1962 IRE International Convention Record

PART 3

Sessions Sponsored by

IRE Professional Groups on

Electron Devices

Microwave Theory and Techniques

at

the IRE International Convention, New York, N.Y.

March 26-29, 1962

The Institute of Radio Engineers

1962 IRE INTERNATIONAL CONVENTION RECORD

An annual publication devoted to papers presented at the IRE International Convention held in March of each year in New York City. Formerly published under the titles CONVENTION RECORD OF THE I.R.E. (1953 & 1954), IRE CONVENTION RECORD (1955 & 1956), and IRE NATIONAL CONVENTION RECORD (1957, 1958, & 1959).

Additional copies of the 1962 IRE INTERNATIONAL CONVENTION RECORD may be purchased from the Institute of Radio Engineers, 1 East 79 Street, New York 21, N.Y., at the prices listed below.

Part	Sessions	Subject and Sponsoring IRE Professional Group	Prices for Members of Sponsoring Professional Group (PG), IRE Members (M), Libraries and Sub. Agencies (L), and Nonmembers (NM)			
			PG	M	L	NM
1	8, 16, 23	Antennas & Propagation	$ .70	$ 1.05	$ 2.80	$ 3.50
2	10, 18, 26, 41, 48	Automatic Control Circuit Theory	1.00	1.50	4.00	5.00
3	1, 9, 17, 25, 28, 33	Electron Devices Microwave Theory & Techniques	1.00	1.50	4.00	5.00
4	4, 12, 20, 34, 49	Electronic Computers Information Theory	1.00	1.50	4.00	5.00
5	5, 13, 15, 22, 29, 47, 54	Aerospace & Navigational Electronics Military Electronics Radio Frequency Interference Space Electronics & Telemetry	1.20	1.80	4.80	6.00
6	3, 11, 31, 35, 42, 45, 50, 52	Component Parts Industrial Electronics Product Engineering & Production Reliability & Quality Control Ultrasonics Engineering	1.40	2.10	5.60	7.00
7	30, 37, 43, 51	Audio Broadcasting Broadcast & Television Receivers	.80	1.20	3.20	4.00
8	7, 24, 38, 46, 53	Communications Systems Vehicular Communications	1.00	1.50	4.00	5.00
9	2, 19, 27, 32, 39, 40, 44	Bio-Medical Electronics Human Factors in Electronics Instrumentation Nuclear Science	1.20	1.80	4.80	6.00
10	6, 14, 21, 36	Education Engineering Management Engineering Writing & Speech	.80	1.20	3.20	4.00
		Complete Set (10 Parts)	$10.10	$15.15	$40.40	$50.50

Responsibility for the contents of papers published in the IRE INTERNATIONAL CONVENTION RECORD rests solely upon the authors and not upon the IRE or its members.

1962 IRE INTERNATIONAL CONVENTION RECORD

PART 3 - ELECTRON DEVICES; MICROWAVE THEORY AND TECHNIQUES

TABLE OF CONTENTS

HIGH FREQUENCY MICRO LAYER TRANSISTOR

R. L. Luce
J. D. McCotter
J. A. Sluss
C. G. Thornton
Philco Corporation, Lansdale Division
Lansdale, Pennsylvania

INTRODUCTION

During the past decade the precision etch process has been used in fabricating several different types of high frequency germanium transistors, including the SBT (Surface Barrier Transistor), MAT (Micro-Alloy Transistor), and the MADT (Micro-Alloy Diffused-Base Transistor). Each new design featured some new element in its structure to impart desirable properties not available in its predecessor.

A thin basewidth structure, first produced in the SBT, was made possible by development of the precision etch process. In the MAT, the addition of micro-alloy electrodes gave increased current handling capabilities without affecting the basewidth control. The use of diffused-base techniques in the MADT permitted significant reduction in electrical basewidth, with resulting gain-bandwidth design extending to 1.0 Kmc.

This paper describes methods and results of fabricating precision-etch PNP transistors having a thin homogeneous base region epitaxially grown on a collector substrate of opposite conductivity type. This new class of devices having a precision etched base layer of sub-micron width in conjunction with collector regions of opposite conductivity type is given the designation MLT, or Micro Layer Transistor.

COMPARISON OF DIFFUSED-BASE AND EPITAXIAL-BASE STRUCTURES

Figure 1 illustrates the diffused-base and epitaxial-base types of precision-etch germanium transistor. The diffused-base, or MADT type, is obtained by the controlled diffusion of an N-type impurity into a high-resistivity N-type substrate, followed by accurate placement of the two micro-alloyed electrodes. In the epitaxial-base structure, an epitaxial N-type base layer is grown on a suitable P-type substrate to form the base-collector junction. Precision etching techniques (the same as those used in MADT fabrication) are employed to obtain the desired thickness of the base and collector regions.

EPITAXIAL GROWTH SYSTEMS

The epitaxial systems used in the growth of the base layer are indicated in Figure 2. Good quality epitaxial layers are obtained by high temperature reduction of $GeCl_4$. Prior to epitaxial deposition, the thin residual germanium di-oxide layer must be completely removed by high-temperature vaporization. Hydrogen is not a suitable ambient for this process since the oxide would be preferentially reduced to amorphous germanium directly on the substrate surface. Therefore, a short high-temperature pre-treatment in an inert ambient is required prior to starting the growth process. The vapor-grown interfaces obtained show no evidence of lattice discontinuity when subjected to rigorous inspection techniques involving cross-sectioning and treatment with structure-sensitive etching solutions. To minimize the effects of diffusion, a short rapid-growth period is used.

The low-temperature disproportionation system with its negligible diffusion effects generally yields smooth single-crystal layers; however, the work to date on this system has not produced good electrical characteristics. Examination of etched cross-sections revealed crystal imperfections at the interface. Poor collector diode characteristics are attributed to these interface imperfections. Figure 3 displays the collector-base diode characteristics obtained by the high temperature process. Note the sharpness of the voltage breakdown on the expanded scale. This high quality collector-base diode is an epitaxially grown junction.

BENEFITS DERIVED FROM COMBINED EPITAXIAL AND PRECISION ETCH TECHNIQUES

Having established the techniques for producing good quality base-collector junctions by epitaxial growth, we now direct our attention toward the benefits to be realized from their use. First, let us consider the delay times associated with the diffused-base structure.

Delay Time Considerations in Diffused- Base Structures

Figure 4 shows the diffused-base structure and its associated carrier delay times. The total delay time for this type of transistor structure has been defined as the sum of three elements: 1) charge time of the emitter barrier capacitance, C_{TE}, 2) transit time through the base region, and 3) transit time through the collector depletion layer. In standard MADT structures designed for gain-bandwidth products up to 1.0 Kmc, the width of the electrical base region is the prime limiting factor in achieving minimum delay time. Calculations show that at higher frequency levels the transit time in the collector depletion layer is expected to become significant. The emitter barrier capacitance, C_{TE}, limits the device only at low current levels. In reducing the emitter area to obtain a lower C_{TE}, current density is increased, and a compromise is necessary. Reduction of emitter area may increase the current density to a point where the initial charge distribution in the base and collector regions is modified, resulting in shrinkage of the collector depletion layer and a consequent increase in the electrical basewidth. The widening of electrical basewidth with increased current density causes a decrease in f_t, the effect being similar to that experienced at very low levels of collector voltage.

Effects of Basewidth Decrease in Diffused-Base Structure

With reference to Figure 5, consider the effects of decreasing the electrical basewidth in the diffused-base device. The process of diffusion requires that the concentration of base impurities obtain a sufficiently large gradient at the diffusion front so that the doping atoms move into the substrate. As the basewidth is decreased by obtaining higher diffusion gradients, that is, by increasing the slope of the base impurity distribution the integrated amount of these impurities is reduced. Since an applied voltage is sustained by the integrated number of impurities, a reduction in dopant atoms in the base layer lowers the maximum voltage which can be applied without causing punch-through.

Compromise in Base Layer Thickness Avoided Through Epitaxial Techniques

The punchthrough compromise in base layer thickness obtained by a diffusion process can be avoided by a process which does not depend upon a concentration gradient to form the base layer. A thin base region that is essentially homogeneously doped is obtained when the dopant distribution is controlled during the process of epitaxially growing the base layer. This is the desirable feature of the short rapid growth of the base layer by epitaxial techniques.

Design for Minimum Transit Time Through Collector Depletion Layer

The final structure takes full advantage of the thin basewidth achieved by the epitaxial layer when the structure is designed to maintain a collector depletion layer transit that represents a minimum contribution to the total delay time. By depositing the base layer on a P-type substrate of suitable conductivity, the expansion of the depletion layer can be controlled to provide maximum frequency performance at practical levels of applied voltages, while maintaining a satisfactory breakdown voltage. Following upon the desired distribution of doping atoms, optimum control of the thickness of the base and collector regions is obtained by using precision-etch techniques.

TYPICAL MICRO LAYER TRANSISTOR

The photomicrograph of Figure 6 shows a cross-section of a typical epitaxial-base transistor, magnified approximately 300X. The diameter of the micro-alloyed emitter junction is 3 mils; collector diameter is approximately 7 mils. The distance between the emitter and the base-collector junction is 1/2 to 3/4 of a micron. The remaining distance between the collector junction and the ohmic contact is controlled to maintain the low series resistance that is characteristic of precision-etch transistors.

The high frequency performance of a typical MLT structure is recorded in Figure 7. Again, a schematic representation of the impurity distribution between the micro-alloyed contacts is included. In (b) of Figure 7, the small-signal current gain at 80 mc is plotted as a function of collector

voltage and current. This plot is for a typical epitaxial-base MLT enclosed and measured to the TO-18 package. A peak f_t of the 2.2 Kmc is obtained by extrapolation from this measurement. Measurements made at 200 mc also extrapolate to the 2.2 Kmc f_t value.

Figure 8 records D-C Beta as a function of collector current at low collector-to-emitter voltages. An additional indication of the quality of the epitaxial-base layer is demonstrated by the level and linearity of the D-C Beta as a function of current.

CONCLUSION

This paper has described incorporation of precision etching techniques in fabrication of epitaxial-base PNP transistors. Evidence has been presented to show that precise control of the base and collector regions results in desirable high frequency performance in the fabricated devices.

It is expected that further refinements of this basic design will lead to still higher frequency performance levels. The intent is to extend the described techniques to the fabrication of devices that will exhibit high frequency performance in high current and high power applications.

ACKNOWLEDGMENTS

The work described in this paper has been, in part, supported by an Air Force Contract under the administration of Lincoln Laboratories. The authors are pleased to acknowledge the assistance given by both Lincoln Laboratories and Philco Corporation personnel.

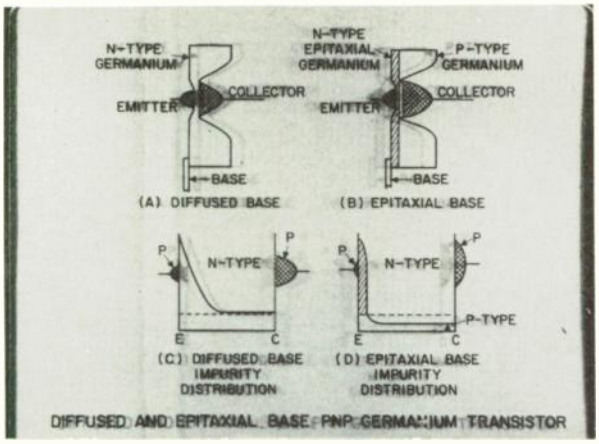


Fig. 1.

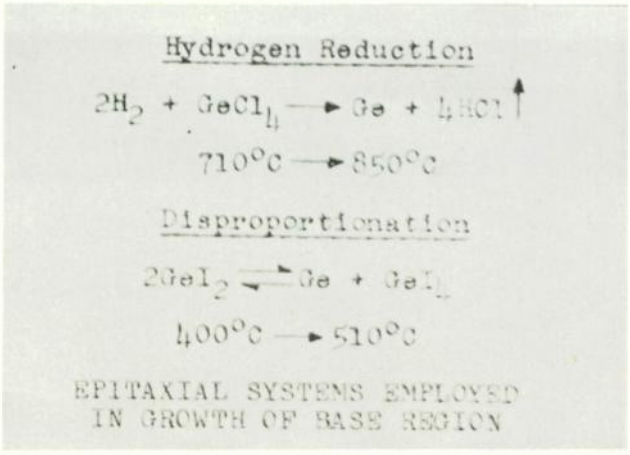
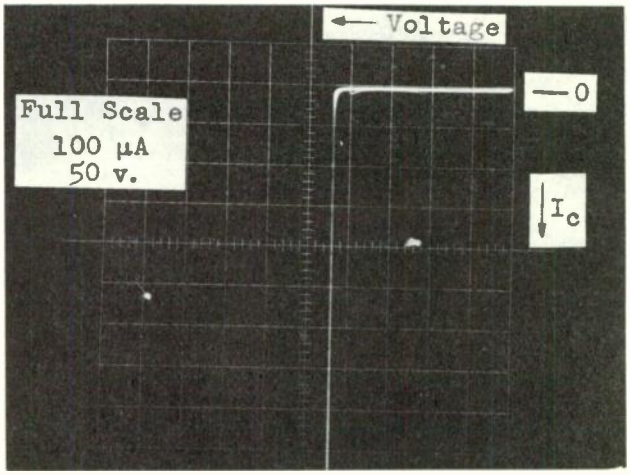


Fig. 2.



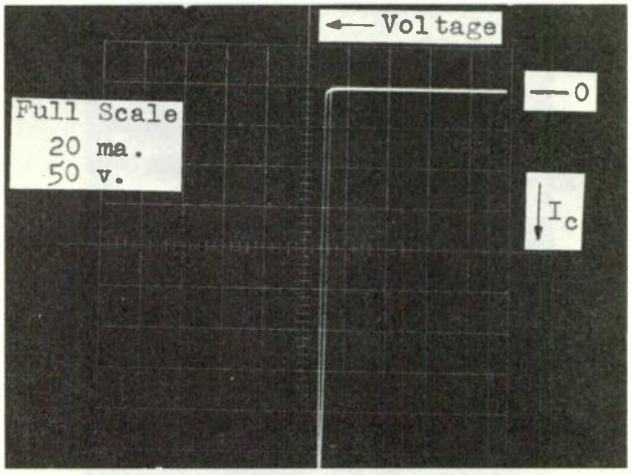


Fig. 3. Epitaxial collector-base junction.

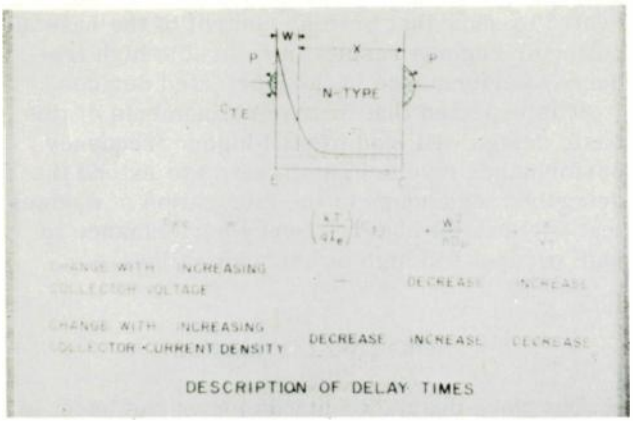


Fig. 4.

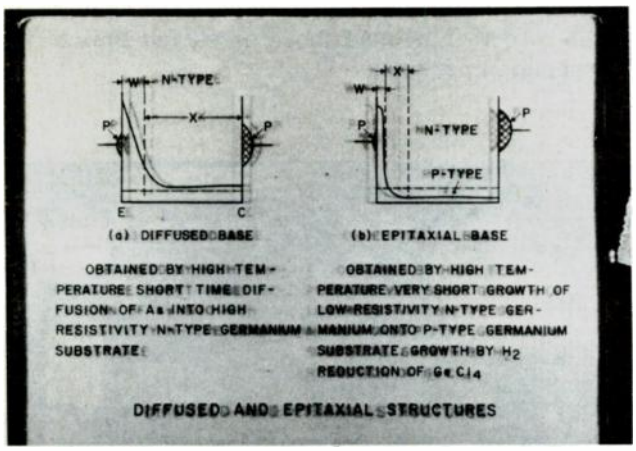


Fig. 5.

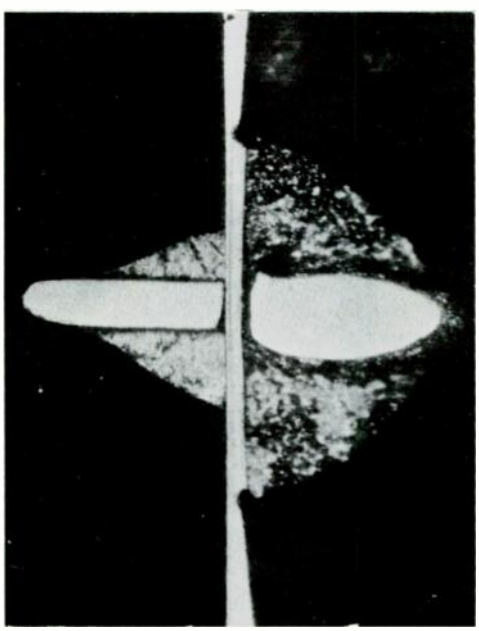


Fig. 6. Cross section of epitaxial base transistor (300X).

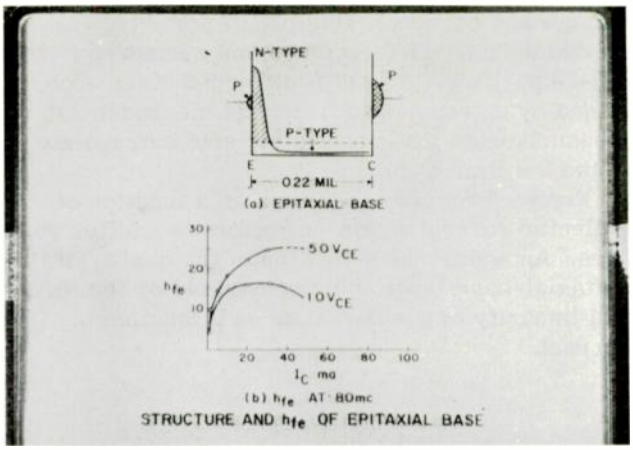


Fig. 7.

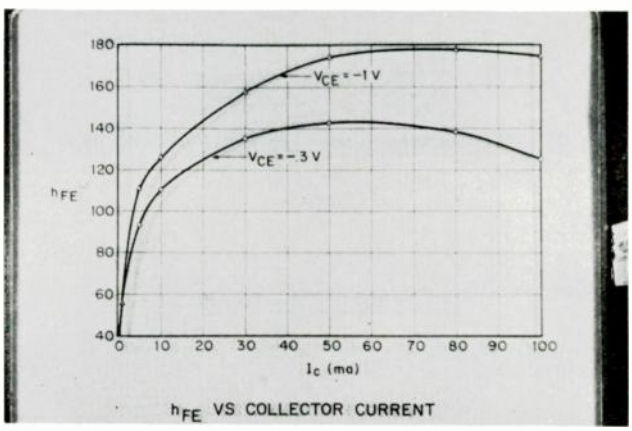


Fig. 8.

A NEW UNIJUNCTION TRANSISTOR STRUCTURE USING SPREADING RESISTANCE MODULATION

V. A. Bluhm, T. P. Sylvan
General Electric Company
Semiconductor Products Department
Syracuse, New York

Summary

A new design approach, using the modulation of the spreading resistance of a small area ohmic contact by injection of minority carriers from a nearby emitter junction, makes possible a unijunction transistor with small size and improved characteristics. Silicon unijunction previously described in the literature have been filamentary in structure with two large area base contacts and have utilized conductivity modulation of a large fraction of the total volume of the filament with a small area emitter junction located near the center of the filament.

The new device geometry permits a reduction of the distance between the emitter and base-one contact for a given value of standoff ratio and hence reduces the turn-on time of the device by a factor one hundred or more owing to the shorter carrier transport time between emitter and base-one. The closer spacing and smaller base-one area reduces the effective volume in which conductivity modulation takes place and thus reduces the emitter saturation voltage, the valley point current, and the peak point current. The lower emitter saturation voltage and valley point current permit bistable circuit applications which have been impractical with conventional unijunction transistors. The lower peak point current makes possible extremely sensitive voltage sensing circuits which require very small values of current for triggering. The lower peak point current combined with a low emitter leakage current makes possible time delay circuits which can operate with a very large time delay using a small timing capacitor.

Devices have been successfully fabricated using a pulse alloy technique for forming both the emitter junction and the base-one ohmic contact. The simplicity of this fabrication process makes it extremely advantageous from the standpoint of cost. Measured device parameters have shown a reasonably good correlation with a first order theoretical analysis.

Introduction

Since the first announcement of the UJT (unijunction transistor) in 1953[1] significant progress has been made in the improvement of its electrical characteristics and in the development of practical circuit applications[2]. The first UJT's, made of germanium, suffered the disadvantage of excessive temperature sensitivity, very low switching speed, and limited range of operating temperature. The availability of high lifetime, high resistivity silicon made possible the introduction of the silicon UJT in 1956. Introduction of the ceramic disc mounting structure in 1958 made possible tighter control of the geometrically dependent electrical parameters with a consequent improvement in the overall specifications. Present applications of the UJT include relaxation oscillators, timing circuits, voltage sensing circuits and firing circuits for SCR's (silicon controlled rectifiers). Although the characteristics of the present design are satisfactory for most of these applications, a significant reduction in switching speed, valley current and saturation voltage would be very desirable in that it would extend the range of present applications and open up new areas of application.

Conventional Structure

A cross-sectional diagram of a conventional UJT is shown to scale in Figure 1A. A ceramic disc having the same thermal expansion coefficient as silicon is used as a mounting platform. A gold antimony film is deposited on both sides of a 10 mil slot in the center of the disc. A single crystal n-type silicon bar having a resistivity of about 120 ohm-cm. and dimensions of about 8 x 10 x 60 mils is placed across the slot and two ohmic base contacts called base-one (B1) and base-two (B2) are formed by alloying between the silicon and the gold. A single p-type emitter (E) junction is formed by alloying a 3 mil diameter aluminum wire on the side of the silicon bar opposite from the base contacts. The emitter junction is generally located closer to base-two than to base-one. Symmetrical units are obtained if the emitter junction is an equal distance from base-one and base-two, but such devices do not have optimum electrical characteristics for most applications.

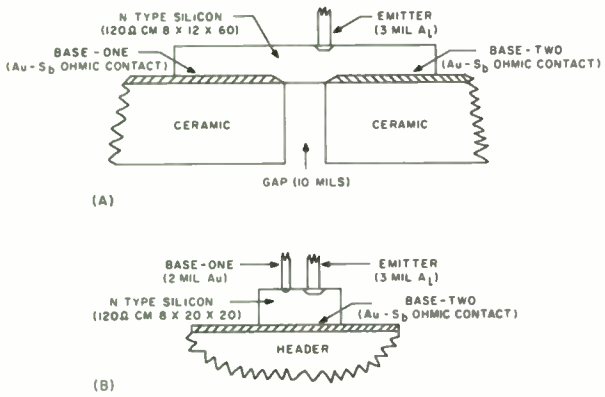


Fig. 1 Cross-sectional details of UJT Structures drawn to scale, (A) Conventional UJT Bar Structure, (B) New UJT Structure (ZJ-85).

Electrical Characteristics

The standard symbol for the UJT together with a static emitter characteristic curve are shown in Figure 2. The emitter curve is exaggerated in the vicinity of the V_E axis in order to show more clearly the details of the curve and the important measurement points. This characteristic is shown for a single value of interbase voltage (V_{BB} = 10v) although generally a complete family of emitter characteristic curves is presented for several fixed values of interbase voltage. To complete the description of the static electrical characteristics of the device, a family of interbase characteristic curves is required which is generally presented in terms of V_{BB} vs. I_{B2} for fixed values of I_E. However, most of the applications of the UJT are based on the emitter characteristics with base-two serving only as a biasing input and consequently the interbase characteristics will not be considered here in any detail.

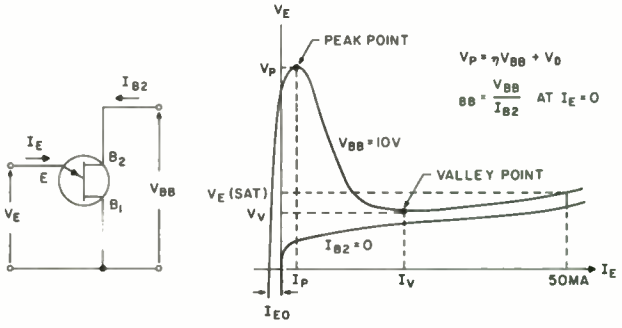


Fig. 2 Standard Unijunction Transistor Symbol With Nomenclature Used For Voltages And Currents; Static Emitter Characteristic Curves Showing Important Parameters And Measurement Points (Exaggerated To Show Details).

In the normal mode of operation, the base-one terminal is grounded and a positive bias voltage, V_{BB}, is applied at base-two as indicated in Figure 2. An interbase current I_{B2} will flow determined by the applied interbase voltage and the interbase resistance, R_{BB}. The interbase voltage establishes an electric field within the silicon bar and produces a voltage on the n side of the emitter junction which is a fraction, η, of the applied interbase voltage. The parameter η, called the intrinsic standoff ratio, is determined by the spacing between the emitter junction and the base contacts. If the emitter voltage, V_E, is less than ηV_{BB}, the emitter junction will be reverse biased over its entire area and only a small reverse leakage current, I_{EO}, will flow in the emitter circuit. If the applied emitter voltage exceeds ηV_{BB} by an

amount equal to the forward drop of the emitter diode, V_D, holes will be injected into the silicon bar. Because of the electric field within the silicon bar, these holes will move towards base-one and being minority carriers will increase the conductivity of the bar in the region between emitter and base-one. As the emitter current is increased, the emitter voltage will decrease because of the increased conductivity so that a negative resistance is observed between emitter and base-one. The essential features of the emitter characteristic curve are shown in Figure 2. The two important points on the characteristic curve are the peak point and the valley point. At these points the slope of the emitter characteristic curve is zero. The current at the peak point is called the peak point current, I_P. For a given unit, the peak point current is inversely proportional to the interbase voltage. The voltage at the peak point is called the peak point voltage and is related to the interbase voltage by the equation;

$$V_P = \eta V_{BB} + V_D \qquad (1)$$

where V_D, the equivalent emitter diode voltage is in the range of 0.6 to 0.7 volts for silicon devices. The current at the valley point is called the valley point current, I_V, and the voltage at the valley point is called the valley point voltage, V_V. For a given unit, both the valley point voltage and current increase with increasing interbase voltage. The emitter voltage at a high value of emitter current is called the emitter saturation voltage, $V_E(sat)$, and is generally specified at I_E = 50 ma, V_{BB} = 10v. The base-two current flowing under the same biasing conditions is called the base-two modulated current, $I_{B2}(mod)$.

The region to the left of the peak point is called the cutoff region, here the emitter junction is reverse biased. The region between the peak point and the valley point is called the negative resistance region, here conductivity modulation of the silicon bar between emitter and base-one is the predominant effect. The region to the right of the valley point is the saturation region. In the saturation region, conduction between emitter and base-one is limited by the surface and bulk recombination rates.

Design Objectives

The optimum UJT design objectives are determined by the particular application, however the following general rules can be given;

(1) The intrinsic standoff ratio should be variable over a range of at least 0.40 to 0.80, but should be capable of close production control.

(2) The emitter saturation voltage should be as low as possible, particular for applications in switching circuits since this parameter determines the switching efficiency of the device.

(3) The interbase resistance should be high enough to prevent excessive interbase power dissipation, but generally the precise value is not important.

(4) The peak point current should be as low as possible, particularly for applications in sensitive time delay circuits and voltage sensing circuits where trigger current sensitivity is important.

(5) The valley current should be designable over a wide range. A high value of valley current is required in oscillator circuits where it permits a high average emitter current or load power. A low value of valley current is required in oscillator circuits where it permits operation at low current and power levels.

(6) A low value of turn on time is required for pulse generators and timing circuits working at short time cycles.

The variables which are available to the device designer to achieve the above objectives include silicon resistivity, bar size, interbase spacing, and emitter location. The resistivity of the metal should be as high as possible to maximize the interbase resistance, but should be low enough to give stable parameters and permit good ohmic contacts to be made. Units have been made using silicon with a resistivity as high as 1000 ohm-cm, but for reasonable stability of parameters, it appears the upper limit should be about 200 ohm-cm. The minimum bar size and interbase spacing which can be achieved are determined by fabrication problems and required tolerances. The location of the emitter junction is determined by a

compromise between the emitter saturation voltage, the intrinsic standoff ratio, and the switching time. The emitter should be located as close as possible to base-one to minimize the emitter saturation voltage and the switching time, but should be located as close as possible to base-two to maximize the standoff ratio. For the bar structure, the value of intrinsic standoff ratio is given to a good approximation by the relationship;

$$\eta = \frac{d_{EB1}}{d_{EB1} + d_{EB2}} \qquad (2)$$

where d_{EB1} is the minimum distance between emitter and base-one and d_{EB2} is the minimum distance between base-two and the point on the emitter closest to base-one. The relationship between intrinsic standoff ratio and the emitter saturation voltage has been found to be given with reasonable accuracy by the following empirical relationship;

$$V_{E(sat)} = \frac{\eta - 0.33}{F} \qquad (3)$$

where F, the figure of merit, is a constant determined by the bar size, the slot width, and the surface and bulk lifetime. Equation 3 can be used to predict the effects that a change in emitter location will have on $V_E(sat)$ and η assuming that all other geometrical values remain fixed. A final compromise which must be made in the device design is the effective lifetime. The surface and bulk lifetime should be as high as possible to minimize the emitter saturation voltage, but should be low to minimize the turnoff time.

The relationships between the more important electrical parameters are summarized in Figure 3. As an example, it is seen from this chart that if R_{BB} is increased by increasing the silicon resistivity t_{on} and $V_E(sat)$ will increase and I_V and I_P will decrease. This chart indicates compromise that must be made when optimizing parameters for a given application.

From the preceding discussion of design objectives and design relationships, it is seen that the present UJT bar structure imposes well defined limitations on the electrical parameters. Once the minimum practical physical size is determined and values for η and R_{BB} are chosen, the values of all the other parameters are fixed.

	t_{ON}	I_V	I_P	R_{BB}	$V_{E(SAT)}$
η	↑	↑	↑	0	↑
$V_{E(SAT)}$	0	↑	↑	↑	
R_{BB}	↑	↓	↓		
I_P	0	↑			
I_V	0				

LEGEND
↑ POSITIVE CORRELATION
↓ NEGATIVE CORRELATION
0 NEGLIGIBLE CORRELATION

Fig. 3 Design Relationships Between Principal UJT Parameters

New Unijunction Transistor Structure

A unijunction transistor structure (designated ZJ-85) has been developed which uses a radical new geometry to achieve an order of magnitude improvement in some of the important electrical parameters of the device. A cross-sectional detail of one version of this new structure is shown in Figure 1B to the same scale as the conventional UJT bar structure in Figure 1A. The essential feature of the new structure is the use of a small area base-one contact which can be formed with a 2 mil diameter antimony doped gold wire pulsed or alloyed into the top surface of a 8 x 20 x 20 mil n-type silicon pellet. The emitter junction is formed by pulsing or alloying a 3 mil diameter aluminum wire approximately 3 mils away from the gold wire. The base-two contact is formed by a broad area gold-antimony alloy contact between the bottom of the bar and the transistor header. Photographs of the conventional UJT in a TO-5 header and the new UJT on a TO-18 header are shown in Figure 4. A cross-section photograph of the new structure is shown in Figure 5.

Because of the geometry, the voltage gradient in the vicinity of the base-one contact is much higher than elsewhere in the silicon, so it is possible to achieve a high standoff ratio with a much closer

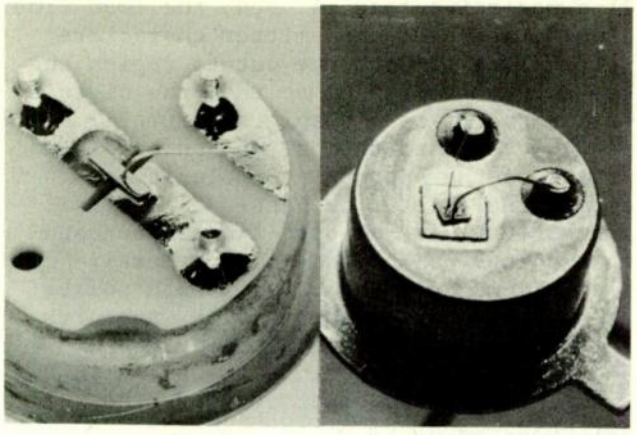


Fig. 4 Photograph of UJT Structures, left- Conventional UJT Bar Structure On TO-5 Header (.370" Diameter), Right - New UJT Structure On TO-18 Header (.225" Diameter).

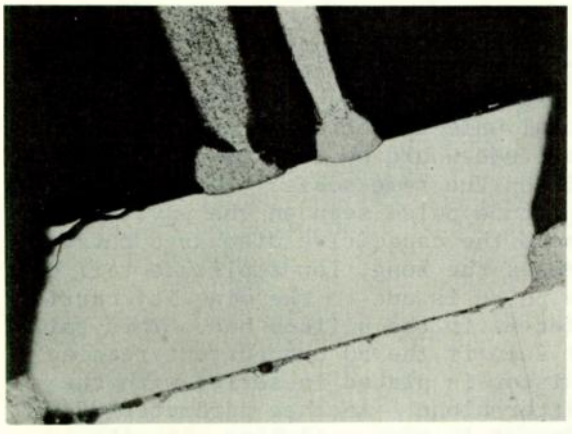


Fig. 5 Cross-sectional Photograph Of New UJT Structure

spacing between emitter and base-one than with the conventional structure. Assuming a hemispherical base-one contact of radius r, the intrinsic standoff ratio is given approximately by;

$$\eta = \frac{d_{EB1}}{r + d_{EB1}} \qquad (4)$$

where d_{EB1} is the minimum distance between the emitter junction and base-one. The interbase resistance is given approximately by;

$$R_{BB} = 63\ \rho \left(\frac{d-r}{rd}\right) \qquad (5)$$

$$\left(\text{error} < \frac{r}{d}\right)$$

where ρ is the silicon resistivity in ohm-cm, r is the base-one radius in mils and d is the pellet thickness in mils. The interbase resistance is determined chiefly by the spreading resistance of base-one contact and the thickness of the pellet has only a second order effect.

For a device having d_{EB1} = 2 mils, r = 1 mil, d = 8 mils, and ρ = 120 ohm-cm, the preceding equations give η = 0.66 and R_{BB} = 6.6K. These values are comparable to the typical values for the conventional structure but the emitter to base-one spacing has been reduced by a factor of 5 so that a significant reduction in the emitter saturation voltage and switching time can be expected.

The static emitter characteristic curves for a conventional UJT structure are shown in Figure 6 together with the emitter characteristic curves of a new UJT structure having the same value of standoff ratio. The lower saturation voltage and valley current for the new UJT structure are evident. The negative resistance and cut-off frequency for the new UJT structure are so high that it is difficult to stabilize the device in the negative resistance region. For high values of emitter voltage in the negative resistance region, the device will oscillate in the relaxation mode with its own case capacity.

The relationship between the intrinsic standoff ratio and emitter saturation voltage for both the conventional and new UJT structures are shown in Figure 7. Measured values on 35 standard production units of each type are given. It is seen that the conventional units follow the relationship given by Equation 3 with a figure of merit, F, between .08 and .10. In the case of the new structure however, the saturation voltage is practically independent of the intrinsic standoff ratio up to approximately η = 0.80. It is evident that the new UJT structure makes possible the achievement of higher values of standoff ratio with reasonably low values of emitter saturation voltage.

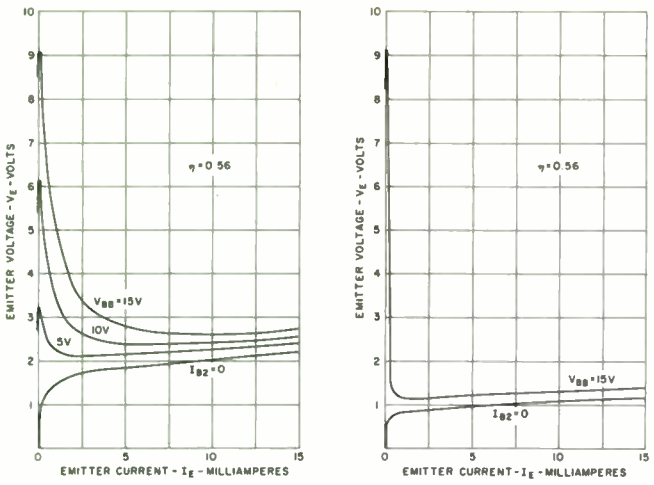


Fig. 6 Comparative Static Emitter Characteristic Curves; Left - Conventional UJT Bar Structure; Right - New UJT Structure.

