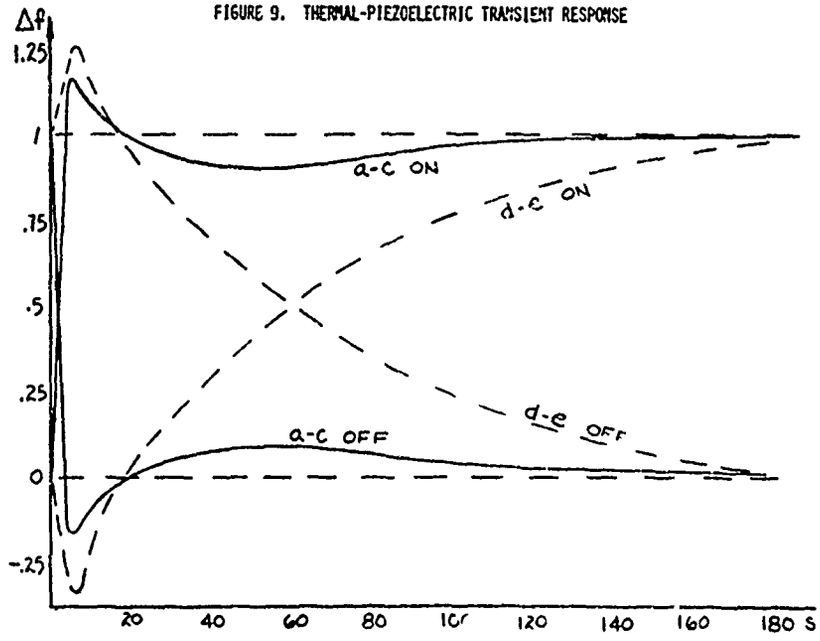


FIGURE 9. THERMAL-PIEZOELECTRIC TRANSIENT RESPONSE



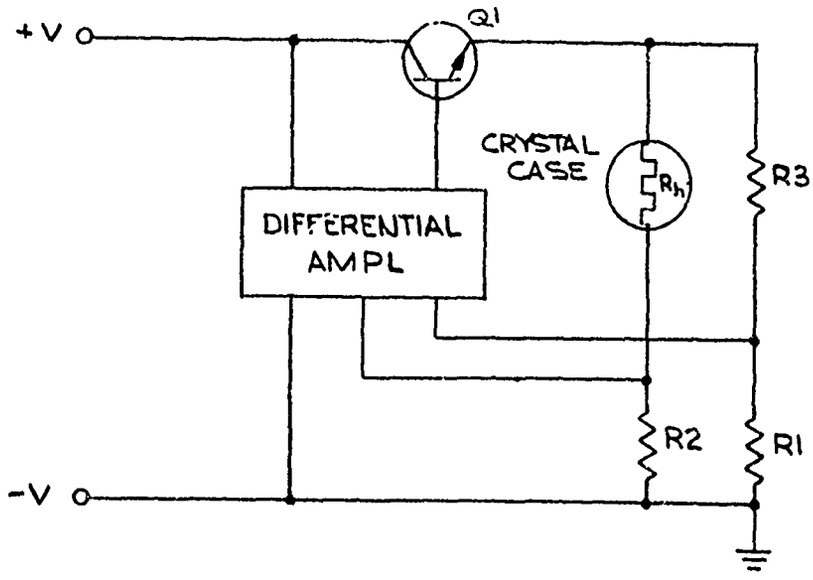


FIGURE 10. TEMPERATURE CONTROL USING INTEGRATED HEATER AS TEMPERATURE SENSOR.

PRECISION MEASUREMENT OF CRYSTAL FREQUENCY BY MEAS OF "CENTER LINE METHOD"

Issac Koga
Research Laboratory, Kokusai Denshin Denwa Company
(Japan's Overseas Radio and Cable System)

Summary

If we plot the square of the crystal impedance (Z) vs frequency (f) of the exciting current, the locus——Center Line——of the mid-point between two points which give equal values of Z^2 is, to a very close approximation, a straight line which intersects with abscissa at the very point of natural frequency, regardless of the shunt capacitance.

Measurement method of impedance at a definite crystal current is presented.

With the aid of this method it is observed that normally the frequency of a crystal increases (decreases) roughly in proportion to the square of the crystal current if the frequency-temperature coefficient of the plate at ambient temperature is positive (negative). It is also noteworthy that crystals of double convex type with very small frequency-temperature coefficients undergo very remarkable amplitude effect.

Key words (for information retrieval) Crystal, Frequency Measurement, Center Line Method, Amplitude Effect.

Introduction

Realizing that (1) the frequency of the motional arm of the equivalent circuit of crystal plate, that is, the frequency determined by L_oC_o in Fig. 1 depends upon the amplitude of vibration, and (2) the shunt capacitance between the electrodes are not definite in accordance with the working condition, the following "Center Line Method" has been devised.

Principle of Measurement

The schematic diagram of measuring circuit is shown in Fig. 2. In this figure Cp, Cq, and Cr represent equivalent capacitances between the electrodes of crystal units and earth, and C_a represents an additional known capacitance referred to later. LC is an electric circuit tuned to the crystal frequency. Current through crystal unit associated with Cr and C_a is practically equal to that through LC, because the resistance of the arm LC is very small compared with the reactance of Cq. The current I through LC is easily measured by reading the terminal voltage across C.

If we plot the square of the crystal impedance (Z) vs frequency(f) of the exciting current as shown in Fig. 3, the locus——Center Line——of the mid-point between two points which give equal values of Z^2 is, to a very close approximation, a straight line which intersects with abscissa at the very point f_o or the natural frequency. The proof is given as follows:

The impedance $\dot{Z}$ between the terminals AB is given by

$$\dot{Z} = \frac{\frac{1}{j\omega C_s}\left[R_o + j\left(\omega L_o - \frac{1}{\omega C_s}\right)\right]}{R_o + j\left(\omega L_o - \frac{1}{\omega C_o} - \frac{1}{\omega C_s}\right)} = \frac{\frac{1}{j\omega C_s}\left(R_o + jX_o\frac{2\Delta\omega}{\omega_o}\right)}{R_o + j\left(X_o\frac{2\Delta\omega}{\omega_o} - \frac{1}{\omega C_s}\right)} \qquad (1)$$

where

$$\left.\begin{aligned} \omega &= 2\pi f \\ \omega_o^2 &= \frac{1}{L_oC_o} \\ X_o &= \sqrt{\frac{L_o}{C_o}} \\ \Delta\omega &= \omega - \omega_o \\ C_s &= C_1 + C_r + C_a \end{aligned}\right\} \qquad (2)$$

Since the quantity $\frac{1}{\omega C_s}$ can be taken as constant in the frequency range of measurement, we put

$$\frac{1}{\omega C_s} = \frac{1}{\omega_o C_s} = X_1, \text{ a constant.} \qquad (3)$$

Let the two values of $\Delta\omega$ be $-\Delta\omega_1$ and $+\Delta\omega_2$ where the impedances are equal to Z. Then Eq. (1) is transformed as follows:

$$\frac{Z^2}{X_1^2} = \frac{R_o^2 + \left(X_o\frac{2\Delta\omega_2}{\omega_o}\right)^2}{R_o^2 + \left(X_o\frac{2\Delta\omega_2}{\omega_o} - X_1\right)^2}$$

$$= \frac{R_0^2 + \left(-X_0 \frac{2\Delta\omega_1}{\omega_0}\right)^2}{R_0^2 + \left(-X_0 \frac{2\Delta\omega_1}{\omega_0} - X_1\right)^2}$$

$$= \frac{\left(X_0 \frac{2\Delta\omega_2}{\omega_0}\right)^2 - \left(-X_0 \frac{2\Delta\omega_1}{\omega_0}\right)^2}{\left(X_0 \frac{2\Delta\omega_2}{\omega_0} - X_1\right)^2 - \left(-X_0 \frac{2\Delta\omega_1}{\omega_0} - X_1\right)^2}$$

$$= \frac{\Delta\omega_2 - \Delta\omega_1}{\Delta\omega_2 - \Delta\omega_1 - X_1 \frac{\omega_0}{X_0}} \qquad (4)$$

Putting

$$\left.\begin{aligned} y &= Z^2, \\ x &= \frac{1}{2}(\Delta\omega_2 - \Delta\omega_1) \\ &= \frac{1}{2}(\omega_0 - \Delta\omega_1 + \omega_0 + \Delta\omega_2) - \omega_0 \end{aligned}\right\} (5)$$

we get the following relation

$$xy = X_1\left(\frac{\omega_0}{2X_0} y + X_1 x\right) \qquad (6)$$

This an equation of rectangular hyperbola with the following asymptotes (see Fig.4)

$$\left.\begin{aligned} y &= X_1^2, \\ x &= \frac{1}{2}\omega_0 \frac{X_1}{X_0}. \end{aligned}\right\} \qquad (7)$$

Since, in Eqs. (5), $\frac{1}{2}(\omega_0 - \Delta\omega_1 + \omega_0 + \Delta\omega_2)$ gives the abscissa of the mid-point between two points for a single value of Z^2 (or y), x gives the distance between the abscissa of the mid-point and ω_0 (or the natural angular frequency). Therefore the intercept of the curve given by Eq. (6) to the abscissa is at the origin (or $x=0$), i.e. the natural angular frequency ω_0.

Since the origin ($x=y=0$), is far from the asymptote $y=X_1^2$, and very near to the asymptote $x=\frac{1}{2}\omega_0 X_1/X_0$, the curvature there is nearly zero, and Eq. (6) is very closely approximated to

$$x = -\frac{1}{2}\frac{\omega_0}{X_0 X_1} y, \qquad (8)$$

that is, the tangent to the curve at $x=y=0$.

If a capacitance C_a is added in parallel with the crystal unit, the inclination of the Center Line will increase, but the intersecting point with the abscissa will not be changed, as shown in Fig. 5, since the latter is independent of C_a.

In case when the natural frequency of a crystal plate is dependent upon the strength of crystal current, as is generally the case, the position of the Center Line will be displaced without the change of inclination, as shown schematically in Fig. 6. Therefore, even a very small change of natural frequency of the order of 10^{-5} can easily be detected.

Thus the straightness of the locus (Center Line) and the coincidency of the intersections of Center Lines with abscissa for different values of C_a will well prove the accuracy of results.

Details of the actual arrangement for measurement will be ommitted here. Care must be taken to avoid mutual static and magnetic inductions between several elements in the measuring circuit. Another important point is that the wave form of the applied voltage V_0 is to be sufficiently free from higher harmonic contents. Otherwise the impedance value calculated from the readings V_0 and V_1 will be detrimental to the accuracy of result.

Practical Example of Measurement

In Fig. 5 an example of the result of measurement is shown. The crystal under measurement was wire-mounted. The crystal current I referring to Fig. 2 was kept constant at about 10 mA by keepting the voltage V_1 at one volt flat. The reactances of L and C were both about 100 ohms. In Table 1 two frequencies (f_1's and f_2's) and their mean values (f_m's) at $V_0 = 0.6, 0.8, \cdots$ volts flat are given for the case of $C_a = 0$. Table 2 gives mean values (f_m's) of frequencies for $C_a = 0$, 15, 25 and 35 pF. These mean values are plotted in Fig. 5. Taking each Center Line as straight line, the intercepting point to the abscissa was determined by means of least square method. The final value of natural frequency or the mean value of the intercepts was

3,142,840.49 ± 0.03 Hz.

Effect of Amplitude upon Natural Frequency

From the results of many measurements, it seemed that, in almost all the cases, the natural frequency of a crystal plate was more or less dependent upon the amplitude of vibration or the crystal current. The following two points amongst others were noteworthy: (1) Some of the authors* have reported that the effect of amplitude upon the natural frequency for AT-Cut is contrary

*For instance, A. seed: Non-Linear Effects in Quartz Resonators, Fourth International Congress on Acoustics, Copenhagen, 21-28 August 1962.

to BT-Cut. However, we interpreted, as far as the normal working conditions were concerned, that the frequency change were mainly caused by the temperature rise from the ambient temperature due to ohmic loss produced by crystal current, because, as a matter of fact, both in AT- and BT-Cut plate the frequency increased (decreased) roughly in proportion to the square of the crystal current when the frequency-temperature coefficient of the plate was positive (negative). The effect due to non-linearity of stress vs strain in quartz crystal during the vibration seems to be negligibly small compared with the effect of heat generation due to ohmic loss.

(2) Crystals of double convex type with very small frequency-temperature coefficients underwent very remarkable amplitude effect. The frequency increased in proportion to the square of crystal current. The amount of change was roughly 0.01 Hz·(mA)^2 for 1 MHz, that is, about 1 Hz and 4 Hz increase at the crystal current of 10 mA and 20 mA respectively from the frequency at zero crystal current.

V_o	f_1	f_2	f_m
1.8	682.5	989.9	836.20
1.6	700.7	973.5	837.10
1.4	718.6	957.2	837.90
1.2	736.5	940.7	838.60
1.0	754.3	924.1	839.20
0.8	771.5	907.8	839.65
0.6	789.2	890.9	840.05

Table 1. V_o vs f_1, f_2 and their mean values f_m in the case $C_a = 0$. Values shown in this Table are to be read for instance 3,142,682.5 Hz for 682.5 at the upper left corner.

V_o \ C_a	0	10 pF	25 pF	35 pF
1.8	36.20	30.45	22.40	16.55
1.6	37.10	32.60	26.20	21.65
1.4	37.90	34.45	29.55	26.05
1.2	38.60	36.05	32.40	29.80
1.0	39.20	37.45	34.80	33.25
0.8	39.65	38.45	36.90	35.80
0.6	40.05	39.35	38.45	37.85

Table 2. V_o in volts vs f_m for various values of C_a. Values shown in this Table are to be read for instance 3,142,836.20 Hz for 36.20 at the upper left corner.

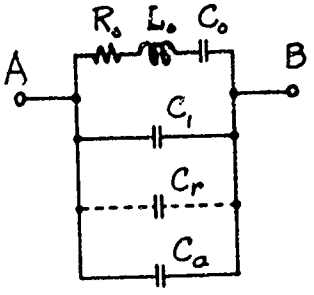


FIGURE 1

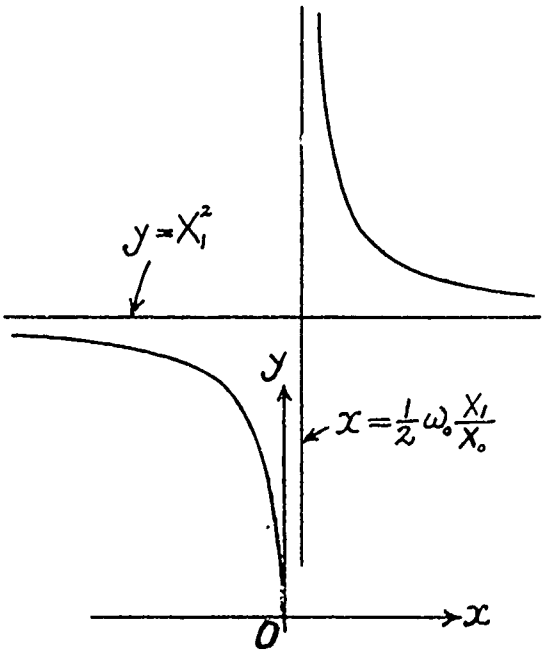


FIGURE 4

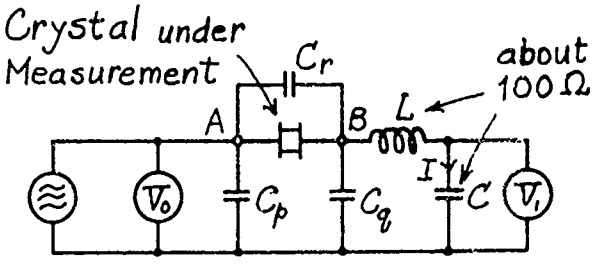


FIGURE 2

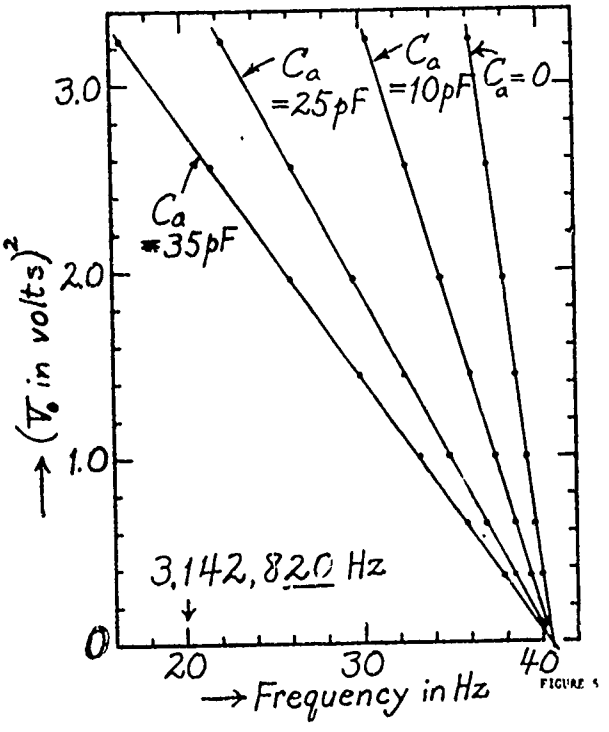


FIGURE 5

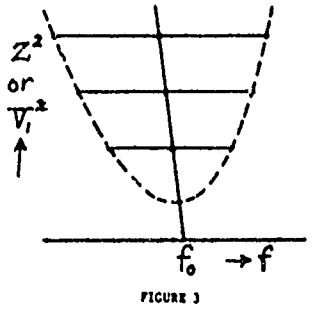


FIGURE 3

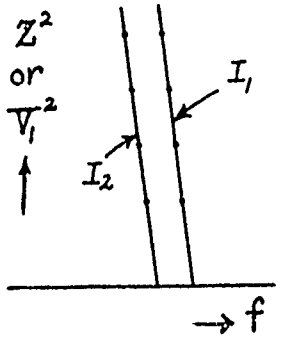


FIGURE 6

A REPORT ON IEC TECHNICAL COMMITTEE TC 49

C. Franx
Crystal Department
N.V.Philips Gloeilampenfabrieken
Eindhoven, Netherlands

Summary

The work of Technical Committee 49 of the International Electrotechnical Commission dealing with devices for frequency control and selection and piezoelectrical materials has resulted in editing Publication nr. 122 entitled Quartz Crystal Units for Oscillators.

Furthermore a standard on filters and a standard on Temperature Controlled Devices are in the printing stage.

Presently TC 49 has a number of objects in discussion:

1 Revision of Publication 122 (Working Group 1)
2 Piezo ceramic resonators and filters (WG 4)
3 Synthetic quartz material (WG 5)
4 Measurements (WG 6)

The paper contains details on the document produced by WG 6 entitled "Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units".

Based on the π network the document specifies a measuring method in the frequency range of 1 - 125 MHz with a frequency accuracy of approximately 1.10^{-7}.

Discussion

This year the 35th IEC General Meeting is scheduled in the U.S.A. namely in Washington. This is a good occasion to focus on the activities of IEC.

In the statutes of IEC the object of the Commission is described as to facilitate the coordination and unification of national electrotechnical standards.

To attain this object the Commission publishes Recommendations which express, as nearly as possible, an international consensus of opinion on the subjects dealt with. Although these Recommendations are not binding upon member organisations, these latter are strongly recommended to follow them when drawing up their national standards so as to unify all national standards.

The work of the Commission covers almost all spheres of electro technology, as well as power, electronics, telecommunications and nuclear energy.

The detailed work is carried out by Technical Committees, each one dealing with a given subject. All National Committees have the right to be represented at meetings of the Technical Committees and take part in the work by correspondence.

Up to December 1969 national organizations of 41 countries are a member of the IEC and 67 Technical Committees have produced approx. 300 publications.

The quartz crystal units are dealt by TC 49, having a scope reading "To prepare international recommendations for piezoelectrical devices for frequency control and selection and for the associated materials".

The work of TC 49 has resulted in a number of publications:

122-1 Standard values and conditions - Test conditions
122-2 Guide to the use of quartz crystal oscillator crystals
122-3 Standard outlines
283 Methods for the measurements of frequency and equivalent resistance of unwanted resonances of filter crystals
302 Standard definitions and methods of measurement for piezo electric vibrators operating over the frequency range up to 30 MHz.

In an advanced state of preparation are recommendations for temperature control devices and for crystal filters.

Presently the work is carried out by four Working Groups:

WG 1 Revision of Publication 122
WG 4 Piezo electric resonators and filters

WG 5 Synthetic quartz material
WG 6 Measuring methods.

The revision of Publication 122 is needed partly because a number of new electrical and environmental tests have been published, partly because it was felt that the component classification system is superfluous.

Working Group 4 has started their activities last year and has produced:

a) Draft recommendation for ceramic resonators (General information and standard values)
b) Draft guide to the use of ceramic resonators
c) Draft recommendation for ceramic filters

Section one . Terms and definitions
Section two : Test conditions

d) Draft guide to the use of ceramic filters.

In Working Group 5 Proposals for a specification for synthetic quartz crystals are in discussion.

Working Group 6 has produced a Draft on the measuring method of Resonance Frequency and Equivalent Series Resistance of quartz crystal units in the frequency range of 1 - 125 MHz with a relative frequency accuracy of approximately $\pm 1.10^{-7}$ and an accuracy on the equivalent series resistance of $\pm$ 2%.

The system is based on the use of the π network (fig 1) and has been described during the 1969 Frequency Control Symposium.

The essential paragraphs of the new standard give a precise definition of resonance frequency and specify the parts of the measuring circuit, particularly the π network, in such a way that all the parts can be calibrated prior to measuring.

As a consequence the standard is compatible with standards based on the use of bridges in the sense that the differences between measurements in π networks and in bridges are very small.

The π network (fig 2) consists of

a. Two resistive attenuators (14.8 dB each) in a metal body
b. Two coaxial connectors (50 Ohms) with sufficiently low contact resistance
c. Two contact plates on which the crystal terminals are pressed
d. A spring loaded plastic block
e. Plastic spacer to determine the location of the reference plane.

The resistive attenuators are asymmetrical and serve to match the 50 Ohm cables to the 12.5 Ohm impedance level between contact plate and body.

The network is calibrated with "zero phase" thin film resistors of the shape of fig 3 by measuring the phase-resistance relationship.

The maximum phase deviation allowed is 0.2^{0} over the resistance range of 0 - 100 Ohms.

Such a small phase deviation can only be achieved if the impedance seen by the calibrating resistor is purely resistive.

For most π networks this impedance has a slightly inductive character. It is possible, however, for frequencies below a certain cut-off, to compensate with a shunt capacitance. This capacitance is created by the dielectric between contact plates and body.

In practice the compensation is performed by measuring the phase-resistance curve at 100 MHz, and varying the thickness of the dielectric.

The dielectric consists of a number of layers Teflon sheets of 17 and of 37 microns thickness. For compensation one only needs to change the number of sheets.

Further requirements concern the reflection coefficients the transducer loss of the π network with a shorting metal blank inserted, the behaviour over the temperature range and the capacitance between contact plates (<1 pF).

Equivalent resistance and power level can be read from the voltmeter of the B channel (fig 1). With a shorting metal blank inserted in the π network the generator is adjusted until the voltmeter of the B channel reads A_0 dB in which

A_0 = nominal crystal drive level - 11.8 dB

A practical reference level is 1 mW corresponding with 0 dB. By adjusting the phase meter to zero reading the initial adjustment procedure is completed both for go-no-go procedure as for individual measurements.

After replacing the shorting blank by the crystal unit the frequency of the generator is adjusted to zero phase reading. The equivalent series resistance R_r is

$$R_r = 25\ (10^{-\frac{A_c - A_o}{20}} - 1)$$

A_c = reading B channel (in dB) with the crystal in the π network.

The power dissipated by the crystal is

$$P = P_{max} \frac{100\ R_r}{(25 + R_r)^2}$$

This function has a very flat maximum for R_r = 25 Ohms (fig 4).

For example: a change of ESR from 9.5 to 65.5 Ohms results in a change of drive level of 20%. In most cases this is acceptable and thus individual drive level adjustment is not necessary.

Reference

C. Franx On precision measurements of frequency and resistance of quartz crystal units Proc. 23rd Annual on Frequency Control Atlantic City, May 1969.

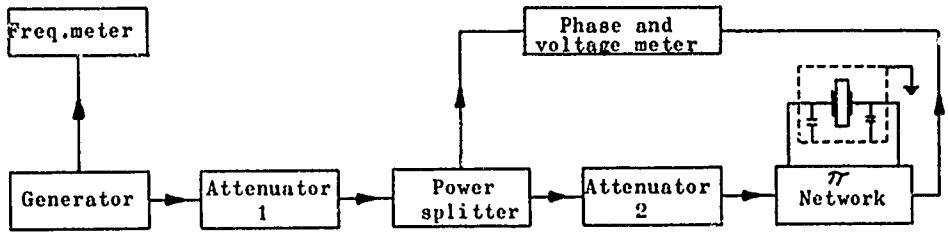

Freq.meter
Phase and voltage meter
Generator
Attenuator 1
Power splitter
Attenuator 2
π Network

FIG. 1

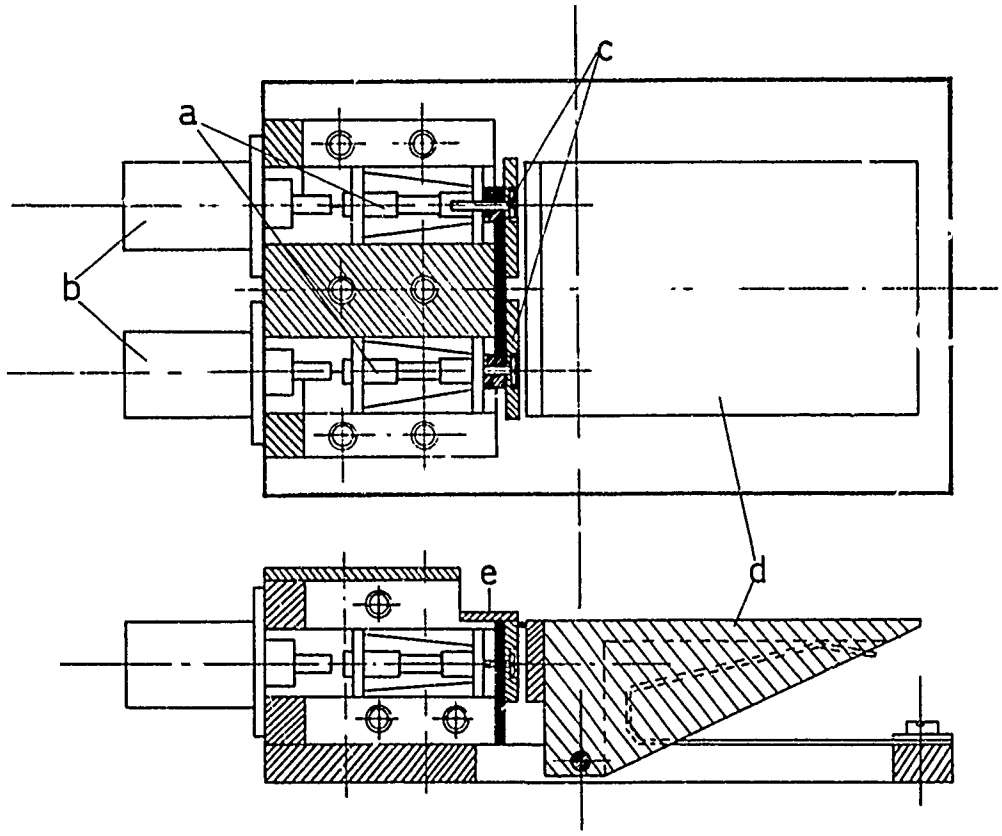
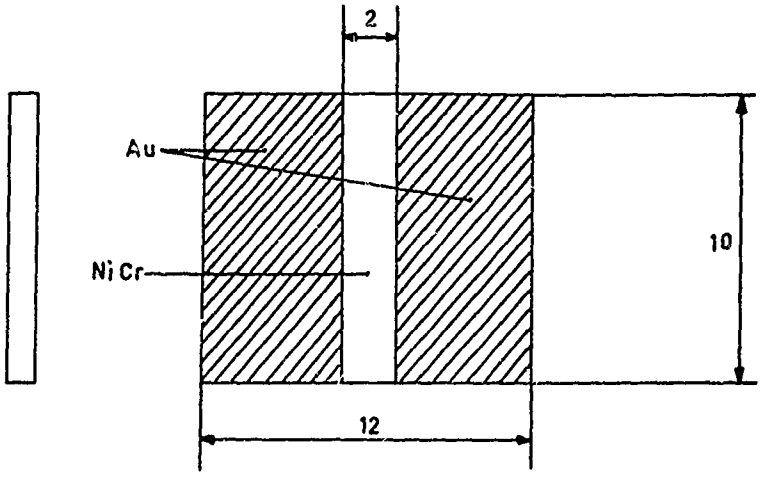

a
b
c
d
e

FIG. 2



Calibrating Resistor

FIG 3

$$\frac{P}{P_{max}} = 100 \frac{R_r}{(25 + R_r)^2}$$

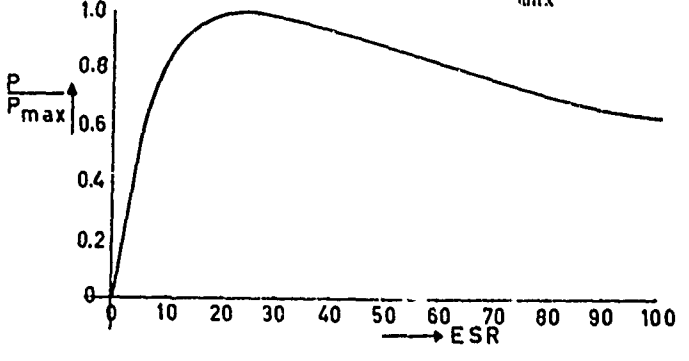


FIG 4

QUARTZ CRYSTAL MEASUREMENTS

Erich Hafner and Arthur Ballato
Electronic Components Laboratory, USAECOM
Fort Monmouth, New Jersey

and
Paul Blomster
Bell Telephone Laboratories
Allentown, Pennsylvania

Summary

The problem of crystal measurements has been considered in the past year by a group of 10 people composed of members from government, crystal users and crystal producers. The group — The IEEE Subcommittee on Crystal Measurements — was charged with producing a draft for an IEEE Standard which could be used as referee in the case of disagreements about the parameter values assigned to a crystal unit. After careful considerations it was concluded that (a) any isolated response of a crystal vibrator is to be characterized by its equivalent circuit parameters and (b) the most suitable techniques for reliable determination of the values of these parameters are based on the use of admittance and/or impedance bridges. A Standards Draft was prepared accordingly. It contains a definition of the parameters to be used for characterizing a crystal unit and it specifies how these parameters should be measured when either an admittance or an impedance bridge is used. An important factor influencing the formulation of the draft has been a round robin experiment participated in by most members of the group.

In the paper available techniques for crystal measurements will be reviewed in broad terms with emphasis on their common aspects. The technique proposed by the IEEE Subcommittee on Crystal Measurements for standard reference purposes will then be discussed in some detail, and the rationale that led to this selection will be presented. The paper will include an analysis of the round robin results.

Introduction

During the past year a group of ten people including crystal users and crystal producers from both government and industry in the USA have been considering the problem of crystal unit measurement. This group — the IEEE subcommittee on crystal unit measurement - was charged with producing a draft which would improve the accuracy and raise the frequency limit covered by IEEE measurement standards from the current limit of approximately 30 MHz to about 300 MHz, and would either supplement or replace the present measurement standard (IEEE Standard No. 177).[1] This paper is an outgrowth of the effort expended by this sub-committee and deals with the considerations which went into the preparation of the standards draft.[2] A round robin experiment (participated in by most members of this subcommittee) was devised to help evaluate the measurement techniques under consideration, including admittance and impedance bridges, π transmission networks and crystal impedance (CI)meters, and has been an important factor influencing the formulation of the draft.

The function of a standard, as interpreted by the sub-committee is to define clearly the parameters that have to be determined, and to specify a particular technique which is most likely to give an accurate measure of the numerical values of these parameters.

The fact that a given measurement technique has been selected for the standard does not mean that no other technique can or should be used in the evaluation of the characteristic quantities. It merely means that anyone using a technique other than that specified in the standard assumes the burden of proof for correlation of his data. And, whenever there is a discrepancy, the values obtained by using the standard will govern.

The function of a measurement standard, therefore, is not to specify the one and only technique to be used for the evaluation of the parameters; rather, the technique specified in the standard should be the most accurate available at the time, so that the standard can serve as a referee in the case of a dispute. And, of course, it should be useful directly where the highest measurement accuracy is required.

Even the best measurement techniques provide accurate results only when proper consideration is given to instrumentation errors. As measurements extend into the upper HF range and beyond, the role of stray circuit elements becomes significant and it becomes difficult to evaluate their effects on the measurements, particularly if the method of measurement requires that circuit elements be placed in series with the crystal unit in order to determine its motional arm parameters. The effect of stray circuit elements on measurements is a problem which must be dealt with no matter what measurement technique is employed. Hence the desire to raise measurement standards to cover the frequency range to 300 MHz has necessitated a look at several measurement techniques with the intent of selecting that technique which either introduces a minimum of stray circuit elements into the measurement network or

provides the ability to balance out these stray elements so that they have a minimum effect on the measurements. It was partly because of the difficulties in assessing the effects of stray elements on measurement accuracy, and the questionable validity of assumptions that would have to be made to do a quantitative analysis of these effects, that the round robin experiment was devised. It was felt that the results of this experiment would provide an insight into the accuracies inherent in the measurement techniques which were under consideration.

This paper proceeds by first considering the equivalent circuit of a quartz crystal unit and the definition of terms.[3] Next the characteristic frequencies of the equivalent circuit are considered along with the effects of added circuit elements on these frequencies. A review of some of the available techniques for crystal unit measurement is included, and finally a statistical analysis of the data taken during the round robin experiment is presented.

Discussion

The electrical behavior of a lightly damped mechanically vibrating system, which is excited piezoelectrically through electrodes, can be represented in the vicinity of a mechanical resonance by the equivalent circuit shown in the upper half of Fig. 1.[4] In this figure the primed elements represent the parameters of the piezoelectric vibrator itself while the unprimed elements represent the parameters of the support structure, an additional feature which becomes important at VHF's. This equivalent circuit assumes (a) that the mode in question is effectively isolated from all other modes of motion, i.e., the admittance of the total network in the frequency range of interest is completely dominated by the equivalent parameters of a single branch, and (b) that the frequency range considered is narrow enough so that the admittance of all the elements in parallel with the mode in question remains essentially constant. The physical construction of the crystal unit is here assumed to possess electrical symmetry; treatment of the case where this is not a valid assumption is straightforward.

When it is true that

$$\frac{R_H}{R_1} << 1, \quad 2R_H R_1 << \left(\frac{1}{\omega C_o}\right)^2$$

$$\frac{R_1}{R_d} << 1, \quad \left(\frac{1}{\omega C_o}\right)^2 << 2R_1 R_d ,$$

then the effects of the resistive elements in the support structure are small, and the equivalent circuit reduces to that in the lower part of Fig. 1 having the following values

$$R_1 = R_1' u^2 \qquad C_o = \frac{1}{au}\left(C_o' + C_c\right)$$

$$L_1 = L_1 u^2 + L_H u$$

$$C_1 = \frac{1}{u^2} C_1' \qquad C_d = C_d' + \frac{1}{a} C_b$$

with

$$a = 1 - \tfrac{1}{2}\omega^2 L_H C_b$$

$$u = 1 - \omega^2 L_H\left(\tfrac{1}{2} C_b + C_o' + C_c\right).$$

The circuit consisting of L_1, C_1, R_1, and C_o has been termed the effective vibrator and except for C_C and L_H it reverts to the actual piezoelectric vibrator at frequencies sufficiently low so that the deviations of a and u from unity become insignificant. The performance of a crystal unit as an electrical circuit element in the frequency range of an isolated mechanical resonance is completely described if the equivalent circuit parameters R_1, L_1, C_1, C_o, and C_d are known.

The accuracy to which the values of these parameters can be determined is in the order of percent, whereas the frequency at which a crystal unit exhibits a certain characteristic value must very often be known to fractions of a part per million. Hence the pair L_1 and C_1 is usually replaced by only one, either L_1 or C_1, together with one of the characteristic frequencies of the crystal unit. For the purposes of the standards draft,[2] f_s, the resonance frequency of the motional arm has been chosen. f_s depends only on the product $L_1 C_1$ and it is independent of the other elements in the equivalent circuit, i.e., C_o, and C_d etc. And it is independent of any additional circuit elements which may be added in series or in parallel with the crystal unit, intentionally or not, when it is placed into a using circuit.

In fact f_s is the only frequency that is truly characteristic of the crystal unit itself, and, when f_s is known, all other so-called characteristic frequencies are readily computed. The sub-committee has decided, therefore, that the quantities which should be used to describe a crystal unit - in the frequency range of a given response - shall be (Fig. 2) R_1, C_1, f_s, C_o and and C_d. Equivalently, either L_1 or Q_o can be used in place of C_1. To complete the definitions part of the standard draft a set of equations, relating the other, conventional, characteristic frequencies to these quantities is needed. And these relations are shown in Fig. 3.

Since the parameters of the equivalent circuit generally depend on the vibrational amplitude, in addition to being functions of temperature and static stress, measurements should be carried out at a specified drive level and care

must be taken to insure that the drive level is held constant throughout the measurement. It will often be found convenient to choose the values low enough such that the drive level effects on the measurements are minimal.

The immittance characteristics of the crystal unit are best examined when the effective vibrator, modified by the addition of a capacitance in parallel with C_o (to account for the support structure) is represented by the network shown in the upper left part of Fig. 3.

The parameters of this equivalent circuit are given as

$$X_1 = \omega_s L_1 \left(\frac{\omega}{\omega_s} - \frac{\omega_s}{\omega}\right)$$

and

$$\bar{X}_o = \frac{1}{\omega (C_o + C_a/2)}$$

when terminal 3 (Fig. 1, lower half) is not connected,

$$\bar{X}_o = \frac{1}{\omega (C_o + C_a)}$$

when terminals 2 and 3 are shorted together. The impedance and admittance plots of this effective vibrator provide an insight into the role of the crystal unit as a circuit element, and are useful in demonstrating the effects of added circuit elements on the characteristic frequencies of the crystal unit. Figure 4 illustrates the fact that the admittance and impedance characteristics of the piezoelectric vibrator describe circles as a function of frequency, with every point on these circles corresponding to a particular immittance. The characteristic frequencies of the effective vibrator are indicated on these diagrams.

Circle representations are valid only if the immittance circle diameter is large compared with the change in reactance due to C_o in the measurement range, i.e., provided $Q^2 >> r$.

Techniques for determining the electrical characteristics of quartz crystal units can be divided into two main categories. In one the crystal unit is used as the principal frequency controlling element in the feedback network of an oscillator while in the other the crystal unit is used as the principal frequency discriminating element in the measurement network of a transmission test circuit. Techniques in this latter category can be further grouped into peak,[1] zero phase,[5,6] or null transmission networks.[4] Many variations of these basic techniques are available. In the most sophisticated case, measuring instruments can be interfaced with computers to provide automatic test sets for determining all the parameters of a crystal unit, and feedback loops can be provided from the output of any of the above mentioned categories of transmission networks to control the frequency of the signal source and thus provide both improved accuracy and increased convenience of measurement.[5-7] Test oscillators are available which can provide precise measurements of crystal unit frequency, but instrumentation errors become difficult to assess when using presently available oscillator techniques in determining motional arm parameters.[8]

Measuring the crystal unit as a passive device in a transmission network, requires the following equipment: (a) a stable frequency generator, (b) a measuring network into which the crystal unit is inserted and (c) a detector.

Two points must be made about these measurements in general: (1) all measurements must be made with respect to a very carefully defined plane of reference,[9] (2) the length of transmission line between the device being measured and this reference plane will alter the characteristics that the device presents to the measurement network unless the electrical length of this line is a multiple of $\lambda/2$ of the applied frequency. * Figure 5 illustrates the network used for determining the frequency of peak transmission. Replacing the peak detector by a phase sensing instrument placed across points A-B provides a measure of the zero phase frequency.

The network used in the null transmission (or bridge) method is, in general, a full lattice with input and output terminations. In practice the circuit configuration for HF and VHF applications is most often that of a Schering bridge, with the unknown terminals in either the series arm (impedance bridge) or parallel arm (admittance bridge). Another null instrument is the transformer ratio-arm bridge, whose external operation is identical to that of an admittance bridge.

The evaluation of the crystal unit equivalent circuit parameters requires in all cases that the impedance which the crystal unit presents at the measuring terminals be determined at a number of frequencies in the response range. These impedance values are then compared to those of the equivalent circuit, which is assumed to represent the crystal unit, and the parameters are determined to obtain a best fit.

In a peak or zero phase transmission network the impedance of the crystal unit has to be extracted from the relationship of magnitude and phase of the output signal to the input signal. This relationship contains the impedances, (resistances and reactances), of all of the circuit elements in the network, each with its associated strays, which must thus be known before the crystal impedance and hence the parameters can be calculated. This process would in general be carried out

*Corrections applicable to bridge measurements in cases where a finite length, other than $\lambda/2$, of transmission line between the measurement and reference planes are given in Ref. 4

with the aid of a computer. Considerable simplification, however, results when all the network elements can be assumed to be purely resistive and when moreover, the stray elements associated with the crystal unit and the series capacitor can be neglected, particularly in the case of the zero phase technique.[5,6]

In a zero transmission (bridge) network, a voltage equal and opposite to that across the branch containing the crystal unit is produced by proper adjustment of reference circuit elements within the bridge. The difference in the settings of these reference elements obtained when first an open (or short) circuit element and then the unknown impedance is placed at the "unknown" terminals of the bridge provides a direct measure of the impedance presented by the device under test. Knowledge of the detailed structure and impedance values of the reference elements is not required in the evaluation of the crystal unit impedance; and the effects of stray circuit elements on the bridge side of the measurement plane are largely eliminated by the initial balance of the bridge. The relationship between the impedance at the unknown terminals and the settings of the bridge controls (reference elements) is established during calibration, a process which is greatly facilitated by the fact that balance for the resistive and reactive components of the unknown are very nearly independent of one another. Although this calibration can in principle be carried out to any degree of accuracy required, practical considerations limit the amount of effort that can be expended in the calibration of individual commercial instruments.

Each approach to the measurement of crystal unit parameters is thus subject to certain kinds of error. Because of the wide range in the character and magnitude of the sources of these errors, any attempt at a theoretical error analysis involved a large number of questionable assumptions, leaving the results subject to endless rounds of criticism. Hence, it was decided to conduct a round robin experiment.

The purpose of the round robin was not only to obtain a measure of the reproducibility of the data attainable on the same crystal, at different geographic locations and using different sets of equipments - each participating member used his own set of instruments; but equally important, the round robin provided each member with an opportunity to obtain first hand experience with the various competing techniques and it thus assured that the final decisions taken concerning the standard draft had been tested in practical application.

Before the round robin began, a two day meeting was held at the Electronic Components Laboratory, USAECOM, Fort Monmouth, where the detailed procedures to be followed with each instrument were experimented with and finally decided on. The crystal units were provided by several of the members. A total of 20 crystals were used ranging in frequency from 1 MHz through 196 MHz. Since it was obviously impractical to use oven control with each crystal, it was decided that the emphasis would be placed not so much on frequency correlation, but mainly on the parameters R_1 and C_1.* Nevertheless the crystals were selected for very low temperature coefficients around room temperature in order to facilitate the C_1 measurement. The latter involves two frequency measurements in the same fixture in reasonably rapid succession.

The instruments used were the Wayne Kerr B801 Admittance Bridge (a transformer ratio-arm instrument) and the General Radio 1606A and 1606B Impedance Bridges in the HF range and the General Radio 1602 or 1609 Admittance Meter for the VHF range. The impedance bridge was necessary in the HF range, even if admittance meters would be preferred, because the range of the Wayne Kerr 3801 is restricted to R_1 values larger than 10Ω.

For the π network measurements two different designs were used, both reported on at the 23rd Frequency Control Symposium. One is that due to Frerking,[5] the other due to Franx,[6] Each design is thought to have a particular advantage over the other. The Frerking π has a minimum number of components and hence, presumably, sources of error — the two voltages V_A and V_B are measured directly across the crystal terminals, whereas the Franx π places great emphasis on impedance matching and isolation of the crystal unit from the rest of the instrumentation. Models of the Frerking π were constructed by each user following detailed instructions. And Mr. Franx has kindly supplied two models of his design; these were used at ECOM. All of the π network measurements were made with HP Vector Voltmeters to establish zero phase, first with the crystal alone, then with the crystal and load capacitor in series to obtain C_1.**

*In independent measurements at ECOM it has been found, using a 5 MHz fundamental mode crystal unit in an oven, that the frequency repeatability is about 1×10^{-10} in several measurements with the same instrument, about 5×10^{-9}, when using different instruments (serial numbers) of the same type. In the VHF range, (110 MHz, 5th overtone, $Q_o = 80000$) the corresponding numbers are 1×10^{-9}, and 2×10^{-8}.

**For the measurements carried out at ECOM, the identical load capacitors were used on the two π designs; this produces a somewhat more favorable weight for the π results than if each had been furnished with a separate load capacitor set. This weight has been neglected in the discussion of results.

The procedure chosen for the admittance bridges was first to establish initial balance with the bridge terminals open, by using the initial balance controls, and — in the VHF case - by adjusting a constant impedance transmission line to $1/2\lambda$. This corresponds to the point B=0 in Fig. 6. Then the crystal was inserted and the bridge was balanced off-resonance by setting the susceptance dial to the value of the $\bar{C}_o$ susceptance; this corresponds to the point (1) in Fig. 6. The conductance dial and frequency were then adjusted to obtain balance at the point (2). The corresponding values are directly $G_{B2} = 1/R_1$ and the frequency is f_s, the motional arm resonance frequency. The bridge then was balanced at two more points on the admittance circle, symmetrically disposed around the susceptance B_o of the parallel capacitance C_o. These points, 3 and 4, are indicated in Fig. 6 to be to the right of the center of the circle. Actually the two points were chosen near the imaginary axis in the case of the Wayne Kerr B801 bridge, and near the top and bottom of the circle in the case of the VHF admittance meter GR1602.

From these data the remaining quantities are determined using, for the B801 bridge

$$C_1 \text{ (in femtofarads)} = \frac{(f_4 - f_3)}{2 f_2 \text{ (in MHz)}} (D_{C3} - D_{C4}) \cdot$$

$$\cdot 10^{-3} \cdot \left[1 + \frac{D_{G3}^2 \times 10^6}{\pi^2 f_2^2 (\text{in MHz}) (D_{C3} - D_{C4})^2}\right];$$

$$\bar{C}_o \text{ (in picofarads)} = D_{C1},$$

where D_{Ck} is the admittance bridge capacitance dial numerical reading at point k without regard to units and D_{Gk} is the like quantity for the conductance dial. (When D_{G3}^2 is very small, the second term in the above relation is negligible). For the GR1602 admittance meter were used

$$C_1 \text{ (in femtofarads)} =$$

$$\frac{D_{G2} \times 10^{-3}}{\pi f_2^2} \cdot \frac{(f_4 - f_3) D_{G3}}{(D_{B3} - D_{B4})};$$

$$\bar{C}_o \text{ (in picofarads)} = \frac{D_{B1} \times 10^{-3}}{2\pi f_2},$$

where D_{Bk} is, similarly, the numerical value of the susceptance arm at point k.

For the measurements with the GR1606 impedance bridge we specified that the bridge should first be initial-balanced with a short circuit substituted for the crystal unit. Then, with the crystal inserted, the resistance dial and generator frequency are manipulated until balance is obtained at the point 6 of Fig. 7. These values are directly the resonance frequency f_r and resistance R_r. Next the reactance dial is changed to a value above and below the setting used for the initial balance, and the bridge is balanced again in each case by manipulating the resistance dials and generator frequency. The resulting Δf and ΔX values are then used to compute C_1 as follows:

$$C_1 \text{ (in femtofarads)} = \frac{2(f_7 - f_6)(f_6 - f_8)}{\pi f_6 (\text{in MHz}) (f_7 - f_8)},$$

when $(D_{X7} - D_{X6}) = (D_{X6} - D_{X8}) = 1000$. D_{Xk} is the 1606 impedance bridge reactance dial setting at point k. It is simply the number scribed on the dial without reference to units. It is related to the actual reactance X_{Bk} at that point (in ohms) by

$$D_{Xk} = X_{Bk} \cdot f \text{ (in MHz)}.$$

For all these measurements the drive level of the crystal units has been kept in the order of 1 μ watt to reduce the frequency-drive level effect.

Although some measurements had been made by some participants using crystal impedance (CI) meters, the quantity of data was small. Because the sample size was not significant, discussion of results obtained with this instrument is omitted; the data appear to vary greatly in quality and a comparison with the passive measurement results appeared inappropriate.

Before a detailed discussion of the round robin results is given, a number of observations should be made. The crystals were first measured at the ECOM laboratories then sent around to the various participants and finally returned to ECOM for a second series of measurements. Although the crystals used were carefully selected in the beginning, some of the units showed obviously poor correlation. On plotting their resistance as a function of time, one sees in Fig. 8 that the poor correlation was not the fault of the equipment used nor was it due to inattentive operators. The units had simply changed with time and very noticeably so. The examples in this figure show a more or less gradual change in resistance. In one other case, however, Fig. 9, there was a sudden jump to a higher value which probably occurred in transit.

Cases like these do occur and when there is a lack of agreement between manufacturer and user data it is not necessarily the fault of the measurements.

In analyzing the data of the round robin, these crystals were excluded. Of a total 20 crystals 5 were eliminated for reasons of this kind. One other set of data was excluded because of an obviously large error in bridge calibration; all resistance values were higher, by about 10 Ω,

from all the other measurements. In this case it was a series of VHF measurements.

Where one or several measurers had made a number of measurements on the same instrument, the values obtained were averaged and the instrument was accorded only one data point to prevent any undue distortion of the results.

In Fig. 10 each point represents the standard deviation in the R_1 values divided by the mean values of R_1 for all measurements performed on a particular crystal at the various locations and/or on instruments of different serial numbers of the same model; in general the averages include from 5 to 12 measurements. (π on top, Z in middle, Y on bottom) In the case of the π network, however, only 4 measurements are represented. Only frequencies up to 20 MHz are shown in the figure. We find that the standard deviations for bridges are comparable to those for the π network measurements.

The results with the π networks are in fact surprisingly good, considering that only one of the two models used was of professional quality, namely that of Franx, whereas the other was homemade, even if made very carefully, and the resistors used for calibration were the standard metal film types. Yet, this is likely to represent the typical application, where, in fact, not only two, but many different types of π networks will probably be used.

On the other hand, the bridges used had not been especially calibrated for the round robin — which was actually a violation of the rules in the standard draft, which specify that the bridges be calibrated. Still, we know now, after analyzing all of the data taken, that not all of the scatter apparent in Fig. 10 is due to errors in bridge calibration. A part at least has been introduced by the less than optimum choice of initial balancing procedure. This has since been corrected in writing the standards draft. In fact, had this improved procedure been used in the round robin, the highest points on Fig. 10 might have been around or even below 3%. Similarly, it is of course quite possible, that improved techniques in making the π network measurements might also result in lower standard deviations.

The resistance data obtained on the admittance meter in the VHF range, Fig. 11, is very much better again. The standard deviations of all values, divided by the mean values are in the order of 1.5 percent.

We consider now the motional capacitance determinations with the various instruments, beginning with the VHF measurements taken on the GR1602 admittance meter.

The normalized standard deviations of the C_1 measurements in the VHF range, shown in Fig. 12 average about 4%, which is to say, there is quite a bit less correlation here than for the corresponding R_1 data. In addition, the points on the figure all cluster about the 4% mark so there is little question about the consistency of the deviation in this frequency range. We suspect that the specific procedure in using the instrument is partly responsible.

The results of the HF measurements of C_1, are shown in Fig. 13, where the normalized standard deviations for the admittance and impedance bridges and for the π networks are plotted (π on top, Z in middle and Y on bottom). Up to 10 MHz the π network points lie between 2 and 3%, but the deviations turn upward and are from 4 to 6% at 20 MHz. For the impedance bridges all deviations save one are less than 2%, averaging about 1.5% over the whole frequency range to 20 MHz, while the admittance bridges are better still. The admittance bridges data points are, but for one point, under 1% over this frequency range, with the average deviation of the order 0.5%.

Conclusions

We conclude from the round robin experiment that the bridge techniques are capable of providing C_1 information with the correlation between different models and serial numbered instruments used at different geographic locations better than 4% throughout the frequency range from 1 MHz to 200 MHz, better than 2% for frequencies from 1 MHz to at least 20 MHz. The correlation in C_1 attained with zero phase π network techniques is about 3% in the range from 1 MHz to 10 MHz, but begins to deteriorate significantly with increasing frequency somewhere between 10 and 20 MHz. The correlation in the R_1 measurements attained by means of bridges is about equal to that attained with zero phase π networks, at least up to 20 MHz.

Admittance and impedance bridges of good quality are available commercially and are required, at the present state of the art, for C_1 measurements of acceptable accuracy, at least for frequencies above 10 MHz. They are equally well suited for the direct measurement of R_1 and f_s. In the round robin experiment synthesizers were used as the generator and, for the most part, communication receivers as the detectors. When phase-sensitive detectors and automatic frequency control on the generator are used, bridge measurements become exceedingly simple, fast and versatile.

In contrast, π networks suitable for zero-phase transmission measurements are not now available commercially. They have to be constructed by the user with the optimum configuration as yet undetermined. While R_r and f_r measurements of acceptable accuracy can be made, at least up to 20 MHz, the designs used in the round robin proved unsuitable for C_1 measurements above 10 MHz.

Further advances in the design of both bridges and π networks can be expected in the future,

however, the relative standing of these two techniques in regards to accuracy and versatility for the determination of resonator parameters is likely to remain as it has been found in the present round robin experiment for some time to come.

The IEEE Subcommittee on Crystal Measurements has decided, therefore, that the next IEEE standard on crystal measurements in the range from 1 MHz to 300 MHz shall be based on bridge techniques, and a standards draft has been prepared accordingly.

Acknowledgement

The work reported on here has been performed by the IEEE Subcommittee on Crystal Measurements of the IEEE Technical Committee on Frequency and Time, Professional Group on Instrumentation and Measurement, with significant contributions from each of its members. The members of the subcommittees and their affiliation are, in addition to the three authors of this paper, Marvin Bernstein, USAECOM, M. E. Frerking, Collins Radio Co., W. H. Horton, Piezo Technology, Inc., Robert Kinsman, CTS Knights, Inc., D. W. Nelson, RCA, J. E. Sherman, Jr. General Electric Co., and Frank Wolf, McCoy Electronic Co.

References

1. "IEEE Standard on Piezoelectric Crystals - Definitions and Methods of Measurement for Piezoelectric Vibrators" - IEEE Standard 177, May 1966.

2. IEEE Subcommittee on Crystal Measurements, "Proposed IEEE Standard Definitions and Methods of Measurement for Piezoelectric Vibrators," March 1970.

3. E. A. Gerber and R. A. Sykes, "Quartz Frequency Standards," Proc. IEEE, 55, 783-791, June 1967.

4. E. Hafner, "The Piezoelectric Crystal Unit - Definitions and Methods of Measurement" - Proc. IEEE, 57, 179-201, Feb. 1969.

5. M. E. Frerking, "Vector Voltmeter Crystal Measurement System," Proc. 23rd Annual Symposium on Frequency Control, U. S. Army Electronics Command, 1969, pp. 93-101.

6. C. Franx, "On Precision Measurements of Frequency and Resistance of Quartz Crystal Units," Proc. 23rd Annual Symposium on Frequency Control, U. S. Army Electronics Command, 1969, pp. 102-110.

7. O. P. Layden, A. Ballato and C. Shibla, "Newly Developed Crystal Measurement Instruments," Proc. 22nd Annual Symposium on Frequency Control, U. S.Army Electronics Command, 1968, pp. 282-297.

8. W.H. Horton & S. B. Boor, "Comparison of Crystal Measurement Equipment," Proc. 19th Annual Symposium on Frequency Control, U. S. Army Electronics Command, 1965, pp. 436-468.

9. L. E. Huntley and R. N. Jones, "Lumped Parameter Impedance Measurements, Proc. IEEE, Vol. 55, pp. 900-911, June 1967.

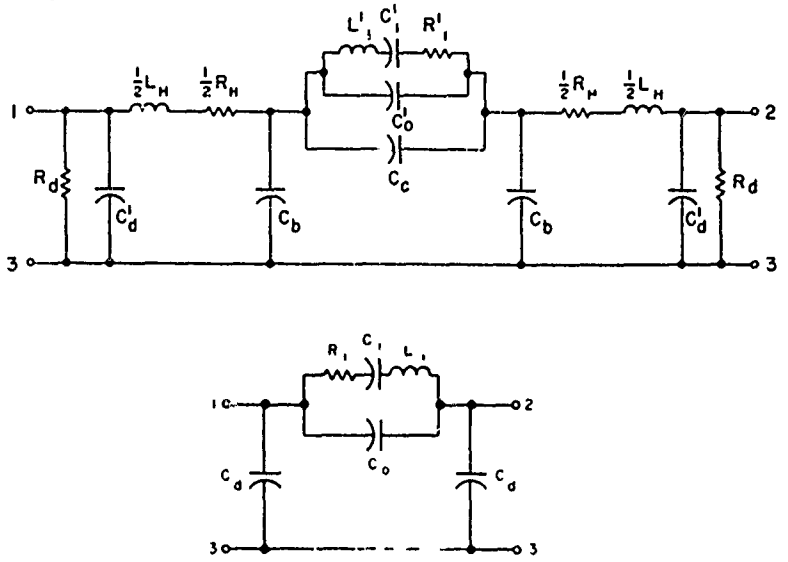
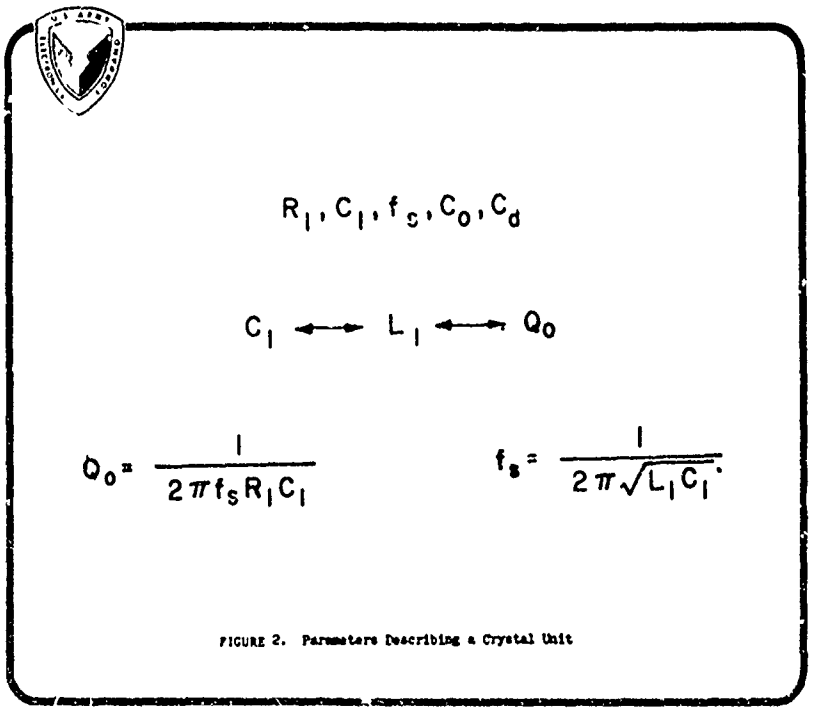


FIGURE 1. Piezoelectric Vibrator Equivalent Circuits



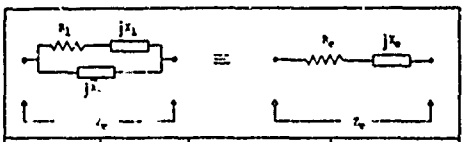
$$R_1, C_1, f_s, C_0, C_d$$

$$C_1 \longleftrightarrow L_1 \longleftrightarrow Q_0$$

$$Q_0 = \frac{1}{2\pi f_s R_1 C_1} \qquad f_s = \frac{1}{2\pi\sqrt{L_1 C_1}}$$

FIGURE 2. Parameters Describing a Crystal Unit

RELATIONS BETWEEN THE CHARACTERISTIC FREQUENCIES OF A PIEZOELECTRIC VIBRATOR AND THE SERIES RESONANCE FREQUENCY OF THE MOTIONAL ARM



Characteristic Frequency		Condition	Exact Relation	Approx Relation
Low Impedance Points	f_s	$X_1(f_s) = 0$	$f_s = \frac{1}{2\pi\sqrt{L_1 C_1}}$	$f_s = \frac{1}{2\pi\sqrt{L_1 C_1}}$
	f_r	$X_0(f_r) = 0$	$\frac{f_r^2}{f_0^2} = \frac{f_s^2}{f_r^2} = 2\delta_1$	$f_r = f_s(1+\delta_1)$
	f_m	$\lvert Z_0(f_m)\rvert = \text{Min}$	$\frac{f_m^2}{f_s^2} = \frac{f_s^2}{f_m^2} = 2\delta_2$	$f_m = f_s(1+\delta_2)$
High Impedance Points	f_p	$R_0(f_p) = \text{Max}$	$\frac{f_p^2}{f_s^2} = \frac{f_s^2}{f_p^2} = \frac{1}{\gamma}$	$f_p = f_s(1+\frac{1}{2\gamma})$
	f_a	$X_0(f_a) = 0$	$\frac{f_a^2}{f_s^2} = \frac{f_s^2}{f_a^2} = \frac{1}{\gamma} + 2\delta_1$	$f_a = f_p(1+\delta_1)$
	f_n	$\lvert Z_0(f_n)\rvert = \text{Max}$	$\frac{f_n^2}{f_s^2} = \frac{f_s^2}{f_n^2} = \frac{1}{\gamma}$	$f_n = f_p(1+$

$\gamma = \frac{C_0}{C_1}$ $\quad \delta_1 = \frac{\gamma}{Q_1(1+\gamma)(2+\gamma)}$ $\quad \delta_2 = \frac{\gamma}{Q_1(1+\gamma)(2+\gamma)}$ $\quad \frac{R_1}{X_1}$

THE "EXACT" RELATIONS CONTAIN NO OTHER APPROXIMATIONS THAN THAT THE IMPEDANCE OF THE VIBRATOR DESCRIBES A PERFECT CIRCLE IN THE IMPEDANCE PLANE

FIGURE 3. Vibrator Characteristic Frequencies

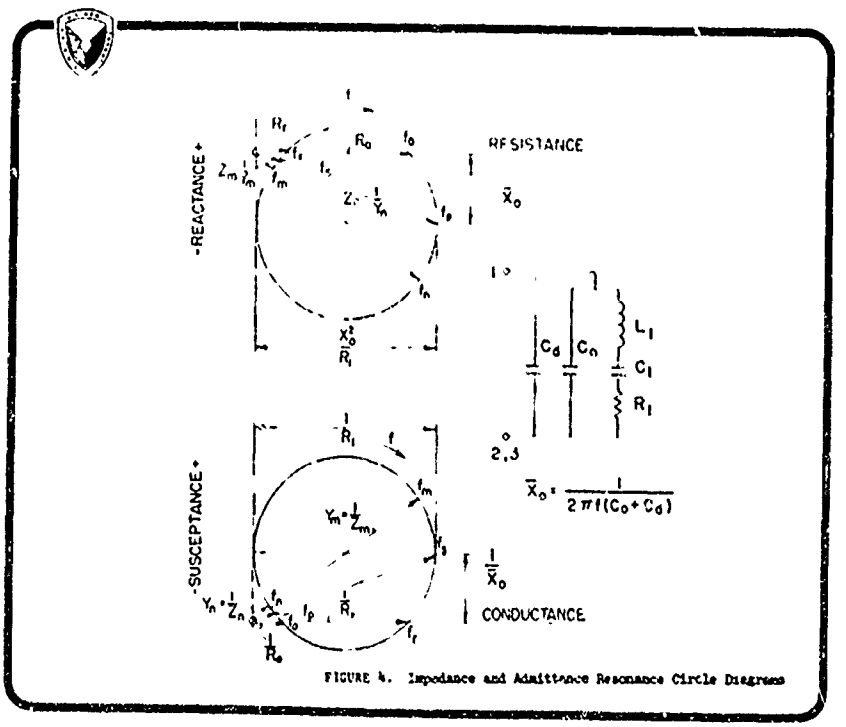


FIGURE 4. Impedance and Admittance Resonance Circle Diagrams

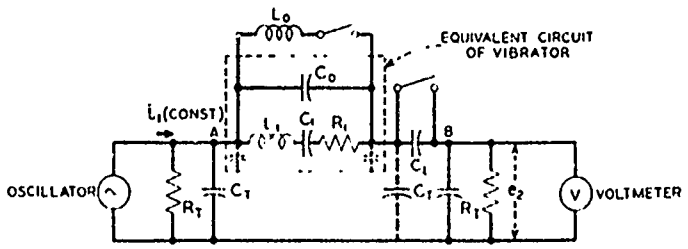


FIGURE 5. Schematic of Transmission Network

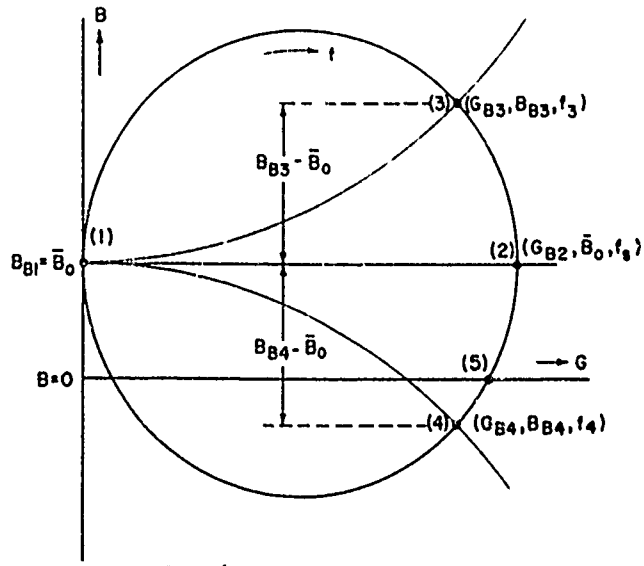


FIGURE 6. Admittance Circle and Points of Measurement

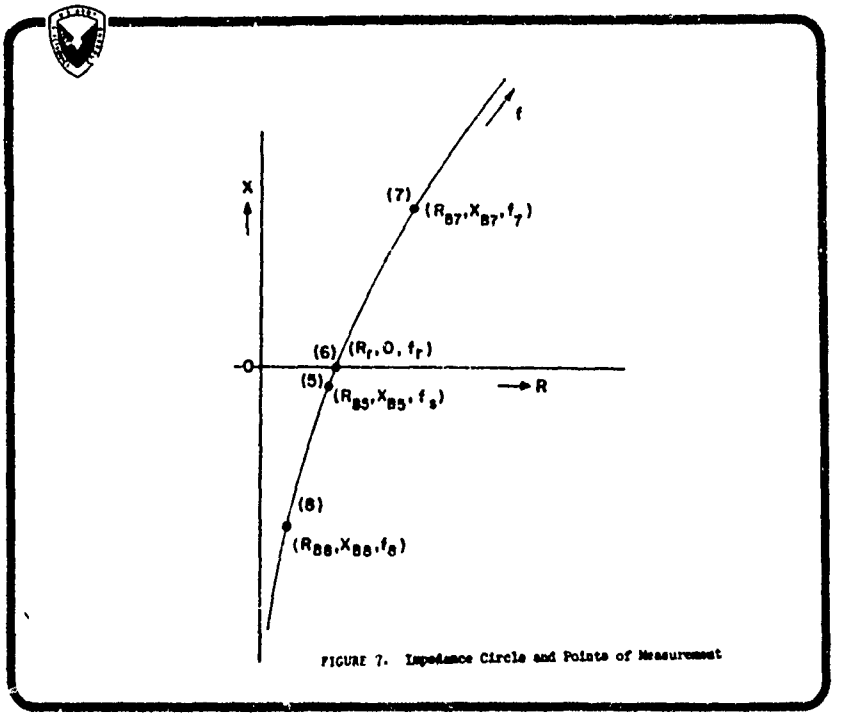


FIGURE 7. Impedance Circle and Points of Measurement

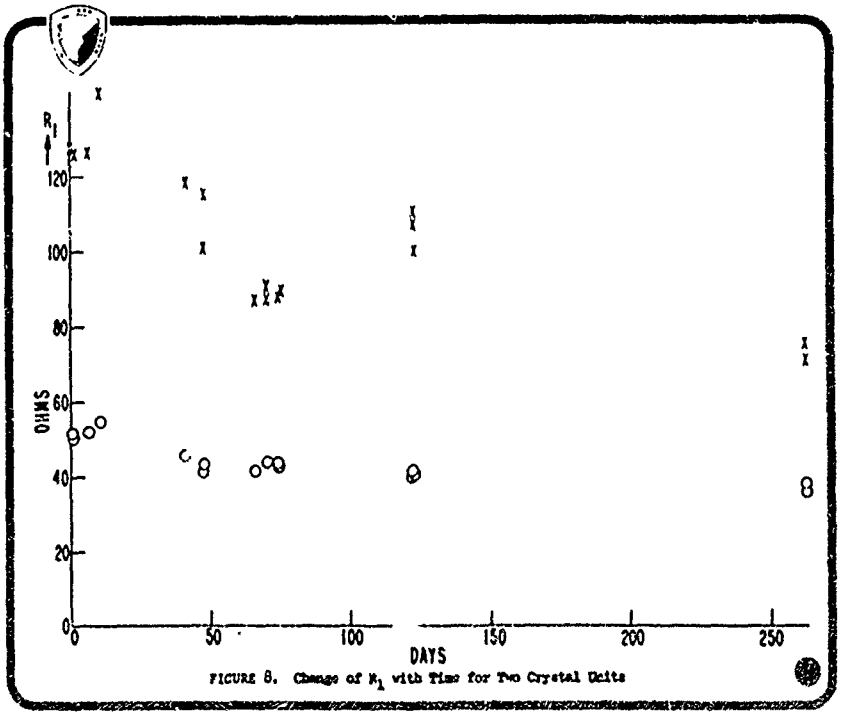


FIGURE 8. Change of R_1 with Time for Two Crystal Units

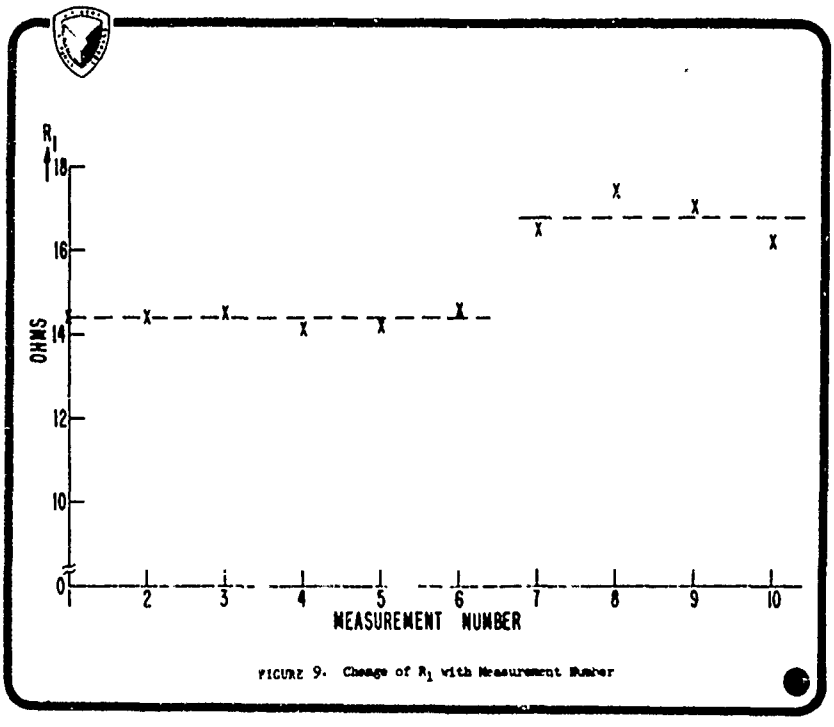


FIGURE 9. Change of R_1 with Measurement Number

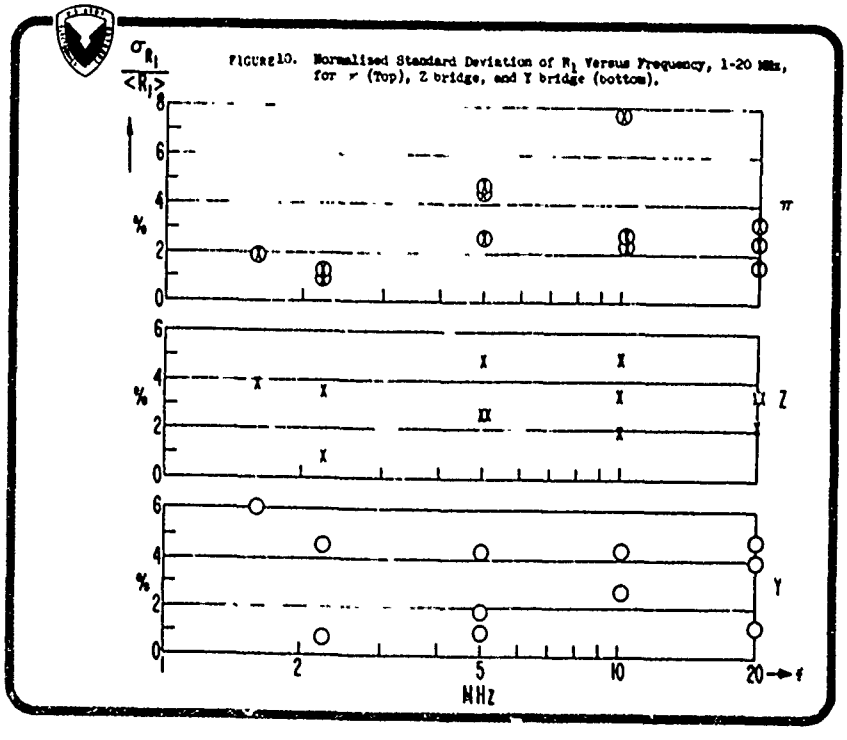


FIGURE 10. Normalized Standard Deviation of R_1 Versus Frequency, 1-20 MHz, for π (Top), Z bridge, and Y bridge (bottom).

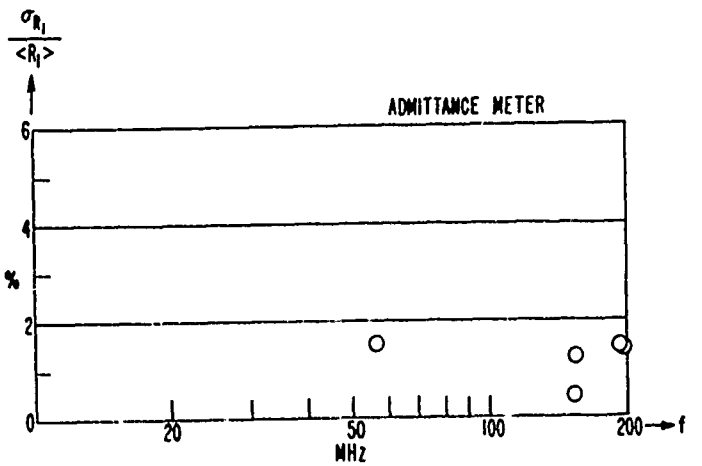


FIGURE 11. Normalised Standard Deviation of R_1 Versus Frequency, 50-200 MHz, for Y Bridge.

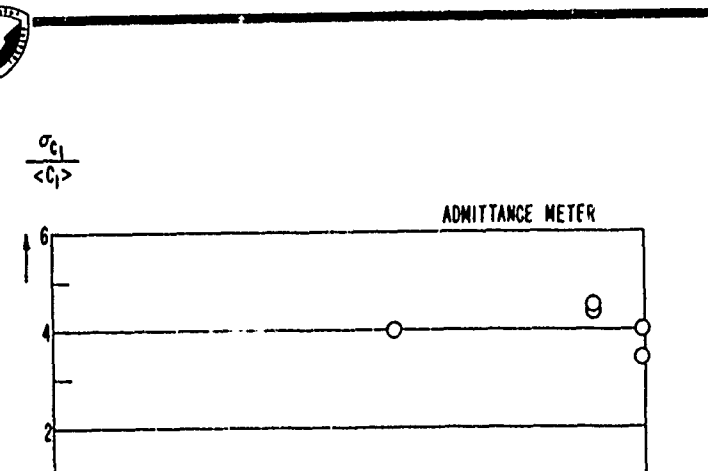


FIGURE 12. Normalised Standard Deviation of C_1 Versus Frequency, 50-200 MHz, for Y Bridge.

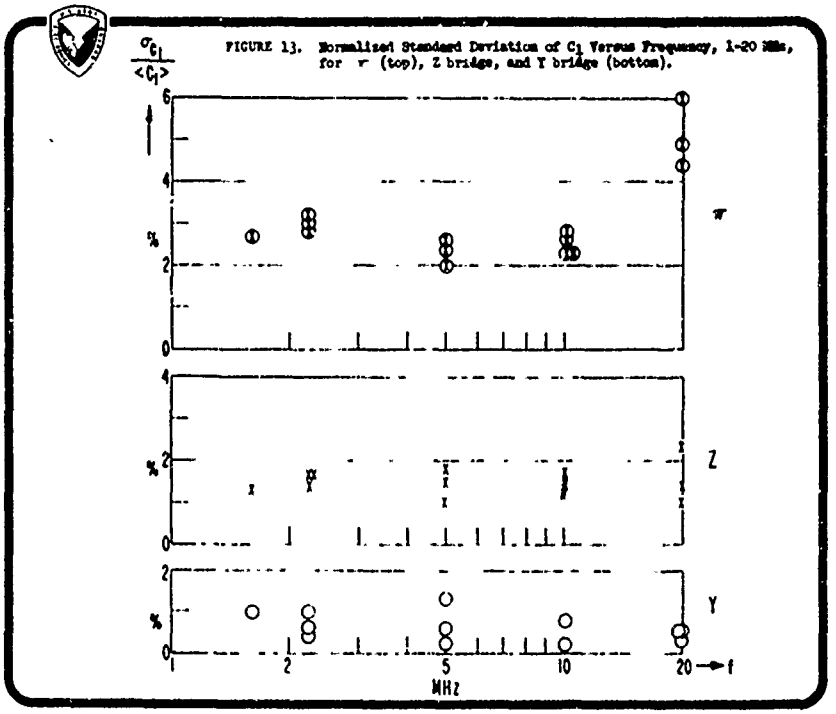


FIGURE 13. Normalized Standard Deviation of C_1 Versus Frequency, 1-20 MHz, for W (top), Z bridge, and Y bridge (bottom).

TEMPERATURE COMPENSATED CRYSTAL OSCILLATORS

Philip C. Duckett, Ralph J. Peduto, George V. Chizak
Bulova Electronics Division
Woodside, New York

Summary

This paper discusses the important aspects necessary for the successful design and manufacture of temperature compensated crystal oscillators with a frequency stability of $\pm 2/10^7$ over a temperature range of -40°C to +75°C and with an input of 50 milliwatts. Four 5 MHz TCXO's with the above specifications, were successfully designed and manufactured at Bulova Electronics Division for the U. S. Army Electronics Command at Fort Monmouth for potential use in suppressed carrier single side band communications equipment, such as the Army AN/PRC-70 tactical transceiver.

Included in this paper is a description of the crystal, oscillator, amplifiers, voltage regulator, compensation network, computer programs, temperature compensation techniques, test equipment and test results.

Introduction

The ever-growing popularity of suppressed carrier single side band equipment in the HF and VHF bands by both commercial and military communication users has brought about the extensive use of temperature compensated crystal oscillators as the signal source in the frequency synthesizers utilized by this type equipment. The frequency stability required of TCXO's used in the Army AN/PRC-70 tactical transceiver is $\pm 2PP10^7$ over the temperature range of -40°C to +75°C. This paper discusses some of the important aspects necessary to successfully and consistently manufacture TCXO's meeting this requirement. In particular the design and manufacturing techniques utilized to fabricate four Bulova Model TCXO-2 temperature compensated crystal oscillators as supplied to the U. S. Army Electronics Command, Solid State and Frequency Division, Fort Monmouth, New Jersey on Purchase Order Number DAAB07-70-M-1356 is presented. These oscillators (Figure 1) were designed and manufactured to meet the following requirements:

- Frequency: 5 MHz
- Frequency Stability:
  - Temperature: $\pm 2PP10^7$ maximum over the temperature range of -40°C to +75°C.
  - Voltage: $\pm 5PP10^8$ maximum for a supply variation of ±5%.
  - Load: $\pm 1PP10^8$ for a load variation of ±20°, ±10%.
- Frequency Aging: $\pm 1PP10^8$ per week at 60°C.
- Output Voltage: 125 millivolts RMS minimum into 50 ohms.
- Input Power: 50 milliwatts maximum at 12 VDC
- Size: 2" x 2" x 1".

Most of the processes to be discussed in this paper were initially developed at Bulova for the design and manufacture of 600 TCXO's, with a stability of $\pm 5PP10^7$ from -40°C to +75°C. These oscillators were used in the AN/GRR-17 radio receiver set which is a Marine Corps. general purpose single side band receiver.

Crystal

The crystals used for the oscillators were designed and manufactured at Bulova. The crystals are 5.0 MHz, fundamental, plano-convex plates with aluminum electrodes, mounted in evacuated HC-27/U holders. They are ordered to the detail requirements of MIL-C-3098D, type CR-90/U with two exceptions.

The first exception is frequency deviation vs. temperature. The MIL specification requires that the frequency

deviation between -40°C and the upper turning point shall be within ±5PPM. In keeping with our computer optimizing procedure to be discussed in a later section, the frequency deviation versus temperature of the crystal is specified in terms of its upper and lower turning points and inflection temperature.

The second exception deals with spurious mode coupling effects. The MIL specification states a maximum allowable change in crystal resistance of less than 1 ohm per 3°C over the temperature range of -55°C to +90°C. We prefer to indicate a maximum allowable frequency discontinuity over the temperature range. It is specified in this manner because a discontinuity in terms of frequency rather than resistance is more meaningful and directly related to the frequency versus temperature stability required of the oscillator.

Circuit

The degree of complexity of a given packaged oscillator circuit is dictated by the required performance characteristics. In the case of these TCXO's the specifications indicate the need for suitable buffering between the oscillator and load and good regulation of the supply voltage both to the oscillator and compensation network. Thus in addition to an oscillator stage and power amplifier, the units incorporate a buffer amplifier between the oscillator and power amplifier and a series regulator.

The crystal oscillator stage is a Pierce oscillator with an isolating resistor in the feedback path. This circuit is ideal for this application since both the crystal drive level and signal voltage across the voltage-variable capacitance diode can be readily adjusted. Typically, the crystal drive level is set to 10 microwatts. The buffer amplifier and the power amplifier are of the common emitter configuration and are bias stabilized with a stability factor of less than 10. The series regulator incorporates a pre-regulator which serves as a constant current source for the reference type zener diode. The entire regulator circuit is temperature compensated. The voltage regulator has a regulation factor of 0.5%, a voltage variation over the temperature range of 50 millivolts and a power consumption of 25 milliwatts.

Compensation Network

The method employed to compensate is best described as continuous compensation, where a network, consisting of resistors and thermistors continuously change the bias voltage of the voltage-variable capacitance diode as a function of temperature. The voltage-variable capacitance diode is in series with the crystal, therefore the varying voltage to the voltage-variable capacitance diode causes a resultant change in its capacitance which in turn changes the crystal frequency by the amount required for compensation. The networks used are classified by the number of points that are used to synthesize the network. Thus, what is termed a five point network, compensates or clamps five points along the crystals temperature coefficient curve, while a six point network compensates or clamps six points. The particular network selected for a given application depends on the required frequency stability and temperature range. Through the use of a computer analysis the optimum performance of the various networks can be computed. Thus, upon reviewing the specific requirements of this contract, the type of network to be used was readily determined.

The computer runs indicated that the requirement could be met with a five or six point compensation network. Using a five point compensation network reduced the allowable spread of crystal upper turning point from the usual ±5°C tolerance to ±2.5°C, when limiting the theoretical error to be no worse than 2/3 of the actual error allowed in the performance requirements. The theoretical error for a five point compensation network with a crystal upper turning point spread of ±2.5° is a minimum of $1.7PP10^{7}$ peak to peak and as shown in Figure 2, a maximum of $3.0PP10^{7}$ peak to peak. In order to utilize readily available crystals with an upper turning point spread of ±5°C the decision was made to use a six point compensation network. Over the temperature range of -40°C to +75°C, the best theoretical error of the six point network was found to be a minimum of $1.0PP10^{7}$ peak to peak and as shown in Figure 3, a maximum of $2.0PP10^{7}$ peak to peak for a crystal upper turning point spread of ±5°. The six points used for compensation were, the crystal's upper and lower turning points, the crystal's inflection point, the minimum and maximum operating temperatures and -20°C.

Physically the network is made up of 1.0% metal film resistors with temperature coefficients of ±50PPM and wafer type thermistors with various negative temperature coefficients measured to a 1.0% tolerance.

Computer Programs

The synthesis of the compensation network and the techniques used to optimize the various parameters affecting the final performance of the TCXO's are carried out through the use of a computer time-sharing service. All of our programs are written in Advanced Basic.

During the initial stages of design the first computer program is used to determine the optimum value of the many variables affecting the compensation. These variables, the crystal upper turning point, the compensation network excitation voltage, the VVCD value, the nominal bias voltage applied to the voltage-variable capacitance diode, and the six temperatures at which the compensation network is to be synthesized, are optimized by the computer ranging these variables over prescribed limits and selecting values that yield a minimum error.