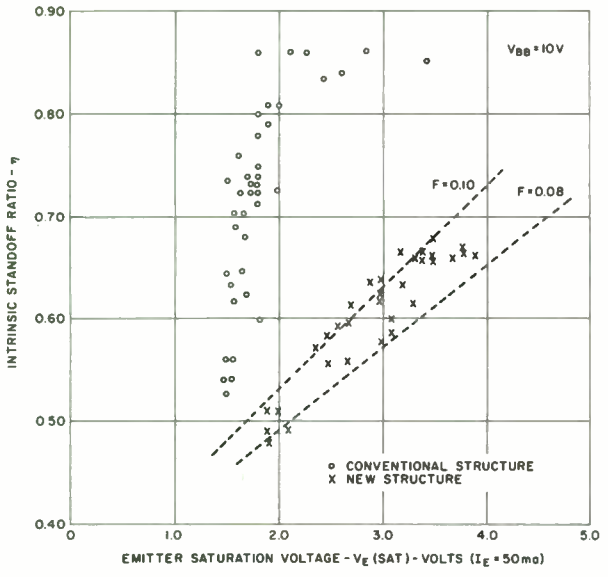


Fig. 7 Relationship Between Intrinsic Standoff Ratio And Emitter Saturation Voltage For Typical Lot Of Conventional UJT's And New UJT's.

Switching Speed

The switching speed of the UJT can be specified with either a resistive or capacitive load in the emitter circuit. If a resistive load is used, the load line should intersect the emitter characteristic curve in both the cutoff region and the saturation region so that the circuit will normally be bistable and it will be possible to turn the UJT on and off with pulses of the appropriate polarity at either the emitter or base-one terminal. A negative pulse at base-two can be used to turn the UJT on but it is generally not possible to turn the UJT off with pulses at this terminal. The turn on waveforms for both the conventional UJT structure and the new UJT structure are shown in Figure 8. The lower trace shows the switching waveform of the new UJT structure alone with the time scale expanded by ten times. The measured 10% to 90% risetimes were 4.0 μsec. for the conventional UJT and 50 nsec. for the new UJT. The rise time can be reduced by proper choice of the value of capacitance in parallel with the load. This capacitance provides a source to rapidly establish the charge concentration required in the emitter to base-one region of the UJT for steady state conduction. The base-one voltage pulse waveform for a new UJT switching a capacitive load is shown in Figure 9; the waveform for a conventional unit is not shown because the amplitude would be too low to be detectable on the same scale. The initial large amplitude pulse seen on the waveform is due to the capacitive discharge current, whereas the long, low amplitude tail on the pulse is due to the minority carrier lifetime in the silicon bar. This tail is not seen if the 50 ohm current reading resistor is placed in series with the emitter alone. Another parameter of interest in some switching circuits is the trigger charge. Measurements of trigger charge on the new UJT structure give a value of about 10 picocoulombs in comparison with a value of 1000 picocoulombs for the conventional structure.

The higher switching speed of the new UJT structure is the result of three factors; (a) for a given applied interbase voltage the average electric field strength in the region between emitter and base-one will be higher because of the smaller geometry, (b) for a given average electric field strength the transit time for both holes and electrons will be less because of the shorter distance between emitter and base-one, and (c) the volume of silicon in which conduc-

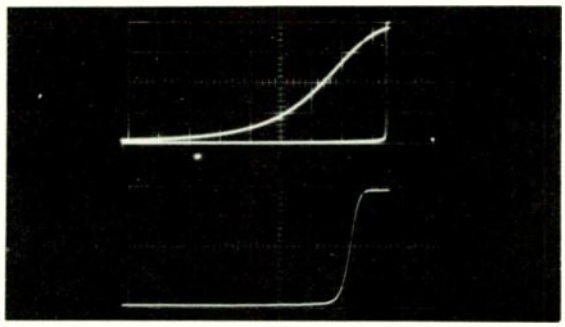


Fig. 8 Comparative Switching Waveforms For UJT Switching 500 Ohm, 200 pf Load In Emitter Circuit; vertical scale is emitter voltage 2.5 volts per major division. Upper Picture - Conventional Structure And New Structure; horizontal scale is time from right to left 1 μsec. per major division. Lower Picture - New Structure Alone; horizontal scale is time from right to left 100 nsec. per major division.

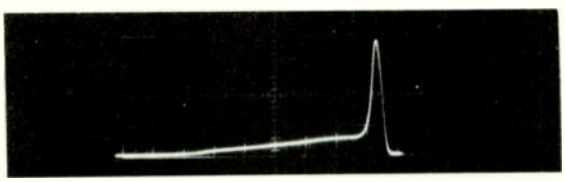


Fig. 9 Base-one Voltage Pulse Waveform Of New UJT In Relaxation Oscillator Circuit, Emitter Capacitor 200 pf, Base-one Resistor 50 Ohms; vertical scale is base-one voltage 0.5 volts per major division, horizontal scale is time from right to left 200 nsec. per large division.

tivity modulation occurs is lower and accordingly the charge which must be established in switching a given emitter current is less.

The turn off time of the UJT is normally defined in terms of the rate at which the peak point voltage approaches its steady state value after the emitter current is cutoff. Note that the turn off time is not equivalent to the reverse recovery time of the emitter diode. The test circuit used to measure the turn off time is shown in Figure 10. In this circuit, current for the emitter is provided by the 470 ohm resistor. The clamping voltage, V_1, is adjusted so that the emitter voltage is below the peak point voltage. The UJT is turned on by the negative pulse at base-two which occurs when the pole of the mercury relay transfers to the left contact. The UJT remains in the conducting state until the pole of the mercury relay transfers to the right contact at which time a negative pulse is coupled to the emitter through the variable capacitor. The emitter voltage will rise towards the positive supply voltage at a rate determined by the time constant of the 470 ohm resistor and the variable capacitor. When the emitter voltage becomes positive, the UJT will turn on (dotted line) if the length of time the emitter is reverse biased is shorter than the turn-off time, otherwise the UJT will not turn on and the emitter voltage will rise until it is clamped by the diode. The variable capacitor is adjusted to determine the minimum time the emitter can be reverse biased and still block when the emitter voltage goes positive. Measured values of turn off time in the circuit of Figure 10 lie between 1.5 and 3.0 μsec. for both UJT structures.

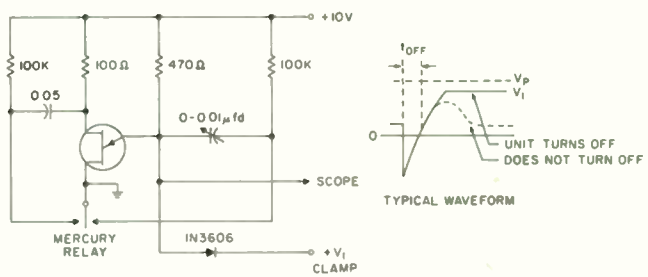


Fig. 10 Test Circuit For Measuring Turn-Off Time Of UJT With Typical Operating Waveforms.

Summary Of Characteristics

Comparative values of the more important parameters for the new UJT structure and the conventional UJT structure are

given in Figure 11, together with the estimated designable range for these parameters. The feasibility of higher values of intrinsic standoff ratio with the new structure has already been demonstrated by Figure 7. The lower limit of intrinsic standoff ratio is limited in the new structure because of the difficulty of accurately maintaining the close spacing required whereas there is no lower limit in the conventional structure. The emitter saturation voltage is obviously lower with the new structure. Generally, the lowest possible emitter saturation voltage is desirable, if higher values are required they could easily be achieved by either decreasing the effective lifetime or by adding appropriate external resistance in series with base-one or the emitter. The range of interbase resistance that can be achieved with both structures is determined primarily by the practical upper limit in the silicon resistivity and the practical lower limit in the size of the base contacts. The lower value of peak point current with the new UJT structure increases the value of the UJT in sensitive trigger circuits. A wider range of valley current is possible with the new structure than with the conventional structure. High values of valley current can be achieved by reducing the surface or bulk lifetime of the silicon. The faster turn on speeds of the new structure have already been described. The thermal resistance to the case of the new structure is about one-half the thermal resistance to case of the conventional structure owing to the fact that the new structure is mounted directly on the header and the heat path is shorter. This will permit higher power ratings with the new structure when used with an appropriate heatsink. The thermal resistance in free air will be approximately the same for the two structures.

	CONVENTIONAL BAR STRUCTURE		NEW STRUCTURE	
PARAMETER	TYPICAL	DESIGN RANGE	TYPICAL	DESIGN RANGE
INTRINSIC STANDOFF RATIO, η	0.60	0-0.75	0.65	0.30-0.90
EMITTER SATURATION VOLTAGE (η=0.60)	3.0V	-	1.5v	-
INTERBASE RESISTANCE R_{BB}	7K	0-30K	7K	0-30K
PEAK POINT CURRENT, I_p (V_{BB}=20V, η=0.60)	1.2μa	-	0.3μa	-
VALLEY POINT CURRENT, I_v (V_{BB}=20v, η=0.60)	15ma	10-20ma	2ma	0.5-20ma
TURN ON TIME (R_L=500Ω)	4μs	-	80ns	-
THERMAL RESISTANCE TO CASE	0.12°C/MW (TO-5)	-	0.06°C/MW (TO-18)	-

Fig. 11. Comparative Values And Design Ranges Of Principal Parameters For Conventional UJT Bar Structure And New UJT Structure.

Applications

The ZJ-85 can be used in most of the standard applications in which the conventional UJT is presently being used, including oscillators, pulse generators, time delay circuits, SCR firing circuits, etc. For applications requiring a high valley current, it is important to design and specify a minimum valley current consistent with the application. In some cases the need for a high valley current can be circumvented by using a small inductance in series with the emitter circuit which produces a ringing action to cut off the emitter at the end of the pulse.

Oscillators

The frequency stability of the ZJ-85 in a relaxation oscillator circuit is lower than the stability of the conventional UJT. However, the stability of the ZJ-85 is satisfactory for many applications. The temperature stability of a ZJ-85 in a relaxation oscillator is shown in Figure 12. In this test, the base-two resistor was adjusted to give the minimum frequency change in the vicinity of 25°C. The UJT alone was subjected to the temperature change, in a practical circuit the temperature stability would also be determined by the relative temperature coefficients of the 20K resistor and the 0.2 μfd capacitor.

One-Shot Multivibrator

The low valley current of the ZJ-85 makes possible a normally on type of one shot circuit such as shown in Figure 13. In this circuit the 3.9K resistor is small enough to provide a current larger than the valley current so that the circuit is stable with the emitter in the conducting state. A positive trigger pulse of the proper amplitude and duration

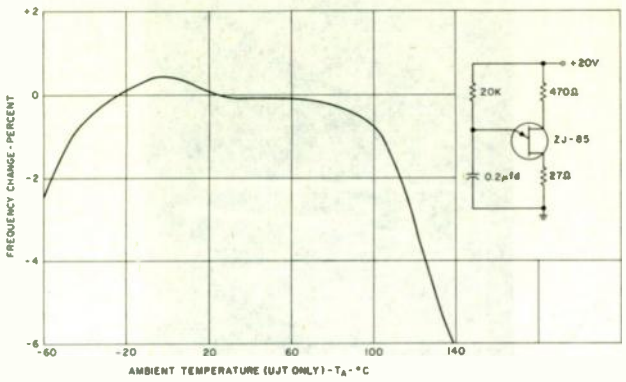


Fig. 12 Measured Frequency Change Versus Temperature For ZJ-85 In Relaxation Oscillator Circuit.

applied to base-one will cause the emitter to cut off. The .001 μfd capacitor will then charge to the peak point voltage at which time the UJT will fire and return to its conducting state. When the UJT fires a negative pulse will be generated at V_0 and a positive pulse will be generated at V_1 owing to the discharge current from the capacitor. For the circuit shown a 1.2 volt, 0.6 μsec. pulse was required to trigger the UJT and the pulse delay was 4.5 μsec.

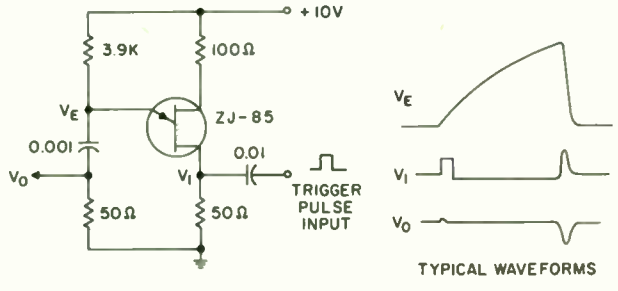


Fig. 13 Normally On One-Shot Multivibrator Showing Typical Operating Waveforms.

Voltage Comparator

Because of its very low peak point current and stable peak point voltage, the ZJ-85 can be used in sensitive voltage comparator circuits. One such circuit is shown in Figure 14. Potentiometer R_4 is adjusted so that the UJT will just fire when the signal inputs are shorted together. If the signal input is negative, the UJT will not fire, but if the signal is slightly positive, the UJT will fire and a pulse will appear at the secondary of the pulse transformer due to the discharge of C_1 through the emitter. The diodes prevent loading of the input circuit prior to triggering and clamp the voltage at the emitter of the UJT when large voltages are applied to the input. Capacitor C_1 is recharged through R_4 so that response time of the circuit is limited by the time constant of R_1, C_2 and C_3. One advantage of the ZJ-85 in this circuit is that small values of emitter capacitor can be used with a consequent improvement in the response time. The location of the pulse transformer is chosen to reduce the possibility of false triggering caused by transients coupled into the circuit through the pulse transformer. Temperature compensation is achieved by adjustment of R_2. It is possible to obtain a firing voltage stable to within ± 2 mv from 25°C to 55°C.

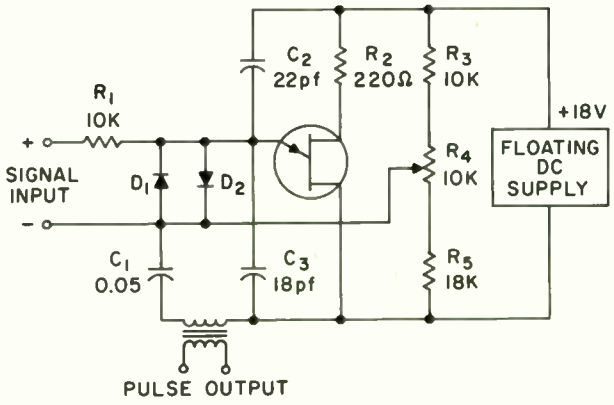


Fig. 14 Sensitive Voltage Comparator Circuit.

Time Delay Circuit

The ZJ-85 can be used in most of the time delay circuits in which the conventional UJT is used and owing to its low peak point voltage, is particularly useful in circuits where a very large time delay is desired with a small value of timing capacitance. In cases where a very long time delay is required over a moderate temperature range, a technique such as that shown in Figure 15 can be used. In this circuit the SCR is chosen in accord-

ance with the maximum supply voltage and maximum load current, diode D_1 is chosen in accordance with the maximum supply voltage, and R_1 is chosen to supply the proper current to zener diode before the SCR turns on and to supply the minimum holding current to the SCR after it turns on. A UJT with a low peak point current and low emitter leakage current is chosen for Q_1. The second UJT, Q_2, serves as a pulse generator, producing negative pulses across the 22 ohm resistor. These negative pulses effectively reduce the value of current required to trigger Q_1 to a value below the peak point current. The required current to fire Q_1 may be as low as 20 na and the ultimate sensitivity of the circuit is limited by the lowest emitter leakage current it is possible to achieve. The time delay for the circuit shown is over one minute even though the timing capacitor is only 1 μfd.

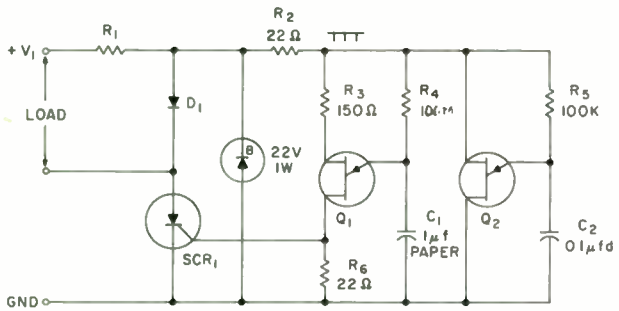


Fig. 15 Time Delay Circuit With Very Long Timing Interval.

Sharp Threshold Device

In applications where a very sharp voltage threshold at low voltage levels is required the ZJ-85 can be used in the inverted connection. The emitter to base-two characteristic of the ZJ-85 with a 5 volt bias on base-one is shown in Figure 16 in comparison with a conventional diode characteristic.

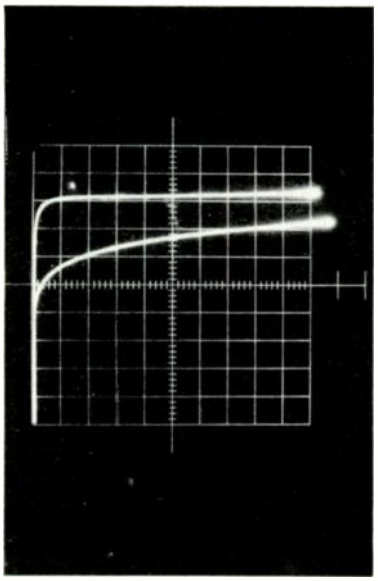


Fig. 16 Sharp Threshold Characteristic Using ZJ-85 In Inverted Connection (Upper Trace) Compared With Conventional Silicon Diode Characteristic (Lower Trace); vertical scale is 0.1 volts per division, horizontal scale is 50 μa per major division.

Acknowledgements

The authors wish to acknowledge to contributions of W. B. Walkup, G. P. Greiner, and B. L. Westcott to the development of processes used in the fabrication of the devices described in this paper.

References

1. Lesk, I.A., Mathis, V.P., "The Double-Base Diode - A New Semiconductor Device", IRE Conventional Record, Part 6, pp. 2 - 8, 1953.

2. Sylvan, T.P., "Notes On The Application Of The Silicon Unijunction Transistor", General Electric Company Application Note 90.10, May, 1961.

FREQUENCY OF MERIT FOR THREE TERMINAL ELECTRON DEVICES*

S. VENKATESWARAN

Department of Engineering, University College of Swansea,
University of Wales, United Kingdom.

Summary. A simple and closely accurate expression for the "frequency of merit" ('figure of merit') of each available three terminal electron device is obtained from its maximum operating power gain potential at higher frequencies as a tuned amplifier, in a suitable configuration (common-source etc.) and system (unilateralisation, mismatch); it is in terms of the parameters of the 'physical equivalent circuit' of the device including its "signal delay time" ('transit time'). Calculations on typical devices show that unlike a vacuum triode or field effect transistor, a graded base junction transistor may have a 'frequency of merit' greater than the inverse of its 'signal delay time'.

List of Principal Symbols.

f_m = "frequency of merit" ('figure of merit')

f_{max} = maximum frequency of oscillation

g_{maxc} = maximum available power gain of stage with conjugate matched terminations (MAG)

k = Stern's 'stability factor'

k_i = Stern's 'inherent stability factor'

p = h, z, y or g

$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}$ = general four-pole matrix of stage

p_1 = total input self parameter (including source immittance)

p_2 = total output self parameter (including load immittance)

s = auther's "invariant stability factor"

U = Mason's U factor

λ = complex variable

ρ = Re (p)

τ_s = "signal delay time" ('transit time')

ω = $2\pi f$

*Based on a Ph.D. thesis[36] of London University

Junction Transistors in Common Base Configuration

C_{c1} = inner collector capacitance

C_{de} = emitter diffusion capacitance

C_{te} = emitter transition layer capacitance

m = field factor in base region

$r_{b'}$ = resistance in series with 'base'

r_d = emitter diffusion resistance

γ = f/f_a

τ_{de} = $C_{de}\, r_d$

τ_e = $(C_{de} + C_{te})\, r_d$

τ_1 = "signal delay time" of 'intrinsic' transistor

τ_I = "signal delay time" of 'extrinsic' transistor

$\omega_a = 1/\tau_a$ = 3 db cut off angular frequency of intrinsic transistor

$\omega_\alpha = 1/\tau_\alpha$ = 3 db cut off angular frequency of extrinsic transistor

Vacuum Triode and Field Effect Transistor in Common Source Configuration.

C_{cs} = control-source charging capacitance

G_{mo} = mutual conductance at low frequencies

R_{cs} = control source series resistance

R_{ds} = drain-source shunt resistance

Y_m = $y_{21} - y_{12}$ = mutual admittance

μ_o = amplification factor at low frequencies

I. Introduction

A three terminal electron device can amplify power in any of its 'configurations'. These configurations correspond to having one each of its terminals i.e. source, control or drain common between input and output terminal pairs. A limitation arises, if the device is used for power amplification, this being its maximum frequency of oscillation, f_{max}; f_{max} is an invariant with respect to configurational

permutation and may in theory, be obtained by equating Mason's U factor[23] to unity. However, no simple and accurate or explicit expression for f_{max} has been obtained so far, for even a single three terminal electron device.[11,13,15,25,27].

It has been previously concluded[35] that no general and simple "frequency of merit" ('figure of merit') relating power gain and bandwidth[16] of active two-port networks exists. The question arises as to whether a simple and accurate frequency of merit exists for each of the commercially available types of three terminal electron devices in terms of its power gain performance at higher frequencies in a suitable configuration, 'system' (unilateralisation or mismatch) and 'matrix environment' (h-, z-, y- or g- type of terminations). An affirmative answer is provided in this paper; this frequency of merit is obtained in terms of important parameters of the 'physical equivalent circuit' of the device including its 'signal delay time' (transit time).

II Maximum Power gain of Active Two-Port Network.

For a wide range of active two-port networks, including valve and transistor amplifier stages, the total self parameters (including terminations), p_1, p_2 and the transfer parameters p_{21}, p_{12} have no poles and zeros either on the imaginary axis or on the finite right half complex plane whether each of their ports be open circuited or short circuited. Only such networks are considered here. Two-port networks not satisfying these conditions may be 'padded' by resistances at the ports and/or external feedback may be applied with passive 'elements' (resistance, inductance, capacitance, transformer, gyrator[32]) such that the modified networks satisfy the above conditions.

At higher frequencies ($f < f_{max}$), conjugate matching with passive terminations at the ports of an active two-port network may be impossible. If so, the 'potentially unstable' network can be rendered 'absolutely stable' by Mason's lossless unilateralisation[23], matrix unilateralisation[3] or mismatch of real parts in port terminations[31].

Mason's Lossless Unilateralisation

It has been shown by Mason[23] that the maximum available power gain of a two-port network after lossless unilateralising feedback is given by U, where

$$U = \frac{|p_{21} - p_{12}|^2}{4\,(\rho_{11}\rho_{22} - \rho_{12}\rho_{21})}, \quad p = z \text{ or } y \text{ and } \rho = \mathrm{Re}\,(p) \qquad (1)$$

As the frequency is increased, U usually decreases and a frequency f_{max} is reached where U equals unity. At this frequency, f_{max}, an invariant for all three configurations, conjugate matching without unilateralisation results in a power gain of unity while matrix unilateralisation results in a power gain not exceeding unity.[36]

f_{max} therefore is a unique frequency of merit for a given three terminal device. Where even a single matrix parameter p_{11}, p_{12}, p_{21} or p_{22} of the device is a quotient of two high degree polynomials involving a number of significant poles and zeros in the complex (λ) plane, an accurate and compact expression for f_{max} in terms of the parameters of its 'physical equivalent circuit' does not seem possible. Such is the situation for every type of three terminal electron device commercially available[11,13,15,25,27] viz. uniform base, graded base and PNIP (or NPIN) junction transistors, field effect transistor and vacuum triode.

Matrix Unilateralisation

The theory of matrix unilateralisation has been developed and applied to junction transistors by Cheng[3], Stern[29] with Aldridge and Chow, Chu[4] and Cote[5]. The type of two-port networks treated by them satisfy the requirements of Section II; as such these networks may be unilateralised by passive external feedback in four basic ways. The maximum available power gain (MAG) after optimised z or y matrix unilateralisation is given by

$$g_{\substack{\text{maxo} \\ \text{opt.unilat.}}} = \frac{|p_{21} - p_{12}|^2}{4\,[(\rho_{11}\rho_{22})^{\frac{1}{2}} \pm \rho_{12}]^2} \qquad (2)$$

with $p = z$ or y

In (2), ρ_{12} is positive for z parameters and negative for y parameters; a positive sign before ρ_{12} is to be taken with impedance matrix and a negative sign with admittance matrix. The upper bound for optimum unilateralised MAG is given by

$$g_{\substack{\text{maxo} \\ \text{opt.unilat.}}} \simeq \frac{|p_{21} - p_{12}|^2}{4\,\rho_{11}\rho_{22}} \text{ with } p = z \text{ or } y \qquad (3)$$

If $\mathrm{Re}\,(z_{12})$ equals zero, z matrix unilateralisation yields a MAG equal to the upper limit, which in this case equals Mason's U factor; similarly for y matrix unilateralisation. This coincidence with U factor should be expected, since the considered matrix unilateralisation is also lossless.

Resistive Mismatch at the Ports

Where Stern's 'inherent stability factor', k_i (i.e. for zero real parts of terminations) of the network is less than unity in a matrix environment, any power gain may be realised[35] by a suitable choice of the network stability factor, k (including terminations). As k tends to unity, maximum power gain tends to infinity.

If one or more regions of 'potential instability' occur over the useful range of frequencies for the device in some configuration, the boundaries of these regions define frequencies of

'marginal stability', where k_i equals unity in all four matrix environments[35]. At any of these frequencies the maximum available power gain equals the 'measure of non-reciprocity', $|p_{21}/p_{12}|$, of the network.

Stern's 'stability factor'[30], k, is not a true index of the gain potentiality of the device as k value varies from one matrix environment to another for the same power gain of the stage[35] except for the 'marginal' case where $k = k_i = 1$. A new stability factor, s, has been defined[37] which is an exact invariant for all its values in the four possible matrix environments; it demarcates the regions of 'potential instability' and 'absolute stability' and is directly related to the maximum available power gain of the two-port network when this gain is finite. In terms of s, MAG is given by

$$g_{max} \text{ at } s \geq 1 = \left| \frac{p_{21}}{p_{12}} \right| \qquad (4)$$

Before these results can be applied to three terminal devices, a knowledge of their equivalent circuits is necessary.

III Equivalent Circuit.

Care must be taken in the choice of a suitable circuit for any three terminal device. A number of different equivalent circuits may yield nearly the same answer for terminal behaviour, but only that equivalent circuit where the parameters are well defined - least frequency dependent - will yield the most compact expression for power gain in terms of frequency. An equivalent circuit that has the least number of frequency dependent parameters is often the one representing the physical processes in the device and is usually called the 'physical equivalent circuit'; it is the circuit favoured by the device designer and any frequency of merit in terms of the parameters of such a circuit may be appreciated by him as well as the circuit designer. In the following two subsections the physical equivalent circuit of the field effect transistor[5] (also of analog transistor[28] and negative grid vacuum triode) valid over a wide frequency range and some circuit features of junction transistors (uniform base, graded base and PNIP types) will be presented.

Negative Grid Vacuum Triode and Field Effect Transistor (Class A Devices)

Llewellyn[22] has derived a simple equivalent circuit for a negative grid vacuum triode which is claimed accurate in predicting its performance till the "signal delay angle" (familiarly known as 'transit angle') equals half a radian. This is illustrated by Fig. 1. In assessing the maximum frequency of oscillation of their field effect transistor, Dacey[6] and Ross have used the very simple equivalent circuit of Fig. 2. For calculations on the performance of this transistor as a tuned amplifier at higher frequencies, a resistance R_{ds} and a capacitance C_{ds} may be placed[9] across the output current generator. In their models the gate or control electrode is a metal slab alloyed to the semiconducting slab with the source and drain terminals on either side. There will therefore be a capacitance C_{gd} between gate and drain and a capacitance C'_{gs} between gate and source; the latter is small compared with C_{gs} which is an intrinsic part of the device. Fig. 3. represents the physical equivalent circuit of the field effect transistor at higher frequencies.

Dishal[10] has neglected R_{gp} (negative) as small compared with the reactance of C_{gp}, when computing the voltage gains of common grid vacuum triode amplifiers at ultra high frequencies. Thus it appears that at higher frequencies, the only significant difference between Figs. 1 and 3 lies in the expression for Y_m, the mutual admittance. In Fig.1, $Y_m = \mu_o/R_p = G_{mo}$, whereas in Fig. 3. $Y_m = G_{mo} / (1 + j\omega C_{gs} R_{gs})$. If in Fig.1. Y_m were modified to $G_{mo} / (1 + j\omega C_{gk} R_{gk})$, it will hardly affect the calculated performance of the vacuum triode when the 'signal delay angle' is less than half a radian. This is because the product $C_{gk}R_{gk}$ is of the order[14] of 10^{-1} of the 'transit time'. So at higher frequencies negative grid vacuum triodes and field effect transistors have physical equivalent circuits as of Fig. 3. This circuit has a physical interpretation. An applied voltage charges the capacitance C_{gs} through the resistance R_{gs}, this process being associated with a time constant $C_{gs}R_{gs}$; a current is generated at output strictly proportional to the instantaneous charge (or voltage) on C_{gs} and independent of frequency for signal delay angles less than half a radian.

Electron devices like vacuum triodes, analog transistors and field effect transistors which have equivalent circuits like that of Fig. 3. may be categorised as "Class A" types. In each of these device types only one polarity (+ or -) of charge carriers plays the significant role.

Junction Transistors (Class B Devices)

The complete physical equivalent circuit of a graded base transistor is complicated due to the following major causes. r_d and C_{de} the diffusion resistance and capacitance respectively are frequency dependent at higher frequencies[8]. The 'extrinsic' elements C_{te}, C_{c1} and r'_b complicate the situation further. Thus the determination of the four-pole parameters for a generalized junction transistor (with or without an electric field in base) is too complicated.

For such a situation h_{12}, h_{21} and h_{22} in the common-base configuration may still be expressed. But the expression for h_{21} is sufficiently complex that judicious approximations are necessary to obtain the external cut off frequency, f_α analytically.

At frequencies less than $2f_a$, h_{21} for the intrinsic transistor, h_{21i}, may be expressed as[34]

$$a = -h_{21i} = \frac{a_o \exp\left[-j(\tau_1 - \tau_a)\omega\right]}{1 + j\omega\tau_a} \qquad (5)$$

where a_o is a, at low frequency, τ_1 the inverse of ω_1 and τ_a the inverse of ω_a. With extrinsic elements C_{te}, C_{c1} and r_b', the expression for h_{21} in common base becomes more complex. At higher frequencies, it is given by [7*]

$$\alpha = -h_{21} \simeq \frac{\dfrac{a_o \exp[-j(\tau_1-\tau_a)\omega]}{(1+j\omega\tau_a)}\left(\dfrac{1+j\omega\tau_{de}}{1+j\omega\tau_e}\right) + j\omega C_{c1} r_b'}{1 + j\omega\, C_{c1}\, r_b'} \qquad (6)$$

neglecting Early's[12] feedback factor, μ, as small. (Refer Appendix 1)

In (6), α_o refers to low frequency α, C_{c1} to collector inner junction capacity [8+] and r_b' to extrinsic base resistance [8#], τ_{de} to the time constant associated with diffusion capacitance and resistance at emitter junction and τ_e to the time constant at emitter junction taking transition capacitance also into account.

It is proved in Appendix 1 that the cut off frequency for a uniform base transistor - with $C_{te}/C_{de} \ll 1$ - is given by

$$\frac{1}{\tau_\alpha} = \omega_\alpha \simeq \omega_a \left\{ \frac{x - \sqrt{x^2 - 4\alpha_o^2(2-\alpha_o^2)\,y^2}}{2y^2(2-\alpha_o^2)} \right\}^{\frac{1}{2}} \qquad (7)$$

$$\frac{C_{te}}{C_{de}} \ll 1$$

where $x = \alpha_o^2 + 4.88\,\alpha_o\, y - (2-\alpha_o^2)y^2$ and $y = \omega_a C_{c1} r_b'$,

Further, if $y \leq 0.1$, for a uniform base transistor

$$\frac{1}{\tau_\alpha} = \omega_\alpha \simeq \frac{\alpha_o\, \omega_a}{\left\{\alpha_o^2 + 4.88\,\alpha_o\, y - (2-\alpha_o^2)y^2\right\}^{\frac{1}{2}}} \qquad (8)$$

$$\frac{C_{te}}{C_{de}} \ll 1 \; ; \quad y \leq 0.1$$

For a practical graded base transistor, $y \ll 1$ but C_{te}/C_{de} is not negligible. As shown in Appendix 1, its ω_α is given by

$$\frac{1}{\tau_\alpha} = \omega_\alpha$$

$$y \ll 1$$

$$\simeq \left\{ \frac{2}{(\tau_a^2+\tau_e^2-2\tau_{de}^2) + \left[(\tau_a^2+\tau_e^2-2\tau_{de}^2)^2 + 4\tau_a^2\tau_e^2\right]^{\frac{1}{2}}} \right\}^{\frac{1}{2}} \qquad (9)$$

At higher frequencies Early's voltage feedback factor is negligible compared with $\omega\, C_{c1}\, r_b'$, so that

$$h_{12} \simeq \frac{jyy}{1 + jyy} \quad , \quad \text{where} \quad y = f/f_a \qquad (10)$$

for both uniform and graded base transistors. With practical transistors $y^2y^2 \ll 1$ in the frequency range f_α to $2f_\alpha$. Therefore

$$\left|h_{12}\right| \simeq \omega\, C_{c1}\, r_b' \simeq yy \qquad (11)$$

$$y^2y^2 \ll 1$$

and increases linearly with frequency over this range. Thus the magnitude of p_{21}/p_{12} ab ω_α is given by

$$\left|p_{21}/p_{12}\right|_{\omega_\alpha} \simeq \frac{\alpha_o}{\sqrt{2}\,\omega_\alpha\, C_{c1}\, r_b'} \qquad (12)$$

$\left|p_{21}/p_{12}\right|$ may be taken to start with the above value at ω_α and decrease at 6 db/octave rate over range ω_α to $2\omega_\alpha$; as shown in Appendix 2 the error in use of this proposed

---

*This assumes that the collector body series resistance, r_{cg}, plays a negligible part; as such it may not be applicable for some VHF/UHF transistors. However it is valid for the epitaxial diffused transistor[33] whose f_a can be in the UHF region.

#Value of r_b', depends upon type of measurement. For the product $C_{c1}r_b'$, in (6) and (11), reliance is made on h_{12} measurements at higher frequencies.

+C_o obtained from h_{22} measurements in common base at higher frequencies is composed of the inner junction capacitance C_{c1} and a terminal stray plus outer junction capacity C_{c2}. For VHF transistors they are comparable.

simple law does not exceed 1 db at $2\omega_\alpha$.

Electron devices that nearly obey the above simple law in their common control configuration may be categorised as "ClassB". Most junction transistors including VHF/UHF transistors of the diffused mesa[20,21], or epitaxial diffused types[33] fall in this category. All these devices depend for their action on two polarities (+ and -) of charge carriers.

Before expressions for the frequency of merit of three terminal electron devices can be obtained in terms of their signal delay times, it is necessary to define signal delay time in a general and useful way.

IV. Signal Delay Time

From sinusoidal response of a given device at its terminals in its common-control configuration and h - environment, it is possible to define "signal delay time" (usually known as 'transit time') in several ways[36]. Of these the one defined from low frequency considerations leads to an expression well related to the 'physical equivalent circuit' of each given device and is hence adopted here.

Class B Devices

Define 'signal delay time', τ_s, of the devices as

$$\tau_s = \left.\frac{d\theta(\omega)}{d\omega}\right|_{\omega \to 0} \qquad (13)$$

where its forward transfer hybrid parameter (in h - environment) in the common-control configuration is given by

$$h_{21} \Big/ h_{21}\Big|_{\omega \to 0} = A(\omega) \exp\ [-j\theta(\omega)] \qquad (14)$$

In (14), $A(\omega)$ is the normalised amplitude response and $-\theta(\omega)$ the phase angle in radians at frequency ω. For the intrinsic transistor in its common base configuration,

$$h_{21i} \Big/ h_{21i}\Big|_{\omega \to 0} = \frac{1}{(1+\omega^2\tau_a^2)^{\frac{1}{2}}} \Big/ -\omega(\tau_1-\tau_a) - \tan^{-1}\omega\tau_a$$

$$\omega \lesssim \omega_a \qquad (15)$$

From (15), $\tau_s = \tau_1$. For the transistor with 'extrinsic' elements C_{te}, C_{c1} and r_b', (6) and (13) yield

$$\tau_s = \tau_1 + \tau_e - \tau_{de} = \tau_1 + C_{te} r_d = \tau_I \qquad (16)$$

The signal delay time depends upon time $\tau_1 = 1/\omega_1$, C_{te} and r_d but is independent of $\tau_a = 1/\omega_a$, C_{c1} and r_b'.

Class A Devices.