A second program performs the actual synthesis of the compensation network. This program utilizes an iterative process to calculate the values of the elements in the network. The program accepts data in the form of the correction voltages and the resistance ratios of the thermistors that are required at each of the six temperatures selected for compensation. The iteration process is then carried out by the computer until the difference between the element value of one iteration is less than .01% of the one preceeding it for all element values. Typically, convergence of all elements of the network is reached after 100 iterations. This entire, rather complex computation is carried out by the computer in less than 5 seconds.

Having synthesized the network, a third computer program then calculates the frequency error that will result at each of the six temperatures for a 1.0% variation of each network element. The data obtained in this program is made use of in the correction process discussed in the next section.

Temperature Compensation Techniques

Compensation of the crystal and oscillator begins with a temperature run of the entire oscillator without the compensation network. This temperature run is performed to verify that the spurious modes of the crystal, if they do exist, are less than the maximum as specified on the crystal specification, and that the upper and lower turning points fall within the range specified. Following this, a second temperature run is performed to determine the voltages required at the six temperatures to compensate the temperature coefficient of both the crystal and circuit. Having determined the voltages, the network required to compensate each oscillator is synthesized using the second computer program. The network element values calculated by the computer are recorded and the required resistors and thermistors are selected from stock and wired into the oscillator. The compensated oscillator is again run over the specified operating temperature range. The frequency error at each measured temperature point after the first run is typically 1-to-10PP10^7. To correct for these errors, use is made of the third computer program. This program indicates the effect on frequency at the six clamp points for a 1.0% variation of each network element. In addition, the effect of each network element for temperatures on either side of the clamp point can be readily extrapolated. Thus, to correct a given network for errors that are in excess of the final requirement, the element or elements having the most effect are changed in the amount and direction indicated from the computer run. For these four oscillators, the average number of corrections required to meet the ±2PP10^7 allowable frequency deviation over the temperature range was eight.

All temperature runs, except the second, are performed automatically in five degree steps using the equipment described in the next section.

Test Equipment

To facilitate the determination of the necessary parameters to temperature compensate a crystal oscillator with a minimum of errors and expenditure of manpower, special semi-automatic test stations were designed and fabricated. These test stations permit an operator to

perform a variety of tests on a large quantity of oscillators in various stages of completion rapidly and reliably with a minimum of attention.

Each test station consists of a Bulova-designed programmable Controller, a Temperature Test Chamber, an Electronic Counter, a Digital Recorder, a Frequency Error Expander and a Digital Voltmeter, as shown in Figure 4.

The precision Temperature Test Chamber utilizes a central low pressure CO_2 system and a proportional type temperature control. These features result in a control accuracy of better than 0.15°C. The Digital Recorders used were modified to print out the unit number, the ambient temperature and the frequency of the oscillator under test upon command from the Controller.

The use of a Frequency Error Expander in the system permits taking frequency data in $PP10^9$ for a one-second time interval. The Controller automatically coordinates and controls the functions of all these instruments, the temperature range over which the tests are performed, the time interval between temperatures, the number of units to be tested and the printing of data.

The Controller permits obtaining and recording most of the data on the oscillators under test automatically and unattended. Override switches on the Controller permit the operator to stop a given test or perform a test manually when necessary.

Test Results

The frequency stability vs. temperature for the four TCXO's is shown in Figures 5 through 8.

The first unit has a stability of +1.4 and -1.9 $PP10^7$
The second unit has a stability of +2.0 and -1.6 $PP10^7$
The third unit has a stability of +1.9 and -2.0 $PP10^7$
The fourth unit has a stability of +2.0 and -2.0 $PP10^7$

Summarizing this data, we can see that all four TCXO's met the frequency stability vs. temperature requirements. The test data for frequency vs. load and supply variation is shown in Figure 9. The data indicates that all the units meet specifications with the exception of one unit which failed the change in load requirement.

Conclusion

We feel that TCXO's with the same requirements as those which have been the subject of this paper can be manufactured in production quantities, reliably and consistently, because these four TCXO's used computer aided design methods, crystals, circuits and automatic test equipment similar to those that have been used at Bulova over the last several years to manufacture several thousand TCXO's with less stringent requirements.

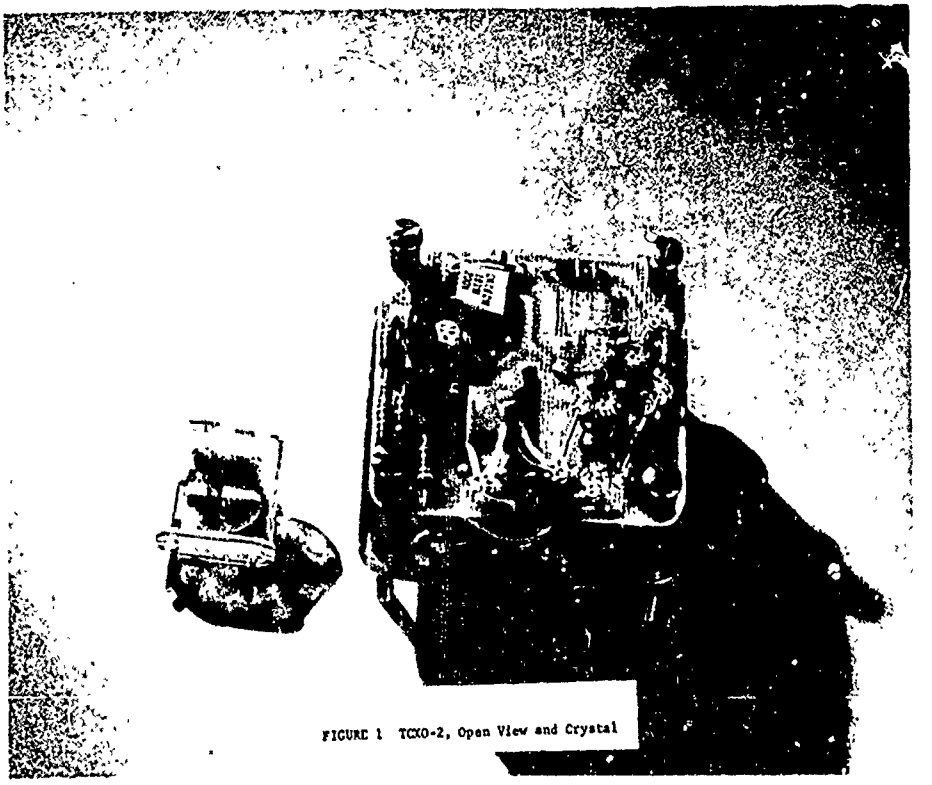


FIGURE 1 TCXO-2, Open View and Crystal

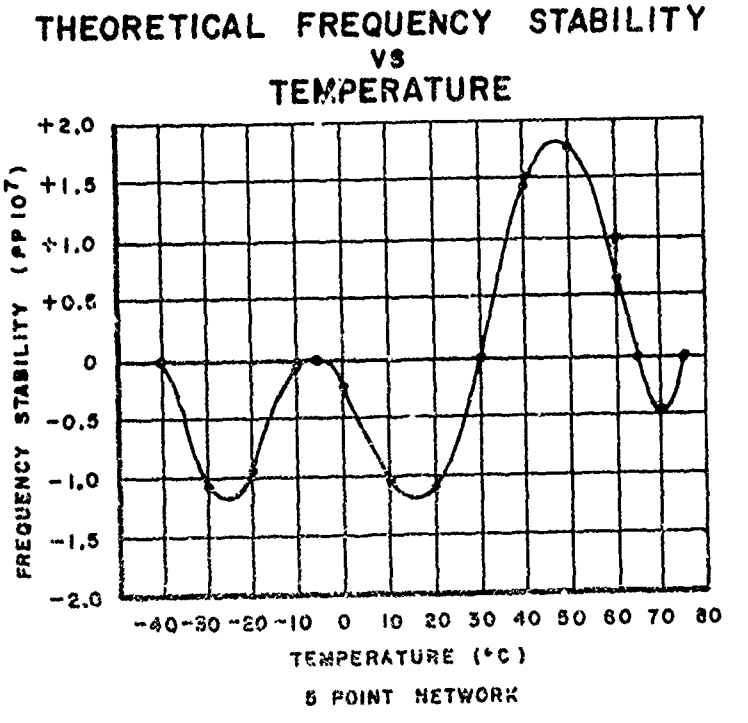


FIGURE 2

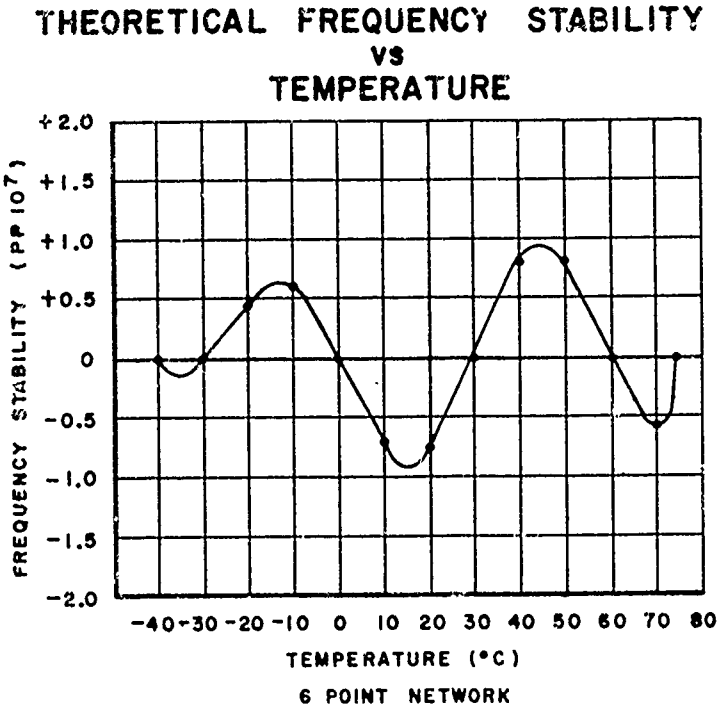


FIGURE 3

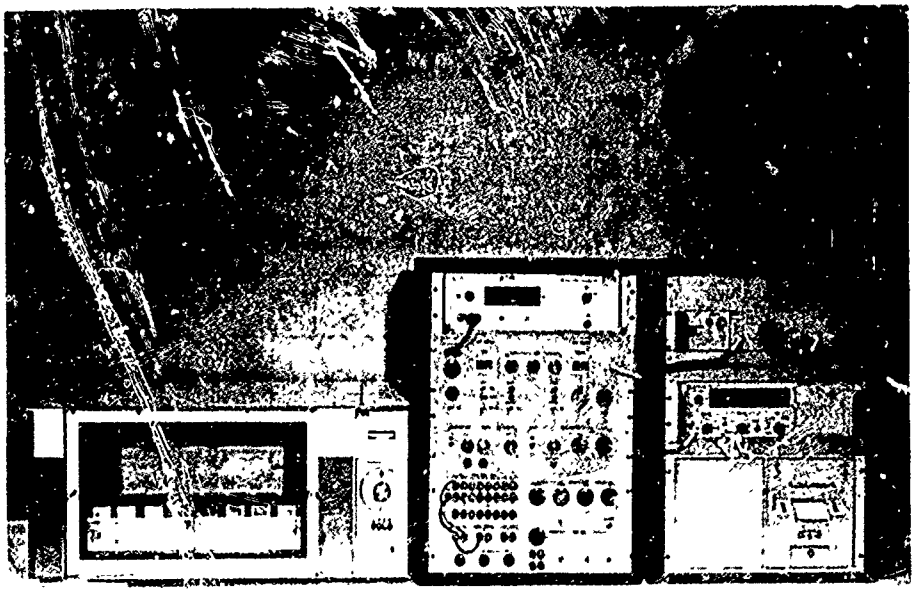


FIGURE 4 TCXO Production Test Station

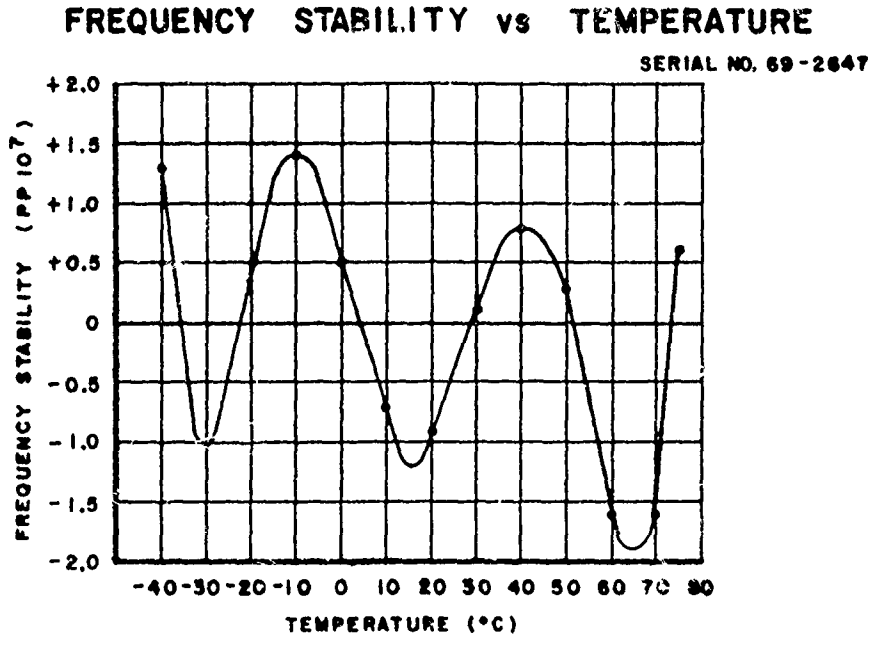


FIGURE 5

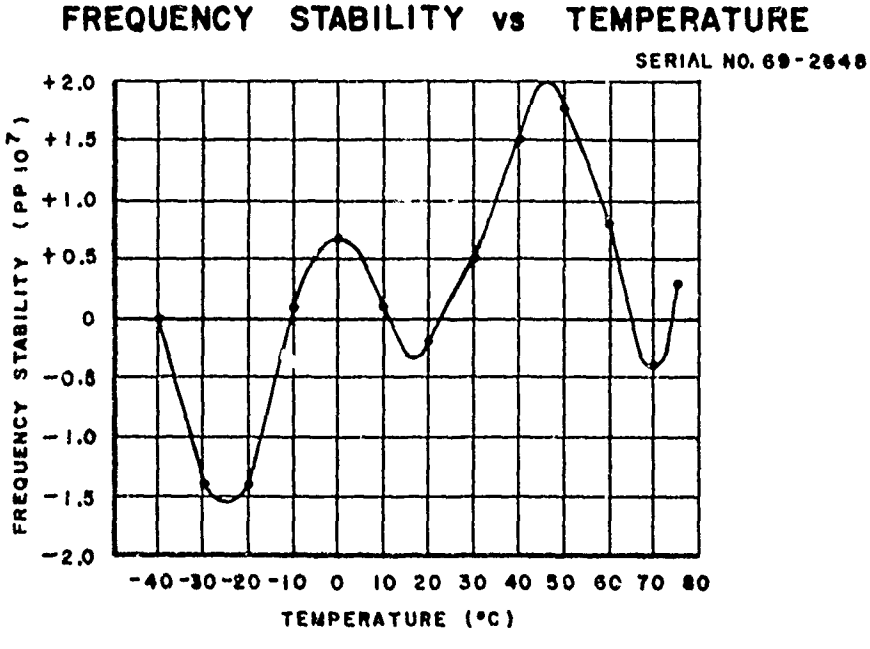


FIGURE 6

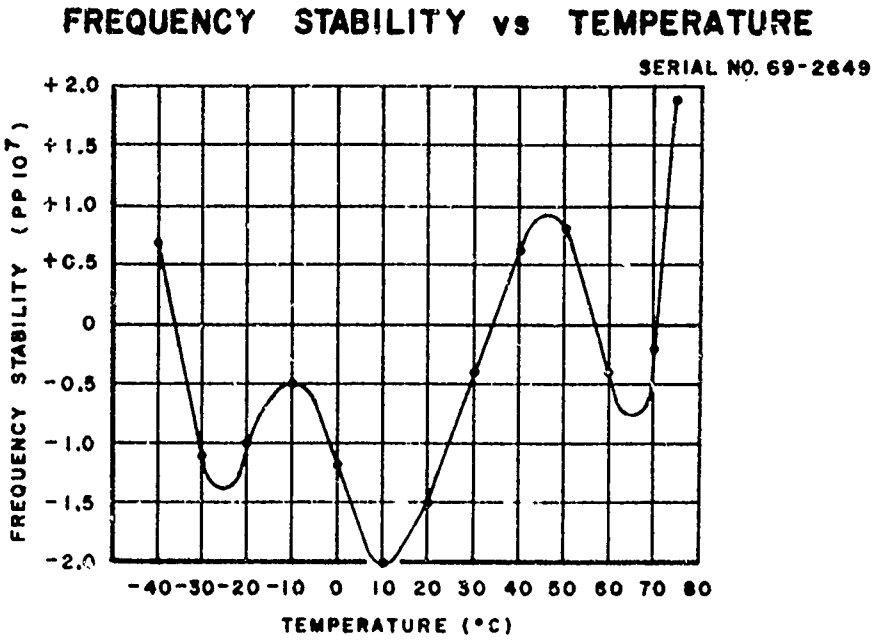


FIGURE 7

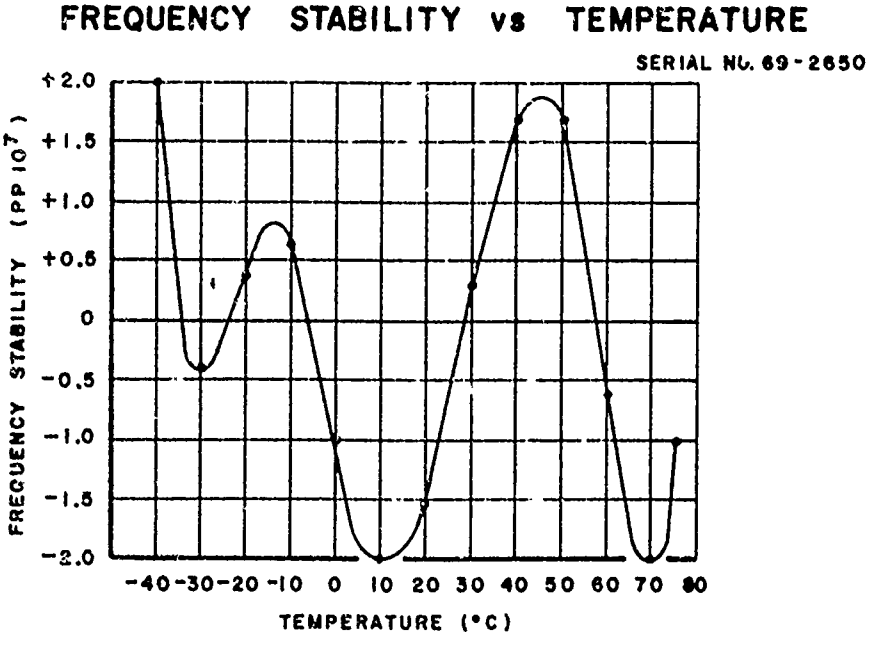


FIGURE 8

FREQUENCY STABILITY

1. VERSUS LOAD VARIATION

LOAD	FREQUENCY STABILITY (PP 10^8)			
50 Ω	69-2647	69-2648	69-2649	69-2650
+10 %,+20°	- 0.9	- 1.9	-0.4	- 1.0
-10 %,-20°	+ 0.8	+2.3	+0.6	+ 0.9

2. VERSUS SUPPLY VARIATION

VOLTAGE (VOLTS)	FREQUENCY STABILITY (PP 10^8)			
	69-2647	69-2648	69-2649	69-2650
11.4	- 2.6	- 3.6	- 2.0	- 3.0
12.6	+ 1.7	+ 3.0	+ 1.3	+2.5

FIGURE 9

A NEW APPROACH TO A HIGH STABILITY

TEMPERATURE COMPENSATED CRYSTAL OSCILLATOR

S. Scholowski
Electronic Components Laboratory, USAECOM
Fort Monmouth, New Jersey

Summary

This paper describes a compensation approach based upon the introduction of controlled nonlinearity into the TCXO's frequency-voltage tuning characteristic by using a pair of varactor diodes and a fixed bias source. The result is considerable simplification of the temperature compensation process for a TCXO using a fundamental AT-cut crystal unit. The simplification comes about by transformation of portion of the classic cubic voltage correction function to a monotonic function which then can be approximated to a high degree of accuracy using one or two "3 point" network segments employing a given thermistor in each. A stability of $\pm 5 \times 10^{-8}$ (5 to 80°C) was obtained in one case where the required monotonic compensation function was generated in two segments.

In another example, a stability of $\pm 1 \times 10^{-8}$ was easily obtained by compensating over a ±10°C range about the upper turnover temperature.

Introduction

Recent investigations in the Electronic Components Laboratory have concerned new methods and techniques for low power quartz reference oscillators. These studies have been directed at meeting the more stringent performance requirements imposed by future tactical SSB equipments employing digital speech transmission. The prime requirement for the reference oscillator in HF equipment of this type is a frequency-temperature stability of $\pm 5 \times 10^{-8}$ over a temperature range of -40°C to 80°C.[1] It is also desired that for maximum system effectiveness the unique characteristics of an ovenless quartz reference oscillator be retained, that is, instantaneous operation and low power consumption.

This paper describes a temperature compensation technique which can provide the means for meeting these requirements over a limited, but substantial, part of the operating temperature range. Unlike previously published compensation techniques,[2-6] this one is based upon an introduction of controlled nonlinearity into the varactor network such as to alter the oscillator's frequency-voltage tuning characteristic. The result is a transformation of a portion of the classic cubic correction function into a monotonic function which can then be accurately approximated using simple thermistor-resistor networks.

Classic TCXO Tuning Method

The method of varactor tuning in a classic high quality TCXO is shown in Fig. 1. Typically, the compensation control voltage, $E_c(T)$, is applied to a single varactor positioned within the transformation network of a modified Pierce circuit. Inasmuch as the values of feedback capacitors C_a and C_b are relatively large, C_v, the varactor capacitance dominates the total effective capacitance seen by the crystal. The crystal is typically a fundamental AT-cut unit rated at a load capacitance of 30 pF. The oscillator has a relatively linear frequency-voltage tuning response essentially described by the equations shown where C_o and r are crystal constants and K, N and ϕ are varactor constants. Linearity, measured as deviation from a straight line, is typically less than 5% for a control voltage range of ±1 volt.

The compensation-tuning process in the classic circuit is shown graphically in Fig. 2. The first quadrant contains the required frequency-temperature compensating characteristic which is the inverse of the oscillator's uncompensated F-T characteristic. A projection of points by way of the frequency-voltage tuning characteristic in quadrant II and the unity transfer voltage line of quadrant III establishes the characteristic curve of quadrant IV. It is this voltage-temperature curve of quadrant IV which is required to be generated by a thermistor-resistor network. The shape of the curve closely resembles the shape of the oscillator's cubic frequency-temperature curve. It is only slightly modified by the minor nonlinearity of the tuning characteristic. The complexity of synthesizing a suitable thermistor-resistor network stems from the high degree of nonlinearity of this function. To reduce the complexity and hence simplify the compensation process it is required that the voltage-temperature curve of quadrant IV be monotonic. To achieve this requirement we could do one of the following:

1. In quadrant I, increase the angle of cut of the crystal such that turnover points are placed outside the operating temperature range. However, an increased crystal angle would invariably increase the slope leaving no net advantage. For instance, a shift of the upper turning point from 80°C to 85°C would result in a slope increase by a factor of 7 and an impractical required compensation factor greater than 500 to 1.

2. We could reshape the frequency-tuning characteristic of quadrant II with nonlinear variable reactance devices[7] such that it has a nonlinearity approximating the curve in quadrant I. The result can be either an increasing or decreasing monotonic function in quadrant IV.

3. We could introduce a function generator stage between the thermistor and varactor networks which would provide a cubic voltage transfer characteristic in quadrant III, again, approximating the curve of quadrant I. A compensation technique based upon this concept was presented at the 1968 Symposium by Mr. Vovelle of the CSF Company in France.[5]

It is the second method, based upon shaping the tuning characteristic of quadrant II, which is the subject of this paper.

Nonlinear TCXO Tuning Method

The circuit of Fig. 3 includes a modification to the classic circuit which provides it with a capability for nonlinear shaping of the frequency-voltage tuning characteristic. The modification simply involves the addition of a second varactor V_2 and a fixed reverse bias source E_b, which is chosen to be always greater than the variable control voltage E_c. This parallel arrangement adds the two separate varactor capacitances to form a total capacitance given by the equation shown. It can be easily seen, that an increase in control voltage E_c results in a decrease in the capacitance of varactor V_1 and simultaneously, an increase in the capacitance of varactor V_2. The capacitance of varactor V_2 increases because its reverse bias, the difference between E_b and E_c, decreases.

The net capacitance variation of the varactor pair is shown in Fig. 4. Here we note a resemblance, at least in dissymmetry, to the shape of the oscillator's cubic frequency-temperature response in the vicinity of the upper turning point temperature. The left handed skewness of the curve results when a shaping factor S, defined as the ratio of varactor constants K2 and K1, is chosen to have a value less than one. In practice, S is determined from the ratio of the rated nominal values of the two varactors. With S equal to one, a perfectly symmetrical curve approximating a parabola, would result. Whereas a shaping factor S greater than one would produce a curve skewed in the opposite direction. In addition, parameters K1 and E_b describe the the vertex of the curve as shown by the equations for the turning point coordinates. E_b, essentially controls the turnover voltage and K1 controls the level of capacitance. Therefore, three degrees of freedom are available for shaping, positioning and adjusting sensitivity of the capacitance voltage response such as to result in an optimum frequency tuning characteristic.

Figure 5 shows the relation between characteristics for a TCXO using nonlinear varactor tuning. Again, quadrant I contains the required compensating characteristic. The slope of 10 ppm represents an optimum design center established for maximizing compensated stability and temperature range. Quadrant II contains the frequency-voltage tuning characteristic resulting from use of the varactor pair network previously described. Again, by graphic projection we obtain, in the fourth quadrant, the voltage temperature characteristic which needs to be generated by a thermistor network. Unlike the required voltage-temperature curve of the classic tuning method this curve is monotonic in the temperature range of 0 to 80°C. It can be very simply and accurately generated as will be described later.

Conditions which must be met in order to assure a suitable monotonic V-T function are the following:

1. Coefficients of the tuning characteristic must be approximately matched to coefficients of the compensation characteristic in the vicinity of turnover. That is, a' in Fig. 5 should be made equal to a and b' equal b at points equidistant from their respective turnovers which are indicated by $\pm\Delta T$ and $\pm\Delta E$. This matching of coefficients simply involves the judicious selection of values for the three shaping parameters previously discussed. The accuracy of match is not critical and it is required to be done only once for a design center curve. This means that the two selected varactors and bias voltage become standard for a given design.

2. The turning point frequencies of the compensating and tuning characteristics must be made coincident. This is a critical requirement. It is accomplished by an initial calibration of the oscillator after it has been stabilized at its turning point temperature, and with the varactor control voltage set to the turnover voltage of the tuning characteristic.

Having satisfied these conditions, measurement data necessary to compute the thermistor-resistor network values is then taken at three temperatures, two at extreme end points of the operating range and the third, at an intermediate point which is one of maximum changing slope.

The Three Point Thermistor Network

The thermistor-resistor network and method of solution used, result in generating exact voltages at the three temperature points. The network is shown in Fig. 6. Only values of resistors R_1, R_2 and R_3 are computed. The thermistor is standard with a nominal 10% tolerance. Standardizing the thermistor in the circuit rather than keeping it a "solved for" quantity, is an absolute necessity since it is not

practicable to duplicate a computed value for a thermistor to the required degree of precision. Its value, however, must be measured with an accuracy better than 0.1% at the three compensating temperatures. Also, it must be mounted in its final position which permits tight thermal coupling with the crystal unit. Thermistor resistance measurement is conducted simultaneously with measurement of the required compensation voltage and the regulated, but temperature dependent, supply voltage E_S. This is to assure that all components are at the same stabilized temperature at the time data is taken. Voltages need to be measured to an accuracy better than ±1 mV. A total of nine measurements, three of resistance and six of voltage, are required prior to computation.

The procedure for computation of the resistor values is based upon a simultaneous solution of the output voltage equation at the three measurement temperatures.[8] Having satisfied constraints imposed by available thermistor sensitivity and supply voltage, an exact solution of resistance is obtained. Then selection and/or adjustment of actual resistance to an accuracy better than 0.1% of the computed value assures that the voltage-temperature function will pass through the measurement points. Further, this exact method lends itself to simple solution using a desk calculator.

The schematic of the TCXO minus the voltage regulator and isolation amplifiers is shown in Fig. . Crystal drive in this self limiting Pierce circuit is maintained below 10 μW primarily as a result of operating the transistor at a low collector current of approximately 500 μ amps and equalizing the two feedback capacitors at 390 pF each. A selected shaping factor of 0.56 is obtained from use of the two varactors; a 1N5140 valued at 10 pF and a 1N5143 valued at 18 pF. Required fixed bias for the 10 pF varactor is obtained from a voltage divider. The thermistor-resistor network is shown in the lower section of Fig. 7. The thermistor is of high Beta, type M material. Typical computed resistance values are given. It will be shown later how the voltage-temperature compensation function can be generated in two segments to obtain greater accuracy. Two segment compensation essentially requires the use of an additional 3 point network similar to the one shown.

Typical stability obtained from this simple straightforward design is shown in Fig. 8. The compensated stability of $\pm 1.25 \times 10^{-7}$ over a temperature range of 5 to 80°C represents an improvement factor of 40 to 1. Frequency data was taken at 5°C increments using the "sequential" measurement method.[9] Results also show that if the crystal slope can be kept to a range of 8 to 12 ppm, a stability better than $\pm 1.5 \times 10^{-7}$ (5 to 80°C) can be consistently obtained. Overall improvement factors approaching 100 to 1 can be obtained by extending the approach to compensation using two segments.

Two Segment Network Compensation

Figure 9 shows how two 3 point networks have been combined to result in an effective 5 point compensation. The dual transistor performs as a voltage comparator or gate which routes the higher of the two base voltages to the varactor network. In this case, the left network, whose output is E_{c1}, provides the control over the range of lower temperatures, shown by segment 1. The right network takes over control at approximately 60°C and provides the upper segment control voltage. The turning point temperature was chosen for crossover, rather than one midway, since it is the point of minimum sensitivity to voltage variations. Therefore, interaction between the outputs of the two networks at voltage crossover, which can be several millivolts, has negligible effect on frequency.

Typical frequency-temperature performance obtained using two 3 point network segments is shown in Fig. 10. As expected, the effective 5 point compensation results in a curve having at least 5 zero crossings. In this case there are seven, indicating that the natural nonlinearities are such that they actually aid the compensation in regions between data points. The achieved stability, $\pm 5 \times 10^{-8}$ over +5 to 80°C represents meeting required frequency performance for approximately 2/3 of the specified operating temperature range. Further, extension of this compensation technique or combination of the technique with others could eventually lead to meeting the requirement over the complete temperature range.

Narrow Range Compensation

It should be noted that this compensation technique, using a single 3 point network, was applied to considerably smaller temperature ranges in the vicinity of the crystal oscillator's upper turning point temperature A typical result obtained is shown in Fig. 11. A compensation to $\pm 1 \times 10^{-8}$ was obtained over a temperature range of ±10°C on either side of a turnover when applied to a crystal with an upper turning point of approximately 65°C. The maximum slope within the compensated temperature range was found to be less than 2×10^{-9}/°C. A potential application of this form of high stability compensation at the turning point exists in those cases where close oven temperature control tolerance must be traded for other performance.

Conclusions

It has been shown that a high degree of precision in temperature compensation can result from addition of another varactor and bias source into the tuning circuit of a TCXO. Varactor and thermistor nonlinearities uniquely combine to realize a close approximation of the actual to the required compensating characteristic over a substantial temperature range. In addition, thermistor measurement and an exact network solution guarantee that the monotonic voltage

temperature correction function will pass through the measured data points.

Application of this technique to a 5 MHz TCXO has successfully realized a $\pm 5 \times 10^{-8}$ stability over a range of +5 to 80°C. Achievement of this stability over the full temperature range of -40 to +80°C now appears more promising. As a spin off, the feasibility of compensating to $\pm 1 \times 10^{-8}$ for a ±10°C temperature range about an upper turnover temperature has been indicated. It is believed that this latter result opens the way for new tradeoff possibilities in oven controlled crystal oscillators.

Acknowledgements

The author expresses his gratitude to Dr. Gunter K. Guttwein for many helpful suggestions and to Messrs. Stanley Emmons and Walter Bowerman who rendered valuable experimental assistance during various phases of the work.

References

1. O. P. Layden, "Frequency Control for Tactical Net SSB Equipment," 1969 Proc. 23rd Annual Frequency Control Symposium, pp. 14-17.

2. G. R. Hykes and D. E. Newell, "A Temperature Compensated Frequency Standard," 1961 Proc. 15th Annual Frequency Control Symposium, pp. 297-303.

3. D. E. Newell and R. H. Pangert, "Frequency-Temperature Compensation Techniques for Quartz Crystal Oscillators," The Bendix Corp. Davenport, Iowa, & USAECOM, Ft.Mon., N.J., Final Report, Contract DA36-039 SC-90782, July 1, 1962 to June 30, 1963, Interim Report, 6th Quarterly, Contract DA36-039 AMC-02282(E), July 1, 1963 to Dec. 31, 1964.

4. C. D. Dominguez and I. E. Hardt, "Frequency Temperature Compensation Techniques for Quartz Crystal Oscillators," The Bendix Corp., Davenport, Iowa, & USAECOM, Ft. Mon., N. J. Final Report, Contract DA36-039 AMC-02282(E), March 1, 1965 to May 1, 1967.

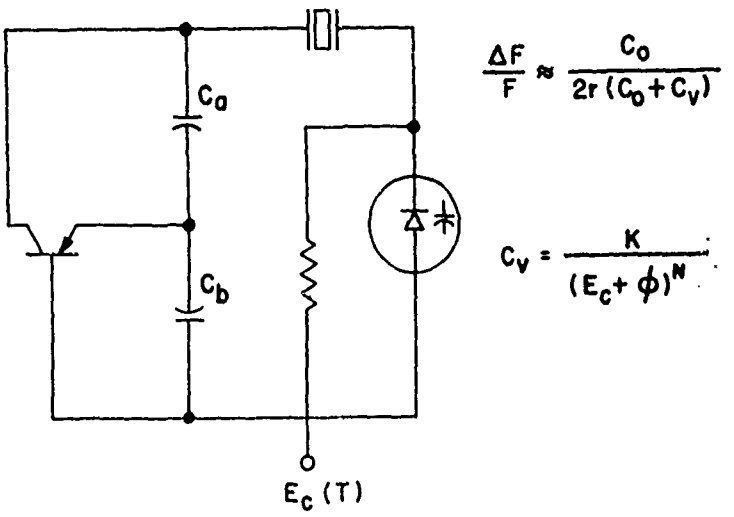
5. P. C. Vovelle, "Recent Improvements to TCXO's," 1968 Proc. 22nd Annual Frequency Control Symposium, pp. 311-324.

6. D. E. Newell and H. Hinnah, "A Report on Segmented Compensation and Special TCXO's," 1969 Proc. 23rd Annual Frequency Control Symposium, pp. 187-191.

7. S. Schodowski, "Capacitance Network for Crystal Compensation," U. S. Patent 3,708,007, 21 September 1965.

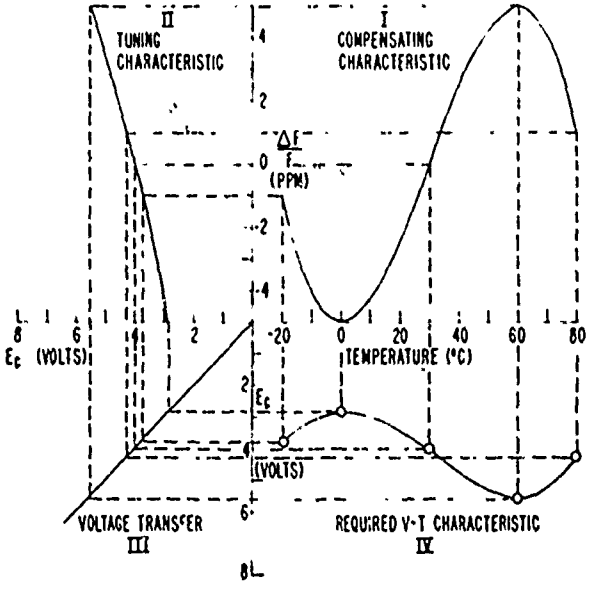
8. V. Weisenbloom, "Ease Resistor-Thermistor Network Design," Electronic Design, Part 1, Vol. 13, 21 June 1967, pp. 60-64 and Part 2, Vol. 14, 5 July 1967, pp. 74-77.

9. S. Schodowski, "Temperature Performance Measurement Methods for Temperature Compensated Quartz Oscillators," USAECOM, Ft. Mon., N.J., Oct. 1967, TR-2896.



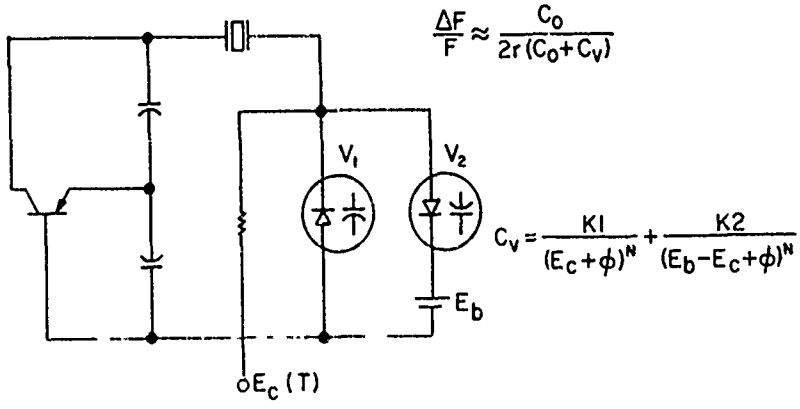
CLASSIC TCXO TUNING METHOD

FIGURE 1



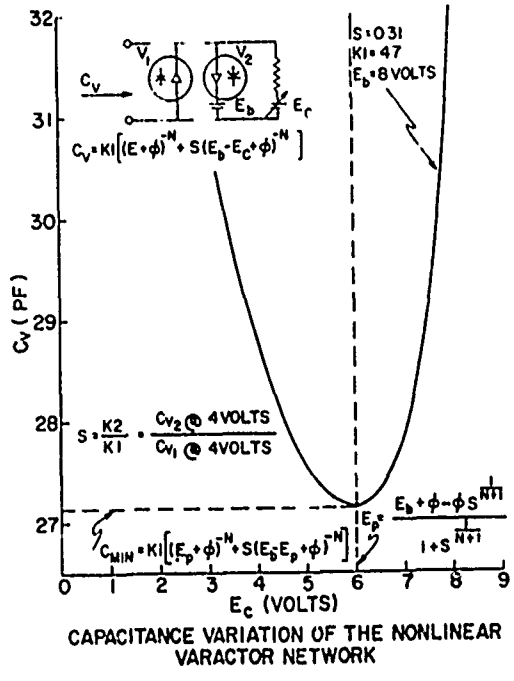
RELATION BETWEEN TCXO CHARACTERISTICS (CLASSIC VARACTOR TUNING)

FIGURE 2



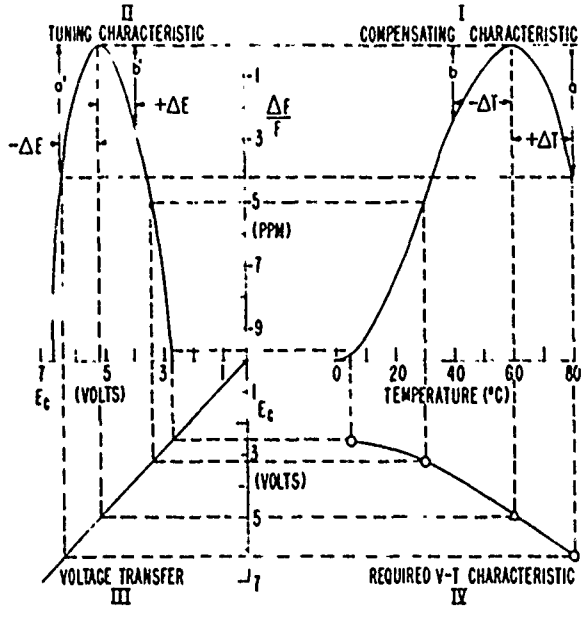
NONLINEAR TUNING METHOD

FIGURE 3



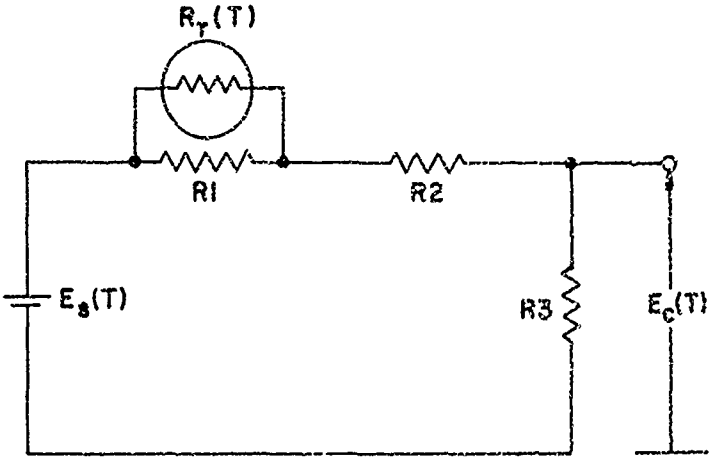
CAPACITANCE VARIATION OF THE NONLINEAR VARACTOR NETWORK

FIGURE 4



RELATION BETWEEN TCXO CHARACTERISTICS (NONLINEAR VARACTOR TUNING)

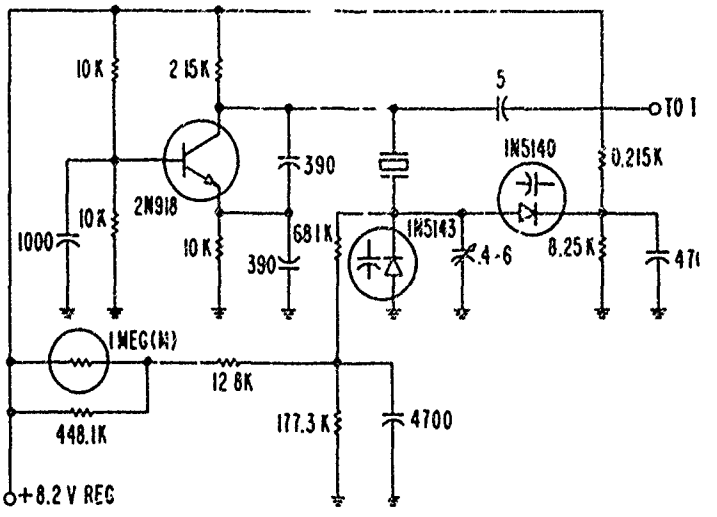
FIGURE 5



$$E_c(T) = \frac{R3\left[R1+R_T(T)\right]E_s(T)}{(R2+R3)\left[R1+R_T(T)\right]+\ R1\,R_T(T)}$$

THREE POINT THERMISTOR-RESISTOR NETWORK

FIGURE 6



5 MHZ HIGH STABILITY TCXO SCHEMATIC

FIGURE 7

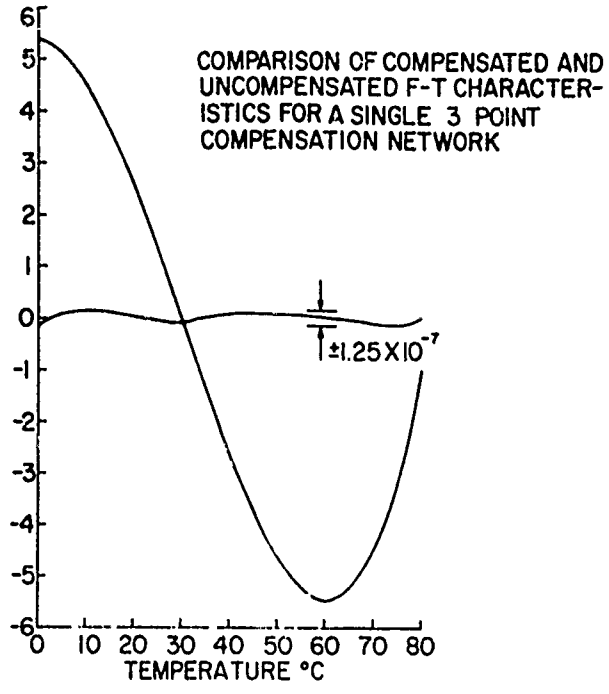
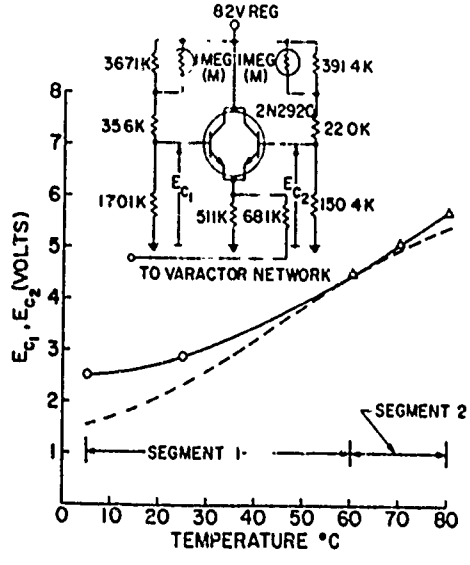
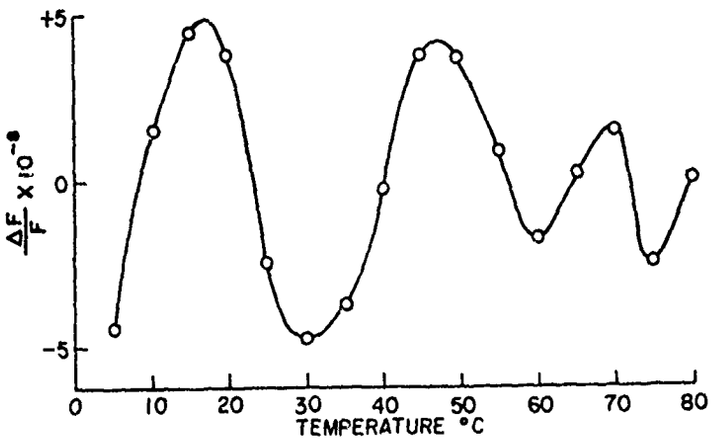


FIGURE 8



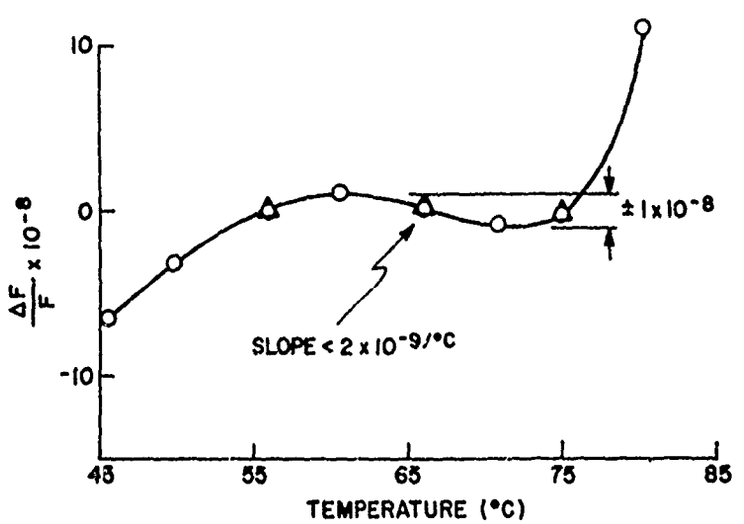
TWO SEGMENT NETWORK COMPENSATION

FIGURE 9



COMPENSATED F-T CHARACTERISTIC
FOR A TWO SEGMENT NETWORK

FIGURE 10



F-T PERFORMANCE FOR NARROW RANGE
COMPENSATION (UTP±10°C)

FIGURE 11

An Ultra Low Noise Direct Frequency Synthesizer

Donald G. Meyer
John Fluke Mfg. Co., Inc.
Seattle, Washington

SUMMARY

The purpose of this paper is to provide a complete technical discussion of the advanced techniques of direct synthesis which have made possible the ultra low noise and spurious performance of the Model 645A Synthesizer. The residual phase noise of the instrument at 50 MHz has a 1 Hz intercept point $\mathcal{L}$ (1Hz) = -100db/Hz. $\mathcal{L}$ (fm) then decreases at 12db/decade to an eventual floor of -137db/Hz. Non-harmonic spurious products are better than 100db down. Frequency programming switching time is < 20 us. The paper has been divided into three sections, which are as follows:

Section I - The basic definitions of phase noise performance are discussed, along with the techniques used to measure them. An attempt has been made to show the relationship between various measures of phase noise performance used by different authors, and to clearly define the most meaningful measures of synthesizer noise performance. Measurement techniques are also discussed.

Section II - The most fundamental problems in building a high performance direct synthesizer are uncorrelated phase noise, correlated phase noise, and spurious signal generation. The basic natures of these problems are examined, and the techniques used to optimize overall performance in the presence of these effects are discussed.

Section III - This section shows how basic design concepts were applied to optimize the overall performance of the unit. The design of each section of the synthesizer is discussed, and the performance of each section is related to the overall performance of the instrument.

Section I - Definition and Measurement of Phase Noise Performance

There has been a noticeable diversity in the nomenclature used by various authors to describe the phase noise performance of signal sources. None of the representations given are incorrect in the least, but it is evident that there are several possible ways of defining phase noise performance so long as the definition of the measure is clearly stated. In the interest of a better common understanding, the relationships between several of the measures which have been used will be defined. Also, the "physical interpretation" of some of these measures will be discussed, so that we can understand more fully their value as measures of synthesizer performance.

Importance of Phase Noise in Direct Synthesis

The concept of phase noise is particularly important in the study of direct synthesis, since the nature of the signal processing performed in a synthesizer usually results in an output signal which has AM noise between 10 and 30 db lower than PM noise over a very large range of offset frequencies from the carrier. Examples of processes which have a tendency to suppress the AM noise on a signal are limiting, division, multiplication, and driving the local oscillator port of a mixer. Since the basic operations which must be performed electronically in a synthesizer are frequency addition, subtraction, multiplication, and division, we can begin to see why the primary output noise contaminant of a synthesizer is PM noise rather than AM.

As an example of this type of behavior, let us consider a single sideband spurious signal which appears at the input of a symmetrical hard limiter. Let us assume that the spurious signal is located 1 KHz below the carrier input to the limiter, and is 50 db below the carrier level in amplitude. Now it is well known that such a signal is composed of one half AM and one half PM. Mathematically such a signal may be expressed as the sum of the carrier plus four modulation sidebands, each of which is 6 db below the level of the original spurious. Two of the sidebands are 56 db down and are phased to represent AM, and two are 56 db down and are phased to represent PM. It is interesting to note that this mathematical representation lacks physical meaning at this point, since observation of the input

signal with the spectrum analyzer will reveal energy is present only at the carrier frequency and 1 KHz below the carrier frequency.

Now if we look at the output signal from the limiter with our spectrum analyzer, we will see two sidebands 56 db down located ±1 KHz from the carrier frequency. Further investigation with a frequency discriminator or phase demodulator will indeed reveal that these sidebands are correlated in phase in such a way as to represent only phase modulation. Notice further that this will be true regardless of the loss or gain incurred in passing through the limiter. The translation of energy to the upper sideband frequency was accomplished by the non-linearity of the limiter, and the AM components were removed because of the insensitivity of the limiter output to variations in input level.

In a similar manner, the amplitude variations in the LO drive to a mixer often do not markedly effect the conversion efficiency of the mixer and hence are not readily transferred to the mixer output, but the phase modulation components present in the input pass through and appear in unchanged sideband to carrier ratios in the mixer output. The effect occurs in multipliers simply because PM sidebands are raised in direct ratio to the multiplication N being achieved, while the amplitude variations are generally somewhat attenuated depending on the particular design of the multiplier. In a divider AM is suppressed and single sideband spurs are converted to PM due to the fact that the divider output amplitude is usually independent of the input signal level, although we must recognize that whatever phase modulation sidebands are present in the input are reduced by the divider ratio at the output.

The RF Spectrum of a Signal

In all that follows we shall make the following assumptions about the general nature of the signal we are measuring; the AM spectrum is much less than the PM spectrum for all offset frequencies of interest, and the RMS phase deviation of the carrier signal is $<<$ 1 radian. Let us now consider the RF spectrum of such a signal. Assuming that we observe its spectrum with a spectrum analyzer having a 1 Hz noise equivalent bandwidth, we would obtain a display similar to that shown in Figure 1. It should be noted that the symmetry of the display about the carrier frequency is a direct result of the assumptions that we have made regarding the character of the signal, since small angle phase modulation side bands must be symmetrical.

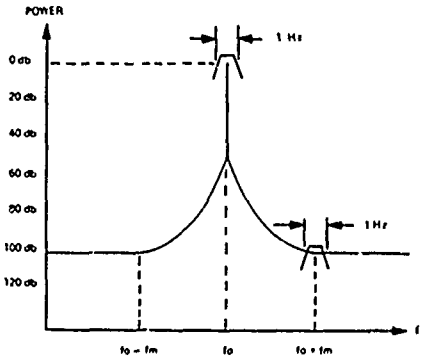


FIGURE 1

RF power spectrum of a signal observed with a 1 Hz bw spectrum analyzer. (Assume AM Noise is negligible).

Definition of $\mathcal{L}$(fm)

If we now wish to speak quantitatively about the amount of noise modulation on the signal, we can tune the analyzer first to the carrier frequency and record a zero db level. Then we can move above or below the carrier by an offset frequency fm, and record the value observed relative to the zero db reference level. We can now speak of the "single side band phase noise in a 1 Hz bandwidth to signal ratio, as a function of the offset frequency fm". This definition has been given by Dr. Donald Halford of NBS[1], and he has assigned to this definition the symbol $\mathcal{L}$ (fm). This definition does not attempt to have a specific mathematical significance, but is an attempt to specify clearly a measure of performance. Since the modulation spectrum is symmetrical, it is only necessary to specify $\mathcal{L}$ (fm) for one side of the spectrum. The normal way of presenting the performance in terms of $\mathcal{L}$ (fm) is with a db vs log fm plot such as is shown in Figure 2. This plot of $\mathcal{L}$(fm) now

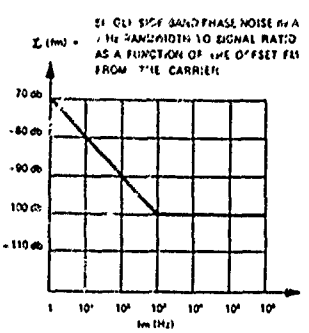


FIGURE 2

Definition of $\mathcal{L}$ (fm).

has a clear physical interpretation in terms of the RF spectral density of a phase noise modulated signal. All phase noise data presented in this paper will use this convention of displaying noise performance. Where db values of $\mathcal{L}$ (fm) are used the units will be noted as db/Hz.

Definition of $S\phi$ (fm)

The most widely used mathematical definition of phase noise density is $S\phi$ (ω), as used by Cutler and Searle[2]. I have taken the liberty here of expressing $S\phi$ as a function of fm instead of ω, since it does not change the meaning of the definition, and measured data is more commonly expressed as a function of offset frequency. Cutler and Searle refer to $S\phi$ (units of rad^2/Hz) as the "two-sided power spectral density" of ϕ (t), where ϕ (t) is found by passing the complete RF spectrum of a signal v (t) through an ideal phase detector. This definition makes $S\phi$ (fm) numerically equal to $\mathcal{L}$ (fm) for a given value of fm. The difference between $\mathcal{L}$ (fm) and $S\phi$ (fm) lies in the fact that $S\phi$ (fm) is the two sided mathematical definition of the power spectral density of the modulation function, while $\mathcal{L}$ (fm) avoids the use of mathematical negative frequencies and provides a more "physical" definition. An example of a db vs fm plot of $S\phi$ (fm) is shown in Figure 3.

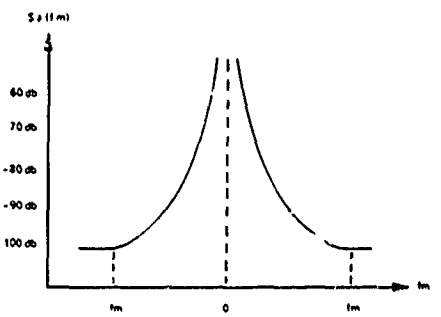


FIGURE 3

$S\phi$ (fm) – Two sided mathematical "Power Spectral Density" of the phase modulation on a signal.

The spectral density $S\phi$ (ω) is necessary and useful in order to derive several important mathematical formulas. The two which are of most interest to us are[2]:

The variance of the phase fluctuations.

$$\sigma^2\ [\phi(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} S\phi(\omega)d\omega \qquad (1)$$

This expression can be rewritten to yield the average value of $\overline{\phi^2(t)}$ as a function of $S\phi$ (fm):

$$\overline{\phi^2(t)} = \sigma^2[\phi(t)] = 2\int_0^{\infty} S\phi(fm)dfm \qquad (2)$$

The expected value of the variance of the frequency departure averaged over a time τ,

$$\sigma^2\ [<\dot{\phi}>\ t,\tau\ ] = \frac{2}{\pi\tau^2}\int_{-\infty}^{\infty} S\phi\ (\omega)\sin^2(\frac{\omega\tau}{2})\ d\omega \qquad (3)$$

This expression can be rewritten and normalized to yield the standard deviation of the fractional frequency fluctuations for an averaging time τ as a function of $S\phi$(fm);

$$\sigma\left[\frac{\Delta f}{f}(\tau)\right] = \frac{1}{\pi\tau f_0}\left[\int_0^{\infty} S\phi(fm)(1-\cos 2\pi\, fm\,\tau)dfm\right]^{1/2} \qquad (4)$$

Using these formulas one can easily calculate values of $\overline{\phi^2(t)}$ and $\sigma[\Delta f/fo(\tau)]$ from measured $\mathcal{L}$(fm) data by use of computer techniques, since $\mathcal{L}$(fm) and $S\phi$ (fm) are numerically equal. In Appendix I, it has been shown that the average squared value of the phase fluctuations has the "physical interpretation" of being the ratio of the total noise power in the phase modulation sidebands to the signal power. For a practical measure, the reciprocal of $\overline{\phi^2(t)}$ is more commonly used as a measure of synthesizer performance, and the measurement bandwidth must also be specified. This is discussed further in the next section.

A convenient definition of the fractional frequency deviation is:

Given a signal $V_{(t)}$:

$$V(t) = A_0 \cos\left[\omega_0 t + \phi(t)\right] \qquad (5)$$

$$\frac{\Delta f}{f_0}(\tau) = \frac{\phi(\tau) - \phi(0)}{2\pi f_0 \tau} \qquad (6)$$

In this definition ϕ (τ) is the value of ϕ (t) at the end of the sample time τ, and ϕ (0) is the value of ϕ (t) at the start of the sample time. The value of $\sigma[\Delta f/fo(\tau)]$ is of course calculated from a large number of individual measurements using some appropriate technique.

A Method of Measuring $\mathcal{L}$ (fm)

Special techniques must often be used in order to measure the residual phase noise performance of signal processing devices which have lower noise than any attainable source. The most common technique uses a test setup which is shown in Figure 4. The example of two multipliers driven by a common source has been used here, but any two identical signal processing devices may be put in place of the multipliers. A line stretcher (or the correct length of coaxial cable) must be introduced into one of the parallel paths so that the output of the phase detector is zero volts DC. The voltage Vx(fm) is now related only to the phase perturbations generated by the two devices under test, since any modulation present on the input signal does not appear at the output due to the symmetry of the test setup.

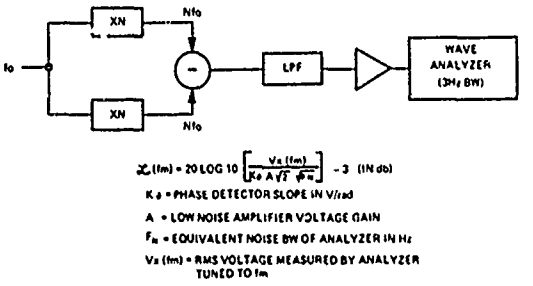


FIGURE 4

The concept of residual noise and its measurement.

Now we can measure the parameters K ϕ, A, and F_N as defined in Figure 4, and we can measure a value Vx (fm) with the wave analyzer tuned to the frequency fm. An expression is derived in Appendix I which relates the experimental parameters to the value of $\mathcal{L}$ (fm) which is due to the effects of both sources. If we assume both sources are identical, we must reduce the value calculated by the expression in Appendix I by 3 db as is shown below.

$$\mathcal{L}(fm) = 20 \log 10 \left[\frac{Vx\ (fm)}{K\phi\, Av\sqrt{2}\sqrt{F_N}} \right] - 3 \quad (7)$$

(Performance of one device assuming identical devices under test.)

A simple modification of this method can be used with basically the identical test setup if the output frequency of the device under test can be made identically equal to the input frequency. One of the devices under test is simply replaced by the proper length of cable to achieve zero volts DC out of the phase detector, and the test proceeds as usual. In calculating the value of $\mathcal{L}$ (fm) from the data in this case we do not subtract 3 db as indicated in equation (7), since all of the phase fluctuations being measured are the result of only one noise source. A little more care must be exercised in the quality of signal source used for this measurement, since the symmetry of the situation is not as complete. A potential problem in this case could occur if a lot of AM was present on the input signal, and the device under test had a pronounced tendency for AM to PM conversion. In this case we might be measuring AM to PM conversion, rather than the residual phase noise of the device under test.

Definition and Measurement of S/N (f_L, f_H)

Another very common method used to specify synthesizer performance is the signal to phase noise ratio. This is usually specified as the "signal to phase noise ratio in a 30 KHz RF bandwidth excluding 1 Hz centered on the carrier". The somewhat odd sounding bandwidth description is simply the result of the fact that the integral given in equation (2) cannot be evaluated from zero to infinity in the practical case. In practice we use a single pole RC low pass filter with a noise equivalent bandwidth f_H of 15 KHz to filter the input to an RMS voltmeter used in place of the wave analyzer shown in Figure 4. The noise equivalent high pass frequency f_L due to the AC coupling of the amplifier and the RMS voltmeter is $\approx$ 0.5 Hz. Let Vx(t, f_L, f_H) be the voltage into the RMS meter, and ϕ(t, f_L, f_H) the phase fluctuations corresponding to the same frequency band. Defining S/N(f_L, f_H), and using equation 2:

$$S/N\ (f_L, f_H) = \frac{1}{\overline{\phi^2\ (t, f_L, f_H)}} = \left[2 \int_{f_L}^{f_H} S\phi(fm)\, dfm \right]^{-1} \quad (8)$$

This can be expressed in terms of the test set parameters by use of equation (22) of Appendix I:

$$\overline{\phi^2\ (t, f_L, f_H)} = \frac{\overline{Vx^2\ (t, f_L, f_H)}}{(K\phi A)^2} \quad (9)$$

If we now define the RMS value of $Vx(t, f_L, f_H)$ as $Vx(f_L, f_H)$, and combine (8) and (9):

$$S/N(f_L, f_H) = \left[\frac{K\phi A}{Vx\ (f_L, f_H)} \right]^2 \qquad (10)$$

For convenience we can express this in db, and if we assume that the phase noise contributed by both devices is equal we must add three db to the value calculated in equation (10) to obtain a value indicative of the performance of one device only:

$$S/N\ (f_L, f_H) = 20 \log 10 \left[\frac{K\phi A}{Vx\ (f_L, f_H)} \right] + 3 \quad (11)$$

(Performance of one device, assuming identical devices under test.)

The three db factor above is added, instead of subtracted as in the case of $\mathcal{L}$ (fm), due to the fact that S/N (f_L, f_H) is a signal to noise ratio rather than a noise to signal ratio. The remarks under the section dealing with the measurement of $\mathcal{L}$ (fm) in the case where the output frequency of the device under test is the same as the input frequency also apply to the measurement of S/N (f_L, f_H). In this case, however we subtract 3 db from equation (11) to determine the performance of the device under test when the symmetrical device under test is replaced with an appropriate length of coaxial cable.

The physical interpretation of S/N (f_L, f_H) is simply the ratio in db of the signal power to the total power contained in all the phase modulation sidebands within the band of frequencies fo $\pm f_H$, but excluding those sidebands which lie within the band of frequencies fo $\pm f_L$.

A Test Set for the Measurement of $\mathcal{L}$ (fm) and S/N (f_L, f_H)

We have developed a very accurate phase noise test set for the measurement of $\mathcal{L}$ (fm) and S/N (f_L, f_H). A block diagram of the test set is shown in Figure 6, and a photograph of it is shown in Figure 5. This set uses basically the same principles illustrated in Figure 4, but has many refinements to allow very accurate measurements to be made. This test set can measure $\mathcal{L}$(fm) at carrier frequencies from 1 to 500 MHz, and at values of modulation frequency from 20 Hz to 50 KHz. From 1-50 MHz, measurements of synthesizers can be made of both $\mathcal{L}$ (fm) and S/N (f_L, f_H) with a one sigma repeatability of $\approx \pm 0.7$ db (70% of readings are within ± 0.7 db of the average value). The absolute calibration accuracy of the set is $\approx \pm 0.8$ db over the 1-50 MHz range. The set uses an analyzer with 3 Hz bandwidth, and has a measurement floor $\mathcal{L}$ (20 Hz) ≈ -140db/Hz and $\mathcal{L}$ (5 KHz - 50 KHz) ≈ -172db/Hz. This test set was used to measure all the noise data presented in this paper.

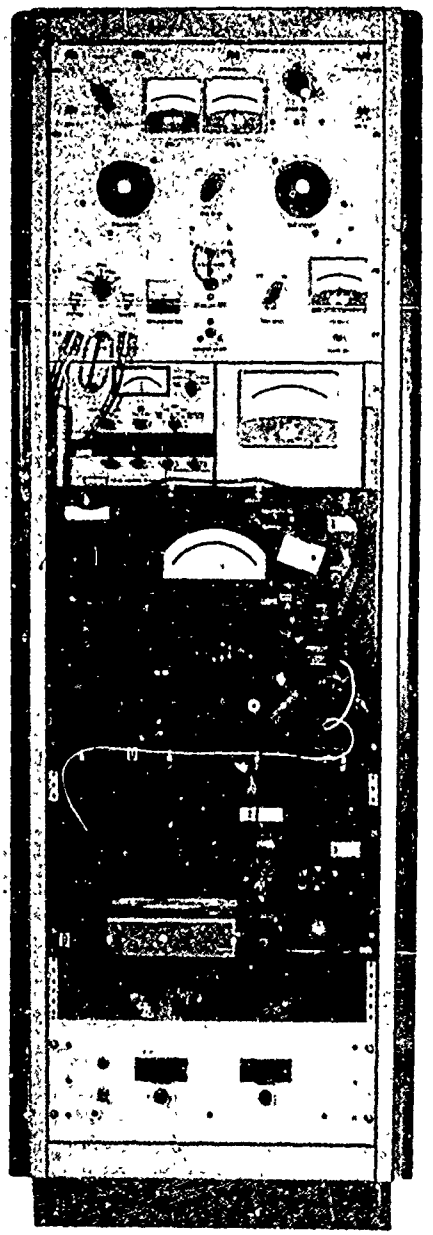


FIGURE 5

Test set used for measurement of $\mathcal{L}$ (fm).

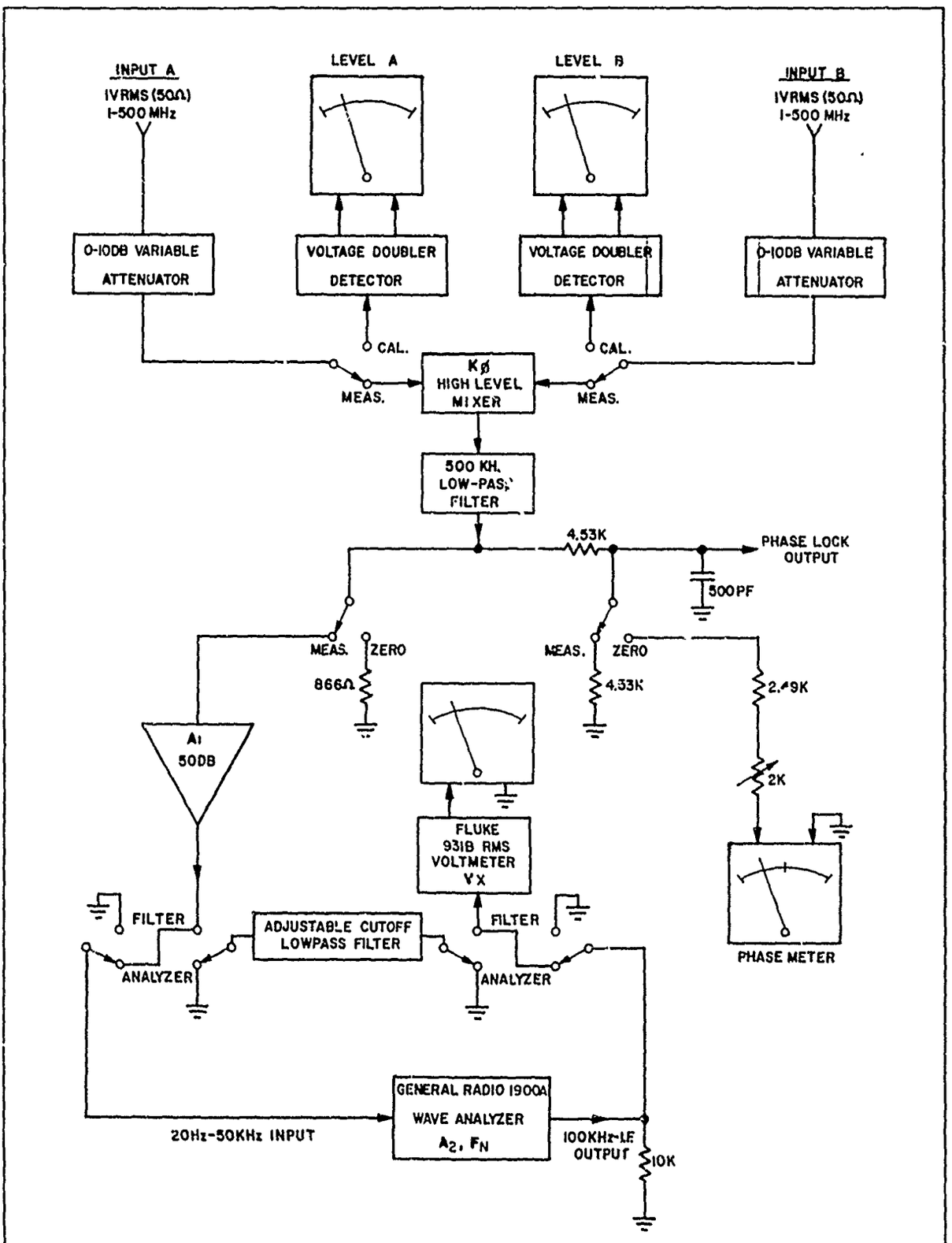


FIGURE 6

Block Diagram of Phase Noise Test Set

Method of Measuring Fractional Frequency Deviation

In actual practice, $S\phi(fm)$ cannot be easily and accurately measured for very low offset frequencies (<20Hz). For this reason it is usually not practical to attempt estimation of $\sigma[\Delta f/fo(\tau)]$ by means of the integral definition given in equation (4), since for longer averaging times an accurate knowledge of $S\phi(fm)$ is needed for extremely small offset frequencies in order to accurately evaluate the integral. An alternate method which is often used because of its good resolution and relative simplicity is the technique of period measurement. A block diagram of a typical test setup is indicated in Figure 7 for the measurement of the residual $\sigma[\Delta f/fo(\tau)]$ of the 645A synthesizer. This test is very similar to the

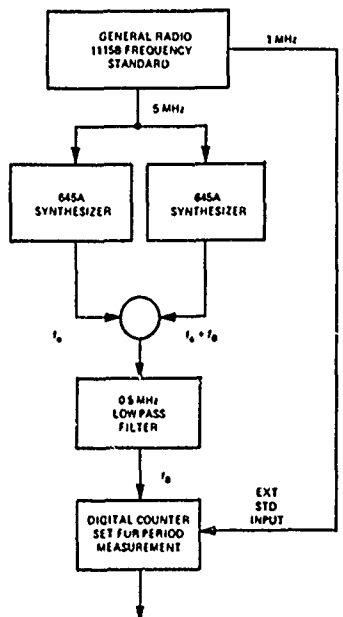


FIGURE 7

Test set up for $\Delta f/f_o$ (τ) measurement.

residual phase noise test indicated in Figure 4, in that the fluctuations of the frequency standard are cancelled out at the mixer output due to symmetry considerations. The resolution of this type of measurement for a given averaging time τ is determined by the time base frequency (f_{TB}) used for the period measurement, the number of multiple periods (M) being measured, and the signal frequency (fo). The value (R) of $\sigma[\Delta f/fo(\tau)]$ which is indicated by one count in the least significant digit of the counter readout is given by equation (12).

$$R = \frac{M}{\tau^2 \, fo f_{TB}} \qquad (12)$$

M = Number of periods measured.
f_{TB} = Time base frequency of counter being used for the period measurement (in Hz).
fo = Synthesizer output frequency in Hz.
τ = M/f_B Sample time of measurement in seconds.
f_B = Offset frequency between synthesizer outputs in Hz.

The standard deviation of the fractional frequency fluctuations must now be computed from the period measurement data as recorded on the printer tape. The most generally accepted variance which can be calculated from this type of data is the Allan variance[4] for N = 2. The general definition given in Allan's paper can be reduced to the form shown in equation (13) for N = 2:

$$\sigma^2(2, T, \tau, f_h) = \frac{(\omega_{i+1} - \omega_i)^2}{2} \qquad (13)$$

τ = Sample time in seconds
T = Time between the start of one sample and the start of the next sample in seconds.
f_h = Bandwidth of the measuring system in Hz.
ω_i = Average angular frequency during the (ith) sample of length τ.
ω_{i+1} = Average angular frequency during the (i+1)st sample of length τ.

This definition represents one estimate of the Allan variance of the angular frequency fluctuations for N=2 made from two samples (ω_{i+1} and ω_i) of the angular frequency averaged over the sample time τ. Notice that the values of T, τ, and f_h must be specified for any given measurement which is made, since the value of the Allan variance depends to some extent on these variables.

Now we must calculate an estimate of the expected value of this variance from a quantity of L measurements, as indicated by the numbers $n_1, n_2, \ldots, n_L$, shown on the output record of the digital printer. In order to do this, we can calculate L−1 estimates of the variance from the L points indicated in the record, and divide the sum of the L-1 estimates by L-1 to obtain the average (expected) value of the variance.

This value can then be normalized to yield the average value of the variance of the fractional angular

frequency fluctuations ($\Delta\omega/\omega o$), which is equal to the variance of the fractional frequency fluctuations ($\Delta f/fo$). The standard deviation of the fractional frequency fluctuations can then be obtained by taking the square root. A computational form can be derived to compute the value of the standard deviation of the fractional frequency fluctuations directly from the measured data. This value has been labeled $\sigma[\Delta f/fo(2, T, \tau, \omega h)]$ to indicate it has been computed by means of the Allan variance for N = 2.

$$\sigma\left[\frac{\Delta f}{f_o}(2, T, \tau, f_h)\right] = R\left[\frac{1}{L-1}\sum_{i=1}^{i=L-1}\frac{(n_{i+1}-n_i)^2}{2}\right] \quad (14)$$

In actual practice, one of the most difficult aspects of making an accurate measurement by this method is in getting the instrumentation ground loops low enough in magnitude to avoid affecting the measurement. It is necessary to use very good isolation transformers and grounding techniques to avoid ground loops due to AC current flow through the line filter capacitors, and primary to secondary winding capacities,of the instruments used in the measuring system. In the test setup shown in Figure 7, the frequency standard is operated on battery power to help reduce ground loops. Another important point is that the internal oscillator of the digital counter must be disabled. If this is not done, phase modulation will occur due to the small offset between the frequency standard being used and the frequency standard in the counter. This modulation is sufficiently large to invalidate the data. This problem usually manifests itself by a periodic sloughing of the values indicated on the counter, at a rate which is related to the frequency difference between the 1 MHz standard frequency and the counter's 1 MHz standard.

Section II - The Fundamental Problems of Direct Synthesis

Building a very high performance direct synthesizer requires an understanding of the various mechanisms of noise and spurious generation which can occur both in individual circuits, and in the overall operation of the synthesizer. The three primary effects which must be considered and dealt with, are uncorrelated phase noise, correlated phase noise, and spurious signal generation. Let us first consider the problem of uncorrelated phase noise.

Uncorrelated Phase Noise

There are two basic sources of uncorrelated phase noise which must be considered in a synthesizer. The first is flicker of phase noise due to modulation of the carrier frequency by the various active devices, and the second is simply due to the additive noise floors caused by the finite signal levels and noise figures which are encountered in practical circuitry. First let us consider the nature of the flicker of phase noise.

Flicker of Phase Noise

Don Halford et al[3] have clearly demonstrated that intrinsic, direct, phase modulation with approximately a 1/f power spectral density (falling at 0 db/decade with increasing fm) is generated when an RF signal is passed through any PN junction semiconductor device. Their measurements indicate a typical value of $\mathcal{L}$ (1 Hz) $\approx$ −115 db/Hz for common emitter (or common source) amplifiers with no negative feedback. They have further demonstrated that negative feedback can be used in transistor circuitry to effect a substantial reduction in the flicker of phase noise due to this effect. Our experience in the design of low phase noise circuitry verifies their conclusion that any type of negative feedback used with an active device tends to reduce the flicker of phase noise. The spectral density of phase noise due to this effect appears to be non-temperature dependent, so that the performance of any system in a range of offset frequencies where this effect predominates essentially does not vary with temperature.

Each active device is a statistically independent generator of this type of noise, and hence each doubling of the number of active devices which contribute directly to the output spectrum should raise the overall spectral density by 3 db (assuming identical devices). In our synthesizer there is one example of this sort of behavior which gives experimental verification to the value of $\mathcal{L}$ (1Hz) $\approx$ − 115 db/Hz measured by Halford et al. In the UHF section of the synthesizer there are 15 transistors used in the common emitter configuration with no negative feedback in order that the maximum amount of gain per stage can be achieved. Referring to the dashed line shown for the residual noise of the UHF section in Figure 20, the portion of the curve for fm < 500 Hz is limited by flicker of phase noise. Extrapolation of this line reveals $\mathcal{L}$(1 Hz) $\approx$ −105 db/Hz, which must be reduced by $\approx$ 12 db to −117 db/Hz to

find the typical noise due to one source, since there are 15 sources. This agrees very well with the performance reported by Halford et al.

In all the 30–70 MHz circuitry which is used in the synthesizer we have applied negative feedback to as great an extent as possible in light of gain and noise figure considerations. We have been able to achieve significant reduction below the $\mathcal{L}$ (1 Hz) ≈–115db/Hz per stage value. An example of a typical circuit used is shown in Figure 8, along with its measured performance. $\mathcal{L}$ (1 Hz) ≈ –122 db/Hz in this case, which indicates that the performance of the two gate FET used as an amplifier is fairly respectable, since it uses no negative feedback, while the other two stages do use forms of negative feedback.

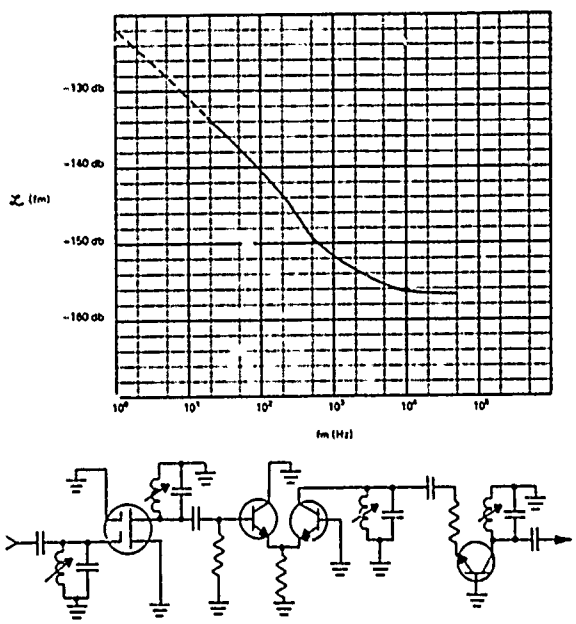


FIGURE 8

Residual noise of a 30 MHz harmonic selector module.

Additive Phase Noise

By using the highest possible signal power levels compatible with low spurious product generation, and optimizing amplifier and mixer noise figures, the noise floors in the various individual circuits have been held from ≈ –148 to –170 db/Hz relative to the carrier level. Where noise floor situations of this type are encountered, the noise is half AM and half PM noise where it first appears. In nearly all cases the resultant signal is passed through several devices which tend to suppress the AM noise before reaching the final output, thus improving the noise floor in terms of signal to total noise power in a given bandwidth by the 3 db of AM noise which is removed. Each place where a phase noise floor like this is encountered it represents an independent source of noise generation, and contributions add in RMS fashion exactly as do the sources of flicker of phase noise at lower offset frequencies.

The only case where limiting is not applied to reduce AM noise is at the output of the final mixer just prior to the synthesizer output amplifier. The overall noise floor at this point is ≈ – 147 db/Hz, which sets the AM noise floor at the output at ≈ –150 db/Hz. The PM noise floor at the output is at ≈ –137 db/Hz, due to the summation of all the independent PM floors encountered in various parts of the instrument.

Correlated Phase Noise

In order to examine the general problem due to correlated noise in a direct synthesizer, let us consider the block diagram shown in Figure 9. This block diagram represents the basic mode of operation of most wideband direct synthesizers. In effect, we start with a common frequency fo, and multiply it up by M & N in two separate multipliers. The ranges of values taken by M and N are appropriately adjusted so that the output frequency range covered by fo (M-N) is as desired. In actual practice, each effective multiplier path has a significant amount of time delay associated with it, and typically these time delays cannot easily be matched. The differential time delay which exists between these two paths has been shown as a perfect time delay $\Delta\tau$ in the M path for simplicity. The following expression has been developed in Appendix II. This expression indicates the relationship between the spectrum of the input signal $\mathcal{L}$ (fm)IN at a given value of fm, and the

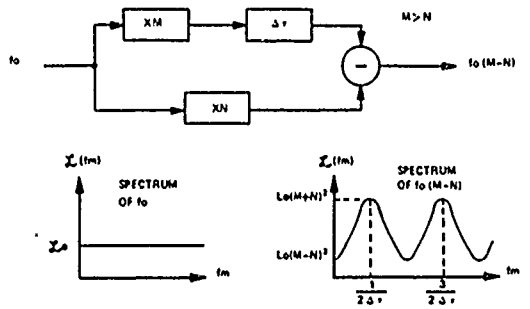


FIGURE 9

The effect of differential time delay in a direct synthesizer.

spectrum of the output signal $\mathcal{L}$ (fm)OUT at the same value of fm:

$$\frac{\mathcal{L}\,(\text{fm})\ \text{OUT}}{\mathcal{L}\,(\text{fm})\ \text{IN}} = M^2 + N^2 - 2\,MN\cos\omega m\,\Delta\tau \qquad (15)$$

Now this expression can be seen to have the following extreme values, depending on the value of $\omega\,m\,\Delta\tau$:

$$\text{for } \omega m\,\Delta\tau \approx o, 2\pi, 4\pi, \cdots, 2N\pi$$
$$\frac{\mathcal{L}\,(\text{fm})\ \text{OUT}}{\mathcal{L}\,(\text{fm})\ \text{IN}} = (M-N)^2 \qquad (16)$$

$$\text{for } \omega m\,\Delta\tau \approx \pi, 3\pi, 5\pi, \cdots, (2N+1)\,\pi$$
$$\frac{\mathcal{L}\,(\text{fm})\ \text{OUT}}{\mathcal{L}\,(\text{fm})\ \text{IN}} = (M+N)^2 \qquad (17)$$

The significance of this result is shown clearly in the plot of Figure 9, which shows the output spectrum which would be obtained from the system shown if a flat input spectrum with power spectral density $\mathcal{L}\,o$ were applied as shown. Ideally we would like to obtain always the power spectral density $\mathcal{L}$ (fm) OUT = $(M\text{-}N)^2$, which is the best possible value in this case since it represents only the effect of an ideal multiplier of value M-N, but in fact it can become as bad as $\mathcal{L}$ (fm) OUT = $(M+N)^2$ due to the effects of the differential time delay.

Several conclusions can be drawn from the preceeding example. In a direct synthesizer one should keep the value of $\Delta\tau$ as low as possible, so that the frequency $fm = 1/(2\Delta\tau)$ occurs well out into the region of the noise floor. In this way the effects of $\Delta\tau$ will be minimized since the input spectral density will be as low as possible, and hence will not "bump up" so much. We should keep the frequency fo as high as possible, and M and N as low as possible, to further minimize the magnitude of the "bumps". And last, we must use the utmost care to keep the spectrum of the "common frequency" fo as clean as possible. Hopefully, if we pay careful attention to these details, the output spectrum of the completed synthesizer will be limited by the sum of the multiplicative and additive noise sources in the remainder of the circuitry. This goal has very nearly been achieved in the 645A synthesizer, as will be shown later.

Spurious Product Generation

Thus far we have discussed only the effects of random phase noise modulation. Another very major problem is that of coherent spurious product generation. A spurious output signal can be of two types; a coherent sinusoidal signal appearing in the synthesizer output signal at a harmonic of the output frequency, or at a frequency which is not harmonically related to the output frequency. The harmonically related spurious outputs are specified as more than 30 db down from the output signal, and are typically 35 to 45 db down depending on the output frequency being generated. Non-harmonically related spurious signals are specified as being at least 100 db down from the carrier, and are typically close to the 110 db down in the worst cases.

Non-harmonically related spurious signals can appear virtually anywhere above or below the output frequency, due to the complexity of the synthesis system. They may appear in symmetrical pairs around the carrier frequency, or as single sideband signals. Some spurious signals remain at fixed frequencies as the output frequency is tuned, while others may move rapidly across the output signal as the frequency is tuned. The behavior of a particular spur depends on the mechanism by which it is generated. Fixed frequency spurious signals commonly arise in the output due to the various fixed frequencies which are generated within the instrument, or at harmonics of these frequencies. Moveable spurious signals generally arise due to mixer crossings of fairly high order, and therefore tend to move very rapidly in frequency as the output frequency is tuned.

Mixer Generated Spurious Products

In order to perform addition or subtraction of frequencies in a synthesizer we must utilize various types of mixers. Ideally, we would like to use mixers which are pure product modulators (i. e., those which have only second order terms in the series expansion representation of their transfer functions). If we could use pure product modulators, the only signals coming out of a mixer having inputs of F_H (the higher frequency) and F_L (the lower frequency) would be F_H, $2F_H$, F_L, $2F_L$, $F_H + F_L$, and $F_H - F_L$. In actual practice all orders of distortion exist to some extent in any mixer so that at the mixer output we obtain signals at all the possible combinations of F_L and F_H, as shown in Figure 10.

The mixer crossings which we most commonly have to deal with are indicated in the spectral plots shown in Figure 10. Usually we are trying to select the sum frequency $F_H + F_L$, or the difference frequency

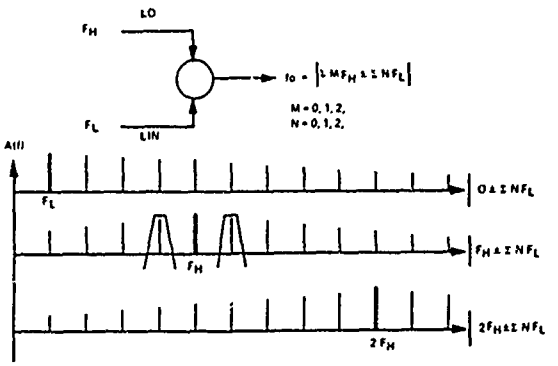


FIGURE 10

Generation of spurious products in a mixer.

F_H - F_L. In either case, it can be seen from Figure 10 that the lowest order classes of spurious signals we will encounter are the harmonics of the low frequency signal (Σ N F_L), and twice the high frequency signal minus the harmonics of the low frequency ($2F_H - \Sigma NF_L$). There are two cases of interest to consider with regard to these types of mixer crossings, fixed input frequencies to the mixer, or variable input frequencies. In the case of fixed input frequencies, very often it is possible to arrange the input frequencies so that all the spurious signals are easily filtered. In the case of variable input frequencies, the spurious signals will often sweep through the desired output signal as the inputs are varied. In these cases no amount of filtering will help, and we must design very carefully to keep spurious product generation as low as possible. Another spurious signal which must always be removed by filtering is the frequency F_H, since it is very close to either of the desired output frequencies and normally is a high level signal. Let us now consider several of the most important design factors which must be considered in order to realize low spurious products in the output of a sum or difference mixer.

Choice of Mixer Frequency Ratios

There are two important frequency ratios to consider in a sum or difference mixer where $F_L << F_H$. The first is the ratio of Fo/F_L (see Figure 10). In the 645A this ratio is kept < 9 in all cases, which results in inband spurious products of order 9 or less being the worst case. The order of the inband spurious is a direct result of this ratio. For example, if the minimum value of Fo/F_L for a given mixer were 7, the worst spurious would be of order 7 or greater. The second ratio of importance involves only the frequency F_L. Let us illustrate the problem which can occur if F_L varies over too wide a band of frequencies by use of an example. If we have a fixed F_H of 30 MHz, and F_L varies from 1 to 2 MHz, the output filter for a sum mixer must run from 31 to 32 MHz. When F_L = 1 MHz, the $F_H + 2F_L$ spurious product will appear within the passband of the filter at 32 MHz, and hence cannot be filtered. In order to be able to properly filter this spurious, the smallest ratio of the band center of F_L to the bandwidth of F_L is approximately 3.9 in the 645A.

Effect of Harmonics of $F_L + F_H$

The signals which are used for the F_L & F_H inputs to a mixer are usually obtained from solid state circuitry, and are fairly rich in harmonics. Since some of the harmonics of F_L fall near the desired sum or difference frequency, they can easily pass straight through the mixer and appear as spurious products in the output band. Harmonics of either signal can also upset the basic symmetry of the drive signal to a mixer, and for this reason can enhance spurious generation since many mixers depend on symmetry conditions for the rejection of certain types of spurious signals. In order to reduce these effects, harmonics of both F_L and F_H are attenuated by the use of low pass filtering in the inputs to all critical mixers in the 645A.

Effect of Reactive Terminations

The spurious product generation of many types of mixers is markedly effected by reactive terminations at the mixer outputs. As an example, the performance of a hot carrier diode mixer on a fifth order spurious signal was observed to be as much as 40 db worse when a tuned filter was used at the output, rather than a broadband resistive termination. The effect appears to be largely due to the impedance seen by the image frequency, as the level of the spurious signal can be changed over wide ranges by "tuning" the line length between the mixer output and the filter input with a line stretcher. The driving impedances seen by the mixer are not nearly as critical, and usually have very little effect on performance. This effect is most pronounced in double balanced, broadband mixers where many of the spurious responses are normally reduced substantially by symmetry conditions.

L. O. Versus Linear Signal Considerations

In a sum or difference mixer where $F_L << F_H$, we always use the F_L signal as the linear signal in order

to obtain lower spurious responses. To see why this is done, let us examine the effects which signal levels have on a given spurious product. The general equation for the amplitude $V_{M,N}$ of a spurious signal which is a result of a product $MF_H \pm NF_L$ is shown in equation (18) below:

$$V_{M,N} \approx K (V_H)^M (V_L)^N \qquad (18)$$

It is clear from this equation that we should reduce the signal levels V_H & V_L in order to reduce the spurious signal voltage $V_{M,N}$. Reducing the voltage V_H by one db causes $V_{M,N}$ to decrease by roughly M db, and similarly a 1 db reduction of V_L causes $V_{M,N}$ to go down by roughly N db. Since in our case $V_L \ll V_H$, the spurious signals we are concerned with have M = 0, 1, 2, but have much higher values of N (see Figure 10). Better results are therefore obtained by using F_L as the linear signal, and decreasing the amplitude V_L until the mixer generated spurious signals are reduced to a satisfactory level.

Proper Choice of Mixer Types

It is extremely important to know the performance characteristics of the different types of mixers which are available, in order to apply them correctly to obtain the best overall performance characteristics. In the 645A we have used transistors, hot carrier diodes, FET's, and varactor diodes as the non-linear elements in mixers. All of the mixers are arranged in either a single or double balanced configuration to obtain as large an improvement in local oscillator feed through as is possible. Transistors are used in some low frequency (30-40 MHz) mixers where spurious generation is not a serious problem. In the more critical low frequency applications, FET's are used to take advantage of their nearly square law transfer characteristics. At higher frequencies (370-420 MHz) hot carrier diodes are used where broad band signals must be accommodated and in-band spurious problems are not so bad because of the synthesizer design. Where inband spurious products exist at higher frequencies and the bandwidth is not so large, varactor diode upper sideband up convertors are used.

The performance of the square law devices is roughly an order of magnitude better in critical applications where inband spurious signals cannot be avoided. In order to maintain a 100 db spurious specification at the instrument output, it was necessary to keep inband spurious products on many of the mixers in the order of 124 db down. This level of performance was achieved while maintaining the mixer noise floors as low as −148 db/Hz, in order to keep from degrading the overall noise performance of the instrument.

Coherent Addition of Spurious Signals

Just as the effects of differential time delay are bothersome in the case of coherent noise modulation, they can be a problem due to coherent addition of spurious signals. Let us illustrate this statement by means of an example. Suppose we have a mixer which has a symmetrical phase modulation sidebands 112 db down on each of its input signals, and let us further suppose that these modulation signals are at ± 1 MHz from each signal. The relative phasing between the two sets of modulation sidebands may be anywhere from 0 to 180°, depending on differential time delays introduced in signal processing. What is then the magnitude of the ± 1 MHz spurious sidebands generated on the output signal? Since we have postulated identical spurious phase modulation, the resultant sidebands could be anywhere from infinitely far down if the phasing were exactly 180° out of phase, to 6 db higher (or 106 db down) if they were exactly in phase.

This problem necessitates the use of a worst case design procedure in order to guarantee the overall performance of the synthesizer. Where the possibility of coherent spurious signal phasing exists, the internal specifications must be raised by 6 db at each input to a mixer above the specification desired at the mixer output. This worst case design procedure was used for the design of the 645A synthesizer, and results in spurious product specifications of as great as 124 db down on some of the internal signals which are fed into the UHF section of the instrument.

Section III - Theory and Performance of the 645A Synthesizer

Now that the basic problems of building a direct synthesizer have been discussed, we are in a position to understand more fully some of the design compromises which were made in the 645A synthesizer to optimize the overall noise and spurious performance. The basic organization of the instrument is shown in Figure 11. The circuitry in the instrument has been divided into

four sections in Figure 11 for purposes of clarity; the input section, fixed frequency generators, sub-synthesizer, and the UHF section. Note that there are three synthesized signals: 30-70 MHz in 10 MHz steps, 30-39 MHz in 1 MHz steps, and 40-41 MHz in 0.01 Hz steps. These signals represent the 10 MHz, 1 MHz, and 0.01 Hz steps respectively which are used to synthesize the final output of the instrument. These three inputs are fed into the UHF section, where they are combined in a drift cancelling loop which operates in the 370-420 MHz region, and then are reduced to the 0-50 MHz output frequency range in a subtractive mixer. Now let us summarize briefly the most salient aspects of the design with regard to noise and spurious products, and then describe the operation and circuitry of each individual section in more detail.

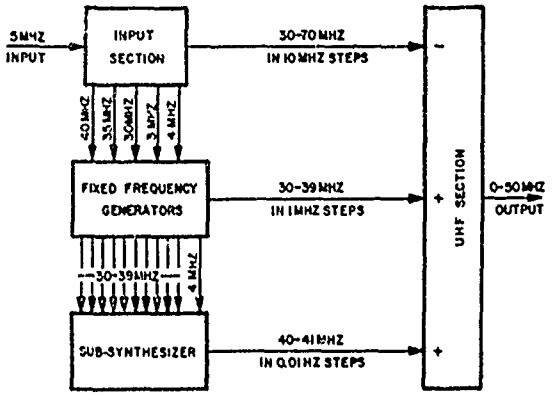


FIGURE 11

Overall Block Diagram of Synthesizer

Designing for Low Correlated Noise

Two important aspects of the overall design for low correlated noise are the use of a high common frequency to start the synthesis process, and the use of low multiplication factors on the synthesized signals. First, let us consider the 10 MHz common frequency. Note that this frequency is generated by multiplying the 5 MHz standard by a factor of two (see Figure 15). There are only 4 active stages through which the signal must pass in order to obtain the 10 MHz common frequency, and each of these has been very carefully designed to keep the spectrum of the common noise at this point to an absolute minimum. In addition, the 5 MHz input signal from the internal or external standard is passed through a two pole crystal filter before it is doubled. This filter serves to attenuate the noise sidebands present on the input signal at offset frequencies greater than 200 Hz, so that for offset frequencies greater than 500 Hz the effects of the frequency standard are negligible. These techniques allow the generation of a common 10 MHz signal with very low residual phase noise.

Now let us consider further the multiplication factors used on the synthesized signals. The effective method of operation of the synthesizer is to multiply the common 10 MHz signal by M = 7-8, and by N = 3-7 (where M and N are defined as used in equation (15) and Figure (9). The effective multiplication M is obtained by summing the 30-39 MHz and 40-41 MHz signals in the UHF section, while the multiplication N of 3-7 is obtained directly by means of the snap diode comb generator. The Nfo signal and Mfo signals are then effectively subtracted in the UHF section, so that the system operation can be approximated by use of equation (15). If we denote the total residual noise on the 10 MHz common frequency as $\mathcal{L}_0(fm)$, the spectrum of the output signal at the frequency fo(M-N) is given by equation (19) below:

$$\mathcal{L}o\,(fm)\ OUT = \mathcal{L}o\,(fm)\,[M^2 + N^2 - 2MN\cos\omega m\Delta\tau\,] \qquad (19)$$

This analysis is somewhat oversimplified since a distinct value of $\Delta\tau$ is difficult to define for the equivalent M = 7-8, but it comes within a few db of predicting actual performance. The important observation is that in the worst case the maximum amount by which the output spectrum can be "bumped up" above $\mathcal{L}_0(fm)$ occurs when M = 8 and N = 7 (10 MHz out), and at an offset of fm = $1/(2\Delta\tau) \approx 50$ KHz in the 645A. The theoretical value of this maximum is given by $20\log_{10}(7+8) \approx 23.5$ db, and in practice measures only about 18.5 db, as will be described later in the section on overall performance. The net effect is that common noise is so effectively reduced that it does not degrade the overall instrument performance by any measureable amount.

Designing for Low Uncorrelated Noise

Now that it has been shown how the effects of correlated noise have been minimized, we should consider how the sources of uncorrelated noise have been minimized. The flicker of phase noise which is generated in all active circuits has been reduced as much as possible by the use of various negative feedback techniques, as recommended by Halford et al[3]. The noise floors have been held as low as possible to avoid degrading the final output. There are several factors which

determine how low the noise floors can be held. The system design has been done in such a way that no worse than 9th order inband spurious products are encountered. This means that higher power levels can be used, thus improving the noise floor. State-of-the-art mixers are used to obtain the best possible spurious performance, which also allows for operation at higher power levels. The noise figures of all the mixers and amplifiers have been optimized to reduce additive noise, State-of-the-art filters have been used to minimize transmission losses where the outputs of mixers must be filtered directly to avoid spurious product generation in active devices. The LC oscillators used in the drift cancelling loop must also have very low noise to avoid degration of the output signal due to the differential time delay in the UHF section. Each of these factors will be discussed further in the descriptions of the instrument.

Designing for Low Spurious Products

There are several basic problems to consider with regard to designing for low spurious product generation. The broad categories are; inband spurious products, out of band spurious products which must be removed by filtering, and spurious products introduced by packaging and cable routing problems. The circuit generated spurious products which are due to effects like mixer crossings, harmonic contamination in multiplier outputs, and local oscillator feed through at the mixer outputs, can be analyzed through the use of normal electrical engineering design techniques. The spurious signals which are generated due to packaging problems of signal leakage from circuit modules by radiation or conduction, or by leakage from RF cable runs, are much more difficult to analyze. Let us consider these two problems in more detail.

Circuit Generated Spurious Products

The two most important factors with regard to circuit generated spurious products which must be considered in the overall system design of the synthesizer are; inband spurious products, and the problem of local oscillator feed through to the output of a mixer. The basic compromise to be made directly involves the order of the in-band mixer crossing, and the difficulty of filtering out the local oscillator feed through. For example, let us examine the case where we have F_L = 3-4 MHz and $F_H \approx 37$ MHz. The output band we wish to generate then covers 40-41 MHz. Notice that the worst in-band spurious product is of order Fo/F_L (where F_L is the highest value used, 4 MHz in this case) or $40.5/4 \approx 10$. Note that the 40-41 MHz filter must reject the local oscillator frequency of 37 MHz, which is 3 MHz below the lower band edge of the filter. Now suppose we use $F_H \approx 38$ MHz, and F_L = 2-3 MHz, to generate the same output band of 40-41 MHz. Now the in-band crossings are of the order $40.5/3 \approx 13$, but the local oscillator feed through is only 2 MHz below the lower band edge of the filter, and is very difficult to filter.

The above example illustrates the major design compromise which must be made. The exact frequencies chosen depend on the state-of-the-art in mixers and filters, not only in the 30-70 MHz range of the low frequency circuitry, but also in the 370-420 MHz range where the UHF section operates. The choices of the various frequencies to be provided to the UHF section, and the frequency that the UHF section is to be operated at, are governed by the identical design compromise. The entire system is interwoven, so that the choice of a frequency in the low frequency section affects the entire system design clear through the UHF section. The overall system design of the 645A has been done in such a way as to obtain no worse than 9th order in-band crossings, and no local oscillator signal closer than ≈ 5 half bandwidths away from the band center of an output filter.

Packaging for Low Spurious Products

The fundamental problems in packaging electronic hardware to avoid cross coupling between various circuits boil down to two primary concerns; how to enclose the circuits themselves, and how to arrange for the minimum usage of physical interconnections between the various circuits. Once these problems have been dealt with, we must face the associated problem of arranging the various modules physically to avoid putting circuits which are sensitive to particular frequencies near any sources of such frequencies. Additional very difficult requirements arise when one is required to fit all the resulting hardware into the smallest possible overall physical space. The 645A was designed to fit in less than one half the rack height of previous synthesizers operating in the same frequency range, and at the same time achieve lower non-harmonic spurious outputs. The combination of these two design goals puts very severe requirements on the packaging techniques used.

One of the important keys to the attainment of the desired goals is the type of module packaging used.

An example of the technique used in the 645A is shown in Figure 12. This module has two symmetrical lids which are clamped down hard into the grooves in the sides by means of screws and spacers (not shown). In very critical applications, RFI gasketing is added down inside the grooves, the lid being designed in such a way as to compress the gasketing the amount recommended by the vendor. Ground plane connection is maintained to the boards all around the edges by means of a lip beneath the board, to which the board is clamped by means of the threaded spacers and screws used to hold the lids on the module. Notice that all inputs and outputs to the modules are by means of coaxial RF connectors, thus making all these inputs and outputs easily accessible as test points, and preventing radiation. Also notice that all DC connections to the module are made by means of feed through capacitors to prevent RF leakage onto the power supply system.

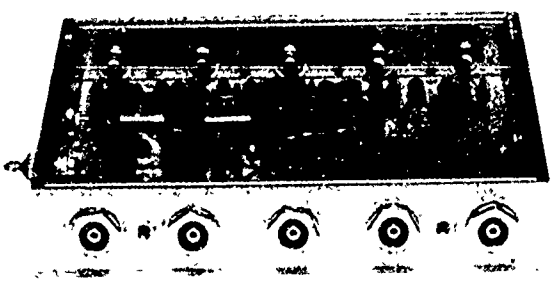


FIGURE 12

Example of module packaging used in 645A.

The physical arrangement and interconnection between the various modules is shown in Figures 13 and 14. Notice that there are relatively few coaxial interconnections, and that the physical arrangement of the modules is exceptionally neat and compact. The modules achieve sufficient integrity so that they can be placed very close together throughout the instrument. All of the circuitry has been placed in the three trays illustrated by the top and bottom views shown in Figures 13 and 14. Note that the UHF tray and the input section tray shown in Figure 14 swing up on hinges and lock, so that easy access is provided to all the interior DC interface areas.

One of the most important parts of the package design which is not immediately obvious is the division of circuit functions within the synthesizer. Early in the program we discovered it was not advisable to attempt more than 70 db of isolation between signals inside of

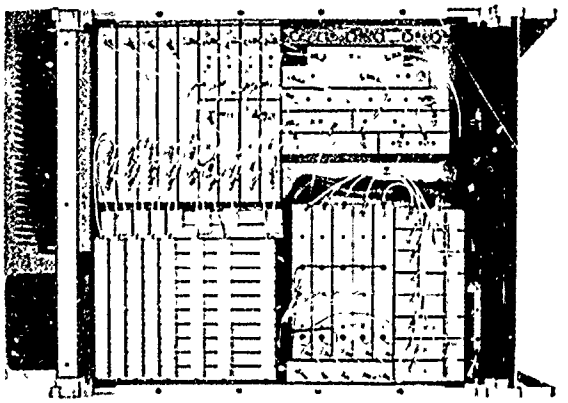


FIGURE 13

Sub-Synthesizer and Fixed Frequency Generator Section of 645A (Top View)

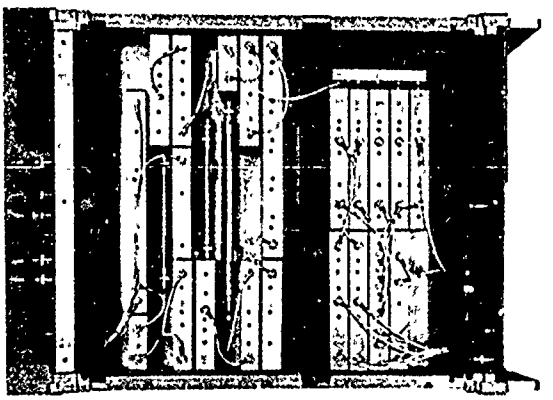


FIGURE 14

Input Section and UHF Section of 645A (Bottom View)

any one module, due to wave guide below cutoff modes and ground loops. All of the critical functions were then packaged according to the necessary electrical specifications in the appropriate manner, separating functions into different modules as necessary.

Theory and Performance of Input Section

The input section applies limiting and filtering to the 5 MHz signal, and then doubles it to 10 MHz (see Figure 15). The 10 MHz signal is then applied to the snap diode comb generator, and the 3rd to 7th harmonics are selected by means of harmonic selector modules (see Figure 8 for typical harmonic selector module circuit and performance). One of the 30-70 MHz outputs is selected by means of gates and serves as the 10 MHz digit input to the UHF section. The only spurious problems at this point are adjacent 10 MHz

harmonics, which are only down ≈ 80 db since an additional 50 db of clean up is provided by a 340 MHz filter in the UHF section. The 70 MHz signal is divided by two to obtain 35 MHz, and the 30 and 40 MHz signals are divided by 10 to obtain 3 and 4 MHz. Outputs are provided at 30, 35, 40, 3, and 4 MHz frequencies for use in the fixed frequency generators.

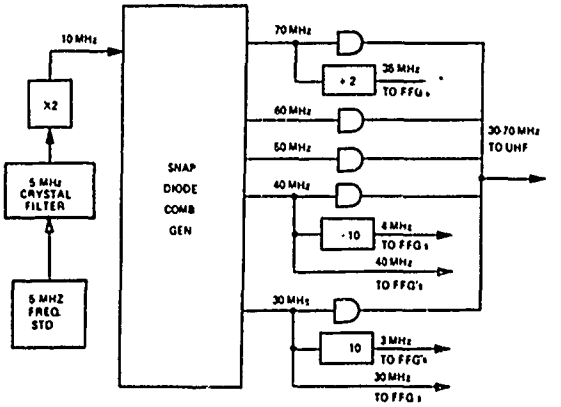


FIGURE 15

Input Section Block Diagram.

The residual noise present in the harmonic selector modules is negligible (see Figure 8) with respect to the multiplied common noise present on the 30-70 MHz outputs, so that there is a very high degree of correlation of the noise sidebands on all these outputs. For all practical purposes the performance of the synthesizer can be nicely predicted by considering the noise on all these outputs as completely common. This noise can easily be measured by driving two synthesizers with a common 5 MHz, and then measuring $\mathcal{L}(fm)$ using the two 70 MHz outputs from the input sections. The output spectra of the 30 MHz and 10 MHz output signals can be obtained directly from this measurement by dropping the 70 MHz curve by 7.5 db and 17 db respectively, as shown in Figure 16. The common noise of the 40-60 MHz outputs can be easily characterized in a similar manner.

Theory and Performance of Fixed Frequency Generators

The operation of the fixed frequency generators is very simple (see Figure 17). The 30 and 35 MHz signals are simply passed through clean up filters, while the remainder of the signals are generated by up and down converting with the various frequencies supplied by the input section. Buffered outputs are provided to drive the gates and associated circuitry in the sub-synthesizer, and a gated 30-39 MHz signal is provided for the 1 MHz digit information to the UHF section.

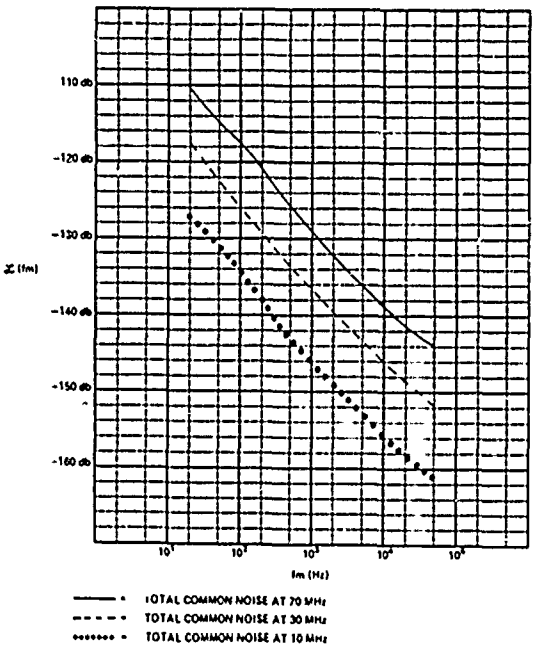


FIGURE 16

Input Section Common Noise

Due to the fact that all the input signals are harmonics of 1 MHz and are synthesized frequencies, the closest spurious signals possible lie at ±1 MHz from the carrier frequencies and are easily filtered. In order to provide a 30-39 MHz signal free of ±1 MHz spurious worse than 118 db down, it was necessary to split the

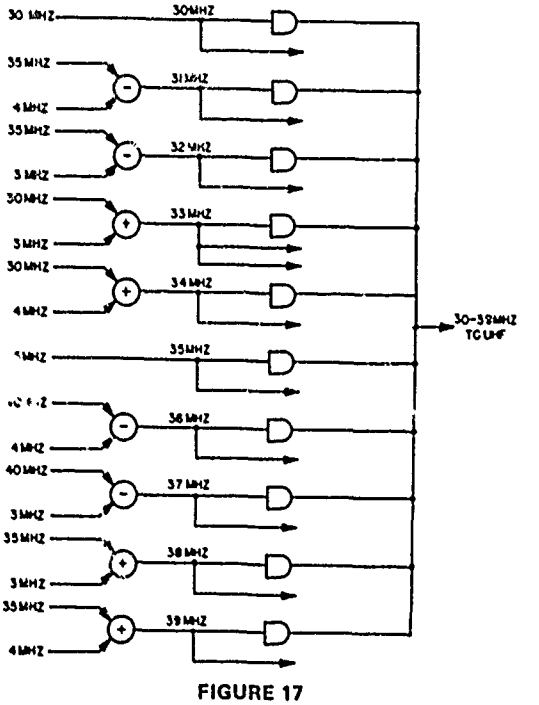


FIGURE 17

Fixed Frequency Generator Block Diagram

gate functions between the fixed frequency filter modules and a separate gate module. The gates which are off are gated off both in the individual filter modules, and in a separate gate module used to sum the outputs into the line going to the UHF section.

The residual flicker noise and noise floors encountered in the fixed frequency generator circuitry are low enough so that they do not contribute appreciably to the noise at the synthesizer output. Therefore, the common noise on the 30-39 MHz outputs can be predicted by the same techniques used in Figure 16, as previously described.

Theory and Performance of the Sub-Synthesizer

The sub-synthesizer and UHF sections involve the most complicated signal processing paths in the instrument, and hence are responsible for most of the uncorrelated residual phase noise present in the output signal. The sub-synthesizer also has some of the most stringent mixer and filter requirements in the instrument. The sub-synthesizer operates by an iterative mix and divide scheme, as illustrated by Figure 18. The frequency selection in each digit is accomplished by gating one of the 30-39 MHz inputs into the appropriate digit unit. Selection of the proper signal is done by means of an electrical 10 line code, which can be supplied either from the front panel or from an external programming source. Note that the gated signal is divided by 10 to provide a 3.0-3.9 MHz signal in 100 KHz steps, which is then used as the linear input to the second mixer in each digit unit. The division by 10 eases the spurious requirements on the digit control gates by 20 db, due to the spurious cleanup during the division process. Additional filtering is also done on the 3.0-3.9 MHz signal to help reduce any sidebands present due to gate leakage.

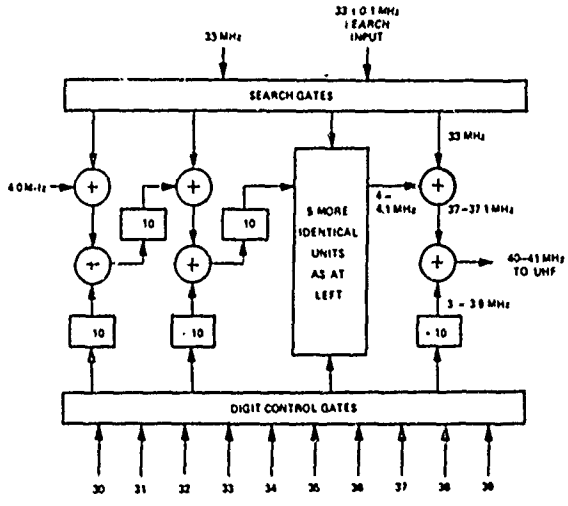


FIGURE 18

Sub-Synthesizer Block Diagram

The digit units all utilize the identical frequencies, but the output of the second mixer in the last digit unit is not divided by 10 in order to provide the 40-41 MHz signal to the UHF section. The actual manner in which the 100 KHz frequency increments are divided as the signal progresses from left to right through the sub-synthesizer is shown more clearly in Figure 19. The V above the digit is used in Figure 19 to trace the progress of a single digit as it passes through the sub-synthesizer, and is shifted by a factor of ten downward to a less significant decimal position each time it passes through a digit unit.

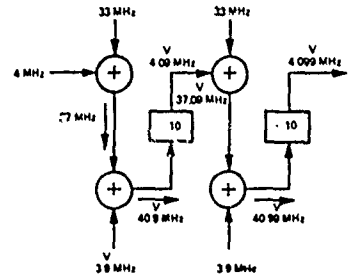


FIGURE 19

Principle of operation of sub-synthesizer.

The design requirements are the most stringent on the 100 KHz digit of the sub-synthesizer, and become less severe for the lower order digits. The first five digit units are identical, and are of relatively simple design. The 1 KHz and 10 KHz digits are of a more sophisticated design. The 100 KHz digit is more sophisticated still, and requires six separate modules in order to obtain the necessary performance. The hardest design problems were to keep the 9th order inband spurious products down in the first mixer, and to remove the 37.1 MHz local oscillator signal from the output of the second mixer. The linear product generation was kept down by the use of a very carefully designed FET balanced mixer, and the filtering was done by the use of two separate Cauer filters connected in series. A full 110 db rejection of the 37.1 MHz local oscillator signal is achieved by the filters alone, and an additional 25 db is provided by the balanced mixer.

The residual independent noise generated in the sub-synthesizer can be measured by using the same 33 MHz signal to drive the sub-synthesizers of two separate instruments, and dialing each instrument to XX.333... MHz by means of the front panel digit switches. The outputs from both sub-synthesizers are at 40.333... MHz when this is done, and $\mathcal{L}$(fm) can be measured by using these two signals as the mixer inputs in the test set up shown in Figure 4. The results of this test are shown as the solid line in Figure 20.

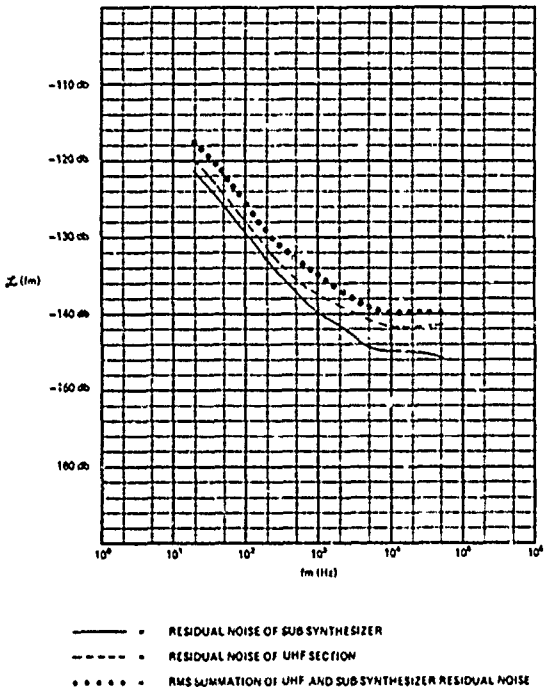
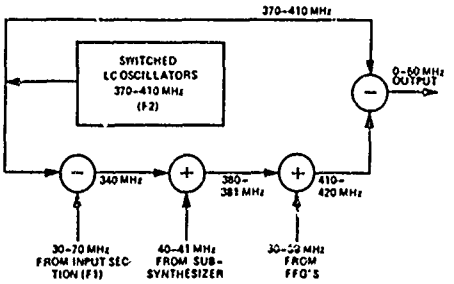


FIGURE 20

Total synthesizer independent residual noise.

Theory and Performance of the UHF Section

The method of operation of the UHF section can easily be understood by reference to the block diagram of Figure 21. The 30-70 MHz signal is switched simultaniously with the 370-410 MHz LC oscillators, each in 10 MHz steps, as is indicated by the table shown in Figure 21. This results in a constant IF frequency of ≈340 MHz out of the first hot carrier diode double balanced subtractive mixer. The carrier frequency, and also 10 MHz sidebands present on the 30-70 MHz input, are sufficiently removed from the mixer output by use of a narrow band 8 section Chebyshev tubular filter at the output of the mixer. The choice of frequencies for this mixer results in a fifth order spurious signal at 350 MHz when the 70 MHz input is used, but has the advantage of having no inband spurious products of less than 15th order. The 5th order spurious is also easily removed by the filter.



OUTPUT FREQ (MHz)	F1 (MHz)	F2 (MHz)
0-10	70	410
10-20	60	400
20-30	50	390
30-40	40	380
40-50	30	370

FIGURE 21

UHF Section Block Diagram

The second and third additive mixers are balanced varactor diode upper sideband up convertors. The second mixer has no inband spurious products less than 15th order which are not filterable, and the third has only an 11th order inband product which is easily held sufficiently low. The removal of the carrier signal at 381 MHz from the 410-420 MHz output signal of the third mixer also requires an 8 section Chebyshev tubular filter. The filter requirements are eased somewhat by good carrier balance and some passive filtering in the up-converter, and also by the very high linear signal levels which can be used without serious spurious product generation.

The output mixer of the UHF section is also a hot carrier double balanced modulator. This type of mixer was used here due to its wide bandwidth capabilities and DC coupled output port The spurious products which are encountered here are of the 15th order in the worst case, and are of the type $7F_H$-$8F_L$. The linear signal must be run at a fairly low level in order to keep the spurious products down, but even so the noise floor at the mixer output due to power level and noise figure considerations can be held at ≈ -147 db/Hz (sum of equal AM & PM noise in this case).

The residual noise performance can be measured by using a common 40 MHz signal at each of the three input ports to the UHF section, and switching on the 380 MHz LC oscillator. This causes 40 MHz to be

synthesized at the instrument output. The 40 MHz output is used as one input to the mixer of the test setup as shown in Figure 4, and the same 40 MHz signal used for the UHF section inputs is used for the other input. The residual noise due to the UHF section can then be calculated by use of equation (7), but the three db factor is not used since there is only one source of noise in this test. The results of this test are indicated by the dashed line shown on Figure 20. Note that the only observable degradation due to the switched LC oscillators occurs in the offset frequency region of 500 Hz to 10 KHz. The slope of ≈-7 db/decade here is due to the -27 db/decade slope of $\mathcal{L}(fm)$ exhibited by the LC oscillators, and the 20 db/decade improvement with decreasing carrier frequency of the common mode noise rejection due to the drift cancelling loop. This latter effect can be understood by letting M=N=1 in equation (15), and using a small angle approximation for the case when $\omega_m \Delta\tau << 1$. At modulation frequencies lower than 500 Hz the flicker of phase noise predominates, as was described in Section II.

A significant advantage of the UHF section over previous designs is the good control of output level which is attained without the use of ALC. The output level from the synthesizer is controlled essentially by the 370-410 MHz signal from the switched LC oscillators into the output mixer. This is true because this signal is used as the linear signal in the output mixer, and the conversion efficiency of the mixer is nearly independent of the LO signal level. The signal from the LC oscillators is passed through a class C overdriven amplifier and a resistive attenuator before being used as the linear signal in the output mixer, and hence the output level can be controlled quite easily by means of adjusting the current in the Class C overdriven amplifier. This method of operation makes possible several modes of output amplitude control, which can be selected by the front panel "OUTPUT" control (see Figure 22).

In the fixed 1 VRMS output mode of operation, DC signals from the LC oscillator module are used to adjust the linear signal level according to which 10 MHz digit is dialed up. The use of this technique means that one internal level setting can be made for each 10 MHz segment of output frequencies, so that at the final output the total level variation from 100 Hz to 50 MHz in this mode is typically held to ±0.4 db. Since the level control is attained without ALC and is very fast acting even when switching 10 MHz digits, the AM switching transients are greatly reduced. The output signal is within ±1 db of its final value less than 20 μs after switching occurs. An ALC mode of operation is also provided, with typical level control of ±0.1 db from 20 KHz - 50 MHz. Level control is provided from 0.2-1.0 VRMS in this mode by means of the front panel "ALC LEVEL" control, or linear voltage control from a remote DC source. Other modes of operation possible are remote leveling using an external detector, and amplitude modulation using an external modulation source.

Search Oscillator and Sweep Functions

An interpolation oscillator is provided which can be used to obtain continuous tuning of the output frequency over a range of ±1 digit, from ±100 KHz to ±0.01 Hz depending on the setting of the search oscillator controls. See the front panel and rear panel photographs of Figures 22 and 23 for details of the controls. The position which is controlled by the search oscillator can be manually selected by means of the decade selector switch on the front panel, or can be remotely selected if the decade selector switch is set to the "REM" position. Note the search digit interpolator controls (see Figure 22), which are also used to control this feature. With the search digit interpolator mode switch in the "LOCAL" position, control of the search is by means of the front panel "DIGIT INTERPOLATOR" dial. Accuracy of this front panel control is ±2% of one digit. With the mode switch in "REMOTE", the search oscillator frequency can be controlled by means of a ±10 VDC signal which can be applied at the rear panel "REMOTE SEARCH INPUT" connector. With the mode switch in "SWEEP", the search oscillator is swept back and forth with an internal linear saw tooth waveform over a range of ±1.0, ±0.5, or ±0.2 of the digit selected. The sweep voltage is available at the rear panel "SWEEP ANALOG OUTPUT" connector for use as the horizontal signal to an oscilloscope. The sweep rate is selectable in a 1-2-5 sequence from 0.01 sec to 50 sec by means of the "SWEEP TIME-SEC" switch.

The most important point to notice about the entire search oscillator feature is that the search or sweep function is always symmetrical about the dialed frequency. This is due to the manner in which the insertion of the search oscillator is achieved, which is illustrated in Figure 18. The search gates are simply used to substitute the search oscillator 33 ± 0.1 MHz signal for the fixed 33 MHz signal in the appropriate digit unit, thus leaving all the digit control gate inputs still active. This feature is very convenient for such

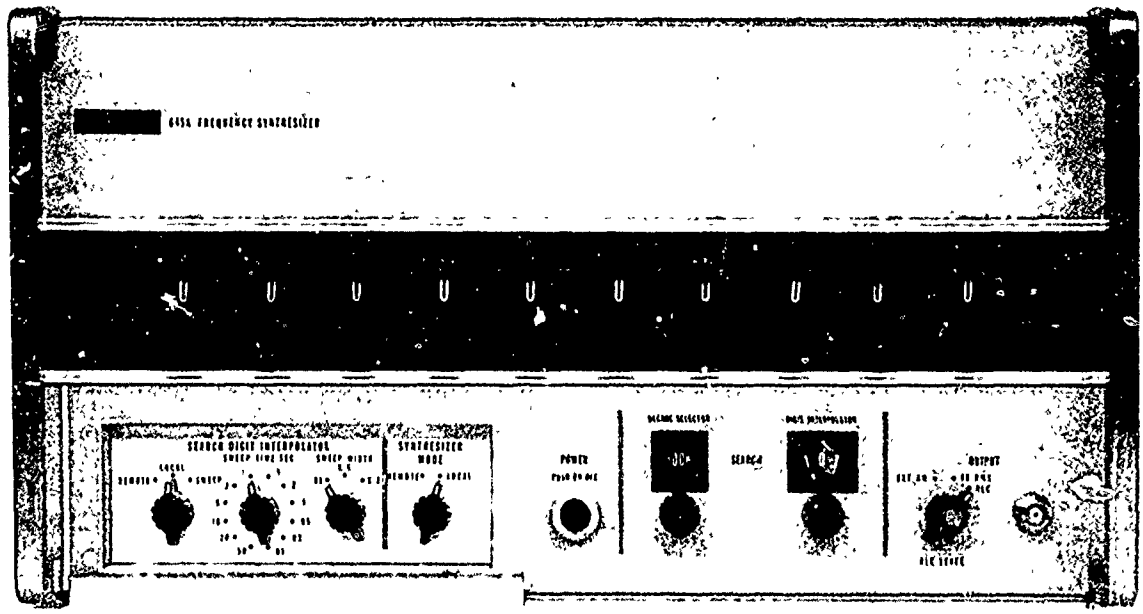


FIGURE 22

Front Panel of 645A Synthesizer

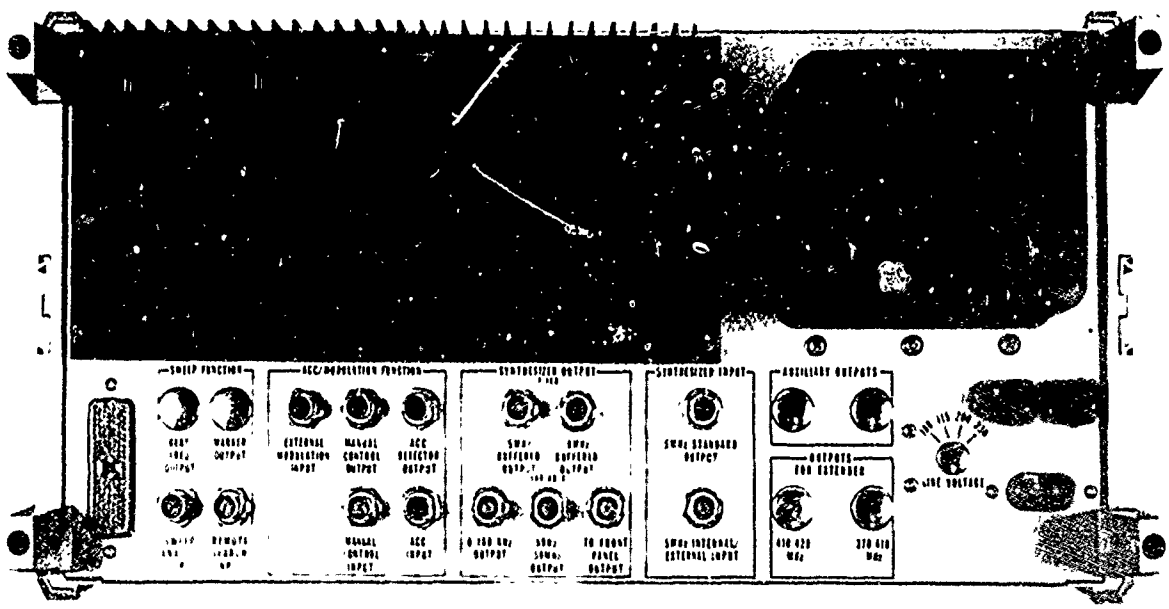


FIGURE 23

Rear Panel of 645A Synthesizer

applications as sweeping crystal filters. For example, if we wish to sweep ±10 KHz around the center frequency of a 1.234 MHz crystal filter, we simply dial up 1.234 MHz on the front panel digit switches and set the controls for a ±1.0 digit sweep in the 10 KHz digit. The 33 ±0.1 MHz signal for the search oscillator is obtained by adding a fixed 30 MHz to the output of a linearized VCO, which is tunable from 2.9-3.1 MHz by means of a ±10 VDC signal.

Overall Synthesizer $\mathcal{L}$ (fm) Performance

The overall residual noise of the synthesizer is easily measured using the technique shown in Figure 4. A common 5 MHz standard input is used for both units, and the final synthesized outputs are used to drive the mixer. The results are reduced by the use of equation 7, since the units under test were previously shown to be very nearly equal in performance. The results of this test for output frequencies of 49.9 MHz and 9.9 MHz are shown in Figure 24.

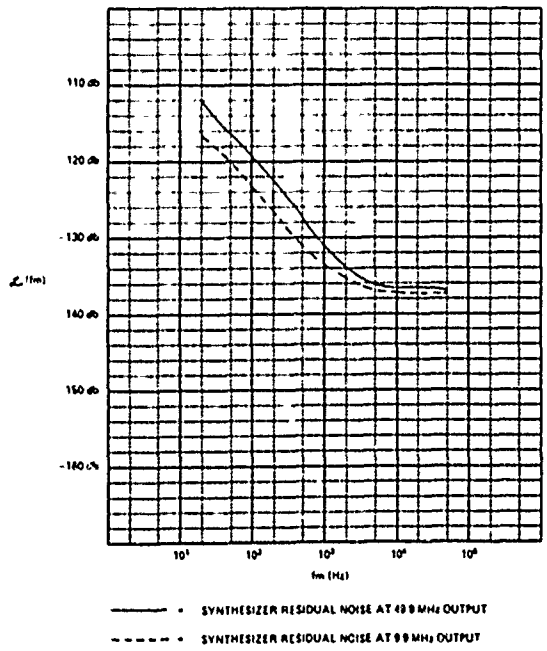


FIGURE 24

Overall synthesizer residual noise.

The performance at 49.9 MHz for offset frequencies fm < 5 KHz is set mostly by the multiplied common noise component. This can be seen by dropping the 70 MHz common noise plot shown in Figure 16 by 4 db (to simulate 50 MHz), and comparing it to the residual noise plot for 49.9 MHz as shown by the solid line on Figure 24. At 9.9 MHz output the performance for offset frequencies fm < 5 KHz is set primarily by the sum of the residual noises due to the UHF section and the sub-synthesizer. This can be seen by comparison of the 9.9 MHz output spectrum with the dotted line shown on Figure 20. For offset frequencies fm > 5 KHz the performance is set by the sub-synthesizer and UHF section residuals in all cases. Since the effect of the common noise is already negligible at 10 MHz, the plot of $\mathcal{L}$(fm) for lower output frequencies will remain essentially constant.

The amount of "bumping" of the common noise which occurs at the output can be predicted by means of a measurement of the transfer characteristics from the 10 MHz common frequency to the final output. A small fixed level of phase modulation is applied to the 10 MHz common frequency, and the amplitude of the resulting spurious sidebands at the synthesizer output is noted as a function of the offset frequency fm for a given output frequency. The number of db by which the spurious signal generated at the output exceeds the modulation put on the 10 MHz common frequency, is a direct measure of how much the common noise on the 10 MHz signal will be "bumped up" by differential time delay. The worst case occurs at ≈ 9.9 MHz output since M+N = 15 in this case, which is the largest value encountered. The common noise component at the output is shown in Figure 25 for this case. The plot was

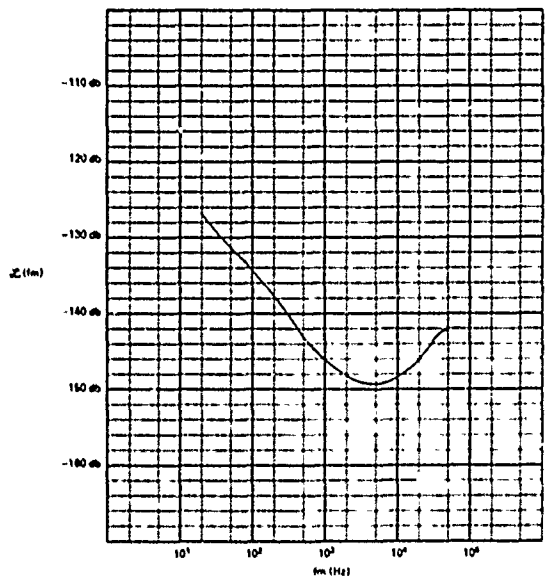


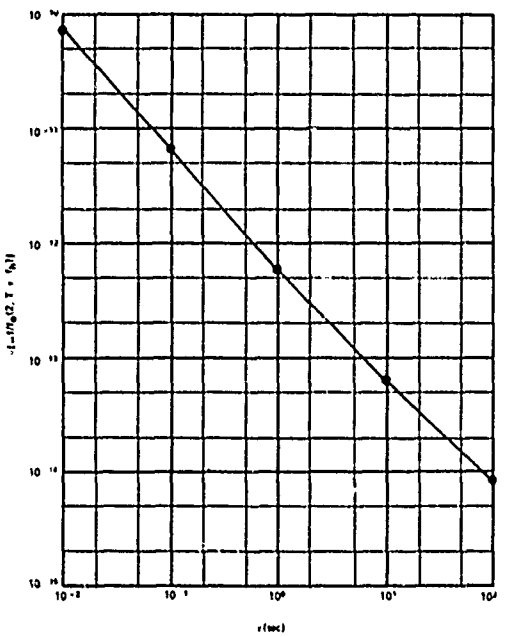
FIGURE 25

Worst case "bumping up" of the correlated noise at the synthesizer output. Occurs at 9.9999....MHz output.

obtained by adding the degradation measured at 9.9 MHz as just described, to the common noise at 10 MHz as shown in Figure 16. It can be seen from this plot that this component is ≈ 6 db below the noise floor at the peak of the "bump" which occurs at fm = 50 KHz, and hence adds very little to the output spectrum.

Residual $\sigma[\Delta f/fo(2, T, \tau, f_h]$ of Synthesizer

The standard deviation of the frequency fluctuations due to the residual noise of the synthesizer can be measured using the technique shown in Figure 7. The data taken, and the test conditions used for the measurement, are shown in Figure 26. Allan has shown that a slope of −1 on the plot of standard deviation versus average time τ, corresponds to a slope of −1 on a plot of $\mathcal{L}$(fm) or Sϕ(fm). Note that the data points shown on Figure 26 indicate very nearly a −1 slope for averaging times τ from 0.01 sec. to 100 sec. This means that the residual noise plot of $\mathcal{L}$(fm) continues at a slope of ≈ 10 db/decade down to offset frequencies at least in the magnitude of 0.005 Hz.



TEST CONDITIONS FOR DATA SHOWN				
f_o (Hz)	f_{fo} (Hz)	M	τ (sec)	T (sec)
100	10^7	1	0.01	0.5
10	10^7	1	0.1	0.5
1	10^7	1	1.0	2.0
1	10^7	10	10.0	20.0
1	10^7	100	100.0	200.0

FIGURE 26

Residual $\sigma[\Delta f/f_o(2, T, \tau, f_h)]$ for 645A at 50 MHz out.

APPENDIX I

Derivation of Expression for $\overline{\phi^2(t)}$

In these derivations the variable $\phi(t)$ represents the phase fluctuations (due to both devices under test) which are being translated to voltage fluctuations at the output of the mixer shown in Figure 4. For this reason the formulas developed in Appendix I represent the effects of the phase noise of both sources, and therefore must be modified as is discussed in the text to obtain expressions for the noise on just one device under test. Using the parameters Kϕ and A as defined in Figure 4, and denoting the output voltage from the phase detector as a function of time as Vx(t), we may write the following equations:

$$\frac{Vx(t)}{A} = \phi(t)\, K\phi \tag{20}$$

$$\phi(t) = \frac{Vx(t)}{K\phi A} \tag{21}$$

Squaring and taking the average value of each side of equation (21):

$$\overline{\phi^2(t)} = \frac{\overline{Vx^2(t)}}{(K\phi A)^2} \tag{22}$$

Physical Interpretation of $\overline{\phi^2(t)}$

Now let us show that $\overline{\phi^2(t)}$ is the total sideband noise to signal power ratio for a phase modulated signal. Taking the coherent sine wave modulation case for simplicity, the vector representation of a small index phase modulated signal shown in Figure 27 below is well known.[5] The peak deviation angle is the modulation index β, and the phase modulation is represented by the sum of the two small counter-rotating vectors each of peak magnitude $V_N/2$. The peak magnitude of the carrier voltage is Vs.

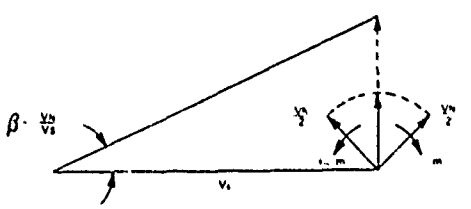


FIGURE 27

Vector representation of small angle phase modulation.

Now the total phase modulation sideband to signal power ratio is easily calculated (assume 1Ω load):

$$\frac{N\phi}{S} = \frac{\left(\frac{V_N}{2\sqrt{2}}\right)^2 + \left(\frac{V_N}{2\sqrt{2}}\right)^2}{\left(\frac{V_S}{\sqrt{2}}\right)^2} = \frac{V_N{}^2}{2V_S{}^2} \qquad (23)$$

substituting $\beta = \frac{V_N}{V_S}$, we obtain:

$$\frac{N\phi}{S} = \frac{\beta^2}{2} \qquad (24)$$

Now since the sinusoidal phase modulation has a peak value of β, it has a mean squared value:

$$\overline{\phi^2(t)} = \frac{\beta^2}{2} \qquad (25)$$

combining (24) and (25):

$$\overline{\phi^2(t)} = \frac{N\phi}{S} \qquad (26)$$

Derivation of Expression for $\mathcal{L}$(fm)

Now if we assume that the relationship of equation (26) holds for random as well as coherent modulation sidebands, we can derive an expression for $\mathcal{L}$ (fm). Let Vx(t, fm, 1 Hz) be the voltage in a 1 Hz bandwidth centered on fm which appears at the input of the analyzer of Figure 4, and let ϕ(t, fm, 1 Hz) represent the phase fluctuations corresponding to the same frequency band. Since $\mathcal{L}$ (fm) is defined in terms of the noise power in only one side band, equations (26) and (22) can be used to write the following expressions:

$$\mathcal{L}(\text{fm}) = \frac{\overline{\phi^2(t, \text{fm}, 1\text{ Hz})}}{2} \qquad (27)$$

$$\mathcal{L}(\text{fm}) = \frac{\overline{Vx^2(t, \text{fm}, 1\text{ Hz})}}{2(K\phi A)^2} \qquad (28)$$

Expressing $\mathcal{L}$(fm) in db/Hz when $F_N \neq 1$ Hz where f_N and Vx(fm) are defined in Figure 4:

$$\mathcal{L}(\text{fm}) = 20\log_{10}\left[\frac{Vx(\text{fm})}{K\phi\, A\sqrt{2}\sqrt{F_N}}\right] \qquad (29)$$

APPENDIX II

Let us derive the expression which relates the spectral density at the output of the system shown in Figure 9 of the text to the density of the input fo. For the derivation we will assume that the input signal fo to the two multipliers has a flat phase modulation spectral density of $\mathcal{L}_o$. If we assume the two multipliers XM and XN are ideal, then the output spectra will have the values M^2 $\mathcal{L}$o and $N^2\mathcal{L}_o$ respectively. We will assume that an ideal time delay is in series with each multiplier and that the time delay τ_M associated with the XM multiplier is greater than the delay τ_N of the XN multiplier. The difference between the two is $\tau_M - \tau_N = \Delta\tau$. We also assume that M>N. The expression will be derived for the small angle sinusoidal phase modulation case. Assuming the input signal is given by:

$$V(t) = A_o \sin(\omega_o t + \beta_o \sin \omega_m t) \qquad (30)$$

The input phase is then:

$$\phi_{IN}(t) = \omega_o t + \beta_o \sin \omega_m t \qquad (31)$$

We will use the substitutions shown below to account for the effects of τ_M and τ_N:

$$t = t - \tau_N \text{ at output of XN} \qquad (32)$$

$$t = t - \tau_M \text{ at output of XM} \qquad (33)$$

Now since the only effect of the ideal multiplier is to multiply the argument by the multiplication factor, we can write expressions for the phase out of the multiplier using the substitutions shown above:

$$\phi_N(t) = N\omega_o(t-\tau_N) + N\beta_o \sin \omega_m(t-\tau_N) \qquad (34)$$

$$\phi_M(t) = M\omega_o(t-\tau_M) + M\beta_o \sin \omega_m(t-\tau_M) \qquad (35)$$

The output phase can then be written as:

$$\phi_{OUT} = \phi_M(t) - \phi_N(t) \qquad (36)$$

$$\phi_{OUT} = \omega_o t(M-N) + \omega_o(N\tau_N - M\tau_M) + \beta_o M\left[\sin\omega_m(t-\tau_M) - \frac{N}{M}\sin\omega_m(t-\tau_N)\right] \qquad (37)$$

The first term of equation (37) represents the angular frequency of the output signal, and the second term represents a fixed phase shift. For the purposes of determining the effects of the system on the modulation sidebands, we need only consider the third term. Using the trigonometric identity shown in equation (38), the third term can be reduced to the form shown in equation (39). The angle δ is actually a complicated function of m,α, & β. The function δ has not been shown since it reduces simply to the form $\omega_m t+\phi$, and for the case at hand ϕ can simply be assumed as zero since it does not affect the result in which we are interested.

$$\sin \alpha + m \sin \beta = [1+m^2+2m \cos(\beta-\alpha)]^{1/2} \sin \delta \quad (38)$$

$$\phi OUT = \beta_o [M^2+N^2-2MN\cos\omega_m(\tau_M-\tau_N)]^{1/2} \sin \omega_m t \quad (39)$$

Now if we consider only the modulation term of the original input phase:

$$\phi IN (t) = \beta_o \sin \omega_m t \quad (40)$$

Combining equations (39) and (40) and substituting $\Delta\tau = \tau_M-\tau_N$, we can obtain the ratio of the magnitude of the output phase modulation to the input phase modulation:

$$\left|\frac{\phi(t)\ OUT}{\phi(t)\ IN}\right| = [M^2 + N^2 - 2MN \cos \omega_m \Delta\tau]^{1/2} \quad (41)$$

Now it can be seen that in the case of an input signal with a spectral density $\mathcal{L}$(fm) IN at a given offset $\pm$fm from the carrier, the ratio of the output spectral density has the same form as equation (41), except that it is squared since $\mathcal{L}$(fm) is a power spectral density.

$$\frac{\mathcal{L}(fm)OUT}{\mathcal{L}(fm)IN} = M^2+N^2-2MN \cos \omega_m \Delta\tau \quad (42)$$

REFERENCES

1. Dr. Donald Halford, Lectures at 1969 NBS Seminar on Frequency and Time Stability.
2. L. S. Cutler and C. L. Searle, "Some Aspects of the Theory and Measurement of Frequency Fluctuations in Frequency Standards", Proceedings of the IEEE, pages 136-154, February 1966.
3. Donald Halford, A. E. Wainwright, and James A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, & Cure", Proceedings of the 22nd Annual Symposium on Frequency Control, pages 340-341, 1968.
4. David W. Allan, "Statistics of Atomic Frequency Standards", Proceedings of the IEEE, pages 221-230, February 1966.
5. Mischa Schwartz, *Information Transmission, Modulation, and Noise*, pages 116-120, McGraw Hill, 1959.

LASER FREQUENCY STABILIZATION TECHNIQUES & ITS APPLICATIONS

H. S. Boyne
Radio Standards Physics Division
National Bureau of Standards
Boulder, Colorado
(Contribution of National Bureau of Standards, not subject to copyright)

Abstract

A review of progress in laser stabilization techniques and laser frequency measurement is given. Methods for relating laser frequencies to the time standard and methods for absolute laser frequency stabilization are described. Experimental information on reproducibility and noise characteristics is reported. Application to frequency and wavelength standards is discussed.

Introduction

Since the discovery of the He-Ne gas laser in 1960 by Javan, Bennett and Herriot,[1] many atomic and molecular laser transitions have been found. The possibility of using one of these transitions as a highly stable frequency or wavelength standard in the visible or near infrared has been envisioned, and many schemes have been developed to try to reproduce accurately the wavelength of one of these transitions.[2,3,4] The usual techniques involve servo-control of the laser cavity to some characteristic of the laser transition. The major problems involved with these attempts were large frequency shift due to gas pressure changes in the plasma, amplitude and frequency fluctuations due to plasma instabilities, and the dependence of the frequency on the current-voltage characteristics of the plasma discharge.

In 1967 there were two important developments in laser spectroscopy that pointed the direction necessary to realize accuracy competitive with frequency standards--saturated absorption of atoms and molecules first reported by Lee and Skolnick[5] and absolute frequency measurement of a laser transition first reported by Hocker et al.[6]

Saturated Absorption

The research of Lee and Skolnick on saturated absorption in Ne led others, namely Barger and Hall,[7] Hanes and Baird,[8] and Rabinowitz et al,[9] into the investigation of saturated absorption of molecular species having absorption transitions coincident with the laser transition. The molecular species that have been reported to date are I_2, CH_4, and SF_6. Molecular species are attractive because molecular vibration-rotation transitions have characteristically long lifetimes, absorption can be obtained from thermally populated states, and there is a much higher probability for coincidences of a molecular transition with a laser transition. The most extensive research to date has been done on I_2 and CH_4.

The principle of saturated absorption can be illustrated by referring to Fig. 1. A cell containing the absorbing gas is placed inside the laser cavity. The absorption transition whose Doppler width is greater than the natural line width interacts with the laser radiation such that absorption takes place at the laser frequency $f = f_o(1 \pm v_z/c)$ where f_o is the molecular transition frequency at rest and v_z is the speed of the molecule along the laser axis. If we think of the standing wave laser radiation pattern as composed of two oppositely directed running waves of equal amplitude, then two holes are "burned out" of the absorption density distribution as a function of velocity. The width of these holes is proportional to the natural width of the absorbing transition so that when the absorption is away from the central peak the two populations act almost independently. As the laser frequency is tuned to the absorption line center, however, the same molecules interact with both running waves and saturation occurs. This leads to an increased power output at frequency f_o. The emission peak at f_o has a characteristic width γ where $1/\gamma$ may be the collision induced lifetime, the natural lifetime, or the transit time of the molecule across the laser beam, whichever is the shortest characteristic time. The output intensity is $I = I_o[1 + \mathcal{L}(f - f_o, \gamma)]$ where $\mathcal{L}$ is a Lorentzian function and $f - f_o$ is the frequency measured from line center. Fig. 2 shows the

power versus frequency curve for saturated absorption of methane at 3.39μ. The absorption is that of a single transition which is the P7 line of the ν_3 vibration band in methane. The laser transition has been pressure shifted to higher frequency by approximately 100 MHz to make the centers of the P7 transition and the 3.39μ Ne transition coincide. The line width is approximately 200 KHz. The broadening of this line has been studied extensively in both pure CH_4 and in CH_4 with the rare gases as buffers. The width of the absorption in pure methane as a function of pressure is approximately 10 KHz per micron. By using special "frequency offset locking" described by Barger and Hall,[7] pressure shift of the methane transition can be investigated. This shift is anomalously small for reasons which are not clearly understood at this time. The shift as a function of pressure is given by Barger and Hall[7] as 75 ± 150 Hz per micron. Two independent methane stabilized lasers have been investigated and a stability plot of the difference frequency between the two lasers as a function of time is shown in Fig. 3. Recent improvements on this stability have been made and will be reported elsewhere.[10] More recent stability measurements show that the methane stabilized laser may be competitive with the hydrogen maser.

Other experiments dealing with the saturated absorption of I_2 at 633 nm and the saturated absorption of SF_6 at 10.6μ have been reported.[8,11] The power tuning curve for saturated absorption of I_2 is shown in Fig. 4. The lower trace in the figure is the derivative of the power curve which has an approximate Lorentzian shape. The several absorptions seen are due to the hyperfine structure in I_2. Fig. 5 shows the absorption of SF_6 from the P18 line of the CO_2 laser.[9] Because of the very high power output, the absorption can be seen with a cell outside the laser cavity in which the radiation is double-passed to achieve a standing wave pattern.

Frequency Synthesis

Table 1 shows the frequencies involved in attempting to establish a frequency locked chain from a stabilized klystron to the near infrared. The initial measurements in this chain were accomplished by Hocker et al[6] who mixed the 12th harmonic of a 74 GHz klystron with the 337μ HCN laser. The next step in the chain has been accomplished by Evenson et al[12] in mixing the 12th harmonic of 337μ with the 27μ water vapor laser. The most recent accomplishment has been to mix the 3rd harmonic of 27μ with the 9.3μ transition in CO_2 by Daneu et al[13] and the mixing of the 3rd harmonic of 27μ minus the 1st harmonic of 78μ to measure both the P18 and P20 transitions of CO_2 laser by Evenson et al.[14] All of these measurements have been accomplished by using a cat whisker diode detector. Fig. 6 shows an exploded view of the cat whisker which is usually made out of tungsten and is pointed by chemical etching techniques. The post which makes contact with the whisker can be either silicon as was the case for the first experiments by Hocker et al,[6] or can be some form of metal post as is the case for the more recent experiments.[13,14] The whisker acts as a long wave antenna as has been shown by Matarrese and Evenson.[15] Number of lobes of the antenna pattern is proportional to the length of the whisker divided by the wavelength of the radiation. Fig. 7 displays this pattern for $L = 7\lambda$.

Attempts are currently being made to mix the 8th harmonic of the 27μ H_2O vapor laser with the 3.39μ He-Ne laser. If this step can be accomplished, a frequency locked chain from the time standard to 3.39μ can be established and the stability of the saturated absorption at 3.39μ strongly suggests that this transition could be used as a new primary length standard. This being the case, the length standard and the time standard could be compared directly. If the primary length and time standard could be derived from the same atomic or molecular transition, the speed-of-light would become a defined quantity and length measurement could be derived from frequency measurements.

Acknowledgement

It is a pleasure to acknowledge cooperation given to me prior to publication by Drs. Barger, Hall, and Evenson in supplying data and helpful suggestions for this report. It is also a pleasure to acknowledge cooperation by Dr. Rabinowitz and Dr. Hanes in allowing photographs of their research to be published.

References

1. A. Javan, W. R. Bennett, Jr., and D. R. Herriot, Physical Review Letters 6, p. 3 (1961).

2. A. D. White, IEEE Journal of Quantum Electronic, QE-1, p. 349 (1965).

3. G. Birnbaum, Proceedings of the IEEE 55, p. 1015 (1967).

4. J. L. Hall, IEEE Journal of Quantum Electronics, QE-4, p. 638 (1968).

5. P. L. Lee and M. L. Skolnick, Applied Physics Letters 10, p. 303 (1967).

6. L. O. Hocker, A. Javan, D. Ramachandra Rao, L. Frenkel and T. Sullivan, Applied Physics Letters 10, p. 147 (1967).

7. R. L. Barger and J. L. Hall, Physical Review Letters 22, p. 4 (1969).

8. G. R. Hanes and K. M. Baird, Metrologia 5, p. 32 (1969).

9. P. Rabinowitz, R. Keller, and J. T. LaTourrette, Applied Physics Letters 14, p. 376 (1969).

10. J. L. Hall and R. L. Barger, to be published. (Also see: R. L. Barger and J. L. Hall, Proceedings of the 23rd Annual Symposium on Frequency Control, p. 306, 1969.)

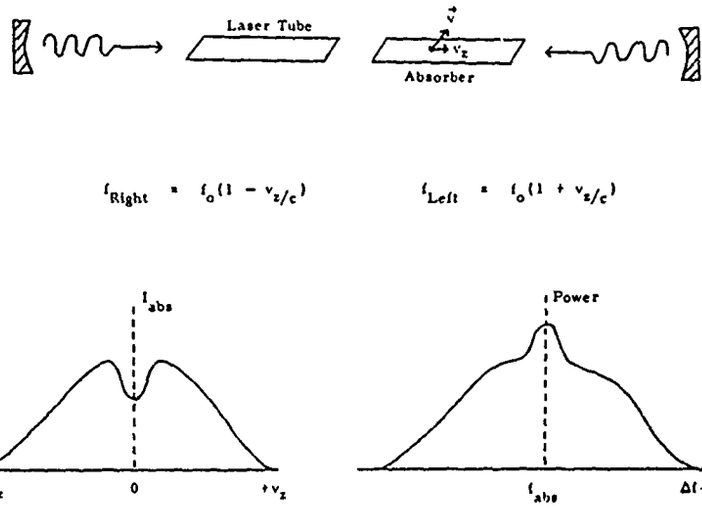
11. G. R. Hanes and C. E. Dahlstrom, Applied Physics Letters 14, p. 362 (1969).

12. K. M. Evenson, J. S. Wells, L. M. Matarrese, and L. B. Elwell, Applied Physics Letters 16, p. 159 (1970).

13. V. Daneu, D. Sokoloff, A. Sanchez, and A. Javan, Applied Physics Letters 15, p. 398 (1969).

14. K. M. Evenson, J. S. Wells, and L. M. Matarrese, Applied Physics Letters 16, p. 251 (1970).

15. L. M. Matarrese and K. M. Evenson, Applied Physics Letters, to be published.



SATURATED ABSORPTION

Fig. 1 - Saturated absorption explained as the interaction of two running waves with the absorbing medium.

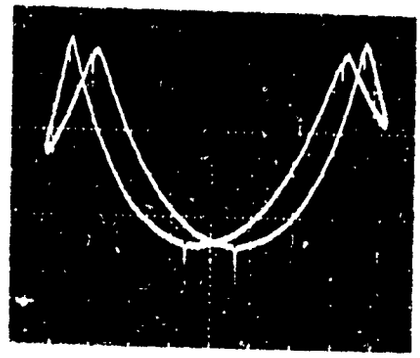


Fig. 2 - Saturated absorption in methane at 3.39μ.

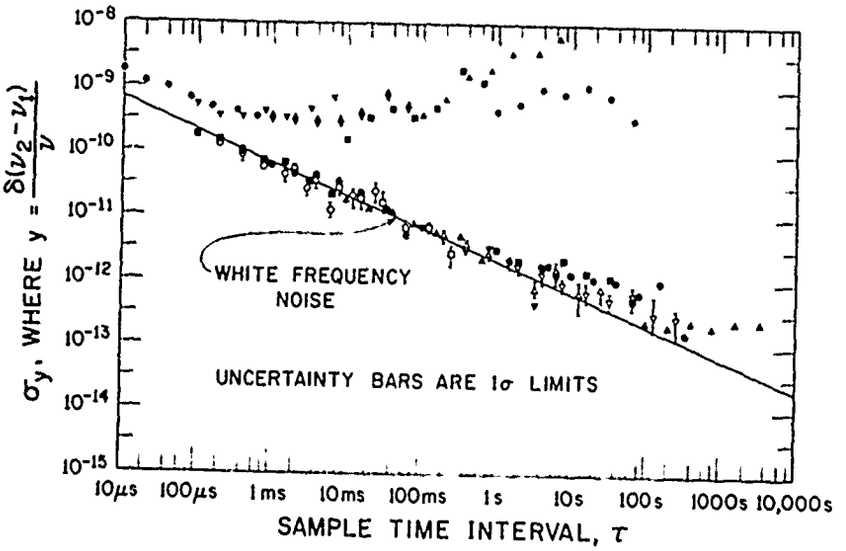


Fig. 3 - Plot of beat frequency stability between two independently stabilized methane lasers as a function of sample time.

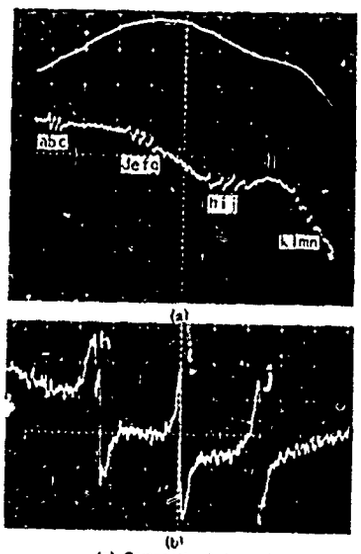


Fig. 4 - (a) Composite photograph showing laser power output versus tuning in the upper trace, and the derivative of power with respect to cavity length in the lower trace. Dispersion is approximately 50 MHz per division. Red is on the right. The labelled discriminant curves are due to the presence of iodine vapour in the laser cavity. (b) Enlarged view of three of the components at 5 MHz/division.

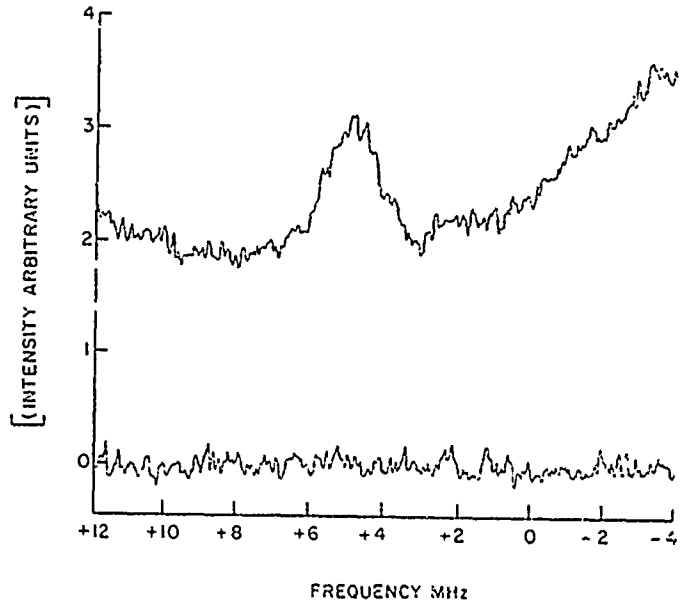


Fig. 5 - Saturated absorption of SF_6 for the P18 transition of the CO_2 laser.

Fig. 6 - View of whisker diode. Metal post is to the left of the whisker. Large vertical post is the mount for the figure.

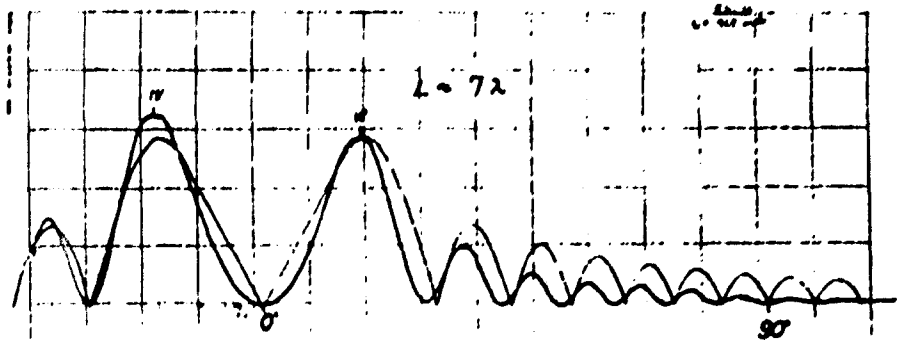


Fig. 7 - Antenna radiation pattern for length of wire 7λ. Solid line is experimental curve. -- line is theoretically predicted curve.

$$\nu_n = n\nu_1 + m\nu_2 + \delta\nu_3 \pm \nu_{beat}\ (30\ \text{MHz})$$

ν_n	λ_n	Laser n	Laser ν_1	Laser m	Laser ν_2	Klystron δ	Klystron ν_3
.9907686[A]	333μ					12	.0762
3.021775[B]	78μ	6	.091	-2	.009	3	.029
10.710073[B]	28μ	12	.091			1	.029
18.339000[B] 18.306251[B]	16.4μ 16.4μ	3 3	10.710 10.710	-1 -1	3.021 3.021	1 -1	.027 .026
32.176000[A] 32.134269[A]	9.3μ 9.3μ	3 3	10.710 10.710			1 -1	.032 .030
60.37637[B]	5.19μ	8	10.710	3	.091	-1	.040

Table 1 - Frequency multiplication steps necessary to achieve frequency synthesis to the near infrared. A--represents measurements done at MIT. B--represents measurements done at NBS, Boulder. B--proposed measurement.

Best Available Copy

AUTOMATIC FREQUENCY CONTROL AND PHASELOCKING OF LASERS

V. J. Corcoran, R. E. Cupp and J. J. Gallagher

Martin Marietta Corporation
Orlando, Florida

Summary

The automatic frequency control and phaselocking of the 890 Hz laser line of HCN to a multiplied frequency standard have been accomplished with a piezoelectric translator or a current control circuit as the control elements in a feedback loop. With a wide bandwidth phaselock loop, line narrowing has been accomplished. Long term stability of the beat of parts in 10^{12} and short time stability of parts in 10^{11} have been achieved.

Introduction

Frequency measurement of laser radiation in the submillimeter region has been accomplished by mixing laser outputs with harmonics of klystron outputs in point contact devices,[1-6] or by mixing laser outputs with fundamentals or harmonics of known laser signals plus klystron signals.[7-8] If the reference in these experiments is stable, then the beat between the laser frequency and the multiplied standard frequency can be stabilized to the standard. Stabilization is accomplished by feedback of a control signal that is generated as the laser drifts.

Automatic frequency control and narrow bandwidth phaselocking of a submillimeter laser have been accomplished using a piezoelectric translator or a current controller as the elements to control the laser frequency.[9-10] In these experiments, a significant improvement in the long term stability of the beat between the laser and the multiplied standard frequency was observed; however, no significant change in the linewidth was observed. With a wideband phaselock loop using a current controller line narrowing has been observed, indicating an improvement in the short term as well as the long term stability of the laser beat.

The experiments to achieve AFC and phaselocking are described in this paper. Also the results achieved on stability of the laser beat for the wideband phase locked laser are given.

General Stabilization Techniques

Before active stabilization can be achieved the signal-to-noise ratio of the IF beat between the millimeter harmonic and the laser must be sufficient (typically 10), and the frequency drift of the laser must be reduced to the point where the signal can be captured by the controlling equipment and held once control has been established. After the laser is passively stabilized and the signal-to-noise ratio is improved the IF beat is typically 2-10 kHz for the 890 GHz line of the HCN laser. The display of this beat on a spectrum analyzer is shown in Figure 1. This signal drifts slowly unless sudden perturbations occur in the system.

The automatic frequency control or phaselocking of this signal to a standard frequency can be accomplished with the aid of any device that can be used to vary the laser frequency. Two techniques that have been used employ a piezoelectric translator or a current controller. When one of the mirrors is attached to the piezoelectric translator, then a correction voltage can be used to change the length of the cavity which, in turn, changes the frequency of the oscillator. If a current controller is added to the laser power supply, then a correction current can be used to change the refractive index of the plasma which results in a frequency change. These techniques for the phaselock case are illustrated in Figure 2. For the case for the piezoelectric translator, the beat between the laser and the multiplied frequency standard is amplified and compared to a reference in a phase detector. The output of the phase detector is amplified and filtered. When the beat frequency is not in phase with the reference signal, then a correction voltage is fed to the piezoelectric translator to set the laser beat in phase with the reference. For the current controller, the filtered output from the phase detector goes to the current controller which is connected to the laser power supply so that any change in the phase detector output changes the current, which changes the frequency. For the automatic frequency control problem, the phase detector is replaced with a frequency discriminator and no reference frequency is used.

Stabilizaton Devices

Initial experiments in AFCing and phaselocking the HCN laser to a standard frequency used a piezoelectric translator on which the output mirror of the laser resonator was mounted. The displacement of the mirror needed to achieve the desired range of frequency variation dictated the use of high sensitivity devices. These high sensitivity devices were constructed of multiple ceramic rings that were mechanically in series and electrically in parallel or a single ceramic tube with the leads on the inner and outer wall. Each configuration resulted in a large capacitance for the device which reduced the theoretical resonant frequency and bandwidth of the device. Furthermore, these devices had unwanted mechanical resonances without electrical resonance below the theoretical resonant frequency of the device.

These factors led to a consideration of frequency control via the power supply current. The first device used a transistor in series with the laser tube and power supply. In this circuit any increase in current would produce a larger voltage drop across the unbypassed emitter resistor which would reduce the current through the transistor, thereby effecting the control. Because of the

lack of range of the transistor circuit, a vacuum tube circuit was used in the wideband loop which was needed to achieve a line narrowed phaselock system. The open loop bandwidth of the system under AFC conditions was greater than 30 kHz using the current controllers, whereas the bandwidth was approximately 1 kHz when the piezoelectric translators were used.

Automatic Frequency Control

The AFC of the HCN laser has been performed by beating the laser output (f_o= 890,758.368 MHz) with the 12th harmonic of an OKI 70 V10 klystron, oscillating at 74,234.864 MHz, to produce a 60 MHz IF signal. The 74 GHz oscillator is phaselocked to a harmonic of an X-band klystron which, in turn, is phaselocked to the harmonic of a 102 MHz crystal controlled oscillator. For accurate frequency measurements, the crystal controlled oscillator is monitored by a General Technology Model 304B Rubidium Frequency Standard, or it could be locked directly to the standard.

The 60 MHz beat signal between laser and klystron harmonic is amplified and a discriminator response is produced in a Micro-Now Model 201 Frequency Stabilizer. The output signal is amplified in a dc amplifier and applied to a piezoelectric element controlling the position of one of the laser mirrors. A Brewster angle window is constructed from polypropylene, and the output mirror, which is mounted on the piezoelectric element, is at atmospheric pressure. Under AFC a significant change in the characteristics of the 60 MHz beat is observed on a spectrum analyzer. Vibrational effects, which are evident in the unlocked free-running operation, are no longer present in the AFC mode. By tuning the microwave harmonic by varying the phaselock reference, it is possible to tune the laser frequency. A variation of the harmonic of the frequency standard shows no variation in the frequency of the IF, indicating that the laser is following the microwave signal when AFC is employed. No significant change is observed in the linewidth, which is on the order of 5 kHz when locked or unlocked. However, the signal remains well within the 5 kHz range for periods of several hours. This corresponds to a stability on the order of 5 parts in 10^9 for this period.

Because of the complications due to mechanical resonances and high voltages associated with the piezoelectric control, a current controller in series with the laser power supply was used to replace the piezoelectric translator as a frequency control. With the AFC using the current controller the output characteristics of the laser were essentially as before.

Phase Lock

Narrowband phaselocking has been accomplished with a piezoelectric translator as a control element or with a current control element in series with the laser power supply.

As in the AFC case, a significant change in the characteristics of the beat is observed on a spectrum analyzer. Vibrational effects, which are evident in the unlocked free-running operation, are no longer present when the laser is phaselocked. When tuning the microwave harmonic by varying the X-band phaselock reference, it is possible to tune the laser frequency. A variation of the harmonic of the frequency standard by as much as 20 kHz shows no variation in the frequency of the IF, indicating that the laser is following the microwave signal when phaselocking is employed. No significant change is observed in the linewidth, which is on the order of 2 kHz when locked or unlocked. However, the signal remains well within the 2 kHz range for long periods of time.

With a narrowband phaselock the beat and reference remain in phase only on a long term average basis, and during rapid fluctuations of the beat frequency the system skips cycles. The narrowband loop, however, does not respond to these rapid variations. As a result the long term stability is improved, but the short term stability is not. In order to improve the short term stability and the linewidth, a wideband phaselock that will follow rapid variations of frequency is necessary.

Wideband phaselock with a consequent line narrowing has been achieved with the setup diagrammed in Figure 3. The HCN laser output at 890,760.238 MHz is mixed with the 12th harmonic of a phaselocked OKI 70 V11 klystron operating at 74,232.520 MHz in a point contact device. The klystron is referenced to a 120 MHz crystal. The difference frequency is sent to a Schomandl FDS 30 phaselock. The output of the phaselock goes to a DC amplifier, then to a current controller which is is in series with the laser power supply. The current controller has an open loop bandwidth of more than 30 kHz under AFC conditions. As a result any change in the laser frequency relative to the microwave harmonics is compensated by a change in the laser current, which controls the laser frequency.

The effect of a wideband phaselock on the beat frequency is shown in Figure 4. The unlocked beat between the HCN laser and the multiplied signal shown in Figure 4 is approximately 5 kHz wide. When the laser is phaselocked, the beat narrows to the 50 Hz limit of resolution of the HP8553L spectrum analyzer as seen in Figure 4.

The long and short term stability of the device can be obtained by measuring the fractional frequency fluctuations as a function of averaging time as shown in Figure 5. For long term averages the data points are noncontiguous 100 pair averages because of the fact that equipment variations tend to cause the device to intermittently drop out of lock. The short term (.1 sec averaging time) stability is seen to be approxiamtely 1 part in 10^{11} and the long term (10 sec averaging time) stability is approximately 2 parts in 10^{12}.

Discussion

Further improvement in the phaselock is

necessary so that contiguous long term pair averages can be obtained. Also the stabilization is of the beat between the multiplied standard and the laser. The multiplied standard is known only by the beat between it and the free running laser to be from 2 to 10 kHz. To determine the width of the laser two phaselocked lasers must be beat together.

The stabilization of a laser to a standard frequency is significantly different from techniques that have been used to stabilize lasers. With the method used here, the reference is an absolute standard, not another unstable device; and no frequency modulation of the laser is used so that the signal is a narrow bandwidth signal with good long and short term stability.

This technique of phaselocking a laser to an absolute standard can be extended to any laser whose frequency can be measured by the techniques first used by Frenkel and his collaborators[1] and can be expected to be extended at least to the CO_2 laser lines in the future.

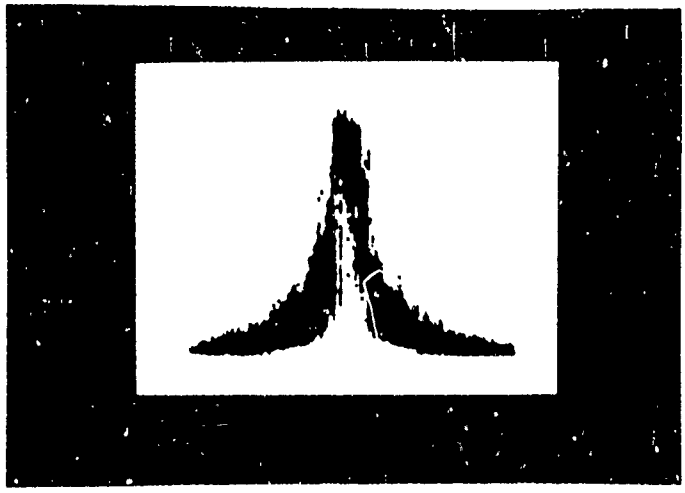
Absolute frequency measurement and control are expected to have applications where precise, stable sources are useful. The fact that the frequency of a laser can be measured accurately means that a velocity of light measurement limited only by the wavelength measurement can be made. The availability of precisely known frequencies in the optical region will result in accurate spectroscopic measurements on molecules that should aid in the understanding of the molecule structure. A knowledge of the frequency of optical sources that are stablized will result in a variety of metrology applications where high accuracy is desirable. Finally, as the optical communications techniques are developed, the laser stabilized to an absolute standard can be expected to be used as a reference in communication links as in satellite communication and as a transmitting source itself.