For Class A devices, as shown in Appendix 3

$$\tau_s = \left(\frac{\mu_o}{\mu_o+1}\right)\frac{C_{cs}}{G_{mo}} \simeq \frac{C_{cs}}{G_{mo}} \quad \text{for} \quad \mu >> 1 \qquad (17)$$

The extreme simplicity of (17) may be observed. 'Transit times' calculated for various Class A and Class B devices [6,14,19] are from the mean velocities of charge carriers in the direction of the electric field and are not obtained from the time of travel of any individual charge carrier. The values thus obtained are coincident with τ_s as given by (13).

V. Frequency of Merit

It is now possible to evaluate a simple and accurate expression for the frequency of merit of each existing type of three terminal electron device; it will be in terms of parameters from its physical equivalent circuit. The power gain maximisation will be carried out in such a configuration (common-source etc.) system (lossless unilateralisation etc.) and matrix environment (h- etc.) as is most convenient.

Class A Devices

The equivalent circuit for these devices is shown in Fig. 3. The y_{12} parameters of such a device at higher frequencies in the three possible configurations are

$$\begin{aligned} &\text{common source } y_{12} \simeq -j\omega C_{cd} \\ &\text{common control } y_{12} \simeq -G_{ds} - j\omega C_{ds} \\ &\text{and common drain } y_{12} \simeq -j\omega C_{cs}/(1+j\omega C_{cs} R_{cs}) \end{aligned} \qquad (18)$$

Of the three configurations and four matrix environments, only the y- environment in the common source configuration yields a lossless reverse feedback. As such y-unilateralisation in common source yields a maximum power gain given by

$$g_{maxc} = U \simeq G_{mo}^2 \Big/ 4\omega^2 C_{cs}^2 R_{cs} G_{ds} \qquad (19)$$

opt.unilat.

as shown in Appendix 4.

A closely accurate frequency of merit, may therefore be defined from

$$\omega_m = \sqrt{g_{maxc}}\ \omega = \sqrt{U}.\ \omega \simeq \sqrt{\frac{\mu_o^2}{4(\mu_o+1)\ \tau_s\ C_{cs}\ R_{cs}}} \qquad (20)$$

where τ_s is given by (17). If $\mu_o >> 1$ as is usually the case, the angular frequency of merit

$$\omega_m \simeq \left\{\frac{\mu_o}{4\ \tau_s\ C_{cs}\ R_{cs}}\right\}^{\frac{1}{2}} \qquad (21)$$
$$\mu_o >> 1$$

(20) or (21) closely defines the maximum power gain with y-unilateralisation in the common source configuration or with Mason's lossless unilateralisation in any of the three configurations, at frequencies where the equivalent circuit of Fig. 3 is valid. According to Fig. 3 the input port admittance due to C_{cs}, R_{cs} has a real part [14,24] $\simeq \omega^2 C_{cs}^2 R_{cs}$ and hence proportional to ω^2 if $\omega^2 C_{cs}^2 R_{cs}^2 << 1$, as in practice. But at frequencies approaching f_{max}, the real part of input admittance increases beyond that given by the above square law due to 'transit time' effects. As a result $f_{max} < f_m$; it may be thought that f_m sets an upper bound for f_{max}, but the real utility of f_m is in its definition of the maximum gain potentiality of the device as a tuned amplifier at higher frequencies in the common source configuration with y-unilateralisation.

Class B Devices

By definition, for a Class B device,

$$\left|p_{21} / p_{12}\right|$$

in its common-control configuration falls nearly at 6db/octave rate from ω_α to $2\omega_\alpha$. For junction transistors in the common base configuration one or more regions* of potential instability occur over the useful frequency range as shown in Figs. 4 and 5, also at least* one frequency of marginal stability ($k_i = s = 1$) occurs near the cut off frequency f_α. Thus as long as the frequency of marginal stability is on the 6 db/octave straight line of fall in measure of non-reciprocity,

$$\omega_m^2 = g_{maxc}\,\omega^2 = \left|p_{21}/p_{12}\right|_{\omega_\alpha}^{k_i=s=1} \omega_\alpha^2 \qquad (22)$$

$$\text{or}\quad \omega_m \simeq \sqrt{\frac{\alpha_o}{\sqrt{2}\,\tau_\alpha\, C_{c1}\, r_b'}}\ \sqrt[4]{1 + \omega_\alpha^2\, C_{c1}^2\, r_b'^2}$$

$$\simeq \sqrt{\frac{\alpha_c}{\sqrt{2}\,\tau_\alpha\, C_{c1}\, r_b'}} \qquad (23)$$

since $\omega_\alpha^2\, C_{c1}^2\, r_b'^2 << 1$ and a fourth root occurs.

Also over angular frequency range ω_α to $2\omega_\alpha$

$$\{g_{maxs}\, s\, \omega^2\}^{\frac{1}{2}} = \omega_m \simeq \left\{ \frac{\alpha_o}{\sqrt{2}\,\tau_\alpha\, C_{c1}\, r_b'} \right\}^{\frac{1}{2}} \qquad (24)$$

The maximum available power gain for a given invariant stability factor, s, obtained through (24) is exact at ω_α and within 1 db at $2\omega_\alpha$. In (24) ω_m is in terms of τ_α and not in terms of signal delay time τ_s ($=\tau_I$). If required ω_m may be expressed in terms of τ_s using (24), (7) or (9) and (16). The complete expression is too complex and will not be presented; but these equations will be presently used in two examples to illustrate this theory.

If an approximate expression for ω_m in terms of τ_s ($= \tau_I$) is required the relation

$$\omega_m = \sqrt{\frac{1}{\tau_s\, C_{c1}\, r_b'}} \qquad (25)$$

may be used for any practical transistor# with a uniform or a graded base. The error+ in its use for a Mullard Type OC44 transistor (uniform base) is + 16% while that for a RCA Type 2N247 transistor (graded base) is -9%. For details on the characteristics of these transistors refer to Section VI.

VI. Verification and Illustration of Theory.

In this Section, it is proposed to verify the behaviour of

$$\left|p_{21}/ p_{12}\right|$$

from ω_α to $2\omega_\alpha$ for a typical uniform base transistor, and a typical graded base transistor. It is also proposed to evaluate $g_{max\,c}$ and

$$\left|p_{21} / p_{12}\right|$$

for these junction transistors in their common base configurations over a wide range of frequencies till ω_{max} values; the ratio of ω_m to ω_{max} and product $\omega_m \tau_s$ will be computed not only for these junction transistors but also for a typical field effect transistor and a vacuum triode.

Junction Transistors

$$\left|p_{21} / p_{12}\right|$$

and $g_{max\,c}$ values are calculated for two typical transistors in their common-base configuration over a wide range of frequencies starting from ω/ω_a of 0.001 to ω/ω_a of nearly 2 for the uniform base transistor, Mullard Type OC44 and ω/ω_a of nearly 10 for the graded base transistor RCA Type 2N247. The following are assumed parameters for these transistors.

---

* These results have not been previously observed by others.

\+ error = (calculated result-exact result)/exact result.

\# $\frac{1}{2}\left\{\frac{1}{\tau_s C_{c1} r_b'}\right\}^{\frac{1}{2}}$ represents ω_{max} of an idealised reference device with specific characteristics. For details refer 1;pp.10.32.

Type OC44* at I_e = 1 mA, V_c = - 6V and K =298°;
$\alpha_o \simeq a_o = 0.99$; $Re(h_{22}) = 0.8\mu\mho$; $\mu_o = 1.11 \times 10^{-3}$
$\omega \to o$
$C_{c1} \simeq C_c = 10.5$pF; $r_b' = 110\,\Omega$ and $f_a = 15$ Mc/s.

Type 2N247 at I_e = 1mA, V_c = -9V and K = 298°.

* $\alpha_o \simeq a_o = 59/60$; $Re(h_{22}) = 0.1\mu\mho$; $\mu_o = 20 \times 10^{-6}$:
$\omega \to o$

† C_{c1} = 1pF; r_b' = 20Ω; f_a = 22.8 Mc/s; f_1 = 13.7 Mc/s;
m = 5; C_{de} = 115 pF and C_{te} = 50pF.

Calculations of $g_{max\ c}$ and

$$\left|p_{21}/\ p_{12}\right|$$

are based on the original equations of Kroemer[1] modified to take into account C_{te}, C_{c1} and r_b'. Figs. 4 and 5 were drawn from a number of spot computations. $g_{max\ c}$ values were obtained from a well known equation.[35]

The extrinsic cut off frequency, f_α, was obtained from (7) or (9) as appropriate. A spot by spot (frequency) evaluation of h_{21} verified the accuracy of these equations. Through the point $(\omega_\alpha/\omega_a,\ \alpha_o/\sqrt{2\omega_\alpha C_{c1} r_b'})$, a 6 db/ octave line is drawn to meet the normalised frequency axis at ω_m/ω_a. It may be seen that

$$\left|p_{21}\ /\ p_{12}\right|$$

follows this straight line closely over a range of at least ω_α i.e. from ω_α to at least $2\,\omega_\alpha$. Also at least one angular frequency of marginal stability is near ω_α, there being two regions of potential instability for the graded base transistor.

The frequencies of merit and of limiting power gain (f_{max}) for these two devices are tabulated in Table 1; also normalised values with respect to signal delay times of 'extrinsic' transistors are given. For both these devices

$$1 < f_m\ /\ f_{max} < 2 \qquad (26)$$

By a remarkable coincidence, both these transistors have the same value for τ_s. Yet the graded base transistor has a higher gain performance and hence a higher frequency of merit; this is due to its smaller value of $C_{c1} r_b'$ achieved by the grading of impurity concentration in base.

* Average values computed from low frequency measurements.

† Average values kindly provided by Das from his high frequency measurements. Reference 8.

‡ From manufacturer's data of average parameters.

Table 1 reveals that for the type 2N247 transistor both f_{max} and f_m are greater than the inverse of τ_s. i.e. for a Class B device f_{max} and f_m may be greater than the inverse of its signal delay time.

Vacuum Triode and Field Effect Transistor

For a negative grid vacuum triode the data selected refer to a RCA Type R-192 vacuum triode as reported by Ferris[14]. The following information was used. μ_o = 12; G_{mo} = 0.645 m℧; R_{in} = 23 kΩ at 30 Mc/s and $\omega\tau_s$ = 0.639 radians at 30 Mc/s. From this data and (17)

$$C_{gk} \simeq 2.4\ \text{pF} \qquad (27)$$

Since $R_{in} \simeq 1/\omega^2 C_{gk}^2 R_{gk}$ additional information in (27) gives

$$C_{gk} R_{gk} \simeq 0.52\ \text{m}\mu\text{s} \qquad (28)$$

From these values and (21) it may be verified that

$$f_m = 200\ \text{Mc/s and } f_m \tau_s \simeq 0.68 \qquad (29)$$

The frequency of oscillation for a Gill-Morrell oscillator *17 may be used to estimate

$$f_{max} \text{ viz } f_{max} \tau_s \simeq 0.5 \qquad (30)$$

From (29) to (30) it follows that

$$f_{max} \simeq 1.5 \times 10^2\ \text{Mc/s and } f_m/f_{max} \simeq 1.3 \qquad (31)$$

For a field effect transistor the data of Dacey[6] and Ross relating to their unit No. 32 and some results from their theory will be used. At a drain voltage of 40V and a gate voltage of -5V, G_{mo} = 1.0m℧, R_{po} = 37kΩ and μ_o = 37. (32)

According to their theory $C_{cs} R_{cs} \simeq \frac{2}{3}\tau_s$ (33)

and $$f_{max} \simeq \frac{1}{\sqrt{2}\ \pi\ C_{cs} R_{cs}} \qquad (34)$$

Their square root mobility theory gives

$$f_{max} \simeq 46\ \text{Mc/s} \qquad (35)$$

From these details, (17) and (21) it may be verified that $C_{cs} R_{cs} \simeq 4.9$ mμs; $\tau_s \simeq 7.3$mμs and $f_m \simeq 81$ Mc/s. (36)

f_m value of 81 Mc/s may be compared with their experimental f_{max} value of 50 Mc/s. Using this value of f_{max}, $f_m \tau_s \simeq 0.59$: $f_{max} \tau_s \simeq 0.37$ and

$$f_m/f_{max} \simeq 1.6 \qquad (37)$$

* Unlike a Barkhausen-Kurz oscillator, the Gill-Morrell oscillator is sensitive to external circuit tuning.

Table 2 summarises these results for both the devices. For both

$$1 < f/f_{max} < 2 \qquad (38)$$

Both f_m and f_{max} values for these "Class A devices" are less than the inverse of the appropriate signal delay times.

VII. Conclusions.

Existing three terminal electron devices are classified under two major headings. Devices which have an equivalent circuit similar to that derived by Llewellyn for the vacuum triode at higher frequencies are categorised under "Class A"; those whose 'measure of non-reciprocity',

$$\left| p_{21} / p_{12} \right|$$

in their common-control configuration follow nearly a 6 db / octave law at higher frequencies are categorised under "Class B". Vacuum triodes and field effect transistors are "Class A devices"; practical junction transistors (with uniform or non uniform base) are "Class B devices".

A "Class A device" may be y-unilateralised in its common-source configuration to yield a MAG equal to Mason's U factor. For a wide range of higher frequencies, this gain is shown to be inversely proportional to the square of frequency. From such an expression a simple and closely accurate 'frequency of merit' has been obtained.

Any electron device yields a MAG inversely proportional to the chosen "invariant stability factor"; for a "Class B device" the product of such a gain and stability factor in the common control configuration follows closely a 6 db / octave law over the frequency range f_α (cut off frequency) to $2f_\alpha$ or higher frequency. Also a frequency of 'marginal stability' occurs about f_α. Both ways a simple and closely accurate 'frequency of merit' has been obtained.

Single delay time may be defined in various ways. Of these, that defined from low frequencies is proved equally adaptable for all existing three-terminal devices; the values thus obtained coincide with those obtained from the mean velocity of charge carries in transit.

Results are computed for a vacuum triode, a field effect transistor, a uniform base transistor and a graded base transistor. These show that both the maximum frequency of oscillation and the frequency of merit are less than the inverse of the low frequency signal delay time for each of the first two electron devices. For the graded base transistor these are greater than the inverse of its signal delay time.

Thus for the same signal delay time, the power gain performance of a graded base transistor may be superior to all other devices. For future 'epitaxial diffused' junction transistors f_{max} is estimated to be pushed up to tens of kMc/s, which is much higher than the estimated maximum of 4 kMc/s for a vacuum triode.[2]

These results on the maximum frequency of oscillation of three terminal electron devices in relation to 'transit time' do not support many of the conclusions of Johnson[18] and Rose.

X. Appendices

1. Forward Transfer Hybrid Parameter in Common-Base Configuration of Junction Transistors.

Taking C_{te}, C_{c1} and r_b' into account

$$-h_{21} = \frac{\dfrac{a_o \exp\left[-j(\tau_1 - \tau_a)\omega\right]}{(1 + j\omega\tau_a)}\left(\dfrac{y_{ee}}{y_{ee} + j\omega C_{te}}\right) + \dfrac{r_b'}{r_c} + j\omega C_{c1} r_b'}{1 + \dfrac{r_b'}{r_c} + j\omega C_{c1} r_b'} \qquad (39)$$

where y_{ee} is the input admittance of the intrinsic transistor in its common base with the output port short circuited and r_c, the collector resistance of the intrinsic transistor with the input port open circuited.

Till the cut off frequency of h_{21}, the variation of C_{de} and r_d with frequency is unimportant*; hence in evaluating ω_α from (39) the low frequency values of C_{de} and r_d may be taken. The factor r_b'/r_c in the numerator and the denominator of (39) may be omitted as it is usually less than 10^{-3}; the low value of r_b'/r_c also makes α_o close to a_o. Thus practically

---

*the cut-off frequency of $C_{de}\, r_d$ is greater than f_a; $\dfrac{\omega_{de}}{\omega_a} = \dfrac{\tau_a}{\tau_{de}} \simeq 1.22$ for $m = 0$ and $\simeq (2m)^{\frac{1}{2}}$ for $m \geq 4$. Reference 7; pp.541.

$$-h_{21} \simeq \frac{\frac{\alpha_o \exp[-j(\tau_1 - \tau_a)\omega]}{(1 + j\omega\tau_a)}\left(\frac{1 + j\omega\tau_{de}}{1 + j\omega\tau_e}\right) + j\omega C_{c1} r_b'}{1 + j\omega C_{c1} r_b'} \qquad (40)$$

where $\tau_{de} \simeq C_{de} r_d \Big|_{\omega \to 0}$ and $\tau_e \simeq (C_{de} + C_{te}) r_d \Big|_{\omega \to 0}$ (41)

It is proposed to evaluate ω_α from (40) separately for uniform and graded base transistors. For uniform base transistors $C_{te}/C_{de} \ll 1$, so that

$$-h_{21} \simeq \frac{[\alpha_o \cos\omega(\tau_1 - \tau_a) - \omega^2 \tau_a C_{c1} r_b'] + j[\omega C_{c1} r_b' - \alpha_o \sin\omega(\tau_1 - \tau_a)]}{(1 + j\omega\tau_a)(1 + j\omega C_{c1} r_b')} \qquad (42)$$

In (42), $\tau_1 - \tau_a \simeq 0.22\,\tau_a$, so that

$$\cos 0.22\omega\tau_a \underset{\omega\tau_a \leq 1}{\simeq} 1 \quad \text{and} \quad \sin 0.22\omega\tau_a \underset{\omega\tau_a \leq 1}{\simeq} 0.22\omega\tau_a \qquad (43)$$

The 3db cut off angular frequency $\omega_\alpha = 1/\tau_\alpha$ is now obtainable from (42) and is given in (7)

For the graded base transistor $\omega_a C_{c1} r_b' \ll 1$ so that (40) may be expressed as

$$-h_{21} \simeq \frac{\alpha_o \exp[-j(\tau_1 - \tau_a)\omega]}{1 + j\omega\tau_a}\left(\frac{1 + j\omega\tau_{de}}{1 + j\omega\tau_e}\right) \qquad (44)$$

$$\left|\frac{h_{21}}{h_{21}}\right|_{\omega \to 0} \simeq \frac{|1 + j\omega\tau_{de}|}{|1 + j\omega\tau_a|\,|1 + j\omega\tau_e|} = \frac{1}{|1 + j\omega\tau_\alpha|} \qquad (45)$$

where $\tau_\alpha = 1/\omega_\alpha$ is given by (9)

2. $\left|p_{21}/p_{12}\right|$ for Junction Transistors in Common-Base over Frequency Range f_α to $2f_\alpha$.

For a uniform base transistor, the law[26]

$$a = \frac{a_o \exp[-j(\tau_1 - \tau_a)\omega]}{1 + j\omega\tau_a} \qquad (46)$$

holds fairly accurately till $\omega\tau_a = 5$. For such transistors $C_{te}/C_{de} \ll 1$ and $y = \omega_a C_{c1} r_b'$ is not negligible (of order of 10^{-1}); but y^2 is negligible. Therefore at higher frequencies $|h_{12}| \simeq \omega C_{c1} r_b'$ and increases at 6 db/octave rate.

If y is negligible, fractional response of α (or a) at $2\omega_\alpha$ ($\simeq 2\omega_a$) will be $1/\sqrt{5}$ according to (46); but the proposed 6 db/octave rate of fall for

$$\left|p_{21}/p_{12}\right|$$

over range ω_α to $2\omega_\alpha$ will yield a fractional response for α at $2\omega_\alpha$ of $1/\sqrt{8}$. Thus the true response of

$$\left|p_{21}/p_{12}\right|$$

at $2\omega_\alpha$ will be $\sqrt{3/5}$ of the proposed response or 1 db higher.

Since y is not negligible, $\omega_\alpha < \omega_a$ and the response of α or $|p_{21}/p_{12}|$ at higher angular frequencies (near $2\omega_\alpha$) falls more rapidly than at lower angular frequencies (near ω_α). By definition response of

$$\left|p_{21}/p_{12}\right|$$

at ω_α will fit; therefore actual response at $2\omega_\alpha$ is less than 1 db above proposed response.

For a graded base transistor, increasing m, the field factor makes the response of, a, at $2\omega_a$ lower and lower[8]; for m = 5, the response is 0.8 db below that indicated by (46) whereas for m = 8 it is 2.1 db. lower. For these transistors y is negligible (of the order of 10^{-3}) but C_{te}/C_{de} is not negligible. If $C_{te}/C_{de} \ll 1$, the proposed response of

$$\left|p_{21}/p_{12}\right|$$

at $2\omega_a$ is 0.6 db below actual response for m = 5 and $\simeq$ 0.1 db above actual response for m = 8.

If $C_{te}/C_{de} \gg 1$ the law

$$\alpha = \alpha_o \frac{\exp[-j(\tau_I - \tau_\alpha)\omega]}{1 + j\omega\tau_\alpha} \qquad (47)$$

is applicable[34] even at $2\omega_\alpha$. In (47) τ_I is the

low frequency signal delay time. For this case the proposed law gives a response at $2\omega_\alpha$ approximately 1 db below actual response.

Usually C_{te}/C_{de} is neither very small nor very great compared with unity. In such a situation, the response at frequencies near $2f_\alpha$ is affected more than that at f_α in relation to the proposed law of response.

This is primarily due to the decrease[8] of C_{de} at higher frequencies unlike C_{te} which is more constant. For the usual value of 5 for m, the error in

$$\left|p_{21}/p_{12}\right|$$

at $2\omega_\alpha$ is therefore less than 0.6 db.

To conclude, under practical conditions, the true response of

$$\left|p_{21}/p_{12}\right|$$

over range ω_α to $2\omega_\alpha$ follows closely a 6 db/octave law; the departure from this law is small and does not exceed 1 db at $2\omega_\alpha$.

3. Signal Delay Time for a Class A Device

Admittance parameters for the circuit of Fig 3 are given by

$$y'_{11} = j\omega C_{de} + \frac{j\omega C_{cs}}{1 + j\omega C_{cs} R_{cs}} \qquad (48)$$

$$y'_{12} = - j\omega C_{cd} \qquad (49)$$

$$y'_{21} = \frac{G_{mo}}{1 + j\omega C_{cs} R_{cs}} - j\omega C_{cd} \qquad (50)$$

and

$$y'_{22} = G_{ds} + j\omega (C_{ds} + C_{cd}) \qquad (51)$$

Therefore

$$\Sigma y' = y'_{11} + y'_{12} + y'_{21} + y'_{22}$$

$$= G_{ds} + j\omega C_{ds} + \frac{G_{mo} + j\omega C_{cs}}{1 + j\omega C_{cs} R_{cs}} \qquad (52)$$

h_{21} in the common-control configuration (i.e. without prime, prime denoting common-source configuration) is given by $h_{21} = - (y'_{21} + y'_{22})/\Sigma y'$

$$= - \frac{G_{mo}+G_{ds}-\omega^2 C_{ds} C_{cs} R_{cs}+j\omega(C_{cs} R_{cs} G_{ds}+C_{ds})}{G_{mo}+G_{ds}-\omega^2 C_{ds} C_{cs} R_{cs}+j\omega(C_{cs} R_{cs} G_{ds}+C_{cs}+C_{ds})} \qquad (53)$$

From (13) and (53), signal delay time

$$\tau_s = \frac{C_{cs}}{G_{mo} + G_{ds}} \qquad (54)$$

and is re-expressed in terms of μ_o in (17).

4. Optimum y unilateralised MAG in Common-Source Configuration for a Class A Device

From (48) to (51)

$$\mathrm{Re}\,(y'_{11}) = \frac{\omega^2 C_{cs}^2 R_{cs}}{1 + \omega^2 C_{cs}^2 R_{cs}^2} \qquad (55)$$

$$\mathrm{Re}\,(y'_{22}) = G_{ds} \qquad (56)$$

and

$$\left|y'_{21} - y'_{12}\right|^2 = \frac{G_{mo}^2}{1 + \omega^2 C_{cs}^2 R_{cs}^2} \qquad (57)$$

Thus from (3), (55) to (57)

$$g_{\text{max c}} \text{ opt.unilat} = U = \frac{G_{mo}^2}{4\,\omega^2 C_{cs}^2 R_{cs} G_{ds}} \qquad (58)$$

and is inversely proportional to the square of frequency.

VIII Acknowledgments

The author is deeply grateful to his research supervisor Dr.A.R. Boothroyd of the Imperial College of Science and Technology, University of London for his guidance, encouragement and advice. He is also indebted to Mr. R. A. King of the same College for many helpful discussions and criticisms.

The personal support by the Ministry of Aviation, United Kingdom is gratefully acknowledged.

IX References

1. BOOTHROYD, A. R., KING, R.A., and DAS, M.B., Notes entitled "Lecture Course on Transistor Electronics", Department of Electrical Engineering, Imperial College of Science and Technology, London. April, 1960.

2. CHAMPEIX, R. "Physique et Technique des Tubes Electroniques" - book, Dunod, Paris, 1960.

3. CHENG, C.C., "Neutralisation and Unilateralisation", IRE Trans. Vol. CT-2, June, 1955; pp 138-145.

4. CHU, G.Y., "Unilateralisation of Junction Transistor Amplifiers at High Frequencies", Proc. IRE., Vol. 43, August, 1955; pp 1001-1006.

5. COTE, A.J. Jr., "Evaluation of Junction Transistor Neutralisation Networks", IRE Trans., Vol. CT-5. June, 1958; pp 95-103.

6. DACEY, G.C. and ROSS, I.M., "The Field Effect Transistor", Bell Sys. Tech. Jour., Vol. 34, November, 1955; pp 1149-1189.

7. DAS.,M.B., and BOOTHROYD, A.R., "Measurement of Equivalent Circuit Parameters of Transistors at V.H.F.", Proc. IEE., Vol. 106 B, Supplement 15, January, 1960; pp 536-549.

8. DAS., M.B., "A Study of the High Frequency Properties of Junction Transistors", London University Ph.D. Thesis, August,1960.

9. DEWITT, D., and ROSSOFF, A.L. "Transistor Electronics" - book, McGraw Hill, New York. 1957.

10. DISHAL, M., "Theoretical Gain and Signal to Noise Ratio of the Grounded Grid Amplifier at Ultra High Frequencies", Proc. IRE. Vol. 32 May, 1944; pp 276-284.

11. DROUILHET, P.R. Jr. "Predictions Based on the Maximum Oscillator Frequency of a Transistor", IRE. Trans. Vol. CT-2; June, 1955; pp 178-183.

12. EARLY, J.M., "Effects of Space Charge Layer Widening in Junction Transistors", Proc.IRE Vol. 40, November 1952; pp 1401-1406.

13. EARLY, J.M. "PNIP and NPIN Junction Transistor Triodes". Bell Sys. Tech Jour. Vol 33, May, 1954; pp 517-533.

14. FERRIS, W.R., "Input Resistance of Vacuum Tubes as Ultra High Frequency Amplifiers" Proc. IRE., Vol. 24, January, 1936; pp82-107.

15. GIACOLETTO, L.J. "Study of PNP Alloy Junction Transistor from DC through Medium Frequencies" RCA Rev. Vol. 15, December, 1954; pp 506 - 562.

16. GIACOLETTO, L.J. "Power Amplification x Bandwidth Figure of Merit for Transducers including Transistors", Jour. of Elec. and Cont., Vol 4, June, 1958; pp 515 -522.

17. GILL, E.W.B. "Electrical Oscillation of Very Short Wavelengths" Phil.Mag., Vol 12, October 1931; pp 843-853.

18. JOHNSON, E.O. and ROSE, A., "Simple General Analysis of Amplifier Devices with Emitter, Control and Collector Functions". Proc. IRE Vol. 47, March, 1959; pp 407-418.

19. KROEMER, H. "The Drift Transistor" Transistors I - book, RCA. New Jersey, March, 1956; pp 202-220.

20. LEE, C.A., and THOMAS, D.E. "A New High Frequency PNP Transistor", IRE Semiconductor Device Research Conference, Philadelphia, June, 1955.

21. LEE, C.A. "A High Frequency Diffused Base Germanium Transistor", Bell Sys. Tech. Jour. Vol. 35, January, 1956. pp 23-34.

22. LLEWELLYN, F.B., "Equivalent Networks of Negative Grid Vacuum Tubes at Ultra High Frequencies", Bell Sys. Tech. Jour. Vol.15, October, 1936; pp 575-586.

23. MASON, S.J., "Power Gain in Feedback Amplifiers", IRE Trans., Vol. CT-1, June, 1954; pp 20-25.

24. NORTH, D.O., "Analysis of the Effects of Space Charge on Grid Impedance", Proc. IRE Vol. 24, January, 1936; pp 108-136.

25. PRITCHARD, R.L., "Frequency Response of Theoretical Models of Junction Transistors", IRE Trans, Vol CT-2, June 1955 pp 183-191.

26. PRITCHARD, R.L., "Electric Network Representation of Transistors - A Survey", IRE Trans, Vol. CT-3, March, 1956; pp 5-21.

27. ROLLETT, J.M. "The Characteristic Frequencies of a Drift Transistor", Jour. of Elec. and Cont. Vol. 5, September, 1959. pp 193-213.

28. SHOCKLEY, W., "Transistor Electronics : Imperfections, Unipolar and Analog Transistors", Proc. IRE. Vol. 40. November, 1952; pp 1289-1313.

29. STERN, A.P., ALDRIDGE, C.A., and CHOW. W.F., "Internal Feedback and Neutralisation of Transistor Amplifiers"., Proc. IRE, Vol.43, July, 1955; pp 838-847.

30. STERN, A.P., "Considerations on the Stability of Active Elements and Applications to Transsistors", IRE Convention Record, Circuit Theory, 1956; pp 46-52.

31. STERN, A.P., "Stability and Power Gain of Tuned Transistor Amplifiers", Proc. IRE. Vol. 45, March, 1957; pp 335-343.

32. TELLEGEN, B.D.H., "The Gyrator, a New Electric Network Element", Philips Res Rept, Vol. 3, April, 1948; pp 81-101.

33. THEURER, H.C., KLEIMACK. J.J., LOAR, H.H., and CHRISTENSEN, H., "Epitaxial Diffused Transistors" - letter, Proc. IRE, Vol. 48 , September, 1960; pp 1642-1643.

34. THOMAS, D.E., and MOLL.J.L., "Junction Transistor Short Circuit Current Gain and Phase Determination", Proc. IRE, Vol. 46, June, 1958; pp 1177-1184.

35. VENKATESWARAN, S., and BOOTHROYD. A.R. "Power Gain and Bandwidth of Tuned Transistor Amplifier Stages", 1959 IEE International Convention on Transistors and Associated Semiconductor Devices, Proc. IEE, Vol. 106 B, Suppl. 15, January, 1960; pp 518-529.

36. VENKATESWARAN, S., "Stability, Power Gain and Bandwidth of Linear Active Four-Pole Networks, with Particular Reference to Transistor Amplifiers at Higher Frequencies" London University Ph.D. Thesis, June, 1961.

37. VENKATESWARAN, S., "An Invariant Stability Factor and its Physical Significance". IEE Monograph No. 468E, September, 1961; to be republished in Proc. IEE, Part C, March, 1962.

Table 1.

Signal delay time and gain performance of a uniform base and a graded base junction transistors.

Device (transistor)	mμ s				$\frac{\tau_s}{\tau_\alpha}$	M c/s		cycles	
	τ_1	τ_s	τ_a	τ_α		f_{max}	f_m	$f_{max}\,\tau_s$	$f_m\,\tau_s$
OC44	12.9	12.9	10.6	12.9	1.09	22.9	34.5	0.30	0.45
2N247	11.6	12.9	6.99	7.68	1.68	176	340	2.3	4.4

Table 2.

Signal delay time and gain performance of a vacuum triode and a field effect transistor.

Device	mμ s		$\frac{C_{cs}R_{cs}}{\tau_s}$	Mc/s		$\frac{f_m}{f_{max}}$	cycles	
	τ_s	$C_{cs}R_{cs}$		f_{max}	f_m		$f_{max}\,\tau_s$	$f_m\,\tau_s$
R-192*	3.4	0.52	0.15	150	200	1.3	$\simeq 0.5$	0.68
32+	7.3	4.9	0.67	50	81	1.6	0.37	0.59

*vacuum triode of Ferris. Reference 14

+field effect transistor of Dacey and Ross. Reference 6.

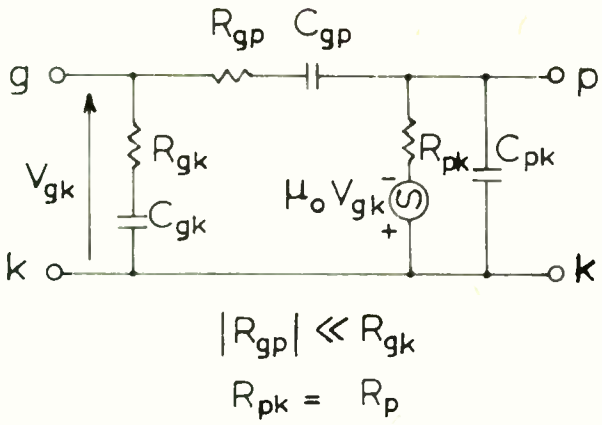


Fig. 1. Equivalent circuit for negative grid vacuum triode as obtained by Llewellyn for $\omega\tau_s \leq \frac{1}{2}$.

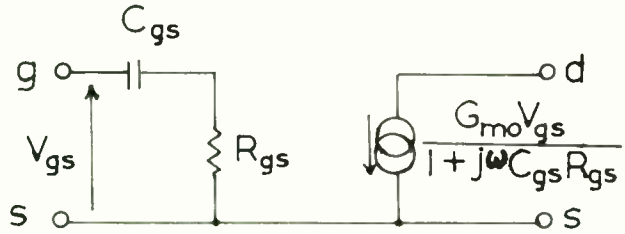


Fig. 2. Equivalent circuit of Dacey and Ross for their field effect transistor.

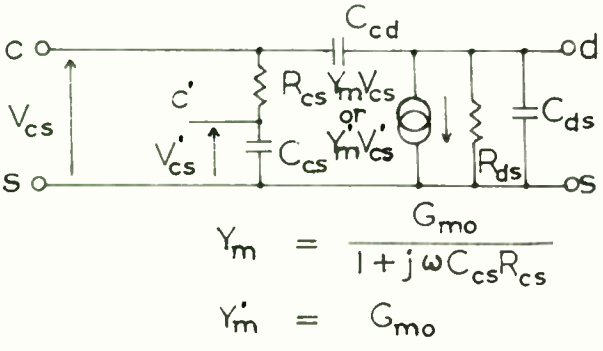


Fig. 3. Common equivalent circuit of Class A devices at higher frequencies.

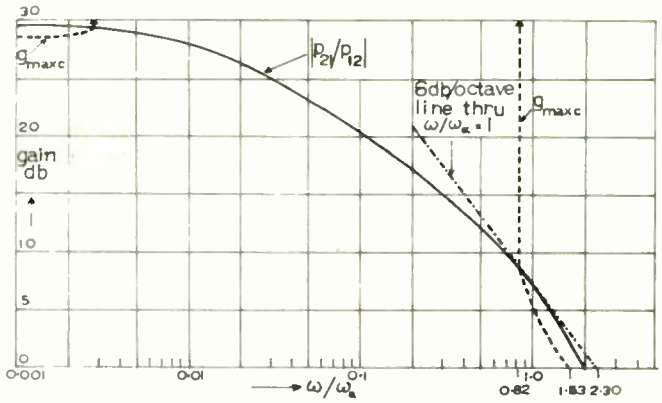


Fig. 4. $|p_{21}/p_{12}|$ and $g_{max\ c}$ for a typical uniform base transistor (Mullard type OC 44) in common base configuration.

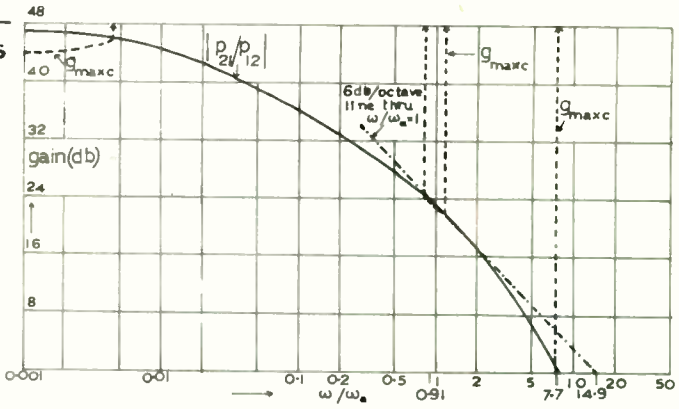


Fig. 5. $|p_{21}/p_{12}|$ and $g_{max\ c}$ for a typical graded base transistor (RCA type 2N 247) in common base configuration.