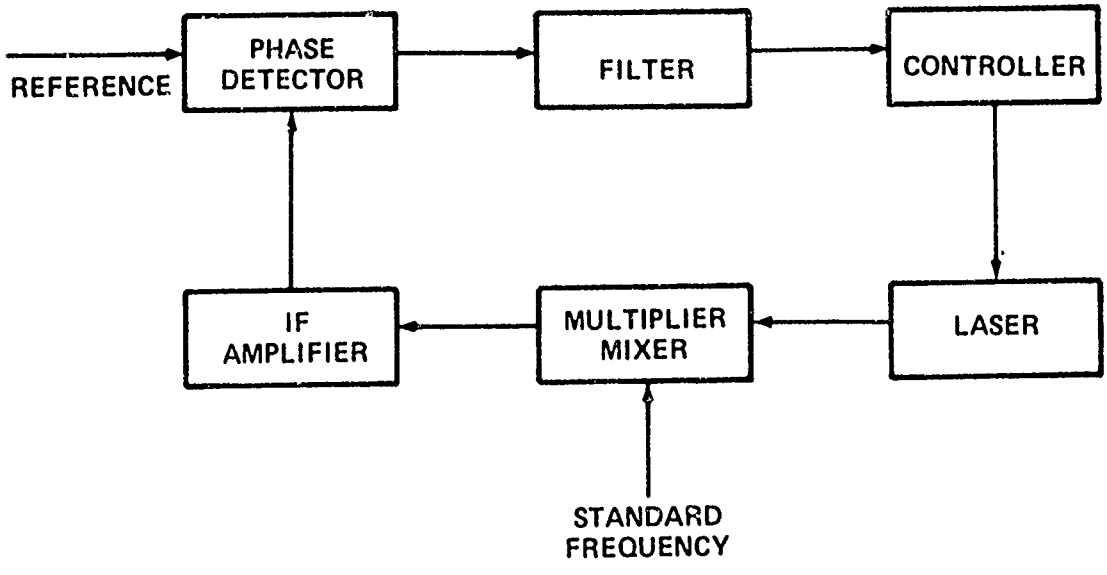
Illustrations

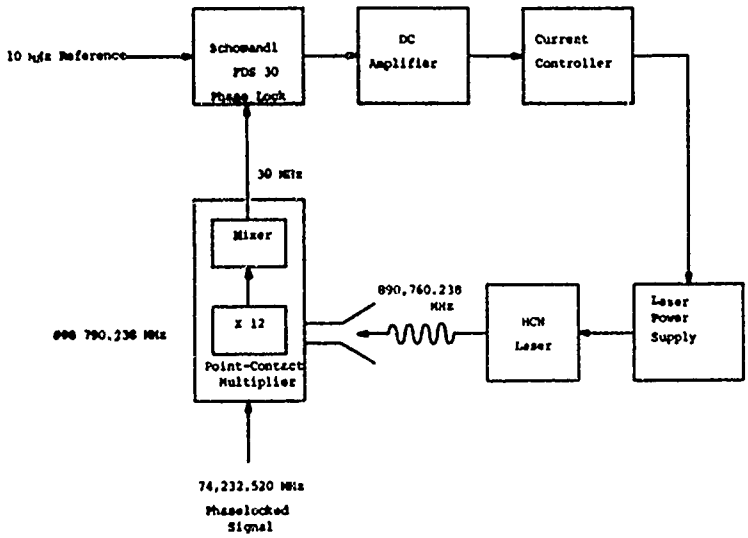
1. Spectrum Analyzer Display of Beat Between Laser and Multiplied Frequency Standard. Scan Width 40 kHz/cm; Bandwidth 30 kHz; Scan Time 10 m sec/cm.
2. Block Diagram of General Method for Phaselocking a Submillimeter Laser Signal to an Absolute Standard.
3. Wide Bandwidth Phaselock Loop Used to Stabilize 890 GHz HCN Laser Line.
4. Spectrum Analyzer Display of Beat Before and After Phaselock.
5. Fractional Frequency Fluctuations as a Function of Averaging Time for Noncontiguous 100 Pair Averages.

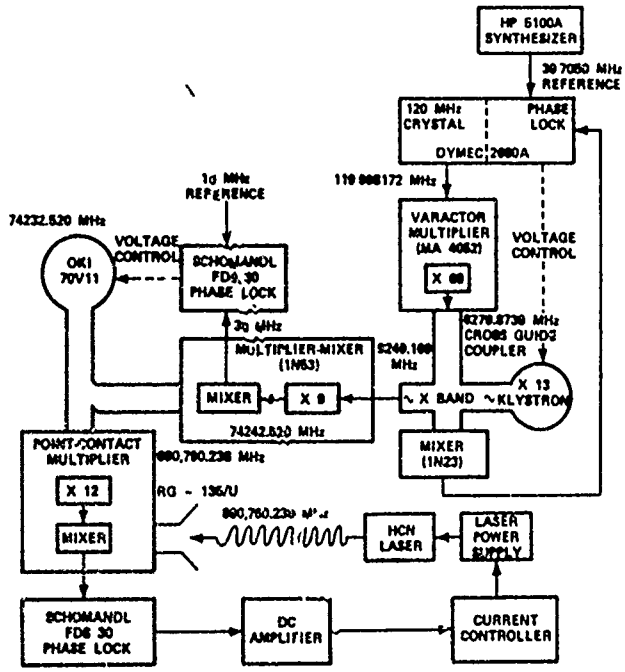
References

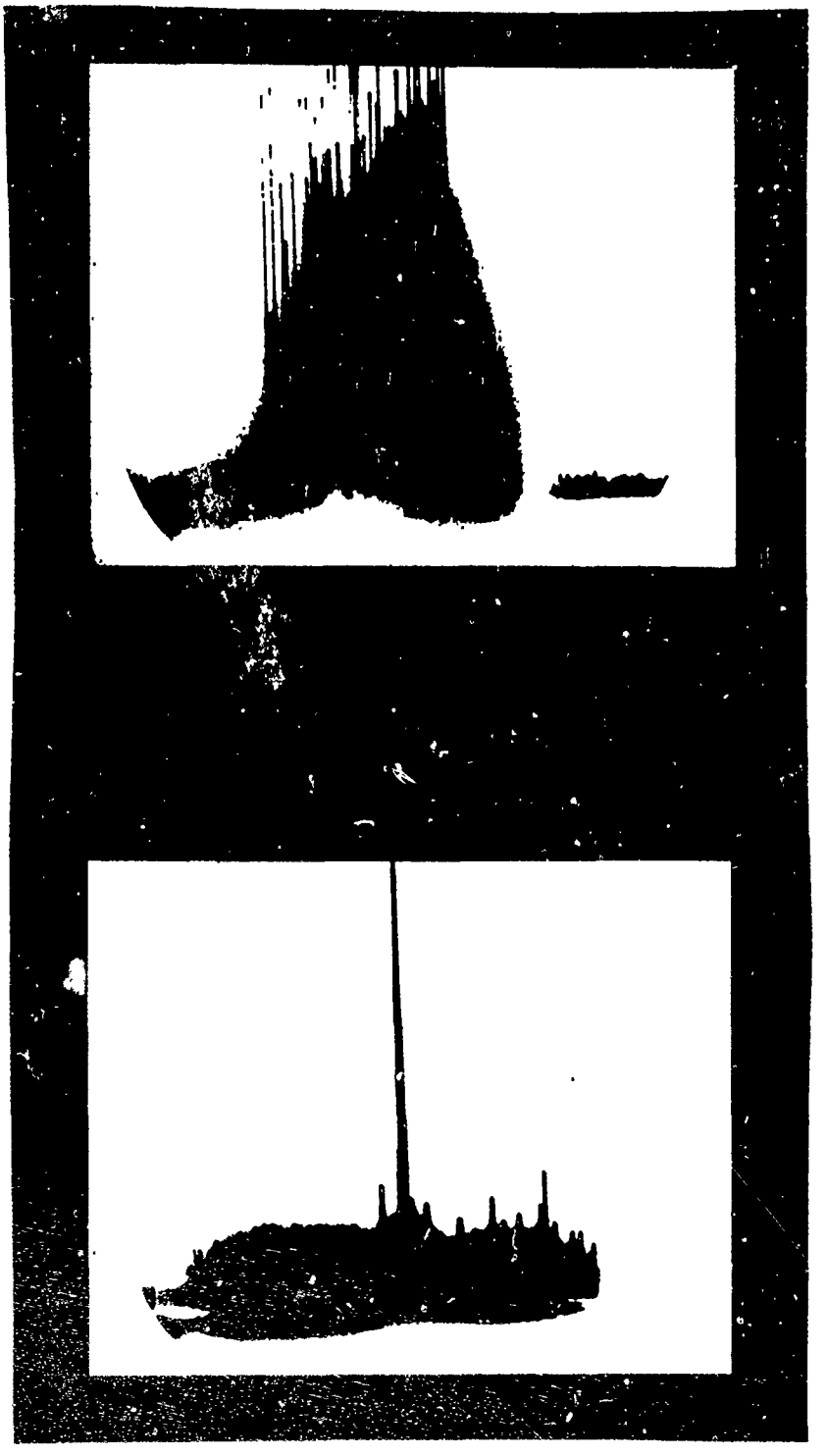
1. L.O. Hocker, et al, "Absolute Frequency Measurement and Spectroscopy of Gas Laser Transitions in the Far Infrared," Appl. Phys. Letters 10, 147 (1967).
2. L.O. Hocker, D. Ramachandra Rao, and A. Javan, "Absolute Frequency Measurement of the 190 μ and 194 μ Gas Laser Transitions," Phys. Letters 25A, 690 (1967).
3. L. O. Hocker and A. Javan, "Absolute Frequency Measurements on New CW HCN Submillimeter Laser Lines," Phys. Letters 25A, 489 (1967).
4. L. Frenkel, et al, "Absolute Frequency Measurements of the 118.6 μM Water-Vapor Laser Transition," Appl. Phys. Letters, 11, 344 (1967).
5. L. O. Hocker and A. Javan, "Absolute Frequency Measurements of New CW DCN Submillimeter Laser Lines," Appl. Phys. Letters 12, 124 (1968).
6. M.A. Pollack, L. Frenkel, T. Sullivan, "Absolute Frequency Measurement of the 200 μ Water Vapor Laser Transition," Phys. Letters, 26A, 381 (1968).
7. L.O. Hocker and A. Javan, "Laser Harmonic Frequency Mixing of 2 Different Far IR Laser Lines up to 118 μ," Phys. Letters, 26A, 255 (1968).
8. L.O. Hocker, J.G. Small, and A. Javan, "Extension of Absolute Frequency Measurements to the 84 μ Range," Phys. Letters, 29A, 321 (1969).
9. V.J. Corcoran, R.E. Cupp, and T.J. Gallagher, "Frequency Lock of the Hydrogen Cyanide Laser to a Microwave Frequency Standard," IEEE Journal of Quantum Electronics, QE-5, 424 (1969).
10. R.E. Cupp, V.J. Corcoran, and T.J. Gallagher, "Phaselock of a Far Infrared Laser to an Absolute Frequency Standard," (Submitted to IEEE Journal of Quantum Electronics).

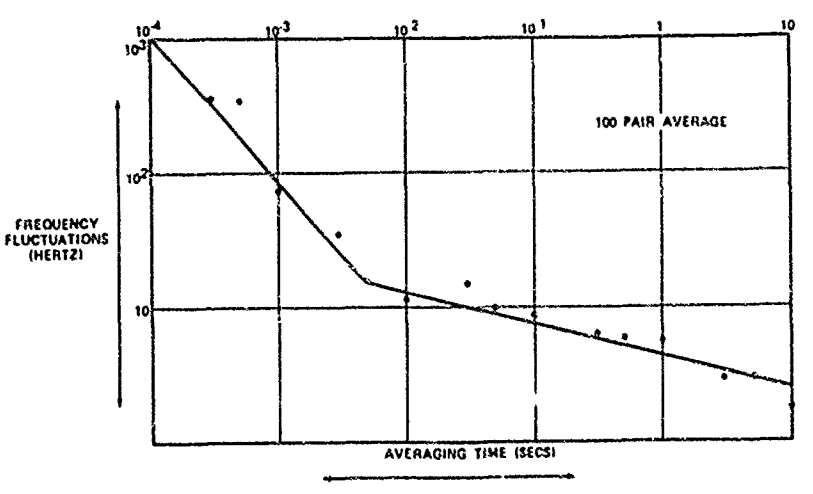



REFERENCE
PHASE DETECTOR
FILTER
CONTROLLER
IF AMPLIFIER
MULTIPLIER MIXER
LASER
STANDARD FREQUENCY


10 kHz Reference
Schomandl
FDS 30
Phase Lock
DC
Amplifier
Current
Controller
30 MHz
Mixer
X 12
Point-Contact
Multiplier
890,760.238 MHz
890 790.236 MHz
HCN
Laser
Laser
Power
Supply
74,232.520 MHz
Phaselocked
Signal


HP 5100A
SYNTHESIZER
30 7050 MHz
REFERENCE
120 MHz
CRYSTAL
PHASE
LOCK
DYMEC 2980A
10 MHz
REFERENCE
119 986172 MHz
74232.520 MHz
VOLTAGE
CONTROL
OKI
70V11
SCHOMANDL
FDS 30
PHASE LOCK
VARACTOR
MULTIPLIER
(MA 4062)
X 60
VOLTAGE
CONTROL
30 MHz
MULTIPLIER-MIXER
(1N53)
MIXER
X 9
74242.520 MHz
8278.8738 MHz
CROSS GUIDE
COUPLER
X 13
X BAND KLYSTRON
MIXER
(1N23)
POINT-CONTACT
MULTIPLIER
X 12
MIXER
890,790.238 MHz
RG - 136/U
HCN
LASER
LASER
POWER
SUPPLY
SCHOMANDL
FDS 30
PHASE LOCK
DC
AMPLIFIER
CURRENT
CONTROLLER




10^{-4}
10^{-3}
10^{-2}
10^{-1}
1
10
10^3
10^2
10
100 PAIR AVERAGE
FREQUENCY FLUCTUATIONS (HERTZ)
AVERAGING TIME (SECS)

AREAS OF PROMISE FOR THE DEVELOPMENT OF FUTURE PRIMARY FREQUENCY STANDARDS

Helmut Hellwig
Atomic Frequency and Time Standards Section
National Bureau of Standards
Boulder, Colorado 80302

Abstract

A primary frequency standard is based ultimately on some fundamental property of nature. The actual realization of an output frequency involves several steps of physical and technical processing. Any of these steps may cause a shift of the output frequency of the standard which can be measured or calculated or both. The resulting corrections are called the biases, and each one has associated with it some uncertainty. The combined uncertainty of all biases is referred to as the accuracy capability of the frequency standard. In this paper possibilities are discussed which may lead to the development of future primary frequency standards of superior accuracy capability. As the reference in performance today's most advanced primary frequency standard, the laboratory type cesium beam tube, is used. Aspects of costs and field-usage are totally neglected.

A review is given of the various methods and techniques which are currently employed in quantum electronic frequency standards or which have a potential usefulness. Various effects which influence the output frequency of a primary standard are associated with these methods. A classification is given:

(a) Effects associated with the interrogation of particles (atoms or molecules)

(b) Effects related to the method of confining the particles

(c) Effects associated with the particles themselves and the way in which they are treated for an effective interrogation by electromagnetic radiation.

These three classes of effects are discussed in detail, and expectation values for the related uncertainties are given. For selected particles certain methods of interrogation, confinement, and particle preparation can be combined such as to minimize the net uncertainty due to all applicable effects. Different technical solutions are the result. A review of existing and proposed devices is given, including quantitative data on the stability and accuracy capability. Aspects of the most promising devices are discussed, and it is concluded that accuracy capabilities of 10^{-14} should be within reach of today's research and development.

Introduction

A primary frequency standard is based ultimately on some fundamental property of nature. The actual realization of an output frequency involves several steps of physical and technical processing. Each of these steps may cause a more or less pronounced shift of the output frequency of the standard. The magnitude of each shift must be evaluated theoretically and experimentally. The resulting corrections are called the biases; however, they are never known exactly but have associated with them uncertainties. The magnitudes of these uncertainties depend on the degree of theoretical understanding, on experimental parameters, and on their measurability. The combined uncertainty of all biases is referred to as the accuracy capability of the frequency standard.

At present the standard with the highest accuracy capability is the cesium atomic beam tube. This fact is reflected in the agreement at the 13th General Conference on Weights and Measures (1967) to define the second as the duration of 9192 631 770 cycles of the unperturbed radiation of the Cs^{133} hyperfine transition $(F = 4, m_F = 0) \leftrightarrow (F = 3, m_F = 0)$. The cesium beam tube, although it still can be improved, is already at a very advanced stage of technical development. An accuracy capability of the order of parts in 10^{13} is now a reality which is achieved in several laboratories around the world.[1-5] However, a natural demand exists for an even more accurate definition, and frequency standards of extreme precision are being needed in various technical applications. New aspects of scientific, technical, and metrological importance have evolved because of recent successes in multiplication of frequencies into the infrared region of the electromagnetic spectrum.[6-10] A substitution of conventional length (wavelength) measurements by much more precise and accurate frequency measurements will soon be possible and may ultimately lead to a unified (frequency) standard

for frequency, time, and length with the speed of light being a defined constant. We may look even further ahead to the possibility that other basic physical quantities, e.g., the volt via the Josephson effect, will be based on the same (frequency) standard. This illustrates the far reaching importance which we have to attribute to the development of primary frequency standards.

In the following we will compare and discuss possibilities in methods and techniques which are at our disposal for pushing the fractional frequency accuracy capability beyond 10^{-13}. This will be done in a way which differs somewhat from the usual approach of comparing different existing devices.[11] Instead we will compare the methods and techniques used in quantum electronic frequency standards, and we will try to synthesize from these an optimized device. As the criterion for excellence in this comparison it appears best to choose accuracy capability as the prime topic in our discussions. The influences affecting the frequency of a quantum electronic frequency standard can be grouped into three classes: (1) Effects associated with the interrogation of the atoms or molecules, (2) effects related to the method of confining the particles, and finally (3) effects associated with the particles themselves and with the way in which they are treated for an effective interrogation by electromagnetic fields.

We will discuss these three groups successively in more detail. Since we are aiming for accuracies of better than the present state-of-the-art (parts in 10^{13}) we will discard as unimportant only those effects which lead to fractional uncertainties of less than 10^{-14}. At this point it should be emphasized that the following discussion is based on our present state of knowledge. There might well be other methods and techniques which could be used, and there might be additional sources of uncertainty of which we are not yet aware.

Discussion

1. Effects on Frequency: Particle Interrogation

In a quantum electronic frequency standard we are not only concerned with the actual quantum transition of the particles involved but also with the production of an output signal. In general two different tasks are to be accomplished: (a) The generation of a signal which is to be compared with the actual "clock" transition, and (b) the generation of a useful standard frequency output. It is obvious that at least one oscillator--in general, more than one--is necessary for these tasks. These slave oscillators may be quartz crystals, klystrons, lasers, or any other suitable devices. They must be locked in frequency or in phase to the "clock" transition. This can be accomplished by four distinctly different methods: The active maser oscillator,[12-15] whereby its output frequency is compared to a synthesized frequency derived from the slave oscillator; the absorption[16-18] or amplification[12] of an electromagnetic frequency by the atomic or molecular transition; the detection and counting of the number of particles having undergone a transition caused by a signal which is derived from the slave oscillator;[1, 19, 20] and finally the observation of transitions by their effect on other transitions of different frequencies to which the "clock" transition is coupled[21] (frequency transformation). In the first column of Table 1 the different interrogation methods are summarized. The second column lists various effects which give rise to bias uncertainties and which are associated with one or more of the methods of interrogation.

The fractional frequency stability[22] of a quantum electronic frequency standard may conveniently be written as

$$\sigma_y \cong \frac{K}{M} \frac{1}{\sqrt{\tau}} \quad . \tag{1}$$

The symbols in (1) have the following meaning: τ is the measurement interval; M is a basic figure of merit* given by

$$M \equiv \sqrt{n_s}\ T_r \quad , \tag{2}$$

where $T_r \equiv$ relaxation time, and $n_s \equiv$ flux of signal particles. The constant K contains the device characteristics. As examples we have for photon detection (with $h\nu \gg kT$) and for particle detection[20, 22]

$$K_{\text{passive}} \gtrsim \frac{1}{2\pi\nu} \quad , \tag{3}$$

and for an active oscillator (with $h\nu \ll kT$)[12, 22] we have

$$K_{\text{active}} \gtrsim \frac{1}{\sqrt{2}\pi\nu} \sqrt{\frac{kT}{h\nu}} \quad . \tag{4}$$

*This M differs from traditional figures of merit found in the literature.

Table 1

Effects on Frequency: Particle Interrogation

Methods		Effects
Active:	Self-Oscillation	Noise in:
Passive:	Absorption or Amplification of Radiation	Detector
		Receiver
	Particle Detection	Resonance Structure
	Frequency Transformation	Multiplier and Synthesizer
		Quantum Transition
		Slave Oscillator Fluctuations
		Spectrum Asymmetry
		Cavity Pulling
		Interrogating Radiation

From these equations we see that the flux of signal atoms and the relaxation time which make up our basic figure of merit M are not the only important parameters in determining frequency stability. The principle which is used in the interrogation of the particles is also important, and is reflected in the value of K. For example a device (with the same M for each version) will have a better noise performance when passive particle detection or optical photon detection is used as compared to using the mode of operation as an active quantum electronic oscillator.[23] This can be seen by comparing (3) and (4). Also, additive thermal noise at any reasonable particle or photon intensity is negligible, and excess noise contributions of particle or photon detectors are typically very low. Thus shot noise of the incoming particles is usually the only limitation. In contrast we almost always will encounter additional noise in an active (maser) oscillator because of performance limitations in its microwave receiver.

Another potential noise source is the frequency multiplier and synthesis chain connecting the slave oscillator to the "clock" transition. The current state-of-the-art is such that this noise contribution is negligible compared to the noise due to the slave oscillator.[1,24] Noise can also be associated with the quantum transition itself, i.e., in the form of spontaneous emission. Such effects will become more serious at higher (optical) frequencies because the probability for spontaneous transitions increases as the cube of the transition frequency.

The short-term stability of the slave oscillator is of equal importance for all methods because its performance determines the stability of the whole standard at short averaging times.[22] The only remedy is the development of good slave oscillators.

The next item in the second column of Table 1 is cavity pulling, which may be written as[12,25]

$$\Delta\nu \equiv G\Delta\nu_c \qquad (5)$$

where $\Delta\nu_c$ and $\Delta\nu$ are the offsets of the cavity and the output frequencies respectively. For an active oscillator the parameter G is approximately the ratio of the widths of clock transition and cavity resonance,[12] for passive methods (sufficiently far from self-oscillation) G is approximately the square of this ratio.[25] Thus an active oscillator is inherently more affected by cavity pulling than a passive device. Moreover a sharp cavity resonance, i.e., a larger G, is a prerequisite for an active oscillator[12] whereas a broad cavity resonance or--at least in principle--no resonance structure at all, i.e., G = 0, can be used with passive methods. Cavity pulling is therefore only of concern to active oscillators.

Spectral asymmetry of the interrogating signal is of concern for all passive devices.[25,26] It can be reduced to unimportance by adequate care in the electronics. Also the choice of a low frequency transition might be helpful. In active devices this problem is practically non-existent.

The last item in Table 1 is the effect caused by the interrogating signal itself. We will not encounter serious limitations until the photon energy of the interrogating radiation becomes large enough to cause new phenomena like, for example, recoil processes. They are discussed in more detail later in this paper under Conclusions.

We summarize: A passive method is to be preferred over an active oscillator. In addition a passive method offers more flexibility and freedom for evaluating bias corrections and accuracy limitations, because a passive technique is not tied to meet an oscillation threshold condition. A low frequency (microwave) transition eliminates some of the problems; however, higher frequencies (infrared) may be used if technical problems associated with the multiplication process* can be adequately solved. Very high frequencies (optical) should not be considered because of effects associated with photons of higher energy.

2. Effects on Frequency: Particle Confinement

A quantum electronic frequency standards is based ideally on the quantum transition of a particle (atom or molecule) which is in an unperturbed state and is interrogated for an indefinite period of time.

In principle this may be achieved by confining an ensemble of particles to that region of space which is filled with the interrogating radiation and to accomplish this confinement in such a manner as to minimize the perturbation of the individual particles. In reality it appears possible to realize the ideal of a free particle to a very good approximation; however, it is at the expense of interrogation time, and vice versa.

Various methods of confinement are at our disposal. We may use a beam of free particles traveling in vacuum through the region of space where interrogation takes place.[18, 19] We may store the particles in the region of interrogation at low particle densities by using storage vessels. The storage vessel has to be coated[27,28] or filled with a buffer gas[13, 29] in order to reduce the interaction between the particles and the wall of the storage vessel and among particles themselves. Or we can eliminate physical walls and use electric or magnetic fields for confinement.[30] Finally, we may just take a gas at a fairly high pressure and observe absorption[16, 17] of the interrogating radiation. A summary of these methods is given in the first column of Table 2. Various effects (listed in the second column of Table 2) which lead to bias uncertainties are introduced by the confinement of particles.

Several effects are associated with the width of the line (line-Q) such as short term stability, cavity pulling, and the performance of the servo electronics. A high line-Q is desirable but not essential because means can be employed to counter the effects of a low line-Q, e.g., one could render cavity pulling totally unimportant, regardless of the line-Q, by avoiding a resonance structure (using a passive method). On the other hand a high line-Q can be obtained with relatively short confinement times by using transitions at high frequencies. For achieving a long duration of confinement, i.e., in excess of one second, storage methods offer the greatest potential, especially the coated vessel or the usage of confining fields (ion storage). Collisions between the particles which lead to spin exchange and shifts of the energy levels are an important source of uncertainty in techniques using comparatively high particle densities. Low density methods like the free beam, the storage in a coated bulb, or the usage of confining fields can be designed to render this effect unimportant. We also may arrange the interrogation mode in such a way as not to look at particles which have experienced a collision, e.g., as in saturated absorption.[16] Wall collisions will introduce uncertainties wherever physical walls are used as a means of confinement. Only the free beam and the confinement by fields are not affected.** However, we may also use an absorption cell and restrict the region of interrogation to a small volume within the cell,[16] e.g., by using a laser beam as the interrogating radiation, thus avoiding wall effects. The first order Doppler effect appears to be mainly a design and engineering problem and thus of no real consequence for this discussion. We may assume that a suitable approach in all of the discussed methods will eliminate first order Doppler frequency shifts and line broadening. This may be done by proper interrogation (e.g., saturated absorption[16]), by storage within dimensions less than one-half of a wavelength of the interrogating radiation,[14] or by a proper mechanical design as required in the case of a free beam traveling through a Ramsey cavity.[19] The second order Doppler shift is of a more fundamental nature. It can be regarded as the problem to know with adequate precision the speed or temperature of the particles. This is difficult in the case of the free beam and is a severe problem in the case of storage in electric or magnetic fields, whereas the use of physical containers allows an adequate knowledge of the temperature, when thermal equilibrium of the kinetic energy of the particle with the container walls can be established. Cavity phase shift offsets occur in the case of the free beam traveling through a cavity.[1-5] For all other methods it is of no concern.

*We may assume that multiplication, in principle, is possible.

**For the case of a magnetic dipole transition this is true when the particles (ions) are confined by electric fields.

Table 2

Effects on Frequency: Particle Confinement

Methods	Effects
Traveling Free Beam	Line-Q
Storage in:	Particle-Particle Collisions
Coated Vessel	Wall Interaction
Buffer Gas	1st Order Doppler
Electric or Magnetic Fields	2nd Order Doppler
Absorption Cell	Cavity Phase Shift

We summarize: The choice of a high frequency transition is to be preferred because this leads to a reduction of the fractional importance of most effects; the Doppler shifts, of course, are excluded. Thus the (second order) Doppler limitation favors the storage in containers where thermal equilibrium exists between the kinetic energy of the particles and the walls of the storage vessel. At low (microwave) frequency we cannot single out a specific method of confinement as being inherently superior to others.

3. Effects on Frequency: Particles and Particle Preparation

The particles which we choose are either atoms or molecules, and we may use any of the different types of electric or magnetic dipole transitions. In order to effect interrogation, the particles are generally prepared so that this can be done efficiently. Preparation usually means the creation of a population difference of the energy levels which differs from (is larger than) the thermal distribution.

Only at high frequencies, above one terahertz, does the preparation of particles lose its importance, because of the large population differences already present in thermal equilibrium. This allows efficient absorption of the interrogation signal without preparation. Possible methods of preparation are listed in the first column of Table 3.

Spatial state selection achieves the change in population difference by eliminating one or more energy states from a particle beam.*,[31] This method is based on differences in the electrostatic or magnetic forces acting in inhomogeneous electric or magnetic fields on the dipole moments which are associated with each energy state. Optical pumping can be used which causes changes in population difference when the pumping light acts on the energy levels selectively by proper use of polarization, filtering,[32] etc. Electron collisions can be used to alter the population of the energy levels from the thermal equilibrium values. As examples, electron beams of well defined energy or a gas discharge can be employed. A change in population can also be accomplished by spin exchange collisions with atoms or molecules which have already been polarized by any of the other methods.[30] Chemical effects may be used such as a chemical reaction or a dissociation which lead to the formation of excited particles.

*This method dates back to the famous Stern-Gerlach experiment.

Table 3

Effects on Frequency: Particles and Particle Preparation

Methods	Effects
Atoms	Electric Fields
Molecules	Magnetic Fields
No Preparation (Absorption)	Majorana Transitions
Spatial State Selection	Neighboring Transitions
Optical Pumping	Coupling of Transitions
Electron Collisions	Collisions
Spin Exchange	
Chemical Reaction	

Again we find various effects associated with the different methods which give rise to bias uncertainties. They are listed in the second column of Table 3. The influence of external electric fields can be made unimportant if magnetic dipole transitions such as the hyperfine transitions in atoms are used.[33] Magnetic fields can be rendered unimportant when certain molecular transitions are taken.[34] Shielding will reduce the effects of both. Majorana transitions may occur if particles travel through regions of varying static field strength and thus are mainly of some importance for spatial state selection. Frequency uncertainties are introduced by transitions from neighboring states. This effect can be minimized by the choice of particles with a simple energy level structure and by carefully avoiding the stimulation of neighboring transitions.

Two classes of effects, listed in Table 3 as collisions and coupling of transitions, represent significant sources of frequency bias (and uncertainty) for all methods other than spatial state selection and absorption. Coupling of transitions occurs when the pumping radiation is present simultaneously with the interrogating radiation. Collisional effects will be encountered when collisions are used as a means of altering the population difference. The very method used for particle preparation is the source of a frequency bias (and uncertainty), and the more efficiently the preparation is made, the more restrictive is its effect on the performance as a primary frequency standard. In principle, we have a solution. In analogy to spatial state selection we may separate spatially the process of particle preparation from the particle interrogation. This can be done, for example, with two storage vessels connected by a diffusion channel or by using a traveling beam. An alternate solution is the separation in time of particle preparation and interrogation (pulsed preparation). However, we must realize that both solutions will encounter serious technical problems in their practical realization although they are fundamentally feasible.

We summarize: An atom or molecule should be used which has a simple energy level structure and properties which reduce the effects of external electric and/or magnetic fields. Obviously, the best choice for a particle preparation method is no preparation at all (simple absorption). However, we must then use transitions in at least the low terahertz range where the population difference in thermal equilibrium is sufficiently large for efficient interrogation. The optimum method of particle preparation at low (microwave) frequencies appears to be the spatial state selection.

1. Existing Concepts for Quantum Electronic Frequency Standards

We will discuss only those quantum electronic frequency standards with a current or potential accuracy capability of better than one part in 10^{10}. Several devices have been developed, and some are even commercially available. They include the forerunner of all, the ammonia maser,[12, 35] the rubidium gas cell,[36] the cesium beam tube,[1-5, 19] and the hydrogen maser.[14, 15] Two more devices have had a preliminary evaluation but are not nearly as mature as those previously mentioned: the thallium beam tube[37, 38] and the rubidium maser.[13] Several more concepts are being proposed or are in the early stages of experimentation. They include saturated absorption of laser radiation (methane,[16] iodine,[17] sulfur hexafluoride,[39] etc.), simple absorption of laser radiation by a traveling beam (iodine[18]), other beam tubes (barium oxide[40]), ion storage (helium,[30] mercury, etc.), and the storage beam tube (hydrogen).[23]

We will now look into the basic operation principle of each device and discuss briefly the chief limitation of each by comparing with Tables 1, 2, and 3. It must again be emphasized that any effect imposing an accuracy limitation of worse than one part in 10^{14} will count as a limitation. No attempt will be made at quoting quantitatively these limitations. For those devices which have been evaluated, the reader is referred to the published data in the referenced literature. For all other devices and concepts it seems futile to quote data which cannot adequately be supported. In order to form his own opinion the reader should consult the referenced literature. Finally it must be emphasized that the limitations quoted here should not be regarded as final. They merely reflect the present state of our knowledge, and future work is likely to change them.

(a) Traveling Beam Tube (Figure 1).

The beam of particles originates at a suitable source or oven and travels through a first state selector which focuses only certain, desired energy states into the cavity region. Here the particles are interrogated by a microwave signal. As a result the distribution of particles into the various energy states is altered when the particles leave the cavity region. This is analyzed by a second state selector which focuses particles in selected energy states on a detector thus generating the error signal for a slave oscillator. Limitations are the spectrum of the interrogating radiation (Table 1); the phase difference between the two cavities, the second order Doppler shift due to the uncertainty in the mean square particle velocity, and the relatively low line-Q (Table 2); and in the case of cesium and barium oxide the effects of fields (Table 3).

(b) Absorption in a Beam (Figure 2).

The beam of particles originates in a source and travels freely through a vacuum chamber. A laser which serves as the slave oscillator radiates perpendicularly onto the particle beam. Changes in the intensity of the laser radiation due to absorption in the beam can be detected (in-line position in Fig. 2), or alternately the intensity of the fluorescence radiated by the traveling beam may be monitored (displaced position in Fig. 2). Limitations are the spectrum of the laser radiation and the radiation itself via photon effects (Table 1); and the first and second order Doppler effects (Table 2).

(c) Traveling Beam Maser (Figure 3).

The only example is the ammonia maser. A beam of molecules leaves the source, is state selected, and enters a cavity which is tuned to the transition frequency. Because of the short interaction time between the particles and the microwave field in the cavity a large particle flux is required to obtain self-oscillation. Also, the delivery of the energy from the molecules to the cavity is in general not uniform and not symmetric along the length of the cavity. This together with the more complex level structure of ammonia leads to the following: Cavity pulling (Table 1); first and second order Doppler shifts, particle collisions, and cavity phase shifts (Table 2); and electric fields, magnetic fields, and the influence of neighboring transitions (Table 3).

(d) Storage Beam Maser (Figure 4).

The only example is the hydrogen maser. A state selected atomic beam is generated in a fashion similar to the beam tubes. The atoms are stored in a coated bulb located within a cavity. The maser will oscillate provided certain conditions are met which relate to cavity- and line-Q, particle flux, and device geometry. Limitations are cavity pulling (Table 1); and wall interaction due to the storage principle (Table 2).

(e) Storage Beam Tube (Figure 5).

The hydrogen storage beam tube is a hybrid between the hydrogen storage beam maser and the traveling beam tube. The operation is analogous to beam tubes; however, the hydrogen storage principle is adopted from the hydrogen maser. The direct flow of atoms from entrance to exit of the bulb must be prevented. A beam stop might be used, for example. The only serious limitation of this device is the wall shift (Table 2).

(f) Optically Pumped Gas Cell (Figures 6 and 7).

Examples are the rubidium maser (Fig. 6) and the rubidium gas cell (Fig. 7). Their basic concepts are quite similar. A filtered light beam optically pumps the rubidium which is stored, together with buffer gases, in a cell within a cavity. The maser will start oscillations under certain conditions. The passive gas cell obtains the error signal for a slave oscillator from a photocell which monitors the transmission of light through the gas cell. Thus a method which was called frequency transformation in Table 1 is used here, since the optical transmissivity is a function of the microwave signal applied to the cavity. Limitations are cavity pulling (maser only) and the spectrum of the interrogating radiation (passive device only) (Table 1); particle collisions, wall interaction, and a relatively low line-Q (Table 2); and the effects caused by the coupling of the clock transition to the optical pump transition which usually are called "light shifts" (Table 3).

(g) Saturated Absorption (Figure 8).

A gas cell is used which is irradiated by intense radiation from a laser which acts as the slave oscillator. Unique advantages of the saturation of an absorption are the significant reduction of the first order Doppler effect and the automatic exclusion from interrogation of most of the molecules which have suffered a collision. Known limitations are the spectrum of the laser radiation as well as photon effects of this radiation itself (Table 1); and particle collisions (Table 2).

(h) Ion Storage Principle (Figure 9).

Ions are confined by an inhomogeneous rf field, and can be stored for fairly long time periods. The confining field may be generated in a quadrupole trap consisting of a hyperbolic doughnutlike ring with hyperbolic caps on top and bottom.* Limitations are imposed by the necessity to ionize and by the method of particle preparation, which, so far, is usually spin exchange with an injected polarized atomic beam. Comments on this aspect were already made in Section 3. The most severe limitation is the second order Doppler effect (Table 2).

*Several other trap configurations are possible.

5. Frequency Stability for One Second Averaging

In the previous discussion we did not include the aspects of random uncertainty. These aspects can be described quantitatively by (1) through (4). In Table 4, an attempt at this is made for the different existing concepts, whereby we choose the best understood actual system in each case, e.g., we use the actual (NBS III)[1] and projected (NBS 5)[1] performances of cesium laboratory standards in the case of the beam tube concept.

The first column lists n_s, the signal particle flux or the number of interrogated particles per second; the second column gives the interrogation time T_r. In the third column the figure of merit M is calculated from (2). With the transition frequencies, which are listed in the fourth column, we can calculate from (1), (3), and (4) the fractional stability for one second averaging time. It must be emphasized that the values given for n_s and T_r and therefore for σ_y are only approximate and sometimes only "educated guesses." However, they may serve to give an approximate feeling for the potential stability performance of each technique. We also must note that the measured stability usually will be worse than the calculated stability because of limitations in the associated electronics, i.e., receiver or detector, slave oscillator, etc. For example the best actual performance of the hydrogen maser is $\sigma_y(1\,s) \approx 8 \times 10^{-13}$ [41] (5×10^{-13} [42]) and of the methane saturated absorption is $\sigma_y(1\,s) \approx 2 \times 10^{-12}$ [16] (2×10^{-13} [43]).

6. Accuracy Capability

At the beginning of this chapter we pointed out the difficulty in quoting quantitative values for the various error sources which make up the accuracy capability. Nevertheless it is of interest to quote those accuracy capabilities which have been determined for actual systems based on experiments. They are listed in the fifth column of Table 4. There are no experimental data which would support accuracy capability claims for the beam absorption and the storage beam techniques. The last column in Table 4 depicts the projected performance which we consider possible from applying and evaluating the thoughts of Sections 1, 2, and 3 of this paper and from interpreting the experimental data. A comparison between the last two columns of Table 4 gives an approximate idea of the stage of knowledge and development of each concept.

Conclusions

The (ammonia) traveling beam maser and the (rubidium) optically pumped gas cell have been studied quite extensively and their limitations are well known. The magnitude of the limitations, stated qualitatively in paragraphs (c) and (f) of Section 4, is such that we cannot expect any improvement in the accuracy capability in the foreseeable future as is indicated in Table 4. Their accuracy capability is also, in comparison, so marginal that they are not competitors in the contest for superior primary frequency standards.

The standard with the best accuracy capability at present is the cesium beam tube. The individual effects contributing to this accuracy capability are also comparatively well understood.[1-5] Therefore it is possible to predict that an accuracy capability of 1×10^{-13} may actually be realized in the foreseeable future.[1] The projected performances of the remaining concepts, which are listed in Table 4, are quoted as better than 10^{-12}. Each of the limiting effects for these concepts, as discussed in the corresponding paragraphs of Section 4, has the possibility of being reduced to that level.

How far it will be possible to push beyond 10^{-12} is quite difficult to predict. As a qualitative rule we may state that the magnitude of an error (bias uncertainty) will decrease when the corresponding bias correction is reduced. Also we may gain an advantage when we choose a concept which involves as few bias corrections as possible. Adopting this philosophy we can reexamine those devices listed in Table 4 which hold promise, and we can try to synthesize an optimum solution following the thoughts of Sections 1, 2, and 3. It is evident that no single superior concept can be found. We have to compromise in order to arrive at a practical solution.

In Section 1 we pointed out that a passive technique should be preferred over an active oscillator (maser, laser), although it is difficult to give a fair judgment on all parameters involved. In this paper we are only concerned with "fundamental" limitations and it is conceded that technical and design aspects may well be more important than the fundamental ones. As an example, it remains to be proven experimentally that the detection of a hydrogen storage beam device does not create sufficient technical problems to render its "fundamental" superiority over the hydrogen storage maser ineffective.

Table 4

Data on Different Concepts for Primary Frequency Standards

	n_s (s^{-1})	T_r (s)	M ($s^{\frac{1}{2}}$)	ν (Hz)	$\sigma_y = \frac{K}{M}\frac{1}{\sqrt{\tau}}$ (τ = 1 s)	Accuracy Capability: Present experimental performance	Accuracy Capability: Projected
Traveling beam tube (Cs)	10^8	10^{-2}	10^2	9.2×10^9	2×10^{-13}	5×10^{-13} [1]	1×10^{-13}
Absorption in beam (I_2)	10^8	10^{-6}	10^{-2}	5.8×10^{14}	3×10^{-14}	-	better than 10^{-12}
Traveling beam maser (NH_3)	10^{13}	10^{-4}	3×10^2	2.4×10^{10}	5×10^{-13}	10^{-11} [44]	10^{-11}
Storage beam maser (H)	10^{12}	1	10^6	1.4×10^9	10^{-14}	2×10^{-12} [45]	better than 10^{-12}
Storage beam tube (H)	10^8	1	10^4	1.4×10^9	10^{-14}	-	better than 10^{-12}
Optically pumped gas cell (Rb)	10^{14}	10^{-3}	10^2	6.8×10^9	10^{-14}	10^{-10} [13]	10^{-10}
Saturated absorption (CH_4)	10^{10}	10^{-6}	1	8.8×10^{13}	2×10^{-12}	10^{-11} [16]	better than 10^{-12}
Ion storage ($^3He^+$)	10^6	10	3×10^3	8.7×10^9	10^{-14}	10^{-9} [30]	better than 10^{-12}

In the search for an optimum passive technique the frequency of the transition is an important parameter. A low frequency is of advantage in the interrogation of particles; however, effects caused by particle confinement are fractionally large, and it is necessary to prepare the particles for effective interrogation. The most effective method for particle preparation which also introduces no adverse effects, if proper care is taken, is spatial state selection of a particle beam. For particle confinement at low frequencies we have to rule out simple absorption. Storage in buffer gases introduces large biases and with it relatively large errors. A third method, ion storage, seems to be severely affected by the second order Doppler effect. A significant reduction of this bias (effective ion temperature) appears difficult and its knowledge to correspondingly better than 10^{-12} seems to be a serious problem.[46] Technological difficulties associated with the creation, injection, and preparation of ions may also restrict the usefulness of ion storage as a method of confinement. However, it should not be considered experimentally impossible, especially if heavy ions are used which reduce adverse heating effects.[47]

Two choices are left. The first is the traveling beam method, i.e., in practice an advanced cesium beam tube. The use of particles other than cesium will not give a substantial advantage. The prospects can be fairly well predicted (Table 4). The second choice is the storage beam tube. So far only atomic hydrogen can be stored effectively in a vessel with coated walls. The only limiting effect is the wall interaction (wall shift). At present the wall shift bias is typically of the order of 10^{-11} and can be measured to about 10 percent.[45] Recent experiments show that the wall shift can actually be made zero at elevated temperatures.[48,49] We also may assume that bulbs of variable size can be used to evaluate the wall shift with a higher degree of accuracy,[50] and that considerably larger storage bulb sizes

can be used.[51] If we assume that the wall shift bias could be reduced by one order of magnitude and could be measured to 1 percent of its value, we would have an accuracy capability of 10^{-14}.

A high frequency transition eliminates most of the problems associated with particle confinement and preparation because we can use simple absorption. Specific problems and technical difficulties may arise in connection with the frequency multiplication and synthesis into the infrared region. We shall assume that these problems can be overcome. Confinement in an absorption cell is possible because the fractional influences of wall and particle collisions are reduced due to the high frequency. Saturated absorption[16] will further reduce adverse effects because only particles with near zero Doppler shift and near zero collisional effects are interrogated. Another choice is the use of a traveling beam[18] which absorbs radiation directed perpendicular to the beam. Effects related to the energy of the interrogating photons are a limitation which becomes more pronounced as the transition frequency increases.* Existing methods like the methane saturated absorption will allow a study of these effects and will give an indication of how well the bias can be determined.**

Following these thoughts one may conclude that such effects are reduced by choosing a transition of lower frequency, e.g., in the far infrared, while still retaining the advantages of a relatively high transition frequency. The optimum frequency for such an approach appears to lie in the low terahertz region. Here the opportunity for gaining full advantage from techniques based on simple absorption still exists (sufficient population and population difference even at thermal equilibrium, and optical techniques can be used), but the biases related to the energy of the interrogating photons are considerably reduced. These effects can be reduced further and the interrogation time can be increased by using molecules of larger mass. Also the traveling beam method might be combined with the saturated absorption technique. Thus accuracy capabilities of better than 10^{-13} may be expected.

In summary we may conclude that among other possibilities the storage beam principle (hydrogen) holds high promise for a quantum electronic frequency standard based on a low (microwave) transition frequency. The other group of techniques which has a potential of competing with or surpassing the projected performance of the cesium beam tube is based on simple absorption in transitions of high frequency, in the terahertz (at least) region. Both approaches hold the promise for accuracy capabilities of 10^{-14} and are within reach of experimental realization.

*The fractional influence is proportional to ν.

**This bias is of the order of 10^{-11} for methane.

Acknowledgements

This paper is based on the work and thoughts of many during the past years as evidenced by the list of references which is by no means exhaustive. I gratefully acknowledge their input to this paper.

I owe special thanks to Donald Halford and John L. Hall who helped shape this paper in many fruitful discussions.

References

1. D. J. Glaze, Proc. of the XVI General Assembly of URSI, Ottawa, Canada, August 1969, entitled Progress in Radio Science, 1966-1969, to be published; also to be published in IEEE Trans. on Instrumentation and Measurement IM-19 (August 1970).
2. P. Kartaschoff and P. E. Debély, Proc. of the Colloque International de Chronométrie, p. A3-1 (1969).
3. G. Becker, B. Fischer, G. Kramer, and E. K. Müller, Proc. of the Colloque International de Chronométrie, p. A2-1 (1969).
4. G. Becker, B. Fischer, G. Kramer, and E. K. Müller, PTB-Mitteilungen 2, p. 77 (1969).
5. A. G. Mungall, R. Bailey, H. Daams, and D. Morris, Metrologia 4, p. 165 (1968).
6. L. O. Hocker, D. Ramachandra, and A. Javan, Phys. Letters 24A, p. 690 (1967).
7. L. O. Hocker, D. R. Sokoloff, V. Danca, A. Szoke, and A. Javan, Appl. Phys. Letters 12, p. 401 (1968).
8. K. M. Evenson, J. S. Wells, and L. M. Matarrese, Appl. Phys. Letters 16, p. 251 (1970).
9. K. M. Evenson, J. S. Wells, L. M. Matarrese, and L. B. Elwell, Appl. Phys. Letters 16, p. 159 (1970).
10. L. O. Hocker, J. G. Small, and A. Javan, Phys. Letters 29A, p. 321 (1969).
11. A. O. McCoubrey, Proc. IEEE 54, p. 116 (1966).
12. K. Shimoda, T. C. Wang, and C. H. Townes, Phys. Rev. 102, p. 1308 (1956).
13. P. Davidovitz and R. Novick, Proc. IEEE 54, p. 155 (1966).

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

15. H. E. Peters, T. E. McGunigal, and E. H. Johnson, Proc. 23rd Annual Symposium on Frequency Control, p. 297 (1969).

16. R. L. Barger and J. L. Hall, Phys. Rev. Letters 22, p. 4 (1969).

17. G. R. Hanes and K. M. Baird, Metrologia 5, p. 32 (1969).

18. S. Ezekiel and R. Weiss, Phys. Rev. Letters 20, p. 91 (1968).

19. R. E. Beehler, R. C. Mockler, and J. M. Richardson, Metrologia 1, p. 114 (1965).

20. R. F. Lacey, A. L. Helgesson, and J. H. Holloway, Proc. IEEE 54, p. 170 (1966).

21. H. G. Dehmelt, Phys. Rev. 103, p. 1125 (1956).

22. L. S. Cutler and C. L. Searle, Proc. IEEE 54, p. 136 (1966).

23. H. Hellwig, Metrologia 6, No. 2 (1970).

24. D. Halford, A. E. Wainwright, and J. A. Barnes, Proc. 22nd Annual Symposium on Frequency Control, p. 340 (1968).

25. J. H. Holloway and R. F. Lacey, Proc. of the Colloque International de Chronométrie, p. 317 (1964).

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

27. N. F. Ramsey, Rev. Sci. Instr. 28, p. 57 (1957).

28. D. Kleppner, N. F. Ramsey, and P. Fjelstad, Phys. Rev. Letters 1, p. 232 (1958).

29. R. H. Dicke, Phys. Rev. 89, p. 472 (1953).

30. E. N. Fortson, F. G. Major, and H. G. Dehmelt, Phys. Rev. Letters 16, p. 221 (1966).

31. N. F. Ramsey, Molecular Beams, Oxford (1956).

32. P. L. Bender, E. C. Beaty, and A. R. Chi, Phys. Rev. Letters 1, p. 311 (1958).

33. C. Schwartz, Ann. Phys. 6, p. 156 (1959).

34. F. S. Barnes, IRE Trans. on Instrumentation I-11, p. 191 (1962).

35. K. Shimoda, J. Phys. Soc. Japan 16, p. 1728 (1961).

36. M. Arditi and T. R. Carver, Proc. IEEE 51, p. 190 (1963).

37. J. Bonanomi, IRE Trans. on Instrumentation I-11, p. 212 (1962).

38. R. E. Beehler and D. J. Glaze, IEEE Trans. on Instrumentation and Measurement IM-15, p. 55 (1966).

39. R. Rabinowitz, R. Keller, and J. T. LaTourette, Appl. Phys. Letters 14, p. 376 (1969).

40. H. Hellwig, R. McKnight, E. Pannaci, and G. Wilson, Proc. 22nd Annual Symposium on Frequency Control, p. 529 (1968).

41. R. Vessot, et al., IEEE Trans. on Instrumentation and Measurement IM-15, p. 165 (1966).

42. H. E. Peters, private communication.

43. J. L. Hall, private communication.

44. J. DePrins, IRE Trans. on Instrumentation I-11, p. 200 (1962).

45. H. Hellwig, et al., Conference on Precision Electromagnetic Measurements, Boulder, Colorado (1970). Submitted for publication in IEEE Trans. on Instrumentation and Measurement IM-19 (November 1970).

46. D. Church and H. G. Dehmelt, J. Appl. Phys. 40, p. 3421 (1969).

47. H. G. Dehmelt, Advances in Atomic and Molecular Physics 5, p. 109 (1969).

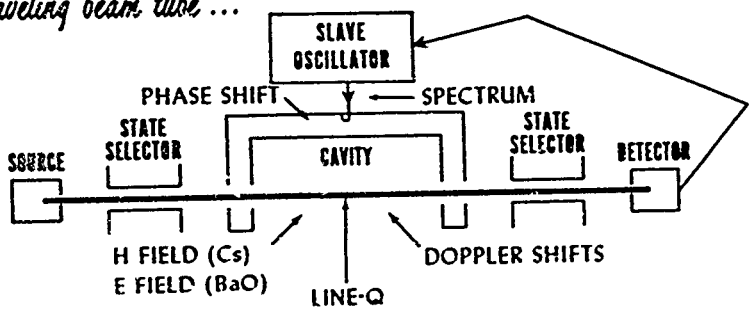
48. P. W. Zitzewitz, Proc. 24th Annual Symposium on Frequency Control (1970). To be published.

49. R. F. C. Vessot and M. W. Levine, Proc. 24th Annual Symposium on Frequency Control (1970). To be published.

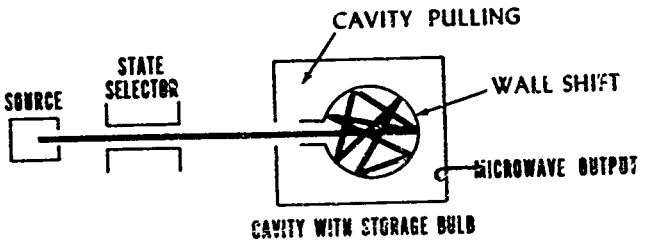
50. P. Debély, Proc. 24th Annual Symposium on Frequency Control (1970). To be published.

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

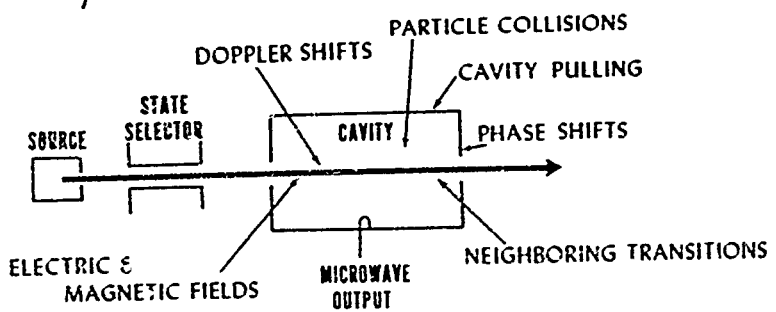
Traveling beam tube ...



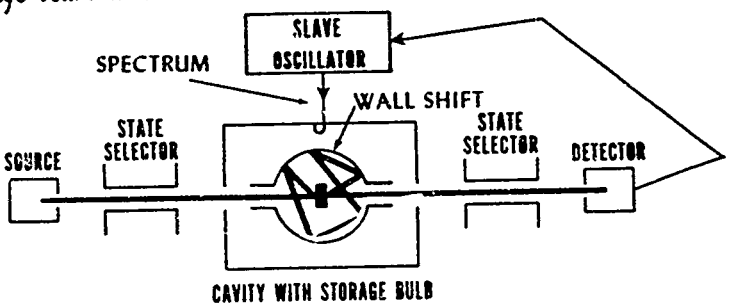
Storage beam maser...



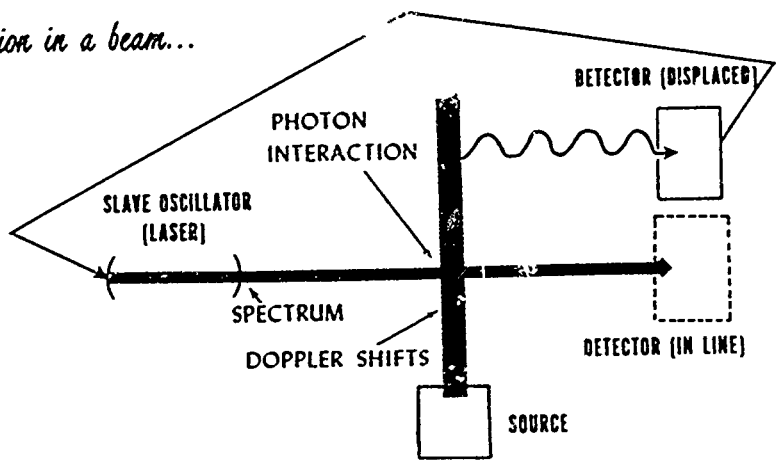
Travelling beam maser...



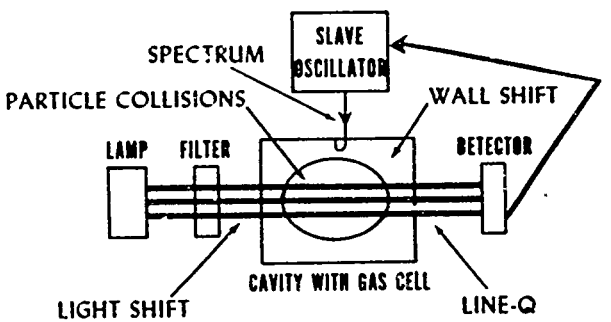
Storage beam tube ...



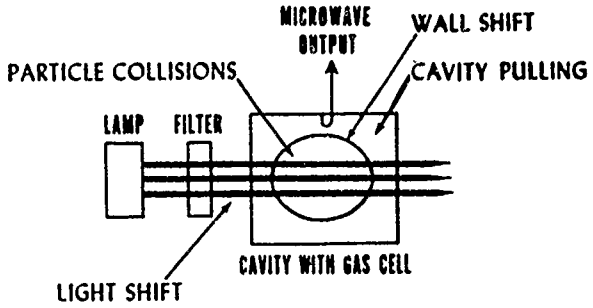
Absorption in a beam...



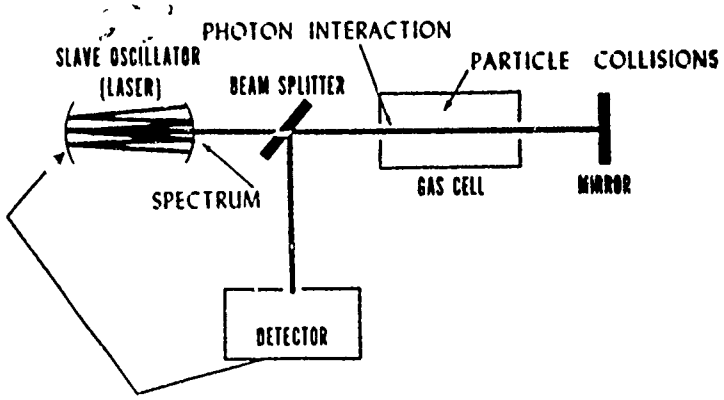
Optically pumped passive gas cell ...



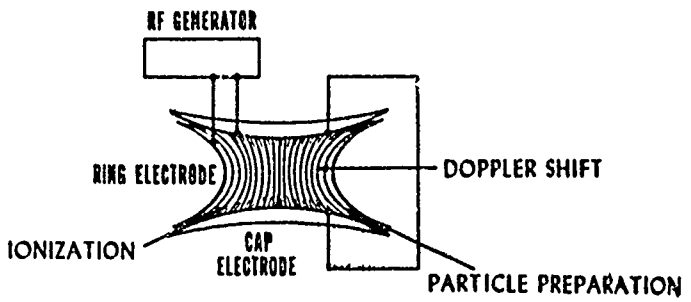
Optically pumped gas cell maser ...



Saturated absorption...



Ion storage principle....



HYDROGEN MASER WITH DEFORMABLE STORAGE BULB*

Pierre E. Debely†
Physics Department, Harvard University
Cambridge, Massachusetts 02138

ABSTRACT

The value which most affects the accuracy of the hydrogen maser is the experimental determination of the frequency shift produced by the collision of the atoms with the surface of the storage bulb. In the current method of using different bulbs with various sizes in order to determine the frequency shift, a large amount of uncertainty is due to the unreproducibility of the FEP Teflon coatings from one bulb to another. A way to get rid of this error is to use a single deformable bulb whose volume can be changed while the same surface is retained. Such a bulb has been developed which can be adapted to any regular maser.

INTRODUCTION

If we call t_o the time between collisions of the atoms with the FEP Teflon coated wall of the bulb, and ϕ an average phase shift which occurs on each collision, we can define a frequency shift $\delta\nu$:

$$\delta\nu = \phi / (2\pi t_o)$$

Since the frequency shift is proportional to the collision rate, it can be determined by measuring the frequency of the maser as a function of the storage bulb size, the mean free path d of the atoms inside the bulb being a function of the volume V and the surface area A of the storage bulb. Let us call N the number of atoms per unit volume inside the bulb and $\bar{v}$ the average velocity of the atoms. We know from gas kinetics[1] that the total number Q of collisions per second is given by the expression:

$$Q = \frac{1}{4} N \bar{v} A$$

Thus: $$d = \bar{v} N V / Q = 4 V / A$$

This result is true for any kind of bulb configuration. Since $t_o = d / \bar{v}$, we have the expression for the frequency shift as a function of the geometry of the storage bulb:

$$\delta\nu = \bar{v} \phi A / 8 \pi V$$

The value of the beat frequency $\Delta\nu$ between two masers (after correction for different magnetic fields and second order Doppler effects) gives us the difference of the frequency shifts due to the different bulb in each maser. Now if we change the bulb of one maser in order to change its ratio A / V, so that it becomes B times smaller, we obtain another value $\Delta\nu'$ for the beat frequency. If we assume that ϕ does not change from one coating to another, B will be the factor of proportionality between the frequency shifts of the bulbs which have been used in the same maser. That is the way the frequency shift determinations have been done by several workers[2-6] and the discrepancies in their results indicate either that there are unrecognized systematic errors in this method or that ϕ varies from one sample of Teflon to another. A way to keep ϕ constant during a frequency shift measurement is to use a flexible bulb in order to be able to change its volume V while keeping the same surface A, as initially proposed by D. Brenner and used by him in some preliminary experiments at Harvard University.[7]

An improved deformable storage bulb has been developed which can be adapted to any regular maser (Fig. 1). A simple device permits both a rapid and accurate variation of the volume of the bulb, as well as the necessary readjustment of the cavity tuning. The high vacuum in the whole system is maintained at all times.

Let us call $\delta\nu_D$ and $\delta\nu_D'$ respectively the frequency shift of the deformable bulb in small and large volume configuration, $\Delta\nu$ and $\Delta\nu'$ being the observed beat frequencies between the two masers. We will call $\delta\nu_R$ the frequency shift of the rigid bulb which is in the reference maser, frequency shift that we wish to determine.

From the two equations:

$$\delta\nu_D - \delta\nu_R = \Delta\nu$$

$$\delta\nu_D' - \delta\nu_R = \Delta\nu'$$

and with the supplementary condition:

$$\delta\nu_D = B\, \delta\nu_D'$$

we obtain the frequency shift of the reference maser $\delta\nu_R$:

$$\delta\nu_R = (\Delta\nu - B\Delta\nu') / (B - 1)$$

DESCRIPTION OF THE DEFORMABLE STORAGE BULB

The deformable bulb is made of three main parts (Fig. 2). The rigid part is a quartz cylinder including a stem in its base, coated with five FEP Teflon layers in the normal way.[2,5] The deformable part of the bulb is a two mil thick FEP Teflon cone which can be inverted; its construction is described below. The base of the cone is held against the upper edge of the quartz cylinder by two rigid Teflon rings, one on each side of the quartz wall. The diameter of the outer ring is adjusted by a nylon screw. The two rings are held together by eight Teflon clamps. The upper part of the cone

is compressed between two Teflon rods, with the surface exposed to the atoms covered with another FEP Teflon film. The whole surface seen by the atoms is therefore made of FEP Teflon. By pulling the main rod up and down, through a sliding o-ring seal in the top of the bell jar, the frustrum of the cone will be either outside or inside the cylinder. The ratio B of the volumes is computed geometrically. In order to make the FEP Teflon cone, a four layer coating is made inside a Pyrex funnel having the desired opening angle. The coating is removed from the Pyrex surface by the action of an acid carefully heated inside the funnel. This acid is a mixture of two parts of concentrated nitric acid, one part of hydrofluoric acid and two parts of distilled water. The hydrofluoric acid diffuses through the film, attacks the Pyrex and releases the Teflon.

The characteristics of the bulb are indicated in the Table I.

EXPERIMENTAL RESULTS

The purpose of the experiment was to determine the frequency shift produced by the bulb which is used in the reference maser of Harvard University. Knowing B, the main characteristic of the deformable bulb:

$$B = 1.299 \pm 0.004$$

and having measured the following values for the beat frequency between the reference maser and the maser with deformable bulb:

$$\Delta\nu = -43.76 \pm 0.27 \quad \text{mHz}$$

$$\Delta\nu' = -28.84 \pm 0.17 \quad \text{mHz}$$

we thus determined the frequency shift of the reference maser which has a 15.25 cm diameter spherical quartz bulb at a temperature of 31.5 °C:

$$\delta\nu_R = -21.07 \pm 2.32 \quad \text{mHz}$$

where 1 mHz = 10^{-3} Hz.

It should be noted that the error on this value can be reduced by choosing a larger ratio B. However, a limitation in the size of B comes from the fact that the conditions required for oscillation in the deformable bulb in the small volume configuration are stronger than those in the case of a spherical bulb, because most of the volume is outside the center of the cavity where the oscillating field strength is concentrated.

COMPARISON WITH OTHER WORK

The result presented in this work is in good agreement with the one obtained by P. W. Zitzewitz[5] who used eighteen different bulbs in order to determine the frequency shift of the same bulb used in the reference maser at Harvard University (Table II). The slight discrepancy may be due to the fact that during the one year interval between the two measurements, the reference maser, which uses mercury diffusion pumps, was not maintained at high vacuum at all times. The error estimate for the earlier measurement was purely statistical without any consideration for systematic effects of undetected anomalous coatings on individual bulbs. Also, no allowance was made for any drift over the year. In the present experiment, the error in $\delta\nu_R$ is obtained by combining the standard deviation of the measured values $\Delta\nu$ and $\Delta\nu'$, and the uncertainty in the knowledge of B, in standard ways. No arbitrary choice is made and all the possible errors are considered. The slight downward adjustment of the frequency shift of the reference maser gives even closer agreement between masers at Harvard and the National Bureau of Standards, Boulder, Colorado.[6]

Next to the fact that this method of frequency shift determination, by using a deformable storage bulb, reduced possible systematic errors, another improvement in regard to the usual method is the tremendous gain of time required for the measurements. A future application of this method is the use of a deformable bulb with this geometry on the Large Storage Box Maser.[8]

ACKNOWLEDGMENTS

The author wishes to thank Professor N. F. Ramsey for his hospitality during this past year, Dr. R. Vessot and the Smithonian Astrophysical Observatory who provided a maser used during the preliminary experiments, and P. W. Zitzewitz, P. C. Gibbons, D. J. Larson and D. J. Wineland for many useful discussions and helpful criticism.

REFERENCES

* Work supported by the "Laboratoire Suisse de Recherches Horlogeres," Switzerland, the U. S. Office of Naval Research, and the U. S. National Aeronautics and Space Administration.

† Present address: Laboratoire Suisse de Recherches Horlogeres, rue Breguet 2, 2000 Neuchatel, Switzerland.

1. N. F. Ramsey, Molecular Beams, p. 11 (Oxford Press, London, 1963).

2. B. S. Mathur, S. B. Crampton, D. Kleppner and N. F. Ramsey, Phys. Rev. 158, 14 (1967).

3. R. Vessot, et. al., I.E.E.E., IM - 15, 165 (1966).

4. Ch. Menoud and J. Racine, Z. angew. Math. Phys. (ZAMP), 20, 578 (1969).

5. P. W. Zitzewitz, E. E. Uzgiris and N. F. Ramsey, Rev. Sci. Instr. 41, 81 (1970).

6. H. Helwig, R. F. C Vessot, M. Levine, P. W. Zitzewitz, H. E. Peters, D. W Allen and D. J. Glaze, to be published at June 1970 Conference on Precision Electromagnetic Measurements.

7. D. Brenner, Bull. Am. Phys. Soc. 14, 943 (1969).

8. E. E. Uzgiris and N. F Ramsey, Phys. Rev. 1, 429 (1970).

Table I: Characteristics of the Deformable Bulb

Parameters	Small Volume	Large Volume
Mean free path d	6.67 cm	8.67 cm
Volume V	1168.20 cm^3	1517.36 cm^3
Surface A	699.95 cm^2	
Volume ratio B	1.299	
Phase shift ϕ	10.7×10^{-6} rad.	

Table II: Comparison of Two Determinations of the Frequency Shift

All values are referred to the same 15.25 cm. diameter bulb, at 31.5°C. K is equal to the product of the frequency shift by the diam. of the bulb.

Work	K (mHz.cm)	ϕ (radian)	$\delta\nu_R$ (mHz)
Zitzewitz	-342.7±7.9	$-(5.67\pm0.13)\times10^{-6}$	-22.47±0.52
Present	-321±35	$-(5.32\pm0.59)\times10^{-6}$	-21.07±2.32

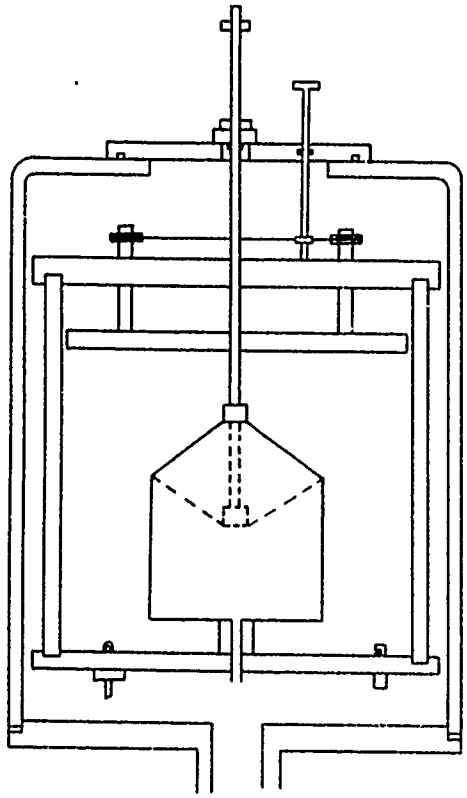


Fig. 1: Upper part of the maser with deformable storage bulb.

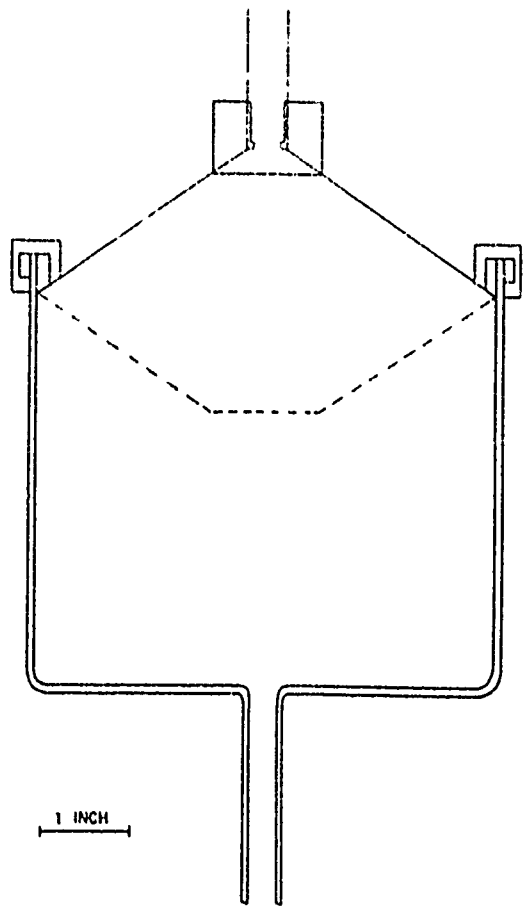


Fig. 2: The deformable storage bulb.

SURFACE COLLISION FREQUENCY SHIFTS IN THE ATOMIC HYDROGEN MASER*

Paul W. Zitzewitz
Department of Physics
Harvard University
Cambridge, Massachusetts 02138

Summary

Series of measurements of the wall shift in the hydrogen maser have been carried out over a wide range of temperatures in order to elucidate the causes of this frequency offset. Coatings of both the Teflon homopolymer (DuPont TFE-42) and the copolymer (FEP-120) have been used, with the same general results. There is a region below about 200 K where the wall shift is rapidly becoming less negative as the temperature is increased. Between about 200 and 360 K the shift becomes less negative at a slower rate, about 1% of the room temperature value per degree. Above about 350 K for FEP and about 390 K for TFE the slope again increases. The wall shift passes through zero near 100°C for both substances.

A theoretical model of the wall shift has been developed based on an adiabatic collision with two-body interactions which predicts the correct room temperature shift, but gives a slope much smaller than observed. The conclusions drawn from this work are that the change in sign of the wall shift suggests that the wall shift is the result of an adiabatic collision rather than physical adsorption. The abrupt changes in slope occur at temperatures at which phase changes are known to occur in bulk Teflon. Thus the wall shift has been linked directly to observable properties of the surface material.

Introduction

The frequency offset in the hydrogen maser due to collisions of the hydrogen atoms with the surface is the systematic shift most difficult to measure and control, and the one that contributes to most of the uncertainty in the hydrogen hyperfine separation. An experimental program has been carried out with aims of both trying to understand the nature of the wall shift of the surfaces now in use and also of preparing the groundwork of searches for new surfaces.

The wall shift can be written as a product of the rate of surface collisions, n, and the phase shift per collision, ϕ. Thus:

$$\delta\nu_w = (1/2\pi)n\phi \qquad (1)$$

The collision rate, $n = \bar{v}/\lambda$, is a function of atomic velocity and bulb geometry alone, while the phase shift is characteristic of the atom-surface interaction, and thus is the quantity of greatest interest.

The result of any experiment is actually the difference in wall shifts of two masers. The calibration of the reference maser was described at this meeting last year.[1]

The experiments carried out in this work were done on a maser which is a Varian H-10 extensively modified to obtain higher pumping speed and to permit operation between 90 and 400 K.[2] In the region above room temperature electrical heaters are used to maintain the desired temperature within a tenth of a degree. Below room temperature a cavity has been used which has hollow walls that can be filled with liquid nitrogen. This allows operations at temperatures near 90 K. Moreover, after the liquid has evaporated special measurement techniques allow measurements as the maser warms up. In this region gradients limit temperature determinations to about 4°C. One further fixed point is obtained by circulating cold (284 K) water in the cavity walls.

Wall Shift Theory

A theoretical model of the hydrogen-surface interaction has been developed which allows calculation of the wall shift for model Teflon surfaces. It is based on two assumptions. First, that the collision is adiabatic, meaning that energy loss at the surface is small enough to prevent trapping, or physical adsorption. Second, that the total hydrogen-surface interaction can be found by means of a summation over separate two-body potentials between the hydrogen atom and individual surface atoms or groups of atoms.

The surface to be treated, Teflon, is a synthetic fluorocarbon which can be purchased either as a homopolymer, polytetrafluoroethylene, such as DuPont TFE-42, or as the copolymer with hexafluoropropylene, as DuPont FEP-120. The former is composed of long chains of $-CF_2-$ groups, the latter of long chains with about 95% $-CF_2-$ and 5% $-CFCF_3-$. In either material the chains are folded back and forth on themselves to form broad, thin crystals called lamellae, in which the chains are primarily perpendicular to the broad crystal faces. Not all material is included in regions of high crystalline order, however, and Teflon is best characterized as a semi-crystalline polymer.

There are two effects of the interaction between the hydrogen atoms and a $-CF_2-$ group. First, the interaction itself determines the motion of the atom. Second, the interaction changes the hyperfine energy of the hydrogen, giving rise to the shift in phase of the oscillating moment. The interaction potential is described by the familiar Lennard-Jones 6-12 potential. The van der Waals attractive force, due to mutually induced electric dipoles, is included in the R^{-6} term. The "Pauli exclusion"

force, causing the repulsion when the electron clouds begin to overlap, is approximated by the R^{-12} term. The two parameters of the potential, well depth and equilibrium separation, are estimated by the use of rules allowing the combination of parameters describing H-H and CF_2-CF_2 interactions. These in turn are found partially from theory and partially from experimentally based empirical potentials.

The dependence of the hydrogen hyperfine energy on the H-CF_2 interaction potential can be seen qualitatively by noting that the hyperfine energy, as described by the Fermi contact term, depends on the density of the electronic wavefunction at the hydrogen nucleus. Thus the attractive van der Waals force which expands the wavefunction lowers the hyperfine frequency. The opposite shift is caused by the compression of the electron cloud by the repulsive forces. The former shift, the negative one, is rigorously proportional to the van der Waals potential, and the proportionality constant can be computed accurately. The positive shift is too complicated to be computed at this time, and so it is assumed to be similarly proportional to the repulsive potential.

The total interaction potential between the hydrogen atom and the surface is then found by summing the pairwise potentials over all -CF_2- groups in the Teflon surface. Because a sum over all groups individually is impractical, but a simple integration over smeared-out interaction centers is inaccurate for forces with such a strong dependence on separation, a compromise is used. The potentials are summed over only seven -CF_2- groups which are assumed to lie on a semicircle, thus representing a fold in the molecular chain at the face of the lamella. The remainder of the semi-infinite surface is treated by an integration. This operation determines the hydrogen-surface potential at any separation, and thus, by simple proportionality, the hyperfine shift.

The total phase shift per collision is then computed by integrating the instantaneous hyperfine shift over the duration of the collision. Since this shift is known as a function of distance rather than time, the interaction potential and classical mechanics are used to replace dt by dz/v, where v is the atomic velocity. Thus the final integral, to be performed numerically, is

$$\phi = \frac{2}{\hbar}\int dE\,\rho(E,T)\int d\Omega\,\rho(\theta)\int_{z_0(E,\theta)}^{\infty} dz\;\delta E(z)\left\{\frac{2}{m}\left[E\cos^2\theta - V(z)\right]\right\}^{-1/2} \quad (2)$$

where V(z) is the atom-surface potential
E initial energy of the hydrogen
m mass of hydrogen atom
θ initial angle of approach
$\delta E(z)$ shift in hyperfine energy
T atom temperature
z_o distance of closest approach

and where the integral is averaged over all angles of approach, weighted by a Lambert Law distribution; and over all initial energies, weighted by a Maxwell-Boltzmann distribution. Because of computer limitations this final energy average was omitted.

The results are shown in Fig 1. The graph showing dependence on angle of approach gives results for the most likely potential depth, ϵ_o = 99.41 K, $\epsilon_o/3$, and $3\epsilon_o$, with the separate contributions of the negative and positive parts of the phase shift indicated. The second graph, showing the phase shift as a function of temperature, is averaged over all angles. The temperature dependence is quite small, about 0.2 percent of the room temperature value per degree. Moreover, inclusion of the energy integration would reduce this dependence even more.

Experimental Results

Data on the temperature dependence of the phase shift per collision in the region above room temperature were obtained for two separate bulbs of each type of Teflon. Coating techniques for FEP-120 have been previously described.[1] The TFE-42 dispersion, diluted 2:1 to 50% TFE-42, 10% Triton X-100 wetting agent, and 40% water, was applied in exactly the same way. The first coat was made easily, but additional coats were much more difficult to apply, and, because of the extremely high melt viscosity of the homopolymer, were often very rough.[2]

The results for both kinds of Teflon over the entire temperature range are shown in Fig 2. A more careful examination of the results is shown in Fig 3, where the residuals from a statistically weighted least square fit of data in the room temperature regions are given. The temperature regions included in each fit are indicated by the solid lines. In the region below room temperature only one coat of each Teflon was tested. The room temperature dependence was first measured, then the wall shift determined from 90 to 250 K, finally the points near 280 K found. Residuals to fits in the room temperature region for the FEP coat are shown in Fig 4, while residuals of a fit of points in the 170-250 K region for the TFE coat are given in Fig 5. For this coat the room temperature results are fit separately, as shown. Parameters for the function ϕ = AT+B for specific temperature ranges are shown in Table I below:

	Bulb	Temps. (K)	A (μrad/K)	B (μrad)
TFE	0.4 s	295-347	0.1294	-48.53
	3.2 s	293-363	0.1248	-46.16
	Lo T	285-311	0.1266	-46.28
	Lo T	170-250	0.1107	-40.35
	Lo T	98-165	0.4680	-100.42

	Bulb	Temps. (K)	A (μrad/K)	B (μrad)
FEP	4"	358-392	0.1081	-40.00
	4"	297-329	0.0747	-28.70
	6.5"	366-411	0.1102	-41.57
	6.5"	297-335	0.0609	-23.83
	Lo T	206-312	0.0719	-28.08
	Lo T	99-163	0.4126	-90.62

Table I Least Squares Fit Parameters

Here bulbs are identified as in the graphs, with Lo T standing for the two bulbs used in the low temperature experiments. Since only one TFE point was definitely above the high temperature break, no line parameters can be given for that region.

Discussion of Results

There are three major conclusions that can be drawn from these results: a) The slope is much larger than predicted by the theory, being about one percent of the 300 K phase shift for both Teflons. As no reasonable variation of the parameters in the theory leads to large enough slopes, some interaction must have been omitted from the model. The theory presented here is essentially equivalent to the two-body interaction causing the pressure shifts in optical pumping, with the summation mainly adding to the magnitude of the negative term. In the optical pumping work experiment and theory of temperature dependence are in reasonable agreement. Thus it seems likely that effects associated with the surface itself (many-body effects) must be added to the present model. b) The phase shift per collision passes through zero near 100°C. The crossing is smooth, without any trace of the asymptotic behavior expected of a process involving physical adsorption of the atoms at the surface. Thus the adiabatic model, described above, in which the total phase shift is the balance of positive and negative shifts, is strongly supported. Our experience with these bulbs further suggests that some of the lack of reproducibility in the phase shift from one coating to the next is smaller at higher temperatures. This is especially true of discrepancies due to increased collision rates due to higher than expected surface areas. In this case the phase shift per collision is essentially rotated about the zero crossing. That is, B/A is independent of collision rate. Further discussion of the implications of the high temperature results will be made below. c) The phase shift per collision undergoes three distinct transitions. These occur at temperatures at which phase changes in Teflon exhibit known relaxations. These relaxations, associated with the onset of molecular motion, occur at temperatures dependent on the time scale of the experiment observing them. The experimental time (and frequency) dependence follows a model of chemical activation:

$$f \propto (kT/2\pi h)\exp(\Delta S^{\ddagger}/R)\exp(-\Delta H^{\ddagger}/RT) \qquad (3)$$

where f is the experimental frequency, $\Delta S^{\ddagger}$ and $\Delta H^{\ddagger}$ the activation entropy and enthalpy of the particular relaxation mechanism. The size of these parameters can be used to find the number of atoms involved in the motion.[3] The highest temperature relaxation, labeled α, is seen as the increase in slope of the phase shift near 350 K in FEP and near 390 K in TFE (Fig 3). This relaxation is associated with the onset of motion of large numbers of $-CF_2-$ groups, and usually is thought to be most important in the polymer regions of lowest crystalline order. It occurs at lower temperatures in FEP because the CF_3 groups replacing the F atoms on a few $-CF_2-$ groups tend to reduce interchain attraction, as evidenced by the lower melting point and smaller melt viscosity in the copolymers. The interchain attraction is primarily due to van der Waals forces, and thus it is likely that this relaxation either causes or is caused by a weakening of these forces. Thus it is not surprising that the negative portion of the phase shift per collision, due to the van der Waals forces, is reduced above this relaxation temperature.

The next relaxation, in order of decreasing temperature, is β. This relaxation is associated with the well known changes in crystalline order occuring at 19 and 30° C. in TFE. In polymer experiments this relaxation is extremely complicated, and usually occurs at temperatures below the crystalline transition temperature. Again, because of the reduced crystal order in the copolymer, it occurs at lower temperatures in this substance. In the phase shift data there is some evidence for a small discontinuity in the TFE data below 280 K, and there is probably a shoulder near 210 K on the low temperature transition in the FEP data. At this time there is no explanation for the phase shift changes in terms of polymer structural changes.

The third relaxation, γ, is associated with the onset of motion of very small numbers of $-CF_2-$ groups, and is almost independent of copolymer concentration. In our data it is seen as the tremendous change in the slope of the phase shift between 170 and 190 K in both Teflons. In terms of the properties of bulk Teflon, this temperature is the primary glass transition temperature, below which the material is a rigid glass-like substance, above which it behaves like a plastic. Again it is difficult to relate these properties to the wall shift, but examination of the theory of thermal accommodation, that is, the probability of loss of energy of the hydrogen atoms during a collision, suggests that below this temperature physical adsorption is more likely to occur than above it. The hydrogen atoms may be trapped on the wall in the lowest temperature region, suffering increased negative phase shifts while in the attractive region of

the potential. An alternative possibility is that impurity gases in the system condense on the wall at these temperatures, changing the nature of the surface adversely. Due to the poor quality of the experimental data in this range, as well as the complicated nature of the system, no conclusions can be reached at this time.

As has been mentioned, the temperatures of all relaxations are dependent on observation frequency. By plotting the temperature-frequency relationship of Eq. (3), using parameters from the literature,[3] as in Fig 6, the type of mechanism acting in the maser can be suggested. The graph clearly indicates a phenomenon near 10^{-2} Hz for all relaxations. Because of lack of polymer experiments in this frequency range, as well as possible breakdown in the theoretical model, no great importance should be attached to this frequency. It does suggest, however, that all measurements are consistent, and that a low, possibly zero, frequency mechanism is involved.

Other Experimental Work

Although this work was not primarily concerned with the measurement of hydrogen atom surface relaxation, some work was done with the new homopolymer sample, TFE-42. Polymer studies[3] have shown that it is likely that the end groups of all molecular chains consist of impurities of the type CF_2H. Since it is known that a hydrogen containing surface has large relaxation rates, and thus leads to large atomic linewidths, it would be advantageous to reduce surface relaxation by choosing a Teflon with as small a proportion of end groups as possible. Since TFE-42 has a molecular weight, and hence a chain length, ten times that of FEP-120, this material would seem a likely candidate. Studies of polymer crystallization, however, suggest that the number of end groups on the surface of the polymer crystals would be reduced by a factor of only three for the homopolymer. Still, a decrease in surface relaxation is indicated.