NEW MICROWAVE TECHNIQUES IN THE MEASUREMENT OF SEMICONDUCTOR PHENOMENA

H. Jacobs, F. A. Brand, J. D. Meindl, S. Weitz
U. S. Army Signal Research and Development Laboratory
Fort Monmouth, New Jersey

and

R. Benjamin
Monmouth College
West Long Branch, New Jersey

Summary. An important aspect in the design and construction of semiconductor devices is the measurement and control of physical and electrical properties of materials. Recently, new techniques have been developed using microwave transmission and reflection to determine the physical constants of crystals such as the bulk lifetime, surface recombination velocity, resistivity and dielectric constant. The new methods appear considerably simpler than some of the more conventional techniques. Sample preparation is minimized with respect to surface conditions and electrodeless measurements are possible. Measurement of crystal characteristics as a function of temperature is also facilitated. One of the most recent developments involving the profiling of ingot resistivity and lifetime by a microwave reflection method promises to be as accurate and less complicated than the more conventional dc conductivity techniques.

Introduction

In the study of the interaction of microwaves with bulk semiconductors, a means exists of studying the electrical properties of materials. Several recent reports have indicated that as microwave energy is transmitted through a semiconductor medium, the presence of excess free carriers can be detected by measurements of transmission, reflection or absorption of the electromagnetic power. Gibson,[1] Benedict[2] and Shockley, Harrick[3] and Ramsa[4] have used the techniques of microwave and infrared transmission to measure such properties as dielectric constant, mobility, conductivity and lifetime. In addition, others[5] have suggested possible device applications inherent in the interaction of microwaves and bulk semiconductors. In what follows, a description will first be given of more recent experiments in the study of physical properties of semiconductors and second, it will be shown that the same techniques can be applied in the development of new devices.

Consider reflections in air due to the presence of a semiconductor medium infinite in the x, z plane and finite in the -y direction. The reflections here are analogous to the reflections from a waveguide filled with a semiconductor in the -y region and air in the +y region as shown in Fig. 1, or with a transmission line which has a propagation constant, Γ_{01}, and impedance, Z_{01}, in the +y region and Γ_{02} and Z_{02} in the -y region. Now if an incident wave is partially reflected, the magnitude and phase of the reflected wave will depend upon the conductivity of the material amongst other factors. We note that if the conductivity is changed by the creation of excess minority carriers due to light or direct current injection, the amount of reflection will change.

Next consider the three medium problem (Fig. 2) where a wave approaches from the right through air and is normally incident upon an infinite sheet of semiconductor of thickness l_2. The analogue for a waveguide with a germanium plug inserted in a section is a transmission line made up of three characteristic sections. In the case describing three media we can see that as the conductivity of medium 2 is changed, that is if the conductivity of the semiconductor is varied, the transmission coefficient, reflection coefficient and absorption will change. If high resistivity semiconductors are used, the changes will be large in both amplitude and phase.

Resistivity and Dielectric Constant by Reflection

To measure resistivity and/or dielectric constant, one of the simplest approaches is the reflection system in Fig. 1. We shall offer some quantitative estimates for a typical case at 10,000 Mc/second. Let us assume an infinite rod of semiconductor material surrounded by the waveguide structure with its own waveguide impedance and propagation constant, Z_{02} and Γ_{02} respectively. The basic equation for the reflection coefficient in a waveguide is given by,

$$r_R = \frac{Z_{02} - Z_{01}}{Z_{02} + Z_{01}} . \qquad (1)$$

To obtain Z_{01} we use the following relations,

$$\Gamma_{01} = j \frac{2\pi}{\lambda_1} \sqrt{\left[1 - \left(\frac{\lambda_1}{\lambda_0} \right)^2 \right]} \qquad (2)$$

and

$$Z_{01} = \frac{j\omega\mu}{\Gamma_{01}} , \qquad (3)$$

and where

$\omega = 2\pi \times 10^{10}$ cycles/second,

$\lambda_0 = 4.56 \times 10^{-2}$ meters, the cutoff wavelength at X-band,

$\lambda_1 = \frac{3.0 \times 10^8}{10^{10}} = 3.0 \times 10^{-2}$ meters, the wavelength in free space,

$\mu = 4\pi \times 10^{-7}$ henry per meter.

To obtain Z_{02} we use,

$$\Gamma_{02} = j \frac{2\pi}{\lambda_2} \sqrt{\left[1 - \left(\frac{\lambda_2}{\lambda_0}\right)^2\right]} - j \frac{\sigma}{\omega\epsilon}, \qquad (4)$$

and

$$Z_{02} = \frac{j\omega\mu}{\Gamma_{02}}, \qquad (5)$$

where

$\omega = 2\pi \times 10^{10}$ cycles/second,

$\lambda_0 = 4.56 \times 10^{-2}$ meters,

$\lambda_2 = 3.0 \times \frac{10^{-2}}{\sqrt{\epsilon_r}}$ meters, the wavelength in lossless germanium,

$\mu = 4\pi \times 10^{-7}$ henry per meter,

$\epsilon = \epsilon_o \epsilon_R$ farad/meter.

In addition, the conductivity, σ, will be varied throughout the useful range of semiconductors.

In Fig. 3 we see the effects of varying the dielectric constant $\epsilon_R = \frac{\epsilon}{\epsilon_o}$, and at the same time changing the conductivity of the semiconductor. In the region of higher conductivity, that is above 10 (ohm meter)$^{-1}$, the reflection coefficient magnitude is independent of the dielectric constant and hence does not depend upon the material. In the region below approximately 10 (ohm meter)$^{-1}$, the reflection coefficient magnitude is more dependent upon the dielectric constant. This indicates that if one would like to measure the dielectric constant for a relatively pure high resistivity material, the determination of the reflection coefficient would provide a rather accurate means. In Fig. 4 the magnitude of the reflection coefficient is given as a function of various dielectric constants with values ranging from 1 to 20. What is proposed then is as follows: the conductivity and/or dielectric constant can be determined at a frequency of 10^{10} cycles per second by inserting the semiconductor rod in the waveguide and measuring the reflection coefficient.

A further approximation can be made in placing a large block, or ingot of semiconductor, over the open end of the air filled section of waveguide. Since the dielectric constant of germanium and silicon is so high, the calculations using a material filled waveguide gives very nearly the same results as the case of reflections from the large crystal not enclosed by waveguide walls.

Experiments have been conducted to test the validity of the calculations. Here, using the technique of placing a large ingot adjacent to the opening of the air filled waveguide, reflection measurements were made to determine the resistivity of crystals measured with four point probe techniques. The agreement between the two methods is good, as shown in Tables I and II. Furthermore, in the resistivity measurements by microwave reflection, the values are not critically dependent upon surface preparation. In these cases, ϵ_R, was taken at 16 for germanium and 12 for silicon. However, in the range studied, the reflection is dominated by conductivity and the dielectric constant plays a minor role.

It was found that these techniques could be applied to lifetime measurements. In this case the surfaces were polished, etched, washed and dried. A pulsed light irradiated the crystal, either at the point of reflection or nearby. By measuring the decay in reflection changes, a measure of lifetime could be obtained. Experiments have been tried correlating the microwave reflection technique with more conventional dc conductivity methods and have shown good agreement (see Table III).

This technique can be applied to frequencies or temperatures where inertial effects come into play. The equations which give the effective conductivity, σ', and permittivity, ϵ', are given in Eqs. 6 and 7. Here, using the Drude-Zener theory of complex conductivity,

$$\sigma' = \sigma_o \frac{1}{1 + \omega^2\tau^2}, \qquad (6)$$

$$\epsilon' = \epsilon \left(1 - \frac{1}{\omega\epsilon} \sigma_o \frac{\omega\tau}{1 + \omega^2\tau^2}\right). \qquad (7)$$

It turns out that one can substitute ϵ' and σ' directly into the nomograph shown in Figs. 3 and 4, or stated conversely, if one measures the reflection coefficient and σ' in the region of inertial effects one can determine ϵ'.

Resistivity and Transmission

Resistivity can also be measured by transmission. Consider a transmission system consisting of a waveguide, in which electromagnetic energy is being propagated in the TE_{01} mode. Assume that a slab of semiconductor is inserted in the waveguide as shown in Fig. 2, where the large surface of the semiconductor slab is oriented perpendicularly to the direction of propagation. The following theory will apply to the measurement of E_o/E_{in}, or the ratio of transmitted electric field to the incident electric

field. This ratio is given by

$$\frac{E_o}{E_{in}} = r_t \left(\cosh \Gamma_{02} l_2 - \frac{Z_{02}}{Z_{ab}} \sinh \Gamma_{02} l_2\right), \quad (8)$$

where

$$r_t = \frac{2Z_{ab}}{Z_{ab} + Z_{01}}, \quad (9)$$

and

$$Z_{ab} = Z_{02}\left(\frac{Z_{03} + Z_{02} \tanh \Gamma_{02} l_2}{Z_{02} + Z_{03} \tanh \Gamma_{02} l_2}\right), \quad (10)$$

where

Z_{01} is the impedance of the line (waveguide) in air and is equal to Z_{03},

Z_{02} is the impedance of the waveguide filled with germanium and is a function of conductivity and wavelength,

Z_{ab} is the impedance of the germanium at the front surface,

Γ_{01} is the propagation constant in air in the waveguide,

Γ_{02} is the propagation constant in germanium in the waveguide,

l_2 is the thickness of the slab of semiconductor.

Typical computations of $\left|\frac{E_o}{E_{in}}\right|$ at 10,000 megacycles are shown in Fig. 5. The use of this type of curve, together with measurements of thickness and frequency, allow the determination of conductivity. By varying the temperature of the sample in the waveguide by means of an oven heating a section of the guide, conductivity has been measured as a function of temperature. Results are shown in Table IV indicating the applicability of the new techniques.

At very reduced temperatures σ' and ϵ' in (6) and (7) can be substituted directly, (8), (9) and (10) to take into account inertial effects of free carriers.

Lifetime and Surface Recombination Velocity

The principle of the measurement can be explained with the aid of Figs. 6 and 7. Here a germanium or silicon sample is cut so that it is rectangular in shape and can be fitted into an X-band waveguide. It is then placed in the thermal control unit and the microwave power is transmitted through the semiconductor block at normal incidence to the largest face. With reference to Fig. 6, as a light pulse is flashed on the side edge of the semiconductor, excess carriers are created resulting in an absorption of microwave power. At the cessation of light, the carriers recombine and the power transmitted is increased. The time constant of the increase in power, as viewed on an oscilloscope, is assumed to be the lifetime of excess minority carriers. The thermal control unit as shown in Fig. 7 was made up of the following elements. A and D represent a large brass block, C is a sliding holder for the semiconductor sample B, and E represents a quartz rod through which the light pulse can pass and irradiate the end of the semiconductor block. It can be noted that for a large block the light does not strike the large planar surfaces and hence after the carriers diffuse into the sample, the microwave energy is absorbed largely by carriers in the bulk material. The temperature was controlled by a container filled with fluid (not shown) soldered to the brass block A. In addition, a thermocouple was inserted by means of a small hole and screw in the brass block D just under the sliding holder. The microwave system was closed by teflon windows inserted in the waveguide and flushing gases or a vacuum could be utilized.

At the outset of the experiments it was determined that for high resistivity germanium, the conventional dc conductivity value of lifetime and the microwave transmission technique of measuring lifetime gave substantially the same result (see Table III). Correlation was established on the same materials in both methods of testing, in air at room temperature. Following this, tests were run in the microwave system with various samples of germanium and silicon in an ambient of dry oxygen and vacuum at various temperatures. In general, thin samples were prepared for surface recombination studies, and from the same crystal thicker samples were prepared in order to determine the bulk lifetime. In following this procedure, two objectives were to be accomplished. First, a check could be made of the work of previous research in the field and second, new physical effects would be examined.

A wide variety of samples was examined and in general, all results were surprisingly consistent with the Shockley-Read theory of bulk lifetime and the Stevenson and Keyes modification for surface recombination. A typical run for the variation of the product of surface recombination velocity and temperature as a function of the reciprocal of temperature is shown in Fig. 8. Here the ambient in oxygen and the microwave technique is used. Similarly in Fig. 9 we see typical results in germanium. Now the Shockley-Read theory predicts that the lifetime for bulk semiconductors is given by,

$$\tau = \frac{\tau_{po}(n_o + n_1) + \tau_{no}(p_o + p_1)}{n_o + n_p}, \quad (11)$$

where τ is the measured lifetime, τ_{po} is the lifetime of holes injected into highly n-type material, τ_{no} is the lifetime of electrons injected into highly p-type material, and n_o, p_o, n_1 and p_1 are the equilibrium density of free electrons, free holes, the density of free electrons if the Fermi level were at the same energy level as the trapping centers, and the density of holes if the Fermi level were at the same level as the trapping

centers respectively. The application of this theory to experimental results gives use to the data in Table V where energy levels are determined as a result of temperature tests. The energy levels ϵ_t, are interpreted as follows. For p-type silicon, $\epsilon_t = E_t - E_v = .065$ ev is the difference in the level of the recombination centers and the valence band. For thin silicon slices $\epsilon_t \cong .15$ ev $= E_t - E_v$ at the surface of the particular silicon sample in oxygen. For thick type germanium, $\epsilon_t = E_c - E_t = .20$ ev, the difference in energy of the conduction band and the trapping centers and for thin germanium in oxygen and vacuum, $\epsilon_t = E_c - E_t \cong .25$ ev and .12 ev represent the differences in energy levels at the surfaces respectively. We conclude here that the microwave techniques represent a simple, electrodeless technique for measuring recombination processes.

Other phenomena such as injection and extraction can also be studied using similar techniques. This has been accomplished by forming the contacts on the side of the semiconductor slug and biasing in the forward or reverse direction. Changes in conductivity can be readily observed by the detector in the microwave transmission system.

Devices

In arrangements similar to that shown in Fig. 10, modulators of microwave carrier frequencies have been developed by biasing the semiconductor so that a diode injects excess minority carriers. The transmission or reflection can be varied to give either amplitude or phase modulation or both, depending on the geometry of the semiconductor block. The limit of frequency is determined by the magnitude of $\omega\gamma$. Preliminary calculations indicate that these devices should operate up to frequencies of 10^{12} to 10^{13} cycles per second.

Conclusion

A general theory of microwave transmission and reflection in semiconductors has been given. Based on this, new techniques for measurement of physical properties of materials has been described i.e., resistivity, dielectric constant, lifetime, surface recombination velocity and injection efficiency. In addition, device possibilities from the microwave region to the far infrared regions of the frequency spectrum are suggested.

Acknowledgment

Acknowledgment should be made to Dr. W. Caldwell, Bendix Semiconductor Division, Holmdel, New Jersey for his suggestions, conversations and encouragement of the work reported here.

References

1. A. F. Gibson, "Progress in semiconductors," John Wiley & Sons, Inc., New York, N. Y., vol. 2, pp. 229-245; 1957.

2. T. S. Benedict and W. Shockley, "Microwave observation of collision frequency of holes in germanium," PHYS. REV., vol. 91, pp. 1565-1566; 15 September 1953.

3. N. J. Harrick, "Metal to semiconductor contacts: injection or extraction for either direction of current flow," PHYS. REV., vol. 115, pp. 876-882; 15 August 1959.

4. A. P. Ramsa, H. Jacobs and F. A. Brand, "Microwave techniques in measurement of lifetime in germanium," J. APPL. PHYS., vol. 30, pp. 1054-1060; July 1959.

H. Jacobs, F. A. Brand, J. D. Meindl, M. Benanti and R. Benjamin, "Electrodeless measurement of semiconductor resistivity at microwave frequencies," PROC. IRE, vol. 49, no. 5, pp. 928-932; May 1961.

5. H. Jacobs, F. A. Brand, J. D. Meindl, and R. Benjamin, "Multiple reflections of microwaves propagating through a semiconductor medium," PROC. IRE, vol. 49, no. 11, pp. 1683-1684; November 1961.

A. F. Gibson, "Infrared and microwave modulation using free carriers in semiconductors," J. SCI. INSTR., vol. 35, pp. 273-278; August 1958.

TABLE I

COMPARISON OF FOUR POINT PROBE AND MICROWAVE REFLECTION TECHNIQUES FOR MEASUREMENT OF RESISTIVITY

Sample Description	Remarks	ρ in Ω cm (4-Point Probe Resistivity)	ρ in Ω cm (Reflection Resistivity)
Sample No. 1 Germanium Ingot	Unpolished, dry surface	20	18.3
Sample No. 2 Germanium Ingot	600 mesh polish, dry surface	1.18 - 1.28	1.20
Sample No. 3 Ge Ingot - 2 GET	Rough polish, dry surface	10 - 15.9	13.0
Sample No. 3 Ge Ingot - 2 GET	Same - opposite end	11.3 - 18.6	14.8
Sample No. 4 Ge Ingot - 2 GE 399	Rough polish, dry surface	10.7 - 16.9	12.7
Sample No. 5 Ge Ingot - 86 AD	Rough polish, dry surface	3.9 - 4.7	4.0
Sample No. 6 Ge Ingot - A 53 EA	Rough polish, dry surface	8.71 - 14.7	11.8
Sample No. 7 Si Ingot - center slice	Opening not completely covered, dry surface	$\approx$ 1.0	1.0

TABLE II

SURFACE PREPARATION AND RESISTIVITY MEASUREMENT
BY REFLECTION TECHNIQUE

Sample Description	Remarks	Resistivity (Ω_{cm})
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	Polished with 600 mesh, standing in air 9 months	1.2
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	Wet with tap water	1.2
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	300 mesh polish, water rinse, left wet	1.18
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	300 mesh polish, water rinse, standing 1 week in air	1.5
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	600 mesh polish, water rinse, heat lamp dry	1.15
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	600 mesh polish, water rinse, heat lamp dry	1.5
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	CP4 A etch, water wash, heat dry	1.5
Sample No. 2 - Ge Ingot .4 x .9 x 1.1 in 4 point probe, ρ = 1.18 - 1.28	Roughed with #32 emory paper, water wash and heat dry	1.6

TABLE III

COMPARISON OF LIFETIME DATA BY METHODS OF DC PHOTOCONDUCTIVITY, MICROWAVE TRANSMISSION VARIATIONS AND MICROWAVE REFLECTIONS

Sample No. and Type	Lifetime by Microwave Transmission in Microseconds	Lifetime by dc Photo-conductivity in Same Holder in Microseconds. Bias Current at 5 ma	Lifetime by Reflection of Microwaves in Micro-seconds	Lifetime by dc Photoconductivity in Same Holder in Microseconds. Bias Current at 2.5 ma
No. 3 - Ge* .257xl.035x2.17 cm 47 ohm cm, n-type	680	800	700	700
No. 4 - Ge* .362xl.035x2.15 cm 47 ohm cm, n-type	1200	1300	1200	1250
No. 5 - Ge* .550xl.210x2.15 cm 47 ohm cm, n-type	1250	1250	750 to 900	1225
No. 6 - Ge** .550x.963x2.08 cm 47 ohm cm, n-type	450	450	475	450 to 475
No. 8 - Ge*/*** .437x.960x2.15 cm 20 ohm cm, n-type	800 at 5 ma 1000 at 3 ma 1225 with no current	900 at 5 ma 1100 at 3 ma	1150	1100
No. 10 - Ge** .241x.750x2.10 cm 8 ohm cm, p-type	90	90	100	110
No. 13 - Si** About .4xl.000x2.00 cm 3000 ohm cm -6000 ohm cm, p-type	275		300, with very bright intens-ity light superimposed on light pulse - 600	With very high intensity light - 600

*Samples exposed to air for one (1) week prior to measurement.
**Surfaces slightly abraded during etching, washing and plating of contacts. Samples tested immediately after washing and drying.
***Sample No. 8 was the only germanium unit which showed variations in lifetime with dc current during test.

TABLE IV

ENERGY BAND GAP AND RESISTIVITY

Trial Run	Experimental Band Gap Near 300°K in ev	Experimental Resistivity in ohm cm at 300°K	Band Gap Near 300°K as Shown in Literature	Resistivity in ohm cm at 300°K as Shown in Literature
1	.70	46.1	.72	47.0
2	.75	45.3	.72	47.0
3	.75	52.5	.72	47.0
4	.75	46.4	.72	47.0
5	.74	48.4	.72	47.0

TABLE V

BULK LIFETIME AND SURFACE RECOMBINATION

Description	Ambient	ϵ_t in ev	Magnitude	Test Run Number
Silicon (thick) .447 cm	Oxygen	.065	τ_o = 620 μsec at 300°K	S-23
Silicon (thin) .077 cm	Oxygen	.142	ST = 3.6×10^5 at $\frac{1}{T} = 3 \times 10^3$.	S-9
Silicon (thin) .059 cm	Oxygen	.153	ST = 4.1×10^4 at $\frac{1}{T} = 3 \times 10^3$	O-14
n-type, 20 ohm cm Germanium (thick) .435 cm	Oxygen	.203	τ_o = 1350 μsec at 300°K	O-21
n-type, 20 ohm cm Germanium (thick) .455 cm	Vacuum	.192	τ_o = 1280 μsec at 300°K	A-5
n-type, 20 ohm cm Germanium (thin) .0283 cm	Oxygen	.265	ST = 6.5×10^4 at T = 300°K	A-26
n-type, 20 ohm cm Germanium (thin) .0354 cm	Oxygen	.244	ST = 4.95×10^4 at T = 300°K	A-23
n-type, 20 ohm cm Germanium (thin) .064 cm	Vacuum	.121	ST = 3.3×10^4 at T = 300°K	T-12

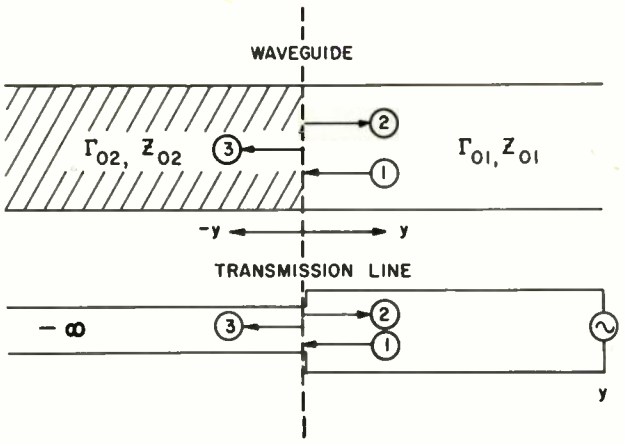


Fig. 1. Transmission line analogue for reflection.

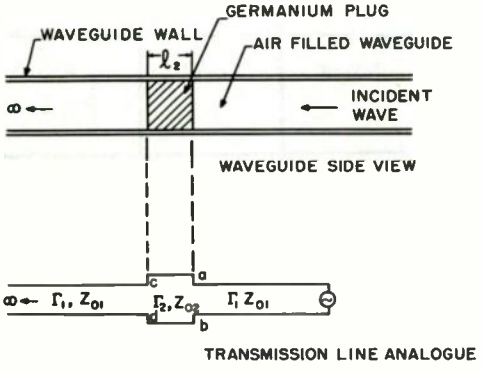
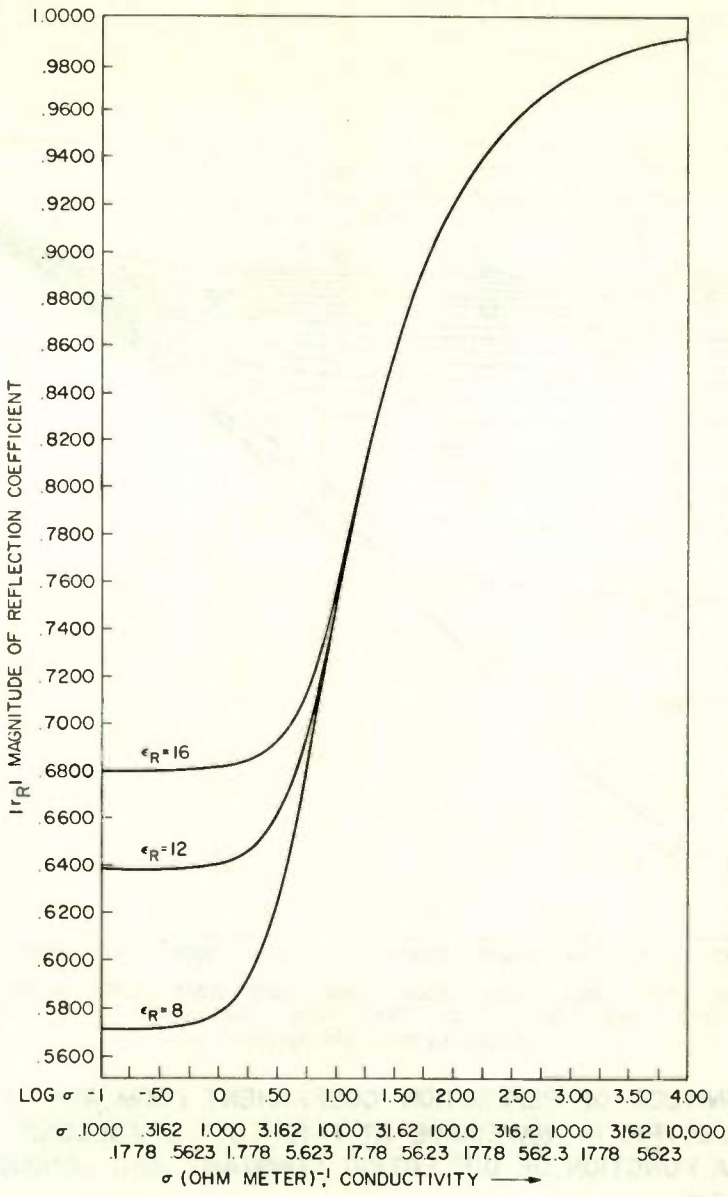
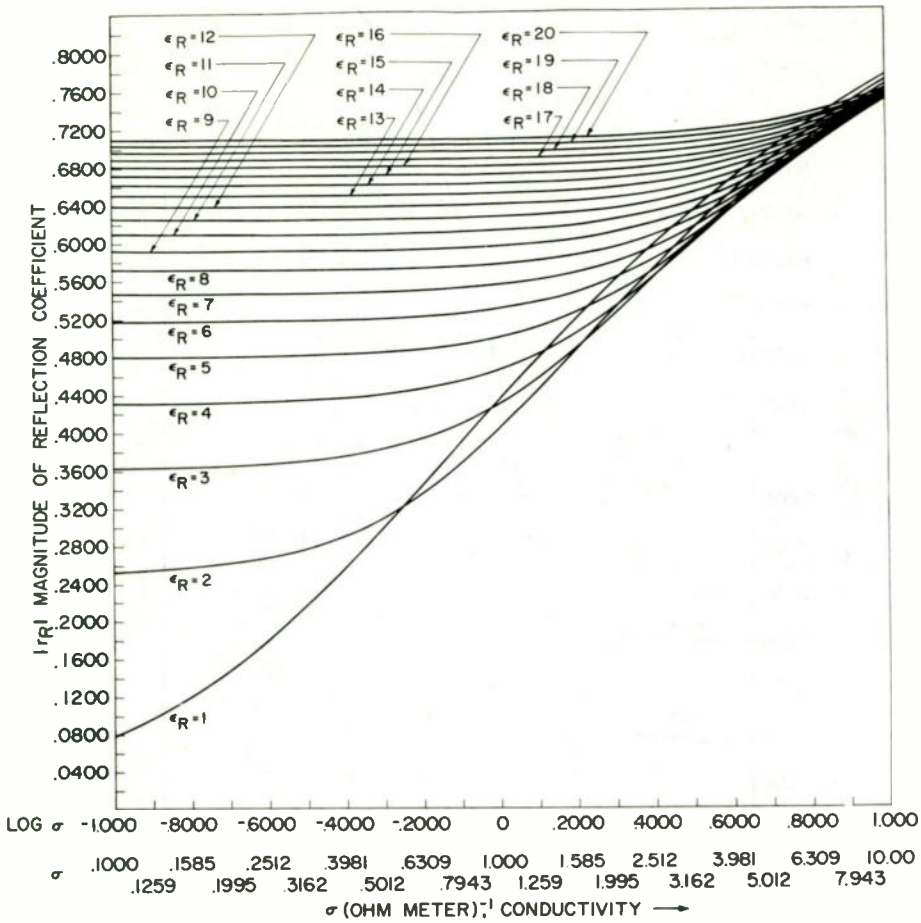


Fig. 2. Transmission line analogue for transmission in air, semiconductor and air.



MAGNITUDE OF REFLECTION COEFFICIENT, $f = 10^{10}$ CYCLES/SECOND, ALL IN WAVEGUIDE AIR TO SEMICONDUCTOR, AS A FUNCTION OF CONDUCTIVITY AND DIELECTRIC CONSTANT.

Fig. 3.



MAGNITUDE OF REFLECTION COEFFICIENT FROM AIR TO DIELECTRIC IN WAVEGUIDE AT $f = 10^{10}$ CYCLES / SECOND AS A FUNCTION OF DIELECTRIC CONSTANT AND CONDUCTIVITY

Fig. 4.

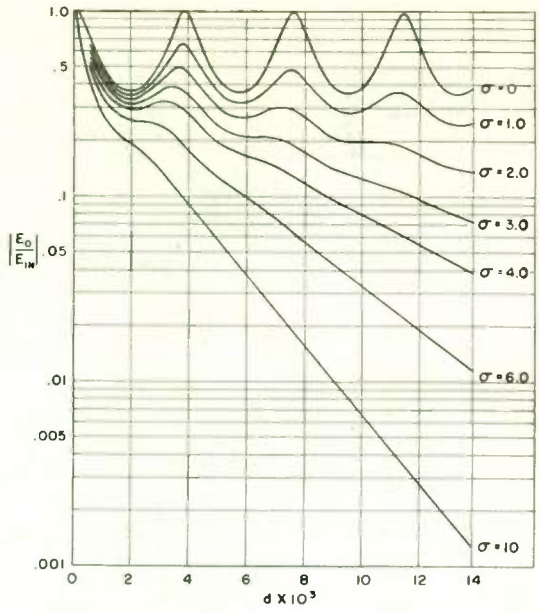


Fig. 5. Magnitude of transmission coefficient for germanium block in X-band waveguide at f = 10^{10} cycles per second. Dielectric constant is assumed to be 16.

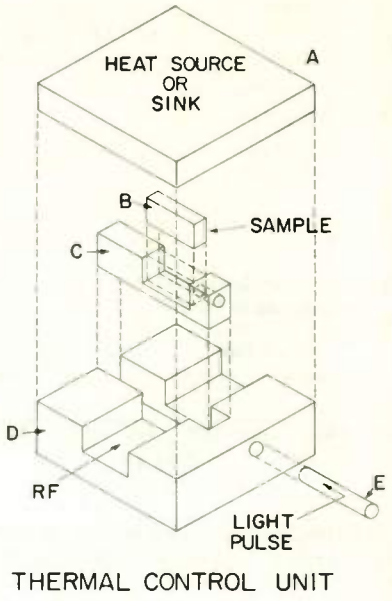


Fig. 7. Thermal control unit and holder for germanium sample.

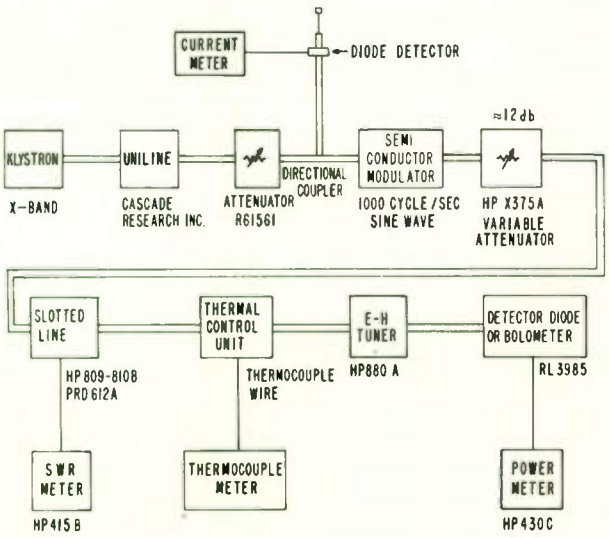


Fig. 6. Experimental arrangement for measurement of transmission coefficient.

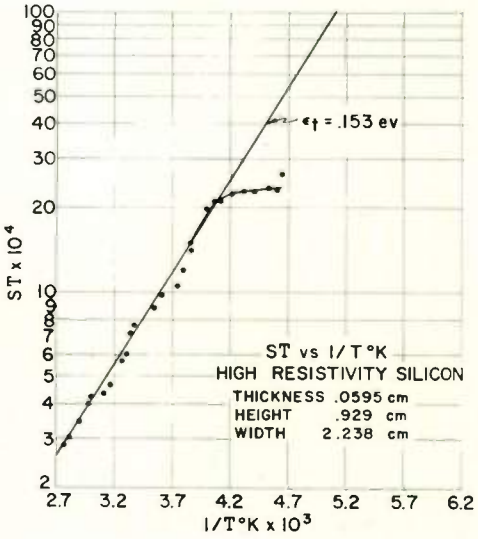


Fig. 8. Product of surface recombination velocity and temperature as a function of the reciprocal of temperature for a thin sample of silicon (resistivity in the range of 3000 to 6000 ohm cm). The ambient is oxygen.

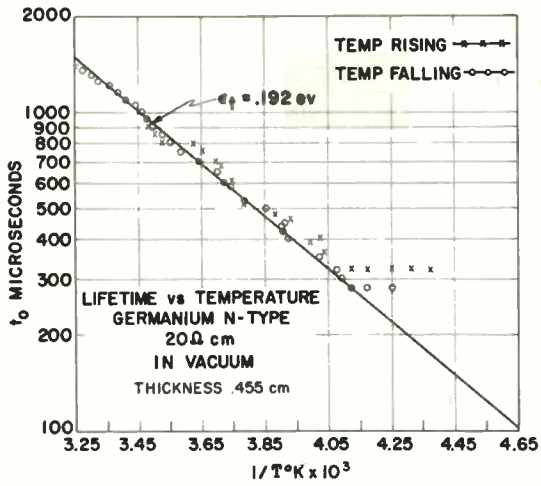


Fig. 9. Lifetime as a function of the reciprocal of temperature for a thick block of 20 ohm cm n-type germanium in vacuum.

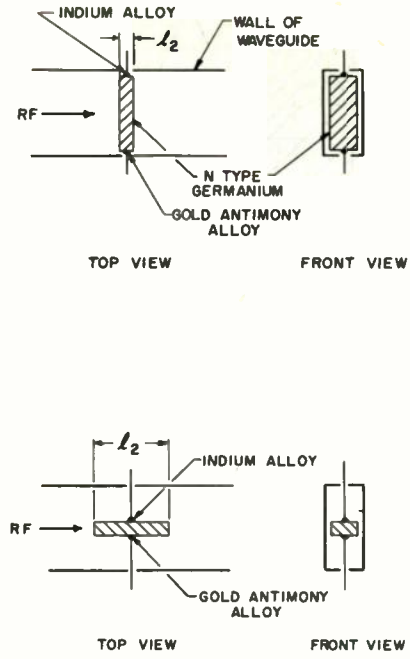


Fig. 10. Design of modulator structures to give amplitude and/or phase modulation of microwaves.

A HIGH-SPEED, MEDIUM-POWER, ALL-DIFFUSED GERMANIUM NPN MESA TRANSISTOR

R. M. Folsom, W. A. Pieczonka,
P. P. Castrucci, and M. M. Roy
Components Division
International Business Machines Corporation
Poughkeepsie, New York

Summary

To improve ferrite-core and thin-film memories, transistors that can switch substantial currents through useful voltage swings in very short periods of time are needed. This paper describes the design and fabrication of one such device, a high-speed, medium-power, all-diffused germanium NPN mesa transistor. The electrical parameters that have a bearing on the transient switching time of the transistor, such as base input resistance, junction capacitances, and gain-bandwidth product, are discussed.

The fabrication process relies on the use of SiO or SiO_2 films as a diffusion mask for N-type impurities. The relative merits of evaporated SiO films and thermally deposited SiO_2 films for device fabrication are discussed and the resulting device performance is described.

Of the electrical performance characteristics, the most important is the switching transient time. The transistors switch one ampere of current through a 20-ohm resistive load in less than 10 nanoseconds, with a V_{CE} of less than three volts at the end of 25 nanoseconds. Other essential characteristics of this device are: collector breakdown voltage, 50 volts; emitter breakdown voltage, five volts; f_T, $\geq$ 150 megacycles; common emitter current gain at one ampere of I_c, > 100; and thermal resistance, $\leq 10^{\circ}$ C/watt.

I. Introduction

This paper describes the development of a germanium transistor for use in nonsaturating solid-state memory circuits to supply the power needed to switch memory storage elements such as ferrite cores and thin films. In such a device, a compromise between power-handling ability and switching speed must be made. At one extreme the voltage swing and the current switched must be great enough to handle a practical number of memory elements, while at the other extreme greater and greater demands on switching speed are being made. To compound the difficulties, the saturation voltage drop must be sufficiently low to prevent destruction of the transistor due to excessive power dissipation while in the "on" state. The transistor developed to do this is an NPN double-diffused device, where the two diffusions produce the base and emitter regions respectively. First some design considerations will be discussed, followed by a description of the fabrication process, the results of the work, and some conclusions.

II. Design Considerations

As mentioned previously, the design of a medium-power switching transistor is usually a compromise between power-handling capacity and switching speed. These two factors are not independent, and specifying one generally sets a theoretical limit on the other.