The temperature dependence of the surface relaxation rate for Drifilm and FEP Teflon was measured by Berg,[4] and thus a similar, though less precise, study was made of the new Teflon. Berg's work showed that the probability of relaxation per collision could be fit by a chemical activation model, $p = a \exp(-b/kT)$, where p is the probability of relaxation per collision, b the energy of activation of the chemical process, and T the temperature. Although the work on TFE is not accurate enough to distinguish between this model and others, the relaxation was larger at higher temperatures, but not as large as FEP. The results for the TFE surface is compared with Berg's work in Table II:

Surface	b (eV)	p (300)K
TFE	0.032 ± 0.021	(1.67 ± 0.17)x10^{-5}
FEP-1	0.089 ± 0.008	(1.20 ± 0.06)x10^{-5}
FEP-2	0.053 ± 0.004	(1.88 ± 0.05)x10^{-5}
Drifilm	0.085 ± 0.002	(22.0 ± 0.2)x10^{-5}

Table II. Surface Relaxation Results

In the table p(300K) is the probability of relaxation at 300 K. One immediate conclusion is that the lower relaxation at high temperature permits much more satisfactory operation of the maser at temperatures where the wall shift is zero.[5] A conclusion that a different relaxation mechanism is involved because of the different activation energy is premature, as the different temperatures for the α relaxation put the TFE data all below that temperature but the FEP on both sides. Also the FEP data was taken on a sample which had a higher wall shift than the samples now in use.[1]

The probability of relaxation is not decreased by the factor of three as hoped. There is one further possibility of decreasing it, however. It has been noted both in this work and in others[6] that the lack of true melting of the TFE dispersion particles leads to rough surfaces which can have an actual surface area much larger than that expected from geometry. Thus the collision rate is higher than calculated and so is the wall shift and relaxation. Thus the development of techniques giving smoother surfaces may lead to both smaller wall shifts and narrower linewidths.

In addition to this work on Teflon, a measurement of the temperature dependence of the Drifilm wall shift was made. The shift becomes more negative at higher temperatures, a result opposite to both the Teflon experiments and the general expectations of the theory. However, the theory considered only kinematical (motional) temperature effects. It is possible that in Drifilm the higher temperature atoms, in penetrating further into the surface material, enter regions of higher perturbing potentials than is the case of Teflon, where the tight fluorine sheath around the carbon core prevents such penetration.

Directions of Future Work

The applications of this study for frequency control work has already been started by other workers.[5] For better understanding of the nature of the hydrogen-surface interaction two sets of future experimental programs are suggested. First, measurement of the surface relaxation and surface frequency shift on the same coating as a function of temperature should help to clarify the nature of both phenomena and

also their relation to the phase changes in the Teflon material. Second, there are several quite pure fluorocarbon substances available that would condense to form solid surfaces at temperatures near 90 K. A comparison could be made between the wall shift (hopefully as a function of temperature) and the pressure shift, obtained at temperatures above the condensation point of the gases. This comparison would illuminate the difference between the two-body collisions leading to the pressure shift and the many-body effects acting in the wall shift.

Acknowledgments

I wish to thank my thesis advisor, Professor Norman F. Ramsey, for support and encouragement during this work. Discussions with Professor Stuart Crampton and Dr. Robert Vessot have been most illuminating. A Danforth Graduate Fellowship during the course of this work has provided financial support, for which I am very grateful.

References

* Research supported by the National Aeronautics and Space Administration, the National Science Foundation, and the Office of Naval Research.

1. E. E. Uzgiris and P. W. Zitzewitz, Proceedings 23rd Annual Symposium on Frequency Control, 1969. P. W. Zitzewitz, E. E. Uzgiris, and N. F. Ramsey, Rev. Sci. Instr. 41, 81 (1970). For a recent re-measurement, see the preceding paper by P. E. Debély.

2. P. W. Zitzewitz, Thesis, Harvard University, 1970 (unpublished).

3. For example, see C. A. Sperati and H. W. Starkweather, Fort. Hochpolym.-Forsch. 2, 465 (1961), and R. K. Eby and F. C. Wilson, J. Appl. Phys. 33, 2951 (1965).

4. H. C. Berg, Phys. Rev. 137, A1621 (1965).

5. R. F. C. Vessot, following paper.

6. H. Hellwig, Private communication.

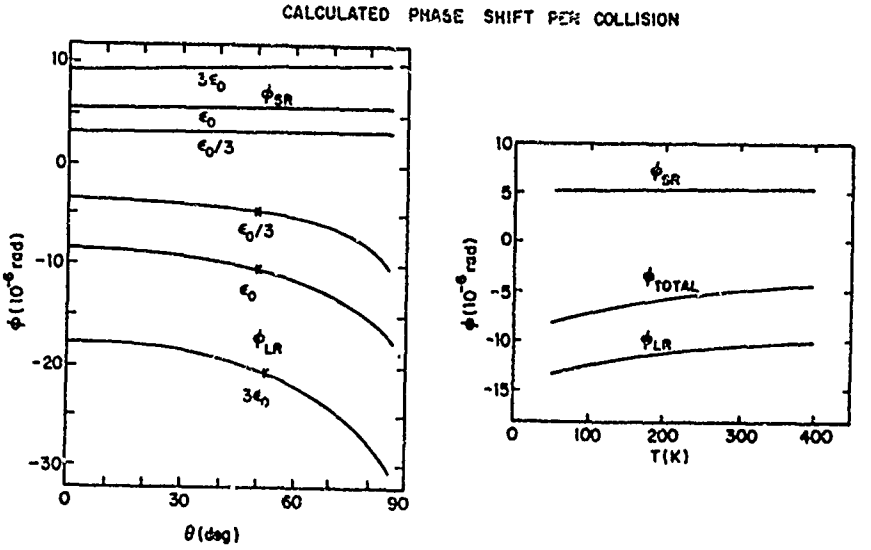


Fig. 1. Calculated Phase Shift per Collision

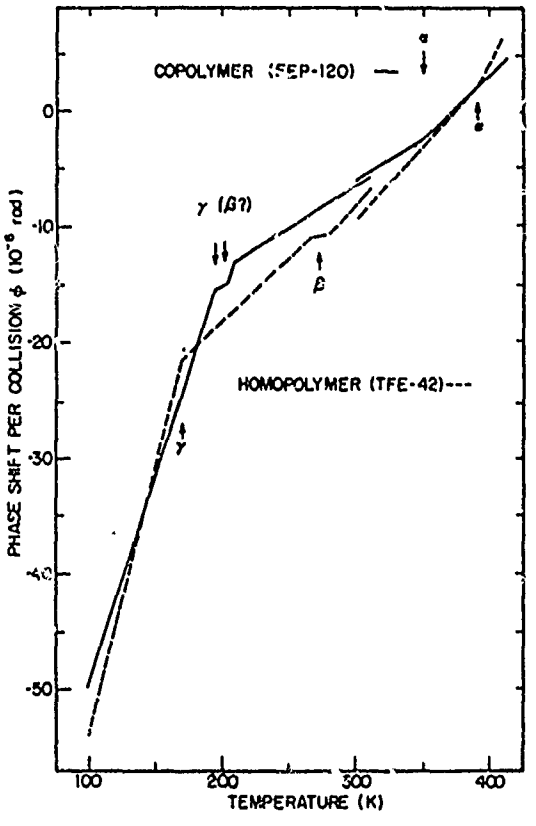


Fig. 2. Phase Shift per Collision as a Function of Temperature

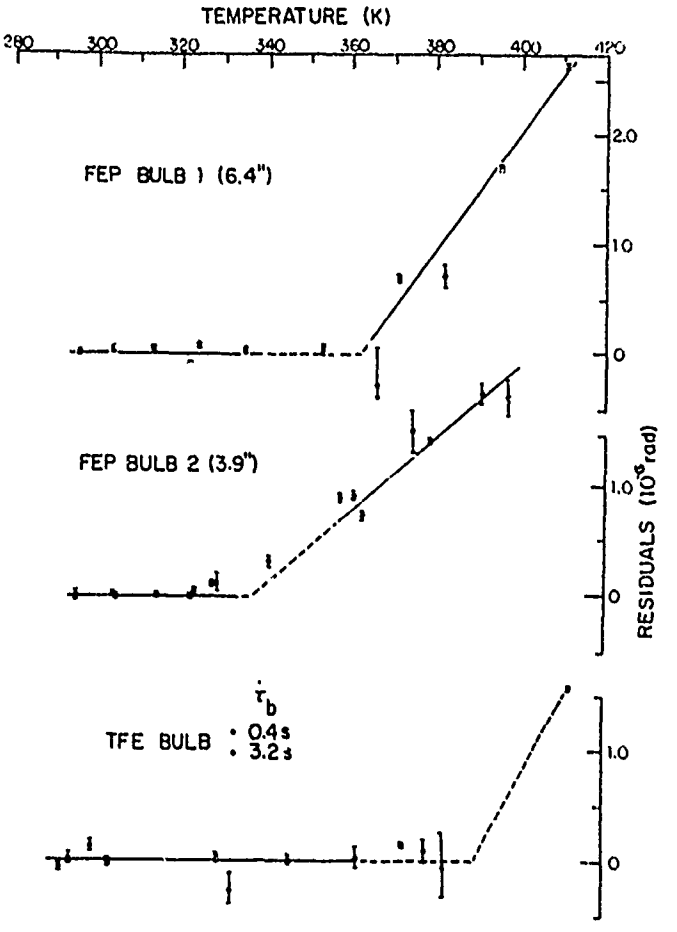


Fig. 3. High Temperature Residuals

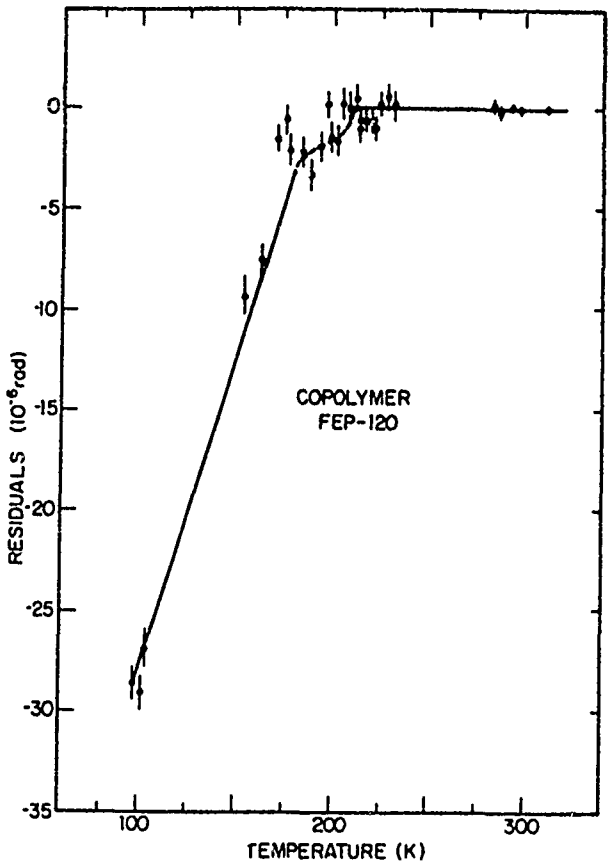
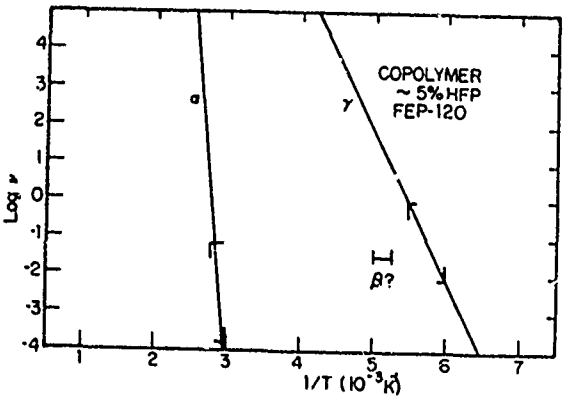


Fig. 4. Low Temperature Residuals - FEP



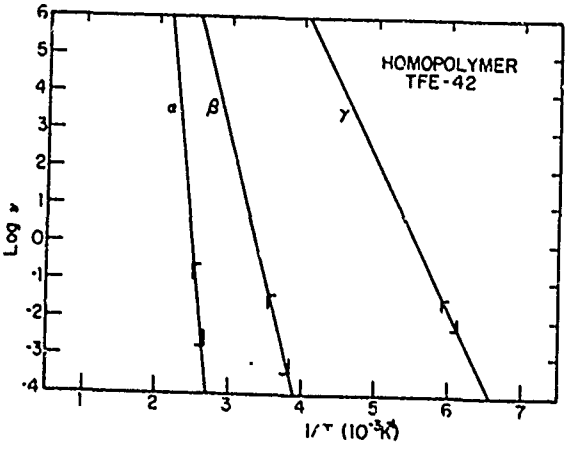


Fig. 6. Teflon Structural Relaxation Spectra

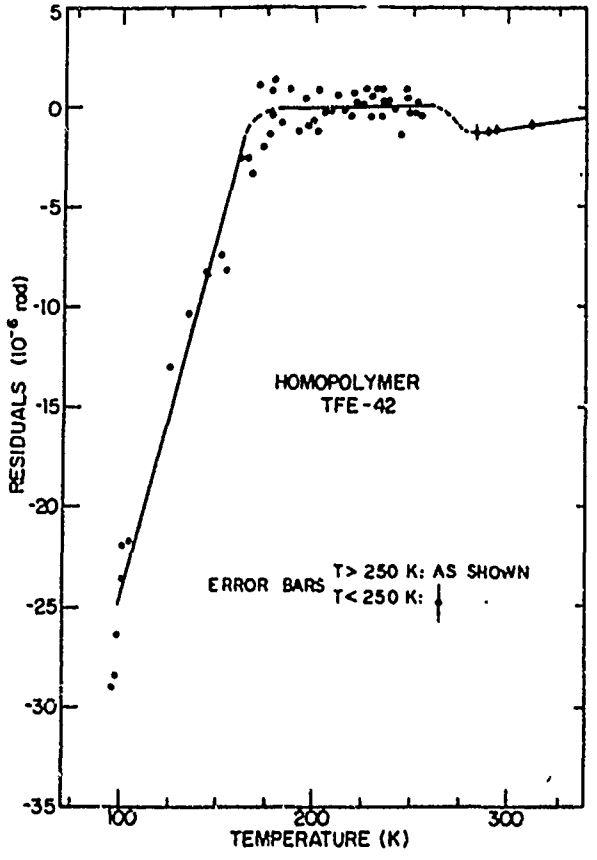


Fig. 5. Low Temperature Residuals - TFE

STUDIES OF HYDROGEN MASER WALL SHIFT FOR HIGH MOLECULAR WEIGHT POLYTETRAFLUOROETHYLENE*

R.F.C. Vessot and M.W. Levine
Smithsonian Astrophysical Observatory
Harvard College Observatory
Cambridge, Massachusetts

Summary

The use of high molecular weight Polytetrafluoroethylene (PTFE) as a wall coating for the atomic hydrogen maser appears to offer some real advantages. By operating the maser bulb at a temperature where the wall shift is zero the effects of bulb size, shape and the surface texture of the coating can be eliminated. Coatings of PTFE can be applied in two quite distinguishable conditions; quenched and annealed. It is believed that the quenched coatings offer good reproductivity of the zero wall shift temperature. The annealed coating has a higher zero wall shift temperature and appears to offer somewhat lower relaxation rate than the quenched coating.

The results from some initial experiments are presented to support these contentions. Further work on wall coatings is in progress.

Introduction

Recent studies[1] of the temperature dependence of the wall shift in the atomic hydrogen maser have shown that the average phase shift per collision can be made positive or negative with respect to the phase of the free hydrogen oscillating dipole moment. We expect that the temperature at which a zero average phase shift per collision occurs is a property of the chemical nature of the walls and not of the mechanical surface conditions. By operating at this temperature, masers with bulbs of any size and surface condition should oscillate at the same frequency, provided that their surfaces are of the same type of material and free of contamination. The surface conditions will affect the relaxation rate and for this reason the wall collision rate should be kept as small as possible.

High molecular weight Polytetrafluoroethylene (PTFE) films cast from water dispersed resin are of particular interest since these materials have, on the average, about ten times the molecular weight of the tetrafluoroethylene (FEP) copolymer most often used as a wall coating material. The PTFE material should have fewer end groups exposed to the atomic hydrogen in the bulb. The end groups, chiefly CF_2H can contribute a significant area of sites containing bound hydrogen that result in larger negative values of phase shift per collision than the fluorine sites and can also contribute substantially to recombination processes (and consequently to relaxation) since the end-group hydrogen atom is much less energetically bound than the fluorine atoms. The tetrafluoroethylene homopolymer can be almost completely crystallized by the very careful and slow annealing through the range 340°C - 295° C. This should result in a 30-fold reduction in porosity to gases and the possibility of much reduced hydrogen collision rates.

Studies of PTFE[2] have shown that the change in the ordering of the amorphous PTFE teflon known as the alpha transition is at a higher temperature than the temperature for zero wall shift making this temperature a viable operating point.

Experimental Techniques

The Reference Oscillator

A reference standard was used to obtain the wall shift data and relate the data taken over a time interval of several months. The reference oscillator used to provide the frequency datum was compared with the Harvard reference maser and with the N.B.S. Maser using a traveling maser frequency standard. This measurement is described by Hellwig et al (1970)[3] and is part of an experiment to determine the unperturbed Hydrogen Hyperfine Transition Frequency. The wall shift of the reference oscillator was determined to be $-.022 \pm .002$ Hz at 319°K. This value has been used in all the measurements.

Working Maser

The working maser used for measuring the effect of temperature on wall shift is a modified Varian H-10 series maser. The r.f. power supply to the dissociator discharge was made to be controllable so as to permit variation of the

* This work is supported by NASA under Contract NSR-09-015-098.

dissociator efficiency and consequently the flux rate of atoms into the bulb. By using automatic cavity tuning techniques[4] much of the effort required to tune the maser has been avoided.

Serious problems with magnetic quenching were encountered at high cavity and bell jar temperature due to thermal gradients across junctions where different metals came into contact. All the oven cylinders, the solenoid and the bell jar components were electrically insulated from each other. The problem has been almost completely eliminated and it is possible to operate over a range of magnetic fields from 0.7 to 100 millioersteds with very little variation of signal amplitude when the maser is oscillating near threshold. For wall relaxation measurements the residual magnetic quenching is still not entirely negligible.

The maser bulb and cavity were both enclosed in the maser vacuum system. The bulb temperature was measured by measuring the cavity temperature by means of a thermocouple after allowing a 4-day period for thermal equilibrium of the bulb-cavity-bell jar assembly.

Maser Bulbs

Bulbs of 7" and 4" diameter were used for taking wall shift data as a function of temperature.

These bulbs were equipped with integral collimators that were coated with the bulbs in order to be assured that hydrogen atoms would encounter only one kind of teflon surface. Figure 1 shows a bulb of 4" diameter with integral collimator.

Coating Procedures

The bulbs were chemically cleaned by using a water solution of nitric acid and rinsed with several changes of hot distilled water. No attempt was made to roughen the interior surface of the bulbs.

The coating material consisting of a mixture of 10% Triton X-100, 50% Teflon 42 (duPont), and 40% distilled water was carefully poured into the bulb to be coated, swirled around to cover all the interior surface and drained out. The wet bulb was then placed on a rack with a small diameter tube extending upwards into the collimator. A flow of approximately 1 cm^3/minute of dry nitrogen, at room temperature was allowed to flow into the bulb to dry the coating.

The bulb was then cured in a horizontal air flow oven for 20 minutes at 360°C. A very small flow of oxygen was piped into the bulb during the curing cycle. The oxygen was of high purity and was passed through a cold trap at -78.5°C (dry-ice) prior to entering the bulb. At the end of the 20 minute curing cycle the bulb was quickly removed from the oven and cooled by a draft of room temperature air. In this condition, the bulb is considered to be quenched.

The subsequent annealing cycle that is required for crystallizing the PTFE coating consists of placing the bulb in a specially controlled oven with the same oxygen flow and cooling the oven, in a steady rate, from 340°C to 290°C. The bulb was removed from the oven at 290°C and cooled in air to room temperature.

Experimental Data

Relation of Wall Shift to Phase Shift per Collision

The quantity of interest is the average phase shift per collision.

This quantity is related to the wall shift by the following expression:

$$\bar{\phi}(T) = \frac{2\pi \Delta\nu_w(T)}{\nu_c} \qquad \text{where}$$

$\Delta\nu_w$ is the wall shift and ν_c is the wall collision rate. The collision rate, ν_c is related to $\bar{v}$, the average velocity and the mean free distance λ by

$$\nu_c = \frac{\bar{v}}{\lambda}$$

For a sphere $\lambda = \frac{2}{3}R$ and $\bar{v} = \left(\frac{8kT}{\pi m}\right)^{1/2}$
The resulting expression relating average phase shift with bulb radius, wall shift and temperature is

$$\bar{\phi}(T) = 5.78 \times 10^{-4} \frac{R}{T^{1/2}} \Delta\nu_w$$

The data are plotted in figure 2 in terms of average phase shift per wall collision. It is seen that the 4" and 7" quenched bulb data lie very nearly on the same straight line. The least squares fit on each set of points is as follows:

7" bulb.

$$\bar{\phi}_7(T) = -53.36 \pm 1.37 + (0.1503 \pm .004)T$$
$$= 0 \text{ at } 355°K$$

4" bulb

$$\bar{\phi}_4(T) = -56.53 \pm 1.5 + (0.159 \pm .005)T$$
$$= 0 \text{ at } 355.1°K.$$

The same 4" bulb was then annealed and data on it are shown in the same figure. Since we wanted to get data in time for the conference we did not have time to allow the bulb and cavity to come to thermal equilibrium and consequently the scatter is larger than for the previous data. We have for the 4" annealed bulb

$$\bar{\phi}_4(T) = -39.09 \pm 3.67 + (.0903 \pm .011)$$
$$= 0 \text{ at } 432.9°K.$$

In both cases the data excluded the 5 circled points in figure 2.

Summary of Data

Table 1 is a summary of these preliminary data and a comparison with other data.

It should be noted here that the Teflon 42 used in all the data shown in this table is from the same sample of Teflon 42 obtained by the Smithsonian Astrophysical Observatory from E.I. duPont de Nemours and Company on July 29, 1969. This sample, which was donated to SAO by duPont, is identified as Blend #8892, % solids 34.3. It should further be noted that the coating procedure used by Hellwig[3] was similar to the quenched procedure described above except for the very rapid cooling employed in the present experiment.

It is possible to extrapolate to zero wall shift from data shown in the "quenched" plot of Figure 2. By using the least square fit to a straight line for the 4" and 7" bulbs quenched data, recalculating the wall shift at 4 temperatures and plotting the results vs. the reciprocal of the bulb diameter one obtains the plot shown in Figure 3.

Figure 3 illustrates an interesting point and leads us to a new method of determining the wall shift. This figure is, in fact, a plot of wall shift versus collision rate. Collision rate depends on the bulb geometry and on the collimator geometry; the effect of the latter is still not well understood.

In making wall shift vs collision rate plots using several bulbs we make the following two assumptions:

1. That the walls of the two bulbs are exactly the same;

2. That we know the collision rate of the hydrogen atoms in each bulb.

The plot of wall shift vs 1/D at different temperatures shows us that the convergence point is not at 1/D = 0 but somewhere near 1/D = .008. Here the convergence is good and lies within 2 mHz of our calibration. Judging from the data it appears that we are making a consistent error in the prediction of the collision rate. However, if we use our knowledge of the temperature at which the observed wall shifts are equal we have no longer to depend on our knowledge of the collision rate. We depend only on the similarity of the teflon used in the two bulbs. The data is in the form

$$\nu(T) = \Delta\nu_e + \Delta\nu_w$$
$$= \Delta\nu_e + \left(\frac{2kT}{\pi^3 m}\right)^{1/2} \frac{\bar{\phi}(T)}{\lambda}$$

where $\Delta\nu_e$ is an unknown offset error and $\Delta\nu_w$ is the wall shift. λ is mean free path of atoms in the bulb and collimator. By changing λ, either by changing bulbs (or by using a deformable bulb as suggested by Brenner[5]) we see that the only condition making $\nu(T)$ the same for the two values of λ is to make $\bar{\phi}(T) = 0$.

To illustrate this technique we have reduced the data discussed earlier in a different way.

We find that

$$\frac{\nu(T)}{T^{1/2}} = \frac{\Delta\nu_e}{T^{1/2}} + \left(\frac{2k}{\pi^3 m}\right)^{1/2} \frac{\bar{\phi}(T)}{\lambda}$$

is quite linear with temperature when $\Delta\nu_e$ is small. For the two quenched PTFE bulbs, we have that

$$\frac{\nu_7(T)}{T^{1/2}} = \frac{\Delta\nu_e}{T^{1/2}} + \left(\frac{2k}{\pi^3 m}\right)^{1/2} \frac{\bar{\phi}_7(T)}{\lambda_7} = C_1 + \alpha_1 T \quad (1)$$

$$\frac{\nu_4(T)}{T^{1/2}} = \frac{\Delta\nu_e}{T^{1/2}} + \left(\frac{2k}{\pi^3 m}\right)^{1/2} \frac{\bar{\phi}_4(T)}{\lambda_4} = C_2 + \alpha_2 T \quad (2)$$

Under the assumption that $\Delta\nu_e$ is the same for both (1) and (2) and that

$$\bar{\phi}_7(T) = \bar{\phi}_4(T).$$

crossover occurs where $\bar{\phi}(T) = 0$ at a temperature $T_0 = C_2 - C_1/\alpha_1 - \alpha_2$.

Recalculating the calibration error, we have

$$\Delta\nu_e = T_0^{1/2}(C_1 + \alpha_1 T_0)$$

From the same data used to plot Figure 2, we obtain a temperature of intercept of 355.3°K and a $\Delta\nu_e$(error)=0.2 x 10^{-3} Hz.

Wall Relaxation

So far, we have made wall relaxation measurements only using 4" bulbs. Our results confirm the data obtained by Zitzewitz for PTFE. At temperatures up to 100°C we see no increase in wall relaxation rate. Relaxation rate measure-

ments are being continued at the Smithsonian Astrophysical Observatory using larger bulbs with very long bulb storage time.

Conclusions

(1) From the data taken so far we conclude that it is possible to operate a maser with its bulb at 355°K (82°C) and substantially reduce the wall shift and its contribution to the uncertainty of the maser output frequency.

(2) It is possible to eliminate the uncertainty due to collision rate in the determination of wall shift by measuring wall shift as a function of bulb temperature using several bulbs of different sizes.

References

(1) "Preliminary Report on the Temperature Dependence of the Wall Shift of the Hydrogen Maser for FEP and TFE-Teflon." Paul Zitzewitz, Harvard University, October 2. 1969.

(2) "Fluorine-Containing Polymers. II. Polytetrafluoroethylene." C.A. Sperati and H.W. Starkweather, Jr. Advances in Polymer Science, Springer-Verlag, Berlin 1961.

(3) "Measurement of the Unperturbed Hydrogen Hyperfine Transition Frequency." H. Hellwig, R. Vessot, M. Levine, P. Zitzewitz, H. Peters, D. Allan, D. Glaze. Conference on Precision Electromagnetic Measurements, Boulder, Colorado, June 1970.

(4) "The Design of an Atomic Hydrogen Maser System for Satellite Experiments." R. Vessot, M. Levine, L. Mueller and M. Baker. Proceedings of the 21st Annual Symposium on Frequency Control. U.S.A.E.C.O.M. Ft. Monmouth, N.J., April 1967.

(5) "Flexible Bulb Method of Measuring Wall Shift of the Hydrogen Maser." D. Brenner, Bull. Am. Phys. Soc. 14, 943 (1969).

Table I

Parameter	Present Data	Previous Data
Average phase shift Temperature	0.15 ± .004 Rad/°K P.T.P.E. quenched	Zitnewitz 1964 0.1296 0.1248
	0.09 ± .011 Rad/°K P.T.P.E. annealed	
Zero phase shift Temperature	355°K quenched 433°K annealed	375°K 369°K
Wall Shift, K' mHz-cm at 297°K	522 Hz-Cm	630.6 Hz-Cm 640.7 570.9 544.8
		Hellwig 1970 520 Hz-Cm

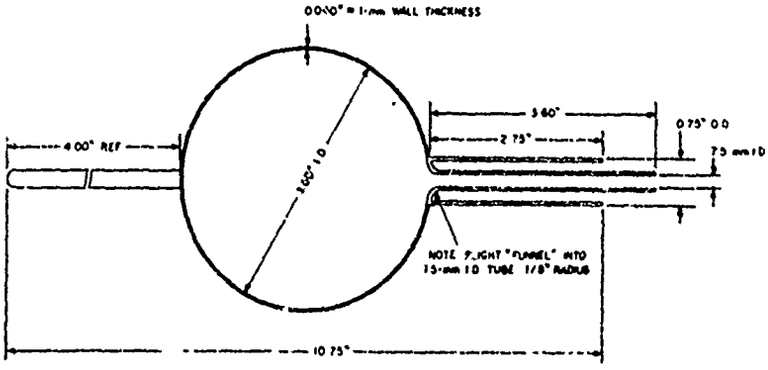


Figure 1 Four-in.-diameter fused silica storage bulb for wall-shift experiments.

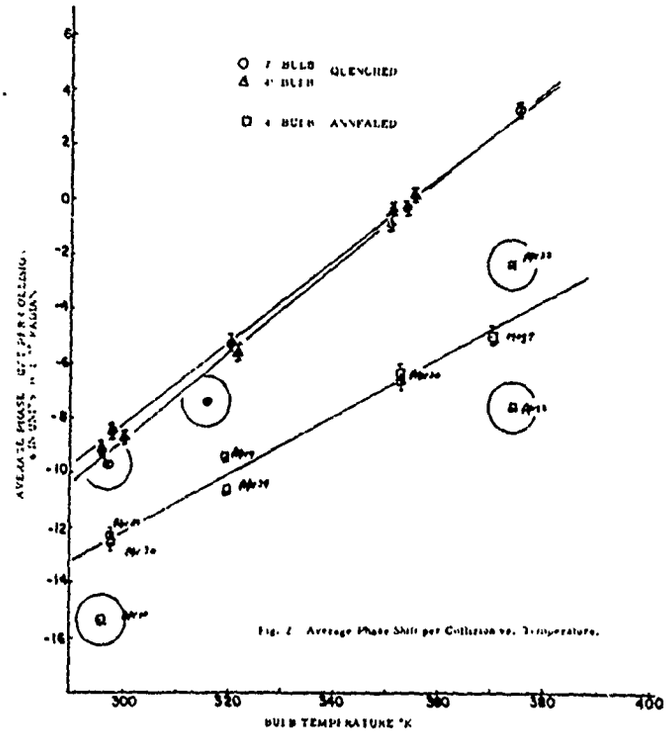


Fig. 2 Average Phase Shift per Collision vs. Temperature.

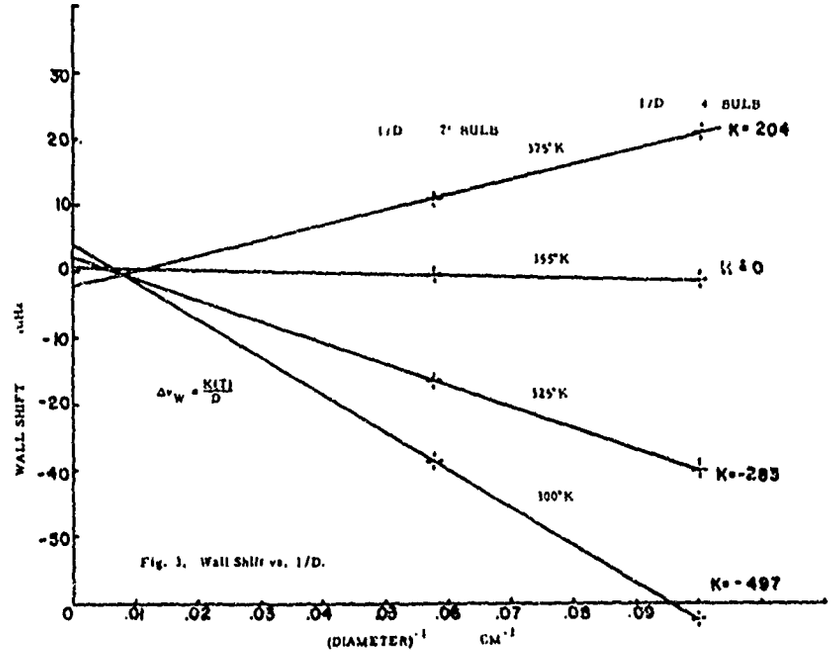


Fig. 3. Wall Shift vs. 1/D.

FREQUENCY STABILIZATION OF CO_2 LASERS WITH RESPECT TO SF_6 AND CO_2 LINE CENTERS

P. Rabinowitz, R. Keller and J. T. LaTourrette
Department of Electrophysics
Polytechnic Institute of Brooklyn
Farmingdale, New York 11735

Summary

Two CO_2 lasers were separately stabilized with respect to either an SF_6 absorption cell or to the maximum of the CO_2 laser power output (the "CO_2 line center"). The beat frequency between the two lasers was measured as a function of the gas pressure and laser input power. A pressure shift in CO_2 was observed which varied from 80 to 150 kHz per Torr for various mixtures of CO_2-He or CO_2-N_2-He. The grating used for transition selection in one CO_2 laser could pull the "CO_2 line center" frequency by as much as ±3 MHz. The pressure shift in SF_6 was not observable.

Introduction

The purpose of this investigation is to assess the relative merits of SF_6 or CO_2 itself as a reference for frequency stabilizing the CO_2 laser. In the case of SF_6, the CO_2 laser is stabilized with respect to the peak in the transmission in an external SF_6 absorption cell which results from the saturation dip in double pass absorption.[1,2] This accurately defines the SF_6 absorption line center.[2] The CO_2 line center was defined by the maximum of the laser output power. In each case, frequency modulation of the laser was employed to find the line center.

Description of the Apparatus

Two CO_2 lasers with good passive stability were used in this work. One mirror of each laser was mounted on a piezoelectric transducer to allow electrical control of oscillation frequency.[3] Each laser could be frequency stabilized with respect to the Lamb dip in an external SF_6 absorption cell.[2] In one of the two lasers, a diffraction grating was employed as one reflector to provide the selection of a single CO_2 transition. The other laser was sufficiently short so that the resonator intermode spacing ($c/2L \approx 230$ MHz) was much larger than the range of dominant mode oscillation on one transition ($\Delta\nu_{Doppler} \approx 50$ MHz). Oscillation on a single transition was obtained due to this frequency discrimination and because of the familiar rotational level cross-relaxation. However, the range of oscillation on any particular transition varied considerably with discharge conditions. In fact, only a few CO_2 transitions would oscillate in this laser. Furthermore, which particular transitions oscillated and their range of oscillation as well as the order in which oscillation occurred as the resonator length was varied (over a half wavelength range) depended on the overall length of the resonator and on discharge conditions. As a result, this laser was generally stabilized to the strong SF_6 dip on the CO_2 P(16) transition, and parameters of the other laser were varied.

The experimental arrangement used to stabilize each laser was similar to that shown in ref. 2, with two changes. The mechanical chopper was eliminated and the output of the lock-in amplifier was amplified and applied to the piezoelectric transducer along with a 5.7 kHz modulation signal. The stabilization feedback loop functions as a simple Type I servo with a cross-over frequency which could be varied up to about 100 Hz.

The 5.7 kHz frequency modulation of each laser was adjusted to be of approximately equal amplitude and phase to minimize the spectral width of the beat frequency between the two lasers. The beat frequency was obtained (see, e.g., ref. 3 but without the AFC feedback) by combining a portion of the output of each laser (obtained from the other end of the lasers) together on a third Ge:Au detector using a germanium plate as a beam combiner.

Stabilization on SF_6

In the first experiment, both lasers were locked to the same Lamb dip in SF_6, the strong dip on the CO_2 P(16). We observed that the beat frequency could be shifted by about ± 100 kHz by introducing a d-c offset voltage (a "bias") at the output of the PAR JB-5 lock-in amplifier. This shift is a measure of the slope of the "discriminant," which in turn is essentially the derivative of the Lamb dip as shown in ref. 2. The magnitude of the frequency shift with bias voltage depends on the gain of the system up to the point of introducing the bias. It thus depends on the laser power, the transmission of the cell and the PAR gain setting. However, when the system was adjusted to zero bias, i.e., to achieve zero correction voltage at the piezoelectric transducer for zero input to the lock-in amplifier, the observed beat frequency was independent of the CO_2 laser power and the electronic system gain. In addition, no consistent shifts in the beat frequency between the two lasers was observed as the SF_6 pressure was varied over 10-100 mTorr. Below ~10 mTorr, the beat frequency jumped to approximately 5 MHz - because the laser shifted over to lock on the CO_2 laser power maximum which is discussed below. This establishes a crude upper limit for the pressure shift in SF_6 of about 600 kHz per Torr which,

compared to the laser frequency of 3×10^{13} Hz represents a relative stability of 2 parts in 10^{11}/mTorr.

Stabilization on the CO_2 Line Center

For the remaining experiments, the SF_6 absorption cell for the "grating" laser was evacuated so that this laser was stabilized to the maximum of the CO_2 laser power output. The pressure in that laser was then varied and the beat frequency was recorded as a function of pressure. We observed, in fact, that the grating had a greater effect on the oscillation frequency than the laser pressure. As a result, we calibrated the tilt angle of the grating, and at each pressure, we recorded laser output power and beat frequency for each of eight or more angular positions of the grating.

The data were first plotted as laser power vs. beat frequency with the grating angle as a parameter, and the maximum of a smooth curve through these points was taken as the "0" grating angle. The measured values of beat frequency vs gas pressure for various grating angles are shown in Figs. 1-4. In Fig. 1, only the helium pressure was varied, with a fixed CO_2 pressure of 3 Torr. The helium pressure was increased by 3 Torr between each set of data points. This data is not inconsistent with the expected pulling of a broad resonator (i.e., the grating) shifting the maximum power output due to the narrower CO_2 laser gain profile. The pulling increases with pressure since the CO_2 line is pressure broadened. The data shown in Fig. 2 was obtained by decreasing the gas pressure (originally 3.0 Torr CO_2 + 15 Torr He) by approximately 2 Torr between sets of readings. As before, the grating pulling appears to vary linearly with pressure. However, it will be noticed that the extrapolation of the lines intersect at a positive pressure.

A similar behavior is obtained for the three component mixture (originally CO_2:N_2:He = (2.5 : 2.0 : 5.4) Torr, respectively) shown in Fig. 3. Several factors must be considered in analyzing this data. The gas temperature in the laser discharge tube causes a variation in Doppler width ($\Delta \nu_D \propto T^{1/2}$) and collision rate ($\Delta \nu_p \propto (nv) \propto p\, T^{-1/2}$). In addition, cataphoresis may cause a concentration gradient in the discharge tube. The connection to the manifold was at the cathode end. As a result, the mixture of the gas could vary as the pressure was reduced. Additional data for the three component (CO_2, N_2, He) mixture is plotted in Fig. 4, using the decrease in laser power from the peak power as a parameter rather than the grating angle. As can be seen, it corroborates the pressure shift deduced from the other data.

Conclusion

These results show that a grating (or other dispersive element) can significantly shift the frequency of a CO_2 laser which is stablized to the CO_2 line center, i.e., to the maximum of its power output. The grating pulling effect would be reduced if the reference provided by the CO_2 were narrower. Such a narrower linewidth CO_2 reference can be obtained from a CO_2 absorption cell as recently reported by Freed.[4] If the grating effect is avoided, the CO_2 line center provides a fairly stable frequency reference, the shift of CO_2 line center with gas pressure is: 80 ± 12 kHz per Torr of He, 75 ± 30 kHz per Torr of CO_2: He = 1.5, and 130 ± 50 kHz per Torr of CO_2; N_2. He = 2.5 : 2.0 : 5.4. The latter result would correspond to a pressure dependence of 4 parts in 10^{12} per mTorr. In comparison, only an upper limit to the pressure shift in SF_6 has been obtained, and it may indeed be much smaller. Also, one operates the SF_6 cell at a lower pressure, so that smaller pressure changes would be encountered. In addition, SF_6 provides a much narrower linewidth reference, so that it is relatively unaffected by a grating pulling effect. Finally, all pulling effects due to the laser could be avoided by using an external frequency modulator.[2]

Acknowledgement

This work was sponsored, in part, by the Joint Services Electronics Program under Contract No. F 44620-69-C-0047.

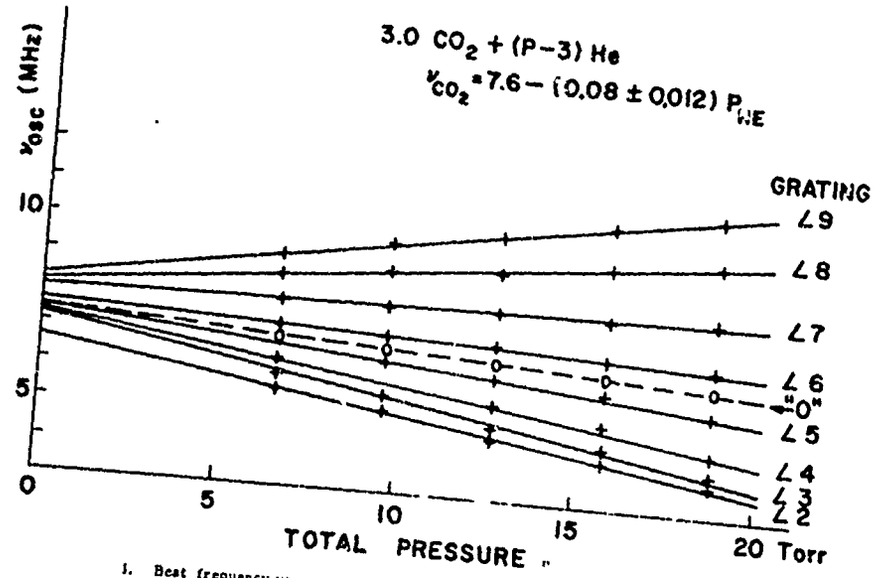
References

1. R. L. Barger and J. L. Hall, "Pressure Shift and Broadening of Methane Line at 3.39 μ Studied by Laser-Saturated Molecular Absorption," Phys. Rev. Letters 22, 4 (1969).

2. P. Rabinowitz, R. Keller and J. T. LaTourrette, "Lamb Dip Spectroscopy Applied to SF_6," Appl. Phys. Letters 14, No. 12, 376 (June 15, 1969).

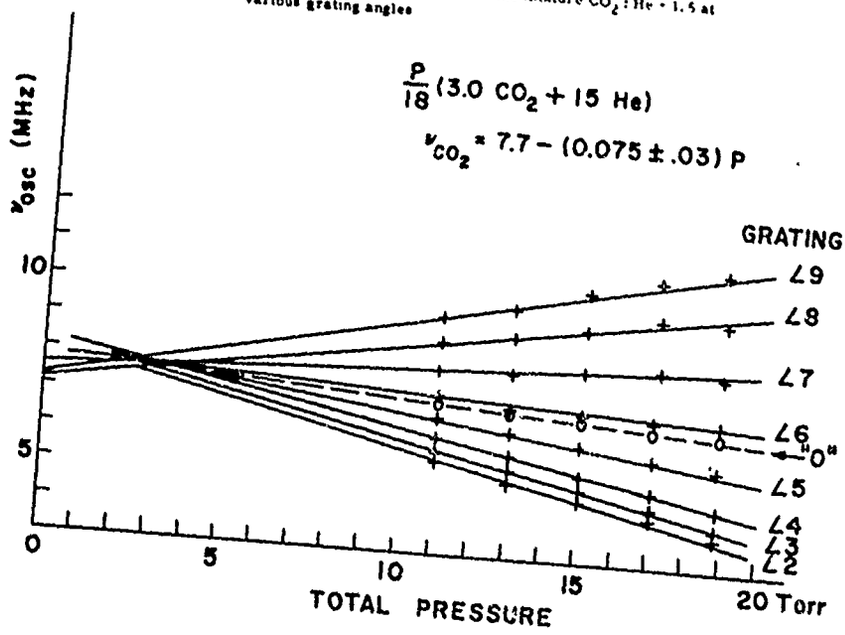
3. P. Rabinowitz, "AFC Optical Heterodyne Detector," Proc. IEEE, 51, 857(1963).

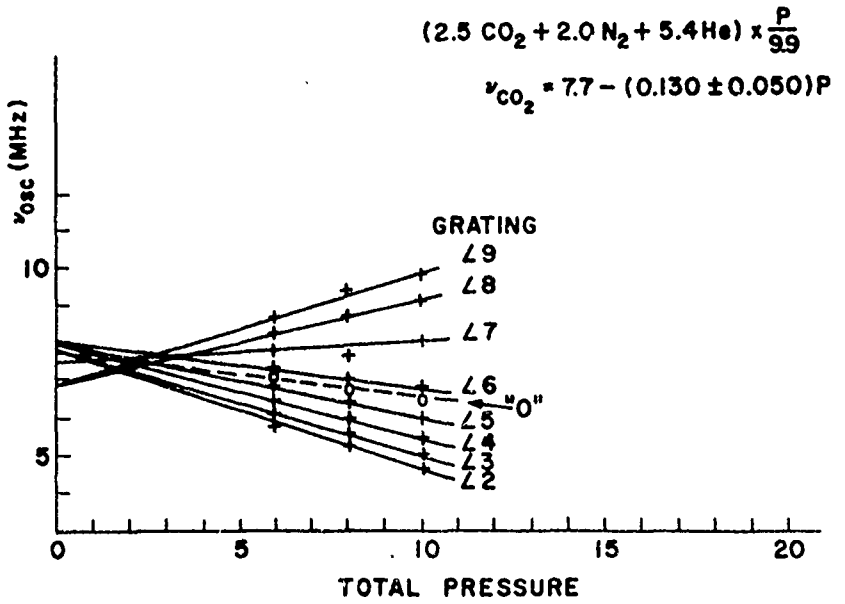
4. C. Freed, private communication; described by H. Boyne "A Review of Laser Frequency Stabilization," in these Proceedings.



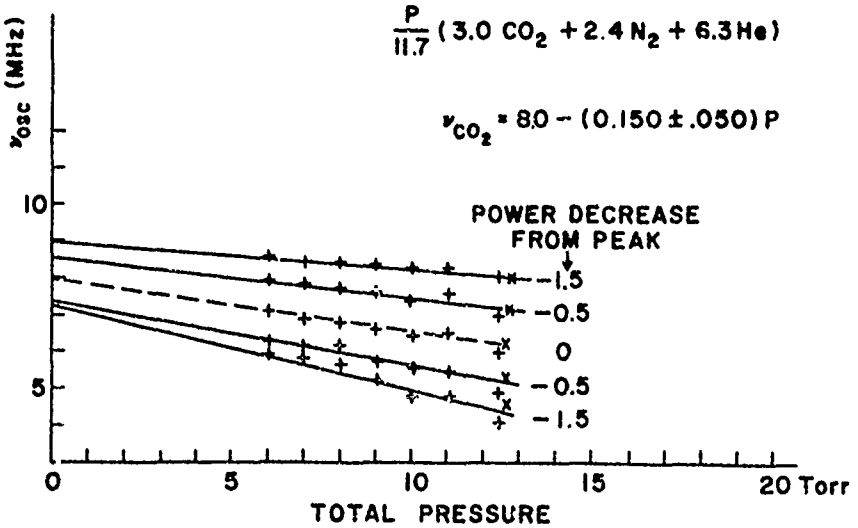
1. Beat frequency vs pressure for fixed 3 Torr CO_2 and variable He pressure, at various grating angles

2. Beat frequency vs. pressure for the mixture CO_2 : He = 1:5 at various grating angles





3. Beat frequency vs. pressure for the mixture CO_2: N_2: He = 2.5:2.0:5.4 at various grating angles



4. Beat frequency vs. pressure for the mixture CO_2: N_2: He = 3.0:2.4:6.3 at various decreases in laser power due to grating tilt

NEW INFORMATION ON THE PHYSICS OF RUBIDIUM GAS CELLS /

P. L. Bender, Joint Institute for Lab. Astrophysics *
V. W. Cohen, Brookhaven National Laboratory

It has recently been found that the Rb^{87} hyperfine transition line width and frequency shift due to collisions can have a contribution from the transient formation of weakly bound rubidium-buffer gas molecules. Evidence for the formation of Rb-Kr molecules was obtained [1-3] at the Ecole Normale Superieure by observing Zeeman transitions at different magnetic fields. The molecules appear to be formed by three-body collisions, and then destroyed easily by two-body collisions.

For the Rb^{87} hyperfine transition, transient molecule formation causes a shift in the hyperfine frequency during the molecular lifetime. It can be shown that this gives an extra contribution to the width which is roughly independent of pressure over a wide range. Hyperfine line widths for helium and neon give no evidence of such a contribution, but line widths for argon, nitrogen, and methane do [4-5]. The clearest case is for argon, where the extra width is roughly 25 Hz. A non-linear frequency shift with buffer gas density is expected at low density and appears to be present. The minimum linewidths obtained for He, Ne, Ar, N_2, and CH_4, extrapolated to zero light intensity and microwave power, were 11, 7, 33, 17, and 28 Hz.

Some comments will also be made on the two commonly suggested possible causes of drift in Rb^{87} frequency standards: (1) change in the light intensity or spectral distribution, and (2) change in the collisional frequency shift.

1. M. Aymar, M.A. Bouchiat, and J. Brossel, Phys. Letters 24A, 753 (1967)
2. M. A. Bouchiat, J. Brossel, and L.Pottier, Phys. Rev. Letters 19, 817 (1969)
3. C.C.Bouchiat, M.A. Bouchiat, and L.C.L. Pottier, Phys. Rev. 181, 144 (1969)
4. P. L. Bender and V. W. Cohen, VI ICPEAC Book of Abstracts, M.I.T. (1969)pp.720-722
5. P. L. Bender and V. W. Cohen, to be published.

/ This work was supported by the National Bureau of Standards and the Atomic Energy Commission.

* of the National Bureau of Standards and University of Colorado

PROGRESS REPORT ON THE RUBIDIUM 85 MASER

J. Vanier, R. Vaillancourt, G. Missout, M. Têtu
CRAM, Département de Génie Electrique
Université Laval
Québec 10, Qué.

Summary

In this paper we give results obtained recently in connection with the operation of a maser based on the hyperfine frequency of the ground state of rubidium 85 (3035 GHz). Studies have been conducted in the following areas:

Efficiency of a rubidium 87 filter used in connection with a rubidium 85 lamp.

Frequency shifts of the hyperfine transition due to various buffer gases.

Relaxation times T_1 and T_2 in the same buffer gases and due to spin exchange interaction.

Oscillation condition in the low field mode.

Introduction

Due to their possible application in systems requiring high spectral purity, rubidium masers have received particular attention in the last few years. That property is expected from rubidium masers owing to their relatively high power output. Successfull operation of the rubidium 87 maser has been reported[1,2]. Operation of rubidium 85 maser with external gain has also been reported[3]. In that last case a stability of the order of 1 in 10^{12} has been obtained for an averaging time of 1 sec. However no report has been given yet of the successfull operation of a rubidium 85 maser without external feedback. In the present paper we give results of measurements of various parameters whose values are critical to the operation of the maser.

Rubidium Maser Basic Principles

The energy levels of the ground state and of the excited P state of rubidium 85 is shown on figure 1. The hyperfine frequency is 3,035 GHz and the optical resonance transitions D_1 and D_2 appear at a wavelength of 7800 A° and 7947 A°. Each of these transitions has an hyperfine structure and through proper filtering one element of each line can be removed. In the proper arrangement line B of figure 1 can be removed and the lamp filtered in such a manner can then be used to produce a population inversion in an absorption cell. This basic idea can be used with either of the two isotopes 85 or 87. However, depending on the isotope used care must be taken in the design of the filter as will be seen below.

The experimental arrangements

The design of the rubidium 85 maser studied in this article is shown on figure 2 along with the associated electronics used to make the measurements. The maser is essentially a rubidium 87 maser except that all isotopes have been interchanged. The configuration is then: Rb^{85} lamp - Rb^{87} filter - Rb^{85} storage bulb in a TE_{011} cavity.

The lamp

The lamp is a 1" bulb filled with argon at a pressure of 2 Torr and traces of rubidium 85. The bulb is placed at the centre of a 3" spherical mirror and is excited with r.f. power at a frequency of about 70 MHz and a power of the order of 20 watts. It can be operated either in a continuous mode or in a pulse mode.

The filter

Filtering of the lamp in order to produce large population inversions is not as easy as in the case of the rubidium 87 maser. However it is still possible to produce an efficient filtering by using the proper buffer gas in the filter cell.

The line profile of a natural rubidium lamp as obtained on a high resolution Fabry Perot interferometer is shown in figure 3; it is seen that line A and B of the two isotopes are not exactly coinciding, a situation which does not lead to maximum filtering efficiency. However it is known that helium shifts rubidium resonance spectral lines to the violet, this shift being accompanied by a relatively more important broadening. Results of measurements made on the amplitude of the two hyperfine components A and B of the D_1 and D_2 transitions are presented on figure 4. It is seen that a much better filtering is obtained for the D_1 line. From the graphs, pressures of the order of 60 to 80 Torr of helium appear to be satisfactory. These results are in qualitative agreement with those of Hartmann[5] and Happer[6].

The storage bulb

In the experiment reported here in connection with the maser storage bulb, three types of measurements were performed. Those are relaxation rates, frequency shifts, spin exchange effects and power output.

Relaxation rates. Relaxation rates were measured in the dark through a technique described extensively in the literature[1]. A light pulse is applied to the maser and produces a large population imbalance. A microwave pulse is applied after the light pulse and produces a stimulated emission signal. This signal decays at a rate γ_2 or $1/T_2$. The initial amplitude of the signal is a measure of the population imbalance and from its behaviour for various times between the end of the light pulse and the end of the microwave pulse, the parameter γ_1 or $1/T_1$ can be obtained. Typical results for various pressures of argon and nitrogen are shown on figure 5. For a three inch bulb the smallest relaxation rate appears at about 10 Torr as in the case of the rubidium 87 maser and that value should be the pressure for maximum gain. It is interesting to note that for argon, the slope of γ_2 versus pressure is smaller than the slope for nitrogen.

Frequency shifts. The frequency shifts due to collisions with buffer gas atoms have been measured as a function of pressure in the dark after the light pulse. Relaxation times were long enough to permit frequency determination with a precision of a few hertz. The results are given on figure 6. Extrapolation to zero pressure gives the unperturbed hyperfine frequency.

$$\nu_{hf} = 3,035\ ,\ 732,440 \pm 10 \text{ Hz}.$$

This is in aggreement with published data obtained through beam techniques[7]. Frequency shifts can be interpreted theoretically by an average phase shift per collision in the atomic wave function. It is thus interesting to note the small phase shift caused by collisions with argon, a fact that is also reflected by the small slope of γ_2 versus pressure for the same gas as reported earlier.

Table I

Frequency shift in Hz/Torr for Rb^{85} in various buffer gases at 27°C

He	Ne	Ar	N_2
328	179	24	241

Spin exchange interaction. Measurement of γ_1 and γ_2 were also made as a function of the rubidium density, through a variation of temperature of the absorption cell. The results are shown in figure 7. On this graph γ_2 is plotted as a function of $(\gamma_1 - \gamma_1^o)$ where γ_1^o is the value of γ_1 at zero rubidium pressure. Consequently $(\gamma_1 - \gamma_1^o)$ is a measure of the spin exchange interaction γ_1^{se} and is effectively equal to $n \bar{v}_r \sigma$ where n is the density $\bar{v}_r$ is the relative velocity and σ the spin exchange cross section. It is somewhat difficult to perform exact measurements of γ_1 and γ_2 as a function of density due to the inherent high gain of the maser; in such a situation radiation damping adds to the relaxation making the measured rates γ_2 appear larger; in order to minimize this effect we have made the measurements for a situation where the cavity is much detuned, simulating a lower gain. The slope of the line of γ_2 versus γ_1^{se} is 2/3, in agreement with theory[8,9].

Power output. Continuous oscillations have been obtained in an experimental arrangement in which the magnetic field was reduced to a value permitting the overlap of several transitions. A power output of the order of 10^{-11} watts was detected; in that experiment the bulb diameter was only 3". It is expected that with a bulb filling entirely the cavity and pumping from both ends of the cavity, a larger power output will be obtained and that oscillation between the field independent levels will be possible.

References

1- P. Davidovits, Appl. Phys. Letters. 5-15, 1969.

2- J. Vanier, Phys. Rev. Vol. 168-129, 1968.

3- F. Hartmann, Phys. Letter, Vol. 28A, 193, 1968.

4- S. Chen, M. Takeo, Rev. Mod. Phys. 29-20, 1957.

5- F. Hartmann, Thesis, Université de Paris, (not published).

6- W.A. Stern and R. Novick, 23rd Annual Symposium on Frequency Control, May 1969, Atlantic City, N.J.

7- S. Penselin, T. Moron, V.W. Cohen and G. Winkler, Phys. Rev. 127. 2, 1962.

8- F. Grossetête, Thesis, Université de Paris, (not published).

9- W. Happer, private communication.

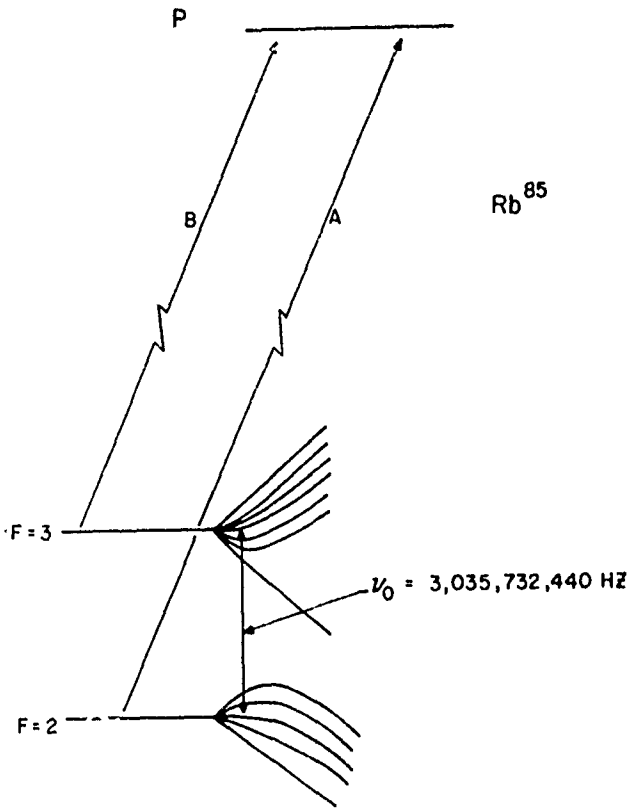
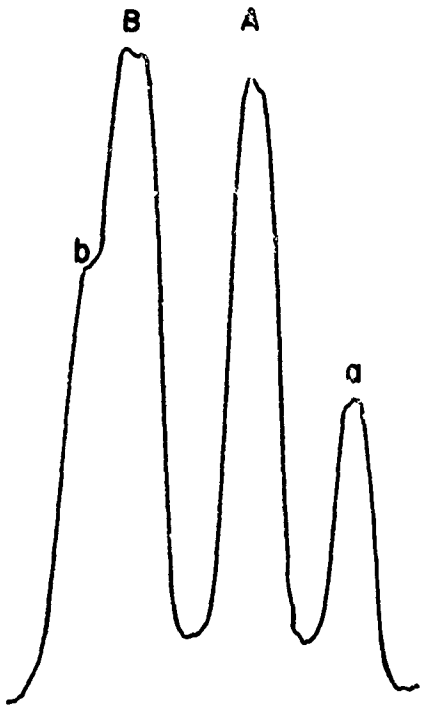


Figure 1 - Energy levels of the ground state and excited P state of rubidium 85.



Rb Naturel
λ = 7800 A°

Figure 3 - Line profile of a natural rubidium lamp at the wavelength 7800 A°.

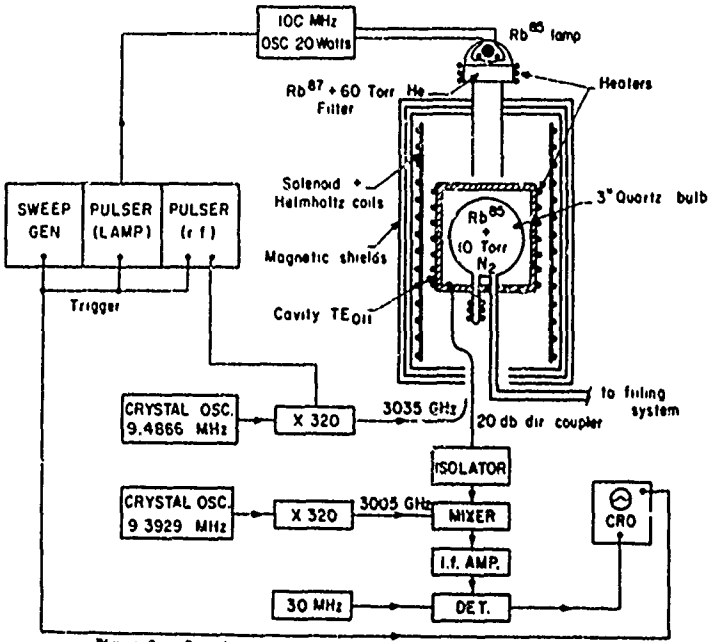


Figure 2 - Experimental arrangement used to study relaxation and frequency shifts in rubidium 85. The same arrangement was used to obtain oscillation.

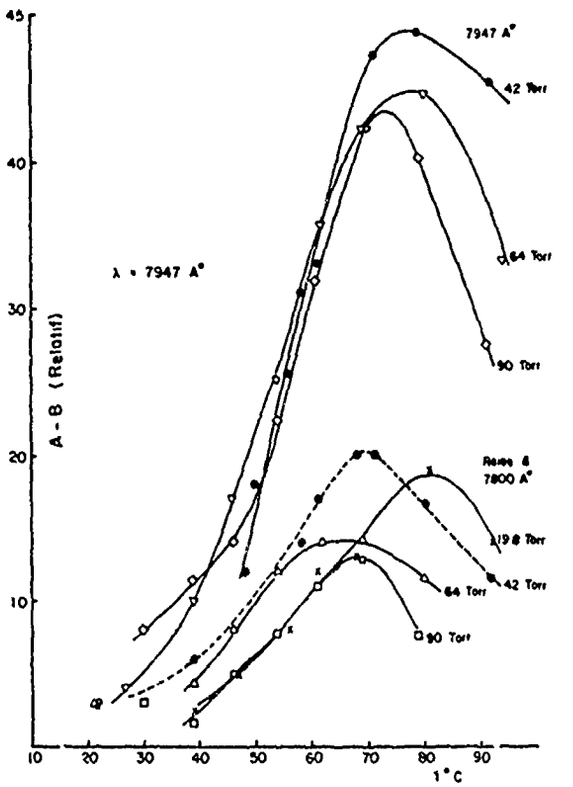


Figure 4 - Variation of the relative amplitudes (A - B) as identified in figure 3, as a function of filter temperature and helium pressure.

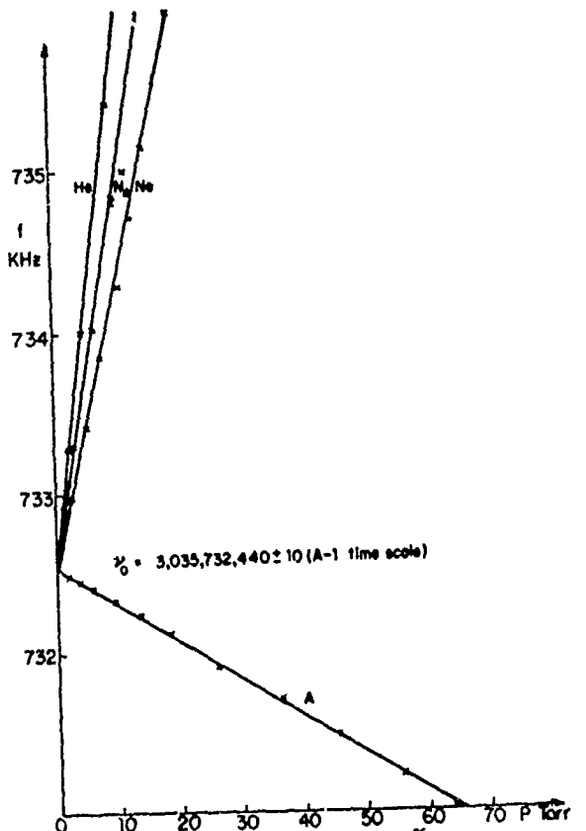


Figure 5 - Variation of the hyperfine frequency of Rb^{85} in various buffer gases. The time scale is 0 T-1.

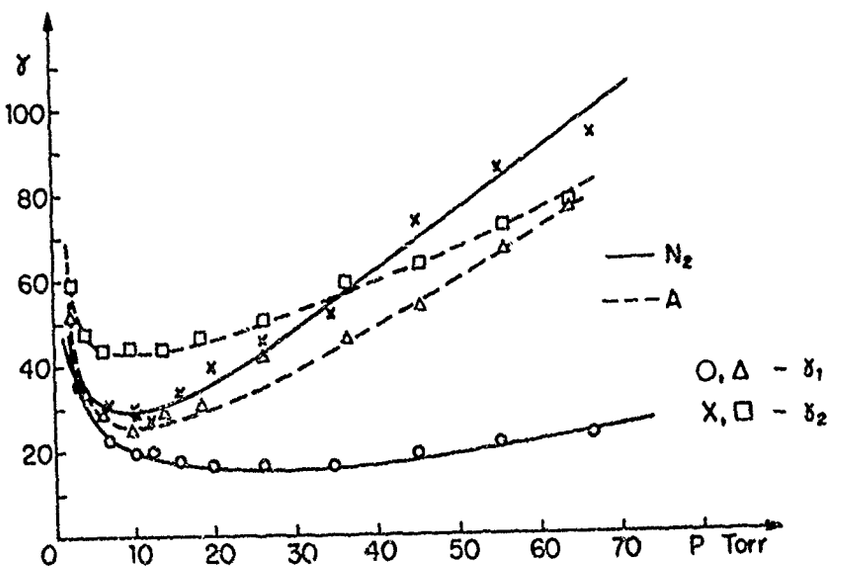


Figure 5 - Variation of γ_1 and γ_2 in rubidium 85 as a function of pressure in nitrogen and argon.

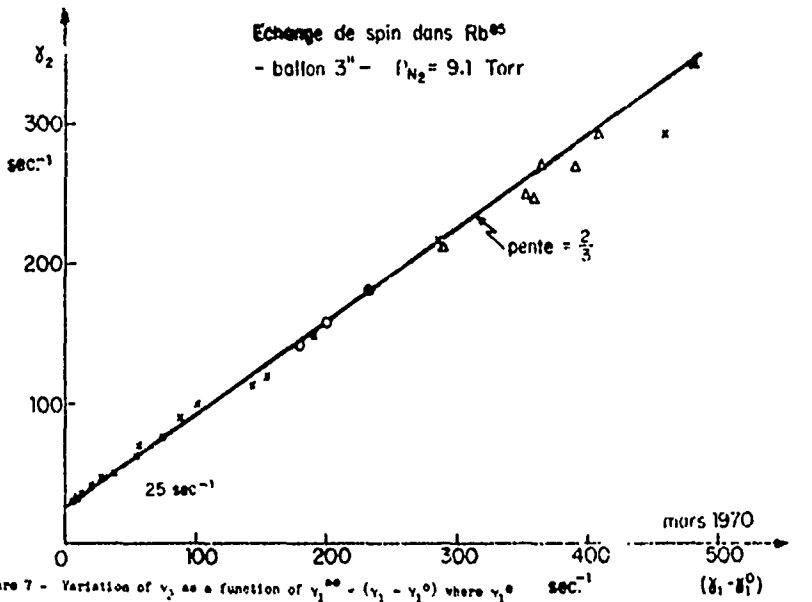


Figure 7 - Variation of γ_2 as a function of $\gamma_1^{se} = (\gamma_1 - \gamma_1^0)$ where γ_1^e is the zero rubidium density population relaxation rate. The density is proportional to $\gamma_1^{se} = n \bar{v}_r \sigma$ where $\bar{v}_r$ is the relative speed, σ is the cross section.

AN OPTICALLY PUMPED PARAMETRIC FREQUENCY CONVERTER*

H. Tang and W. Happer
Columbia Radiation Laboratory, Department of Physics,
Columbia University, New York, New York 10027

Summary

The observation of a new type of nonlinear optical effect in optically pumped Rb^{87} vapor affords the possibility of constructing several new devices, such as microwave spin oscillators, microwave parametric oscillators, and microwave photon counters. Recent experiments on a Rb^{87} traveling wave frequency converter have yielded conversion efficiencies on the order of one percent. Similar frequency converters can be constructed using Cs^{133} vapor. The feasibility of microwave spin oscillators operating on the Rb^{87} or Cs^{133} hyperfine transition frequencies rests on the availability of efficient microwave photodetectors.

Introduction

For weak light intensities, a nonlinear component of the dielectric susceptibility of an optically pumped vapor arises when the oriented atoms are driven by a resonant magnetic field, causing them to undergo magnetic resonance transitions. These magnetic dipole type nonlinear susceptibilities can be calculated using a semiclassical model of the interaction between the atomic system of the vapor and the electromagnetic field of the light.[1] We have observed parametric frequency conversion in a vapor of Rb^{87} atoms which were optically pumped to establish a population imbalance between the ground-state hfs levels and which were excited by a resonant microwave magnetic field at the F = 1; $m_f = 0 \rightarrow F = 2$; $m_f = 0$ transition frequency. Optical pumping of the hyperfine levels of Rb^{87} is achieved by a well-known method.[2] Figure 1 shows the ground-state hyperfine structure of the two isotopes of rubidium. The D_1 transition lines of both isotopes are shown as an example. Rb^{85} has a $5^2S_{1/2}$ ground state with two hyperfine levels formed by the coupling of the electron spin, J = 1/2 and the nuclear spin I = 5/2. These levels correspond to total angular momenta F = 3 and F = 2. Rb^{87} has a similar pair of hyperfine levels in its ground state, which are formed by the coupling of its electron spin J = 1/2 and its nuclear spin I = 3/2. The corresponding total angular momenta are F = 2 and F = 1. The hyperfine levels of the $5^2P_{1/2}$ first excited states of both isotopes are not well-resolved, so that the D_1 lines of each isotope consists practically of two hyperfine components. The separation frequency of the components (A) and (B) is 3035 MHz, while the components (a) and (b) are separated by 6835 MHz. The components (A) and (a) nearly coincide and actually overlap. Thus, when Rb^{87} vapor is illuminated by Rb^{85} light, only the (a) transition of the Rb^{87} vapor is excited. Because the Rb^{87} atoms decay from the excited state to both ground-state hyperfine levels, the Rb^{85} resonance light will tend to deplete the F = 2 level while populating the F = 1 level of the Rb^{87} ground state.

Vapor Susceptibility

The following is a brief discussion of how the modulated components of the dielectric susceptibility of a Rb^{87} vapor is calculated. According to the optical pumping theory of Happer and Mathur,[3] the polarizability of an optically pumped alkali vapor may be expressed as the expectation value of a suitable polarizability operator α

$$\langle \alpha \rangle = \text{Trace} [\alpha \rho] , \qquad (1)$$

where ρ is the ground-state density matrix of the atomic system. The polarizability is related to the susceptibility $\langle \chi \rangle$ of the vapor by

$$\langle \chi \rangle = N \langle \alpha \rangle , \qquad (2)$$

where N is the number density of the atoms in the vapor. For weak light intensities, the polarizability of an alkali vapor can be shown to have the form[4]

$$\begin{aligned} \langle \alpha \rangle = {} & \alpha_{eq} + \alpha_{hfs} \langle \underset{\sim}{I} \cdot \underset{\sim}{J} \rangle \\ & + i \sum_{F_g F_g'} \alpha_{gt}(F_g F_g') \langle \underset{\sim}{J}(F_g, F_g') \rangle \times \\ & + \sum_{F_g F_g'} \alpha_{br}(F_g F_g') \langle \underset{\approx}{Q}(F_g F_g') \rangle \end{aligned} \qquad (3)$$

The first term in (3) is the equilibrium polarizability α_{eq}, which is independent of the state of the vapor. It would be the only nonvanishing term in (3) if the atoms of the vapor were randomly distributed among the sublevels of the ground state. The second term in (3) is the hyperfine structure polarizability, which is proportional to the quantity $\langle \underset{\sim}{I} \cdot \underset{\sim}{J} \rangle$. This quantity is a measure of the population imbalance of the ground-state

hyperfine levels. The two terms α_{eq} and $\alpha_{hfs}\langle \underset{\sim}{I}\cdot\underset{\sim}{J}\rangle$ comprise the isotropic polarizability α_o of the vapor. with

$$\alpha_o = \alpha_{eq} + \alpha_{hfs}\langle \underset{\sim}{I}\cdot\underset{\sim}{J}\rangle \ . \qquad (4)$$

The third term in (3) consists of the gyrotropic polarizabilities, which are proportional to the expectation values of the electronic angular momentum operator $\langle \underset{\sim}{J}(F_g,F_g')\rangle$ within each hyperfine level, $(F_g = F_g')$, and between different hyperfine levels, $(F_g \neq F_g')$. When transitions are induced either between Zeeman levels within a hyperfine level or between different hyperfine levels, the ground-state density matrix will contain rapidly oscillating off-diagonal elements. Therefore, the angular momentum observables $\langle \underset{\sim}{J}(F_g,F_g')\rangle$ will also have rapidly oscillating components. As a result, the gyrotropic susceptibilities will contain oscillatory components which cause light modulation or the mixing of light waves with radio-frequency fields or with microwaves.

The fourth term in (3) contains the birefringent polarizabilities, which are proportional to the expectation values of the quadrupole operator $\langle \underset{\sim}{Q}(F_g,F_g')\rangle$. These terms may also be rapidly oscillating when magnetic resonance transitions in the vapor are excited; however, it may be shown that when only the 0 - 0 hyperfine transition is driven, the birefringent susceptibility does not have an oscillating component.

The expectation value of the electronic angular momentum operator J can be written as

$$\langle \underset{\sim}{J}\rangle = \text{Trace}\ [\underset{\sim}{J}\,\rho] \ , \qquad (5)$$

when the 0-0 transition is driven. The oscillating parts of (5) are

$$\begin{aligned}\langle J(12)\rangle &= \langle 10|\underset{\sim}{J}|20\rangle\rho_{20,10};\\ \langle J(21)\rangle &= \langle 20|\underset{\sim}{J}|10\rangle\rho_{10,20}\ ,\end{aligned} \qquad (6)$$

where $\langle 10|\underset{\sim}{J}|20\rangle$ is the matrix element of angular momentum between the F = 1, $m_f = 0$ and the F = 2, $m_f = 0$ states. The term $\rho_{20,10}$ whose magnitude is often referred to as the coherence, is the off-diagonal density matrix element between the F = 2, $m_f = 0$ and the F = 1, $m_f = 0$ states. It can be easily shown that

$$\rho_{20,10} = \rho^*_{10,20} \ . \qquad (7)$$

In the case where a small, static magnetic field establishes the z-axis of the coordinates, and a microwave magnetic field $\underset{\sim}{H}_1$ which excites the 0-0 transition also lies along the z-direction, we can write

$$\begin{aligned}\langle 20|J_z|10\rangle &= 1/2\ \hat{k}\\ \langle 10|J_z|20\rangle &= 1/2\ \hat{k}\end{aligned} \qquad (8)$$

where $\hat{k}$ is a unit vector along the z-direction. Since the Rb^{87} vapor is placed inside a microwave cavity, the microwave magnetic field will have the form of a standing wave. A standing wave can be considered as composed of two traveling waves propagating in opposite directions. We assume that the light wave can only couple with one of the traveling waves, and that the other one may be neglected. We therefore write the magnetic field in the form

$$\underset{\sim}{H}_1 = \underset{\sim}{\mathcal{K}}_1 e^{i(Kz - \omega_m t)} + \text{c.c.} \qquad (9)$$

where c.c. denotes the complex conjugate, and K is the guide wavevector of the microwaves. Then the off-diagonal density matrix element can be shown to be

$$\rho_{10,20} = \frac{(\rho_{10,10} - \rho_{20,20})\ \frac{\mu_o \mathcal{K}_1}{\hbar}\ e^{i(Kz - \omega_m t)}}{(\omega_m - \omega_o + \Delta\omega_\ell) - i(\Gamma/2 + \gamma_2)}, \qquad (10)$$

where $\rho_{10,10}$ is the diagonal density matrix element of the F = 1, $m_f = 0$ state, and $\rho_{20,20}$ is the diagonal density matrix element of the F = 2, $m_f = 0$ state. We define:

μ_o = Bohr magneton;

ω_o = pressure-shifted transition frequency;

$\Delta\omega_\ell$ = light shift;

Γ = optical pumping rate;

γ_2 = transverse relaxation time.

A plot of the square of the magnitude of $\rho_{10,20}$ versus the square of the magnitude of the microwave field for the on-resonance case $(\omega_m = \omega_o + \Delta\omega_\ell)$ is shown in Fig. 2. Typical experimental values of Γ, γ_1, and γ_2 were used. The quantity $|\rho_{10,20}|^2$ is seen to increase linearly with $|H_1|^2$ for small values of $|H_1|^2$. $|\rho_{10,20}|^2$ eventually reaches a maximum and goes to zero for large values of $|H_1|^2$, where saturation occurs. From (2), (3), (6), (7), (8), and (10), we see that the oscillating part of the gyrotropic susceptibility is given by

$$\langle \chi\rangle = \frac{1}{4}(\rho_{10,10} - \rho_{20,20}) \times N\mu_o \left\{ \frac{\alpha_{gt}(12;\omega)\mathcal{K}_1 e^{i(Kz - \omega_m t)}}{(\omega_m - \omega_o + \Delta\omega_\ell) + i(\Gamma/2 + \gamma_2)} + \frac{\alpha_{gt}(21;\omega)\mathcal{K}_1^* e^{-(Kz - \omega_m t)}}{\omega_m - \omega_o + \Delta\omega_\ell) - i(\Gamma/2 + \gamma_2)} \right\} \hat{k}\hat{X}. \qquad (11)$$

We can define from (11) an up-coupling susceptibility χ_+ of the form

$$\chi_+ = \frac{i}{4}(\rho_{10,10} - \rho_{20,20}) \times N\mu_o \left\{ \frac{\alpha_{gt}(12;\omega)}{(\omega_m - \omega_o + \Delta\omega_\ell) + i(\Gamma/2 + \gamma_2)} \right\} . \quad (12)$$

This relates an electric polarization $\underset{\sim}{P}(\omega_+)$ to a magnetic field

$$\underset{\sim}{\mathcal{K}}_1 e^{i(Kz - \omega_m t)}$$

and an electric field

$$\underset{\sim}{\mathcal{E}} e^{i(Kz - \omega t)} ,$$

according to

$$\underset{\sim}{P}(\omega_+) = \chi_+(\omega_+ = \omega + \omega_m)\underset{\sim}{\mathcal{K}}_1 \times \underset{\sim}{\mathcal{E}} e^{i[(K+k)z - (\omega + \omega_m)t]} \quad (13)$$

Similarly we can define from (11) a down-coupling susceptibility χ_- of the form

$$\chi_- = \frac{i}{4}(\rho_{10,10} - \rho_{20,20}) \times N\mu_o \left\{ \frac{\alpha_{gt}(21;\omega)}{(\omega_m - \omega_o + \Delta\omega_\ell) - i(\Gamma/2 + \gamma_2)} \right\} , \quad (14)$$

which relates a polarization $\underset{\sim}{P}(\omega_-)$ to a magnetic field

$$\underset{\sim}{\mathcal{K}}_1^* e^{-i(Kz - \omega_m t)}$$

and an electric field

$$\underset{\sim}{\mathcal{E}} e^{i(Kz - \omega t)} ,$$

according to

$$\underset{\sim}{P}(\omega_-) = \chi_-(\omega_- = \omega - \omega_m)\underset{\sim}{\mathcal{K}}_1^* \times \underset{\sim}{\mathcal{E}} e^{i[(K-k)z - (\omega - \omega_m)t]} \quad (15)$$

The susceptibilities χ_+ and χ_- are functions of optical frequency, as $\alpha_{gt}(12;\omega)$ and $\alpha_{gt}(21;\omega)$ are functions of optical frequency. Figure 3 shows the frequency response functions $\alpha_{gt}(12;\omega)$ and $\alpha_{gt}(21;\omega)$.

Wave Equation

Having defined the nonlinear polarizations, we may incorporate them as source terms in Maxwell's Equations to obtain the sideband solutions. We begin with the wave equation,

$$\nabla \times \nabla \times \underset{\sim}{E} - \frac{1}{c^2}\frac{\partial^2 \underset{\sim}{E}}{\partial t^2} = -4\pi \frac{\partial^2 \underset{\sim}{P}}{\partial t^2} . \quad (16)$$

If the carrier light wave is represented by a superposition of plane harmonic waves propagating in the positive z-direction, the electric field of a particular Fourier component will have the form

$$\underset{\sim}{E}(\omega) = \underset{\sim}{\mathcal{E}}(\omega,z,t)e^{i(kz - \omega t)} + \text{c.c.}, \quad (17)$$

where $k = \omega/c$. If we assume that the amplitudes of all the light waves propagating in the vapor are slowly varying with respect to time and distance, then (13) can be simplified to the form

$$\frac{\partial \underset{\sim}{\mathcal{E}}(\omega)}{\partial z} + \frac{1}{c}\frac{\partial \underset{\sim}{\mathcal{E}}(\omega)}{\partial t} = \frac{2\pi i}{\omega c}\frac{\partial \underset{\sim}{P}(\omega)}{\partial t} e^{-i(kz - \omega t)} . \quad (18)$$

The polarization $\underset{\sim}{P}$ can be written as the sum of two parts:

$$\underset{\sim}{P} = \underset{\sim}{P}^L + \underset{\sim}{P}^{NL} . \quad (19)$$

The linear part of the polarization has the form

$$\underset{\sim}{P}^L(\omega) = \chi_o(\omega)\underset{\sim}{E}(\omega) , \quad (20)$$

where $\chi_o(\omega)$ is the isotropic susceptibility at frequency ω given by

$$\chi_o(\omega) = N\langle \alpha_o(\omega) \rangle . \quad (21)$$

The term $\langle \alpha_o \rangle$ is defined in (4). We can see from Fig. 3 that the up-coupling and the down-coupling susceptibilities are sizable only in the vicinity of the atomic transitions. Therefore, for a carrier wave of frequency ω which is near a Rb^{87} transition frequency, we can assume that coupling takes place in the vapor only to an upper sideband of frequency ω_+ and to a lower sideband of frequency ω_-. All other sidebands may therefore be neglected, and we may write the nonlinear parts of the polarizations as

$$P^{NL}(\omega_+) = \chi_+(\omega_+ = \omega_m + \omega)\underset{\sim}{\mathcal{K}}_1 \times \underset{\sim}{\mathcal{E}}_1 e^{i[(k+K)z - (\omega + \omega_m)t]} ;$$

$$P^{NL}(\omega) = [\chi_-(\omega = -\omega_m + \omega_+)\underset{\sim}{\mathcal{K}}_1^* \times \underset{\sim}{\mathcal{E}}_1 - \chi_+(\omega = \omega_m + \omega_-)\underset{\sim}{\mathcal{K}}_1 \times \underset{\sim}{\mathcal{E}}_1] e^{i(kz - \omega t)} ;$$

$$P^{NL}(\omega_-) = \chi_-(\omega_- = -\omega_m + \omega)\underset{\sim}{\mathcal{K}}_1^* \times \underset{\sim}{\mathcal{E}}_1 e^{i[(k-K)z - (\omega - \omega_m)t]} \quad (22)$$

Substituting (20) and (22) into (18), we obtain a set of three coupled equations for the steady-state amplitudes of the carrier wave and the two sideband waves:

$$\frac{d}{dz}\underset{\sim}{\mathcal{E}}(\omega_+) = 2\pi ik\left[\chi_o(\omega_+)\underset{\sim}{\mathcal{E}}(\omega_+) + \chi_+ \underset{\sim}{\mathcal{K}}_1 \times \underset{\sim}{\mathcal{E}}_1(\omega) e^{i(K-k_+ +k)z}\right] ;$$

$$\frac{d}{dz}\underset{\sim}{\mathcal{E}}(\omega) = 2\pi ik\left[\chi_o(\omega)\underset{\sim}{\mathcal{E}}(\omega) + \chi_- \underset{\sim}{\mathcal{K}}_1^* \times \underset{\sim}{\mathcal{E}}_1(\omega_+) e^{-i(K-k_+ -k)z} + \chi_+ \underset{\sim}{\mathcal{K}}_1 \times \underset{\sim}{\mathcal{E}}_1(\omega_-) e^{i(K+k_- -k)z}\right] ;$$

$$\frac{d}{dz}\mathcal{E}(\omega_-) = 2\pi ik\left[\chi_o(\omega_-)\mathcal{E}(\omega_-) + \chi_- \underset{\sim}{\mathcal{K}}_1^* \times \underset{\sim}{\mathcal{E}}_1(\omega) e^{-i(K+k_- -k)z}\right] , \quad (23)$$

where $k_+ = \omega_+/c$ and $k_- = \omega_-/c$. From the cross products in (23), we can see that for a linearly polarized carrier wave, the sideband waves are generated with a linear polarization which is orthogonal to that of the carrier wave. By manipulating (23), we can show that the optimum coupling takes place between the carrier and the upper sideband when the condition

$$n(\omega_+)k_+ = n(\omega)k + K \quad (24)$$

is satisfied. The terms $n(\omega_+)$ and $n(\omega)$ are the indices of refraction of the vapor at frequency ω_+ and at ω, respectively. Similarly, it may be shown that optimum coupling between the carrier and the lower sideband takes place upon satisfaction of the condition:

$$n(\omega_-)k_- = n(\omega)k - K . \quad (25)$$

The term $n(\omega_-)$ is the index of refraction of the vapor at frequency ω_-. The index of refraction $n(\omega)$ is related to the real part of the isotropic susceptibility by

$$n(\omega) = 1 + 2\pi\,\chi_o'(\omega) . \quad (26)$$

Since χ_o is a function of the number density of atoms in the vapor and the population imbalance between the hyperfine levels, the phase-matching conditions (24 and 25) depend on the temperature of the resonance cell and the extent to which the vapor is optically pumped. The guide wave vector K of the microwaves can be adjusted to meet the conditions of either (23) or (24). It is usually not possible to phase match to both the upper and lower sidebands.