In the present case, the power-handling capacity was specified by circuit applications; that is, the device must switch one ampere through a 20-volt swing. For an NPN germanium device to do this, BV_{CBO} should be at least 2 1/2 to 3 times this value, or about 60 volts, to provide a margin of safety with respect to BV_{CEO}, which should be equal to or greater than V_{CC}. The design problem, then, was to meet this breakdown voltage and the one-ampere current specifications while minimizing rise and fall times and V_{CE}. With these criteria in mind, the physical parameters for this device may be chosen in the following way.

First, consider the parameters that control the rise and fall times of a switching device and the limits imposed on these parameters by switching current, switching voltage, and V_{CE} requirements. The turn-on time (t_{on}) of a switching device can be approximated with reasonable accuracy by the equation:

$$t_{on} = \frac{Q}{I_b} = \frac{Q_1 + Q_2 + Q_3}{I_b} \qquad (1)$$

In other words, there exists a charge Q made up of three components that must be supplied by the base current during the rise time. Q_1 is the charge that must be replaced in the emitter depletion region and is given by:

$$Q_1 = \int_0^{V_{BE}} C_e dV_{be} \qquad (2)$$

Q_2 is a similar charge that has been swept out of the collector depletion region and is given by:

$$Q_2 = -\int_{V_{CC}}^{V_{CE}} C_c dV_{cb} \qquad (3)$$

Q_3 is the charge necessary to neutralize the injected minority carriers in the base region, and is given by:

$$Q_3 = \frac{I_c W_b^2}{2D} \qquad (4)$$

where W_b is the base width and D the minority carrier diffusion constant.

Assuming linearly graded junctions as a result of the double-diffusion process, and substituting for C_e and C_c as follows,

$$C_e = \frac{\epsilon A_E}{d_e}, \quad C_c = \frac{\epsilon A_C}{d_c}, \qquad (5)$$

where ϵ is the permitivity of germanium, A_E and A_C are the emitter and collector junction areas respectively, and d_e and d_c are the emitter and collector junction depletion widths respectively. Poisson's equation, $\nabla^2\psi = \frac{\rho}{\epsilon}$, can be solved for the junctions and in this particular case results in a final engineering equation:

$$t_{on}(ns) \approx \frac{2.88 \times 10^{-11}}{I_b}\left[A_E a_e^{1/3} V_{BE}^{2/3} + A_C a_c^{1/3}\left(V_{CC}^{2/3} - V_{CE}^{2/3}\right)\right] + \frac{I_c}{I_b}\,\frac{6.4 \times 10^3 W_b^2}{2D} \qquad (6)$$

where

A_E = Emitter area in mils^2
A_C = Collector area in mils^2
A_e = Emitter junction impurity gradient in cm^{-4}
A_c = Collector junction impurity gradient in cm^{-4}
W_b = Base width in mils
D = Minority carrier diffusion constant in cm^2/sec

From Equation (6) we can see that the device parameters (in a first-order approximation at least) that have a bering on the turn-on time of the transistor, and the ones that can be varied readily, are the junction areas and the base width. However, the first term in Equation (6) represents stored charge in the emitter transition capacity and results in a turn-on delay, which in the present device has proved to be very small and thus can be neglected. Thus for all practical purposes the only parameters that affect the turn-on transient or rise time are collector area and base width.

Obviously, then, the collector area and base width should be made as small as possible. As mentioned previously, however, other requirements impose limits on these two parameters; they are switching voltage and current, and V_{CE}. Thus the collector junction depletion width in the base, which is dependent upon operating voltage, fixes the minimum base width. The necessary collector junction area is dictated by the maximum current density at which the device will perform satisfactorily, and by the allowable extrinsic collector series resistance that will determine the value of V_{CE} at a particular collector current. The relationship between V_{CE} and other device parameters is given by the expression:

$$V_{CE} \approx 4 \times 10^2 \; I_c \; \frac{\rho W_c S}{A_C} \qquad (7)$$

where

ρ = Resistivity of the collector bulk in Ω-cm.
W_c = Thickness of the collector bulk in mils
A_C = Collector area in mils^2
I_c = Collector current in amps
S = Spreading factor ($\approx$ 0.8)

This expression is only approximate because of the effect of two other factors, which are difficult to assess. One is that the current density across the collector junction cannot be considered to be uniform, because an emitter area smaller than the collector area as well as transverse currents in the base may cause emitter crowding. The other factor is spreading resistance in the extrinsic collector bulk. These two factors have opposite effects on V_{CE} and tend to cancel each other although the former probably has the greater effect.

Again referring to Equation (7), the resistivity of the collector bulk also has an effect on V_{CE}, such that by reducing this resistivity an improvement in V_{CE} should be realized. However, the problem arises, if ρ is reduced, of obtaining the necessary collector-junction breakdown voltage, so that a minimum value of ρ is set primarily by practical processing techniques.

Upon weighing the previous considerations, the following physical parameters were chosen:

ρ = 0.1 Ω-cm
A_E = 150 mils^2
A_C = 400 mils^2
W_b = 0.08 mil
W_c = 5 mils

If 1 ampere is switched through a 20-volt collector swing with a base drive of 0.1 ampere, the preceding equations predict:

$$\tau_r = 4.2 + 2.8 = 7 \text{ nsec}$$

$$V_{CE} = 1.0 \text{ v}$$

For reasons mentioned previously, the value for V_{CE} should be a little higher.

Equations (1) through (7), and definitions of the terms used in these equations, are summarized in Figures 1 and 2.

III. Device Fabrication

The following is a brief outline of the process steps used to fabricate the double-diffused transistor.

Starting Material

The starting material is an antimony-doped, 0.1 Ω-cm, germanium wafer etched to 10 mils thickness and cut into a square 0.480 inches on edge (see Figure 3a).

Diffusion Blotter

A film of pure tin, ≈0.01 mil thick, is evaporated onto one side of the wafer. This serves to "blotter" such contaminants as copper during the base diffusion.

Base Diffusion

The wafer is placed in a one-zone dynamic indium diffusion furnace for a period of 20 hours at 860°C. The source is a tin-indium alloy and the carrier is forming gas. Through suitable control of the furnace conditions, an indium diffusion results with the following parameters, as shown in Figure 3b.

C_o = 1.5 x 10^{17}/cm^3
x_{jc} = 0.25 mil

Mask for Emitter Diffusion

The second diffusion, the emitter diffusion, produces the emitter junction and determines the base width of the transistor. In order to prevent undesirable channeling due to n-type inversion layers on the base region, the emitter must be of a planar structure. To accomplish this, the side of the wafer with the indium, or base diffusion, is coated with a film of SiO_2 approximately 2000 Å thick, as shown in Figure 3c. By using standard photolithographic techniques, an opening is etched in the SiO_2 film to allow arsenic to diffuse into the wafer in a selected area. Before the deposition of the SiO_2, the tin blotter is lapped off.

Emitter Diffusion

The emitter diffusion is carried out in an evacuated quartz capsule containing the wafer and a few milligrams of pure arsenic. The diffusion is done in 30 minutes at 700°C, with the following results (as shown in Figure 3d):

C_o = 6 x 10^{19}/cm^3
x_{je} = 0.17 mil
W_b = 0.08 mil

Figure 4 is a drawing of the final net impurity concentration in the wafer.

Ohmic Contacts

The SiO_2 film is removed from the wafer and ohmic contacts are deposited through a metal mask on the emitter and base regions, by standard vacuum-evaporation techniques, and alloyed with the wafer (see Figure 5a). The contacts are silver alloys. The contacts are covered with a thin film of evaporated SiO during the alloy step to prevent spreading of the metal across the emitter junction (see Figure 5b). The wafer is then lapped to a thickness of 5 mils.

Moat Etching

The contact side of the wafer is then coated with a thin film of Apiezon black wax, and a 0.002-in. wide chisel is used to plow grooves in the wax, outlining the collector area of the transistor, as shown in Figure 5c. Etching in CP_8 produces moats that bare the collector junction. The advantage of moat etching over mesa etching for laboratory fabrication is that it is more versatile, since it does not employ evaporation masks (see Figure 5d).

Final Processing

The rest of the process is rather straightforward and simple. The remainder of the black wax is rinsed from the wafer, and the wafer is cut into individual units. The units are then soldered to a header containing a copper plug; gold wires are thermocompression-bonded to the base and emitter

contacts; and the transistor is then etched electrolytically, washed, dried, baked, and capped. After baking 100 hours at 100°C, the unit is ready for electrical tests.

IV. Results

A germanium core-driver for use in nonsaturating memory circuits has been developed which will switch one ampere through a 20-volt swing. Experimental values of V_{CE} and rise time are close to the theoretically limiting values. The electrical characteristics of this device are summarized in Figure 6. The circuit used to measure rise time is also shown in this figure.

$$t_{on} = \frac{Q}{I_b} = \frac{Q_1 + Q_2 + Q_3}{I_b} \qquad (1)$$

$$Q_1 = \int_0^{V_{BE}} C_e \, dV_{be} \qquad (2)$$

$$Q_2 = \int_{V_{CC}}^{V_{CE}} C_c \, dV_{cb} \qquad (3)$$

$$Q_3 = \frac{I_c W_b^2}{2D} \qquad (4)$$

$$C_e = \frac{\epsilon A_E}{d_e}, \quad C_c = \frac{\epsilon A_C}{d_c} \qquad (5)$$

WHERE: ϵ = THE PERMITIVITY OF GERMANIUM
A_E = EMITTER AREA
A_C = COLLECTOR AREA
d_e = EMITTER DEPLETION WIDTH
d_c = COLLECTOR DEPLETION WIDTH

Fig. 1. Charge equations.

$$t_{on}(ns) = \frac{2.88 \times 10^{-11}}{I_b}\left[A_E \alpha_e^{1/3} V_{BE}^{2/3} + A_C \alpha_c^{1/3}\left(V_{CC}^{2/3} - V_{CE}^{2/3}\right)\right] \qquad (6)$$

$$+ \frac{I_c}{I_b} \frac{6.4 \times 10^3 W_b^2}{2D}$$

$$V_{ce} \approx 4 \times 10^2 \, I_c \frac{\rho W_c S}{A_C} \qquad (7)$$

WHERE: A_E = EMITTER AREA IN MILS^2
A_C = COLLECTOR AREA IN MILS^2
α_e = EMITTER JUNCTION IMPURITY GRADIENT IN CM^{-4}
α_c = COLLECTOR JUNCTION IMPURITY GRADIENT IN CM^{-4}
W_b = BASE WIDTH IN MILS
D = MINORITY CARRIER DIFFUSION CONSTANT IN CM^2/SEC
ρ = RESISTIVITY OF THE COLLECTOR BULK IN Ω-CM
W_c = THICKNESS OF THE COLLECTOR BULK IN MILS
I_c = COLLECTOR CURRENT IN AMPERES
S = SPREADING FACTOR (≈0.8)

Fig. 2. Design equations.

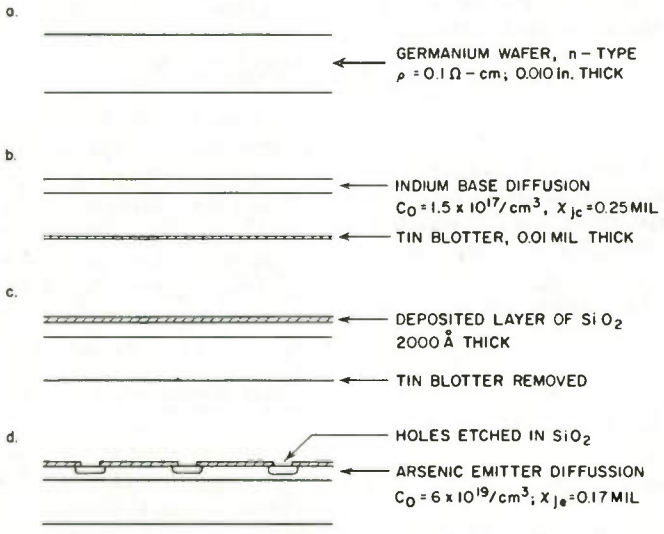


Fig. 3. Diffusion steps.

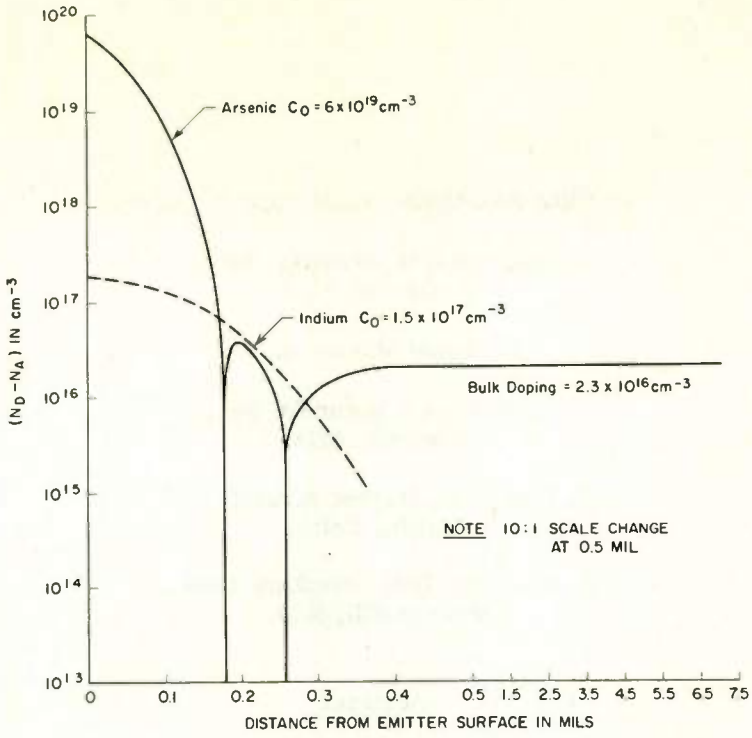


Fig. 4. Net impurity concentration in wafer.

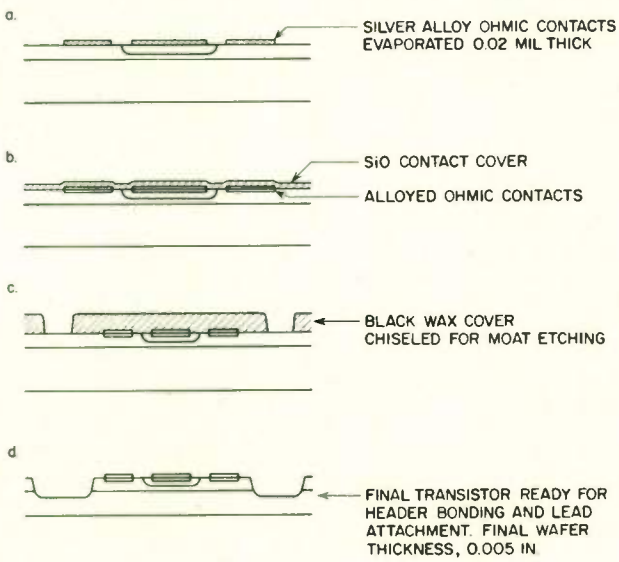
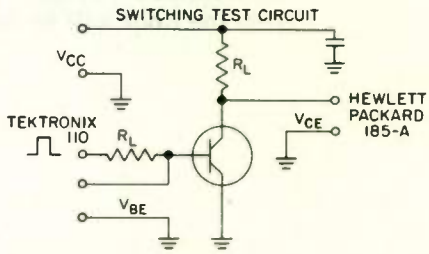


Fig. 5. Ohmic contact and mesa etching steps.



DOUBLE - DIFFUSED, NPN, Ge, CORE DRIVER
EXPERIMENTAL ELECTRICAL CHARACTERISTICS

B V_{CBO}	@ 100 μa	60 v
B V_{EBO}	@ 100 μa	8 v
B V_{CEO}	@ 1 ma	20 v
V_{CE} (SAT)	@ I_C = Ia, I_b = 0.1a	1.0 v
V_{BE}	@ I_C = Ia, I_b = 0.1a	1.0 v
τ_r	@ I_C = Ia, I_b = 0.1a	10 ns
β	@ I_C = 0.5a	150
f_T	@ V_C = 5v, I_e = 30ma	250 Mc

Fig. 6. Summary of electrical characteristics.

PANEL: BROADENING DEVICE HORIZONS

Chairman: W. Webster, RCA,
Princeton, N. J.

Panel Members:

I. A. Lesk,* Motorola, Inc.,
Phoenix, Ariz.

G. Birnbaum, Hughes Aircraft Co.,
Malibu, Calif.,

J. M. Early, Bell Telephone Labs.,
Murray Hill, N. J.

Abstract

This panel session consists of a group who were invited to talk on significant new developments in the field of electron devices which will soon increase the scope of the electronics industry. Frontier areas which will be discussed this year include generators, modulators and detectors of coherent optical radiation, semiconductor devices in the microwave region and integrated electron devices. The speakers will review developments in these fields and discuss some of the future potentialities in these forefront areas. An opportunity for comments and questions from the floor will be an important feature of the session.

*Paper follows.

INTEGRATED ELECTRON DEVICES

I. A. Lesk
Motorola, Inc.
Semiconductor Products Division
Phoenix, Arizona

Summary

Definitions previously advanced[1] to cover the solid-state device aspects of microsystems electronics[2] are slightly modified and extended into the circuitry area to cover those aspects relating to integrated circuits. The design of integrated circuits may be broken down into four steps--conventional component circuitry, compatible component circuitry, hybrid integrated circuits, and single or multiple block integrated circuits. Surface passivation is an extremely important requirement for integrated circuits. Advantages of integrated circuit applications are apparent in such areas as cost, size and shape, speed, weight, power utilization, and thermal matching. The improved reliability potentially obtainable by integrated circuits is a significant factor in their eventual large scale use, but it must be very carefully designed into them, and guaranteeing it requires quantitatively determined process reliability and control comparable to that now demonstrated on high quality conventional components.

The introduction of integrated circuits into operating electronic equipment has become gradual, starting in many cases with mixed component systems to capitalize on the advantages of particular integrated circuits as they are developed. Although standard integrated circuits will be available as catalog items, the varying requirements on integrated circuits for different applications, the continually appearing innovations in devices and circuits, and the use of proprietary circuits by equipment manufacturers indicate that a majority of integrated circuits will be special items.

Introduction

The great current interest in integrated circuits stems partially from their demonstrated as well as potential advantages, partially from their novelty, and partially from the uncertainty as to what they actually are. The term integrated circuits (and others similar to it) has been publicly applied all the way from magnetic thin films to assemblies of plug-in rectifiers. To add to the confusion, a great many process technologies that can be used and are contemplated are being identified with it. The purpose of this paper is to try to define the various device and pertinent circuit aspects of microsystems electronics in a consistent manner, in order to present a clearer perspective of integrated circuits and their present status. As much as possible, existing terms are used.

Definitions

Table 1 shows the device[3] and pertinent circuit terms to be defined for the purposes of this paper. A basic device is defined as the simplest useful device exhibiting a basic solid-state phenomenon. There are a large number of basic devices, including the various types of diodes, resistors, capacitors, inductors, etc. Figure 1 shows schematically some of the forms a resistor may take, depending on the processes used in its fabrication, and its required properties. In use, care must be taken so that the types employing p-n junctions are not appreciably forward biased, or reverse biased to exceed the breakdown voltage. The shaped types generally employ higher resistivity semiconductor material to obtain higher resistance values, hence, have larger temperature coefficients. In Figure 2 are two types of capacitors, the junction type which can have higher (but voltage variable) values, and the film type.

A complex device is a device containing within its structure more than one recognizable basic or simpler complex device, utilizing the properties of the interconnecting medium to perform a function other than electrical connection or isolation. Complex devices include the various types of transistors and many multi-junction devices. For example, in Figure 3, by putting together various combinations and types of the diode schematically illustrated on the left, we can arrive at many other commonly used structures. In the centre we may

have, for example, a bipolar, unipolar, avalanche, tunnel emitter, unijunction, or tetrode transistor, or a compensated avalanche diode, in each case only by utilizing the properties of the interconnecting (base) region other than connection or isolation. On the right in Figure 3, by designing the structures to make proper use of both base regions, we can have a controlled rectifier, or a hook collector or remote base transistor.

A distributed device is a device having within its structure more than one recognizable basic or complex device, utilizing for its performance the distributed nature of their geometries and interconnections. For example, Figure 4 shows a distributed resistance-capacitance network.

An integrated device is defined as a device containing within its structure more than one recognizable basic or complex device, in which the interconnection medium is used only for electrical connection or isolation, restricting the interconnecting medium to be of the same (or slightly modified) material as the rest of the structure. Figure 5 shows schematically integrated semiconductor diodes and resistors. Because the interconnecting medium is common to that of the device regions themselves, there can be undesired interactions other than the connection (or isolation) desired. For example, in Figure 5 top, there is not only electrical connection between the n-type regions of the diodes, but, if improperly designed, measurable transistor action as well. Hence, this would result in the forward biased on one diode producing an increase, the leakage current of the reverse biased ones can increase. In Figure 5, bottom, the contacting geometry must be proper in order that the total resistance be equal to the sum of the parts. Figure 6 is a photomicrograph showing part of a silicon wafer containing diffused resistor integrated devices. Each 20 mil square pattern contains 20 resistors in values of 100 or 1,000 ohms, all connected in series.

In the area of circuit terms, we will define a functional electronic block (FEB) as a circuit formed in such a way as to have all of its component devices within its (solid) structure, having at least some of its interconnections and/or isolations made of materials different from those comprising the devices. Figure 7 shows schematically two functional electronic blocks incorporating the same circuit. In both cases the diode (left) and transistor (center) are isolated by diffused p-n junctions. The resistor (right) is isolated from the substrate by its own p-n junction. Required interconnections are made by means of adherent metal films (black) over the insulating surface layer (cross-batched). Examples of FEBs are shown in Figures 8-11. Figure 8 shows a silicon flip-flop of the R-C coupled type (mounted on an 8-pin TO-5 header). Figure 9 shows a wafer of 5-stage Darlington amplifiers before separation into individual blocks. In Figure 10 we have a thin film FEB, the R-C pattern for a bistable circuit. Figure 11 shows a Series 51 Solid Circuit Semiconductor Network (Texas Instruments) which is encapsulated in a small package having dimensions about 250 x 125 x 35 mils. The structures shown in Figures 8, 9, and 10 we shall call monolithic, i.e., the series of processes used to make them resulted in a completely solid structure. Although the component in Figure 11 is not monolithic because of the few wire interconnections from point to point on the FEB, other units in the TI Series 51 are monolithic.[4]

We now have all the components for building integrated circuits, and will define three types in an attempt to have terms that will be valid well into the future. An integrated circuit is defined as an assembly in a single package of device and FEB component parts, containing at least one FEB or integrated device. The integrated circuit then becomes associated with the smallest "throw-away" (or replaceable) circuit item, although integrated circuits can be permanently interconnected in groups to make a larger throw-away component. A "single block integrated circuit" is one having a single FEB or integrated device in a package. Figure 12 is an illustration of this type, one of the Fairchild micrologic elements (half-shift register in this case). Figures 8 and 11 are other examples of single block integrated circuits. A "multiple block integrated circuit" will be defined as one having more than one unpackaged FEB and/or integrated device interconnected in the same package. There are many technical programs under way to develop multiple block integrated circuits, and their practical application in the near future is certain. The third category is the "hybrid integrated circuit," in which case we have at least one FEB or

integrated device, together with one or more basic, complex, or distributed devices, interconnected in the same package. Examples of hybrid integrated circuits are shown in Figure 13 (a transistor and a four-diode integrated device to make a NAND gate) and Figure 14 (a transistor, capacitor, resistor, and an integrated resistor device to make an R-C coupled amplifier stage).

There are cases where mounted devices are interconnected with FEBs to produce a more complex circuit function. We will call these "combination circuits," illustrated in Figure 15 for the case of a dual AND gate.

The term molecular electronics[5,6,7] (molectronics, molecular engineering) is becoming one more and more defining an attitude on the part of the research and development worker in that it emphasizes the utilization of many broad fields of endeavor (i.e., solid state research, process technology) toward the attainment of new electronic application. The results of work in this area may, however, be classified in terms of the definitions given above. Functional advances in solid state technology are continually increasing our store of devices, primarily the basic and complex ones. These, in turn, can be combined to give us other devices, FEBs, and integrated circuits. Advances in process technologies, for example, epitaxial crystal growth, are beneficial to all areas shown in Table 1.

Some Integrated Circuit Design Considerations

In the design of integrated circuits it is necessary to consider some process strengths and limitations. It is generally advisable to maximize the number of transistors and diodes, since they are highly developed and occupy little space. High value resistors and capacitors will take more room, depending on the types and processes used, so should be minimized. Only small values of inductance are presently available, so large inductors (and transformers) must be separate components, perhaps put inside the package of a hybrid integrated circuit. The magnetic field coupling of thin film inductors to low resistivity substrates and other components places stringent requirements on their location. Devices requiring tight control through final processing (i.e., closely matched peak current tunnel diodes) are to be minimized. The interconnection and isolation parasitics in an integrated circuit will affect the high frequency performance. In general, interconnection inductances are considerably reduced, inter-component capacitances often increased, compared to standard circuitry. Tight frequency control or filtering will still require piezoelectric elements. Power dissipation of the package, heat removal from the assembly of packages, and the topology of external circuit connections are other important design factors to be considered from the early stages.

The design of a new integrated circuit (whether special or standard) will generally proceed through the four steps shown in Table 2. The first step is the circuit design and construction using conventional components, with consideration given to the component limitations in integrated circuits. Next comes duplication of the circuit using compatible components. Compatible components are devices made in exactly the same way that they will appear in the integrated circuit. For examples, Figures 16-19 show a diffused resistor, silicon dioxide capacitor, metal film inductor, and an integrated two-diode device respectively, each in a separate TO-18 can. They will each have the same component values and internal parasitics as when made in integrated circuit form. In the future, the more commonly used integrated devices and FEBs will find increased use as compatible components at this stage.

The next step is to bring the separate components together in the same package, resulting in a "separate chip circuit" if all elements are separate, or a hybrid integrated circuit when some integrated devices or FEBs are used. The trend for the future is more commonly into a hybrid integrated circuit directly at this design stage. Figure 23 shows the completed assembly. Some redesign may be necessary at this point to allow for the changed interconnection parasitics, and several forms of the hybrid integrated circuit may result as separate devices are consolidated. Figures 20-23 show an example of the hybrid stage of integrated circuit development, a 12 Mc I.F. amplifier, the schematic of which is shown in Figure 20. Figure 21 shows the two .003 and the .001 μf junction capacitors mounted on a 10-pin TO-5 header. Figure 22 shows the rest of the circuit mounted on a preferentially metallized ceramic piece that is mounted over the capacitors on the same header.

In many cases, i.e., for small sample sizes, if isolation parasitics in a FEB are too high, if initial cost is a prime factor, if a very short time cycle must be adhered to, or if the particular reliability considerations dictate, the hybrid integrated circuit version will be the one used in actual operation. In many cases, the single (or multiple) block integrated circuit will have advantages in cost, size, reliability, etc. that make its design and fabrication steps important and the future will see an increasing use of this component.

The Role of Surface Passivation

Surface passivation is used in the fabrication of many devices and FEBs, and is absolutely necessary to any practical integrated circuits approach. Surface passivation refers to the formation on the surface of a device a film that has three properties:

1. Results in desired electrical properties of the device.
2. Keeps the electrical properties stable under specified thermal, electrical, shock, and environmental conditions.
3. Isolates the device from its immediately surrounding ambient.

The above properties are necessary if the structure is to have interconnections by means of metal surface films over the passivating (isolating) layer, be measured (and perhaps preselected) before encapsulation, and pass required reliability tests. The present state of the art[8,9] is that the above requirements can be met on structures in glass to metal hermetic packages, and there is promise of extension to non-hermetic packages.

Advantages of Integrated Circuits

The large scale application of integrated circuits, as with any electronic element, depends upon one or more advantages it has over the previously used methods. The same factors cannot be applied to all integrated circuits, but rather, different integrated circuits will be used in different situations for different reasons.

Cost is an important consideration in most contemplated uses. A comparison between integrated and standard circuits should include the comonent, inventory, test, interconnection, and replacement costs. The number of a specific type made will have a bearing on its cost, although in view of the large usage of special integrated circuits, manufacturing versatility and process control will have a large bearing on the matter of cost. Although generally the opposite, in some ways integrated circuits have a yield advantage over conventional components. For example, since the end use conditions are known, unneeded parameters or parameter ranges needn't be obtained. Also, sometimes a structure can be specially designed to be used for more than one FEB by using, for example, a different surface film metallizing pattern to produce different circuit functions.

The size, weight, and shape advantages of well-designed integrated circuits are obvious today. Also, because of the close proximity of different parts of a circuit, thermal matching (i.e., keeping parts of a circuit at the same temperature so their electrical performance will be the same) can be a decided advantage. Speed can be improved in some circuits, since distances between circuit elements (and, hence, delays) are reduced. Wiring inductances and capacitances are also reduced, although intercomponent capacitances (due to isolation parasitics and close proximity) may be increased. Reductions in power dissipation are obtained mainly through novel circuit design, and apply to standard circuitry as well. However, integrated circuits can contribute to lower power and temperature through reduced interconnection parasitics, and better utilization of package dissipation capabilities.

Package Types

There are several package types being used at the present time for integrated circuits, and others are certain to follow. Figure 24 shows pancake and regular multi-lead TO-5 packages. Similar versions of the TO-18 are used. Figure 25 shows a 10-pin TO-5 package of standard height on which five layers of circuitry are possible. All except the bottom stage are mounted on metallized ceramic wafers, inter-wafer connections being by means of the header riser wires. Figure 26 illustrates the flat package used in the Texas Instruments Solid Circuits.

Reliability Considerations

Separate electronic devices have reached high levels of reliability, so high (i.e., better than .001%1,000 hour failure rate in some cases) that it is

impossible to quantitatively measure reliability under operating conditions with reasonable numbers of units or in reasonable lengths of time. It is necessary, then, to measure component reliability under a variety of high stress conditions, and establish acceleration factors to extropolate the reliability numbers obtained to those to be expected under operating conditions. It is important to note that the devices actually used in equipment are not those on which reliability data are obtained. (most manufacturers employ some sort of ageing cycle on all units. This is considered part of the production process and not for reliability determining purposes). Hence, what is guaranteed is the control and stability of all the processes that went into making a particular component. Since many failure modes are directly attributable to surface changes, and surfaces are not at all well understood at the present time, a small process change can lead to rather catastrophic reliability changes. This underscores the need for reliable surface passivation in integrated circuits. Also, proven process reliability and control must be used to make reliable integrated circuits in the same way that they are used to make reliable devices. Since large enough numbers of each type of integrated circuit may not be made for reliability determination, especially since the cost can become prohibitively large, and since the component devices in an integrated circuit cannot always be put under high electrical stress conditions because of their interconnections, it is necessary to determine process reliability on each type of compatible component that goes to make up an integrated circuit separately, and then use exactly the same processes to make the integrated circuit. Reliability of the isolation and/or interconnection scheme must be determined separately in a similar way. Also, the package reliability must be ascertained.

Conclusions

It is possible to categorize the device and pertinent circuit aspects of micro-systems electronic in a manner consistent with our defined integrated circuit terms. Some integrated circuits are practically useful today, and the future will see their increasing use as further developments are made. Single block, multiple block, and hybrid integrated circuits will all be used, depending upon the particular application. Also, some applications will at least start as mixed systems, in which both standard component and integrated circuits are used to take advantage of particular integrated circuits as they are developed.

Acknowledgment

The work of many people and several companies contributed to this paper. Examples, except where specific reference is made, are by the Motorola Semiconductor Division and Solid State Systems Division. Much of the work covered in this paper was done under support of the U.S. Air Force, Aeronautical Systems Division, Wright-Patterson Air Force Base, Dayton, Ohio, under contract AF33(616)8276.

Figures

1. Some forms of resistors (basic devices)
2. Two types of capacitors (basic devices)
3. Basic and complex devices
4. Distributed device
5. Integrated devices
6. Part of a wafer of integrated resistor devices--20 resistors in series per device, 20 mil centers.
7. Functional electronic blocks
8. Silicon flip-flop, R-C coupled type
9. A wafer of completed 5-stage Darlington amplifiers.
10. R-C pattern for a bistable circuit. 8 nichrome resistors, 4 silicon monoxide capacitors.
11. A Series 51 Solid Circuit Semiconductor Network (Texas Instruments)
12. Half Shift Register (Fairchild)
13. NAND gate
14. R-C coupled amplifier
15. Dual AND gate. 12 nichrome resistors, 2 silicon monoxide capacitors, 8 diodes, 2 transistors.
16. Compatible component - diffused resistor.
17. Compatible component - silicon dioxide capacitor
18. Compatible component - thin film inductor
19. Compatible component - two diode integrated device
20. Circuit diagram - 12 Mc I.F. amplifier
21. Bottom level - part of 12 Mc I.F. amplifier
22. Top level - part of 12 Mc I.F. amplifier
23. Completed (two level) 12 Mc I.F. amplifier
24. Multiple pin TO-5 packages--pancake and standard

25. Five level circuitry in 10-pin TO-5 package-standard
26. Flat package (approximately 250 x 125 x 35 mils). Texas Instruments.

Table 1

SOME DEVICE AND CIRCUIT CATEGORIES

Devices

Basic
Complex
Distributed
Integrated

Circuits

Functional electronic blocks
Integrated circuits
  single block
  multiple block
  hybrid
Combination circuits

Table 2

STEPS IN THE DESIGN OF INTEGRATED CIRCUITS

1. Circuit design using conventional components.
2. Circuit design using compatible components.
3. Hybrid integrated circuit versions.
4. Single or multiple block integrated circuit.

References

1. I. A. Lesk, N. Holonyak, Jr., R. W. Aldrich, J. W. Brouillette, and S. K. Ghandi, Proceedings of the IRE, Vol. 48, p. 1833 (1960).

2. Term adopted by the AIEE Solid State Devices Committee, Sub-committee on Micro-Systems Electronics, used broadly to cover work applicable towards achieving small electronic equipment.

3. For extensive, but still partial, lists of basic complex and integrated devices, see reference No. 1.

4. An ingenious aspect of the TI Series 51 Solid Circuit Semiconductor Network appears to be that the structure in the silicon is the same for all units, different numbers of the series being obtained by means of different interconnection schemes.

5. MOLECULAR SCIENCE AND MOLECULAR ENGINEERING, Arthur R. von Hippel, published jointly by the Technology Press of MIT and John Wiley & Sons, Inc., New York, 1959.

6. MICROMINIATURIZATION AND MOLECULAR ELECTRONICS, Richard D. Alberts, 5th AGARD Avionics Panel "Micro-miniaturization," Oslo, Norway, July, 1961.

7. Molecular Electronics--New Concepts for Control, H. V. Noble, Control Engineering, Vol. 8, No. 9 p. 183, September, 1961.

8. Semiconductor Surfaces and Films; The Silicon-Silicon Dioxide System, by M. M. Atalla, Conference on the Properties of Elemental and Compound Semiconductors published by Interscience, 1959, pp. 163-182, Boston.

9. Surface Passivation Techniques of Micro-Electronics, S. S. Flaschen, R. J. Gnaedinger, Jr., Transactions, 1961, Semiconductor Symposium AIMME, Los Angeles, California, In Press.

Accelerated Thermal Oxidation of Silicon, D. A. Kallander, S. S. Flaschen, R. J. Gnaedinger, Jr., Camille M. Lutfy, Journal of Electro-Chemical Society, May, 1961, Spring Meeting, Extended Abstract.

Accelerated Thermal Oxidation of Germanium, S. S. Flaschen, D. A. Kallander, Camille M. Lutfy, Journal of Electro-Chemical Society, May, 1961, Spring Meeting, Extended Abstract.

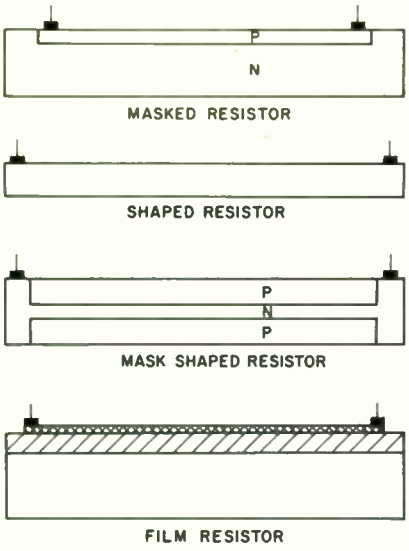


Fig. 1.

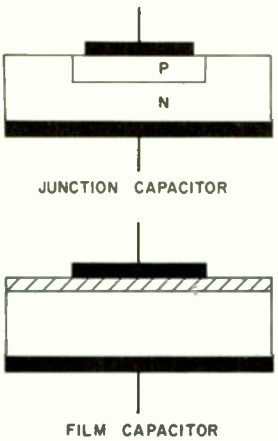


Fig. 2.

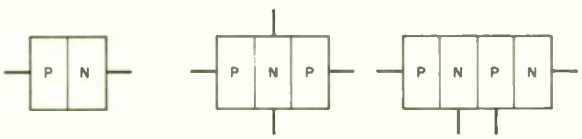


Fig. 3.

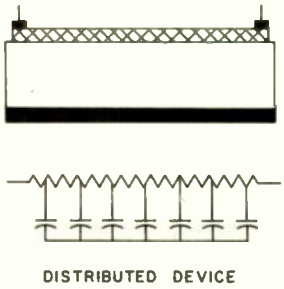


Fig. 4.

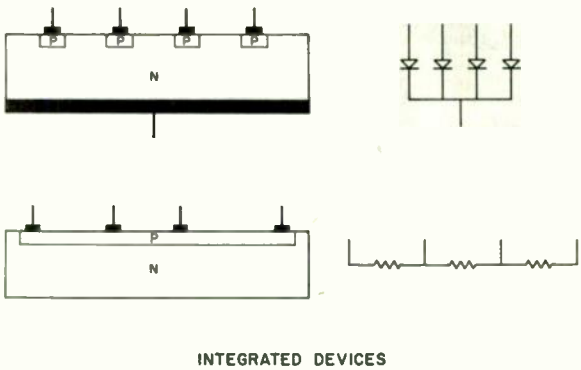


Fig. 5.

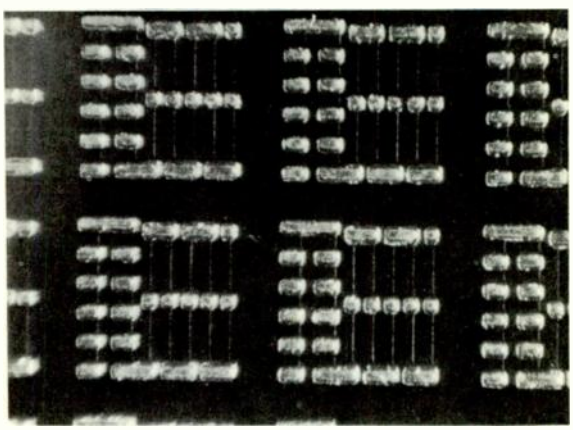
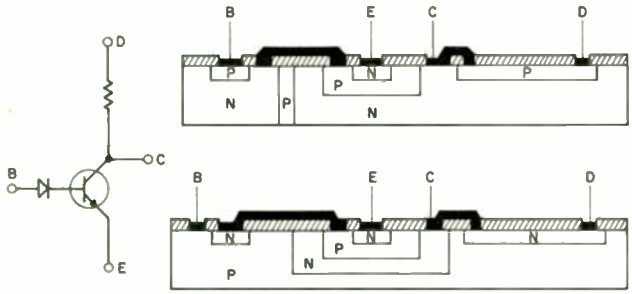


Fig. 6.



FUNCTIONAL ELECTRONIC BLOCKS

Fig. 7.

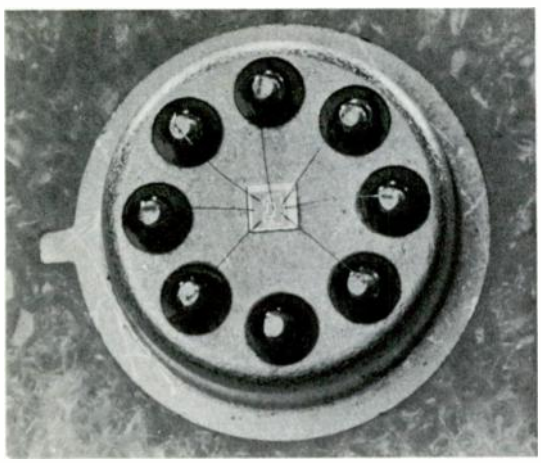


Fig. 8.

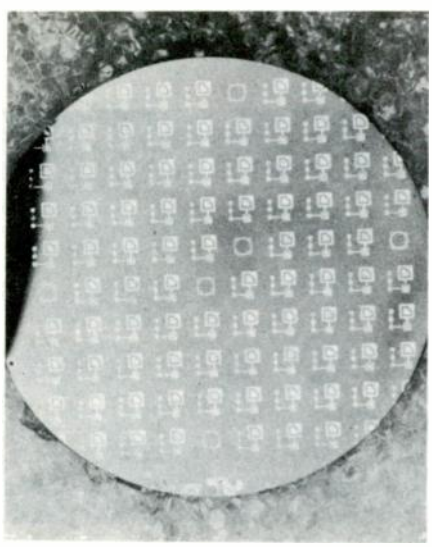


Fig. 9.

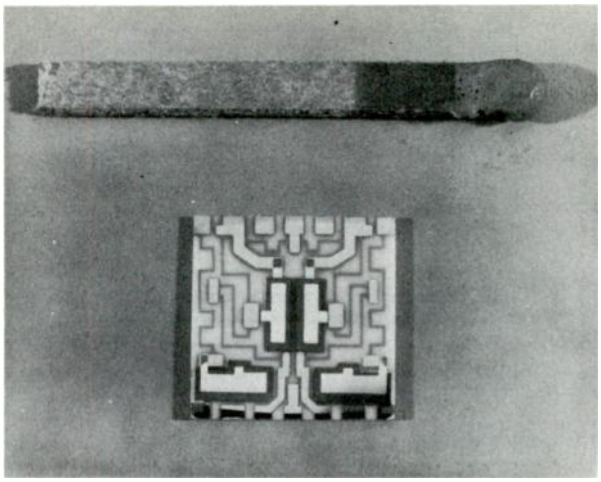


Fig. 10.

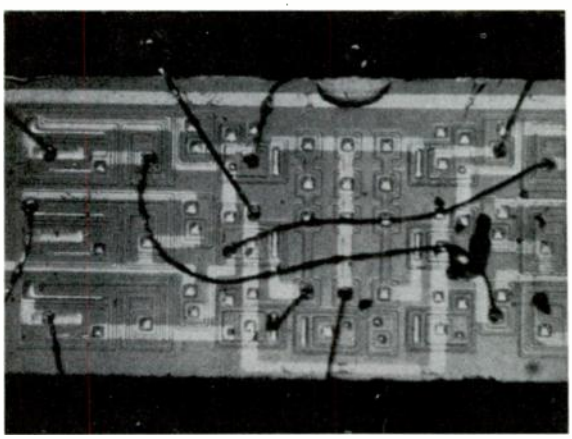


Fig. 11.

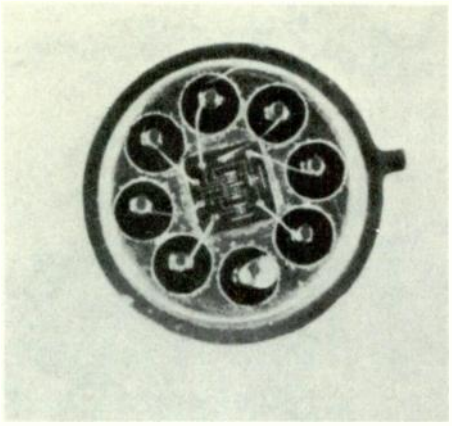


Fig. 12.

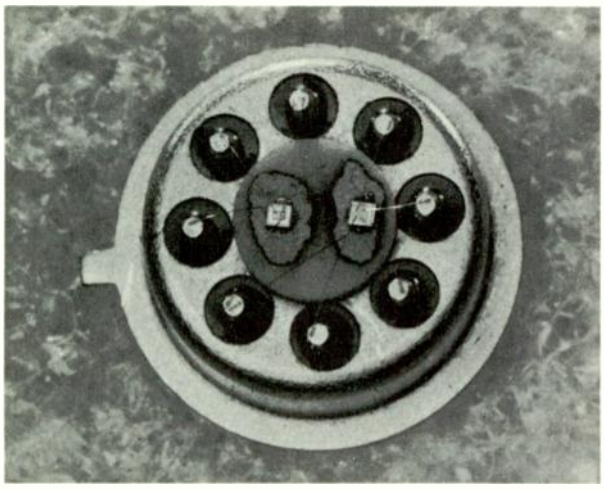


Fig. 13.

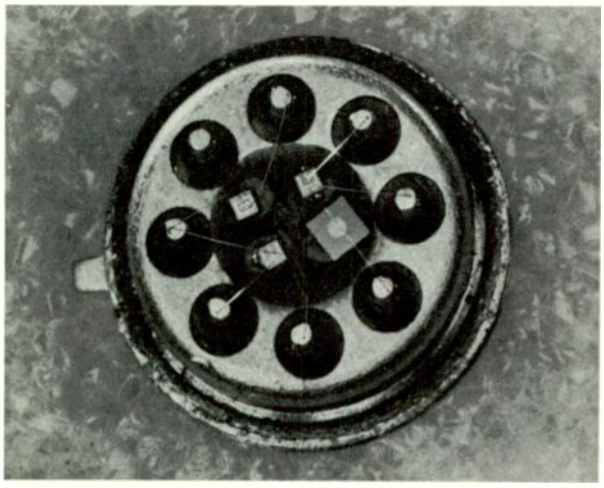


Fig. 14.

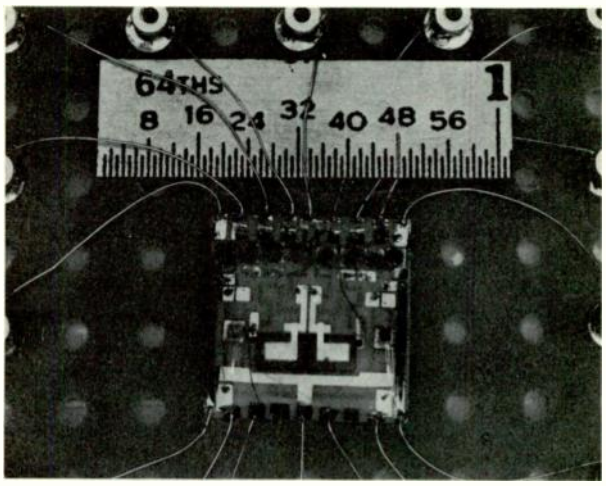


Fig. 15.

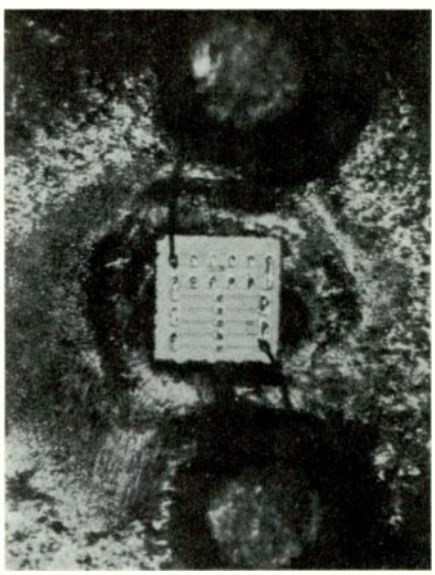


Fig. 16.

Fig. 17.

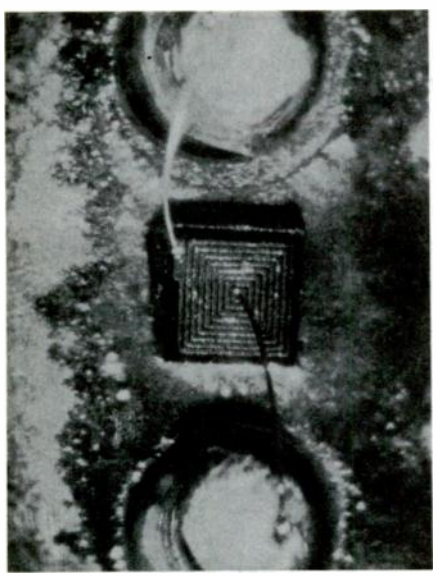


Fig. 18.

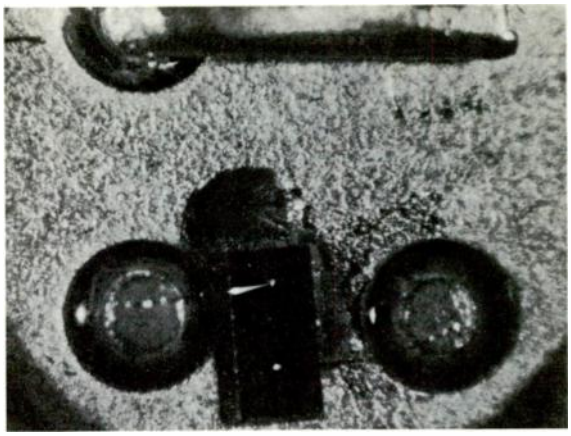


Fig. 19.

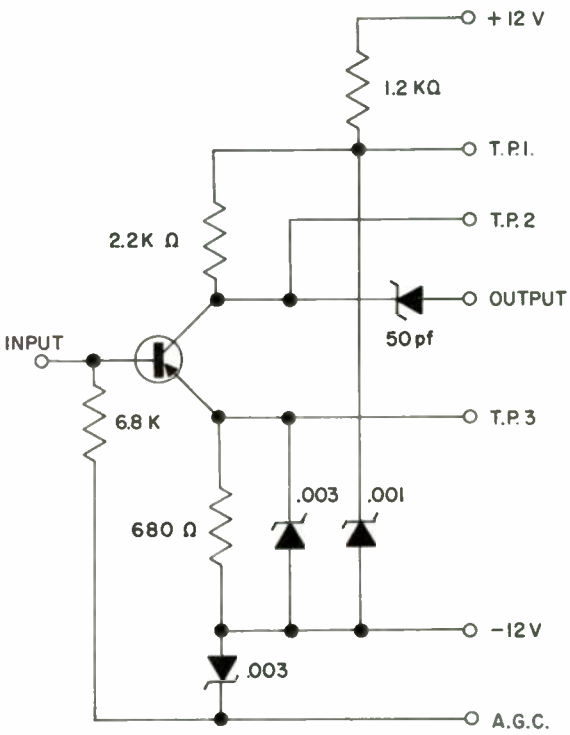


Fig. 20.

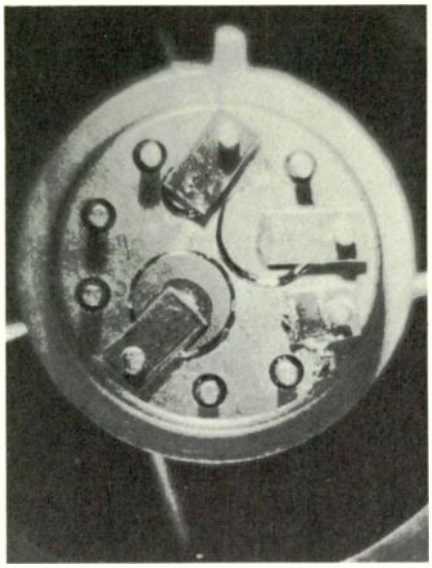


Fig. 21.

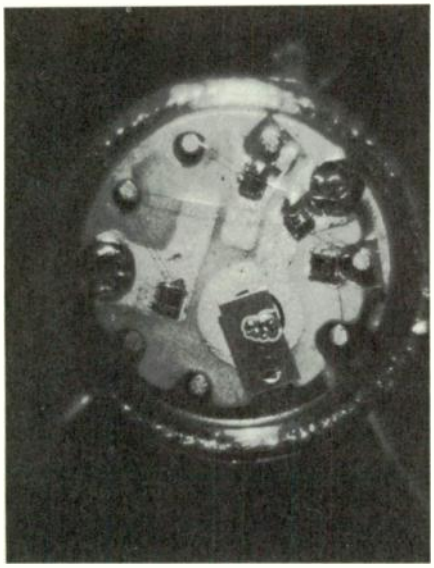


Fig. 22.

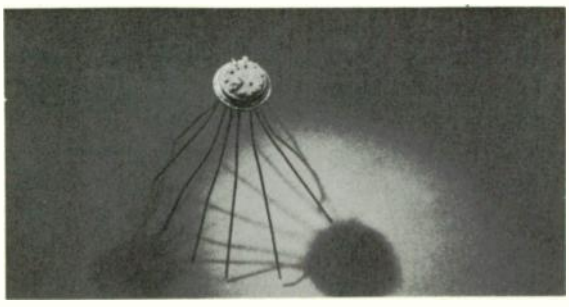


Fig. 23.

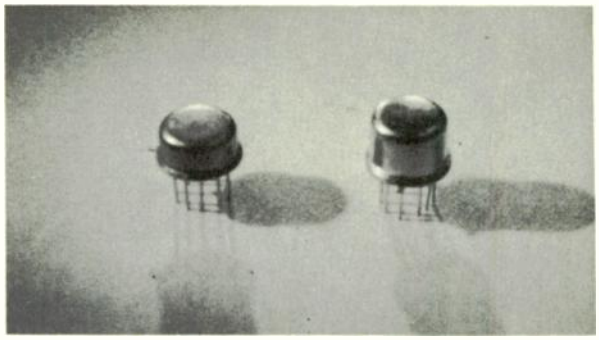


Fig. 24.

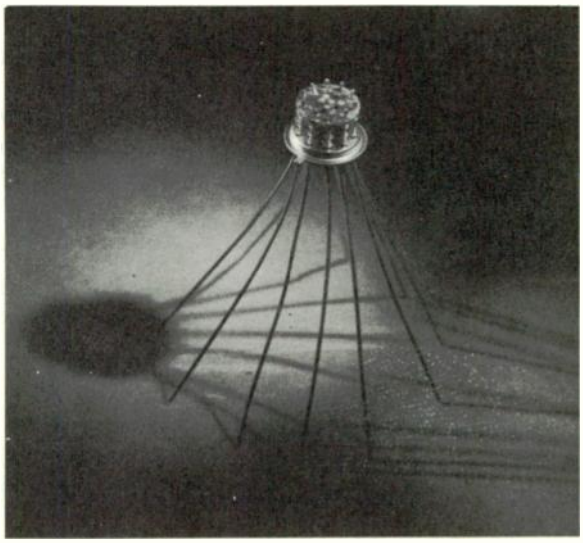


Fig. 25.

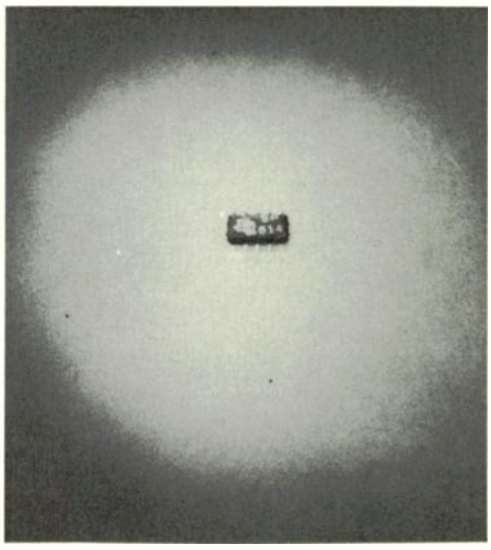


Fig. 26.

OPTICAL MASERS *

G. Birnbaum
Hughes Aircraft Company
Malibu, Calif.

SPEED IN SEMICONDUCTOR DEVICES

J. M. Early
Bell Telephone Laboratories
Murray Hill, N. J.

*The test of this paper was not available at the time of publication.

A WIDEBAND MICROWAVE DEFLECTION-AMPLIFIER TUBE

H. J. Wolkstein and R. W. McMurrough, Electron Tube Div., RCA, Harrison, N. J.

Abstract

This paper describes a novel wide-band microwave amplifier tube which bridges the gap between conventional low-frequency amplifier tubes and broad-band microwave traveling-wave tubes. This small, light-weight, rugged, preamplifier tube will be useful in the frequency range from 100 to 1200 Mc.

This experimental device uses a novel wide-band, synchronous, slow-wave input-deflection circuit in conjunction with a similar wide-band output circuit. It can be operated in two ways: the output can be taken from the slow-wave circuit as in a conventional transverse-field traveling-wave tube, or the output can be taken from a high-impedance output target on which the modulated collector current impinges.

THE MULTIPLE-BEAM KLYSTRON

M. R. Boyd, R. A. Dehn, J. S. Hickey, T. G. Mihran
General Electric Company, Schenectady, N. Y.

Summary

The multiple-beam klystron (MBK) is a device for extending klystron power generation capability at a given frequency by a factor of ten or more. The MBK utilizes a multiplicity of electron beams in conjunction with multiwavelength waveguide circuits. These circuits are periodically loaded and operate in the /2 mode. The efficiency, bandwidth, gain, and stability of the MBK are equal to or better than that of the single-beam prototype klystron. Furthermore, the MBK permits generation of a given power level at an unusually low voltage, thus minimizing insulation, x-radiation, and modulator problems.

Two MBK amplifiers have been built. The first at 750 Mc demonstrated the principle of the device at 10 kilowatts output from ten beams. Initial tests of a ten-beam X-band amplifier have resulted in 35 kw cw output at 8350 Mc. A single main magnetic field is used to focus all beams simultaneously. Bandwidth and gain are identical with those of the single beam prototype operation. Second harmonic output of the X-band MBK was measured as 56 db below fundamental. No disruption in operation was experienced when several beams were shut off, indicating that end of useful life of an MBK will not be reached until more than one or two beams have ceased functioning.

Introduction

The continually increasing sophistication of radar and communication systems requires ever-increasing transmitter power output. At this time the most energetic stable power source in the microwave region is the klystron power amplifier. The klystron can be characterized as being a high gain amplifier of moderate bandwidth and efficiency and having a more robust mechanical structure than other types of generators. In the search for devices with higher and higher power capability, the klystron interaction represents a logical starting point. This paper discusses a unique configuration of electron beams and circuits whereby the power generation capability of amplifiers at a given microwave frequency can be increased by ten- to one-hundred fold while retaining other desirable operating characteristics such as gain, efficiency, and bandwidth.

Conventional klystrons, either pencil or hollow-beam, have a limiting power capability which depends on frequency, cathode emission density, and the thermal dissipation required on various portions of the rf circuits. This limitation results, in part, from the necessity for limiting maximum resonator dimensions to values less than one free-space wavelength and also by the requirement for relatively small drift tube diameters to achieve good coupling between electrons and resonator fields. For pencil beams, the result is a drift tube diameter of the order

$$d = \frac{\lambda_o}{\pi} \cdot \frac{v}{c}$$

where

λ_o = free space wavelength
v = electron axial velocity
c = velocity of light

Use of a hollow beam increases this limiting diameter by the factor c/v. However, at high power levels, where large-beam voltages must be used, this factor will, in general, be less than two. Obviously, no order-of-magnitude increase in power is available by this expedient.

One solution to the problem of very high power generation is to employ circuit structures with extended dimensions normal to the electron velocity. In such a case at least one dimension of the electron beam must be restricted so that good coupling to circuit fields are maintained. This leads to use of either a sheet electron beam, or a multiplicity of pencil beams. The circuits used with these beams may be either resonant or non-resonant. Devices using a multiplicity of beams in conjunction with resonant circuits have been proposed in recent years,[1] but no devices representing a reduction to practice have been reported. The adjacent mode interference problem was probably responsible to a great degree for holding back development of this idea. In the work to be described this difficulty has been

greatly minimized by utilizing a mode with the maximum frequency separation between it and adjacent modes.

Multiple- or sheet-beam devices utilizing non-resonant circuits have also been proposed for obtaining high power output.[2,3] However, while these studies investigated the principles of traveling-wave klystron interaction, neither resulted in a practical demonstration of a significant amount of power output.

At the present time, the most successful practical approach to the problem of increased power output is to combine the outputs from an even number of tubes by the use of hybrid couplers. The MBK constitutes an alternate and significantly different approach to "harnessing" beams together which is much less complex and bulky than the hybrid scheme. Furthermore, it is much less sensitive to disruption of operation in the event of beam failure.

In the following sections, the basic principles governing the choice of the operating mode and the design of MBK circuits are outlined. This is followed by a description of experiments in which an order-of-magnitude increase in power output was obtained by the use of MBK principles.

Circuits for Multiple-Beam Resonators

Essential components of any circuit suitable for a multiple-beam klystron are the interaction gaps. In general, these gaps are small compared with wavelength and can be treated as capacitors. The geometrical configuration of the gaps and tunnels is largely determined once the power level per gap and the frequency are specified. Thus, in a specific MBK circuit design, these capacitors are fixed, and the circuit problem reduces to determining the optimum method for coupling power-producing interactions from an array of these gaps. Factors to be considered are the location of the gaps relative to the field maxima of a given mode in the resonator and the frequency separation of the various possible resonant modes.

A multi-mode resonator may be derived from a conventional klystron cavity by tightly coupling a number of such cavities in an extended array. The resulting structure may be considered as a length of waveguide periodically loaded with capacitors as shown in Figure 1. The modes of this circuit can be obtained from a knowledge of the propagation characteristic of the loaded waveguide as a transmission line. In Figure 2, a theoretical frequency-phase shift (ω-β) curve of a capacitor-loaded guide is shown. Also given are frequency points corresponding to the resonances of an eight-section shorted length of such a loaded line. In addition, Figure 2 shows the spatial voltage patterns of the respective modes. The experimental data of Figure 2 were obtained using a circuit designed for the four-beam MBK tests to be described later. The theoretical curve was plotted using an approach outlined reference 4.

From the spatial mode plot of Figure 2, it is evident that only the mode at $\beta D = \pi$ has equal maximum voltages at all gap locations. This is a desirable characteristic if electron beams passing through all gaps are to be used most effectively. Reference to the ω-β curve, however, shows that in the vicinity of π -mode, the modes tend to be crowded together in frequency and therefore adjacent mode interference would be a serious problem. Similar crowding exists for zero-mode (not shown), i. e., the frequency at which the waveguides just cut on. From the ω-β diagram of Figure 2, it is evident that the mode at $\beta D = \pi/2$ represents the optimum mode configuration for two reasons. First, maximum frequency separation of adjacent modes exists here. Second, all voltage maxima are of equal amplitude. In this particular mode, alternate capacitors are located at voltage nulls, hence electron beams are not used at these positions, that is, alternate gaps would be dummy capacitors having the same capacitance as tube gaps.

In the $\beta D = \pi/2$ mode each electron beam is situated in precisely the same electrical environment as a beam in the single-mode resonant cavity of a conventional klystron. This situation allows each of the beams of an MBK to deliver energy to the resonator fields fully as efficiently as in the single beam case. Furthermore, in the MBK the individual interactions or outputs of all beams are phase-locked because the fields are very tightly coupled throughout the resonator. This advantage of strapping does not exist in a system where outputs of isolated klystrons are combined externally.

The use of a multipoint drive and a particular mode in a multimode resonator has an important advantage with regard to klystron harmonic output. Current bunches arriving at a klystron output gap are extremely rich in harmonic content. In a single resonant cavity,

for harmonic power to be developed it is only necessary to have impedance present at the harmonic frequencies. In an MBK, however, in addition to this impedance requirement, there also is a phase requirement that must be satisfied to obtain harmonic output. In this case harmonic generation requires not only that the harmonic frequency occur in a higher passband at a resonance of the loaded circuit, but also that the phase shift per section be harmonically related to that of the fundamental input signal. Minimum harmonic output will be achieved by designing so that harmonic frequencies occur in the stop bands of the circuit. This minimum output should be significantly less than that of a single beam klystron of the same total fundamental output.

Tuning and coupling to multimode resonators are somewhat more complicated than in single mode resonators. For efficient operation klystron output circuits are usually loaded externally to an impedance level that permits a voltage swing approximately equal to the beam voltage to be developed across the interaction gap. To load the $\pi/2$-mode of an n-section MBK resonator to the same Q as the prototype klystron requires that the load be coupled to the resonator n times as tightly as in the single cavity case. In practice it was found that iris coupling such as that shown in Figure 1 could be used to obtain any degree of loading. With iris coupling it was found, incidentally, that as the loading is increased it is necessary to move the position of the iris in the direction of the gaps to avoid distortion of the fields at the gap locations. For the limited number of beams used in our low frequency test setups, loop coupling or probe coupling was used. For instance, in the ten-beam experimental model MBK a probe coupler section was used in the center of the resonator. This was done to permit symmetrical trim-tuning by moving two shorts in the end resonators simultaneously.

Mechanical tuning of a multimode resonator can also be achieved by varying other cavity dimensions. End-wall tuning is restricted to narrow frequency ranges because the mode pattern tends to become distorted. Side-wall tuning of each cavity section is required for wide tuning ranges or large numbers of beams.

Demonstration of MBK Principle

Description of Equipment

For flexibility and convenience, the basic principles of the MBK were first tested by using conventional commercially-available external circuit klystrons inserted into multi-wavelength waveguides. This setup has proven to be very useful in that modifications can be made in the rf circuitry without disturbing the dynamics of the individual electron beams.

The tube chosen for this purpose was the 3K3000LQ klystron. This is a three-gap klystron, and operates in the frequency range from 610 to 985 mc. Although these tubes are rated at 2 kw rf output, they were operated conservatively at the reduced beam voltage of 7 kv and a power output of 1.0 kw.

Our goal was to show that ten such tubes properly interconnected could operate as an MBK and deliver 10 kw of rf power to a single load, representing an order of magnitude increase in power capability. To do this, we first constructed a four-beam MBK, and when this was found to operate satisfactorily, a ten-beam model was built. A cross-sectional view of the four-beam model is shown in Figure 3(b). The prototype cross-section is shown in Figure 3(a). The tubes are detachable from the circuitry just outside the ceramic insulators, shown in dotted cross-section. Loading capacitors are added midway between tubes in order to obtain $\pi/2$-mode operation, as described in the previous section. Tuning is accomplished by movable shorts located at both ends of the individual waveguides.

The ten-beam model was arranged and tuned in much the same manner as the four-beam model, except that the circuits were horseshoe-shaped rather than linear. A photograph of an assembled ten-beam MBK is shown in Figure 4. The high-voltage seals of the tubes can be seen beneath the table. The electron beams travel vertically upward through the three waveguide circuits and are collected by water-cooled collectors. These collectors are insulated from the tube bodies, making it possible to monitor individual body current on each tube. The output coupling cavity utilizes probe coupling and is located directly below the double-slug tuner shown in the upper part of the photograph. Input coupling is obtained by a loop in the corresponding cavity of the input deck. The corresponding cavity of the center deck is not used for coupling but is loaded by a capacitor equal to the average tube gap capacity to preserve $\pi/2$-mode. Tuning is accomplished by means of double tuning pistons with gear mechanisms placed in the cavities diametrically opposite the coupling cavities. Thus the 12-unit circuit is composed of ten units enclosing tubes, one coupling unit, and one tuning mechanism unit.

The four beams shown in Figure 5 and the ten beams in Figure 4 were focussed by the single solenoid and iron shell shown in Figure 5. In addition, the particular tubes utilized in this demonstration normally require focus coils near the cathode end as shown in Figure 3. These coils would not be used for electron guns of the types currently in widest use.

Test Results

Power Output, Efficiency and Gain. The performance of the ten-beam MBK is most readily evaluated by comparing it with single-tube performance. This comparison is made in Figure 6 where rf power output is plotted as a function of rf power input. The upper curve represents ten-beam MBK performance; the lower curve was obtained by taking data on a single-beam klystron and subsequently scaling the efficiency at a drive level of six watts to the average of the individual efficiencies of the ten tubes used in the ten-beam MBK as determined by data supplied by the manufacturer. In taking these curves, the middle or penultimate circuits were optimized at each drive level. The shapes of the two curves are very similar except that the ten-beam MBK gives ten times more power output at a drive level ten times that of the corresponding prototype point. Note that this is done not by raising beam voltage but by increasing current, or, in effect, perveance by a factor of ten.

Bandwidth. The bandwidth of the ten-beam MBK is compared with single-tube bandwidth in Figure 7. Two measurements of bandwidth were made. In one case all circuits were optimized at midband and a curve of power output versus frequency was taken with no retuning. These curves are shown as the dashed lines marked "overall". The curves shown as solid lines were taken by keeping drive level constant and retuning both input and middle circuits. The output circuit remained fixed in this test, hence this curve is indicative of output circuit bandwidth. If additional stagger-tuned bunching cavities were available, presumably the dashed curves could be made to approach the solid curves as in conventional broadband klystrons.

Comparing MBK and a single-tube bandwidths, we see the output circuit of the MBK performs as well as the single-klystron output circuit, both bandwidths being about 3.5 mc out of 720 mc. This represents a bandwidth of 0.5 percent, which is proper for a beam whose dc beam resistance is about 20,000 ohms working into an output circuit with an R/Q of about 100. Thus, the MBK output circuit operates over the same bandwidth as the prototype klystron output circuit. The reason for the apparent improvement in the overall bandwidth of the MBK, shown by the dashed curves, is not clear. It probably can be attributed to different relative placement of the resonant frequencies of the input, middle, and output circuits used in the two cases.

Effect of Beam Failures. The MBK is relatively insensitive to beam failures. If a beam is lost, the power output drops by an amount roughly equal to the power the lost beam was contributing. There is some reaction on the performance of the other tubes, but it is not major. For instance, dropping out one beam reduces dc input power by 10 percent, hence output power would be expected to drop off by a similar factor. When this test was made, output power dropped by 14 percent. The additional drop-off over that expected is due to a combination of factors, such as reduced effective R/Q, lower circuit efficiency, and improper output impedance. This last factor can be eliminated by reoptimizing the output impedance after a tube is dropped out. A test was made in which three beams of the ten-beam MBK were biased nearly off, such that there was a 28 percent reduction of dc input power. Output power dropped 40 percent under these circumstances. By reoptimizing the output load impedance, this drop-off was reduced to 35 percent, which compares favorably with the expected drop-off of something greater than 28 percent.

Ten-Beam X-Band MBK

The successful demonstration of the MBK principle has led to the design and construction of a ten-beam amplifier for CW operation at 8.4 Gc. Here all beams and resonators are contained in one vacuum enclosure. A cutaway section of this amplifier is shown in Figure 8. The ten individual gun assemblies are precisely located in a water-cooled reference bar in the enclosure at the top. The beams from these ten guns are focussed through corresponding tapered apertures in a magnetic pole piece after which they traverse the four multiple-gap resonators shown in the center of the photograph. A single magnetic coil and yoke assembly keeps the beams focussed in this region. After output power extraction in the lower resonator, the spent beams enter the tapered collector shown at the bottom. This collector has a shape similar to a gable roof.

The rf drive is coupled from the standard RG51/U waveguide at the left through a vacuum seal and iris aperture to the input resonator. Output is coupled also by an iris at the end of

the lower resonator to a tapered waveguide and through a circular window section to WC 150 circular waveguide operating in TE_{11} mode.

The individual electron beams have a perveance of 0.7×10^{-6} which means an overall perveance of 7×10^{-6}. The tube is designed for 150 KW CW input. Preliminary test results on an amplifier of this design are shown in Fig. 9. The tube was tuned for high efficiency operation with a corresponding gain ranging from 45 to 50 db. The top point of 35 KW output represents the point at which failure of the waveguide load caused the output seal to fracture. In pulse tests, 51 KW output was reached at 41% efficiency.

The second harmonic level has been measured at 56 db below fundamental. This is to be compared with a 41 db second harmonic measured with the same equipment on a single beam klystron at 1 KW output, 8.4 Gc. Similar measurements on third harmonic indicated the MBK to have approximately ten db less harmonic level than the equivalent single beam tube. Absolute values of harmonic level remain to be determined and will require proper summation of power components in the twenty-six transmission modes possible in WC150 guide at 25 Gc.

Conclusions

A multiple-beam klystron has been demonstrated which promises to extend the power capability of microwave tubes embodying klystron-type interactions by at least one order of magnitude. As the name implies, this device has the gain, efficiency, bandwidth, and stability characteristic of conventional klystrons. However, by using a multiplicity of electron beams and a carefully selected mode of resonance, the power output of the MBK can be made to exceed that of the prototype klystron by a factor equal to the number of beams used without sacrifice of any klystron advantages. The concept of reduced harmonic generation because of the phasing requirement in the MBK appears to be valid from preliminary tests on the X-band amplifier.

The MBK is fundamentally a low-voltage, high-current approach to superpower. Therefore, problems such as high-voltage power supplies, voltage breakdown, and x-radiation protection are simplified. In an MBK each beam operates under the same conditions as the beam of the prototype klystron. Consequently, the electrical and thermal stresses impose no new limits on available materials. The basic unit of the MBK is a conventional klystron beam; therefore, an MBK design requires no advance over the present state of the art for the design of the single-beam prototype unit. Furthermore, any future advances in klystron technology may be incorporated into MBK designs.

The ultimate limits of power capability of an MBK are not precisely known. The first limit to be encountered will probably be due to adjacent mode interference. This limit will depend upon circuit design, individual beam impedance, and the ability to control mode excitation in the input circuit by selective loading or strapping. The best present estimate for typical existing klystrons is that the limitation will occur somewhere between 40 and 100 beams. A more fundamental limitation will be encountered at power levels where the circuit losses are comparable to the unit power being developed. At present the latter is not a practical limit.

Since the failure of individual beams in a multiple-beam klystron does not result in a serious reduction of output power, the useful life of an MBK extends beyond the point where one, or even several, beams become inoperative. The MBK concept thus permits the effective coupling of many electron beams in a stable, reliable manner. This is accomplished without the cumbersome equipment and redundancy of operating controls characteristic of systems whereby the output powers from several amplifiers are combined.