Figure 4 shows the theoretically calculated results of passing a carrier beam of Rb^{85} D_1 light through a 40-cm length of optically pumped Rb^{87} vapor with hyperfine coherence. This figure shows the spectral profile of the emerging light from the cavity as the vapor temperature was varied from 50°C to 70°C, while the guide wavelength of the microwaves was adjusted for perfect phase matching between the carrier and the upper sideband at each temperature. Typical parameters were again assumed for the vapor. Since the phase-matching condition can only be maintained for a frequency. This frequency was chosen as -40 mK. The initial spectral profile of the Rb^{85} D_1 light is represented by two rectangular hyperfine components, each of which is of unit height and 50 mK in width. After passing through the Rb^{87} vapor, the Rb^{85} lines are seen to be partially absorbed as a result of overlapping with the Rb^{87} absorption lines. However, well-resolved sidebands have also appeared in the spectral profile of the emerging light. One sees from Fig. 4 that at 70°C, the upper sideband intensity can be as large as 10% of the initial carrier intensity.

Experiment

The experimental apparatus used for our investigations is shown in Fig. 5. A cylindrical microwave cavity was constructed from a stack of brass plates, spaced 1 cm apart. A hole of 6.5-cm diameter was cut in the center of each plate. A photograph of the cavity is shown in Fig. 6. The cavity was terminated at one end by a plate with a smaller diameter hole (~3 cm) and at the other end by a section of ordinary cylindrical waveguide and a perforated shorting plate. The cavity operated in the TE_{01n} mode. A cylindrical quartz cell, filled with a small amount of Rb^{87} metal and nitrogen buffer gas at a pressure of about 12 Torr, was placed in the cavity. The cell temperature was maintained at 65°C. The dimensions of the quartz tube were chosen so that the dielectrically loaded cavity had the proper guide wavelength to satisfy the phase-matching condition (24). The Rb^{87} cell was pumped by resonance radiation from a long Rb^{85} lamp, which was parallel to the cavity axis. An elliptical reflector focused the Rb^{85} light from the long lamp onto the Rb^{87} absorption cell. Microwaves from a stabilized klystron oscillator were used to induce the 6.835 GHz field-independent transition of the Rb^{87} atoms and to produce the microwave coherence in the vapor. The microwaves were square-wave modulated on and off to permit observation by phase-sensitive detection of the accompanying changes in the probing beam intensity.

Optical carrier waves were provided by a second Rb^{85} lamp, situated at one end of the cavity. The carrier beam was linearly polarized by the polarizer P_1. After emerging from the microwave cavity, the probing beam passed through a linear

polarizer P_2, which was orthogonal to the input polarizer P_1. The polarizer P_2 served to remove the carrier waves. Finally, the D_1 component of the probing beam was frequency-analyzed by a scanning Fabry-Perot interferometer. The free spectral range of the interferometer was 1000 mK and the finesse was about 35. The intensity of the central fringe of the interferometer was recorded with a cooled EMI 9558QA photomultiplier tube, and the photomultiplier signal was detected in synchronism with the 13-Hz microwave modulation frequency.

Some typical observations are reproduced in Fig. 7. Figure 7a gives the spectral profile obtained with the apparatus of Fig. 5. The signals of Fig. 7a were obtained with a 30-sec integration time and a 20-min scan of the interferometer. Some of the width of the peaks is results from slight drifts of the interferometer during the sweep.

Figure 7b shows the spectral profile of the Rb^{85} probing lamp. For comparison, a Rb^{87} spectral profile is shown in Fig. 7c. Two peaks are visible in Fig. 7a. One peak is displaced to the high-frequency side of the F = 3 component of the Rb^{85} profile by $\approx$225 mK, which is the microwave frequency, while the other peak is displaced by $\approx$225 mK to the low-frequency side of the F = 2 component of the Rb^{85} profile. The spectral components of Fig. 7a do not coincide with the spectral components of either a Rb^{85} lamp profile (Fig. 7b) or a Rb^{87} lamp profile (Fig. 7c). The peaks in Fig. 7a did not result from fluorescently scattered light from the long, Rb^{85} pumping lamp, since they disappeared when the probing lamp was removed. We conclude that the peaks of Fig. 7a are upper and lower sidebands, which were generated by the microwave coherence of the vapor. The sideband power is several parts per thousand of the carrier power.

In order to support the interpretation of the peaks in Fig. 7a as sidebands, we removed the Fabry-Perot interferometer and recorded the integrated intensity of sidebands as a function of microwave power (see Fig. 8a), while keeping the microwave frequency on the center of the resonance. The sidebands attained a maximum intensity at a field strength of about 1.5×10^{-4} gauss. The sidebands decreased in amplitude for greater or lesser field strengths. Thus, the sidebands of Fig. 7a have the same qualitative dependence on microwave power as does the microwave coherence of the vapor.

As a final check of the nature of the peaks in Fig. 7a, we replaced the Fabry-Perot interferometer with a Rb^{85} filter cell,[5] and we recorded the intensity of the probing beam as a function of the filter-cell temperature (see Fig. 8b). There was no noticeable attenuation of the probing beam when the polarizer P_2 was orthogonal to the polarizer P_1, i.e. when only the sidebands passed through the filter cell. The intensity of the probing beam decreased rapidly as a function of filter-cell temperature when the polarizer P_2 was parallel to P_1, i.e. when only the carrier reached the filter cell. The experiment confirms the fact that the peaks of Fig. 7a lie well outside the absorption profile of Rb^{85} vapor.

Figure 7a represents some of the first data taken in our experiment. The observed conversion efficiency was lower than we expected. This was because some of the parameters in our apparatus were not optimized. The phase matching condition of (24) was not precisely met and the buffer gas pressure in the resonance cell was much higher than necessary. By lowering the N_2 buffer gas pressure in the cell from 12 Torr to 5 Torr we were able to get an improvement of a factor of five so that we are now able to obtain a conversion efficiency of about 1%. By also correcting the phase mismatch, we should improve the conversion efficiency even more.

Possible Applications

Perhaps the most promising applications for this microwave-to-optical-frequency converter is its use in a microwave spin oscillator. Figure 9 shows schematically how such an oscillator might work. As we have already mentioned, for linearly polarized carrier light, the sideband waves are polarized orthogonally with respect to the carrier wave. However, if circularly polarized carrier light were used, then the sidebands would also be circularly polarized in the same sense. Beats would therefore occur between the carrier waves and the sideband waves, and the transmitted light from the frequency converter would be intensity modulated[6] at the microwave frequency. Hence, for the spin oscillator, a circularly polarized carrier beam would be used, and the modulated light from the frequency converter would be detected by a microwave photodetector. The microwave output from the photodetector would be amplified, phase shifted by 90°, and fed back to the cavity where it drives the coherence of the vapor. When the loop gain of the

feedback system is adjusted to unity, self-oscillation would occur, and the frequency of oscillation would be fixed by the field-independent 0-0 transition frequency. We are in the process of constructing such an oscillator. Our main problem has been in obtaining a suitable sensitive microwave photodetector. We are presently using a crossed-field photomultiplier tube which has a sufficiently large bandwidth, but it has a low quantum efficiency ($\eta \sim 10^{-4}$) for rubidium D_1-light and a small load impedance (50 Ω). Even though we have sizable light intensity emerging from the cavity and the modulation index of this light is on the order of 10%, the power output from the photomultiplier tube is very low ($\sim 10^{-12}$ W). About 70 dB of additional amplification would be needed in order to achieve oscillation. Therefore, we are considering the construction of a new crossed-field photomultiplier tube with a GaAs photocathode ($\eta \sim$ 8%) and a resonant load. A much larger power output would be obtained with this new tube so that little or no additional amplification would be necessary in order to achieve oscillation.

Since this optically pumped frequency converter under optimum conditions takes microwave photons and converts them into optical sideband photons on a one-to-one basis, a sizable power gain takes place in this conversion process. Because the presence of microwaves is necessary for the appearance of sidebands, one may use this device as a sensitive low noise detector of microwaves. The ultimate sensitivity of this device would probably be limited by how well one can discriminate the sidebands from the carrier rather than by the thermal noise in the microwave cavity. For good discrimination of the sidebands from the carrier one can take advantage of the fact that for linearly polarized carrier photons, the sideband photons are linearly polarized in an orthogonal direction. Further discrimination can be obtained by using a Rb^{85} filter cell. Figure 10 shows an arrangement for using the Rb^{87} up-converter as a microwave photon counter. Microwave photons are picked up by an antenna and the optical sideband photons are detected by a high-gain low-background photomultiplier tube. Such an up-converter would be a very narrow band device but could be made tunable over a range of several megahertz by operating on a field dependent hyperfine transition.

References

*This work was supported by the Joint Services Electronics Program (U. S. Army, U. S. Navy, and U. S. Air Force) under Contract DAAB07-69-C-0383.

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

2. H. Tang and W. Happer, Phys. Rev. Letters 24, 551 (1970).

3. W. Happer and B. S. Mathur, op. cit., p. 12.

4. W. Happer, "Light Propagation and Light Shifts in Optical Pumping Experiments," Progress in Quantum Electronics, eds., K. W. H. Stevens and J. H. Sanders (Pergamon Press, Inc., Oxford, 1970), p. 551.

5. P. Davidovits and N. Knable, Rev. Sci. Instr. 35, 857 (1964).

6. B. S. Mathur, H. Tang, B. Bulos, and W. Happer, Phys. Rev. Letters 21, 1035 (1968).

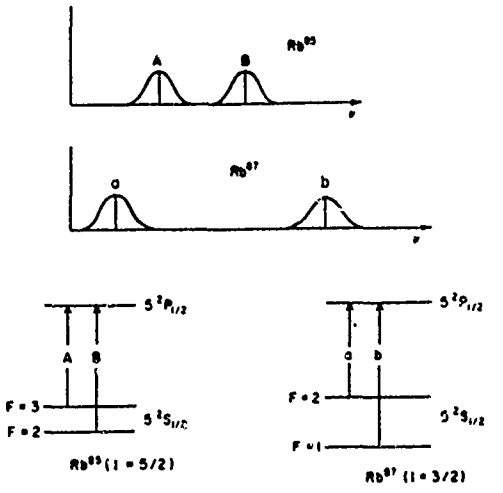


Fig. 1. Hyperfine structure of the $D_1(^2S_{1/2} - {}^2P_{1/2})$ resonance line of Rb^{85} and Rb^{87}.

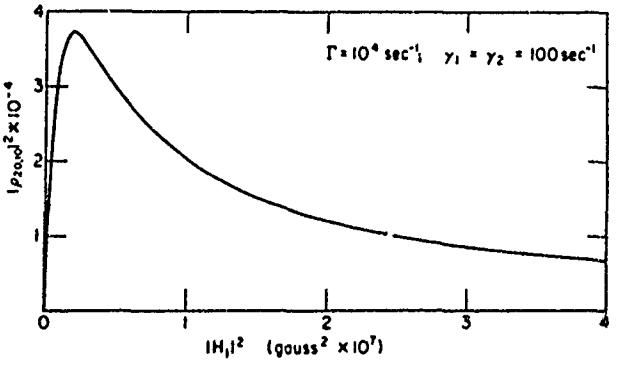


Fig. 2. The square of the magnitude of the off-diagonal density matrix element for an optically pumped Rb^{87} vapor versus the square of the magnitude of the microwave magnetic field which excites the 0 - 0 hfs transition. The vapor parameters are assumed to be: Γ(optical pumping rate) = 10^4 sec^{-1} γ_1(longitudinal relaxation rate) = γ_2(transverse relaxation rate) = 100 sec^{-1}.

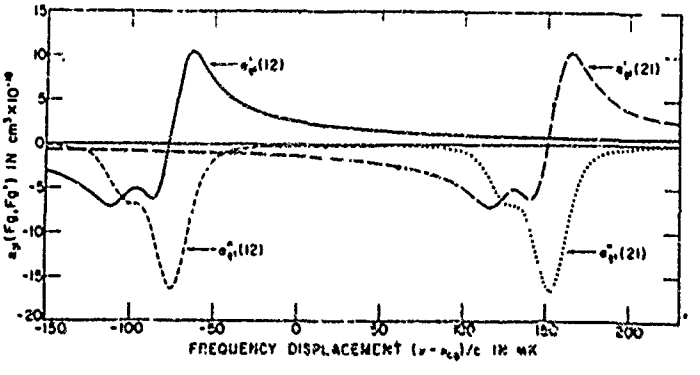


Fig. 3. The frequency response functions of the gyrotropic up-coupling, $\alpha_{gt}(12)$, and down-coupling, $\alpha_{gt}(21)$ polarizabilities of a Rb^{87} vapor. The horizontal scale measures the frequency displacement of the optical carrier frequency from the center of gravity of the Rb^{87} optical absorption lines.

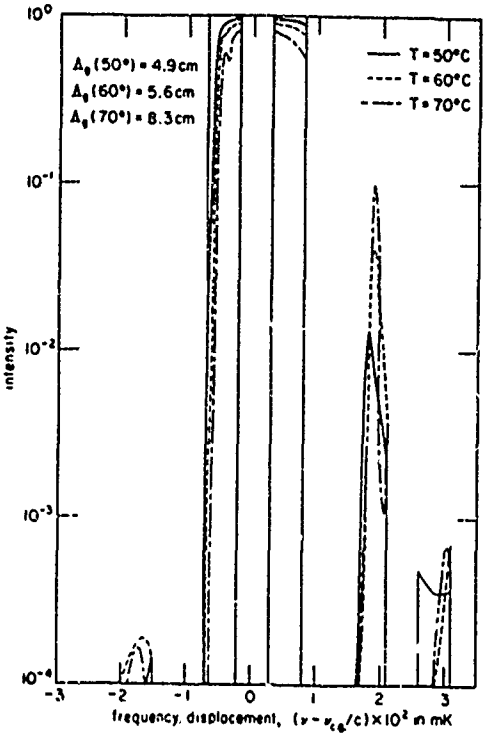


Fig. 4. Predicted modification of the D_1-line of Rb^{85} resonance light after passing through a 40 cm length of optically pumped Rb^{87} vapor with hfs coherence. The initial spectral profile of Rb^{85} light is represented by two hyperfine components of rectangular shape, and 50 mK in width. The quantity λ_g is the guide wavelength of the microwaves which is adjusted to meet the phase matching condition of Eq. (24) at each value of the cell temperature. The following vapor parameters are assumed:

Y(pressure broadening parameter)	=	0.10
$\langle I \, J \rangle$	=	-0.923
Γ(optical pumping rate)	=	10^4 sec^{-1}
γ_1(longitudinal relaxation rate)	=	100 sec^{-1}
γ_2(transverse relaxation rate)	=	100 sec^{-1}

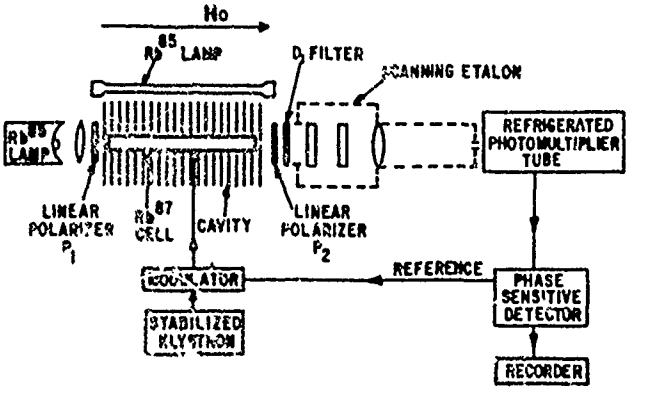


Fig. 5. Schematic diagram of the apparatus used to generate and to observe the sidebands.

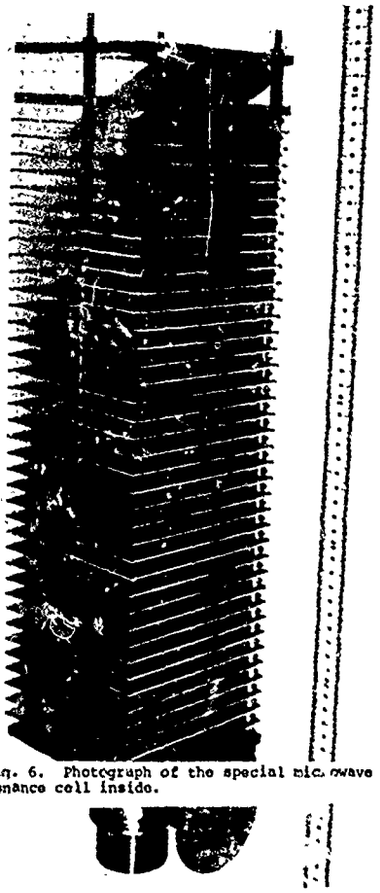


Fig. 6. Photograph of the special microwave cavity with the long resonance cell inside.

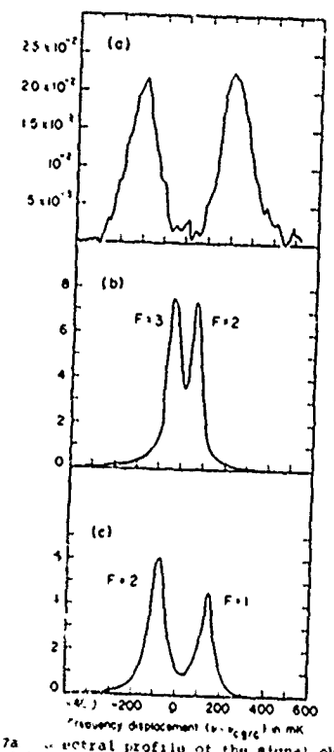


Fig. 7a. Spectral profile of the signal observed through the scanning Fabry-Perot interferometer with polarizers P_1 and P_2 crossed.

Fig. 7b. Spectral profile of the P_1 line of a Rb^{85} lamp before passing through the Rb^{85} vapor.

Fig. 7c. Spectral profile of the D_1 line of a Rb^{87} lamp.

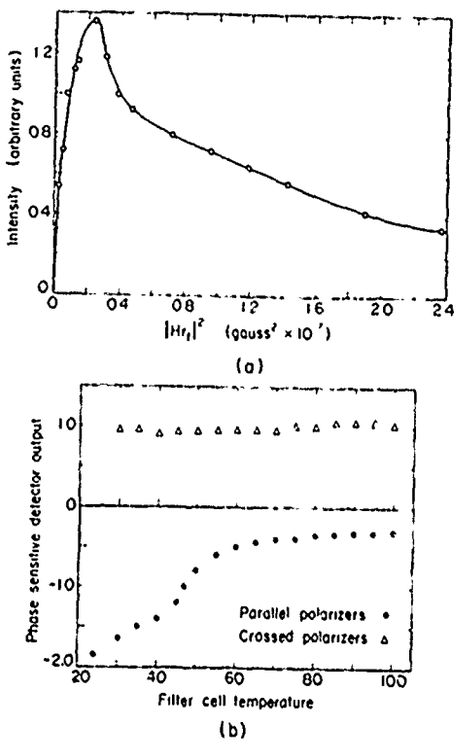


Fig. 8a. Total intensity observed with the phototube directly behind the polarizer P_2 and the D_1-filter versus the square of the magnitude of the longitudinal microwave magnetic field. The polarizers P_1 and P_2 were crossed.

Fig. 8b Total intensity observed through a Rb^{85} filter cell with the polarizers P_1 and P_2 crossed (sideband intensity) and with P_1 and P_2 parallel (carrier intensity).

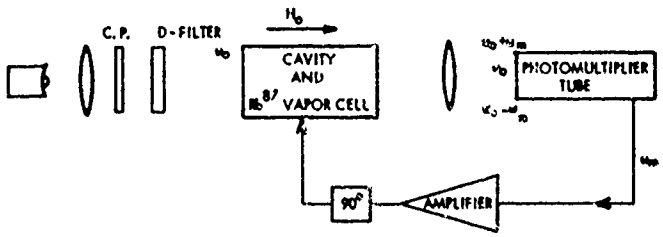


Fig. 9. Schematic diagram of a microwave spin generator.

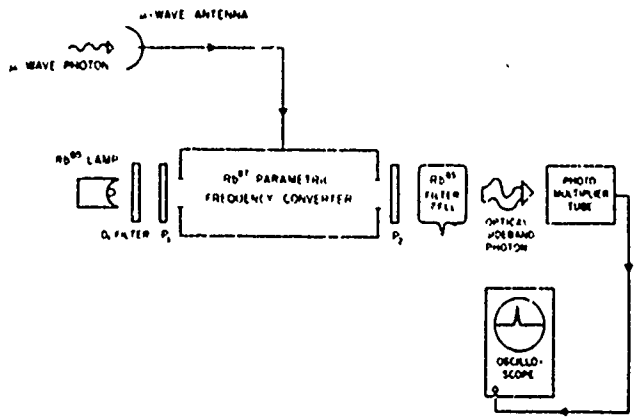


Fig. 10. Schematic diagram of a microwave photon counter.

CESIUM BEAM SERVO SYSTEM USING SQUARE WAVE FREQUENCY MODULATION

H. Daams
National Research Council of Canada
Ottawa, Ontario

Summary

This paper describes a new modulation and servo system which is being tested in the NRC long beam cesium frequency standard.

Square wave frequency modulation is employed: it is introduced by switching between two separately generated frequencies.

The synthesizer producing these frequencies and the problems caused by switching transients will be described.

Since in a laboratory standard systematic errors play a very important role these errors are discussed in the description of the servo system. The resulting systematic error as calculated from results of experiments on its components is less than 1 part in 10^{13}

Introduction

The 2.1 meter cesium beam standard of the Physics Division of the National Research Council of Canada currently acts as primary standard of time and frequency for Canada. A cesium beam tube of this length can provide high intrinsic accuracy. This is a report on progress in developing a solid state servo system to enable us to realize this accuracy. As in the earlier system described in Proceedings of the 1965 Frequency Control Symposium[1], square wave frequency modulation is used but the modulation frequency is increased by a factor 80. Modulation, which is in effect frequency switching, takes place in the 12.632 MHz synthesizer which is second reference in the final phase lock loop.

The transients resulting from the switching give rise to certain potential systematic errors. These will be discussed in detail. The aim has been to reduce errors in the servo loop to less than an acceptably low level of $\Delta f/f = 5.10^{-14}$ for an averaging time of 10,000 seconds. A servo system for a cesium beam standard may employ sinusoidal or square wave modulation. In the case of sinusoidal modulation either frequency or phase modulation can be used; the resultant r.f. output signal being the same in both cases. An important consideration is the effect of harmonic distortion in the modulation, in particular due to even harmonics. Fig. 1 a) shows how an undistorted frequency modulated signal centered on a symmetric resonance produces zero beam signal at the fundamental of the modulation frequency. If however second harmonic distortion is present as shown on an exagerated scale in Fig. 1 b), then although the carrier frequency of the r.f. signal is centered on the cesium resonance, a signal is produced at the fundamental modulation frequency. The servo system tries to compensate for this signal and thus produces a systematic error. These problems are discussed by Shirley[2].

Both phase and frequency square wave modulation can be employed. Square wave phase modulation, discussed in detail by Searle, e.a.[3, 4], makes use of the apparent phase shift between the two r.f. cavities which an atom observes when the phase is switched during the time of flight between the two cavities. Since a signal is produced only by atoms which are between the two cavities when the phase change occurs square wave phase modulation has to be used at a high enough modulation frequency that its period is similar to the inter-cavity time of flight.

Square wave frequency modulation has been used in the NRC cesium standard since 1965 and is also used in the new standard of the PTB. In this system the modulation frequency can be much lower than the frequency deviation. In the present NRC system 0.1 Hz is used as the modulation frequency and 41 Hz as the frequency deviation.

The inherent switching transients which cause an unwanted phase modulation occur for only .01 second of the 10 second modulation period and can be neglected. In the new system at present being tested 41 2/3 Hz frequency deviation is still used but the modulation frequency is increased to 8 1/3 Hz. Transients now have to be considered and either they must be eliminated or measures must be taken to prevent systematic errors. As mentioned above the beam current transients arise from the fact that atoms which left the first cavity before a frequency change and arrive at the second cavity after the change observe in effect a varying phase difference of large magnitude between the two cavities.

In Fig. 2 this phase difference is shown for a beam of atoms all of the same velocity; atoms arriving at the second cavity after the moment of switching observe that the phase increases linearly with the time between the moment of switching and the arrival time:

$$\Delta\phi = \Delta\omega(t_a - t_s)$$

t_a = arrival time

t_s = moment of switching

The value of $\Delta\phi$ is for a mean time of flight of 5.10^{-3} s and $\Delta\omega = 2\Pi.2.41\ 2/3 \simeq 500$:

$$\Delta\phi \simeq 2.5 \text{ radian.}$$

The velocity distribution in the beam smears out the transient. The transients in Fig. 2 are both shown as positive signals; that they both have the same polarity follows by considering (Fig.3) Ramsey resonance patterns for different values of phase difference between cavities. In one case frequency of excitation in the first cavity was low before switching frequency and the apparent phase difference increases afterwards, in the other case the frequency was high and the apparent phase difference decreases.

In case of a symmetric resonance pattern and distortionless modulation therefore the beam current error signal only contains even harmonics when the carrier is centered on the resonance pattern. These harmonics are the equivalent of the second harmonic signal produced in sinusoidal modulation. A symmetric cesium resonance pattern implies zero phase difference between cavities. If this condition is not fulfilled or if asymmetry is caused by C field distortion, the transient and "steady state" signal should permit evaluation of cavity phase difference.

Experimental

A. Beam excitation signal

The system used for excitation of the cesium resonance in the NRC standard is shown in Fig. 4. A phase locked loop is used to add 12.632 MHz from a frequency syntnesizer to 9180 MHz multiplied from 5 MHz. The 12.632 MHz signal contains the square wave frequency modulation, which is therefore transferred to the 9192.632 MHz excitation signal. In Fig. 5 the principle of synthesis of 12.632 MHz is shown in more detail. 12.5 MHz obtained from 5 MHz is mixed with 132 kHz containing the modulation. The 132 kHz center frequency is obtained from a voltage controlled crystal oscillator, and is phase locked by division by 66 and phase comparison with a 2 kHz reference derived from 5 MHz.

From this 2 kHz reference are obtained the frequency offset for modulation of 41 2/3 Hz and the modulation frequency of 8 1/3 Hz. These values were chosen because they are subharmonics of 2 kHz and are not too harmonically related to the 60 Hz line frequency.

In Fig. 6 the modulation system is shown. Two voltage controlled crystal oscillators are phase locked 41 2/3 Hz above and below the 132 kHz reference, and modulation takes place by switching between the outputs of the two oscillators. The modulation frequency of 8 1/3 Hz and the frequency deviation of 41 2/3 Hz are both derived from the same source. They and the 132 kHz signals are present as square wave signals. Since they are all synchronous the zero crossings of 8 1/3 Hz coincide not only with those of the 132 kHz but also with those of the frequencies 41 2/3 Hz higher and lower. Therefore when switching occurs between the latter two signals there is no phase discontinuity since it occurs at either 0° or 180° phase. There is thus also no phase discontinuity in the output of the mixer where 12.5 MHz is added to the 132 kHz signal. The importance of phase continuity was demonstrated by using a non synchronous 8 Hz signal for frequency switching. Serious transients in the 12.632 MHz output phase and amplitude occurred. They were of sufficient amplitude to upset the 9192, 632 MHz klystron phase lock loop and caused the servo to produce systematic frequency errors.

The synthesizer was constructed mainly from digital intigrated circuits. There are only two tuned circuits at 12.5 and 12.632 MHz and for both tuning is not critical. Since 132 kHz side-bands in the 9192,632 MHz signal would excite the $(4,3\leftrightarrow3,3)$ and $(4,-3\leftrightarrow3,-3)$ field dependent microwave resonances, removal of these sidebands is essential. This is achieved by passing the 12.632 MHz signal through a crystal filter of a few kHz bandwidth. Phase stability over a period of 2000 sec is better than 0.1 radian. A systematic error in measuring the cesium resonant frequency of less than 1 part in 10^{15} would be produced by gradual phase shift of such magnitude.

B. Servo system

The servo system consists.of an electrometer preamplifier, an active bandpass filter, demodulator and integrator. The integrator output feeds the varicap diodes in the voltage controlled 5 MHz crystal oscillator. The servo loop is characterized by an integration and one principal time constant. In the new system the latter is the 80 millisecond response time of the 2 Hz wide servo bandpass amplifier. In the earlier system it was the time constant of the hot wire and electrometer combination used which varied from 1 to 10 second. From the formula (Fig.7) for the open loop gain for the system

$$KG = \frac{K_{osc}\ K_{det}}{(1 + s\tau)sT}$$

where K_{osc} = oscillator sensitivity rad/volt sec

K_{det} = detector sensitivity volt/sec/rad

τ = principal time constant

T = integrator time constant

and the condition for critical damping:

$$K_{osc}\ K_{det} = \frac{T}{\tau}$$

the maximum loop gain is found to be

$$(KG)_{max} = \frac{1}{(1 + s\tau)s}$$

It will be noted, that the integrator time constant T has disappeared.

We are most interested in averaging times for which

$$\left|\frac{1}{s}\right| >> \tau$$

so that $(KG)_{max} = \frac{1}{s\tau}$

which shows the importance of a small time constant in the loop.

Limits to this time constant, that is to the bandwidth of the servo amplifier are however set by the presence of noise. The main noise source is shot noise in the beam current. If the bandwidth chosen is too wide the noise amplitude will be such that the amplifier operates in the non-linear region and noise is mixed into the desired signal. In the design it was assumed that noise peaks exceeding shot noise by a factor of 10 should not drive the operational amplifiers in the active bandpass amplifier into saturation. Care was also taken that clipping occurred only in the final stage of amplification, that is in the untuned amplifier and phase inverter between bandpass amplifier and synchronous detector.

The beam current error signal is first amplified by a Keithley electrometer operational amplifier with a 10^{11} ohm feedback resistor. This is followed by the bandpass amplifier which uses an active 3 pole 3 stage filter 2 Hz wide centered at 8 1/3 Hz. An untuned amplifier and a phase inverter then supply the error signal to a balanced MOSFET detector. It was found that to avoid non-linearity in the MOSFET "on" characteristics the signal had to be limited; operational amplifiers which are used in the preceeding amplifier stages appear to have quite symmetric limiting characteristics. The integrator following the detector is chopper stabilized and has a 20 second time constant.

C. Errors

The sources of error signals in the servo loop are shown in Fig. 8. They are beam noise, demodulator unbalance, integrator offset voltage, oscillator offset and multiplier output phase variation. These are sources not only of random noise but also of dc signals as for instance integrator operation amplifier offset and drift. These sources can each be translated into fractional frequency deviations. They are counteracted by the feedback in the loop, although not all by the same amount and some not at all. Beam noise can not be eliminated since it enters at the same point as the cesium reference. Another error contribution which is not compensated is a gradually increasing phase shift in multiplier and synthesizer chains.

For a systematic error not exceeding

$$\frac{\Delta f}{f} = 10^{-14}$$

for a 10 hour interval an upper limit can be placed on all the factors shown in the diagram. Examples are the offset voltage of the integrator operational amplifier, drift of the 5 MHz oscillator, gradual phase shifts in the excitation signal and synchronous detector zero signal output.

Calculation of the different inherent sources of error in the system shows that the possible error does not exceed

$$\frac{\Delta f}{f} = 10^{-13}.$$

An error appeared at first because of the 8 1/3 Hz reference signal leaking into the electrometer amplifier through ground loops. When it is realized that a frequency offset of 10^{-13} corresponds to an error signal of 10^{-17} amp rms, an interfering signal of this magnitude can be generated by milliamps of reference signal flowing through a cable shield.

A second error arose from the 8 1/3 Hz chopping frequency for the synchronous detector and harmonics of this signal not being adequately filtered by the integrator. Frequency modulation at 8 1/3 Hz of the 5 MHz oscillator produced an error signal 90° out of phase with the wanted error signal. The integrator output was therefore passed through two 3 pole active low pass filters of 2 Hz and 3 Hz cut off frequency. These filters also remove effectively any 60 Hz line frequency components which might be present in the oscillator control signal.

D. Results

The short term stability of the system was tested by comparing the 5 MHz output signal against one of the two NRC hydrogen masers. At the same time the beat between the two masers was monitored. Short term stability was found to be

$$\frac{\Delta f}{f} = 2 \times 10^{-12}$$

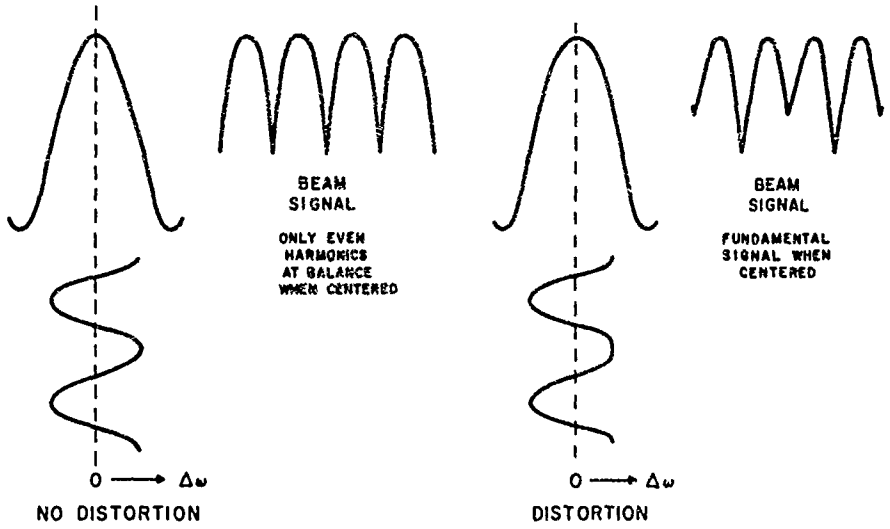
for 28 sec averaging time.

Old and new servo systems agree to within 1 part in 10^{13}. However variations using the new system are considerably smaller and noise pulses in the beam current which upset the old system considerably do not affect the performance of the new system.

Data currently published by NRC will continue to be based on the old system until further evidence has been obtained to justify transfer to the new system.

References

1. H. Daams and A.G. Mungall: Proc. 19th Annual Frequency Control Symposium, Atlantic City, 1965, p. 332.
2. G.H. Shirley: G. Appl. Phys., 34, 789, 1963.
3. R.S. Badessa, V.J. Bates and C.L. Searle: IEEE Trans. Instr. and Meas. IM-13, 175, 1964.
4. C.M. Bell and D. Babitch: IEEE Trans. Instr. and Meas. IM-17, 155, 1968.



EFFECT OF SECOND HARMONIC DISTORTION

FIG 1

Effect of second harmonic distortion

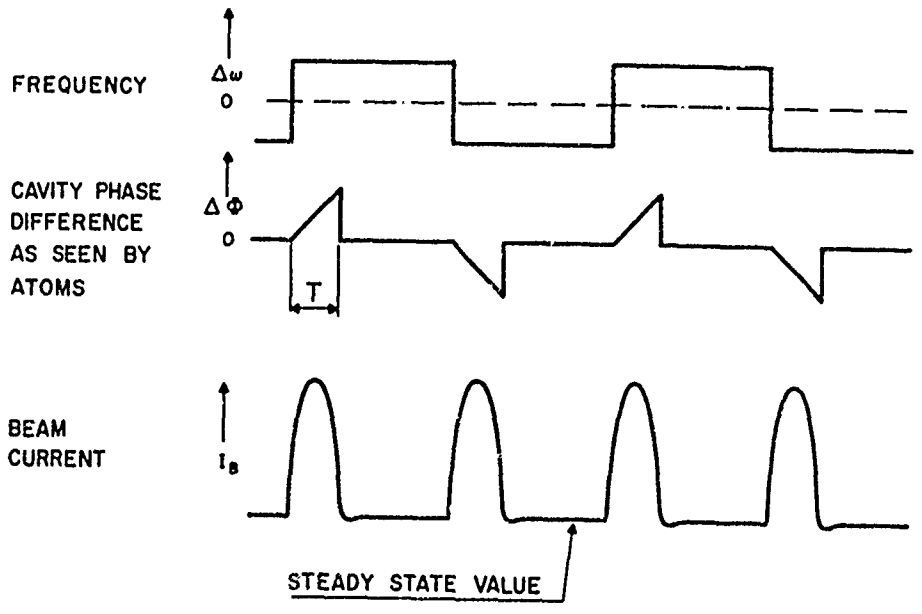


FIG. 2

Beam current transients due to frequency switching.

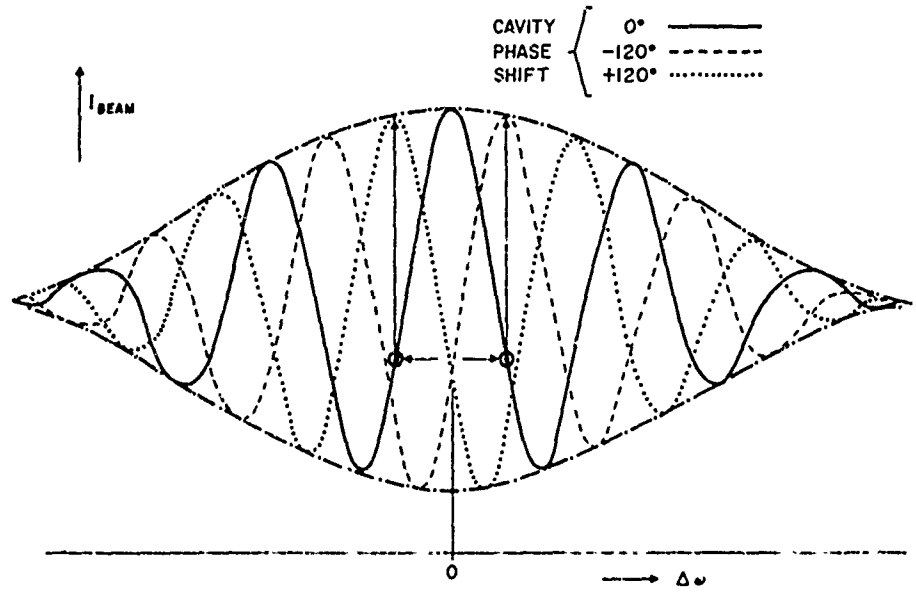


FIG 3

Ramsey resonances for different values of cavity phase shift

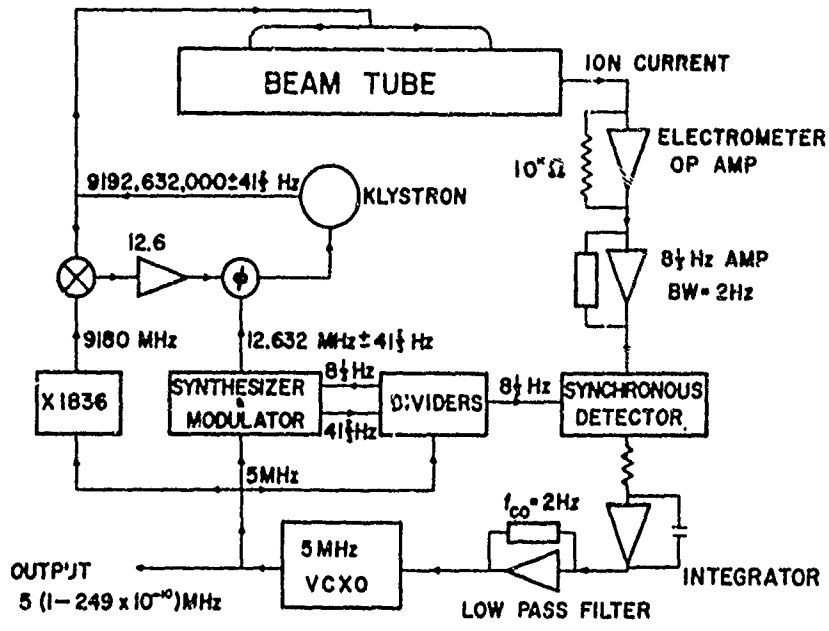
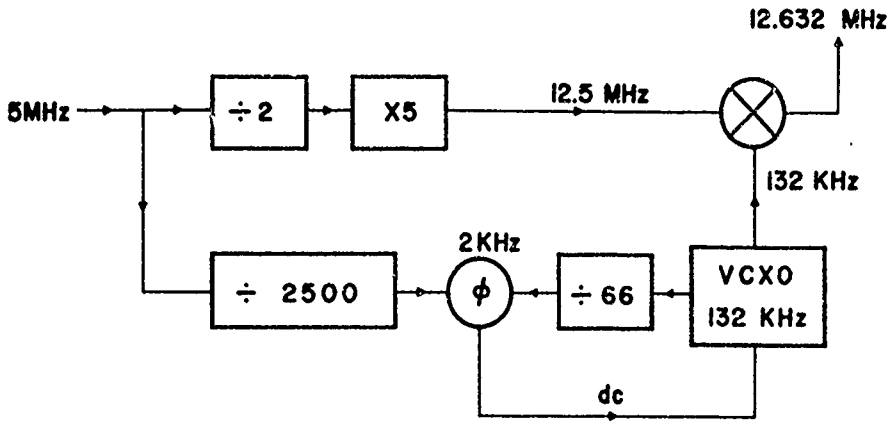


FIG 4

Block diagram of cesium beam standard.



SYNTHESIS OF 12.632 MHz

FIG 5

Synthesis of 12.632 MHz

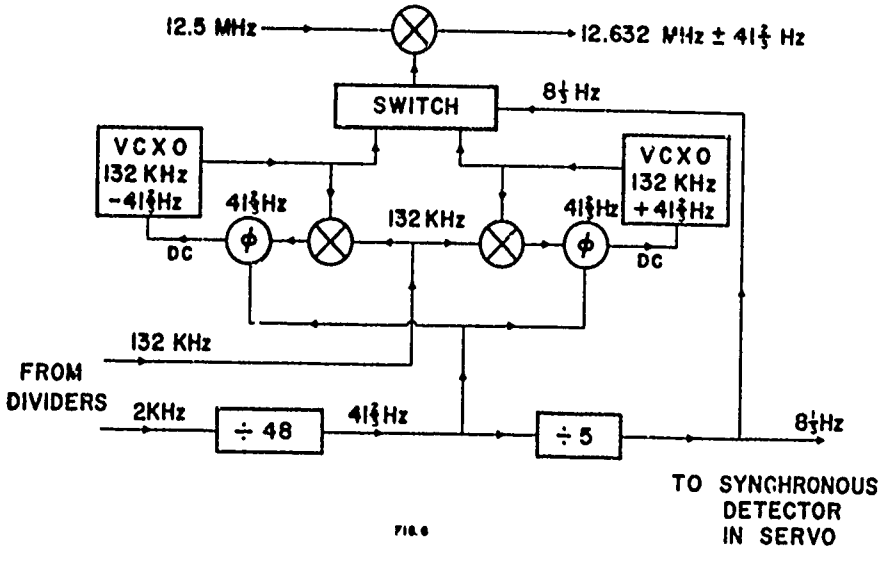


FIG 6

Frequency modulation system.

OPEN LOOP GAIN:

$$KG = \frac{K_{OSC}\ K_{DET}}{(1+s\tau)\ sT}$$

K_{OSC}: $^{RAD/SEC/VOLT}$

K_{DET} : $^{VOLTS/RAD/SEC}$

STABILITY:

$$K_{OSC}\ K_{DET} \leqslant \frac{T}{\tau}$$

τ = AMPLIFIER TIME CONSTANT

T = INTEGRATOR TIME CONSTANT

MAX. OPEN LOOP GAIN:

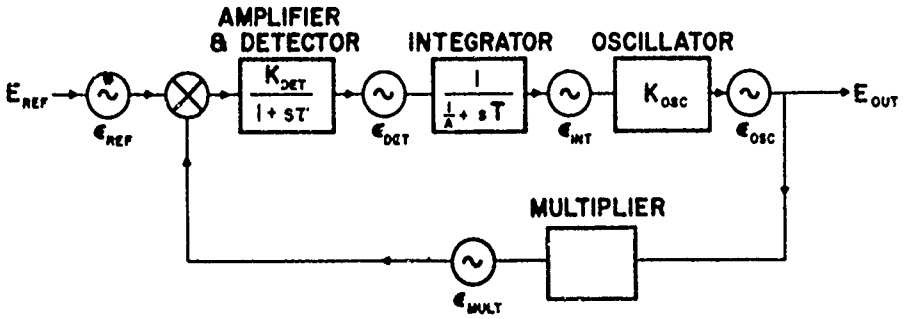
$$(KG)_{MAX} = \frac{1}{(1+s\tau)\ s\tau}$$

LONG AVERAGING TIME:

$$\frac{1}{s} \gg \tau \longrightarrow (KG)_{MAX} = \frac{1}{s\tau}$$

FIG 7

Open loop gain of system.



$$E_{OUT} = \frac{KG}{1+KG}\left[\left(E_{REF} + \epsilon_{REF} - \epsilon_{MULT}\right) + \frac{1+s\tau}{K_{DET}}\ \epsilon_{DET} + \right.$$

$$\left. + \frac{K_{OSC}}{1+KG}\ \epsilon_{INT} + \frac{1}{1+KG}\ \epsilon_{OSC}\right]$$

FIG 8

Sources of error in servo system.

FREQUENCY COMPARISON OF FIVE COMMERCIAL STANDARDS WITH A NASA EXPERIMENTAL HYDROGEN MASER

A. R. Chi,
F. G. Major,
J. E. Lavery
Goddard Space Flight Center,
Greenbelt, Maryland

Abstract

Recent long term phase comparisons between five commercial cesium beam frequency standards (Hewlett-Package Co. Model No. CO2-5061A, a special modification for NASA Manned Space Flight Network) and a NASA experimental hydrogen maser, designated as NX1, have provided data for a determination of the difference in frequency between these standards. The result of this comparison is a mean value for the NX-1 frequency that is 6.7 parts in 10^{12} lower than the commercial Cs units with a statistical uncertainty in the measurements (1σ) of 1.6×10^{-12}. The NX-1 maser synthesizer frequency is set according to the Vessot, et al. (1966) value of the hydrogen frequency with corrections applied for magnetic field, second order Doppler and wall shift according to coefficients derived from previous studies of Vanier, Peters, and Vessot. On applying the magnetic field correction to the Cs standards and assuming the manufacturer's data on the dispersion in frequency among the commercial units and their 1.3×10^{-12} agreement with NBS III, it is concluded that the frequency of NX1 is 1420405751.7767 ± 0.0024 Hertz.

Introduction

Recent measurements of the long-term frequency stability of five commercial cesium beam frequency standards[1] relative to a NASA experimental hydrogen maser,[2] constructed at GSFC and designated NX-1, have made possible a new comparison of these two types of atomic frequency standards. The result of this comparison is a value for the hydrogen maser frequency that is 6.77 parts in 10^{12} lower than that given by Vessot et al.[3] in 1966.

The frequency comparison of the atomic standards is based on continuous phase difference measurements, sampled hourly, over the 123 days between May 7 and September 7, 1968. The 2952 hourly phase difference data thus obtained were punched on cards for computer analysis. Frequency data were obtained by taking the time derivative of successive five-data point least-square-fits of the phase data, using Koenig's approximation formula.[4] The frequency data were then reduced to yield the mean fractional frequency difference between each cesium standard and the hydrogen maser. These figures were finally corrected so as to obtain the results in terms of the A-1 second.

Data Reduction

To obtain the results shown in Table 1, the frequency data were least-square fitted (LSF) to a linear equation for frequency as a function of time. The fact that the total frequency change over the entire period of measurement due to the time dependent term was less than 1σ indicates that the cesium beam standards exhibit no systematic linear frequency drift. However, in calculating the mean frequency difference, it was convenient to take the sum of the LSF frequency intercept at $t = 0$ and one-half of the total frequency change due to the LSF time dependent term (these two components are given in Table 1). That this procedure correctly yields the mean of the frequency difference data can readily be proved. An alternative way for calculating the mean frequency difference is simply taking the difference in phase change between the final and initial points and dividing it by the elapsed time. This was correctly pointed out by Barnes of NBS. The results, of course, are the same.

Corrections

The frequency output of the NX-1 hydrogen maser is synthesized to the value given by Vessot et al.[3] in 1966, corrected for differences in bulb shape, second order Doppler shift, and applied magnetic field. For these corrections see appendix. To achieve frequency outputs which are in terms of the coordinated universal time, this synthesized NX-1 output and the output of each of the cesium standards are offset by an amount which is nominally -300×10^{-10}, the UT-C offset for 1968, relative to the defined cesium resonance frequency based on the SI second.[5]

In the NX-1 this offset is maintained at exactly -300×10^{-10}. In the cesium standards, however, the amount of the offset was found to be slightly larger. Hence the measured fractional frequency difference between the cesium standards and the hydrogen maser must be corrected for the difference between the actual offset in the cesium standards and -300×10^{-10}.

The offset in the cesium standards is maintained by an internal electronic synthesizer and a magnetic field fine tuner. The first two columns of Table 2 show the contribution to the offset of each of these sources. The third column shows the effect on the offset of the magnetic field drift which was observed during the measurements. (The approximately linear behavior of this drift

with time, shown in Figure 1, was used to compute its contribution to the offset.) The remaining columns of Table 2 show the total offset of each cesium standard and the appropriate correction to account for the difference between this offset and -300×10^{-10}.

After applying the corrections discussed above, the total fractional frequency difference between each cesium standard and the hydrogen maser,

$$\left(\frac{\overline{\Delta f}}{f}\right)_{C_s - HM},$$

is as given in Table 3. The mean fractional frequency difference for the five units is 6.77×10^{-12}. The deviation, δ, of

$$\left(\frac{\overline{\Delta f}}{f}\right)_{C_s - HM}$$

for each cesium standard from this mean is also given in Table 3, yielding a δ_{RMS} of $\pm 1.11 \times 10^{-12}$.

Estimated Errors

If the total error in the mean fractional frequency difference is taken to be the root-sum square of σ_{RMS} from Table 1, and δ_{RMS}, from Table 3, a total error of 1.66×10^{-12} is obtained. This is consistent with the result obtained by combining the estimated absolute errors of the Hewlett-Packard cesium standards (1.33×10^{-12}) and the NX-1 hydrogen maser (1.0×10^{-12}), which yields a total of 1.64×10^{-12}. These individual errors are compared in Table 4 with those of other similar units.

Conclusions

Our measured value for the fractional frequency difference between the commercial cesium standards and the NASA experimental maser is thus $(6.77 \pm 1.66) \times 10^{-12}$. Since the frequency of the NASA maser is synthesized to the 1966 measurement by Vessot et al., our value for the hydrogen maser frequency is therefore $f_{HM} = 1{,}420{,}405{,}751.7767 \pm .0024$ Hertz.

If this value is compared with two other measurements also made in 1968, those of A. G. Mungall et al.[10] and C. Menoud et al.[11] as shown in Table 5, it is evident that these three independent measurements agree within their stated errors.

It is also interesting to note that the three results shown in Table 5 fall within the error limits of the earlier results of Winkler[12] and of Peters et al.[13]. The discrepancy between these five measurements and the measurements of Vessot et al. indicates that further comparisons are desirable.

Acknowledgment

The authors wish to express their appreciation to H. E. Peters for providing the hydrogen maser physical parameters and for calculating the corrections for the NX-1 maser and to Drs. J. A. Barnes, R. F. C. Vessot and A. G. Mungall for their helpful comments.

Notes and References

1. The commercial cesium beam standards are manufactured by Hewlett-Packard Company. The first four units, which are all model 5061A, bear the serial numbers 136, 138, 137 and 139 respectively and were furnished to us for the determination of their long-term frequency stability by the Manned Flight Support Directorate. The fifth unit, model 5060A, serial number 152, is a laboratory standard.

2. H. E. Peters, E. H. Johnson, and T. E. McGunigal, "Atomic Hydrogen Standards for NASA Tracking Stations," Proc. 23rd Annual Frequency Control Symposium (May 1969).

3. R. Vessot, H. Peters, J. Vanier, R. Beehler, D. Halford, R. Harrach, D. Allan, D. Glaze, C. Snider, J. Barnes, L. Cutler, "An Intercomparison of Hydrogen and Cesium Frequency Standards," IEEE Trans. on Instr. & Meas. IM-15, 4, 165 (December 1966). Also, R. Beehler, D. Halford, R. Harrach, D. Allan, D. Glaze, C. Snider, J. Barnes, R. Vessot, H. Peters, J. Vanier, Proc. IEEE 54, 2, 301 (February 1966).

4. D. M. Koenig, "Relative Phase Difference Analysis Program," Final Report, ASEE NASA Summer Faculty Fellowship Program, NASA Document X-520-68-179, pp. 77-88 (April 1968).

5. The 13th General Conference on Weights and Measures in 1967 adopted the definition of the second as that length of time which results in a value of exactly 9, 192, 631, 770 Hertz for the transition frequency between the $F = 4$, $M_F = 0$, and $F = 3$, $M_F = 0$, hyperfine levels of cesium 133.

6. Private communication with L. N. Bodily. Based on the measurements of 150 units of Hewlett-Packard cesium standards during factory test, the 3σ fractional frequency distribution of all the units measured relative to the house standard is 4×10^{-12}. Through VLF phase comparison, the fractional frequency difference of the house standard relative to the frequency of the NBS III cesium standard is known to be within 1×10^{-12}.

7. Private communication with R. E. Beehler. Under normal care, the accuracy of the NBS III cesium standard is $3\sigma = 5.6 \times 10^{-12}$. When the Varian H-10 hydrogen maser was brought for comparison with NBS III in 1965 special care was given to

the operation and error measurement of the cesium beam standard. The best estimate of the error of the NBS III during the 50 days of that comparison was $1\sigma = 1.1 \times 10^{-12}$. Since 1966 NBS III has been improved in its vacuum system and frequency multiplier chains. The accuracy of this frequency standard is at least as quoted if not better. See "Improvements in Atomic Cesium Beam Frequency Standards at the National Bureau of Standards" D. J. Glaze 16th General Assembly of the International Scientific Union (URSI) Ottawa, Canada, August, 1969.

8. A. G. Mungall, R. Bailey, H. Daams and D. Morris, "A Re-Evaluation of the NRC Long Cesium Beam Frequency Standard," Metrologia 4, 4, 165 (October 1968).

9. Private communication with H. E. Peters. This value of σ is based on his repeated comparisons of the Varian H-10 and NASA NX-1 hydrogen masers. The second order Doppler shift, magnetic field and wall coating corrections were applied in all of his measurements.

10. A. G. Mungall, D. Morris, H. Daams, and R. Bailey, "Atomic Hydrogen Maser Development at the National Research Council of Canada," Metrologia 4, 3, 87 (July 1968). Also private communication with A. G. Mungall. The ±0.0020 Hertz uncertainty in the NRC measurement is based on the RMS uncertainties of $2\sigma = 1.5 \times 10^{-12}$ for NRC Cs III and $1\sigma = 1 \times 10^{-12}$ for hydrogen masers No. 1 and No. 2.

11. C. Menoud and J. Racine, "Stabilitat und Genauigkeit der Frequenz von Wasserstoff - Masern," Z. Ang. Math U Phys. 20, 4, 5-78-584 (1969).

12. Private communication with G. M. R. Winkler. His result for the hydrogen maser frequency, based on measurements carried out during 1965-66 between the U.S. Naval Research Laboratory's Varian H-10 Maser and the U.S. Naval Observatory's cesium maser clock, using a telephone link and averaged over a 20-day period, is 1420, 405, 751.7765 ± .0020 Hertz.

13. H. E. Peters, J. Holloway, A. S. Bagley, L. S. Cutler, "Hydrogen Maser and Cesium Beam Tube Frequency Standards Comparison," Appl. Phys. Lett. 6, 2, 34 (Jan. 15, 1965). This comparison of a Varian hydrogen maser with Hewlett-Packard cesium standards yielded a value of 1420, 405, 751.778 ± .016 Hertz for the hydrogen maser frequency. (The estimated error in this case was large enough to include both our result and the Vessot's result.)

14. Private communication with Vanier and Vessot.

15. H. E. Peters, T. E. McGunigal and E. H. Johnson "Hydrogen Standard Work at Goddard Space Flight Center" pp. 464-491 Proc. of the 22nd Annual Symposium on Frequency Control, (April 1968).

Appendix: Frequency Output of NASA Experimental Hydrogen Maser, NX-1.

The hydrogen frequency used to synthesize the NX-1 maser is based in the value given by Vessot et al.[3] in 1966. This frequency is:

$$\nu = 1{,}420{,}405{,}751.7864 \pm .0017 \text{ hertz.}$$

The NX-1 synthesizer was set in accordance with the following corrections:

1. Wall Shift – Following Vanier and Vessot's measurement of 1964, the wall shift correction for a bulb coated with FEP Teflon is given by

$$\delta\nu_w = \frac{K}{D}\left[1 + \alpha\,(T - T_1)\right] \qquad (1)$$

where

$\delta\nu_w$ = wall shift correction in hertz

K = wall shift constant in hertz-inch

α = wall shift temperature coefficient relative to T_1

T = absolute temperature of storage bulb

$T_1 = 40 + 273.16 = 313.16°K$

Based on their measurements, using five different diameter bulbs (4", 5-1/2", 6", 7", and 8") and operating temperature in the range of 27 to 40°C, normalized to 40°C, Vanier and Vessot gave

$$K = -(208 \pm 2) \times 10^{-3} \text{ Hz - in}$$

$$\alpha = -(5 \pm 1) \times 10^{-3} \text{ deg}^{-1} \text{ K.}$$

The value of the wall shift correction for the NX-1 maser is

$$\delta\nu_w = -.02807 \pm .00031 \text{ Hz.}$$

The pertinent physical parameters used for this calculation were furnished by H. E. Peter[15] et al. These parameters are:

Bulb shape: cylindrical with spherical ends

Diameter of cylinder: 5.5 in.

Length: 17 in.

Surface Area: $S = 1863 \pm 3$ cm^2 (calc.)

Volume: $V = 5722.8 \pm 1$ cm^3 (Meas.)

Bulb Temperature: $t = 44.3 \pm 0.5$°C

Equivalent spherical diameter: $D = 7.25 \pm .0104$ in.

The wall shift uncertainty of ± .00031 hertz is due to the uncertainties in the determination of the quantities given in Table A1.

2. Second-order Doppler correction is given by

$$\delta\nu_d = -\frac{3}{2}\frac{kT}{Mc^2}\nu \qquad (2)$$

where

$\delta\nu_d$ = second-order Doppler correction

k = Boltzmann's constant

M = Mass of hydrogen atom

c = speed of light

ν = hydrogen frequency

$$\delta\nu_d = -1.9597 \times 10^{-4}\ T$$
$$= -.06221 \text{ Hz.}$$

The uncertainty in $\delta\nu_d$ due to ΔT is

$$\Delta\,\delta\nu_d = -\delta\nu_d \frac{\Delta T}{T} = .00010 \text{ Hz.}$$

Therefore,

$$\delta\nu_d = -.06221 \pm .00010 \text{ Hz.}$$

3. Magnetic field correction is given by

$$\delta\nu_H = 2750\ H^2 \qquad (3)$$

and

$$H = \frac{\nu_z}{1.4 \times 10^6} \qquad (4)$$

where

$\delta\nu_H$ = magnetic field correction

ν_z = Zeeman frequency

H = magnetic field

The magnetic field of NX-1 maser was measured at intervals on dates given in Table A2. The root mean square of the magnetic field was 4.35 milligauss with an uncertainty of ± 6.57 microgauss. Therefore

$$\delta\nu_H = .05205 \pm .00016 \text{ Hz.}$$

The total correction applied to the NX-1 maser is -0.03823 hertz with an uncertainty of ±.00036 hertz. A summary of the corrections is given in Table A3.

TABLE 1

CALCULATED MEAN FRACTIONAL FREQUENCY DIFFERENCE BETWEEN Cs STANDARDS AND NX–1 MASER (IN PARTS PER 10^{12})

Cs UNITS	(a) $\left(\frac{\Delta f}{f}\right)_{t=0}$	(b) $\delta\left(\frac{\Delta f}{f}\right)_{t=\frac{T}{2}}$	(c) $\left(\frac{\overline{\Delta f}}{f}\right)$	(d) σ
1	6 63	- 046	6 58	1 07
2	6 32	- 281	6 04	1 48
3	5 97	- 047	5 92	0 90
4	3 12	- 758	2 36	1 21
5	4 25	- 556	3 69	1 49
σ_{rms} = 1 25×10^{-12} (RMS FOR ALL FIVE Cs UNITS)				

(a) $\left(\frac{\Delta f}{f}\right)_{t=0}$ IS THE INTERCEPT OF AN LSF LINEAR EQUATION FOR FREQUENCY AS A FUNCTION OF TIME

(b) $\delta\left(\frac{\Delta f}{f}\right)_{t=\frac{T}{2}}$ IS THE AVERAGE FREQUENCY CHANGE DURING THE TOTAL PERIOD OF MEASUREMENT. T=2952 HOURS

(c) $\left(\frac{\overline{\Delta f}}{f}\right)$ IS THE MEAN FRACTIONAL FREQUENCY DIFFERENCE BETWEEN EACH CESIUM STANDARD AND THE NASA HYDROGEN MASER NX-1

(d) σ IS THE STANDARD DEVIATION OF $\frac{\Delta f}{f}$ FROM ITS MEAN $\frac{\overline{\Delta f}}{f}$

TABLE 2

MAGNETIC FIELD AND SYNTHESIZER OFFSETS IN CESIUM STANDARDS AND CONSEQUENT CORRECTIONS

(THE OFF-SET FREQUENCY OF UT-C FOR 1968 RELATIVE TO THE A-1 TIME SCALE IS -300×10^{-10} EXACTLY.)

Cs UNITS	(a) SYNTHESIZER OFFSET (PARTS PER 10^{10})	(b) H-FIELD OFFSET (PARTS PER 10^{10})	(c) δH-FIELD OFFSET (PARTS PER 10^{12})	TOTAL OFFSET (PARTS PER 10^{10})	CORRECTION TO BE APPLIED TO FRACTIONAL FREQUENCY DIFFERENCE (PARTS PER 10^{12})
1	-301 8722	1.860	0.544	-300.0068	0.68
2	-301.8722	1.850	0.419	-300.0180	1.80
3	-301.8722	1.851	0.251	-300.0187	1.87
4	-301.8722	1.847	-0.084	-300.0260	2.60
5	-301 8722	1.849	----	-300.0230	2.30

(a) THE SYNTHESIZER OFFSET IS FURNISHED BY THE MANUFACTURER

(b) THE MAGNETIC FIELD OFFSET IS CALCULATED FROM THE BREIT-RABI FORMULA, i.e $\frac{\Delta f}{f}$ = 427H^2 WHERE H = $\frac{\nu_z}{350\times10^3}$ ν_z BEING THE ZEEMAN FREQUENCY

(c) THE APPLIED MAGNETIC FIELD IN THE TRANSITION REGION BETWEEN THE RAMSEY CAVITIES WAS FOUND TO EXHIBIT A SMALL UPWARD DRIFT, AS SHOWN IN FIGURE 1

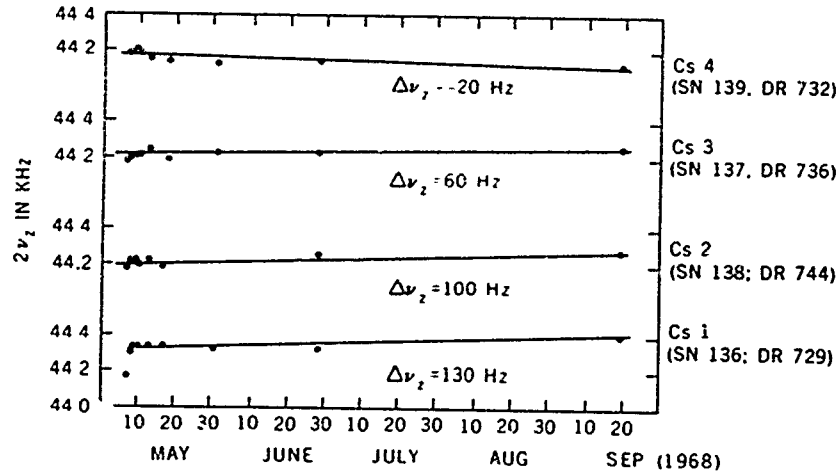


FIGURE 1

MAGNETIC FIELD MEASUREMENTS OF CESIUM STANDARDS

$\Delta\nu_z$ IS THE TOTAL ZEEMAN FREQUENCY CHANGE DUE TO THE MAGNETIC FIELD VARIATION OVER THE 130 DAYS OF FIELD MEASUREMENT; THE SERIAL NUMBER (SN) AND CURRENT CONTROL POTENTIOMETER DIAL READING (DR) ARE GIVEN FOR EACH UNIT.

TABLE 3

CORRECTED MEAN FRACTIONAL FREQUENCY DIFFERENCE BETWEEN CESIUM STANDARDS AND NX–1 MASER (IN PARTS PER 10^{12})

Cs UNITS	$\overline{\frac{\Delta f}{f}}$	$\left(\frac{\Delta f}{f}\right)_c$	$\left(\overline{\frac{\Delta f}{f}}\right)_{Cs\text{-}HM}$	δ
1	6.58	0.68	7.26	0.49
2	6.04	1.80	7.84	1.07
3	5.92	1.87	7.79	1.02
4	2.36	2.60	4.96	-1.81
5	3.69	2.30	5.99	-0.78

$\left(\overline{\frac{\Delta f}{f}}\right)_{Cs\text{-}HM} = 6.77 \times 10^{-12}$ (MEAN FOR ALL FIVE Cs UNITS)

$\delta_{RMS} = 1.11 \times 10^{-12}$ (RMS FOR ALL FIVE Cs UNITS)

TABLE 4

ERRORS IN CESIUM AND HYDROGEN MASER STANDARDS

TYPE OF STANDARD	(a) σ 10^{-12}	SOURCE	COMMENT
Cs (HP5060A-5061A)	1 33	H/P[6]	BASED ON FACTORY TEST OF 150 UNITS RELATIVE TO THE COMPANY'S CESIUM HOUSE STANDARD $3\sigma = 4 \times 10^{-12}$ THE FREQUENCY OF THE H P HOUSE STANDARD IS WITHIN 1×10^{-12} OF THAT OF NBS-III
Cs (NBS III)	1119	NBS[37]	UNDER NORMAL CONDITIONS $3\sigma = 5.6 \times 10^{-2}$ DURING THE 1966 MEASUREMENTS WITH THE VARIAN H-10 MASER THE NBS III RECEIVED SPECIAL CARE AND ITS ERROR WAS ESTIMATED AT $1\sigma = 1.1 \times 10^{-2}$
Cs (NRC III)	75	NRC[8]	BASED ON REEVALUATION OF NRC Cs III $2\sigma = 1.5 \times 10^{-1}$
H MASER (NASA NX-1)	1 0	NASA[9]	BASED ON REPEATED COMPARISONS OF NASA NX 1 AND VARIAN H MASERS AND VANIER AND VESSOT WALL SHIFT CORRECTION
H MASER (NRC)	0 9	NRC[10]	BASED ON INTERCOMPARISON OF TWO HYDROGEN MASERS HM 1 AND HM 2
H MASER (H-10)	0 47	H/P[3]	BASED ON VARIAN H-10 #4

(a) IT SHOULD BE NOTED THAT THE 1σ ERRORS GIVEN IN THIS TABLE ARE NOT DIRECTLY COMPARABLE SINCE THE HEWLETT PACKARD σ ONLY REPRESENTS THE DISPERSION OF A LARGE NUMBER OF PRODUCTION CESIUM STANDARDS WHILE THE OTHERS ARE ESTIMATES BASED ON MEASUREMENTS OF PHYSICAL PARAMETERS WHICH ARE KNOWN TO AFFECT THE FREQUENCY WITH THE ERRORS COMPUTED IN A DIFFERENT WAY IN EACH CASE

TABLE 5

COMPARISON OF RECENT RESULTS OF HYDROGEN MASER FREQUENCY MEASUREMENTS (SINCE 1968) WITH THE VALUE GIVEN BY VESSOT et al. (IN 1966)

(VESSOT et al VALUE = 1,420,405,751.7864 ± 0017 HERTZ)

SOURCE	HYDROGEN MASER FREQUENCY 1.420.405.751 +	IDENTIFICATIONS
GSFC	7767 ± 0024	H/P5060 & 5061 vs NASA NX-1
NRC	7763 ± 0020	NRC HM-2 vs NRC CsIII [a]
LSRH	7782 ± 0036	LSRH H-2 vs EBAUCHES Cs [b] (MODEL B 500SN3)

(a) NRC = NATIONAL RESEARCH COUNCIL, OTTAWA, CANADA

(b) LSRH = SWISS LABORATORY FOR HOROLOGICAL RESEARCH, NEUCHATEL, SWITZERLAND

TABLE A1

CALCULATION OF UNCERTAINTIES IN WALL SHIFT FREQUENCY CORRECTION

SOURCE OF ERROR	FORMULA $\Delta\delta\nu_w$	VANIER - VESSOT	
		FRACTION OF ERROR	$\Delta\delta\nu_w$
D	$\delta\nu_w \frac{\Delta D}{D}$	$\frac{.0104}{7.25}$	.00004
K	$\delta\nu_w \frac{\Delta K}{K}$	$\frac{2}{208}$	.00027
α	$\left(\delta\nu_w - \frac{K}{D}\right)\frac{\Delta\alpha}{\alpha}$	$\frac{1}{5}$	.00012
T	$\left(\delta\nu_w - \frac{K}{D}\right)\frac{\Delta T}{T-T_1}$	$\frac{5}{43}$	.00007
$\overline{\Delta\delta\nu_w} \equiv \sqrt{\Sigma\Delta\delta\nu_w^2}$			.00031

TABLE A2

MAGNETIC FIELD MEASUREMENTS

DATE	ν_i Hz	$\Delta\nu_i = \lvert\nu_i - \bar{\nu}_i\rvert$ Hz
NOV. 14, 1967	6095	4
FEB. 20, 1968	6093	2
JULY 18, 1968	6080	11
NOV 8, 1968	6080	11
DEC 21, 1968	6109	18

$\bar{\nu}_i$ (RMS) = 6091 ± 9.4 Hz

$\bar{H}$ = 4 35 ± 0 007 milligauss

$\delta\nu_H$ = 05205 ± .00016 Hz

TABLE A3

SUMMARY OF FREQUENCY CORRECTIONS

SOURCE	$\partial\nu$ Hz	$\Delta\partial\nu$ Hz
WALL SHIFT	-.02807	±.00031
DOPPLER	- 06221	±.00010
MAGNETIC FIELD	+ .05203	±.00016

$\Sigma\partial\nu$ = -.03823 Hz

$\sqrt{\Sigma\Delta\partial\nu^2}$ = ±.00036 Hz

A NEW PRIMARY CESIUM BEAM FREQUENCY STANDARD

Fred Kupersmith, Clarence Thornburg and John Ho
Atomichron, Inc.
New Hyde Park, New York

SUMMARY

A New Primary Cesium Beam Frequency Standard has been developed utilizing a two loop system. This Standard has all of the usual features, plus several important new features such as the ability to make very small "setability" adjustments or time scale changes without readjusting the C-field.

INTRODUCTION

This paper describes some recent developments in primary standards for use in time and frequency control applications. A New Cesium Beam Frequency Standard has been developed using a Cesium Beam Resonator in a frequency comparison and control system to generate highly stable 5 MHz, 1 MHz and 100 kHz signals.

This Standard, a front view of which is shown in Figure 1, is a primary standard. That is, a standard capable of producing, without reference to any other standard whatever, frequencies of 5 MHz, 1 MHz and 100 kHz accurate to within one part in 10^{11}.

A technical description is presented here which describes a double loop control system which incorporates several new important features previously not available in primary cesium standards, such as:

1. Digital frequency or (Time-Rate) setability with high resolution.
2. Time scale changes without readjustment of the C-field.
3. A modular design concept.

The use of these new features in some typical applications is described, and the operating data available on those standards delivered are presented.

The specifications of this standard are shown in Figure 2. In general, these specifications are typical of a Primary Cesium Beam Standard of this quality, with a few notable exceptions. These are the specifications on "Setability" and "Time Scale". The most important feature of these two quantities, are their precision reproducability, and freedom from the necessity of readjusting the C-field, all of which will be fully discussed below.

TECHNICAL DESCRIPTION

The Two Loop Concept

A Block Diagram of the standard is shown in Figure 3. From this diagram, it can be seen that the standard is divided into two sections, a primary loop, and a secondary loop.

In this technique, the primary loop consists of a quartz oscillator operating at a direct subharmonic of the cesium transition frequency and locked to that atomic resonance. A secondary loop containing a synthesizer is locked to the output of this primary oscillator. It is important to note that the secondary loop has no affect on the frequency reference which has been established by the primary loop. The accuracy ($1:10^{11}$), reproducability ($5:10^{-12}$), long-term stability ($5:10^{-12}$), and short-term stability for averaging times longer than 100 seconds - which covers almost all timing applications, are all determined solely by the primary loop. The secondary loop contains a very high quality low noise quartz oscillator from which a signal is synthesized that is phase-locked to the output of the primary oscillator. It is extremely small frequency steps available in this synthesizer which make several of the innovations possible.

The short-term stability (less than 100 seconds), and the residual noise on the outputs, are a function only of the quartz oscillator used in the secondary loop. The long time constant of the secondary loop eliminates the inherently poor short-term stability of the primary loop.

Since the two loops are separated, one primary loop may be used to operate two or more secondary loops. For example, one primary loop may be used to furnish accurate frequencies referred to the International Second through one secondary loop, while a second secondary loop furnishes signals to a timing generator to generate 1 pps signals referred to the UTC Time Scale.

Cesium Beam Tube

The Cesium Beam Tube is a militari-

zed 20" tube, designed and tested to MIL-E-5400. The tube has a Ramsey resonance linewidth of 500 Hz and incorporates a microwave structure that does not require temperature compensated matching over the range of 0°C to 60°C. The tube exhibits a minimum signal-to-noise ratio of 45 dB in a 3 Hz bandwidth, which is equivalent to a primary loop short-term stability of 5×10^{-11} RMS for 1 second averages. The short-term stability of the outputs is much better than this number because of the long secondary loop time constant and the excellent short-term stability of the secondary oscillator. The tube has a guaranteed operating lifetime of three years.

Primary Loop

The operation of this loop (Figure 4), is typical of those used in Cesium Standards and has been described in many previous articles. Several features of this loop are worth noting, however.

First, it utilizes square wave frequency modulation obtained by applying the modulating signal to a separate voltage control circuit of the primary oscillator.

Secondly, the loop time constant of the primary loop is short (less than 1/2 second). In other words, the primary oscillator is tightly locked to the atomic resonance. And finally, the primary loop has sufficient DC loop gain to reduce the effect of three years aging of the primary oscillator to 1×10^{-12}.

Secondary Loop

The secondary loop (Figure 5), consists of a very high quality low noise quartz oscillator from which a signal is synthesized that is phase-locked to the output of the primary oscillator. The synthesizer uses the rate multiplier technique in which the pulse trains going into the OR-gate combiner are selected by means of thumbwheel switches, the "TIME-RATE SELECTOR". A unit change in the least significant thumbwheel corresponds to a change of 5.580875×10^{-13} in the output frequency.

In order to realize the excellent short-term stability of the secondary quartz oscillator, the secondary loop time constant is quite long (about 200 seconds).

SPECIAL FEATURES

Setability

The largest recent advance in precision timekeeping has been made in the development of a synthesizer which digitally steps the output frequency in such small steps that the time-rate may be set to any reference within less than 25 nanoseconds per day. This setability corresponds to setting the frequency of the Cesium Beam Standard to any other reference to within 3 parts in 10^{13}. The frequency can be offset while the unit is operating, with no ill effects on the output of the standard.

This offset is accomplished in practice by adjusting the "TIME-RATE SELECTOR" to the desired number. A total range of more than $\pm 1 \times 10^{-7}$ (from frequencies referred to the atomic time scale) is available. Furthermore, the offset is an integral number of discrete known steps and therefore, is not only accurately known, but can be reproduced or removed with precision.

A most important feature of this setability is that it requires no C-field readjustment. By necessity, the C-region must be magnetically shielded from ambient magnetic fields which, in practice, is accomplished by surrounding the C-region with envelopes of magnetic shielding material until the remaining field is essentially zero. Then a known field is superimposed on the region by means of passing a carefully controlled current through wires situated in the C-region. This C-field is modified in aspect and value by the surrounding shields. Because the magnetic shielding material has magnetic hysteresis and because magnetic domain orientation is essentially a thermodynamic process, the C-field exhibits both non-reversability and time-dependent (aging) effects. The ideal situation is obtained by setting the C-field once, letting it age well, and never adjusting it again. This situation has been made possible by restricting all fine frequency adjustments to the secondary loop, the primary loop (and C-field) stay fixed.

Example: At the present time, the East Coast Loran-C chain is losing about 50 nanoseconds per day. After monitoring the chain for sufficient time to establish the time-rate exactly, the master station could be directed to change all

of their clocks by decreasing the "Time Rate Selector" one unit, and every clock would increase its time-rate by precisely 48.21876 nanoseconds/day. The whole transaction could be completed remotely by telephone with full knowledge by the directing authority of the exact results of their command.

If this adjustment were made by changing the C-field of each clock, a month would be required for the C-fields to "settle down". Then several weeks data would be required to establish exactly the new offset for each clock.

Time Scale Changes

Since the total range is $\pm 1 \times 10^{-7}$, time scale changes can be made with the time-rate selector switches. To a precision system containing many cesium standards, a time scale change formerly was a small disaster. The C-fields of every clock would have to be changed requiring a "settling down" period as described above. Then a recalibration period would be required during which the exact new offset of each clock was determined.

Using the technique described above, relative calibration is maintained from one time scale to another since the C-field remains unchanged.

1. If you know the absolute accuracy of a standard on one time scale and the time scale changes, you will know accuracy with the same precision after the adjustment.

2. If a system has a group of clocks (perhaps in different geographical locations) and the time scale changes, the clocks will have exactly the same relative accuracy after adjustment as before.

Short-Term Stability

The two loop approach makes possible the elimination of the inherently poor short-term stability (for averaging times less than about 100 seconds), of the primary loop. A high quality output oscillator together with a long time constant in the secondary loop combine to produce an excellent short-term stability. This standard exhibits a typical short-term stability of 3×10^{-12} for 1 second averages.

C-Field Provisions

The C-field is supplied by ovenized constant current source. This supply is good to .075% over the temperature range from 0°C to 50°C and has an aging characteristic better than .01% per year. For reference .1% corresponds to a change in the output frequency of 3.3×10^{-13}. In order to calibrate the C-field, a built-in Zeeman frequency source is provided. No additional equipment (audio oscillators, counters, etc.) is required to set-up this standard.

PERFORMANCE DATA

The performance data of the first ten standards delivered is as follows:

(a) Accuracy:

Average	-0.1×10^{-11}
Standard Deviation:	0.35×10^{-11}

(b) Short-Term Stability (1 Second):

Average	3.1×10^{-12} RMS
Total Spread:	2.5×10^{-12} (Min.) to 4.0×10^{-12} (Max.)

(c) One Hour RMS Stability:

Average:	7.4×10^{-13} RMS
Total Spread:	6.0×10^{-13} (Min.) to 1.0×10^{-12} (Max.)

It is probable that a large portion of the spread in the measured accuracy can be attributed to the reference system which utilized VLF transmissions.

CONSTRUCTION

The Standard is ruggedly constructed to withstand extreme vibration and shock such as may be encountered in military and industrial applications. An aluminum housing combined with a cast front panel provide the proper combination for a dependable yet lightweight instrument. A photograph of the general layout is shown in Figure 6.

All of the major functional units in the Standard are plug-in modules. Having the unit divided into plug-in modules serves two major functions. First, repair of the Standard is reduced to the lowest echelon with a modular construction of simple design permitting rapid replacement of major modules with a minimum of down-time. Second, only those modules which are necessary to the intended use need be supplied with the equipment. Others can be added later and simply plugged-in by the user.

OPTIONS

The following items are furnished as optional equipment with the Standard:

1. Time Standard Module (shown in Figure 1). This optional plug-in

module provides:

a. 1 pps clock pulse output.
b. Automatic synchronization.
c. Direct-reading thumbwheel time delay.
d. 24-hour clock with stepping sweep second hand.

2). Internal 1-Hour Battery.
3). 8-Hour Battery Pack.
4). Inverter for converting 12 VDC to 115 VAC permitting operation of the Standard and charging of the 8-Hour Battery Pack.
5). Miscellaneous other options such as a 100 MHz output for easy comparisons.

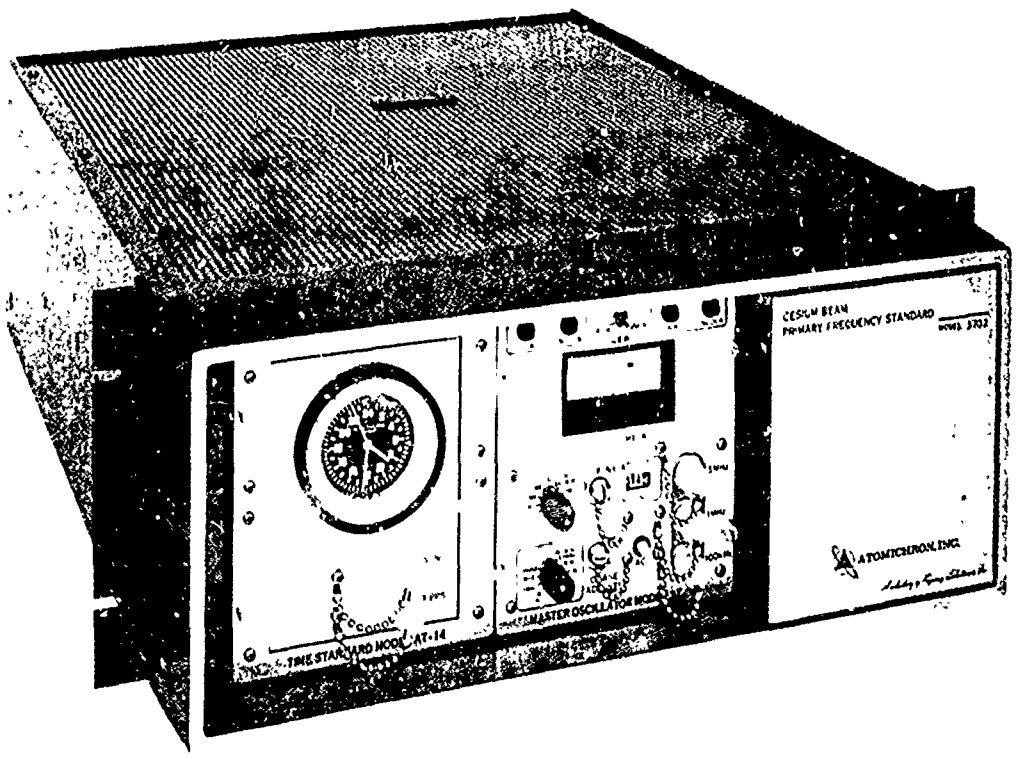


Figure 1 -- Front View, Cesium Beam Primary Frequency Standard

FIGURE 2. -- TECHNICAL SPECIFICATIONS

ACCURACY:	$\pm$ 1 X 10^{-11}
REPRODUCABILITY:	$\pm$ 5 X 10^{-12}
SETTABILITY (Freq.):	Within 3 X 10^{-13} with no C-field readjustment. $\pm$ 1 X 10^{-7} total range.
SETTABILITY (Time-Rate):	Within 25 nanoseconds per day with no C-field readjustment.
LONG-TERM STABILITY:	$\pm$ 5 X 10^{-12} for life of Cesium Beam Tube.
SHORT-TERM STABILITY:	1 X 10^{-11} RMS for 1 second averages. 3 X 10^{-12} RMS for 1000 second averages.
OUTPUT SIGNALS:	
Frequencies:	5 MHz, 1 MHz, 100 kHz.
Voltage:	1V RMS into 50 ohms.
Harmonic Distortion:	-40 dB
Non-Harmonic Distortion:	-80 dB
Terminals:	Type BNC Connectors
TIME SCALE:	Adjustable with thumbwheel switches from zero to -1000 X 10^{-10}. No C-field readjustment is required. Relative calibration is maintained from one time scale to another. Accuracy of time scale offset is $\pm$ 3 X 10^{-13}.
CESIUM BEAM TUBE LIFE:	3 years guaranteed
ENVIRONMENTAL:	
Temperature:	$\pm$ 5 X 10^{-12} over 0°C to 50°C.
Humidity:	95%.
Magnetic:	$\pm$ 1 X 10^{-11}, 2 gauss field, any orientation.
Vibration:	MIL-STD-167.
POWER:	
Voltage Required:	115 or 230 VAC $\pm$ 10%, 50 to 400 Hz, or 22 to 30 volts DC.
Power Consumption:	75W @ 25 VDC or 90W @ 115 VAC.
SIZE:	7" High, Fits 19" Rack. Requires 22" Depth behind panel (includes connectors).
WEIGHT:	62½ Pounds (without Options).