Acknowledgement

We acknowledge with appreciation the support of the Electron Tube Branch, U. S. Army Signal Engineering Laboratory under Contract DA-36-039-sc-78178.

References

1. R. Warnecke and P. Guenard, "Tubes a Modulation de Vitesse," Gauthier-Villars, pp. 721-730, 1955.

2. T. G. Mihran, "The Duplex Traveling Wave Klystron," Proc. IRE, Vol. 40, pp. 308-315, March 1952.

3. V. A. Heathcote, P. A. Lindsay, J. Barraclough and J. R. Newby, "The Periodically-Loaded Travelling-Wave Multiple Beam Klystron," Proc. Inst. of Elec. Engineers, Vol. 105, part B, pp. 952-965, May 1958.

4. R. C. Honey, E. M. T. Jones, "A Wide-band UHF Travelling-Wave Variable Reactance Amplifier" - Trans. IRE PGMTT, Vol. MTT-8, pp. 351-361, May 1960.

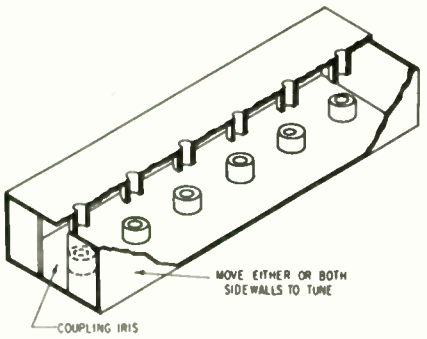


Fig. 1. Distributed-beam resonator.

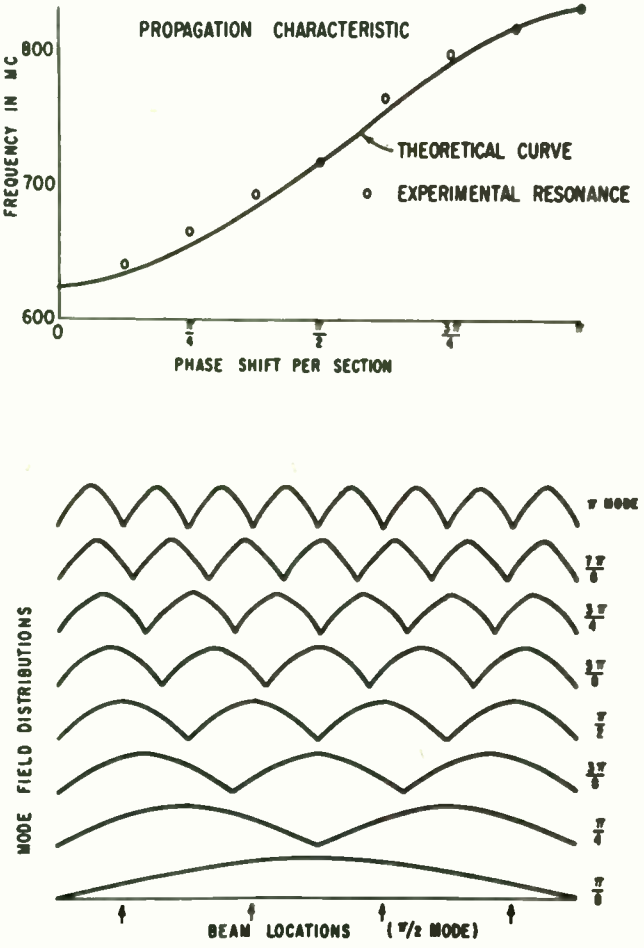


Fig. 2. Modes in periodic resonator.

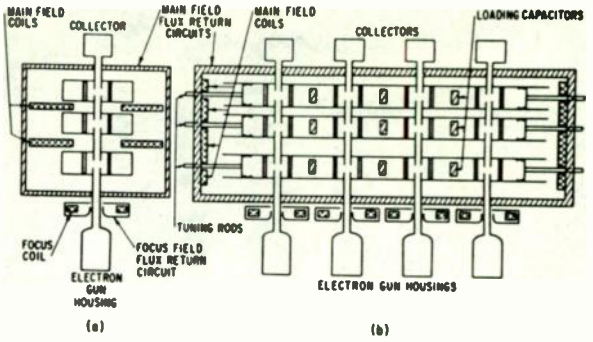


Fig. 3. Cross-sections of single and multiple-tube amplifiers.

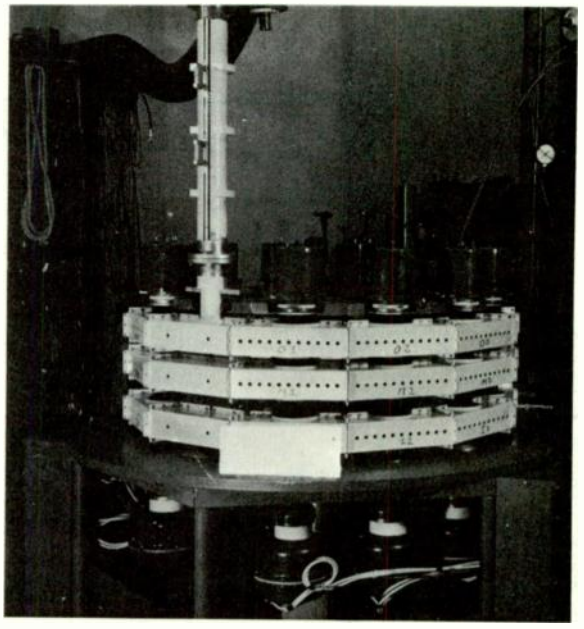


Fig. 4. Ten-beam MBK amplifier without magnetic beam.

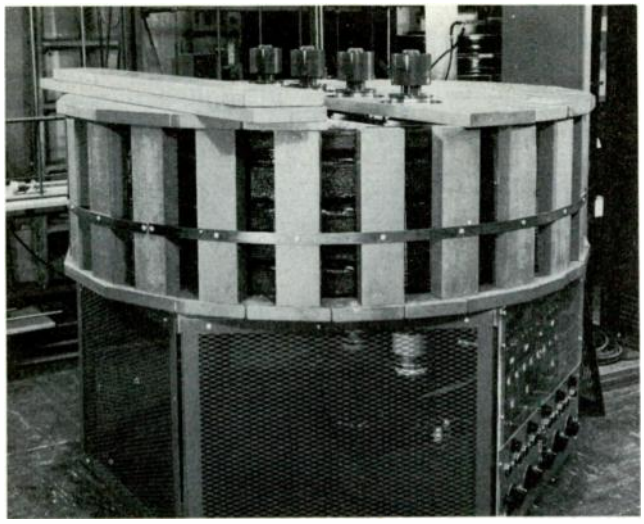


Fig. 5. Four-beam MBK amplifier.

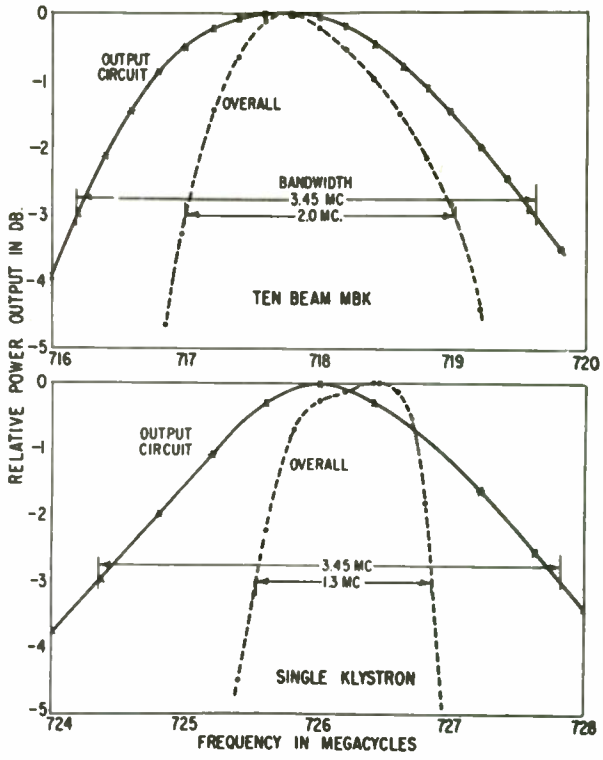


Fig. 7. MBK bandwidth.

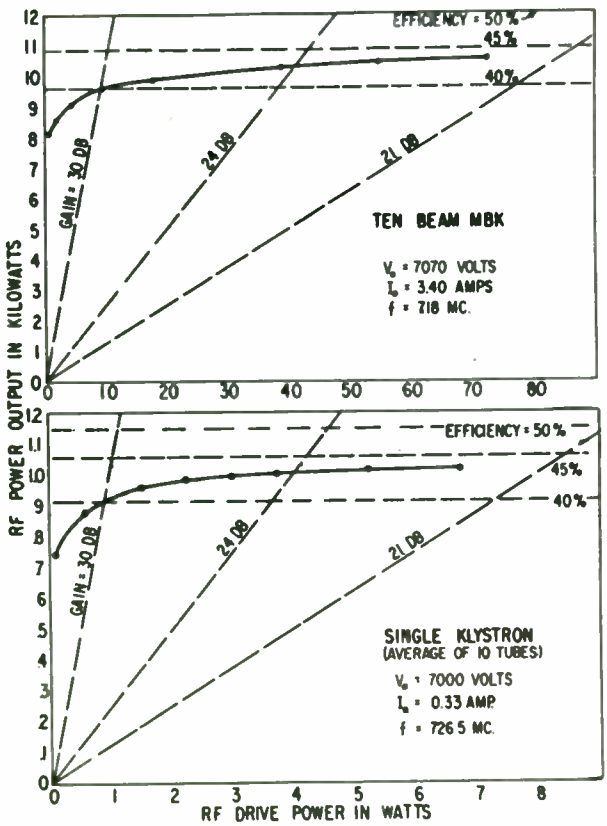


Fig. 6. MBK power output.

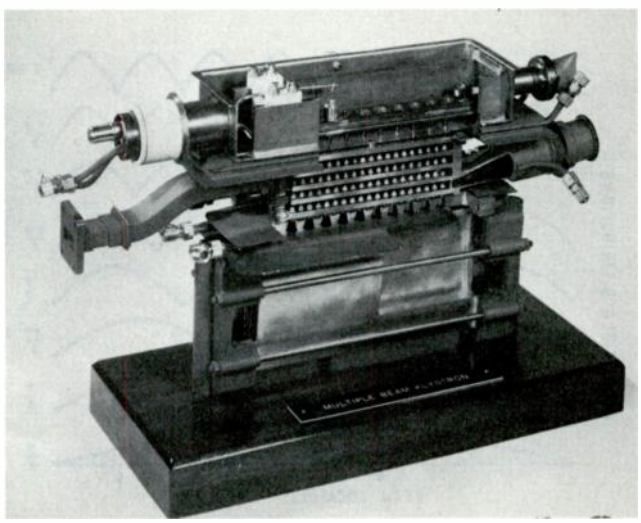


Fig. 8. X-band MBK cut-away model.

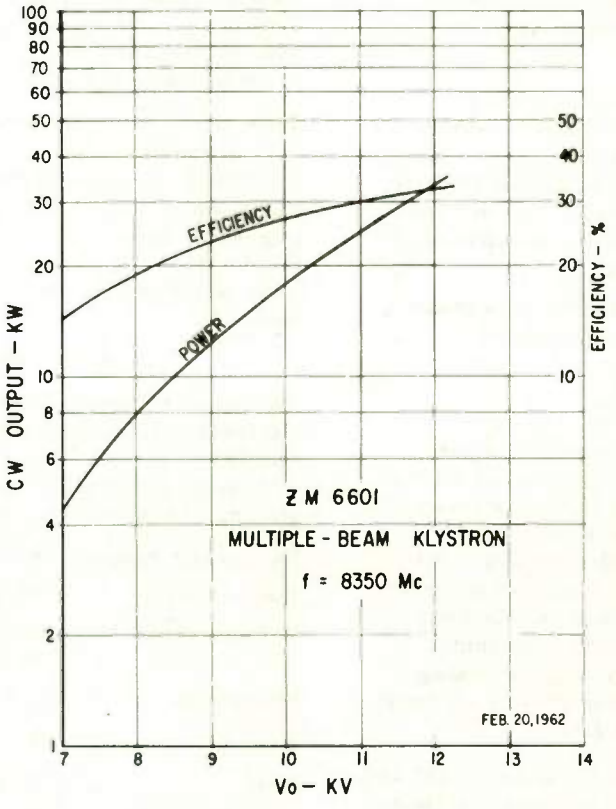


Fig. 9. Power output—X-band MBK.

A LIGHT WEIGHT, COMPACT BACKWARD-WAVE OSCILLATOR FOR X-BAND

C. Biechler and K. R. Evans

Microwave Associates, Inc.
Burlington, Massachusetts

SUMMARY

A backward-wave oscillator with the physical properties of a reflex klystron has been developed. As a voltage tuneable oscillator, it has bandwidth capability of 20 per cent with power output exceeding 10 milliwatts. Forty milliwatts over a 3 per cent bandwidth has been observed. The nominal operating frequency is 8.5 to 9.6 kmc.

These backward-wave oscillators utilize electrostatic-periodic focusing and a unique r.f. matching technique to achieve light weight, small size and the similarity to reflex klystron oscillators. The design is simplified and the reliability improved by the use of a convergent-flow solid beam electron gun.

This device is still in the experimental stage and minor problems yet exist. Most of the design goals have been met.

INTRODUCTION

There are many microwave systems in use today which employ a reflex klystron as a local oscillator. The klystron, in some instances, must be tuned in order to track a frequency-varying received signal. Utilization of a reflex klystron as a tuneable local oscillator has been only partially successful because of the problems encountered in mechanical gap tuning with simultaneous reflector voltage tracking. Further, the mechanical tuning imposes a limit on the speed of tuning and on life.

Yet the klystron, because of its size, weight and low voltage requirements is, in many instances, more attractive to the system designer than fast, electronically swept backward-wave oscillators as they exist today. Permanent-magnet focused BWOS are comparatively heavy and large and, in addition, are sensitive to magnetizable materials placed near them. Hence, the development of an inexpensive, light weight, electronically tuneable oscillator seemed very desirable.

Such a device form has been largely achieved. The application of periodic-electrostatic focusing to an "o" type backward-wave oscillator and the development of a radio frequency slow-wave circuit-to-waveguide match have allowed the realization of a electronically tuneable oscillator which looks like a klystron.

X-band was chosen as the frequency range for the development program for convenience and for application to general system requirements. A modest tuning range of 10 per cent was set as a goal in order to concentrate on the inherent electron-optical problems.

Although the development of this backward-wave oscillator is by no means complete, enough progress has been made to date to predict the achievement of the original design goals.

This paper is not intended as a theoretical analysis of the operation of a backward-wave oscillator nor of periodic-electrostatic focusing since these topics have been the subject of numerous papers.[1] Instead, it is our purpose to report on the details of the development of a backward-wave oscillator to meet the objectives set forth above.

DESIGN OBJECTIVES

As explained previously, the tube was to have characteristics, both physical and electrical, as similar to a reflex klystron as possible. The following specifications were outlined at the beginning of the development program.

Frequency range	8.5 - 9.6 kmc
Power output	20 mw (minimum)
Voltage range	all applied potentials to be below 500 v
Efficiency	0.5 per cent (minimum)
Weight	4 - 5 oz. (maximum)
Size	3" long (maximum)
Configuration	Waveguide output perpendicular to tube axis
Cost	Competitive with reflex klystron with mechanical tuning.

These objectives determined the design approach.

DESIGN

This design, as with many, is a compromise between complexity and hence high cost, and electrical specifications. Since weight, size and voltage range are of primary importance, efficiency and tuning range specifications are more modest than for a typical permanent-

magnet backward-wave oscillator.

It was important to choose a lightweight focusing system in order to eliminate the heavy and bulky permanent magnets usually used to focus backward-wave oscillators. Periodic-electrostatic focusing in the form of a bifilar helix was chosen for its light weight and potential for low voltage operation. The bifilar helix performs a dual function; that of the focusing structure and of the r.f. slow-wave circuit for the tube.

Figure 1 shows a d.c. potential difference applied between adjacent turns of the helix assembly. The resultant static field forms alternate convergent and divergent electrostatic lenses perpendicular to the direction of beam flow. An electron, because of the axial decelerating force spends more time under the more negative conductor and hence is influenced more by the convergent lens. An electron, transversing the series of lenses receives a net radially inward force, which under the proper conditions can be made to counterbalance the radially outward force exerted by space charge in a beam of electrons. The electron beam within such a focus system drifts at an average potential

$$V_o = \frac{V_1 + V_2}{2}$$

where V_1 = voltage on helix #1,

V_2 = voltage on helix #2,

$V_f = V_1 - V_2$ = focus voltage.

Figure 2 is a block diagram of the power supply arrangements which are used to establish the required potentials and operate the tube.

Utilization of the bifilar helix as an r.f. structure and a focusing structure simultaneously is a complicated compromise between r.f. performance and focusing properties. Considering the average voltage V_o, as the "tuning" voltage (i.e., that drift potential determining the frequency of operation) and choosing a minimum drift potential consistent with focusing requirements, resulted in these values for V_o:

V_o = 280 at 8.5 kmc,

V_o = 440 at 9.6 kmc.

The frequency at which operation in both the -1 space harmonic (backward-wave) and the 0^{th} space harmonic (forward-wave) can exist simultaneously is termed the upper cutoff frequency. This occurs when the parameter

$$\frac{ka}{D} = 0.5,$$

where $k = 2\pi/\lambda_o$, free space propagation constant,

a = radius of helix in meters,

D = dialetric loading factor.

Most helix type backward-wave oscillators exhibit a discontinuity in power output at this point although they will operate at even higher frequencies. A value of $ka/D = 0.43$ at midband was chosen for this tube in order to allow the helix diameter to be as large as possible without operating too close to the cutoff frequency.

A value of 0.600" was chosen for the length of the slow-wave circuit, assuming that the current for start-oscillation was available from the beam. Subsequently, this length has been increased to decrease the starting-current requirements.

For simplicity and cost considerations, the electron gun used is one furnishing a solid beam. Because of the rapid decrease in r.f. fields radially inward from the helix radius, only the outside edge of a solid beam is effective in interaction with the slow-wave circuit. Electrons near the center of the beam are not modulated and hence decrease the d.c. to r.f. conversion efficiency. This problem is solved in the permanent-magnet focused BWO by using a hollow cylindrical beam formed by a comparatively expensive electron gun. The problem solves itself with the periodic-electrostatic tube because of the radial distribution of focusing force.

Just as the r.f. fields are an inverse function of radius, so too are the focusing fields. Electrons near the beam edge are in a strong focusing field region; whereas electrons near the axis are practically unaffected by focusing fields, but are influenced by a radially outward space charge force. Hence these inner electrons drift toward the beam edge until they reach an equilibrium radius where focusing force balances space-charge repulsion. A small percentage of the beam never reaches equilibrium because of the short transit time through the slow-wave circuit. The most important condition for focusing a solid beam with this system is that the edge electrons be at their equilibrium radius with no transverse velocity at the entrance to the bifilar helix.

Figure 3 demonstrates the effect of electrostatically focusing a solid beam. This picture was drawn from observation of beam interception with a thin carbon screen placed across the beam exit of the focusing structure. With no focusing potential applied, the major part of the beam drifts unperturbed through the bifilar

helix. Application of the proper focusing potential separates the beam into three thin segments forming a circle slightly smaller than the helix diameter. The sementation is thought to be caused by the electrostatic charging of the three helix support rods by stray electrons. The field so produced creates an improper focusing potential in the vicinity of the support rods and hence defocuses and scatters the beam passing close by. Better shielding of the support rods from stray electron interception can no doubt decrease this effect with a consequent slight increase in efficiency.

Electron beam optics have, in this tube, been determined purely by a cut-and-try process. Some theoretical analysis could perhaps have been performed but boundary conditions and primary assumptions were so indefinite that the experimental determination of proper beam shape and entrance conditions seemed to be the more practical approach.

To provide the backward-wave oscillator with the objective configuration of a reflex klystron, we borrowed the concept of the standard waveguide-to-coaxial transmission line transition. In place of the usualy dipole antenna is a cylinder housing the bifilar helix assembly. The cylinder projects thru the waveguide leaving a small gap to the opposite wall. The usual waveguide short is placed behind the transition for matching purposes.

Electrical and r.f. connections are made to the slow-wave circuit by welding connections to the bifilar helix across the gap formed between the helix cylinder and opposite wall of the waveguide. The dipole-like cylinder is isolated from the waveguide by a ceramic insulator. An r.f. choke prevents energy leakage around the insulator.

Some special studies were conducted on the resultant r.f. match to determine if we were indeed coupling to the backward-wave space harmonic of the bifilar helix. Field solutions of the various space harmonics for a helix (single or multi-filar) show that

$E_{zo} \neq 0$ at $r = 0$

E_z = longitudinal electric field

r = radius

whereas

$E_{z-1} = 0$ at $r = 0$

for the -1 space harmonic or backward wave.

A conducting rod down the center of the helix will short out any electric field existing on the axis. Measurements of the r.f. match showed no change with the presence of this conducting rod. A typical match pattern is shown in Fig. 4 where VSWR is displayed as a function of frequency. It should be noted that the irregularities are due to some reflections from the collector end of the helix assembly which was not terminated when this data was obtained.

Electrons intercepted by the helix turns tend to heat the molybdenum tape and increase its r.f. loss. The helix loss has a large effect on efficiency and beam starting current requirements. To provide a better heat sink, the helix support rods are made of beryllia, an insulator with the thermal conductivity of brass. The helix assembly is shrunk fit into a precision broached bore in the dipole cylinder further enhancing the heat drain from the helix. Heat is carried from this cylinder through the ceramic insulator and out through the collector. Figure 5 is a cross section drawing of the experimental tubes as they have been constructed at this writing. There are many parts of the tube which were drawn from existing assemblies for expediency rather than for compactness. Further design refinement will reduce the overall length of the tube approximately one inch.

Backward-wave oscillators require an r.f. termination at the collector end of the slow-wave circuit. This usually takes the form of sprayed graphite in a helix-type tube, but because of the electrical d.c. isolation required turn-to-turn in this tube, the normal termination was impractical. Various d.c. isolated terminations have been used with only partial success. Further effort is being expended to develop an inexpensive reliable termination.

OPERATING DATA

The marriage of a solid beam to a periodic-electrostatically focused backward-wave oscillator has produced some gratifying results. Experiments with gun and beam testers verified approximate theoretical calculations of focus voltage requirements. They also revealed a need for lower starting current requirements. The data discussed in this section may be favorably compared to the design objectives described earlier.

The present tube uses a gun with an area convergence of 8 or 9:1 and a fairly high perveance of

$$K = \frac{I_o}{V_1^{3/2}} = 6 \times 10^{-6} \text{ perv.}$$

where I_o = beam current,

V_1 = anode and helix #1 potential.

Gun perveance, of course is a function of anode-cathode spacing. This spacing also affects the beam entrance conditions for focusing to the extent that the entrance conditions determine the perveance. In order to change the perveance appreciably, an extensive redesign of the gun and entrance optics would be necessary. Figure 6 is a plot of power output vs frequency over the X-band range of interest. Power output is a function of focus voltage and is plotted for two different values in Fig. 6. By judicious selection of focus voltage it is possible to achieve larger bandwidths at a sacrifice of power level. It is also feasible to program the focus voltage such that as the tuning voltage is varied, the power output curve may be made a predetermined value (within limits) versus frequency. The power output characteristic presented here is from an early experimental tube without a termination.

Figure 7 shows the average tuning voltage V_o vs frequency of the same early tube. We have since decreased the overall voltage requirements by approximately twenty per cent.

Beam currents for the power levels shown are quite high leading to disappointing conversion efficiencies as shown in Figure 8. Much work is in progress to raise these efficiencies and in fact one very recent tube had better than double the efficiency of the tube represented by Fig. 8. It is yet too early to predict the pratical limitations in efficiency as the design is undergoing considerable refinement both on focusing and thermal stability.

Figure 9 is a photograph of our experimental tube, waveguide down, next to a typical reflex klystron. As mentioned in the section on constructional details the tube may be made considerably shorter by proper design refinement of the collector-insulator region.

CONCLUSION

Although development is not yet complete, the program has produced results up to or exceeding the original design goals. The feasibility of a compact lightweight, voltage-tuneable oscillator has been demonstrated.

The device is reasonably simple to construct and with proper production techniques, should be relatively inexpensive. The weight, size and configuration are those of a reflex klystron with the possibility of lower values with proper refinement.

Some basic refinements remain to be performed which should lead to a marketable device in the near future.

REFERENCES

1. See, for example, H. Heffner, Analysis of the Backward-Wave Traveling-Wave Tube", Proc. I. R. E. Vol. 42, pp. 930-937; June 1954.

   and J. R. Hechtel, "Electrostatic Focusing of Microwave Tubes", Microwave Journal, p. 86; Nov. 1960

ACKNOWLEDGEMENTS

We wish to extend our appreciation to the following: to Mr. Tony King for helpful discussions regarding the r.f. match, to Messrs. John Fronduto, George Byron and Oley Cunningham for patient and accurate planning and assembly of the device, and to countless others at Microwave Associates for help and cooperation along the way.

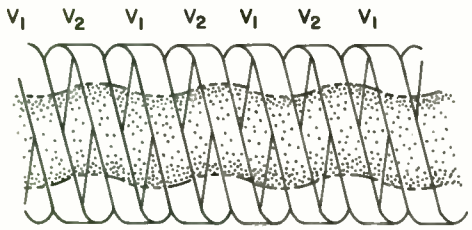


Fig. 1. Bifilar helix for focusing an electron beam.

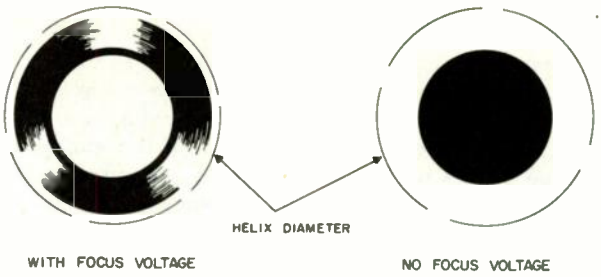


Fig. 3. Beam shape at collector end of helix.

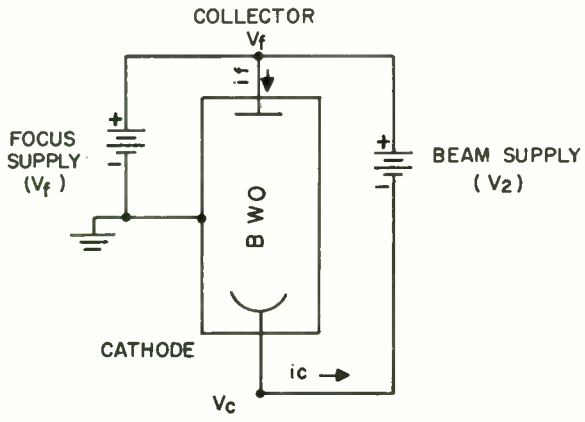
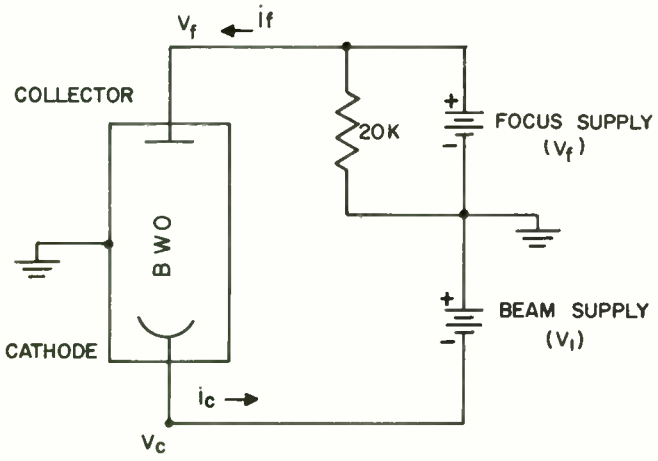


Fig. 2A. Low voltage circuit.

Fig. 2B. Standard circuit.



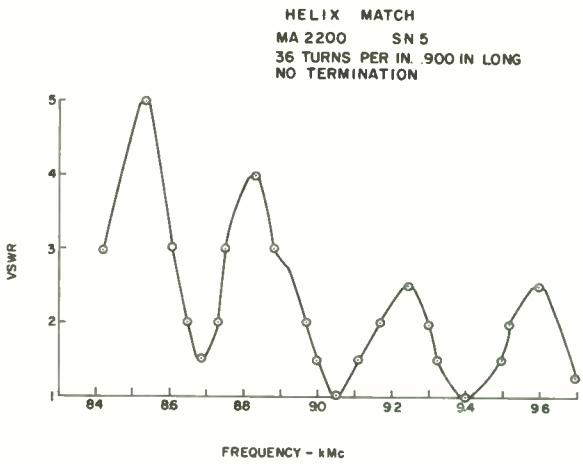


Fig. 4.

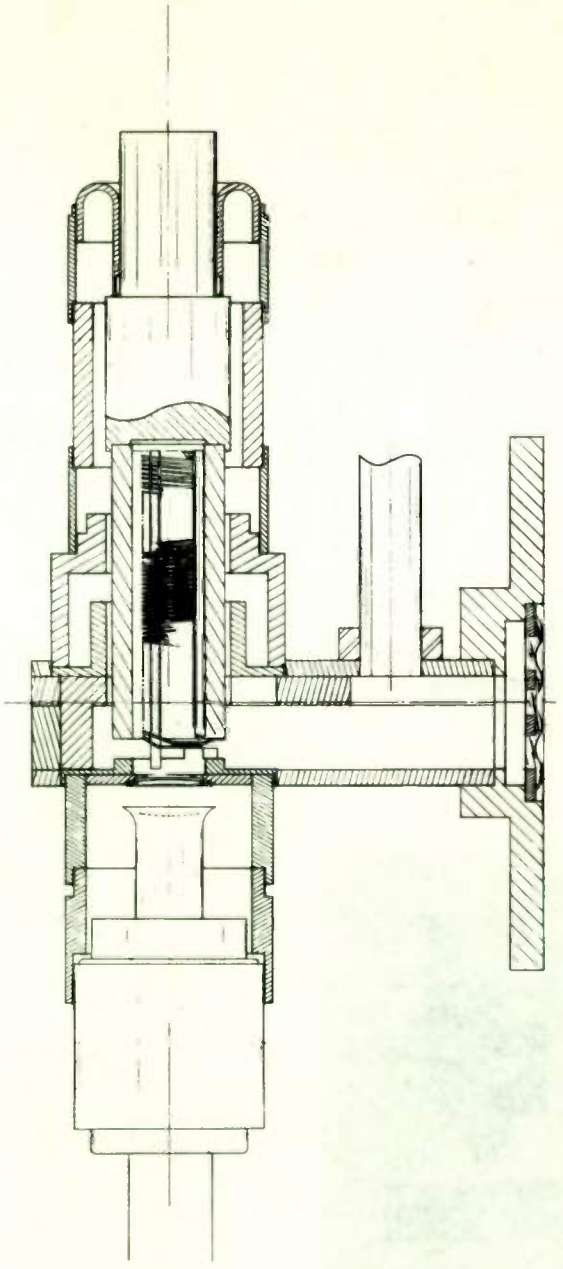


Fig. 5. Crossection of backward-wave oscillator.

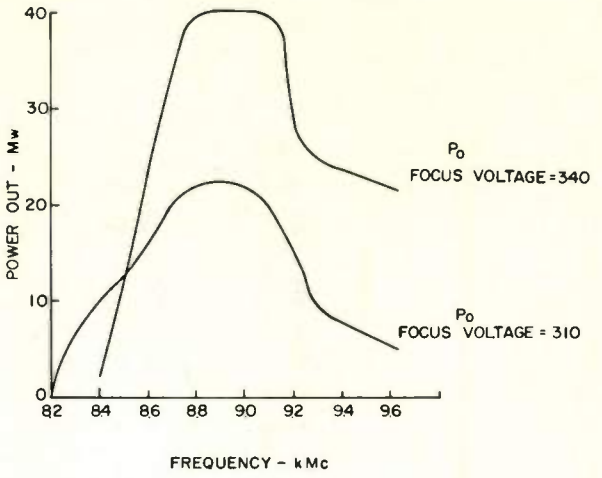


Fig. 6. Operating characteristics MA 2200 SN 5.

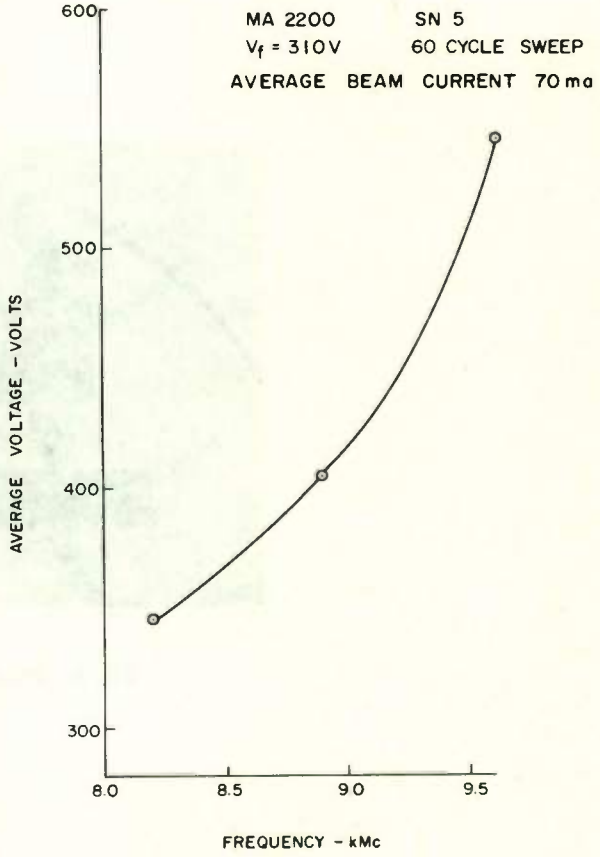


Fig. 7. Tuning voltage.

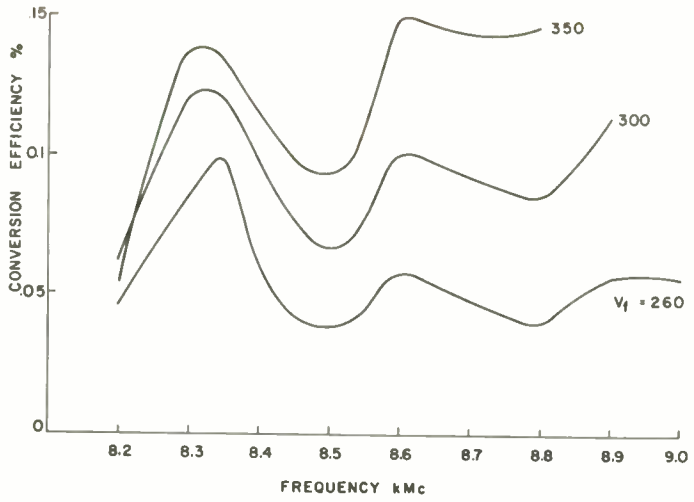


Fig. 8. MA 2200 #5 CW operation.

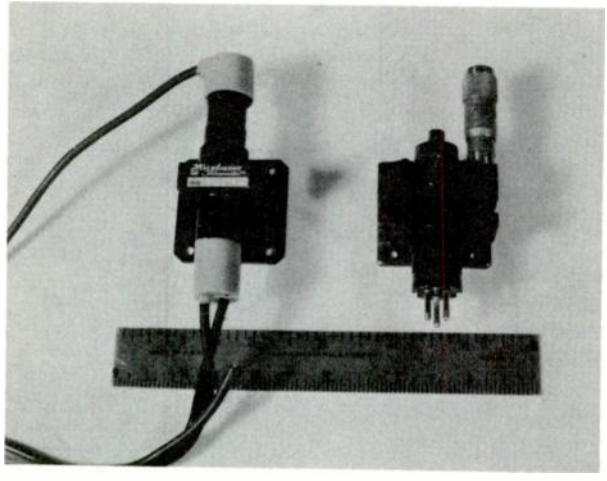


Fig. 9. Physical comparison.

THE PROPERTIES OF THERMO-ELECTRIC ELEMENTS AS MICROWAVE POWER DETECTORS

S. Hopfer, N. H. Riederman, L. A. Nadler
PRD Electronics, Inc.
Brooklyn, New York

I Introduction

The development of low level, microwave power measuring devices has been restricted in the past almost exclusively to the design of bolometric elements. This is surprising, in view of the impressive amount of work which has been done on thermo-electric devices as radiation detectors. In fact, the state of the art in this field has progressed to a point close to the theoretical limit[1], and although highly sensitive bolometric radiation detectors have been developed, the thermo-electric devices have found[2] more extensive use by far in radiometric measurements. Accordingly, we shall show in the following discussion, that the use of thermo-electric elements in the microwave range is not only feasable but offers many advantages over the bolometric systems.