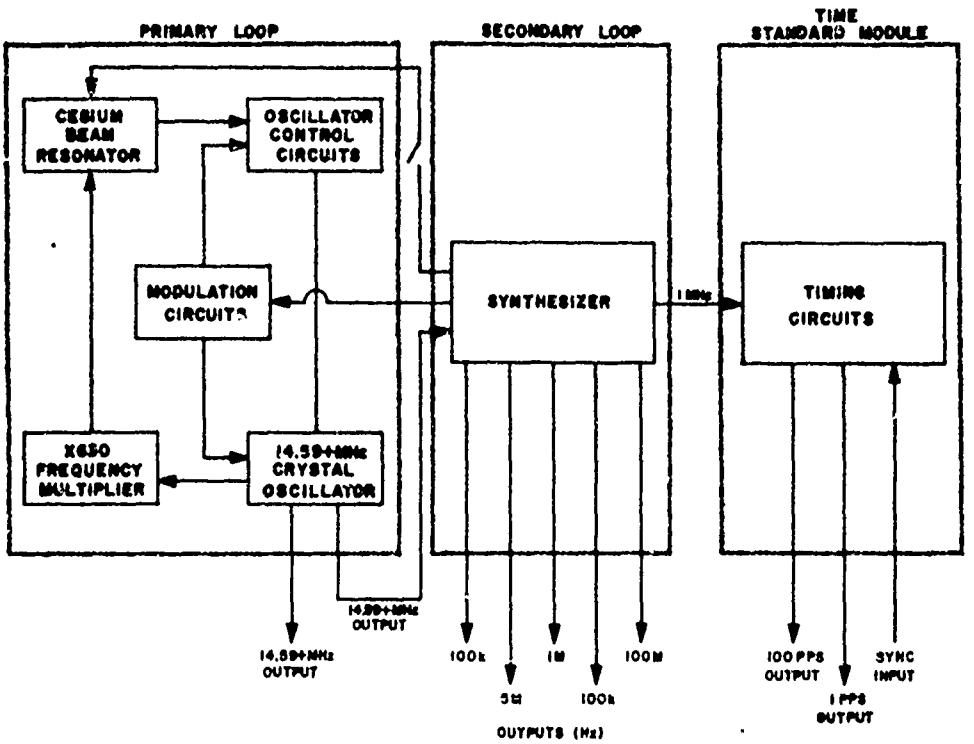


Figure 3 -- Simplified Block Diagram, [illegible] Standard

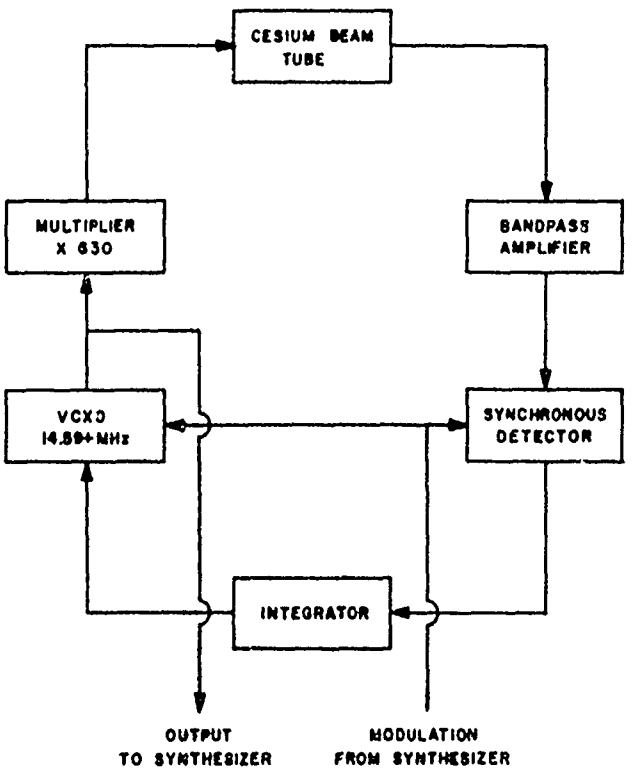


Figure 4 -- Block Diagram, Primary Loop

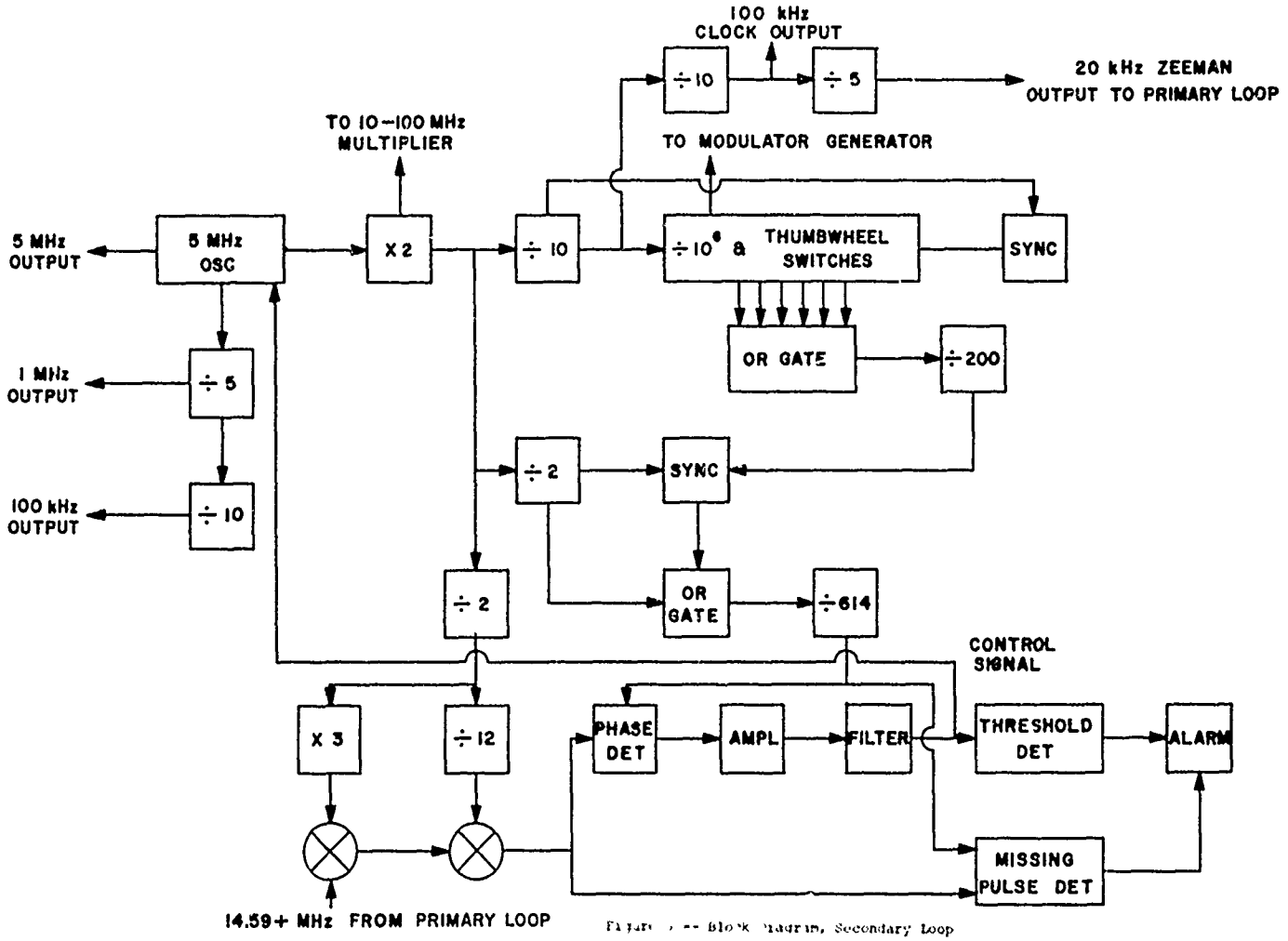


Figure 5 -- Block Diagram, Secondary Loop

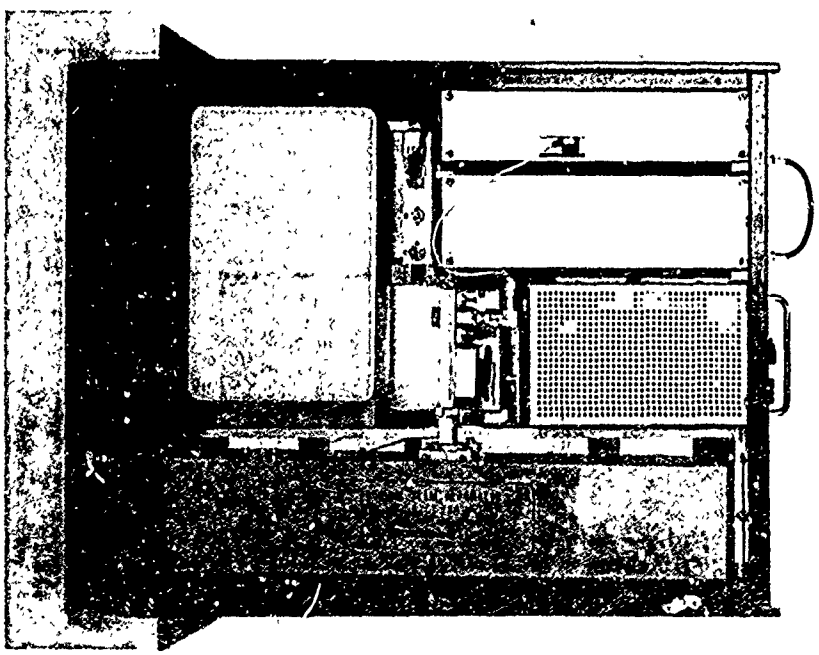


Figure 6 -- General Layout.

TIME FREQUENCY TECHNOLOGY IN SYSTEM DEVELOPMENT

Robert E. Perkinson
McDonnell Douglas Astronautics Company - Eastern Division
St. Louis, Missouri

Summary

For the last nine years, McDonnell Douglas has been working on what has come to be known as time/frequency technology. This technology is based on very precise synchronization of time in moving stations -- time synchronization so precise that range measurements can be made essentially in a one-way sense, and frequency synchronization so accurate that closing rate can be measured by doppler shift in a one-way sense also. This system, of course, imposes a requirement for very exact frequency references and very exact synchronization and control of time.

The use of "one-way" techniques is a very powerful communication technique because it allows a "broadcast" mode of transmission; i.e., a single transmission gives range, range rate and digital data to all airplanes and ground stations within receiving range.

The synchronization process which permits this system is well established. The accuracy in frequency and time keeping which is required by the time/frequency approach is attainable by good quality crystal oscillators for systems where resynchronization is readily available and by atomic resonance frequency sources where time must be held for periods of hours to days. For the first time a real requirement is imposed for synchronizing remote stations to accuracies of approximately .1 microsecond. Several techniques are available which can achieve this, but much work is required before we can maintain this accuracy on a routine basis.

The last 10 years have seen a quiet revolution come through the technology of communications, navigation and identification. The development of very precise frequency standards and the simultaneous development of integrated circuit logic elements which allow precise counting at very high rates have enabled us to divide time to an astounding precision and to hold this precise time over very long periods. (Figure 1). This audience knows well these fantastic accomplishments in precise time and frequency keeping, since they have been responsible for many of these advancements. I would like to address myself to some of the practical applications of this precise time and frequency capability which led to new approaches to some of the old problems of communicating and navigating.

One of the problems that has plagued the aircraft community for many years is the mid-air collision. The first recorded mid-air collision was in 1910 in Milan, Italy (Figure 2). At the time, there were approximately 50 airplanes in existence. The collision problem has continued to haunt the flying community and becomes ever more serious as the density of the airplane traffic increases.

One would think this is a problem that could be easily solved with today's technology. Almost any sort of radar transponder could be used to provide protection for one airplane against one other airplane. The real requirement is to provide continuous protection to an airplane in a large community of aircraft and mutually provide this same protection to all other aircraft. The basic constraint is a communication problem -- a requirement that each of the airplanes obtain information on the range, altitude, and range rate of all the other airplanes in its vicinity. Trying to do this with normal interrogate/respond techniques results in a very limited capacity. With the use of very precise time and frequency synchronization, this capacity problem has been overcome, and literally thousands of aircraft can be accommodated on a common frequency.

This increase in capacity is accomplished by assigning a specific time or "message slot" to a particular unit knowing that that unit can reliably keep its transmissions in that alotted time. Double-slot occupancy can be guarded against by monitoring by an external agency, or, as is done in the time/frequency collision avoidance system (CAS), the equipment can be made to periodically inhibit transmission and listen for any other transmission in its time slot. If other transmissions are heard, the time slot is moved to the next available unoccupied time.

In the CAS system we have the capability of accommodating 2,000 1-1/2 millisecond slots in a three-second time interval without interference.

A guaranteed interference-free message slot allows the use of long signals which allow direct doppler measurement on a single pulse and also permits 100 or so bits of digital data. (See Figures 3 and 4).

By knowing exactly when a signal is transmitted from an airplane, a one-way measurement of the elapsed time from transmission to reception of the signal can be made. This of course is a measurement of range. The advantage to measuring range in this manner is that a single transmission permits range measurements to be made by all other stations listening to the signal. As a matter of fact, it allows measurement of range by an unlimited number of passive stations. This is the first system where all of the participants can measure range to all of the other participants. In addition, range rate can be determined "one way" (as it is in the CAS system) by measuring the difference in frequency between the received signal and the onboard frequency source.

The sync technique is quite simple. The aircraft are aligned in coarse synchronization (for message slot purposes) by the master station transmitting a coded message that indicates the cycle start. This is used to start a listening member's counter and the member then counts down to his message slot and transmits his normal signal. (See Figure 5). In the "fine synchronization" process, the synchronizing station detects the normal transmissions of a station and notes apparent propagation time. If the transmitting station has a time different from the synchronizing station, this apparent propagation time will be:

$$T_{Prop_{App}} = T_{Prop_{True}} \pm T_{Error} \quad \text{(See Figure 6)}$$

The total apparent propagation time is subtracted from a fixed "synchronization time" toward the end of the message slot, and the synchronization station transmits a sync pulse at this time:

$$T_{Sync} - T_{Prop_{App}} = T_{Sync} - \left(T_{Prop_{True}} \pm T_{Error}\right)$$

Since the sync pulse is transmitted within a few milliseconds of the range pulse, the propagation time is essentially unchanged and the sync pulse arrives at the airplane propagation time later at:

$$T_{Sync} - T_{Prop_{App}} + T_{Prop_{True}} = T_{Sync} \mp T_{Error}$$

Thus in the airplane being synchronized, the pulse arrives early or late with respect to sync time by the amount the airplane clock is late or early with respect to the synchronizing clock. This is recognized, and a digital correction made to make the airplane time agree with the sychronization station time. The airplane can recognize how early or late it is with respect to master time, and since the only thing that can cause a constant early (or late) time error is a difference in frequency between the synchronizer and the synchee, this information can be used to pull the airplane oscillator on to the same frequency as the synchronizing station.

This technique and the associated hardware are well established. We have been using it daily in an operational collision avoidance system in St. Louis since 1956. (Figure 7). Typical synchronization accuracies are shown in the histogram. (Figure 8). The accuracies achievable are dependent on the oscillator frequency accuracy, digital count resolution and pulse arrival time accuracy. The accuracies shown in Figure 8 are typical of the CAS system. Much better accuracies have been achieved in military systems where precise ranging is important.

The system we have been using in St. Louis since 1965 uses a local time which is provided from a master station located on the ground. Exact time is by this means readily provided to all the aircraft within line of sight range of the ground station. An advanced version of this equipment, which is designed to have very much less dependence upon ground stations, has been developed for the Air Transport Association. (References 1, 2 and 3). The flight demonstration of this equipment to the airlines was completed last fall. (Reference 4). With this system, airplane to airplane time synchronization is provided. It is easy to synchronize one airplane from another, the only difficulty involved in synchronizing a group of airplanes is in deciding on a single source of time so that all aircraft can be synchronized to the same time and to the same frequency.

In the ATA system, this is done by setting up a hierarchy of time. Ground stations are designated as "hierarchy zero", and it is assumed that all ground stations are synchronized to each other to less than 1 microsecond. Airplanes which are being synchronized by "hierarchy one" airplanes are "hierarchy two", and so forth out to "hierarchy 40". The hierarchy of the airplanes is announced within the airplane's transmission by means of a digital code. The aircraft desiring a synchronization, then, looks for an airplane which has a better time and requests a synchronization from that airplane over the digital link and is subsequently synchronized.

In addition to this "chaining", the systems also demote in hierarchy as a function of time and quality of the onboard frequency source. If an airplane has not been synchronized for a period of time, the system hierarchy will be demoted at an interval based on the quality of its frequency source. Table 1 shows the demotion rate as a function of oscillator quality in terms of total time to demote through the 40 steps of the hierarchy chain. It should be noted that in the time/frequency systems we have been describing, this is the only item which is affected by quality

of the oscillator. When synchronization is available, ranging accuracy and range rate accuracy are essentially unaffected by oscillator quality as long as the oscillator is stable within one part 10^8. Good quality temperature-stabilized crystal oscillators can achieve this stability within 25 to 30 minutes after turn-on; and the frequency synchronization techniques which we use in conjunction with the time synchronization will allow us to pull an oscillator into frequency agreement with the master oscillator to an accuracy of one part in 10^8 within 2 to 3 minutes.

Some airlines have indicated that they intend to carry cesium beam standards on their CAS-equipped airplanes. This is particularly true for the airlines which have overseas operations so that they can maintain correct time on long, over-water flights. At least one equipment manufacturer (Hewlett-Packard) has announced intentions of building a cesium beam standard in a 1/2-ATR rack size for this application.

The time hierarchy approach imposes a requirement for very precise time synchronization between ground stations. Ground stations should be synchronized to one another to accuracies better than 1 microsecond. The highly accurate frequencies available from atomic resonance references provide a solid foundation for maintaining correct time after synchronization is achieved.

Achieving the correct time synchronization will be greatly facilitated by the atomic resonance oscillator-equipped airplanes with their ground to air and air to ground synchronization capabilities. Airplanes with these capabilities will be flying across the country almost continuously and will be making many many trips across the oceans and around the world on a daily basis. Since these airplanes will be capable of synchronizing any ground station in line of sight as they fly over, the CAS system itself should add tremendously to the capability of obtaining initial time synchronization. These capabilities were first demonstrated in a joint U.S. Naval Observatory/McDonnell experiment conducted on 6 August 1966. In his report, Dr. William Markowitz concluded that: "Any two ground stations on the earth could be synchronized to 0.1 microseconds by flying an atomic clock between the stations in an aircraft which need not land at either station." Even so, it would be highly desirable to be able to monitor clocks on the ground on a continuous basis and check them against some master time standard.

One technique we have been using with some success for this purpose is to monitor the LORAN-C transmissions. During the Test and Evaluation program in the summer and fall last year in Baltimore, we monitored the Cape Fear LORAN-C transmissions and compared them to an atomic clock, which had been set to the U. S. time standard at the Naval Observatory. The results were quite satisfactory. We are continuing to monitor the LORAN-C transmissions since we have returned to St. Louis and have had similar results in tracking the transmissions from the Dana, Indiana, station. It appears that the LORAN-C tracking technique can be used to good effect for monitoring the time accuracy of atomic resonance sources at the ground stations, particularly if special care is exercised to insure that no net time shifts are allowed to occur in the transmitted LORAN-C signals. The experiments that have been conducted in Europe using television signals as a timing reference look very interesting, and it is my understanding that similar experiments are being conducted by the U.S. Time and Frequency Service group in Boulder, Colorado. Perhaps ultimately we can obtain exact time synchronization at the same time we are watching the Late, Late Show. Certainly a synchronization system that would allow an excuse for monitoring the Midnight Movie on the second shift should obtain broad acceptance at the technician level.

The technique of ranging by exact time synchronization is useful for many applications in addition to CAS. In our St. Louis operation we monitor the CAS transmissions at our ground station measuring range by the time-ranging technique described previously and measuring bearing by a phased array antenna. The resulting information is displayed on a PPI scope (Figure 9). The ground station has the capability to monitor 1,000 airplanes with a 2-second update rate and is completely free of fruit and overlap problems. Any number of ground stations can monitor these signals without interference because the ground monitors do not interrogate but merely listen.

In a time synchronized system, range can be read to all units, from all units. This means that all the CAS equipped airplanes can read range to a ground station and/or to another airplane, thus it can be used as a DME and/or as an air to air station keeper. (Figure 10). As a matter of fact, multiple range measurements can be made simultaneously, which allows the ready accomplishment of multilateration navigation and/or surveillance systems. The time/frequency technique is a new, powerful tool which is just beginning to be recognized as an effective answer to many of the problems which plague present interrogate/respond systems.

I have tried to call attention in this paper to a requirement for very exact time synchronization at ground stations scattered across the United States and ultimately around the world. The synchronization accuracies that are needed are greater than are required for present navigation or time keeping purposes. This exact time will be most immediately useful for the ATA collision avoidance system but will ultimately form the basis for a new family of navigation, identification and digital communication systems.

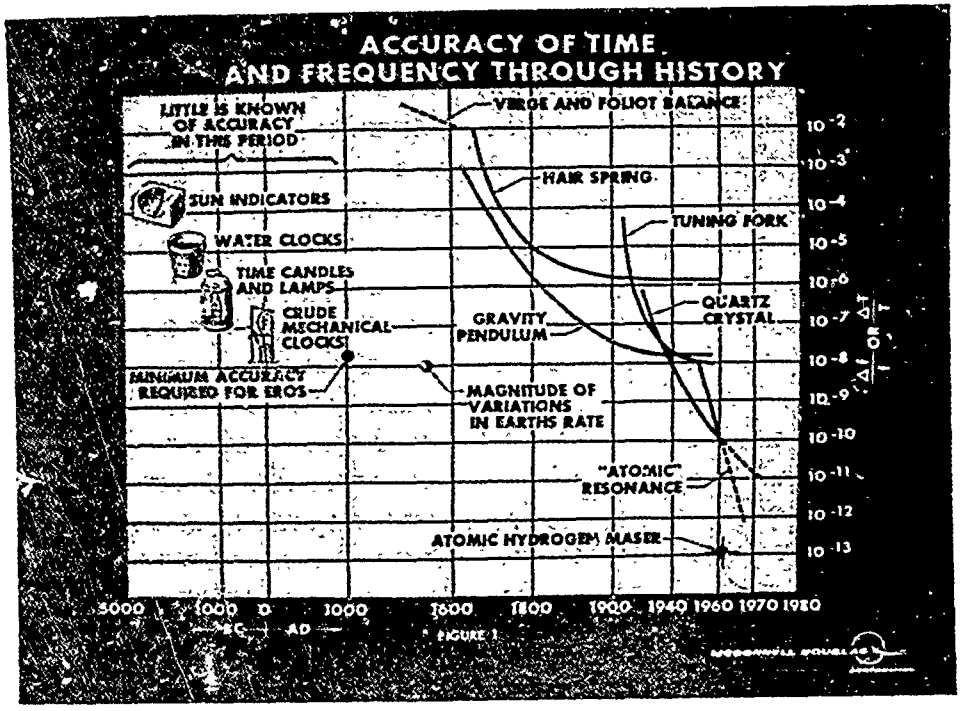
REFERENCES

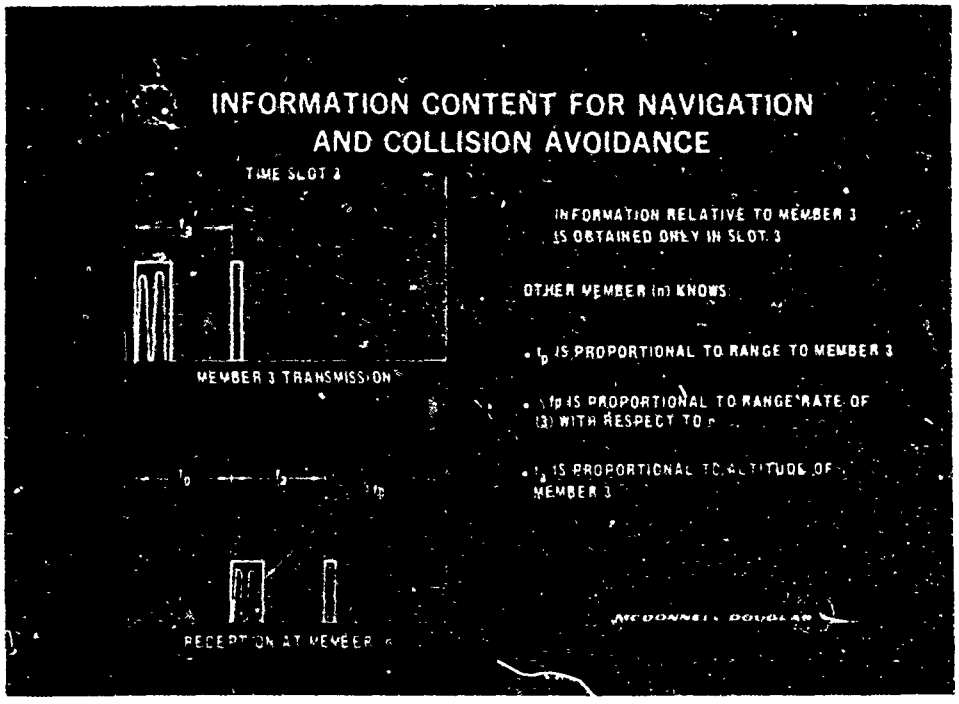
1. "Airborne Collision Avoidance System - Statement of Airline Policy and Requirements and a Technical Description of the System", Airline Air Traffic Control Division, Air Transport Association of America, ANTC Report No. 117, Revision 9.

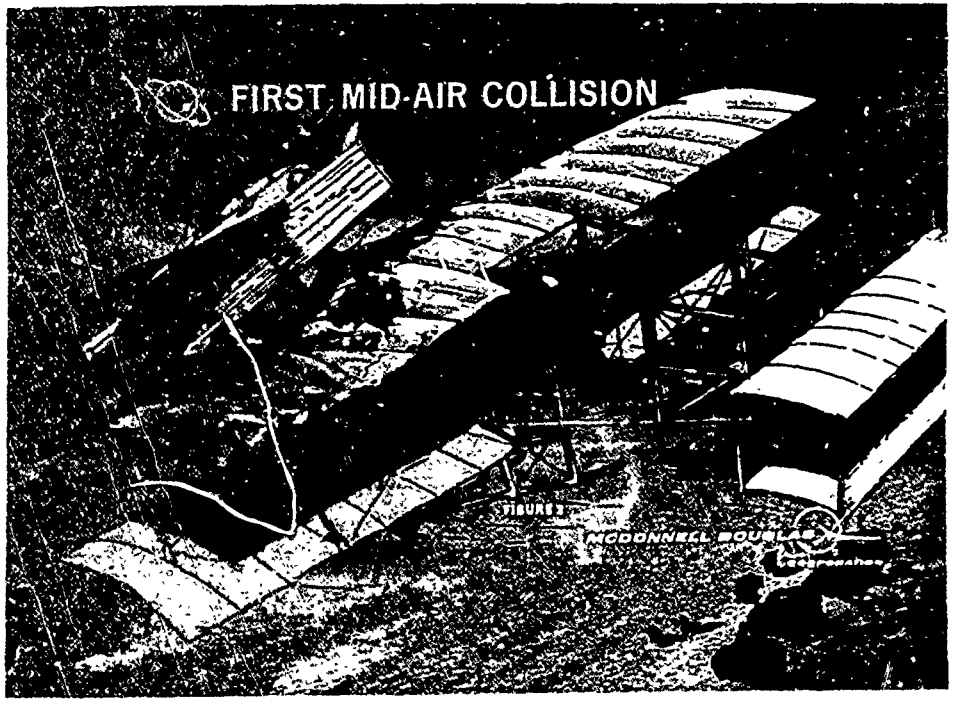
2. "Air Transport Collision Avoidance System", Aeronautical Radio Inc. Project Paper No. 587, Draft No. 2, 24 February 1970.

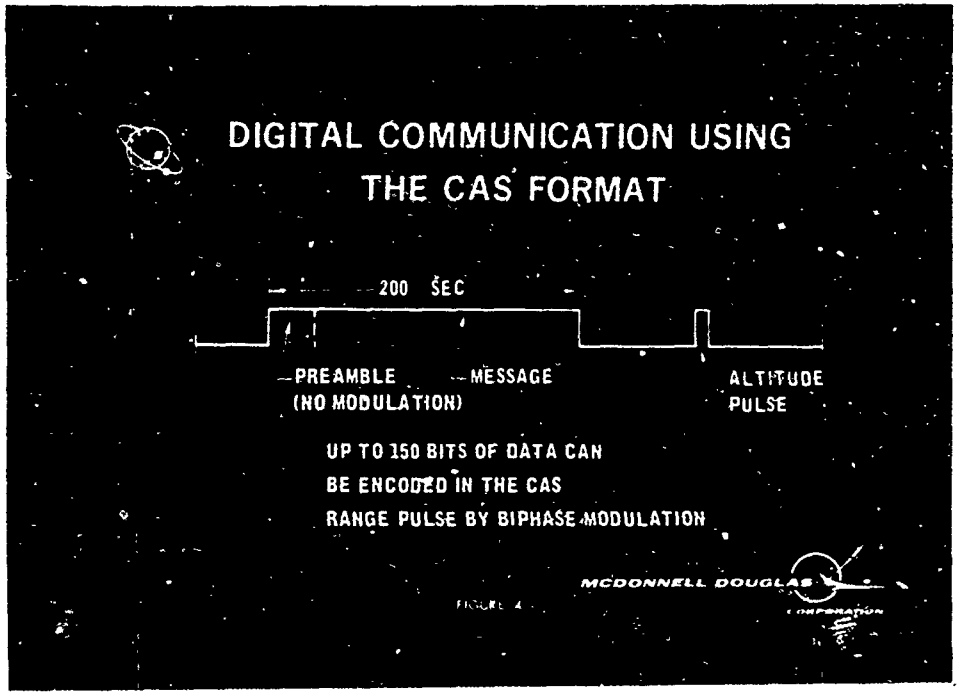
3. "Collection of Papers on Time Frequency CAS," IEEE Transactions on Aerospace and Electronics Systems, Volume AES-4, Number 2, March 1968.

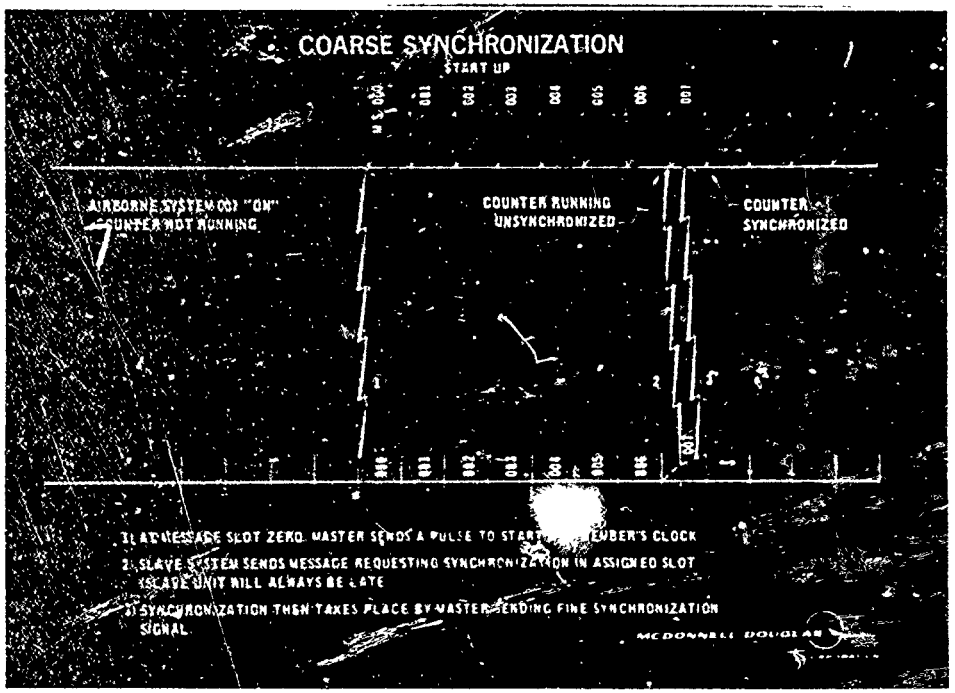
4. "Flight Test and Evaluation of Airborne Collision Avoidance System Defined by ATA/ANTC Report 117," by Martin Marietta Company for the Air Transport Association, 20 March 1970.

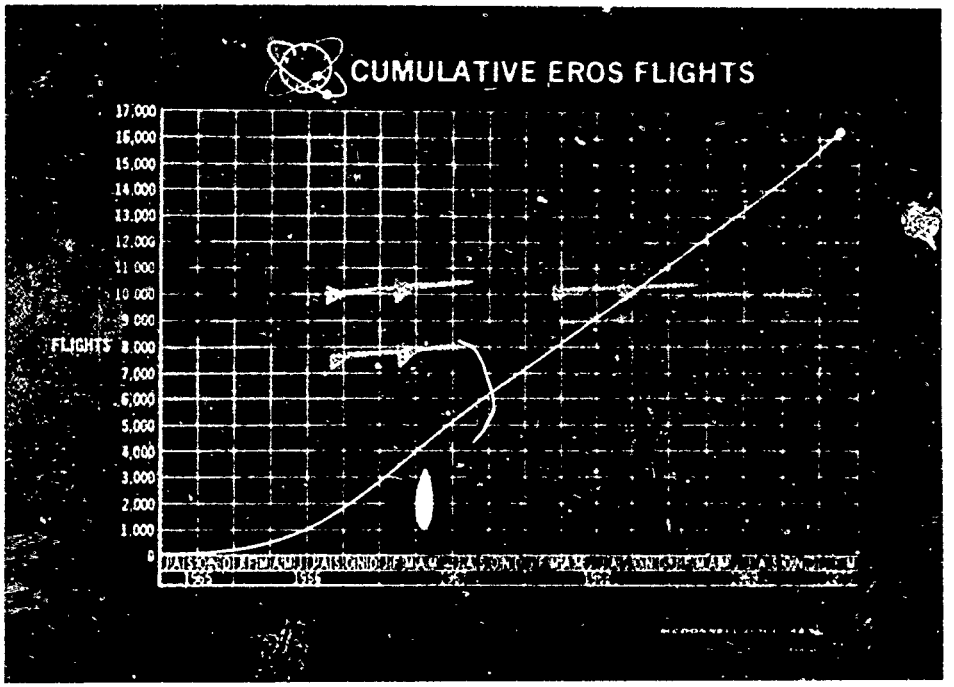

ACCURACY OF TIME AND FREQUENCY THROUGH HISTORY
LITTLE IS KNOWN OF ACCURACY IN THIS PERIOD
VERGE AND FOLIOT BALANCE
HAIR SPRING
TUNING FORK
SUN INDICATORS
WATER CLOCKS
TIME CANDLES AND LAMPS
CRUDE MECHANICAL CLOCKS
GRAVITY PENDULUM
QUARTZ CRYSTAL
MINIMUM ACCURACY REQUIRED FOR EROS
MAGNITUDE OF VARIATIONS IN EARTHS RATE
"ATOMIC" RESONANCE
ATOMIC HYDROGEN MASER
10 -2
10 -3
10 -4
10 -5
10 -6
10 -7
10 -8
10 -9
10 -10
10 -11
10 -12
10 -13
Δf OR ΔT / T
3000
1000
0
1000
1600
1800
1900
1940
1960
1970
1980
BC
AD
FIGURE 1
McDONNELL DOUGLAS

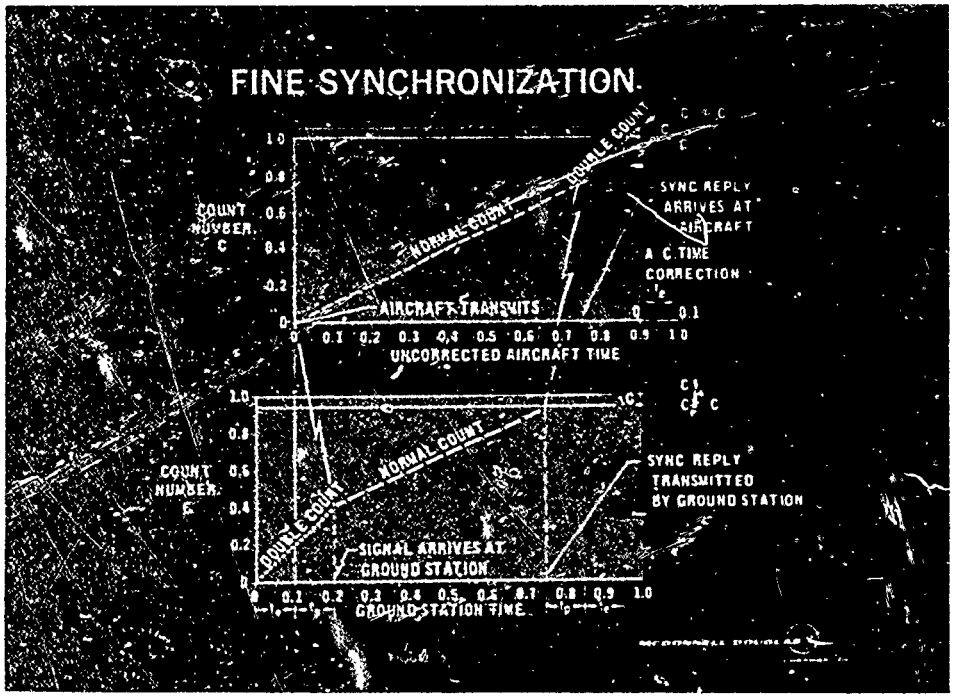

INFORMATION CONTENT FOR NAVIGATION AND COLLISION AVOIDANCE
TIME SLOT 3
MEMBER 3 TRANSMISSION
INFORMATION RELATIVE TO MEMBER 3 IS OBTAINED ONLY IN SLOT 3
OTHER MEMBER (n) KNOWS
• t_b IS PROPORTIONAL TO RANGE TO MEMBER 3
• t_p IS PROPORTIONAL TO RANGE RATE OF (3) WITH RESPECT TO
• t_a IS PROPORTIONAL TO ALTITUDE OF MEMBER 3
RECEPTION AT MEMBER
McDONNELL DOUGLAS

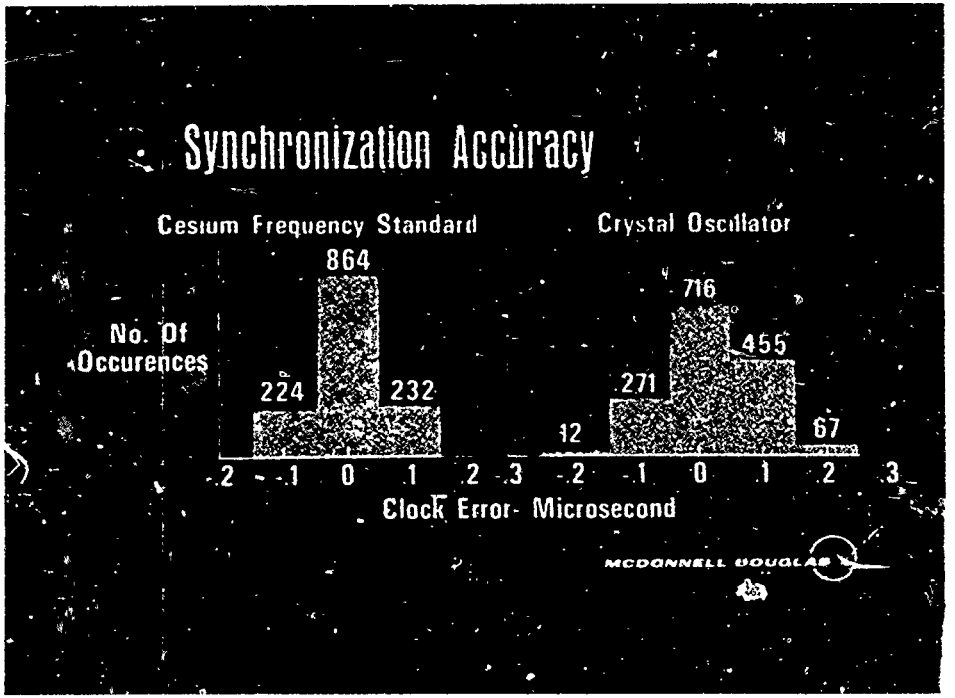

FIRST MID-AIR COLLISION
McDONNELL DOUGLAS


DIGITAL COMMUNICATION USING THE CAS FORMAT
200 SEC
PREAMBLE (NO MODULATION)
MESSAGE
ALTITUDE PULSE
UP TO 150 BITS OF DATA CAN BE ENCODED IN THE CAS RANGE PULSE BY BIPHASE MODULATION
McDONNELL DOUGLAS


COARSE SYNCHRONIZATION
START UP
AIRBORNE SYSTEM (S) "ON"
COUNTER NOT RUNNING
COUNTER RUNNING
UNSYNCHRONIZED
COUNTER
SYNCHRONIZED
1) AT MESSAGE SLOT ZERO, MASTER SENDS A PULSE TO START
2) SLAVE SYSTEM SENDS MESSAGE REQUESTING SYNCHRONIZATION IN ASSIGNED SLOT
3) SYNCHRONIZATION THEN TAKES PLACE BY MASTER SENDING FINE SYNCHRONIZATION
SIGNAL
MCDONNELL DOUGLAS


CUMULATIVE EROS FLIGHTS
FLIGHTS


FINE SYNCHRONIZATION
COUNT NUMBER C
NORMAL COUNT
DOUBLE COUNT
SYNC REPLY ARRIVES AT AIRCRAFT
A C TIME CORRECTION
AIRCRAFT TRANSMITS
UNCORRECTED AIRCRAFT TIME
SYNC REPLY TRANSMITTED BY GROUND STATION
SIGNAL ARRIVES AT GROUND STATION
GROUND STATION TIME

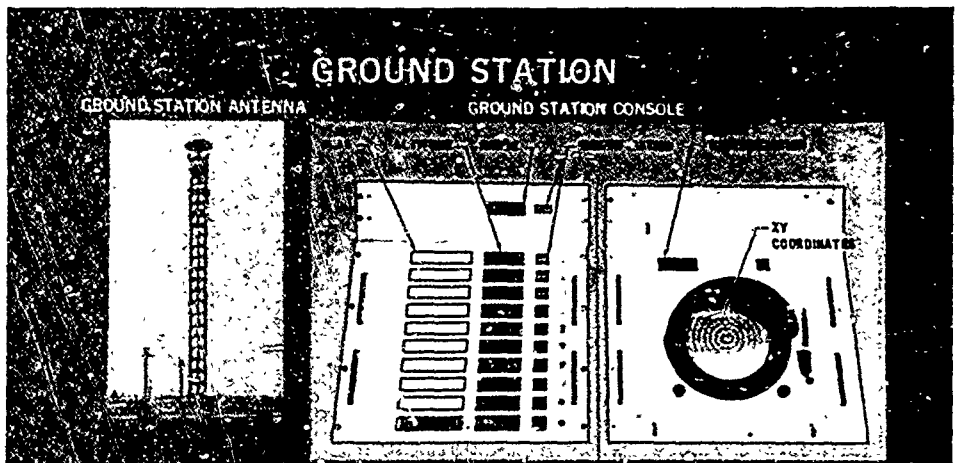

Synchronization Accuracy
Cesium Frequency Standard
Crystal Oscillator
No. Of Occurences
224
864
232
12
271
716
455
67
-.2 -.1 0 .1 .2 .3
-.2 -.1 0 .1 .2 .3
Clock Error- Microsecond
MCDONNELL DOUGLAS

HIERARCHY DEMOTION RATES

OSCILLATOR FREQUENCY OFFSET	TIME PER HIERARCHY	TIME IN SYNC MODE
1×10^{-8} (CRYSTAL)	5.0 SEC	3.3 MIN
1×10^{-9} (CRYSTAL)	50.0 SEC	33.3 MIN
5×10^{-11} (CESIUM)	16.7 MIN	11.1 HR
1×10^{-11} (CESIUM)	1.4 HR	55.6 HR

BASED ON COUNT DOWN FROM HIERARCHY #1 WITHOUT RESYNCHRONIZATION

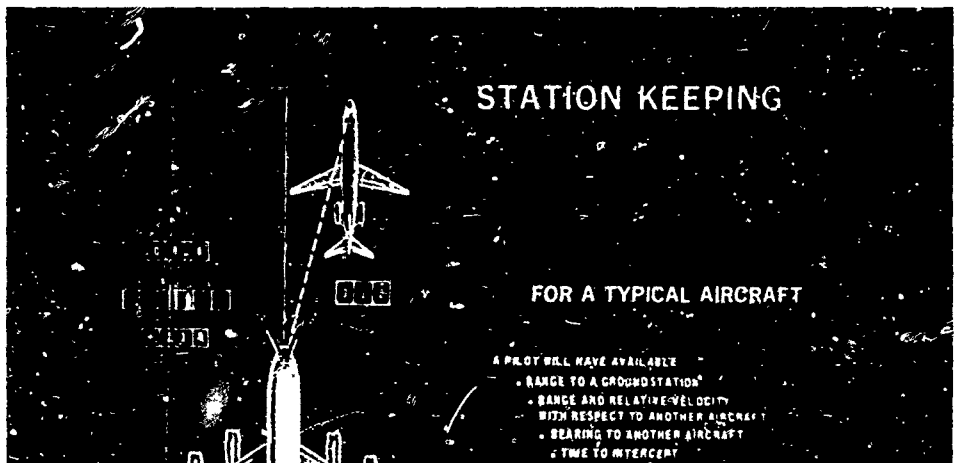
GROUND STATION



FOR A TYPICAL EROS GROUND STATION

A CONTROLLER WILL HAVE AVAILABLE
- RANGE AND BEARING FROM THE STATION OF ANY AIRCRAFT WITHIN COVERAGE
- ALTITUDE OF THE AIRCRAFT
- IDENTIFICATION OF THE AIRCRAFT
- XY COORDINATES
- SIGNAL STRENGTH POWER INFORMATION

STATION KEEPING



FOR A TYPICAL AIRCRAFT

A PILOT WILL HAVE AVAILABLE
- RANGE TO A GROUND STATION
- RANGE AND RELATIVE VELOCITY WITH RESPECT TO ANOTHER AIRCRAFT
- BEARING TO ANOTHER AIRCRAFT
- TIME TO INTERCEPT

A SURVEY OF TIME AND FREQUENCY DISSEMINATION TECHNIQUES

J. L. Jespersen
Frequency-Time Dissemination Research Section
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80302

There are a number of common elements which characterize most time and frequency dissemination systems. A simple VLF timing system can be used to illustrate these characteristics. First, suppose there is a sinusoidal signal available at 10.0 kHz and that no other signals are available. A certain positive going zero-crossing of this signal leaving the transmitter at 0000 UT will reach the user at a later time equal to the propagation delay. The user must know this delay because the accuracy of his knowledge of the time can certainly be no better than his knowledge of this delay.

Since all cycles of the signal are identical, the signal is ambiguous and the user must somehow decide which cycle is the "on time" cycle. This means he must already know the time to plus or minus half the period of the signal or ± 50 μs in this case.

Finally, the user may want to use this system, say once a day, for an extended period of time to check his clock. However, it may be that the delay will differ from one day to the next and if he does not know how this value differs from day to day, then his accuracy will be limited by the lack of repeatability of arrival time of the signal. These three characteristics are primarily related to the design of the signal and to the propagation characteristics of the medium. Some important practical considerations are cost to establish and maintain the service, ease of use, and cost to the user.

We can classify time dissemination systems in a number of ways but one useful division is to consider those systems whose carrier frequencies are low enough frequencies to be reflected by the ionosphere and those whose carrier frequencies are at higher frequencies which penetrate the ionosphere. The former may be observed at great distances from the transmitter but suffer from ionospheric propagation anomalies that limit accuracy, the latter (except for long-range groundwave systems such as Loran-C) are restricted to line of sight but suffer little or no deterioration in the signals due to propagation anomalies. The most accurate systems tend to be in the latter category while the former are used by the greatest number of users. Some systems, such as WWV, were built expressly for the purpose of time and frequency dissemination while others, such as Loran-C, a navigation system, can also be used to disseminate time and frequency.

Almost any radio signal with an unambiguous characteristic and which can be observed simultaneously at two locations can be used to synchronize clocks at those two locations. At the present time, TV is used extensively in this way. For example, suppose the arrival time of a synchronization pulse is measured with respect to the clocks at the two locations. If clocks are synchronized then the difference between the two clock measurements will simply be the radial difference in propagation delay as shown in Fig. 1. If this differential delay is known in advance of the measurements, then any difference between the measured differential delay and the known delay is the time difference between the two clocks. Measurements of this sort have shown that TV signals have the potential for disseminating time in the sub-microsecond region.[1,2] This has led NBS[3] to develop an experimental TV time dissemination system which incorporates a code in the transmitted television signal during the vertical blanking interval. Upon reception, this signal is decoded and displayed on a commercial TV set. The top line of the display gives hours, minutes, seconds and the bottom line gives the difference between a local clock 1 pps and the TV 1 pps. (See Fig. 2.) The resolution in the second line of display may be as small as 1 nanosecond.

There are a number of questions and problems for time dissemination which are not resolved. These are largely related to uncertainties concerning the configuration of the electronic systems of tomorrow which depend heavily upon time and frequency techniques. Examples are complex navigation and communication systems which are now being considered by both the civilian and military sectors. It is not clear how these systems could be used to disseminate time, but if there are a large number of independent systems, each generating and disseminating its own frequency and time information,

we may find it to be much more efficient, in terms of maximizing the communication capacity of the systems, to have one or two systems provide frequency and time information for the others. As an example (which is related to our current experience) consider TV which utilizes a significant portion of its communcation capacity to transmit sync pulses. We have a situation here where inexpensive clocks (or oscillators as they are usually called) have to be synchronized quite often at the expense of channel capacity. On the other hand, if TV sets had synchronization information from some other source, there would be more capacity for communications.

Alternatively, the TV set could have a better oscillator. Of course, the reason that we use inexpensive oscillators in TV receivers is that it is more economical than efficient use of the spectrum. However, the spectrum is a limited natural resource which will become more expensive as greater demands are made upon it. In fact, we might find at some future time that it will be more economical, in terms of spectrum conservation, to use both better clocks and common sources of time and frequency information. We might even go so far as to think of the latter as a time and frequency utility which could be "plugged into" at will.

References

1. J. Tolman, V. Ptáček, A. Souček, and R. Stecher, "Microsecond Clock Comparison by Means of TV Synchronizing Pulses," IEEE Trans. Instrumentation and Measurement, Vol. IM-16, No. 3, pp. 247-254, September 1967.

2. J. B. Milton, "Standard Time and Frequency: Its Generation, Control, and Dissemination from the National Bureau of Standards Time and Frequency Division," NBS Tech. Note 379, pp. 1-17, August 1969.

3. D. D. Davis, J. L. Jespersen, and G. Kamas, "The Use of Television Signals for Time and Frequency Dissemination," Proc. IEEE (Ltr.), Vol. 57, No. 6 (in press).

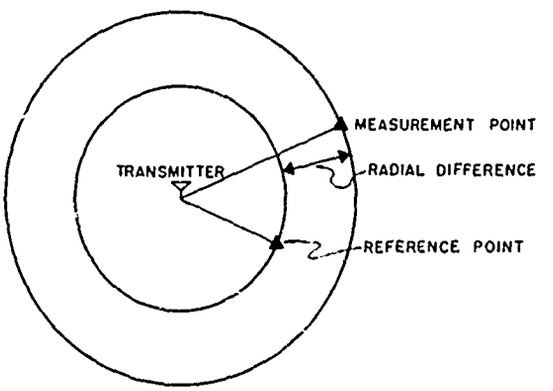


Figure 1. Illustrates use of a TV signal to synchronize clocks at two locations.

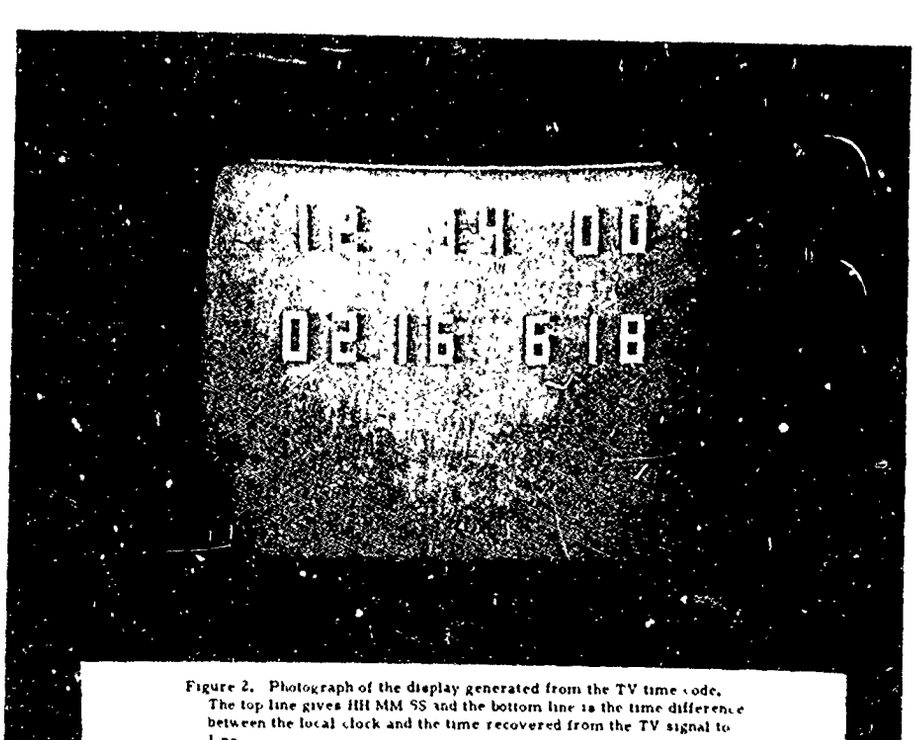


Figure 2. Photograph of the display generated from the TV time code. The top line gives HH MM SS and the bottom line is the time difference between the local clock and the time recovered from the TV signal to 1 ns.

TIME AND FREQUENCY TRANSFER VIA MICROWAVE LINK

D.H. Phillips, R.E. Phillips, and J.J. O'Neill
Naval Research Laboratory
Washington, D.C.

Summary

A two-way microwave link between the Naval Research Laboratory and the Naval Observatory has been established to transfer both time and frequency information with high accuracy. Time and frequency information is transmitted simultaneously by algebraically adding together a precision 1 MHz signal and 1 pps, thus providing continuous phase information for frequency comparison and epoch time by the pulse. Phase resolution is better than 10 nanoseconds and determination of epoch time better than 0.1 microseconds. Seven precision standards have been intercompared via the microwave link. Continuous phase recording since September 1969 has shown no diurnal dependence and very little effect due to seasonal and temperature changes.

A second microwave link has been established between The Naval Observatory and the Naval Research Laboratory's satellite research communications terminal in Waldorf, Maryland.

Introduction

This paper is concerned with the transmission of precise time and frequency information derived from the highest quality standards. The microwave link has proven to be a low noise, highly accurate method of time and frequency transfer which does not degrade the accuracy of the transmitted signals. This paper will describe techniques utilized in and results obtained from two existing microwave links with respect to establishing and/or maintaining epoch time at remote locations where precision frequency standards are maintained. The results also show great promise for application of microwave links to worldwide time transfer.

Microwave Link

Use of a microwave link for time transfer has yielded highly accurate results (better than 0.1 microseconds). Figure 1 shows a two-way microwave link which has been established between Naval Research Lab and the Naval Observatory and which transferred both time and frequency information. The path between the two locations was practically line of sight and its length was 11750 meters (7.3 miles). The transmitting frequency from the Naval Research Lab was 7137 MHz and the radiated power was 1 watt. The emission bandwidth was 20 MHz and the modulation frequency was 1 MHz which was derived from a hydrogen maser, the Laboratory's standard. At the Naval Observatory the transmitting frequency was 7415 MHz and the radiated power was 1 watt divided between a link to NRL and a second link to Waldorf, Maryland. The emission bandwidth was 20 MHz. The modulation was 1 MHz and 1 pps from the Observatory's Master Clock.

Figure 2 is a block diagram of typical instrumentation at a microwave terminal. At the Naval Observatory the 1 MHz and the 1 pps were combined and fed to the transmitter where a reflex klystron was repeller modulated. This produced a wideband frequency modulated (FM) signal which was fed by waveguide to a 4 foot paraboloid. The energy traveled through air with a delay of 39.3 microseconds to a receiving paraboloid at Naval Research Lab where the signal was fed through waveguide to the receiver for amplification and demodulation. The video output of the discriminator was very similar to the signal fed to the transmitter at the Naval Observatory. The unprocessed video signal was fed to the stop gate of a start stop counter. The counter was started by a pulse (tick) from the standard clock which was run by the phase shifted output of a cesium beam. The local tick starts the counter and the delayed tick stops the counter. The system was calibrated as having a delay of 40.6 microseconds, which included 0.5 microseconds delay at the Naval Observatory, 0.4 microseconds delay in the transmitter and receiver, 0.4 microseconds cable delay at the Naval Research Lab and 39.3 microseconds propagation time in air. The calibration was achieved with the use of a cesium beam traveling clock. The video signal was fed to a 1 MHz distribution amplifier with a 300 cycle bandpass which filtered out the tick information and delivered a pure sine wave for phase comparison to the hydrogen maser reference.

Perturbations to the Link

Figure 3 shows a section of a continuous phase recording. The green trace is the Naval Observatory's Master Clock via the microwave link compared against the NRL hydrogen maser. The blue trace and the purple trace are phase recordings of inhouse cesium beams against the

maser. The traces were so similar that it was as if the Observatory's master clock was in our own laboratory. Continuous phase recording, 24 hours a day, showed the technique to give results which were independent of diurnal shift. The link has been in operation since September 1969, and does not appear to be affected by seasonal and temperature changes. The only weather condition which seriously affected the link was one of heavy wet snow accumulating on the paraboloid which caused it to act as a microwave absorber. The transmission path was directly over National Airport which permitted evaluation of the effects of high density air traffic. The presence of individual airplanes was repeatedly observed by the amplitude of the signal on an oscilloscope but was not detectable on the phase recording.

Simultaneous Time and Frequency Information

Simultaneous time and frequency information was transmitted by sending a precision 1 MHz signal and 1 pps, algebraically added together. Figure 4 shows the oscilloscope presentation of the composite signal. A storage oscilloscope triggered by an external precision tick was used. Thus, continuous phase information was available for frequency comparison and epoch time was provided by the tick. The correlation of the tick and the phase yielded a method of cross checking time which assured freedom from timing errors at the receiving end. In instances where the output of the 1 MHz was multiplied by 10 to 1000 times phase resolution was observed to be better than 10 nanoseconds. Figure 5 shows a phase recording of the signal multiplied by 100 times. Full scale is 10 nanoseconds. An alternate method of obtaining high resolution would be the transmission of a higher frequency (5 MHz or 10 MHz).

A crystal oscillator was phase locked to the 1 MHz output of the distribution amplifier which yielded an alternate method of preserving time and phase measurements through minor interruptions. The frequency of the oscillator was adjusted to that of the standards so if the link were interrupted the oscillator continued to provide continuity for determination of time.

Frequency Standards Compared

Frequency standards of the highest caliber were used at all terminals of the link. Figure 6 is a photo of the hydrogen masers at Naval Research Lab. Maser I on the right has operated for 6 years as the NRL standard. Figure 7 is a photo of a cesium beam primary frequency standard. A combination of these two types of standards permits implementation of an excellent method for the determination of precision time. Figure 8 shows data obtained at the Naval Observatory of a phase comparison of Naval Research Lab's hydrogen maser via the microwave link against the Naval Observatory's maser. The green trace showed a frequency adjustment of the Naval Observatory's maser. The smooth trace showed the excellent short term stability of the link as well as of both the standards. The Naval Observatory's data showed the USNO H10 #10 maser had a standard deviation of 5 nanoseconds for an averaging time of 5 days for readings taken every 3 hours. NRL H10 #1 maser via microwave link had a standard deviation of 8.2 nanoseconds for the same conditions. For 1.5 day averaging time USNO #10 maser had a standard deviation of 2.2 nanoseconds NRL #1 maser + link had a standard deviation of 3.6 nanoseconds. Figure 9 shows data obtained and graphed at the Naval Observatory. These graphs show the performance of Naval Research Lab hydrogen maser via microwave link against the Naval Observatory's computed average. The top graph shows the frequency offset from 1 January 1970 to 2 March 1970 where full scale is 5 parts in 10^{12}. The lower graph shows the standard deviation in nanoseconds of a 1-day averaging time of readings taken every 3 hours from 1 January 1970 to 2 March 1970. It was noted that the sigma values fall below 8 nanoseconds.

Microwave Gear

Navy surplus microwave equipment was utilized for the link. Separate transmitter and receiver klystrons were used. Figure 10 shows a photo of the equipment at a microwave terminal - a transmitter and receiver connected by a diplexer. A diplexer was used to allow transmitting and receiving from the same 4 foot microwave paraboloid. The frequencies being used were in the 7 GHz range which allowed the use of highly directive parabolic antennas. Low transmitter power was utilized and interference was minimized by the narrow beams. Carrier frequency stability was not critical allowing the use of unsophisticated microwave terminals. The narrow beams reduced the multipath error to a minimum. The multipath which did occur came from sources close to the direct path and the timing error remained small. The first information arriving was the direct path so that in cases where multipath did occur the leading edge of a pulse was used to give time information. A start stop counter was used to minimize the effects of the multipath.

Waldorf Link

A second microwave link was recently established in conjunction with satellite timing experiments between the Naval Observatory and the satellite communications terminal at Waldorf,

Maryland. Figure 11 shows a photo of the 6 foot microwave paraboloid on top of a 100 foot tower with the satellite communications antenna in the background. The 20 mile link was practically line of sight. Phase recordings have been made of the Naval Observatory master clock versus cesium clock #208 at Waldorf. The phase record of this link is shown in Figure 12. The chart has a full scale deflection of 1 microsecond and showed phase resolution to better than 10 nanoseconds.

Conclusions

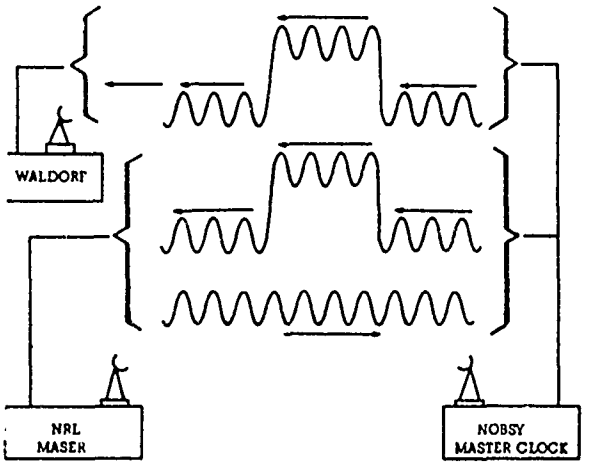
The concept of using a microwave link to transfer precise time and frequency information with a high order of accuracy has been satisfactorily implemented. The capability to transfer data from remotely located standards is particularly important in eliminating systematic errors such as caused by temperature changes and power failures. At present, 7 precision standards are being intercompared via the microwave link.

The microwave link is in current use as part of an experimental program to disseminate time worldwide. At present, epoch time is being transferred to 2 satellite systems. Time has been transferred from the Naval Observatory to the Naval Research Laboratory's satellite research communications terminal at Waldorf. A description of one satellite time transfer experiment, between Brandywine and Hawaii, was recorded in Naval Research Laboratory's Memorandum Report 2110, "Time Transfer by Communication Satellite" of March 25, 1970, by R.R. Stone, J. Murray, D. Phillips, and D. Pritt.

Acknowledgements

The authors would like to acknowledge with gratitude the assistance of the Naval Observatory in obtaining data pertaining to the microwave link and the contributions to this experiment by Mr. Phillip Hooten, Mr. Tom Leonard and Mr. John Bowman of Naval Research Laboratory.

Key words (for information retrieval): time transfer, microwave link, precision frequency standards



Equipment: Raytheon Television Microwave Relay Model KTR-1000G

Naval Research Lab	Naval Observatory
Trans. freq. - 7137 MHz	Trans. freq. - 7415 MHz
Power - 1 watt	Power - 1 watt
Emission bandwidth - 20F9	Emission bandwidth - 20F9
Transmit hydrogen maser derived 1 MHz signal	Transmit 1 MHz and 1 pps from Master Clock

Results

Total time delay - 40.6 microseconds. (0.5 microseconds cable delay at Naval Observatory, 0.4 microseconds delay in transmitter and receiver, 0.4 microseconds cable delay at NRL. 39.3 microseconds as time traveled in air. This is equivalent to 7.3 miles.)
Precision - 2.3 nanoseconds for one day average.
Phase accuracy - 10 nanoseconds.
Tick jitter - 22 nanoseconds tick to tick (average error over 1 minute was 5 nanoseconds).
Time transfer to better than 0.1 microseconds have been achieved.

Figure 1. MICROWAVE LINK

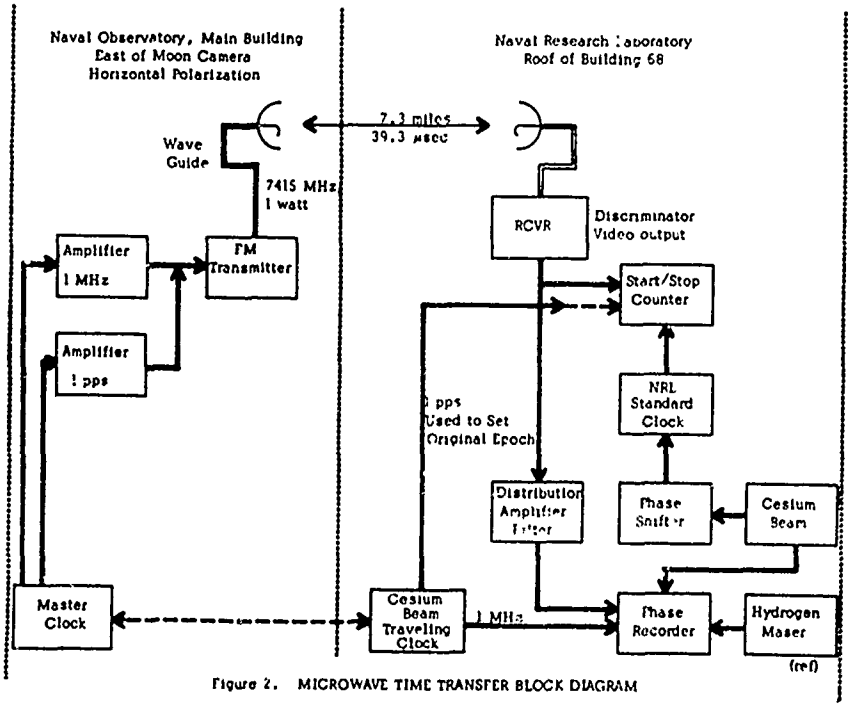


Figure 2. MICROWAVE TIME TRANSFER BLOCK DIAGRAM

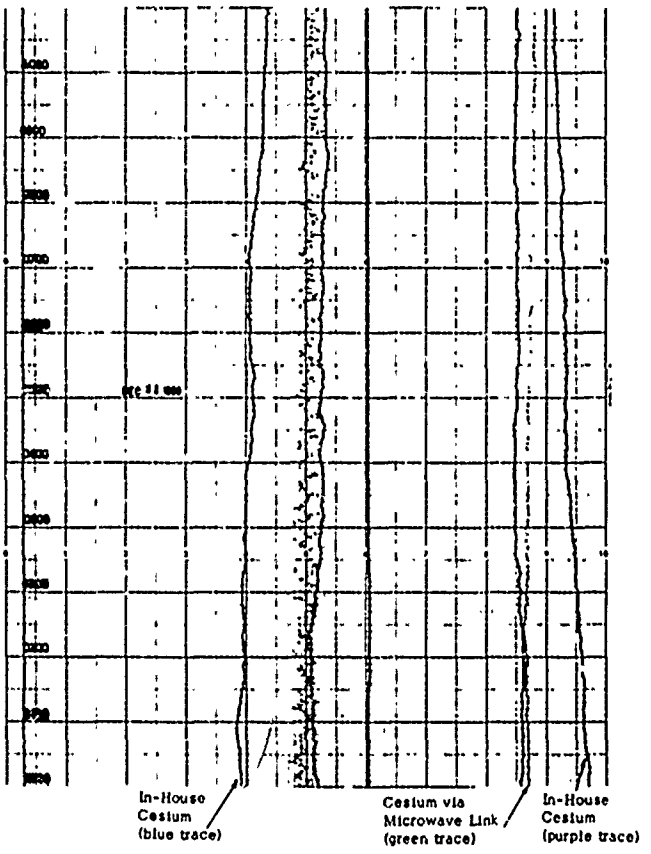
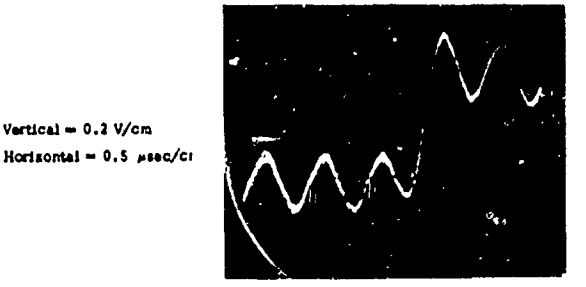
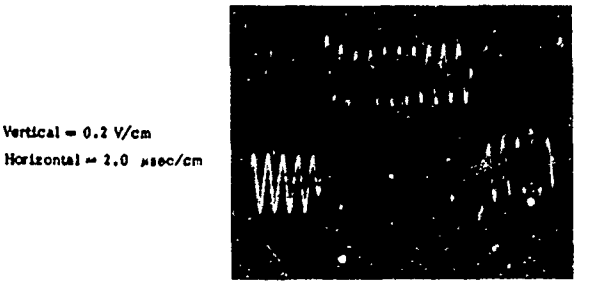


Figure 3. TWELVE-HOUR MULTIPLE PHASE RECORDING AT NAVAL RESEARCH LABORATORY



Vertical = 0.2 V/cm
Horizontal = 0.5 μsec/cm

Attenuation:
Pulse - 2 db
1 MHz - 4 db



Vertical = 0.2 V/cm
Horizontal = 2.0 μsec/cm

Figure 4. STORAGE OSCILLOSCOPE PRESENTATION OF RECEIVED PULSE (Microwave Link Pulse and 1 MHz Transmitted)

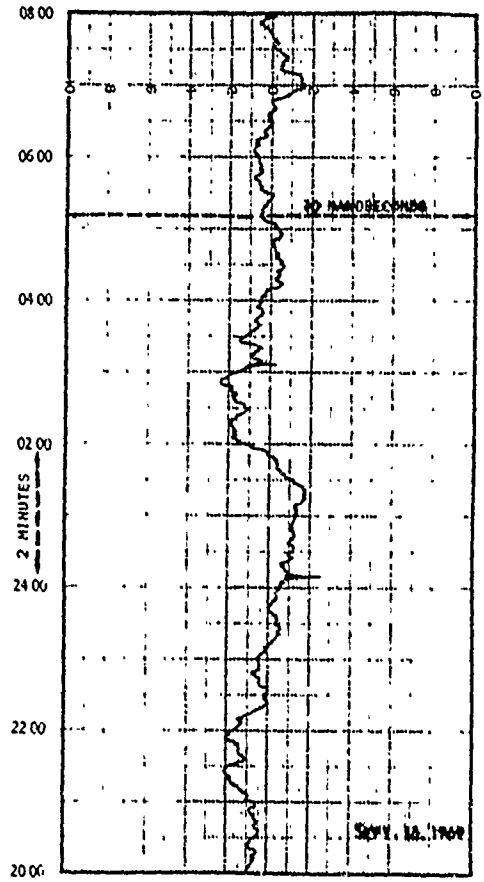


Figure 5. SHORT-TERM PHASE COMPARISON OF 1 MHz SIGNAL WITH MASER (Difference Between Master Clock at Naval Observatory and Standard at Naval Research Lab)

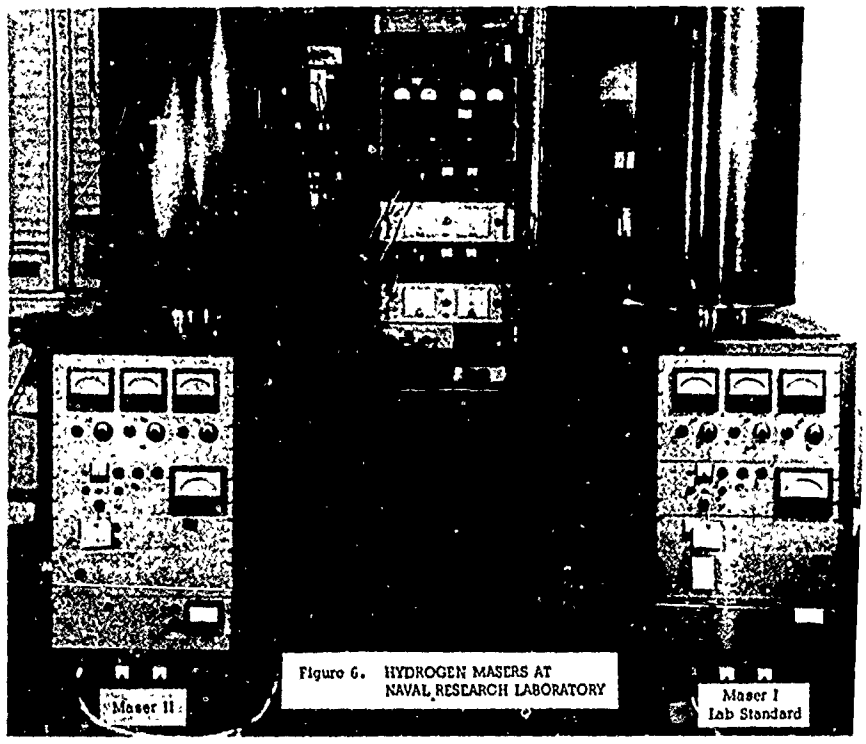
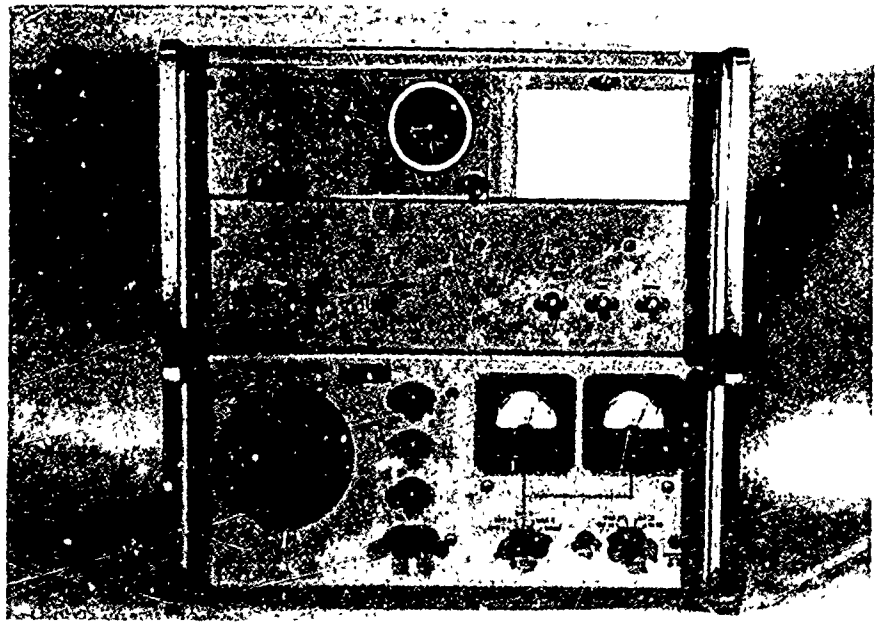


Figure 6. HYDROGEN MASERS AT NAVAL RESEARCH LABORATORY

Reproduced from best available copy.



CESIUM BEAM

Figure 7. TRAVELING CLOCK TIME AND FREQUENCY STANDARD

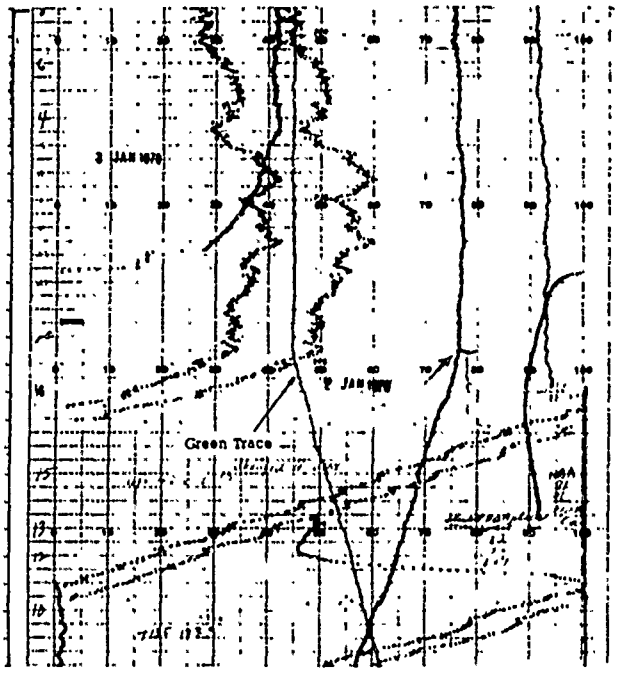


Figure 8. TUNING OF USNO H10 #10 MASER WITH NRL H10 #1 MASER VIA MICROWAVE LINK

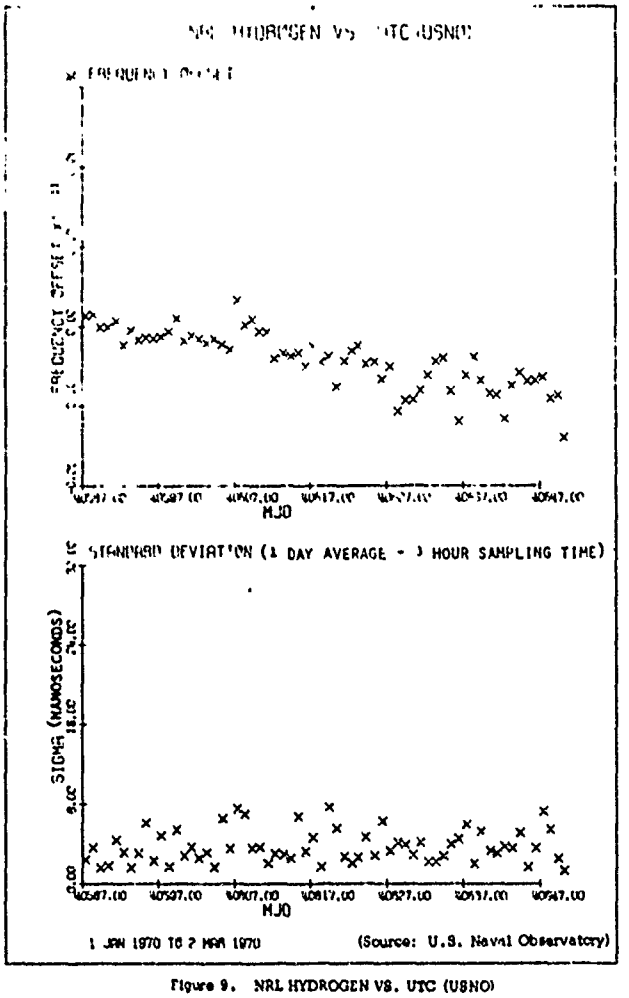


Figure 9. NRL HYDROGEN VS. UTC (USNO)

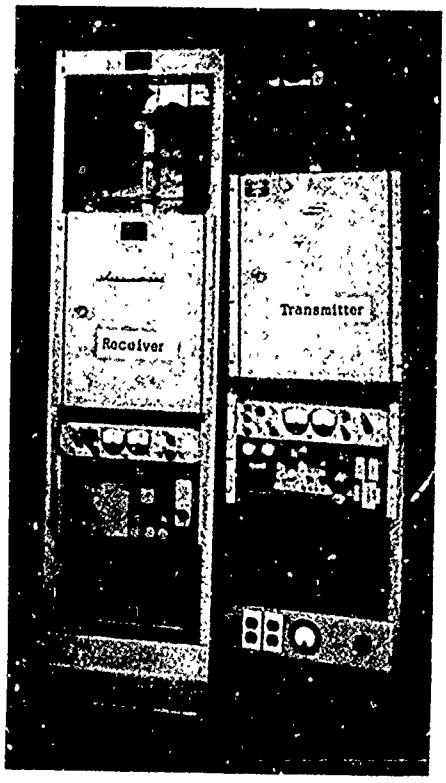


Figure 10. MICROWAVE GEAR

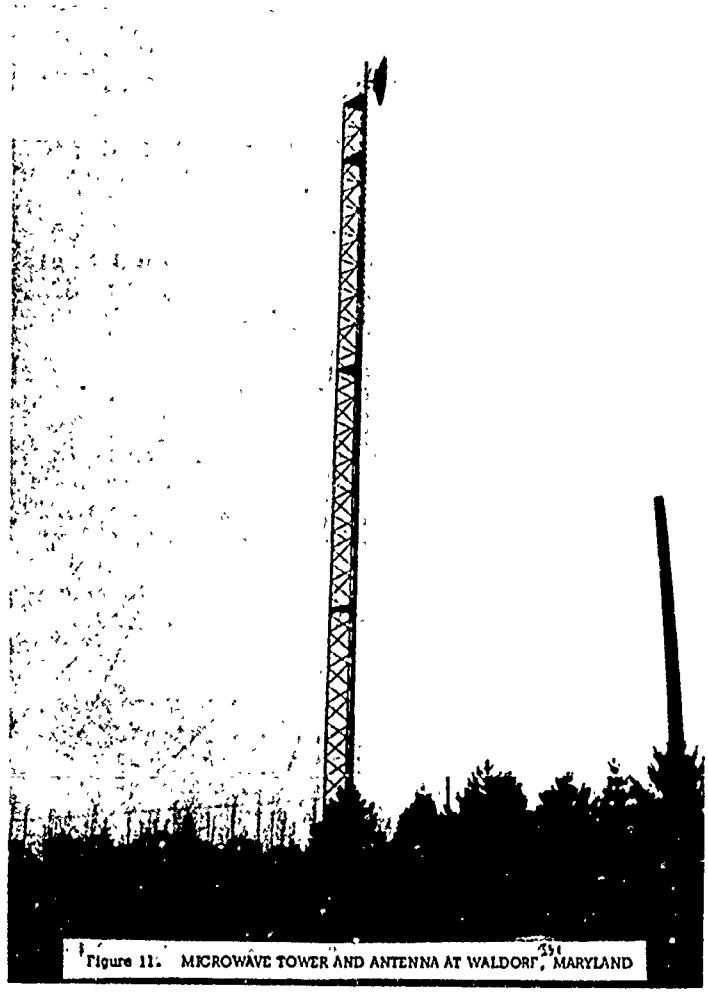


Figure 11. MICROWAVE TOWER AND ANTENNA AT WALDORF, MARYLAND

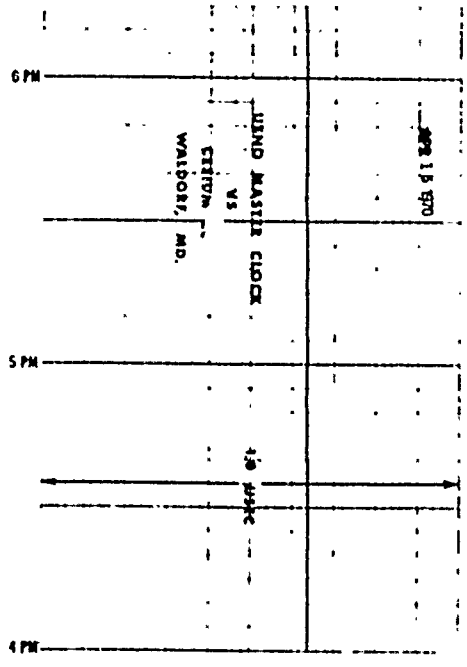


Figure 12. PHASE RECORDING AT WALDORF, MARYLAND

DIURNAL PHASE CONSTRUCTION OF VLF SIGNALS NEAR ANTIPODE OF A TRANSMITTER

Andrew R. Chi
Goddard Space Flight Center
Greenbelt, Md. 20771

Abstract

During a field trip in March 1967, to a NASA tracking station in Carnarvon, Western Australia for testing time synchronization using dual VLF transmissions from NBS Radio Station WWVL, Fort Collins, Colorado, it was found that the phase of the received signal did not exhibit the familiar diurnal phase patterns. It was further observed that the phase tracking receiver locked at different positions during and near the sun-rise and sun-set periods.

From relative amplitude measurements and phase records of the received signals, these observed phase changes during and near the sun-rise and sun-set periods were identified to be due to reception of other signals. These other signals were postulated to be those signals which were propagated along the path in the opposite directions of the transmitter-receiver path. Using this hypothesis and assuming that the total phase delay of the emitted signals from the transmitter to the receiver over the long and short propagation paths is constant, the diurnal phase near the antipode of the transmitter can be constructed.

This paper will present the results of diurnal phase construction from the phase records obtained in March 1969 in Carnarvon which is near the antipode of the WWVL transmitter.

Background

Since Pierce[1] identified that the phase variation of a received VLF signal is diurnal, and that the pattern is characteristic of the propagation medium, path length, etc. the application of VLF transmission for precision frequency comparison became immediately obvious. The diurnal phase pattern also becomes a signature for identifying the receiving of a VLF signal.

During a field trip in March 1967, to a NASA tracking station in Carnarvon, Western Australia for testing time synchronization using dual VLF transmissions[2] from NBS Radio Station WWVL, Fort Collins, Colorado, it was found that the phase of the received signals did not exhibit the familiar diurnal phase patterns. It was further observed that the phase tracking receiver (Tracor Model 599) locked at different positions during and near the sun-rise and sun-set periods.

The irregular phase variation of the received signal during the sun-rise and sun-set hours was postulated to be that due to two signals received from the transmitter propagated along a great circle in two opposite directions. The hypothesis was not tested until March, 1969 when a group from GSFC headed by D. A. Premo went to Carnarvon to test the optical tracking for Geodetic Satellite, GEOS II. The results of this report bear out the hypothesis and also provide a new technique for constructing the diurnal phase from the two received signals near the antipodal point of a transmitter.

Instrumentation

Nearly identical instrumentation and procedure were furnished to the second group who went to Australia to re-test the reception of the dual VLF transmissions. These include two phase tracking VLF receivers (Tracer Model 599), phase coherent frequency generator for receiver calibration,[3] loop and whip antennae with a cardiod unit, a rubidium portable clock (Tracer Model 307A) for time reference, and a rubidium gas cell frequency standard (H/P Model 5065A) for frequency reference and comparison with the portable clock.

WWVL Transmissions

The 19.9 and 20.0 KHz carrier frequencies which had been transmitted by WWVL since 1965 were changed to 20.0 and 20.5 KHz on October 7, 1968. For our test these two carrier frequencies were changed again to 19.9 and 20.5 KHz in February and March 1969 by the National Bureau of Standards. The transmission was also changed to permit positive identification[4] of the received VLF signals to a format follows:

1. Single frequency transmission at 19.9 KHz from 2300 UT February 25 to 1700 UT March 1, 1969.

2. Single frequency transmission at 20.5 KHz from 1700 UT March 1 to 1700 UT March 4, 1969.

3. Dual Frequency transmission at 19.9 and 20.5 KHz alternately at 10 second intervals from 1700 UT March 4 and terminating at 1700 UT March 11, 1969.

Experimental Result

The actual phase records of 19.9 and 20.5 KHz reduced in size are shown in Figure 1. It is obvious that these curves do not exhibit the recognizable diurnal phase pattern which is typical of a VLF signal propagated by a single path. Using the 19.9 KHz phase record, if the

signal phase during 0900 and 1800 UT February 26, is received during night time along the short path between Fort Collins and Carnarvon, the phase record from 2200 to 0200 UT February 26 and 27 would have to be that along the long path during night time since during this time the signal phase is increased in delay.

Hypothesis

In order to construct a diurnal phase of a VLF signal, the received phase of the signal propagated by a mixed paths must be converted to that by a single path, i.e., the signal phase propagated along the long path must be converted to the equivalent signal phase along the short path.

The conversion of the signal phase propagated by the long path to that by the short path is made with the following assumptions:

a. The sum of the propagation delay of a VLF signal along the short path and along the long path is constant.

b. The difference in the propagation delay for day time and night time paths can be approximated by using two propagation phase delays per unit length to account for the variation of the height of the signal reflection medium, D-layer of the ionosphere.

Using the above assumptions, we can consider the positions of the transmitter and receiver located in a great circle as shown in Figure 2 by T and R respectively. The signal time delay along the short path, TAR, is given by equation (1) and the signal time delay by the long path, TBR, is given by equation (2).

$$t_s = x\rho_d + y\rho_n = n_s\tau + \frac{\Delta\phi_s}{\omega} \qquad (1)$$

$$t_l = (l_d - x)\,\rho_d + (l_n - y)\,\rho_n = n_l\tau + \frac{\Delta t_2}{\omega} \qquad (2)$$

In equations (1) and (2) ρ is the propagation delay per unit length; l, x, and y are propagation path lengths as shown in Figure 2; n is an integer; τ is $1/f$; ω is circular frequency; Δt is phase angle less than 2π; and the subscripts s, l, d, and n denotes short path, long path, day, and night respectively.

Add equations (1) and (2) we obtain equation (3). K is a constant under assumption (a).

$$t_s + t_l = l_d\rho_d + l_n\rho_n = K \qquad (3)$$

From equations (2) and (3) we can also write equation (3)

$$\Delta\phi_s = \omega[K - (n_s + n_l)\,\tau] - \Delta\phi_l. \qquad (4)$$

Let

$$\varphi_k = \omega\,[K - (n_s + n_l)\,\tau]$$

which is also a constant of n_s and n_l do not change. Equation (4) can then be written as

$$\Delta\phi_s = \varphi_k - \Delta\phi_l. \qquad (5)$$

Equation (5) implies that the signal phase along a long path can be converted to an equivalent short path phase by a mirror reflection along the axis φ_k.

In VLF phase tracking it has been observed, however, that n_s and n_l do not remain constant.[5] This change occurs whenever the signal phase is changed at a rate faster than that of the servo system of the receiver. When such phase change occurs, the phase record usually indicates a phase jump which, due to receiver design, takes the steps of a whole integer of cycles of the carrier frequency unless there is a discontinuity in operation. In phase conversion from a long path to a short path, the integer of cycles of phase jump becomes an integer multiple of a half cycle due to mirror reflection.

Phase Conversion

If the signal of a received VLF transmission is propagated along a short path and then along a long path, the signal phase along the long path can be converted to that along the short path by a mirror reflection at the time when the transition took place. This is based on the fact that the total propagation delay of the signal along a great circle is constant. If there is a change in phase due to the signal strength dropping below the receiver threshold, such as those indicated by integer cycle separation in Figure 3, the phase conversion can be made by the same technique. Of course, the mirror position is shifted if the phase change is not corrected as indicated by the solid lines. When the phase change is corrected, the mirror position is moved by multiples of one-half cycle as indicated by the dashed lines.

This phase processing technique can be used only if the change of reception of the signal via long path to short path is known. The detection of the change of the signal must be made by observing a recording of the signal amplitude as shown in Figure 4. In Figure 4, the lower

curve is the amplitude record of the 20.5 KHz signal from WWVL as received in Carnarvon and the upper curve is the phase record of the same signal. As is evident from the amplitude record, the received signal amplitude is gradually dropped to -10db at 0700 UT on May 3 and then increased to +10db at 1000 UT. The increase in the amplitude of the received VLF signal after sun-rise is contrary to the known behavior of the field strength of a VLF signal propagated by a single path.[6] Since the signal amplitude is attenuated more during the day time propagation than during the night time propagation, the received signal after sun-rise must be that which is propagated during the night time path and is along the opposite direction. The phase jump of one cycle as indicated by the upper curve on Figure 4 is due to the phase drift caused by noise in the receiver in the absence of a signal, i.e., the signal was below the threshold of phase tracking receiver which is -20db. This was the case at 1800 UT. on May 3 as shown by the lower amplitude record, and the same analysis is applied for the 20.5 KHz signal as shown on the upper curve in Figure 3.

When the 10 second duty cycle was resumed in the WWVL transmissions on March 4, the phase records were made for 19.9 and 20.5 KHz signals. These have been reconstructed in accordance with the technique described and are shown in Figures 5 and 6 respectively. The consistent phase advance in the 19.9 KHz signal in Figure 5 was proven after the test to be due to the malfunction of the phase detector of the receiver.

We have developed a theory to explain the phase conversion technique and are still collecting data to test the validity of this technique.

Acknowledgment

The author wishes to express his appreciation to Dr. J. A. Barnes and R. E. Beehler and P. P. Viezbicke of the National Bureau of Standards for their cooperation in the transmission of 19.9 and 20.5 KHz signals in a new format from WWVL, and to his colleagues at Goddard Space Flight Center in particular D. A. Premo for collecting the data in Australia and W. E. Fizell for reducing the data.

References

1. Pierce, J. A., "The Diurnal Carrier Phase Variations of a 16 Kc/s Trans-Atlantic Signal," Proc. IRE 43, 584-588, 1955.

2. Chi, A. R. and Witt, S. N., "Time Synchronization of Remote Clocks Using Dual VLF Transmissions," Proc. 20th Annual Symposium in Frequency Control 588-611, April 1966.

3. Chi, A. R. and Witt, S. N., "Time Synchronization of Remote Clocks Using Dual VLF Transmissions," Proc. National Telemetering Conference 211-214, 1966.

4. The test transmission of 15.5 KHz from U. S. Naval VLF transmitter, NWC, at Northwest Cape in 1967 generated an interfering signal. The difference between the 4th harmonic of 15.5 KHz signal and the 3rd harmonic of the local oscillator frequency generated in the receiver at 21.0 Khz produced the same IF frequency as the receiver when tracking the 20.0 KHz signal. This interference signal was removed when the local oscillator frequency was tuned to 19.0 KHz.

5. Walker, D., "Phase Steps and Amplitude Fading of VLF Signals at Dawn and Dusk," Radio Science Vol. 69D, 11 (November 1965).

6. Snyder, F. P. and Bickel, J. E., "Measured Amplitude Variations of the 19.8 KHz Field of NPM Near Its Antipode," Radio Science Vol. 2, 7 (July 1967).

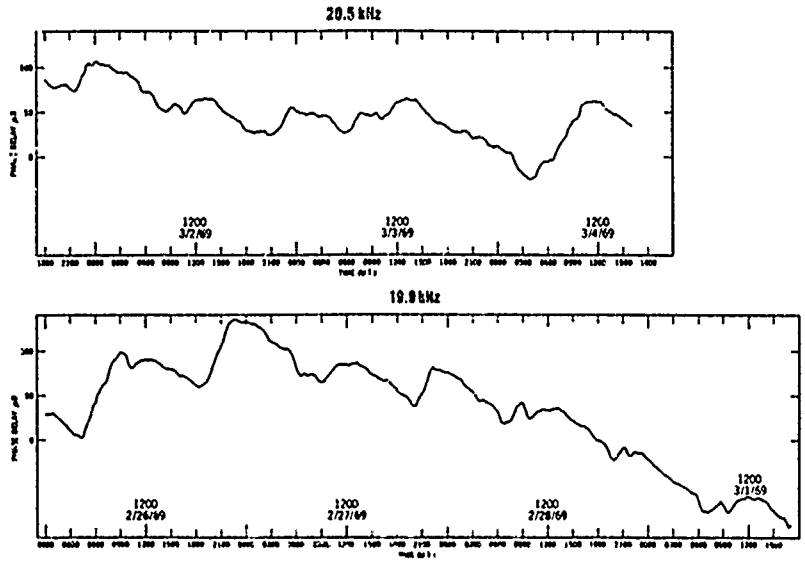
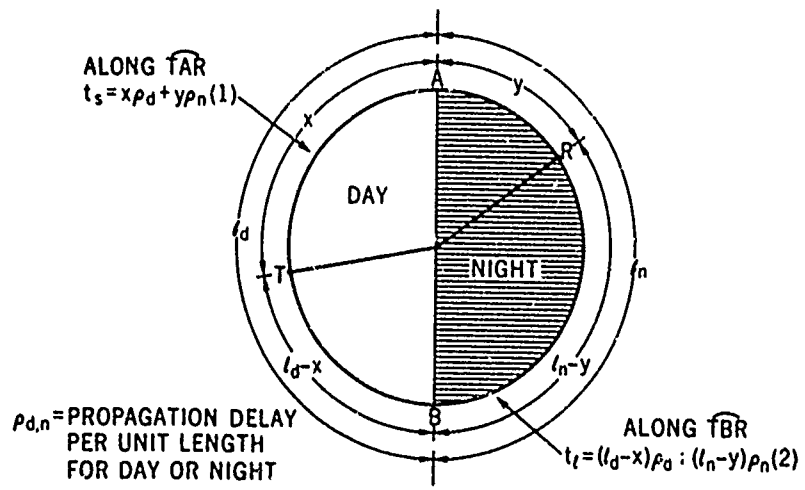
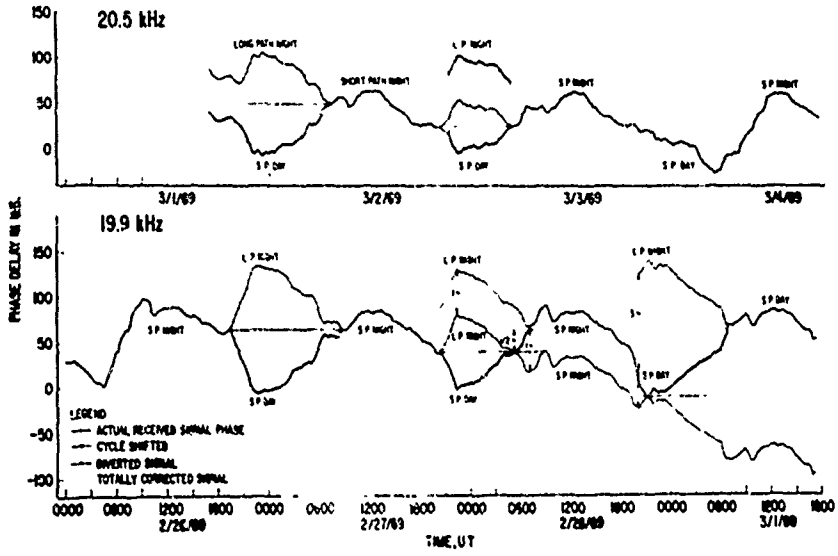


FIGURE . PHASE RECORDS OF 20.5 AND 19.9 KHz SIGNALS FROM WWVL AS RECEIVED AT CARNARVON WESTERN AUSTRALIA

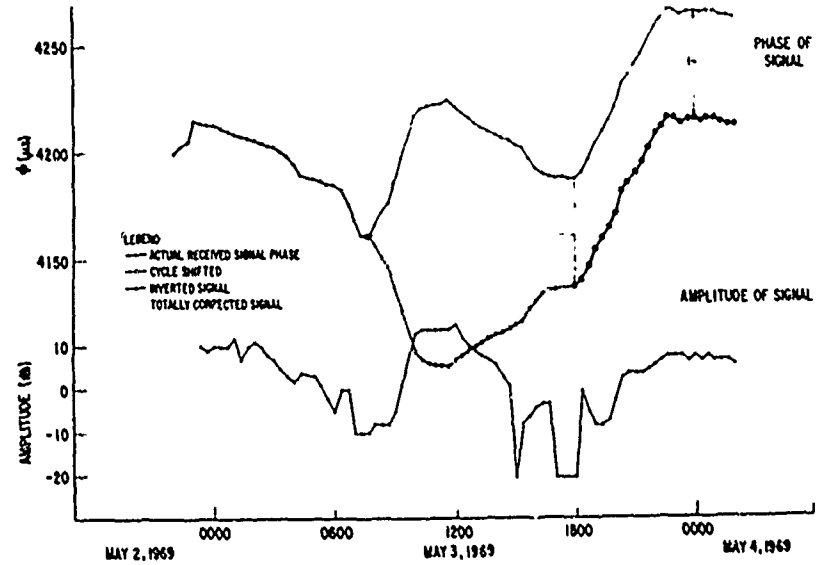


RECEIVER, R, IS NEAR THE ANTIPODAL POINT OF THE TRANSMITTER, T. AB IS THE DIVIDING LINE FOR NIGHT AND DAY TIME PROPAGATION. x IS THE DAY TIME PATH AND y IS THE NIGHT TIME PATH ALONG $\overset{\frown}{TAR}$. l_d AND l_n ARE THE TOTAL DAY TIME AND NIGHT TIME PATH LENGTH RESPECTIVELY

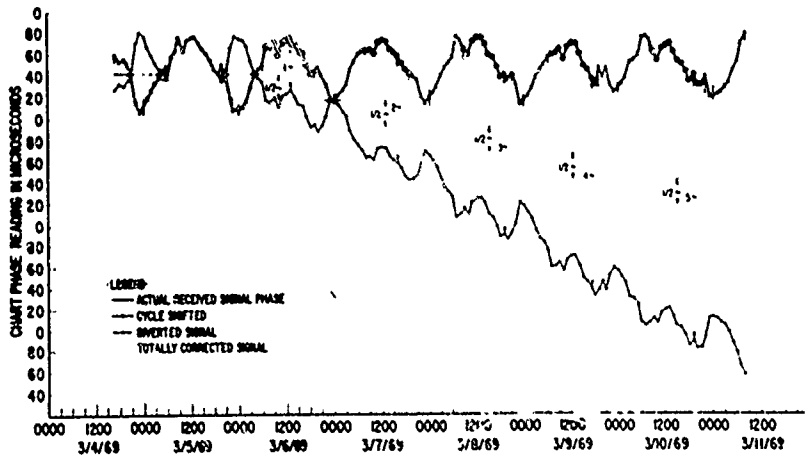
RECEIVED FROM 19.9 kHz AND 20.5 kHz DUAL FREQUENCY EXPERIMENT



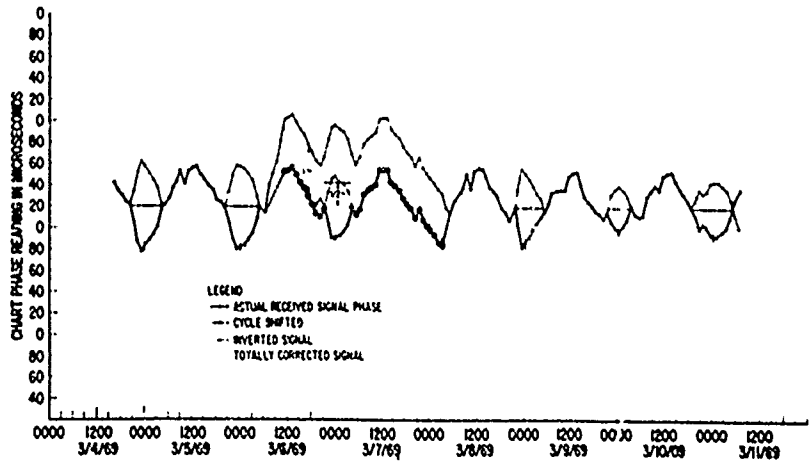
PHASE OF SIGNAL BROADCAST AT WWVL, FT. COLLINS, COLO. & RECEIVED AT CARNARVON, AUSTRALIA



19.9 kHz RECEIVED PHASE: TRANSMITTED FROM WWVL, FT. COLLINS, COLO. RECEIVED AT CARNARVON, AUSTRALIA



20.5 kHz RECEIVED PHASE: TRANSMITTED FROM WWVL, FT. COLLINS, COLO. RECEIVED AT CARNARVON, AUSTRALIA



DELAY TIME, t_S ALONG SHORT-PATH $\overset{\frown}{TAR}$,

$$t_S = x\rho_d + y\rho_n = n_S\tau + \frac{\Delta\phi_S}{\omega} \qquad (1)$$

DELAY TIME, t_l ALONG THE LONG-PATH $\overset{\frown}{TBR}$,

$$t_l = (l_d - x)\rho_d + (l_n - y)\rho_n = n_l\tau + \frac{\Delta\phi_l}{\omega} \qquad (2)$$

WHERE $\rho_{d,n}$ = PROPAGATION DELAY PER UNIT LENGTH FOR DAY OR NIGHT

$n_{S,l}$ = INTEGER OF CYCLES ALONG SHORT OR LONG PATH

$\Delta\phi_{S,l}$ = PHASE ANGLE LESS THAN 2π FOR SHORT OR LONG PATH

ADD Equations (1) and (2)

$$\boxed{t_S + t_l = l_d\rho_d + l_n\rho_n = K \text{ (CONSTANT)}} \qquad (3)$$

ALSO FROM Equations (1) AND (2)

$$(n_S + n_l)\tau + (\Delta\phi_S + \Delta\phi_l)\frac{1}{\omega} = K \qquad (4)$$

FROM Equation 4

$$\Delta\phi_S = \omega\,[K - (n_S + n_l)\tau] - \Delta\phi_l$$

$$\boxed{\Delta\phi_S = \phi_K - \Delta\phi_l} \qquad (5)$$

WHERE $\phi_K = \omega\,[K - (n_S + n_l)\tau]$

ϕ_K IS ASSUMED TO BE CONSTANT

FOR THE SAME LOCATIONS OF T AND R.

A SECOND SATELLITE OSCILLATOR EXPERIMENT

R. Easton, C. Bartholomew and J. Bowman
Naval Research Laboratory
Washington, D.C.

and .

M. Bloch
Frequency Electronics Inc.
New Hyde Park, N.Y.

Summary

The performance of a quartz crystal oscillator in the Timation I satellite has been reported in a previous symposium on frequency control and will be reviewed briefly as a basis for comparison. The principal problem with the oscillator in this first satellite was temperature coefficient. The effect of radiation on this oscillator has been deduced from subsequent data and will be presented.

A second experimental navigation satellite was launched in September 1969 in which an improved quartz crystal oscillator was included. Pre and post flight performance of this oscillator will be compared with the earlier unit.

Introduction

This paper discusses the performance of quartz crystal oscillators as frequency standards for earth satellite experiments. The preliminary results of the first experiment were reported at the symposium two years ago. The oscillator used for the first experiment used a selected 5 MHz fifth overtone glass enclosed crystal in a double proportional oven system. The input power required was 1.5 watts at 20°C. This unit was provided with an electromechanical tuning device that allowed frequency adjustments in increments of one part in 10^{11}. The operating temperature range was from -10 to +35°C.

Timation I

Figure 1 is a history of frequency of the oscillator versus temperature for the first eight months of orbital life. The temperature coefficient determined from this data was approximately $2pp10^{11}$ per degree Celsius. This coefficient was higher than expected and the primary limiting feature in the oscillator's performance. If the frequency is considered at various points of equal temperature the apparent aging rate is on the order of $+2pp10^{12}$ per day. This value is lower than that indicated before launch. Radiation is suspected as being the cause of this favorable compensation. Figure 2 shows segments of frequency versus time and temperature at later periods. It is seen that the apparent aging rate changed from a small positive rate to a negative one of approximately -1 part per 10^{11} per day.

At this time the backup oscillator in the laboratory had reached an aging rate of approximately +1 part per 10^{11} per day. If one assumes that the flight oscillator had reached a similar real aging rate then the effect due to radiation was -2 parts in 10^{11} per day. Figure 3 shows the aging rate performance of the backup oscillator and the performance of the flight oscillator. The values for the flight oscillator after launch were obtained by compensating the apparent aging rate with the inferred radiation effect. It appeared that if the shielding about the quartz crystal could be increased by a factor of two, the radiation effect could be reduced at least by that amount with the real value depending on the number-energy distribution of particles in this environment. The preflight tests on oscillators for Timation II indicated fairly constant aging rates of approximately 1 part per 10^{11} per day. It appeared that the total shielding about the oscillator in Timation I was equivalent to 1 cm of aluminum. To double this value approximately 0.24 cm of lead was included in the oscillator package for Timation II.

Timation II

The Timation II oscillator includes two major improvement factors. The first is a crystal using thermal compression bonded leads, cold welded enclosure seal and a high temperature bakeout under high vacuum. Crystals prepared using these techniques attained aging rates of a few parts in 10^{11} per day in days compared to months for the crystal type used in Timation I. The second improvement is a triple proportional oven system designed to improve the temperature coefficient by a factor of ten. In addition, the tuning mechanism gear ratio was increased to reduce the tuning steps to approximately 3.6 parts per 10^{12}. While increments of this magnitude are beyond the linearity specifications of the best piston type capacitors, this value is consistent when 50 or more increments are averaged.

A proportional controlled thermal electric temperature control system was developed during the design of Timation II as an experiment to determine if a critical satellite component could be maintained at a relatively constant temperature over the expected ambient variation. This device was installed with the oscillator to provide improved thermal control.

Figure 4 is a graph of percent time in sunlight of the satellite, satellite temperature and oscillator temperature versus time. The thermal

electric system maintained the oscillator temperature at $24^{o} \pm 0.1^{o}C$ over the variations in temperature of the satellite. The vertical lines for satellite temperature for each day show the range of temperatures measured during from three to five passes a day from a telemetering station. These data and the oscillators measured temperature coefficient of 1-2 parts per 10^{12} per ^{o}C indicate that the oscillators performance is independent of ambient temperature.

Figure 5 is a graph of frequency versus time for the satellite oscillator. The lower curve shows the measured frequency and the times and increments of tuning operations. During the early phase the aging rate is positive and gradually approaching zero. Optimistically, it was hoped that the oscillator was recovering from the effects of launch vibration and zero g environment. The low aging rate observed prior to launch (approximately 1 part per 10^{11} per day) and the expected radiation effect with the increased shielding could possibly yield a long term aging rate of a few parts in 10^{12} per day.

The balance of the data shows our optimism was short lived. The effect of radiation is apparent and soon became approximately constant. It appears that the oscillator has reached a relatively low constant aging rate but the effect of radiation has been miscalculated. The upper curve is the normalized frequency versus time curve and provides a more graphic view of the results.

Subsequent studies have provided the probable answer. Different pieces of quartz exhibit different frequency variations with radiation dosage. For example, Figure 6 is a graph of frequency versus radiation dosage carried approximately to saturation. It is significant that the slope of the curve (extrapolated) for low dosage is highest. This particular crystal showed a maximum value of 4 parts per 10^{11} per rad initially and decreased from that point. The Q of this unit changed from 1.4×10^{6} before radiation to 0.7×10^{6} after near saturation. It would appear that components which show reduced radiation effects such as swept quartz, are probable answers. Radiation tests might be made on a particular crystal to provide the value of shielding necessary for compensation in a known environment. Sufficient shielding to make the radiation effect negligible is straightforward but a weight consideration for many applications.

Figure 7 is a plot of aging rate versus time to show the typical performance of the Timation II oscillator after a tuning operation (prior to flight) the dotted curve is the long term aging rate based on the total accumulated time versus time. The second curve is a day to day aging rate determined by the difference in offset frequency for one day (based on the accumulated time for that day) subtracted from the next day. While the long term aging shows an acceptable device for long term timekeeping, the day to day variations limit the predictability of time (or phase) for short periods. The problem cause appears to be one of oven stability and time constant. Present design efforts are being directed in this area.

THE INFLUENCE OF SUNLIGHT ON TEMPERATURE AND FREQUENCY

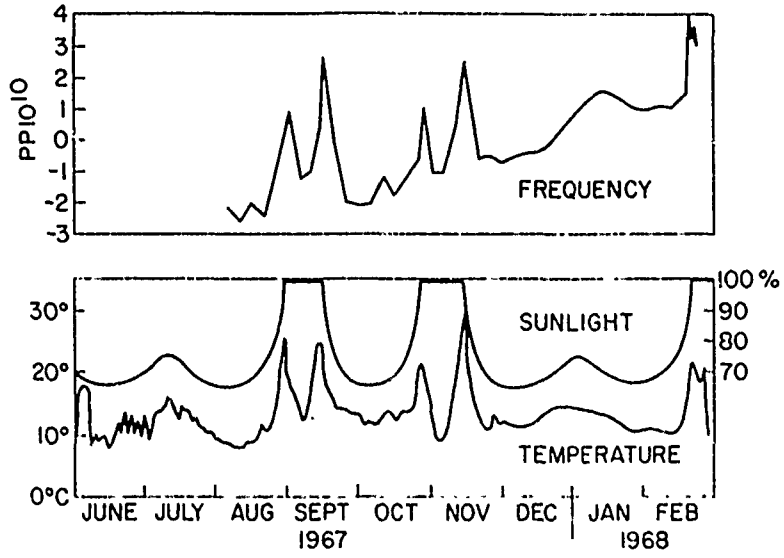


Figure 1 The Influence of Sunlight on Temperature and Frequency

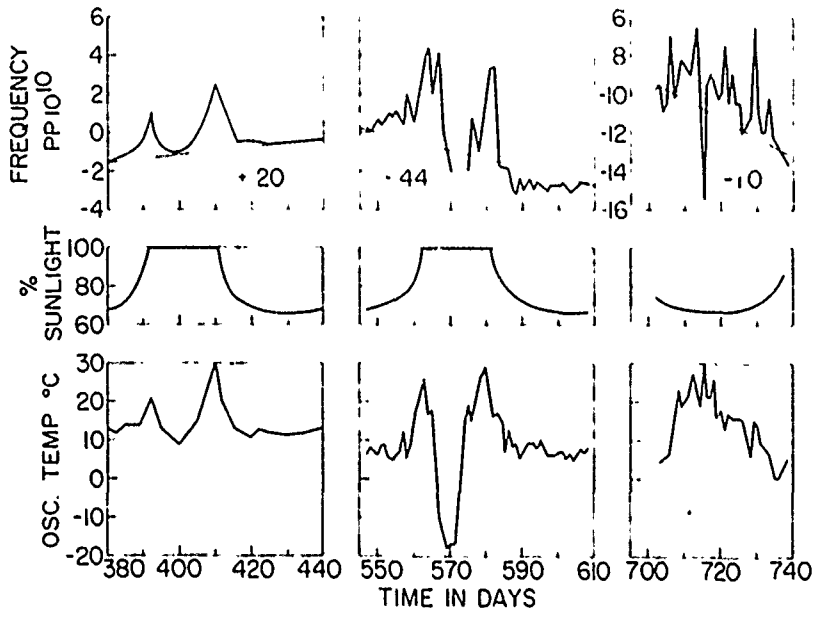


Figure 2 Apparent Aging Rate at Different Times

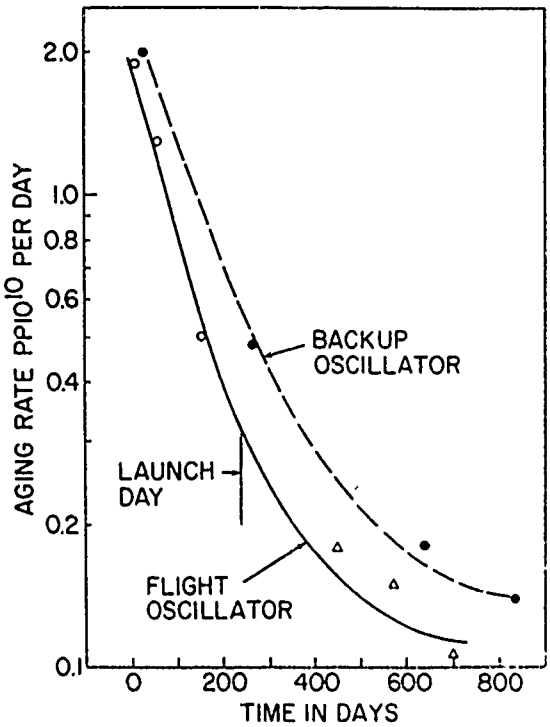


Figure 1 Aging Rate versus Time

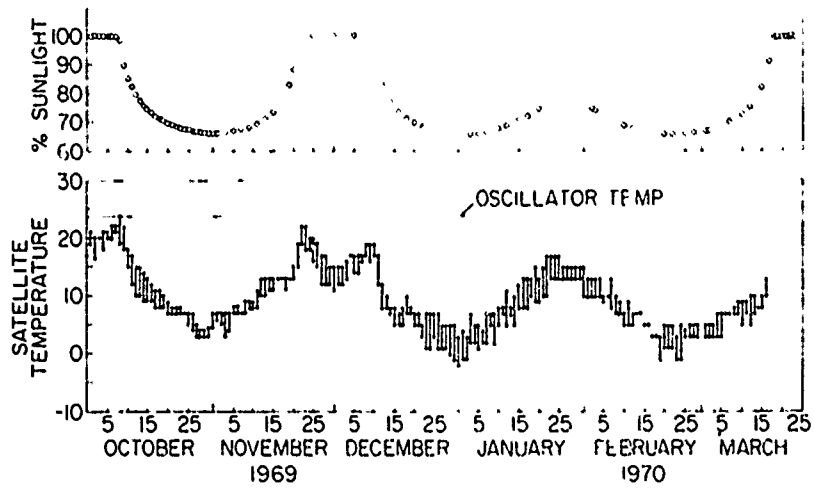


Figure 4 Satellite and Oscillator Temperature and Percent Sunlight versus Time

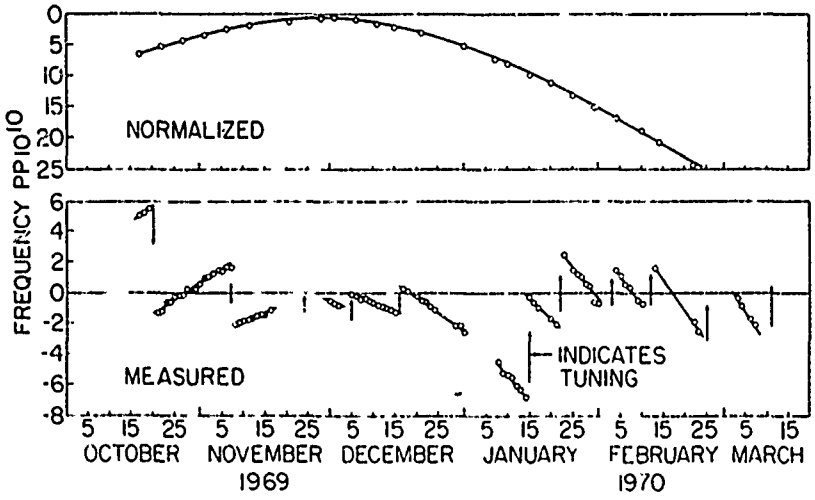


Figure 5 Oscillator Frequency versus Time

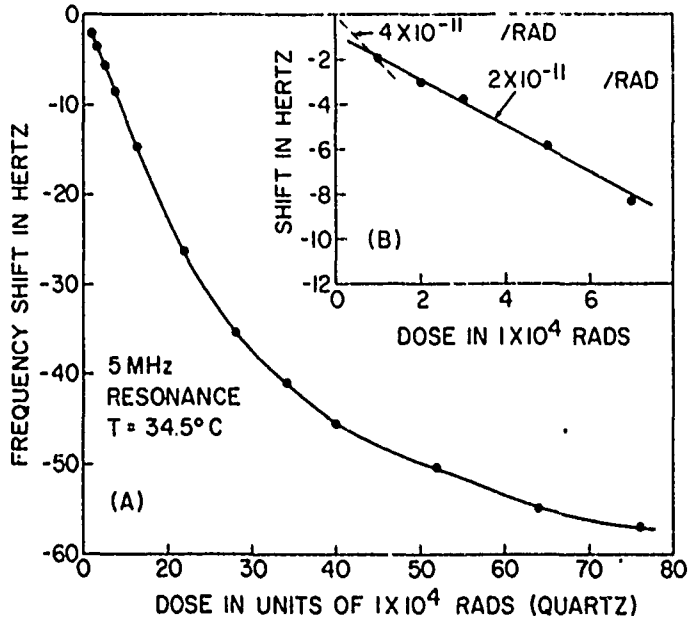


Figure 6 Quartz Crystal Frequency versus Radiation Dosage

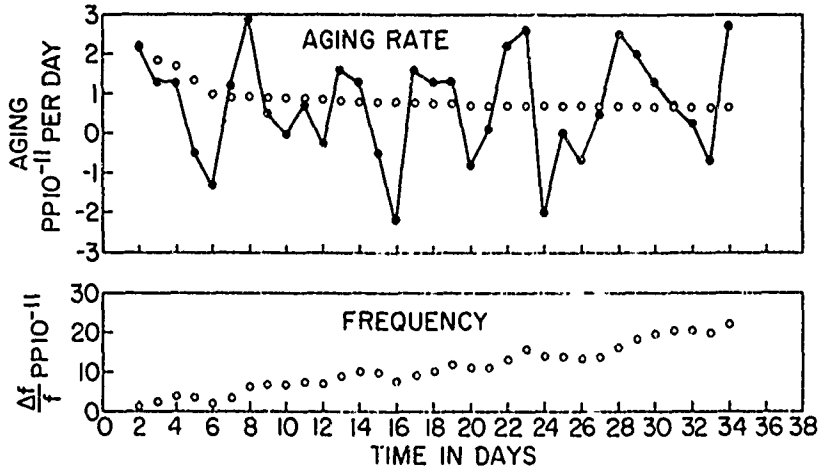


Figure 7 Aging Rate and Frequency versus Time

THE OMEGA NAVIGATION SYSTEM AS A SOURCE OF FREQUENCY AND TIME

Winslow Palmer, PE
68 Lucinda Drive
Babylon, NY 11702

Summary

Because it necessarily provides a highly-stable world-wide radio signal field synchronized to UT, with a format adapted to ready determination of the phase of the signals of at least three VLF Radio Stations at every place on earth, the OMEGA Radio Navigation and Position Locating System is inherently also a world wide source of Standard Frequency and Time of exceptional quality.

There is NO essential conflict between the use of OMEGA for position locating and as a source of Time and Frequency. The same receiving equipment can be utilized for either purpose without modification.

As presently being implemented, the OMEGA Signal Format can be tracked to yield a continuous 24 hour-a-day Standard of Frequency and Time anywhere on earth, with a phase variation of the order of a few microseconds RMS or better; but does not, however, yield an unambiguous epoch of time without initial setting.

A completely unambiguous time service could be provided in the OMEGA Signal Format, with minimal impact on all present users of Navigation equipments, by lowering the OMEGA RF frequencies by 0.75% (0.0075 per unit) and adding two additional components to the Multiplexing Pattern.

Introduction

All radio navigation systems are inherently also sources of Frequency and Time.

For an area coverage radio navigation system to function, it must provide highly stable synchronized signals from at least three more or less widely spaced locations, which can be received throughout its coverage area; with a signal format adapted to ready determination of relative phase or timing with a minimum of ambiguity.

These characteristics are precisely those required of a radio source of Standard Frequency and Time.

The OMEGA Navigation System is to be a world wide radio navigation or radio position-locating system in which the "Phase Field" of the signals or eight VLF (10 - 14 kHz) radio transmitting stations, distributed around the world in a more or less uniform net, as in Figure 1, provide a world wide grid of stationary iso-phase contours as position coordinates whereby the location of any point on earth can be identified.

For the Phase-Field of the OMEGA Signals to be stationary so as to give a stable and precise indication of position:

- The signals of the several stations must be synchronized to within a very few microseconds
- The Format of the signals must be designed to allow the signal phase to be determined readily and without significant ambiguity, and
- The signals of at least three (preferably four) of the eight stations must be receivable at every point on earth.

As a purely positioning device, there is no requirement that the OMEGA signals conform to any particular scale of Standard Time, since for position locating it is only necessary that the components of the signal format system be synchronous among themselves. However, to guarantee the stability, reliability and phase continuity of the OMEGA signals, the overall timing of the OMEGA Signal Field is to be derived from the concensus of thirty-two Atomic Frequency Standards (four per station at eight different locations distributed more or less uniformly around the world) all operating under essentially ideal conditions of stable environment and undisturbed operation, with careful maintainance by trained crews dedicated to achieving the utmost in stability and continuity of operation.

These inherent characteristics of the OMEGA Signal Field:

- o Ultra Stable Frequency
- o Ready determination of phase with no significant ambiguity
- o Signals of three or more stations receivable everywhere on earth

are precisely the qualities required of a World Wide Standard of Frequency and Time.

System Status (1970)

There are presently (1970) four OMEGA stations in operational status. Station A (Norway); B (Trinidad); C (Hawaii) and D (New York, USA)

Construction of a station in North Dakota, USA to replace the station in New York has been started, and it is understood that negotiations are underway for the installation of the four other stations to complete the world wide network.

Possible, (not official) locations of the other four stations might be southern South America (Chile or Argentina), Japan, New Zealand or Tasmania and La Reunion Is.

Power

Each station is to radiate a signal at a power level of 10 KW, such that at any place on earth at least three (3) and possibly up to five (5) of the eight stations can always be received at a useable signal to noise ratio under the worst conditions of natural atmospheric noise and ordinary man-made interference.

Signal Format

The format of the signals emitted by one OMEGA station is a recurring pattern of three successive one-second transmissions of pure continuous wave (CW); consisting of a one second (more or less) transmission at 10.2 kHz; followed by another one-second transmission at 13.6 kHz and then a third at 11 1/3 kHz; repeating every ten seconds, with gaps of 0.2 second between transmissions and six to seven seconds between groups, as illustrated in Figure 2. Each transmission is a segment of a continuous smooth sinusoidal wave of that frequency. That is, if extrapolated over the ten-second interval between transmissions at that frequency, each transmission would fall precisely in phase with the preceeding and following transmissions from that station at that frequency, as sketched (not to scale) in Figure 3.

Interlaced Transmission Pattern

The eight stations of the OMEGA network are to transmit their segments at each frequency in turn. The total of all transmissions at all three frequencies form an interlaced pattern of twenty-four one second transmission on three frequencies, as in Figure 4.

All transmissions in a given segment of the sequence are of the same length, but the length of transmission varies from segment to segment of the pattern, to form a repeating length-coded pattern identifying individual segments within the sequence. The lengths of the successive segments in seconds, are:

A	B	C	D	E	F	G	H
0.9	1.0	1.1	1.2	1.1	0.9	1.2	1.0

with a gap of 0.2 second between transmissions, for a total of exactly 10 seconds for the full sequence.

Signal Synchronization

For the Phase-Field of the OMEGA Signals to be stationary, the transmissions of all stations must be synchronous to within a fraction of a microsecond. To achieve and maintain system synchronization to this precision, in the face of transmission uncertainties and signal outages, etc., the transmissions of each station are timed by the consensus of four (4) Atomic Frequency Standards, and four (4) independent cross-checking signal format generators, providing very precisely timed, stable and reliable driving signals for the transmitters, for which the probability of loss of phase continuity is essentially negligible.

A system monitor associated with each station determines, more or less continuously, the phase difference between the signals of that station and the signals of other stations within range. The phase difference information from all monitors is then collected at a central location, and analysed to determine the state of the system, including

- o phase retardation over the distance between each pair of stations

- o the deviation of the phase of each signal from the mean phase of the system (corrected for retardation)

- o phase difference between the system mean phase, and Standard Time.

From time to time, the phase and frequency of each station is then adjusted so as to minimize its deviation from the mean of the system; and all stations are adjusted to steer the mean phase of the system to Standard Time.

Because of the extreme stability of the station timers ($1 : 10^{11}$ or better) the observations of phase difference with respect to any one signal can be restricted to the most stable transmission periods for that signal; and phase differences observed during periods of sun-spot activity, magnetic storms, and other propagation disturbances can be and are ignored.

At the present writing (1970) the evidence seems to be that the residual systematic synchronization error between stations is of the order of one microsecond or less; with a random variation of a few nano-seconds.

Relative Phase of the Omega Field Components.

The timing of all components of the OMEGA Signal Field is controlled such that every RF component as radiated from every station (if extrapolated to a continuous sinusoid) passes through zero with positive slope concurrently with the beginning of the first or "A" Segment of the multiplexing sequence, exactly on each minute and half minute of Standard Time.

Each segment of the multiplexing sequence (Figure 3) begins on an exact one-tenth (0.1) Second epoch of Standard Time. That is, the A Segment always begins exactly on an even ten second epoch of standard time. A B Segment then begins exactly 1.1 seconds later, A C Segment starts 2.3 seconds after each even 10 seconds, etc; the complete sequence being

A	B	C	D	E	F	G	H
0.0	1.1	2.3	3.6	5.0	6.3	7.4	8.8

The 10.2 kHz and 13.6 kHz components, being exact harmonics of 10 Hz, pass through zero with positive slope at each 0.1 second epoch, and hence have a zero coincident with the beginning of every segment of the multiplexing sequence. Unfortunately (for a time signal that is) the 11 1/3 kHz component not being a harmonic of 10 Hz, is a maverick that does not begin every segment of the sequence at zero phase as do the 10.2 and 13.6 kHz components. However, since its period is commensurate with 0.3 second, the 11 1/3 kHz component does pass through zero with positive slope concurrently with the start of each multiplexing segment at least once every three sequences, ie., 30 seconds, viz

A	D	F	B	C	E	G	H
0.0	3.6	6.3	11.1	12.3	15.0	17.4	28.8

The three RF frequency components, as transmitted from each station are controlled so as to have a triply concurrent zero (if extrapolated to continuous sine waves) as in Figure 5, exactly on the minute and half minute of Standard Time. The same relative phase (triply coincident zero) repeats at the least common multiple of the individual RF periods, or every 882+ microseconds; ie., 17 periods per 15 milliseconds. The triply coincident zeros of the three RF components also coincide with the beginning of every segment of the multiplex sequence at least once in every three sequences (30 seconds)

Thus, the OMEGA Signal Field is a high-stable multiply-periodic function of time whose various components subdivide the flow of time into sequences of identifiable periods commensurate with Standard Time, viz

10.2 kHz	51 Periods per 5 Millisec
13.6 kHz	68 Periods per 5 Millisec
11 1/3 kHz	34 Periods per 3 Millisec
Triple Zeros	17 Periods per 15 Millisec
Multiplex Sequence	1 Period per 10 Seconds

Propagation Considerations

The OMEGA Signal Field is a highly stable multiply periodic function, whose various components subdivide the flow of time into a set of related sequences of identifiable periods, whereby the phase of one or more components of the field at the instant of any event of interest

characterizes the time of the event.

The timing of the signals is controlled such that the instantaneous phase of every component of the signal being radiated from each station is an exact representation of Standard Time, as radiated.

Out in the field at any distance from the stations, however, the phase of the OMEGA signal is retarded with respect to the phase at the station by the time required for the signal to traverse the intervening distance by some 6 microseconds per nautical mile. With the transmission distances of OMEGA, where only eight stations provide a signal field encompassing the entire earth, the delay or retardation can be as much as 0.03 second, and has a systematic diurnal variation up to several hundred microseconds.

According to contemporary theories, the propagation of radio signals at VLF can best be described mathematically in terms of "Waveguide Mode Fields" in a "Waveguide Duct" bounded by the surface of the earth and the lower boundary of the ionosphere. In these terms the signal field at any distance is the resultant of one or more "Mode Fields" whose phase difference or "retard" relative to the signal as transmitted, is a more or less linear function of the distance from the transmitter, or source.

The rate of change of phase with distance, or "retardation" of a waveguide mode field depends upon the width of the waveguide - in this case, the height of the ionosphere. This in turn depends on the intensity of ionization by solar energy and varies as a function of many variables, particularly with the zenith angle of the sun, - but to a lesser degree with latitude. The retardation also depends on the direction of propagation relative to the magnetic field of the earth, and the electrical properties of the boundary surface.

The magnitudes of these effects are substantial - the principle factor being the diurnal change in ionosphere height introducing a phase change of up to two or even three cycles of phase at 1012 kHz, ie., 200 to 300 microseconds over some distances. Fortunately, the effects are largely systematic, and are accounted for as follows.

Retardation Prediction.

The total phase retardation of the OMEGA Field at any point can be separated into a large constant part - equal to the retardation of a surface wave traveling the geodesic distance from the station to the point of observation over sea water; and a much smaller systematic time variable part dependent upon the locations of transmitter and receiver, and the composition of the path between; and varying with time of day and season.

The large constant part is obtained by computing the geodesic distance from the geographic positions of transmitter and receiver, corrected to the geoid by standard (eg., Andoyer-Lambert formulae) methods, expressing the distance in equivalent electrical time units. The systematic time-variable part for the given location, date and time can be obtained from established OMEGA skywave Correction Tables (HO 224) published by the U.S.Navy Oceanographic Office (NAVOCEANO)

Diurnal Phase Variation

The magnitude of phase change between night and day is a function of the distance, appearing as a diurnal variation of the phase, generally trapesoidal in shape. The phase varies slowly about a more or less constant minimum value for the duration of daylight over the path between transmitter and receiver, It rises to a maximum night-time value as the evening terminator sweeps across the signal path, changing by up to several cycles of phase (i.e., several hundred microseconds) over the longer path lengths. It varies slowly about a more or less constant maximum value during the night and returns to the daytime value as the morning or sunrise terminator sweeps the path.

The rate of change of phase during the morning and evening transitions depends on the station bearing relative to the terminator great circle, as well as the total change between day and night. Where the station lies more or less to the east or west so that the line of bearing is perpendicular to the great circle of the terminator, the transition phase may last for three or four hours, as the terminator traverses the distance from the transmitter to the receiver. Where the station is to the north or south, sunrise and sunset occur more or less simultaneously over the entire path; and the transition phase lasts for only a half hour or so.

At high latitudes, there may be no transition during some seasons when all day or all night conditions prevail continuously over the entire path.

Phase Prediction

Fortunately, the variation is quite systematic, suprisingly so for such a complicated function of so many variables. At any given location, the signals of e each station repeat the same phase at the same time of day with an RMS variation of only a very few microseconds, in accordance with a systematic function of time of day and season of year.

A mathematical model of ionosphere electrical benavior has been developed, containing some thirty to forty terms - some polynomial, some trigonometric - accounting for the many systematic influences affecting the apparent velocity of propagation of VLF signals. A regression analysis to establish undetermined coefficients of this model has been performed with respect to many thousands of observations of the diurnal pattern of phase shift, taken over the past two decades at many places around the world, for all times of day and seasons of the year.

Experience in navigating by the OMEGA system with later versions of this model of ionospher electrical behavior appears to indicate that the model will usually predict the diurnal variation at any given location, time and date, to within a very few microseconds.

OMEGA Skywave Correction Tables

For use in navigation and position locating by OMEGA, the US Navy Oceanographic Office (NAVOCEANO) publishes a set of "OMEGA Skywave Correction Tables" HO Pub 224., giving predictions of the diurnal variation to be expected throughout the coverage area for every hour of each day of the year. Adding the tabulated data to the phase of the signals as observed, reduces the indicated values to what they would have been if the propagation velocity of the signal were the assumed "Ground Wave Phase Velocity" used in preparing the OMEGA Charts.

The Tables consist of a number of volumes, each volume giving predictions for one year for one OMEGA Frequency over a part of the OMEGA Coverage.

Each Volume of the Tables is arranged with each pair of facing pages displaying four separate tables (one for each of the present four operational stations) giving the corrections for an area of four degrees latitude by four degrees longitude. Each Table is a rectangular matrix of diurnal shift data, in per cent of a cycle at 10.2 kHz, arranged in twenty-four rows, one row for each semi-monthly period of the year (eg., 1-15 Jan; 16-31 Jan; 1-15 Feb; 16-28 Feb; etc.) and twenty-four columns, one column for each hour of Greenwich Mean Time.

The expected diurnal shift for any location, date and hour of the day can thus be obtained, by turning to the Table for the given area, within ± 2 degrees of latitude and longitude, and taking out the per-cent of a cycle, plus or minus, at the intersection of the row for the date (within the current 15 day period) and the column for the present hour of GMT.

During all daytime and all night-time conditions over the propagation paths, no interpolation is required, the tabular interval being less than the accuracy of prediction. During sunrise and sunset transitions, interpolation may improve the accuracy of the predictions by a distinguishable amount.

Time from OMEGA Signals

In theory the Standard Time of an occurrance of any event could be determined by observing the instantaneous phase of the several components of the OMEGA field at that instant, and applying the appropriate conversions and corrections. In practice however, radio signals are always immersed in noise and interference to the extent that it is usually difficult, if not impossible, to observe the instantaneous phase of a signal with any accuracy. Thus, Standard Time is always obtained by providing a LOCAL TIME SCALE or CLOCK whose instantaneous phase characterizes the timing of events; with a means for determining the mean phase difference between Local and Standard Time, over some reasonable period sufficient to minimize errors due to noise fluctuations etc., to a tolerable level. Such a system is as diagrammed in Figure 6, a "Tracking Filter" consisting of a local clock function of some kind, a source of remote or Standard Time, some means for determining the phase difference between the local clock indication and the Standard, and some means of applying a correction to either the local clock itself, or to its indication. Methematically such systems are "Low Pass Filters" providing a smoothed representation of the noisey standard, a representation whose phase can be observed more readily than that of the noisey standard.

Superheterodyne Tracking Filter

In many applications of Standard Frequency and Time, it is only a standard of Frequency that is of interest, ie., it is only the local time-difference between events that is important - absolute standard time is immaterial. Where this is the case, a precise standard of frequency can be derived from the OMEGA Field, in Hertz and decimal multiples thereof, by means of a Superheterodyne Tracking Filter, as shown in Figure 7

In this embodiment, a 1 MHz Frequency Standard and a set of decimal frequency dividers, provide a local clocking function, in decimal multiples of Hz; including in particular, a 10 kHz sine wave; a 200 Hz sine wave and a one Second pedestal every 10 seconds.

An RF amplifier tuned to 10.2 kHz picks up the 10.2 kHz components of the OMEGA Signals, which are mixed with the 10 kHz clock signal to produce a 200 Hz IF Signal. The 200 Hz difference signal is further amplified and compared in phase with the 200 Hz local clock signal. The phase error thereby obtained is applied to control the frequency of the 1 MHz Frequency Standard.

The 1 second switching pedestals, at 10 second intervals commutate the input to the frequency control, so that control is derived from the signals of one station only.

With a minimum of circuitry, this system tracks the phase of the 10.2 kHz component of one station, which can be any of the eight stations which happens to deliver a good signal to the receiver location. With a reasonably good local clock (frequency standard) and by limiting operations to that portion of the day when transmissions are most stable; a frequency calibration of $1 : 10^{10}$ or better can be realized with little difficulty.

With a truly precision frequency standard (e.g., atomic or rubidium) capable of maintaining phase within ± 50 microseconds from one day to the next; this system will deliver a frequency calibration over a period of several days, approaching that of the OMEGA System itself, ie., $1 : 10^{12}$ or better

Although the superheterodyne tracking filter tracks the phases of the incoming signals on a 1 : 1 time-shift basis (meaning a 1 usec shift of the incoming signal produces a 1 usec. shift of the local clock) it does not yield a reliable epoch of time, because of the 51 fold internal phase ambiguity inherent in the superheterodyning tracking filter system.

If a Time Epoch must be maintained, the local clock must operate at some multiple of the signal frequency, so as to avoid internal ambiguities in the system.

Audio Time Signal

A simple three or four transistor radio with a BFO or regenerative detector, as in Figure 8, tuned to one of the OMEGA Frequencies (10.2, 13.6 or 11 1/3 kHz) will provide an audio tone signal whereby a watch or chronometer, once set, can be corrected to within a fraction of a second of UT at any time of the day or night, anywhere on earth.

The signal of each station picked up by such a receiver produces a one-second audio tone every ten seconds, whose beginning falls exactly on a one-tenth second epoch of GMT, as in Figure 2 and 3. Thus, by marking the OMEGA Epochs (time each segment begins) on the seconds scale of a watch or chronometer, as in Figure 9, its error, to within a fraction of a second can be determined at any time, anywhere on earth, by simply noting the position of the second hand at the beginning of each OMEGA Signal tone from the receiver.

It is of course necessary to be able to recognize the signal of a particular station which can be accomplished in a number of ways. First, a rought estimate of relative station distance will indicate which station provides the stronger signal. Secondly, with even a slight acquaintaince with the pattern of OMEGA signals, a subconscious familiarity with the cadence of the length coding sequence is developed, whereby the start of each sequence is immediately apparent.

It is also necessary to know GMT to within plus or minus five (5) seconds, in order to eliminate the 10 second ambiguity, which is well within the capability of contemporary mechanical time pieces of even moderate quality, if corrected daily. On the other hand, if the local epoch is lost, any source giving time to within ± 5 seconds (radio time signal, or even a sextant observation of lunar distance) can re-establish the epoch, which can then be corrected as above to a tenth second or so.

Thus, this technique of audible reception of the OMEGA signals provides a means whereby a chronometer can be corrected to GMT with more than sufficient accuracy for celestial navigation; any time, day or night and anywhere on earth,

with nothing more than a simple transistor radio adapted to receive the OMEGA signals.

Automatic Sequence Tracking System

Many contemporary OMEGA Navigator's equipments include some form of automatic sequence-tracking function, to keep the multiplex commutating function aligned in phase with the incoming sequence of signals. Such tracking systems are generally similar to that shown in rudimentary form in Figure 10, consisting of one or more RF amplifiers tuned to the OMEGA frequencies, and a detector and signal combining function, producing a "best representation" of the signal envelopes.

A local clock and frequency divider system produces a stable 10 Hz clocking signal, which drives a "Correlating Function Generator" producing a "local correlating waveform". Cross-correlation of the local waveform with the signal envelope produces a null type error signal, that is applied to control the frequency (phase) of the clock, so as to keep the local correlating function aligned in phase with the sequence of incoming signals.

Systems of limited sophistication may require initial signal acquisition to me accomplished by manual adjustment of local time; from whence they acquire phase lock and track the incoming signal sequence phase within some tolerance.

By the use of computer technics, more sophisticated systems are possible having the capability of acquiring initial phase alignment automatically, in one ten-second sweep of the signal sequence.

Present indications are that these envelope tracking systems are capable of tracking the signal envelope phase with a resolution bettern than one (1) millisecond (1000 microseconds)

The absolute accuracy is of course dependent upon how well the retardation over the path from the transmitters can be predicted. This can be as much as 30 milliseconds, and can apparently be predicted to bettern than a milliseconds, so that simply tracking the phase of the signal envelope function yields standard time to a millisecond or better.

Readout is ambiguous beyond the ten second period of the multiplex sequence. To establish the epoch, the timing must be set initially from some other source to be correct to within ± 5 seconds. Once set however, the automatic tracking function maintains the epoch within a millisecond or better, indefinitely.

Phase Tracking

An epoch of Standard Time can be maintained to within a few micro - seconds, (with an ambiguity of plus or minus 441 microseconds) by tracking the RF phase of the OMEGA Signal Field, which can be accomplished, with an OMEGA Navigator's equipment, with no modification of the equipment whatever, by the system of Figure 11.

In principle, the method is to provide a local clock signal in a format matching the emissions of one of the (more distant?) OMEGA Stations, and use the OMEGA Navigators equipment to indicate the phase difference between the local signal and the phase of OMEGA Signal Field.

The lower part of Figure 11 includes a 1.02 Megahertz Frequency Standard or Clock, and three frequency dividers dividing the clock frequency by 100, 75 and 90 respectively, to produce the three OMEGA frequencies, 10.2, 13.6 and 11 1/3 kHz. Additional logic circuitry re-clocks the 90/1 divider to ensure that all three counter outputs cross zero together at the appropriate common frequency (1133 1/3 Hz) and provides gating waveforms to pass a one second block of each RF component in succession every ten seconds, so as to generate the signal format of one OMEGA station (Figure 2)

The output of this local OMEGA Signal Generator is added to the input of an OMEGA Navigator's Receiver, which is adjusted to read the phase difference between the local signal, and the signals of one or more of the stations of the OMEGA Network.

The combination of the three RF components establishes a sequence of local time epochs, with a resolution of a microsecond or better, and a period of 882+ microseconds (1133 1/3 Hz)

The phase readout of the OMEGA receiver indicates the phase difference between the local time scale and the OMEGA signal field, which, when corrected for retardation over the distance from the station, and for diurnal variations (by HO 224) yields Standard Time to an accuracy of a microsecond or so, and an ambiguity of ± 441 usec.

Standard Time by Signals from Three Stations

By observing the phase differences between the signals of three or more stations and a local timer, as in Figure 11, Standard Time can be maintained without precise knowledge of the actual location of the receiver, by a process very similar to the reduction of a three-star fix in celestial navigation. An outline of the steps involved are as follows.

1. Assume a Geographical Position for for the receiver, in latitude and longitude, more or less in the vicinity of the actual location.

2. Compute, by standard methods of celestial navigation (eg., HO 211 or HO 214 et al.) the bearing and distance of each station from the Assumed Position. Correct the distances so obtained to Geodetic Distance (Andoyer-Lambert or equivalent formulae) and convert to electrical units (Units of phase retardation, eg., per cent of a cycle etc)

3. Look up the Skywave Corrections (HO 224) for the Assumed Position, Date and Time of Day. Subtract (repeat subtract, no add) the corresponding correction for the distances.

4. Compute the Mean Retardation (Mean Corrected Distance) and the Deviation from the Mean expected of each signal at the assumed Position.

5. Observe the Phase of the Signals being received from the three stations with respect to the local time signal (Figure 11) and adjust the timing of the local signal, such that the mean phase with respect to the three signals being received is zero. Record the phase differences ϕ_1, ϕ_2 and ϕ_3

6. Plot the Assumed Position on a Mercator Plotting Sheet, and lay off Lines of Bearing to the three stations. as in Figure 12.

7. Determine the "Intercept Distance" by which the distances to the Assumed Position must be increased or decreased to agree with the observed phase offsets from the mean, viz

$$\text{Intercept} = \phi - (d - \bar{d})$$

8. Plot a line of position perpendicular to each line of bearing, offset from the assumed Position by the corresponding intercept distance and extend the three Lines of Position to form a triangle, as in Figure 12.

9. Inscribe a circle in the triangle bounded by the three position lines.

The Mean Retardation of the local timer with respect to common Standard Time at the Stations, is then exactly the electrical equivalent of the Mean Distance of the Assumed Position (Step 4) plus the radius of the inscribed circle, and the center of the circle is the actual position of the receiver.

Resolution of RF Ambiguity

The overall format of the presently operational OMEGA signals, does NOT (repeat NOT) provide an un-ambiguous epoch of time. With presently available equipmrnts it is only possible to obtain a scale of Standard Timew with a period of 882+ microseconds (17 periods per 15 milliseconds) and a resolution of a microsecond or so; and a time scale with a period of ten (10) seconds, and a resolution of about 1 millisecond.

In both cases, resolution of the epoch ambiguity must be obtained by reference to some other source of Standard Time. In the case of the 10 second period, this is within the capability of ordinary mechanical time pieces, by reference to broadcast time signals, by a portable clock brought up from some permanent source of standard time, or even by astronomic observationd (lunar distance). Resolution of the RF pattern ambiguity requires a source of Standard Time with a resolution of ± 441 microseconds, which is well within the capability of a portable quartz crystal clock, particularly if more or less continually reset by reference to the OMEGA phase pattern.

Once an epoch is established, it is maintained indefinitely by reference to the OMEGA Signal Field, at any location on earth

Means for resolving these ambiguities could be added to the OMEGA Signal Format with no interference whatever with the OMEGA Navigation function, by including other one-second transmissions at additional frequencies in the segments of the multiplex pattern where the stations are now silent. For example, adding a fourth component of 12 750 Hz (12 3/4 kHz) to the transmissions of each station, increases the unambiguous period of the composite RF phase by a factor of four

to 3.53 milliseconds, and adding a fifth component at 12 112.5 Hz extends t e unambiguous period by another factor of four to 14 Milliseconds, which it is reasonable to resolve with the 1 millisecond epochs deriveable from the multiplex sequence phase.

It is unfortunate that the presence of the 11 1/3 kHz component in the OMEGA format, makes the common period of the RF components non-commensurate with even seconds, so that the fundamental period of the entire OMEGA Format is 30 seconds not 10 Seconds.

A commensurate pattern could be obtained, with possibly minimal impact on present navigators equipments, by reducing the OMEGA Frequencies by 0.75% (0.0075 per unit) while keeping the same ratios of 9 : 10 : 12 obtaining frequencies of 10.125, 11.250 amd 13.5 kHz.

Additional components at 12.0 and 10.8 kHz then yields a common period of 13 1/3 milliseconds (75 Hz) This is an exact factor of one second, so that the pattern then repeats exactly i.. each multiplex period.

Conclusions

1. The essential characteristics of the signal format of a world wide radio navigation system, viz

- o Signals receivable from at least three stations anywhere on earth
- o Highly stable periodic signal format providing for ready determination of relative phase and reduction of ambiguity
- o Compensation for vagaries of signal transmission

are precisely the characteristics required for a world wide radio source of Frequency and Time

2. As a Navigation and Radio-Locating Facility, the OMEGA Navigation System, when fully implemented will provide a world wide radio signal field with a highly periodic signal format, whereby anywhere on earth, the relative phase of the signals from at least three of the eight stations can be determined to within a few microseconds with more or less complete reduction of ambiguity.

3. The signal field of the OMEGA System, being timed by the concensus of thirty-two Atomic Frequency Standards, in continuous operation in controlled environments under supervision of trained crews dedicated to maintaining the highest standards of reliability and accuracy of synchronization; is inherently a universally available source of Frequency and Time of exceptionally high quality.

4. An OMEGA Navigation Receiver of contemporary design can be used without modification to determine the phase lag of a local frequency standard and time code generator with respect to the OMEGA Signal Field, thereby providing a correction of the local time code to Standard Time.

5. No other source of Frequency and Time of like quality is, or is likely to be, available on a world wide basis.

6. The only conflict between OMEGA as a radio locating or navigation facility, and as a source of Frequency and Time, is between the need of OMEGA and its lane ambiguity resolving functions to operate in the natural scale of the assigned frequencies; while it is deemed preferable that a time service be compatible with even seconds of time.

7. Conflict between OMEGA as a radio locating facility and as a source of Frequency and Time is more apparent than real, since Decimal Time in Seconds can be obtained from OMEGA Phase by division by a constant.

8. The present OMEGA Signal Format does NOT provide a complete Time Service, in that it does not, of itself, provide information giving total time elapsed since any recognized epoch of Time Zero.

9. The present Operational OMEGA Signal Format of only three RF frequencies transmitted in a ten-second multiplex sequence doew NOT provide an unambiguous time scale, even within the ten second multiplex period.

11. A more desireable signal pattern for both OMEGA and Time Service would be obtained by shifting frequency from 10.2 to 10.125 kHz; with additional frequencies of 10.8, 11.25, 12.0 and 13.5 kHz; to obtain a pattern commensurate with the 10 second multiplex pattern, so as to eliminate the 3 : 1 ambiguity inherent in the format including 11 1/3 kHs; and provide for complete ambiguity resolution within the 10 second multiplex period.

11. An indication of Total Time from a recognized time zero could be included in the signal format by gap-coding the signal transmissions, with negligible interference to the radio-locating function.

12. The Phase of the OMEGA Signal Field (OMEGA Time) is a retarded version of Standard Time at the stations, the phase of the signals being retarded by the time required for the radio signal to traverse the distance from the transmitter to the point of reception.

13. Present OMEGA Skywave Correction Tables (HO 224) provide information f for compensating diurnal variation of OMEGA Phase to within a few microseconds.

14 There is NO compelling reason for Standard Time to be other than Continuous Atomic Time. The differences between Atomic Time and Mean Solar Time are totally un-noticeable in the ordering of the daily affairs of mankind. All highly precise scientific, engineering and commercial uses of Time are always adjusted for the particular use. It is more important that the recognized Scale of Standard Time be a smooth continuous function, thant that is Correspond to any particular physical phenomena.

RECOMMENDATIONS

1. It be recognized that the essential characteristics of the Signal Format of a world wide radio navigation system and radio locating system are precisely the characteristics required of a world wide source of Standard Time and Frequency.

2. It be recognized that the Signal Field of the OMEGA System is inherently a universally available Scale of Frequency and Time of exceptionally good quality.

3. The policy of step adjustment of OMEGA Time to follow GMT be dropped, and the program of supervision and maintainance of the OMEGA Station Timers be modified to provide, as nearly as can be realized, within the state of the art, an absolutely smooth continuous linear time scale

4. The PHASE of the OMEGA SIGNAL FIELD as transmitted be recognized officially as a World Scale of Standard Time.

5. Developement of the OMEGA System as a world wide source of Frequency and Time be continued, including the developement of OMEGA Clock Systems specifically designed to yield Standard Time to the high est accuracy attainable in the state of the Art.

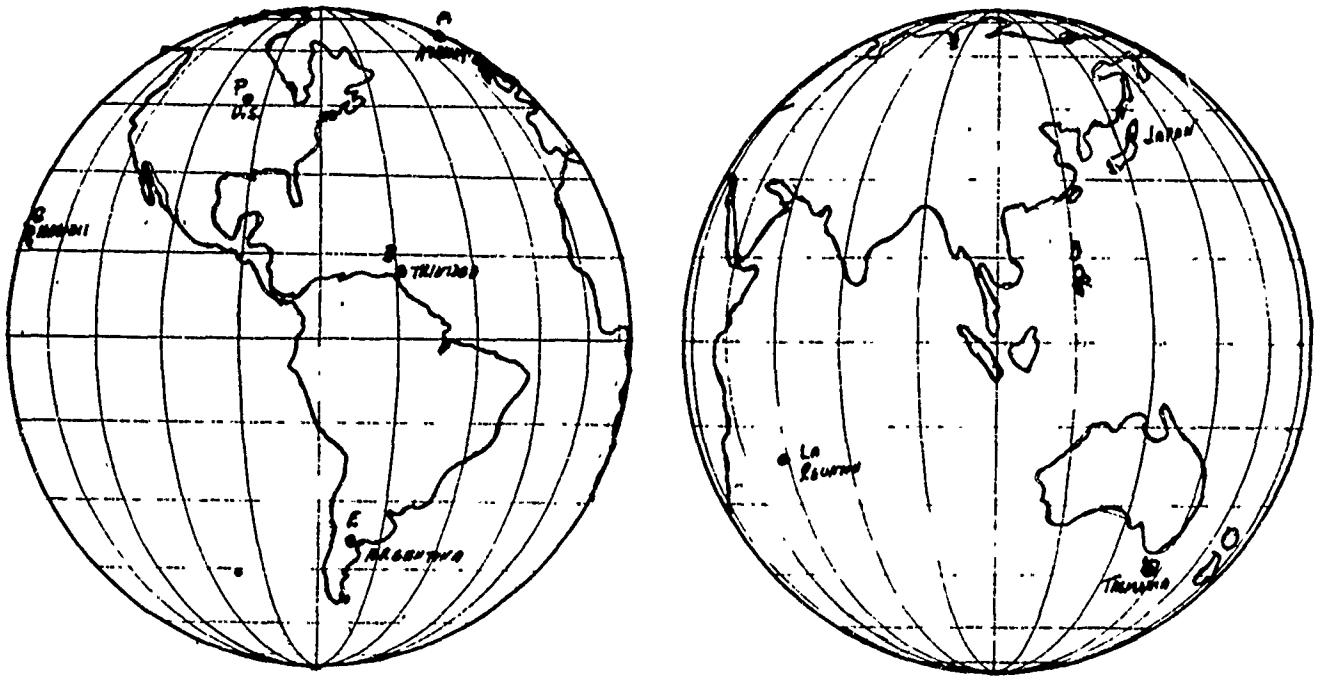


FIGURE 1 World OMEGA Network

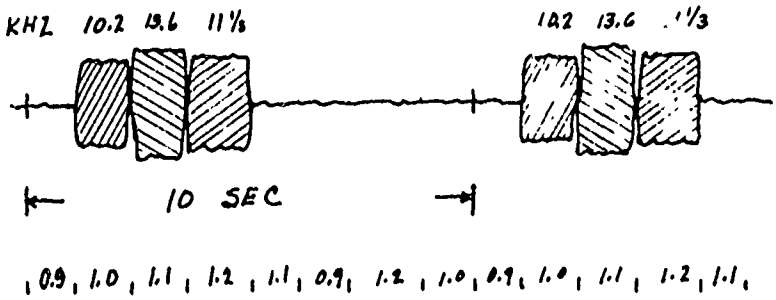


FIGURE 2 OMEGA Signal from One Station

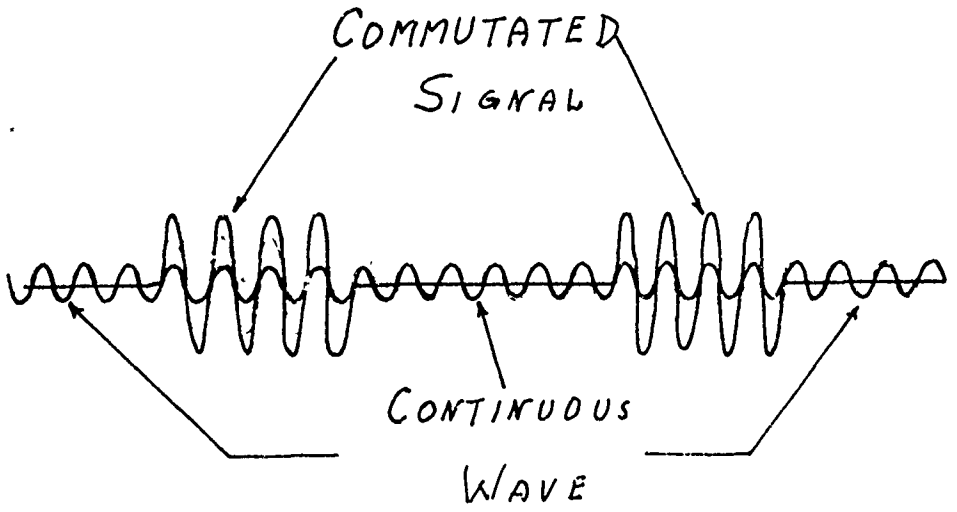


FIGURE 3 Extrapolation to Continuous Wave

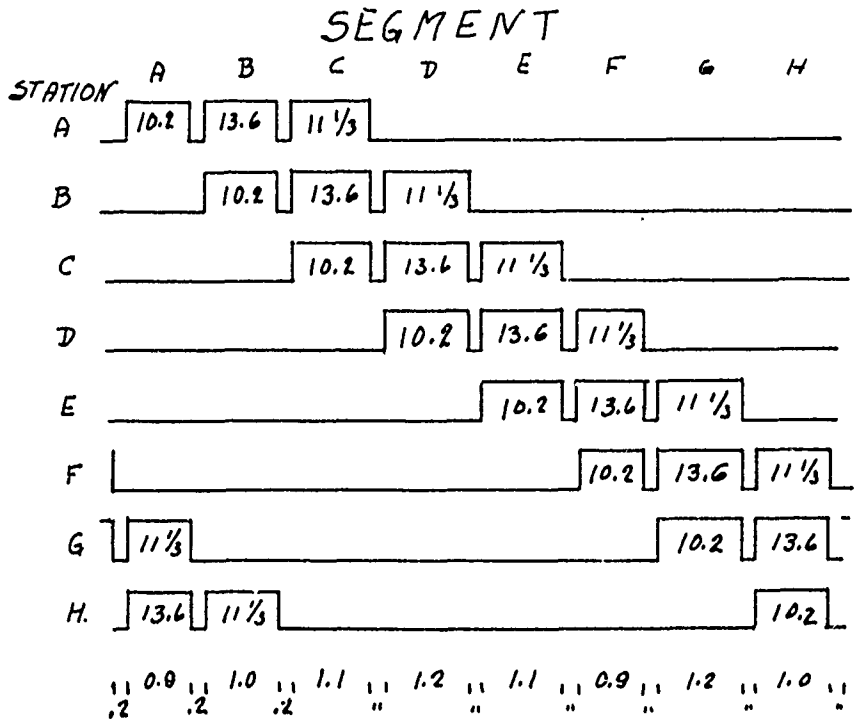


FIGURE 4 OMEGA System Signal Format

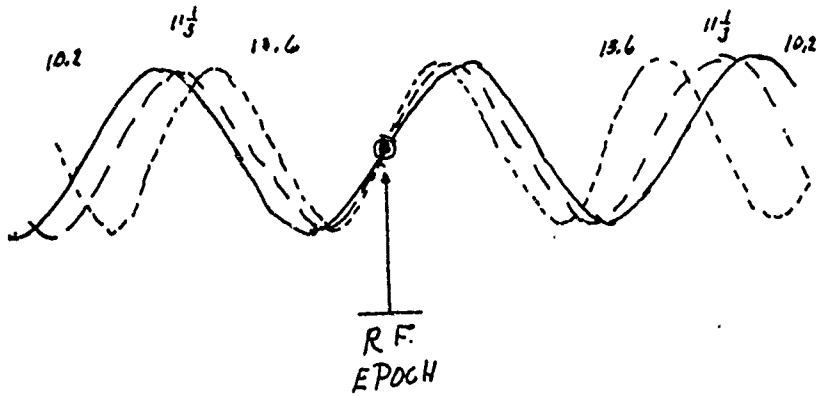


FIGURE 5 RF Epochs

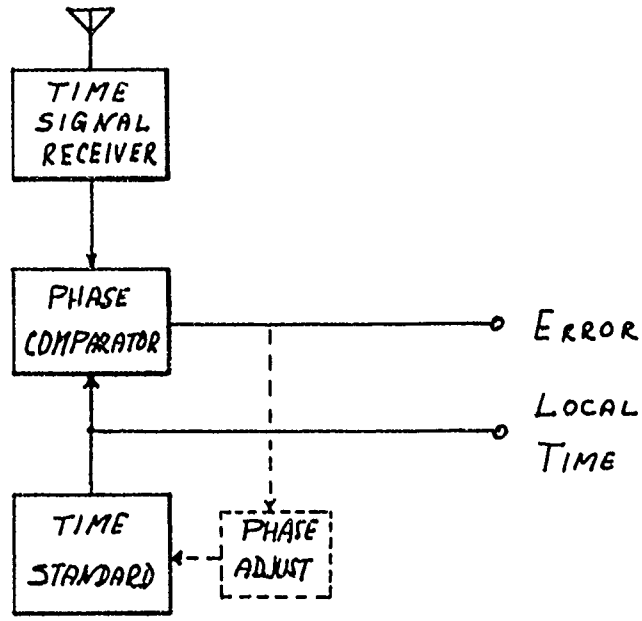


FIGURE 6 Rudimentary Remote Clock System

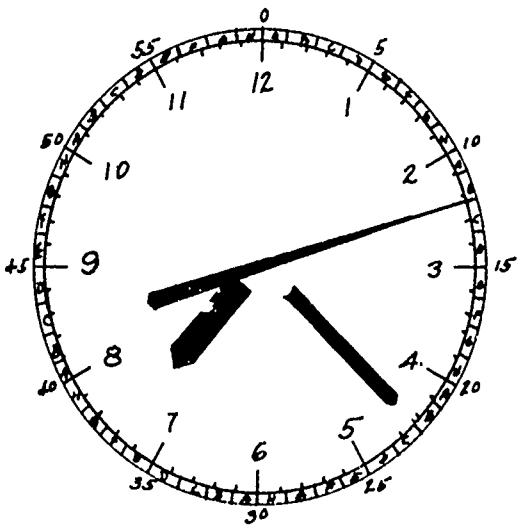


FIGURE 9 Clock Face

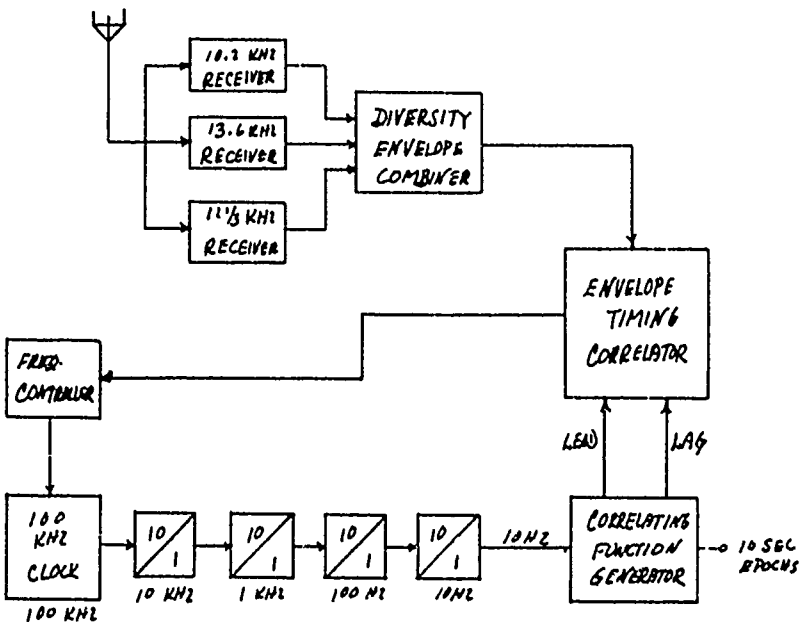


FIGURE 10 Envelope Sequence Tracking System

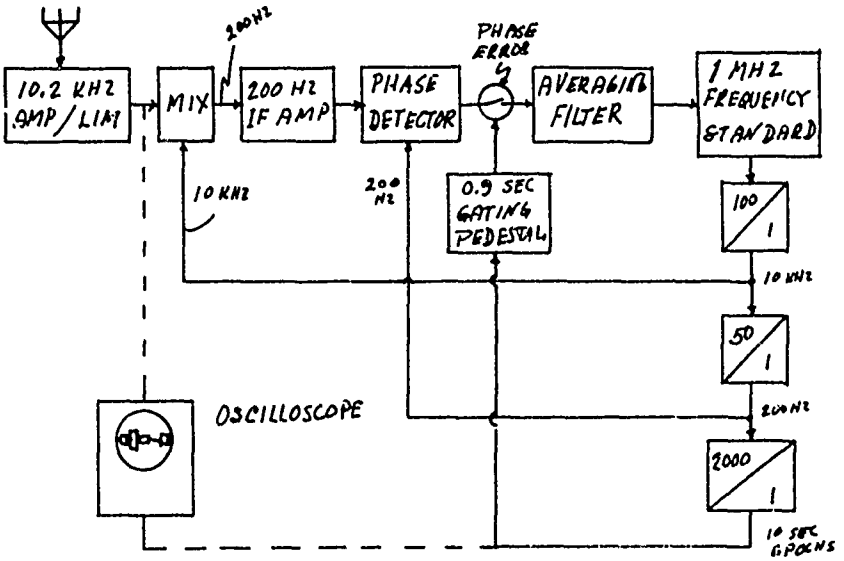


FIGURE 7 Superheterodyne Tracking Filter

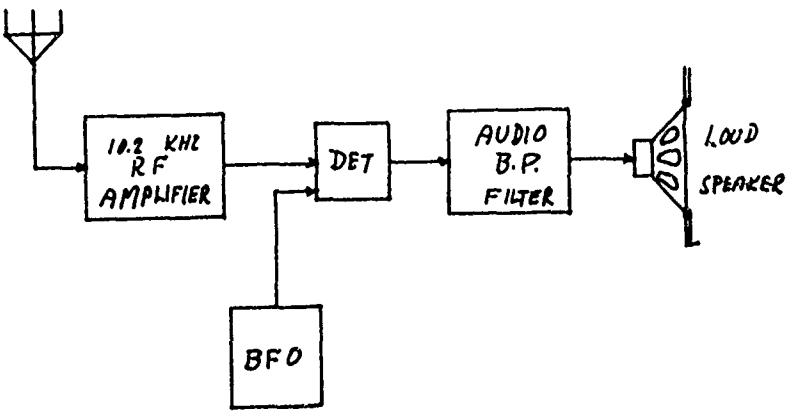


FIGURE 8 Audio OMEGA Time Signal Receiver

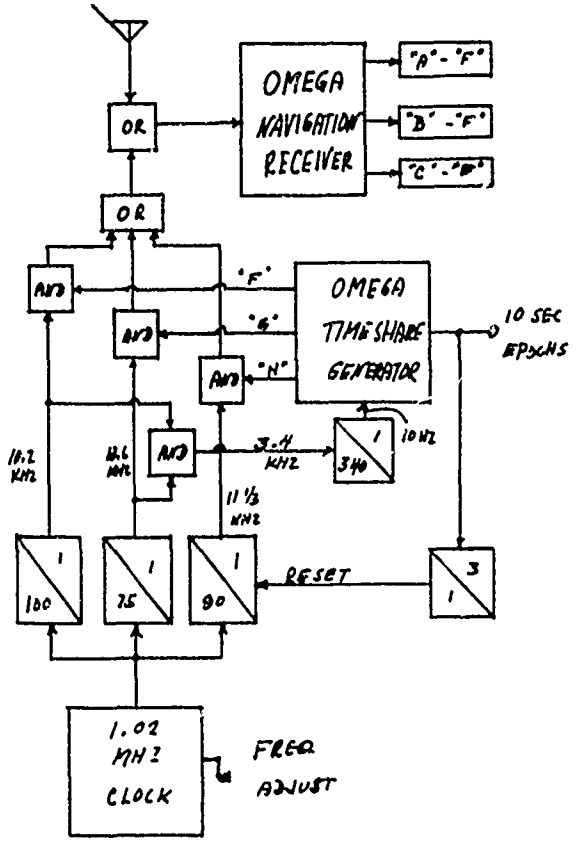


FIGURE 11 OMEGA Navigators Receiver as OMEGA Clock

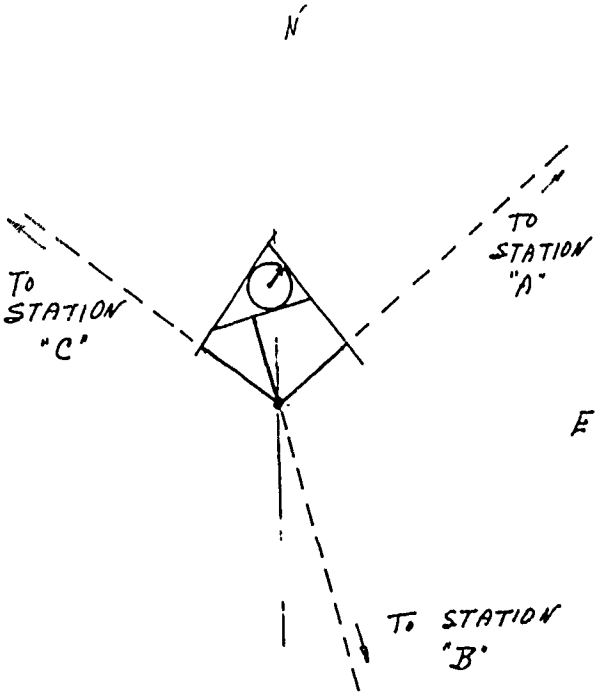


FIGURE 12 Mean Time of Three Signals

THE NBS ATOMIC TIME SCALE SYSTEM: AT(NBS), SAT(NBS), AND UTC(NBS)

David W. Allan, James E. Gray, and Howard E. Machlan
Atomic Frequency and Time Standards Section
National Bureau of Standards
Boulder, Colorado 80302
Tel. (303) 447-1000, ext. 3208

Summary

The Atomic Time Scale at NBS Boulder has its rate determined by periodic calibration with the NBS primary frequency standard, currently NBS III. Because of the increased demand for accurate and precise timing, a study was made of the clock ensemble which generates the time scale AT(NBS) to determine the precision and accuracy with which time and frequency may be generated.

The accuracy involved both the 5 part in 10^{13} uncertainty in NBS III and the frequency dispersion of the ensemble after calibration. To determine the precision capabilities, noise models were established for the time dispersion characteristics of each clock in the ensemble. Some theoretical considerations based on these models yielded very encouraging predictions on both how to process the time data from each clock and on the precision capabilities of the ensemble.

A computer program incorporating the above ideas was written with the capability to self-optimize the statistical weight given to each clock in the ensemble. The program also determined the clock time, frequency, and precision of each member of the ensemble. Reasonable agreement was achieved between the predicted results and those measured.

The accuracy of the second used for AT(NBS) time scale is estimated to be within 9 parts in 10^{13} (1 sigma limit) of the internationally agreed upon definition employing the cesium atom. The internal estimate of the precision capabilities of the AT(NBS) time scale gives the following equation for the time dispersion:

$$\epsilon_{rms} = [k_1 \tau + k_2 \tau^3]^{1/2}$$

where $k_1 = 5.3 \times 10^{-22}$ seconds, $k_2 = 3 \times 10^{-37}$, and where ϵ_{rms} and τ are the possible root-mean-square time error accumulation (1 sigma limit) in seconds and the elapsed time in seconds respectively.

SAT(NBS) and UTC(NBS) are coordinated time scales, i.e., controlled by international regulation. In addition a cooperative effort between the United States Naval Observatory and the National Bureau of Standards has led to a synchronization of the times of the UTC(USNO) and UTC(NBS) time scales to within 5 microseconds.

* * *