II Thin Film R.F. Thermocouples

Microwave heating in conductors is caused by surface currents. In order to assure proper conversion of the incident microwave energy into heat by means of these currents, the thickness of the thermo-electric elements must be small relative to the skin depth at the highest operating frequency. Since it is extremely difficult, if not impossible, to draw thin enough wires of suitable materials or to fuse or solder them to form the necessary junctions, the use of thin film thermo-electric elements is implied. Accordingly, we shall restrict our discussion to thin film devices only.

The basic situation we wish to describe is illustrated in Fig. 1. It is seen that r.f. currents generated in the metallic boundaries by the microwave fields are made to flow capacitively through the substrate material and then conductively through a thin metallic film consisting of metals A and B to complete the r.f. circuit. Owing to the Joule heating which takes place in the film, its temperature is changed, leaving the center of the film at the junction of the two metals at some elevated temperature T_1 and the ends of the film which are in contact with relatively large thermal masses at the ambient temperature T_a. As a result of this temperature distribution, a d.c. current will flow in the external circuit as shown in Fig. 1. This non-conventional use of the thermocouple raises some important questions.

1. Does the r.f. requirement of the use of thin metallic films have any adverse effect on the magnitude of the thermal emf?

2. Is the thermal emf affected by the r.f. currents in the film other than through the thermal effects which they produce?

Effect of Film Thickness

As is well known, the material constants of metals such as electrical conductivity, thermal conductivity, etc. can not be treated as constants any more when used in connection with thin films. Since the electronic mean free path in this case is comparable to the thickness of the film, those characteristics of the metal which depend upon it become functions of the thickness. In particular, the thermal emf should change with film thickness, as shown by Meyer[3] and others. If we now demand that the film thickness t, shall not exceed a given fraction p of the skin depth δ, we can determine a lower limit of the film resistance R_o in Ω per unit square, provided we know the dependance of the resistivity on the film thickness for any particular material. In order to obtain a feel for the numbers involved, we have chosen the factor 5 for the ratio of film to bulk resistivity independant of material and have plotted $R_{o\ min}$ for various materials as a function of frequency. This graph is shown in Fig. 2. It is evident that our interest lies in reasonably thin films. We have measured the thermal emf of Bismuth-Antimony (Bi-Sb) combinations for films varying in resistance from 7 $\Omega/\square$ to 200 $\Omega/\square$. The corresponding film thicknesses are estimated[4] to range from 5000 Å to 300 Å. Fig. 3 shows the results of these measurements. It is seen that within the accuracy of our measurements, the thermal emf for the Bi-Sb combination does not show any thickness dependance for the range of thicknesses considered. The seemingly lower value of sensitivity of about 90 $\mu v/°C$ compared to the bulk value of about 110 $\mu v/°C$ should not necessarily be considered a thin film effect, since the magnitude of this difference can be attributed to a number of other causes. It is worth noticing that the reproducibility of the thermal emf in the various samples is quite good, remembering that these samples have been made at different times and no doubt at conditions of evaporation which cannot be considered identical. This seems to indicate that these films are polycrystalline showing no particular ordering, since otherwise one should expect differences in the thermal emf owing to its strong dependance upon the anisotropy of the Bismuth and Antimony crystals. A view of the experimental set up for the above measurements is shown in Fig. 4.

Effect of R.F. Currents

The presence of high frequency currents in the film should theoretically have no direct influence upon the thermal emf. This follows from the nature of the thermal emf, which according to Bridgman[5] is a driving emf, located in the individual conductors and generated by the tendency of the electrons to restore thermal equilibrium. This tendency being greater in one conductor than the other produces the resultant emf. Thus, the overall process may be pictured as follows: the electromagnetic field induces currents in the metallic film delivering energy in the form of i^2R losses to the film. This energy in turn is almost entirely lost to the surroundings through thermal conduction, convection and radiation, leaving the film in a non-equilibrium state with a certain temperature distribution. The tendency of the electrons to restore equilibrium gives rise to the thermal emf driving a current through the circuit which is sustained by the absorption of heat at the junction in the form of the Peltier heat and along the leads in the form of the Thomson heat. Evidently, the process described above is completely insensitive to the manner in which the state of thermal non-equilibrium is maintained. This is clearly shown in Fig. 5 where deviations in the thermopile output for constant r.f. power are plotted against frequency. The apparent decrease in sensitivity with frequency is due to r.f. mount losses which can be largely eliminated by proper design.

III The Thermo-Electric Load

As mentioned previously, there is an extensive literature[6] on the optimum design of thermocouples and thermopiles for radiometric measurements. Many of these results apply directly to our case, yet there are others which are unique to the present application. In particular, we shall comment in the following on the problems of thermopile versus thermocouple, and the shape and drift characteristics of the thermo-elements.

Thermopile Versus Thermocouple

Roess and Dacus[7,8] have shown that if properly designed thermocouple amplifier systems are used, the various thermal detectors may be compared on the basis of their respective signal to noise ratios. Under these conditions the signal to noise ratio of a thermocouple is ideally equal to, but in practice better than that of a thermopile. Consider now the thermocouple of Fig. 1. Its signal to noise ratio is proportional to $E_1/\sqrt{R_1}$ where E_1 is the thermocouple voltage and R_1 is its resistance, assumed to be equal to Z_o, the characteristic impedance of the transmission line which it terminates. If we now use n thermo-electric branches instead of a single one, the branch resistance R' must be so chosen that all the n branches, being in r.f. parallel and sharing the power equally, are now terminating the line, i.e. $R' = nZ_o$. But the d.c. resistance R_n of the pile is at least nR', since some extra path length is always involved in turning the resistive path around so that

$$R_n \geq n^2 Z_o .$$

Similarly, we find that the thermal emf E_n of the pile is given by

$$E_n \leq nE_1$$

since the power is now shared by n branches and the thermal losses of the n branches to the surroundings is at best as low, but practically larger than that of the single branch. Thus, we have

$$\frac{E_1}{\sqrt{R_1}} \geq \frac{E_n/n}{\sqrt{R_n/n^2}} = \frac{E_n}{\sqrt{R_n}}$$

which is the result stated by Roess and Dacus. This conclusion, however, is only valid when the Johnson noise of the thermocouple is the limiting factor, otherwise the thermopile offers better sensitivity. Furthermore, there are practical considerations such as drift, thickness of film and others which greatly favor the use of a pile over a couple.

Shape Factors

Consider the truncated film strip of Fig. 6. It is desired to find the temperature of the strip along $X = 0$, assuming that at $X = \pm a/2$ the film is at the ambient temperature T_a. If $\tau(x) = (T(x) - T_a)$, is assumed to be small, so that the convective and radiative losses are directly proportional to τ, equation (1) is a statement of the heat balance which has to hold. For computational simplicity we are ignoring the y dependance of $\tau(x,y)$ which should actually enter into equation (1) through the divergence of the conductive heat flow. This approximation seems valid if the angle of the strip is within about 20° to 30° and should hold particularly well along the line $y = 0$, or almost anywhere else on the strip if the total conduction loss is only a small part of the overall thermal loss to the surroundings. Two special cases are of practical interest for the type of films we deal with. Case 1 is the condition of free convection where the film element is mounted in a microwave housing at atmospheric pressures. Under these conditions it can be demonstrated that the convective heat loss is by far the greatest, justifying the dropping of the conduction term as far as $\tau(o)$ is concerned. Case 2, covers the vacuum condition with no convection losses but only conductive losses. This situation may not always be as readily realized in practice as that of case 1, since the radiation losses may become comparable to, or even predominate over the conduction losses. Since our application does not require the use of a blackened receiver as do the radiometric devices, the low emissivity of the desirable metallic films of reasonable thickness will keep the radiative losses considerably below the conduction losses. However, if the film is thin enough to become transparent, the approximation of case 2 is not applicable. Another

point of caution in the use of the expression for $\tau(o)$ is the fact that both K and t enter into it. Since in a matched system the total strip resistance must remain constant, t must change with ζ and so will K due to its thickness dependance. As K and t are treated as constants in our expression, the formula applies strictly speaking only if the substrate backing plays a considerable part in the conduction process. In any event, the above expression represents a lower limit of what can actually be expected under the conditions of case 2. Defining the gain factor G as the ratio of $\tau(o)$ of the truncated strip and $\tau(o)$ of a rectangular strip of width d, we can plot these factors as a function of the variable ζ. This is shown in Fig. 7. It is seen that in the case of free convection G increases as ζ increases, whereas under vacuum conditions G decreases slightly as ζ increases. In cases where both types of loss mechanisms are present one would expect the gain factors to fall somewhere between the two curves G_1 and G_2. It may be concluded, therefore, that the detection sensitivity of the thermocouple or pile is not criticially dependant upon its shape. This conclusion is fortunate in view of the undesirable microwave effects which result from employing large ratios of (b/d).

Drift Characteristics

As mentioned previously, the thermal emf is a driving emf caused by a state of thermal non-equilibrium. This implies that uniform changes in ambient temperature will by themselves not generate a thermo-electric output. In practice, however, changes in the ambient temperature are usually accompanied by the generation of thermal gradients, which in turn give rise to output drift. Fortunately, this type of drift is largely eliminated in our application. This is demonstrated in Fig. 8 which depicts two thermo-electric branches in thermal contact with two isothermal metallic masses at temperatures T_1 and T_2 respectively. Due to non-equality of T_1 and T_2, a temperature gradient is set up along the branches. If properly designed, the junction temperature T_j can be made to correspond to the average temperature of T_1 and T_2. Under these conditions it is easily seen that cancellation of the partial emf terms will occur resulting in zero output. On the other hand, when current flows through the branches, T_j will be larger than $(T_1 + T_2)/2$, resulting in the desired output e_o. In microwave structures one usually deals with metallic boundaries which shield the internal parts from direct interaction with the outside. Thus, outside thermal interaction with the thermopile takes place via the metallic boundaries, which permits the design of internally balanced structures as demonstrated above. Fig. 9 shows several balanced film structures and Fig. 10, a microwave housing for them.

IV Conclusions

Our work with thermo-electric films has shown that these elements can be readily used for the detection and measurement of microwave power. Furthermore, these elements show decided advantages over bolometric elements with regard to broadbanding, drift characteristics, simplicity of operation and overall accuracy.

References

1. Havens, R., Jour. Opt. Soc. Am. -- 36, 355 (1946)

2. Strong, J., Proc. in Exp. Phys. -- Prentice-Hall, New York -- p. 308 (1938)

3. Meyer, Structure and Prop. of Thin Films -- Wiley and Sons, Inc., New York -- p.240 (1959)

4. University of Denver Research Institue, Final Report -- N123(60530)10049A -- ASTIA No. AD-250-090 (November 1960)

5. Bridgman, P. W., The Thermodynamics of Electrical Phenomena in Metals -- Mac Millian Co. New York (1934)

6. PRD Engineering Report 5.2-1 -- W 33-038-ac-15474 (November 1946)

7. Roess and Dacus, Rev. Sci. Ins. -- 16, 164 (1945)

8. Roess, L. C., Rev. Sci. Ins. -- 16, 172 (1945)

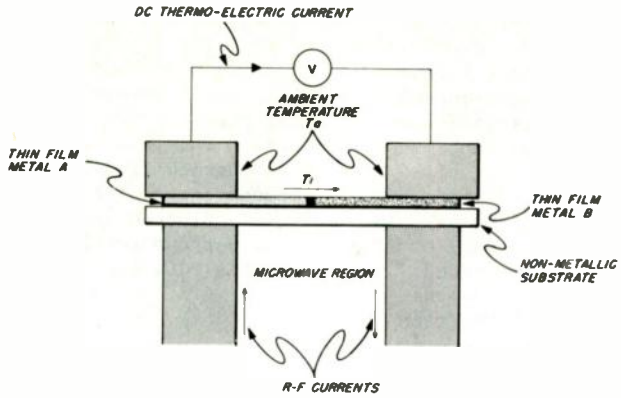


Fig. 1. Schematic of high-frequency thermocouple arrangement.

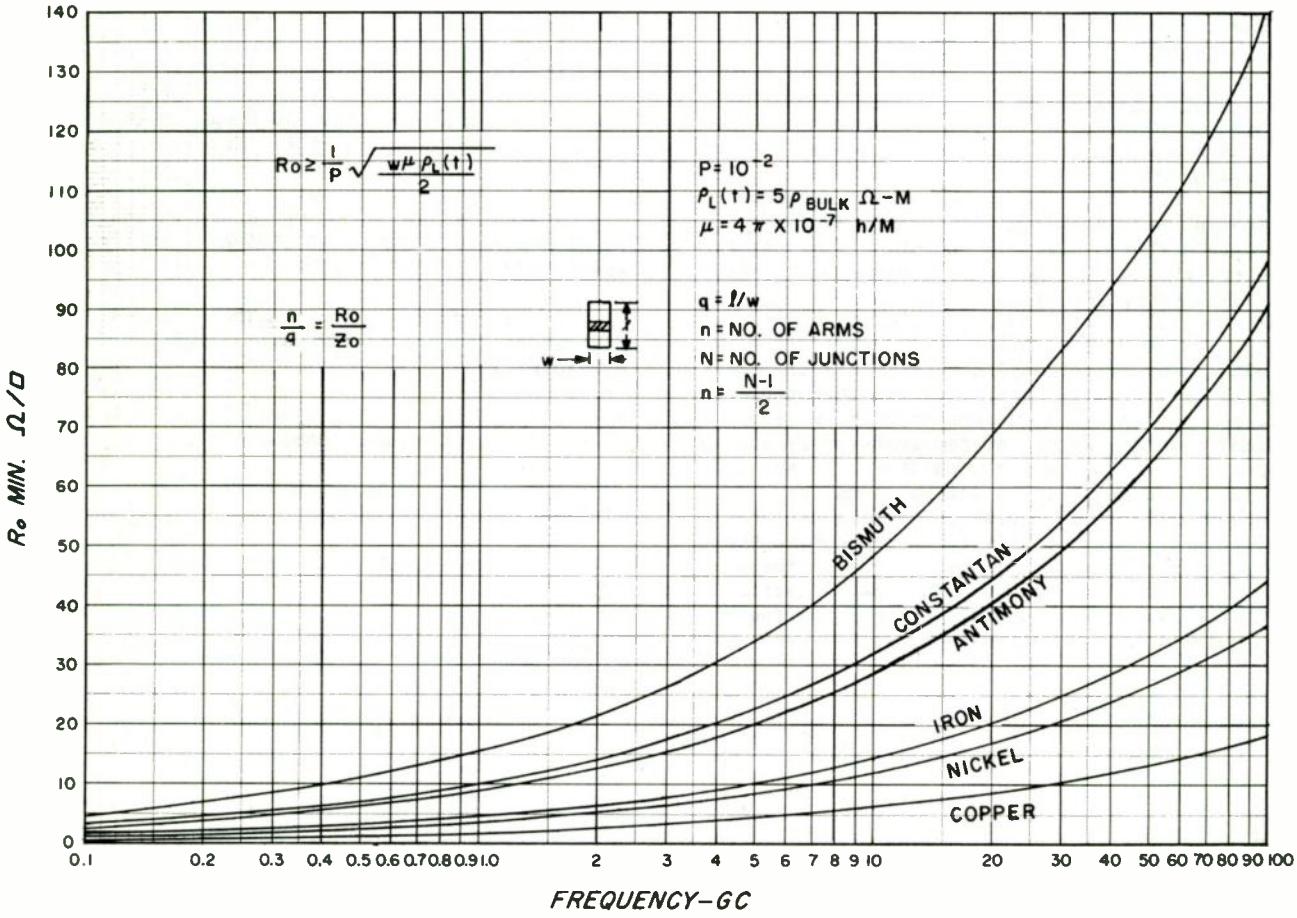


Fig. 2. Film resistance vs. frequency.

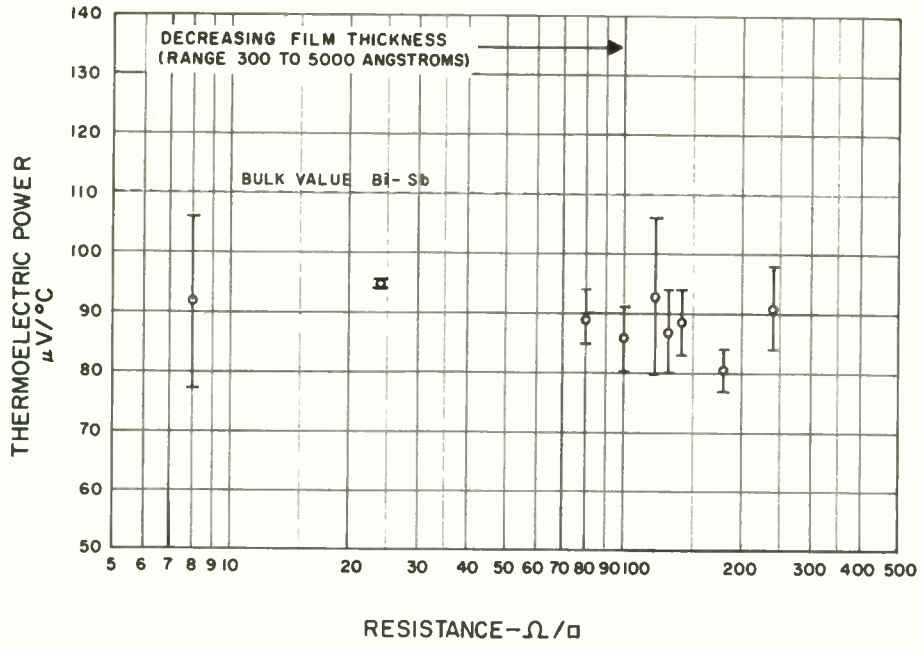


Fig. 3. Thermo-electric power vs. resistance.

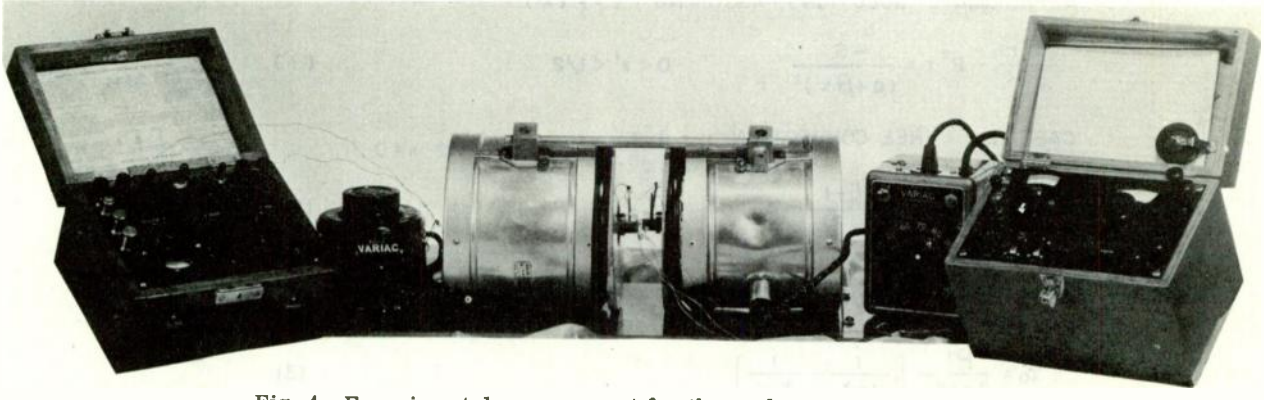


Fig. 4. Experimental arrangement for thermal EMF measurements.

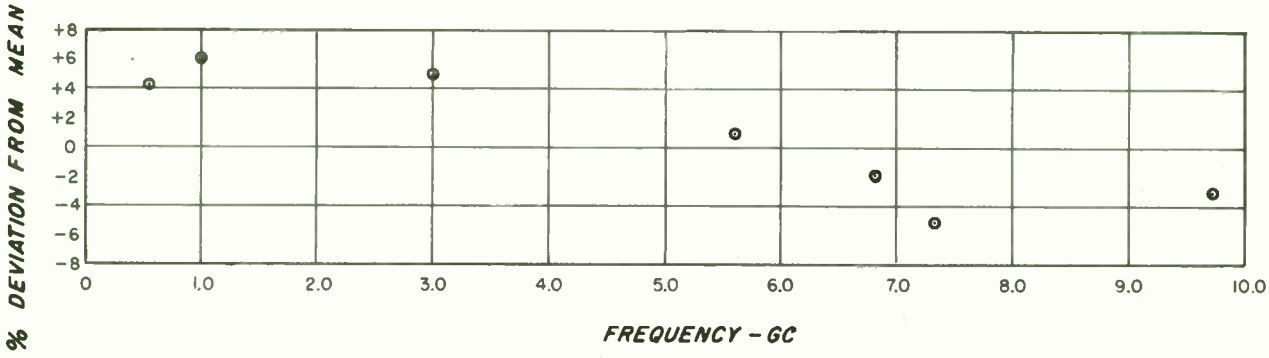
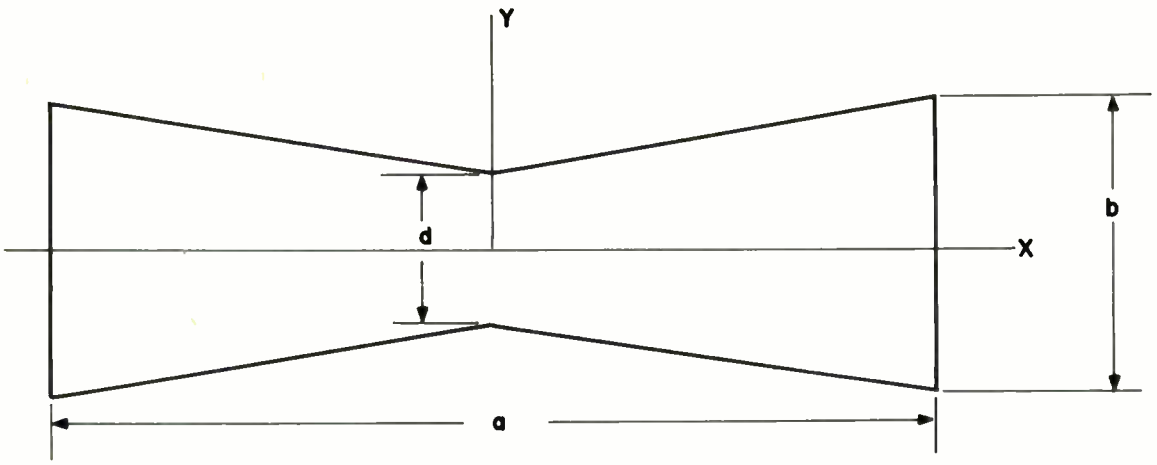


Fig. 5. Power sensitivity vs. frequency. (Ref. power standard-PRD type N670 calorimeter-absolute accuracy ± 2%)



APPROXIMATE SOLUTIONS, ASSUMING $\tau = f(x')$

$$\frac{d^2\tau}{dx'^2} - P^2\tau = \frac{-S}{(\alpha+\beta x')^2} \qquad 0 < x' < 1/2 \qquad (1)$$

CASE 1: FREE CONVECTION $\frac{d^2\tau}{dx'^2} << P^2\tau$ at $x' = 0$

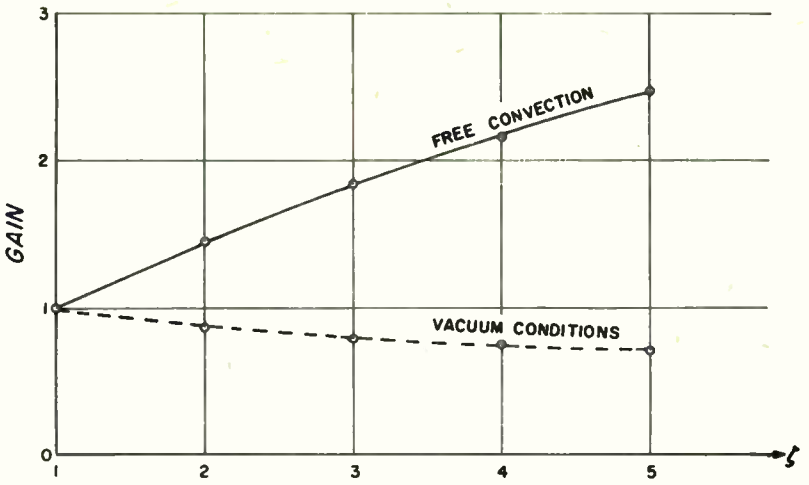
$$\tau_0 = \frac{P'}{qad}\ \frac{\zeta-1}{\ln\zeta} \qquad (2)$$

CASE 2: VACUUM CONDITIONS $P^2\tau << \frac{d^2\tau}{dx'^2}$

$$\tau_0 = \frac{P'}{4\kappa td}\left[\frac{1}{\ln\zeta} - \frac{1}{\zeta-1}\right] \qquad (3)$$

P' = POWER DISSIPATED IN STRIP
q = CONVECTIVE LOSS CONSTANT
κ = THERMAL CONDUCTIVITY OF STRIP
t = FILM THICKNESS
ζ = (b/d)

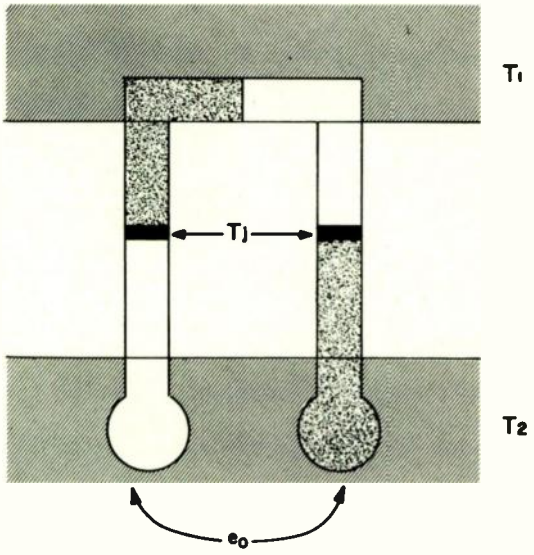
Fig. 6. Truncated strip design.



$$G_1 = \frac{\zeta - 1}{\ln \zeta} \text{ (FREE CONVECTION)}$$

$$G_2 = 2\left(\frac{1}{\ln \zeta} - \frac{1}{\zeta - 1}\right) \text{(VACUUM CONDITIONS)}$$

Fig. 7. Truncated strip gain factors.



IF $T_1 \neq T_2$ $e_0 = 0$ IF $T_j = \frac{T_1 + T_2}{2}$

IF CURRENT FLOWS THROUGH STRIP

$T_j > \frac{T_1 + T_2}{2}$ AND $e_0 > 0$

Fig. 8. Internal balance of thermocouple.

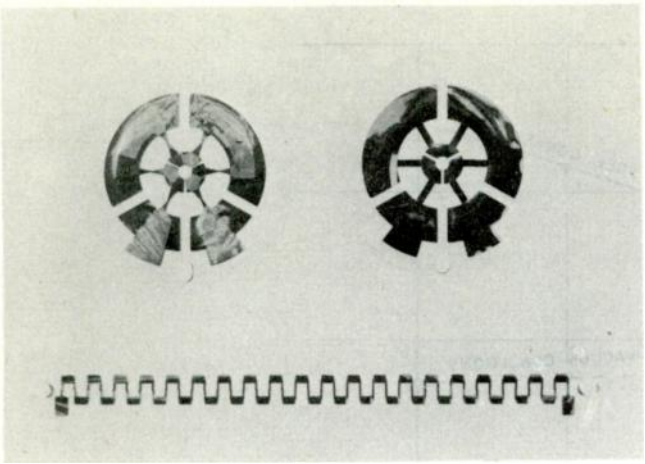


Fig. 9. Thermo-electric film elements.

Fig. 10. Coaxial test mount.

A BROAD-BAND UHF PARAMETRIC AMPLIFIER

Richard LaRosa
Hazeltine Research Corporation
Little Neck, New York

Introduction and Summary

This paper describes a fixed-tuned diode parametric amplifier for the frequency range 406 to 450 Mc. The one-port was chosen in preference to the other types of parametric amplifiers for the following reasons:

1. The diode mount of a one-port amplifier requires external ports only for signal frequency and pump frequency. This makes the construction quite simple compared to a mount which must provide an external port for one of the sidebands.

2. Two or more one-port stages are easily cascaded with resultant increase in gain, bandwidth, and noise-figure safety margins. The ease of cascading is due to the fact that the signal output frequency is the same as the input frequency, allowing identical or similar design of the successive stages.

The amplifier to be described uses two stages to achieve an over-all noise figure under 2.0 db when a vacuum-tube post amplifier is used. The gain per stage is only about 9 db. This reduces sensitivity to pump-level changes and temperature drift of the components. A block diagram of the amplifier is shown in Fig. 1.

Circulator System

The success of this type of amplifier is critically dependent on the availability of broad-band low-loss circulators. Such circulators for the 400-Mc range became available commercially in mid 1960. Three of these are used: one in conjunction with each of the two one-port amplifiers to provide separate input and output ports for each stage, and one to serve as an isolator between stages. A compact field-displacement or absorption-type isolator might have been used in place of the interstage circulator. However, better over-all performance was obtained by specifying the entire nonreciprocal network as a single assembly. Manufacturers who quoted on the assembly invariably preferred to use Y-junction circulators throughout.

Main Signal Path

Fig. 2 shows the intended signal path. A wave of unit amplitude incident on the input port of the first circulator is routed with complex voltage transmission coefficient t_1 to the first parametric one-port. The complex reflection coefficient ρ of the one-port has a magnitude of about 3. The reflected wave from this one-port is routed to the next circulator, which in turn routes it on to the third circulator. Assuming that the circulators are identical, the reflected wave from the first one-port suffers the complex transmission coefficient t_1^3 in making its way to the second one-port.

The second one-port (identical to the first for simplicity) also has a reflection coefficient ρ with magnitude of about 3. The reflected wave from this one-port leaves the output port of the last circulator and proceeds to the post amplifier.

All told, the forward signal path makes five passes through circulators and undergoes two reflections. The forward gain is therefore:

$$t_1^5 \rho^2 \qquad (1)$$

Feedback Loops

Fig. 3 shows the six principal feedback loops which are present in this amplifier. Fig. 3a shows that a signal traversing the full forward path of the amplifier can be reflected by the post amplifier, whose reflection coefficient is ρ_2. This reflected wave is transmitted back to the second circulator which has been included to stop just such backward transmission. Most of the backward wave is absorbed in the termination of this circulator, but the circulator allows some backward transmission, specified by the complex number t_2, toward the first circulator. The first circulator routes the backward wave to the source. The reflection coefficient ρ_1 of the source completes the feedback loop. The complete loop of Fig. 3a contains seven forward passes through circulators and one backward pass, as well as four reflections. The loop gain is therefore:

$$\text{Path (a): } t_1^7 \, t_2 \, \rho^2 \, \rho_1 \, \rho_2 . \qquad (2)$$

The feedback path shown in Fig. 3b is similar to the above, except that the reflected wave from the second one-port returns via the reverse transmission t_2 of the last circulator. Thus, two forward passes are eliminated and one reverse pass is substituted for the post-amplifier reflection. The loop gain is therefore:

$$\text{Path (b):}\quad t_1^5 t_2^2 \rho^2 \rho_1. \qquad (3)$$

In the feedback loop of Fig. 3c, the post-amplifier reflection is included but the source reflection is eliminated by reverse transmission through the first circulator. The loop gain is

$$\text{Path (c):}\quad t_1^5 t_2^2 \rho^2 \rho_2. \qquad (4)$$

The feedback path of Fig. 3d eliminates both source and load reflections by reverse transmission in each circulator. The loop gain is therefore:

$$\text{Path (d):}\quad t_1^3 t_2^3 \rho^2. \qquad (5)$$

Fig. 3e shows a feedback path which is confined to the front end of the amplifier. The loop gain in this case is therefore:

$$\text{Path (e):}\quad t_1 t_2 \rho_1 \rho \qquad (6)$$

Similarly, Fig. 3f shows a local feedback path at the output end of the amplifier whose loop gain is

$$\text{Path (f):}\quad t_1 t_2 \rho_2 \rho \qquad (7)$$

Observed magnitude values for the various parameters are given in Table I.

TABLE I. Typical Magnitudes of Parameters

Parameter	Voltage Ratio	Equiv. db	Nomenclature
t_1	0.98	- 0.2	Forward Loss
t_2	0.22	-13.0	Isolation
ρ_1	0.33	- 9.6	Source Reflection Coefficient (2:1 VSWR)
ρ	3.0	9.5	One-port gain
ρ_2	0.20	-14.0	Load reflection Coefficient (1.5:1 VSWR)

The value of t_2 shown in Table I applies only at the band edge and at a temperature extreme; it is much better at mid-band. The other parameter values apply across the band. The forward gain and the loop gains for all the paths of Fig. 3 are given in Table II, computed from the values of Table I.

TABLE II. Forward and Loop Gains

Path	Voltage Gain	Equiv. db	Frequency where important
Forward	7.9	18.0	Flat across band
Loop (a)	0.11	-19.0	Band edge only
Loop (b)	0.13	-17.6	Band edge only
Loop (c)	0.079	-22.0	Band edge only
Loop (d)	0.093	-20.6	Band edge only
Loop (e)	0.22	-13.3	Band edge only
Loop (f)	0.13	-17.7	Band edge only

The separate loop gains listed in Table II show that the interstage isolator has broken up the most serious of the feedback loops, leaving (e) as the most troublesome.

The loop gain for path (e) can only be reduced by lowering the first-stage gain and/or providing an input isolator. There are already three ferrite devices in the main signal path of the amplifier, so that we were quite reluctant to add another. Furthermore, isolator loss at the input of the amplifier adds directly to the noise figure. Path (f) is not affected by the interstage isolator but it is affected by load reflection coefficient. This has therefore been made small by designing the input circuit of the vacuum-tube post amplifier for impedance match rather than best noise figure.

In the actual amplifier, the effect of the feedback loops can be seen on an automatic reflection-coefficient plotter when the input reflection coefficient is observed over the frequency band. There is about one revolution on the Smith Chart over the 44-Mc band (406 to 450 Mc) and the magnitude is dependent on the reflection coefficients of the two one-ports. Moreover, there is some ripple observed when a mismatched antenna on a long transmission line is simulated and over-all gain is plotted against frequency. However, while these effects can be observed in laboratory measurements, they are harder to find in a practical installation and are not noticeable in operation. The triple-circulator assembly described has therefore proved adequate.

Design of One-Ports

In order to present a reflection coefficient greater than unity in magnitude, the one-port must contain a signal-controlled active device. In the present case, a nonlinear capacitance is used to control the flow of RF pump power into the external circuit at signal frequency. In accordance with well-known principles,[1] power at the lower-sideband (idler) frequency must be dissipated. In this amplifier, idler power is entirely dissipated in the diode internal resistance.

Fig. 4 is a cross section of the one-port showing external terminals for signal, bias, and pump power. A small silicon varactor diode about 1/8" long and 1/8" in diameter is located at the

center of the drawing. One end of the diode cartridge presses against a spring-loaded piston which is surrounded by a double concentric set of spring fingers. These fingers form a low-inductance path between the piston and the fixed conducting sleeve. Movement of the fingers controls the inductance in series with the diode.

The outer conductor of the variable inductor is interrupted by a capacitive gap across which the pump voltage appears. At pump frequency, this capacitor resonates with a coaxial-cavity resonator approximately $\lambda/2$ long surrounding the diode mount, as shown in Fig. 4. The coaxial line from the pump source feeds the cavity by means of a probe at the center. A capacitive screw at the center of the cavity tunes the system to resonance at the fixed pump frequency of 5930 Mc. The resonant pump feed allows appreciable pump voltage to appear across the capacitive gap in spite of its low reactance.

At the lower-sideband or idler frequency (which lies in the range of 5502 $\pm$ 22 Mc) the pump feed resonator is almost an open circuit across the pump feed capacitor. Therefore, only the small capacitive reactance of this gap appears in series with the diode at idler frequency. Thus, the resonator pump feed causes negligible disturbance to the diode circuit at idler frequency and upper-sideband frequency. At signal frequency, the small inductive reactance of the pump feed cavity short-circuits this gap and there is no disturbance to the signal circuit. It is, however, necessary to control the gap capacitance carefully to avoid a parallel resonance with the feed cavity near the signal-frequency band.

The diode is so located in the coaxial line that the pump feed gap is at about the mid-point of the diode cartridge. One end of the diode is connected to the outer conductor by the spring-loaded piston and contact fingers already described. The other end of the diode presses against a stationary contact whose capacitance to the outer conductor completes the diode circuit as far as pump, lower side-band and upper-sideband frequencies are concerned.

Connection to the UHF circuit box is made across this latter capacitive diode contact. The lead from the contact is very thin and becomes the center conductor of a coaxial trap one quarter-wave long at pump frequency. This choke prevents pump, idler, and upper-sideband currents from flowing into the UHF box by introducing an open circuit in the lead coming from the diode contact. The short-circuited end of this quarter-wave choke is connected via a button mica capacitor to the center conductor of a male type-N panel-mounted connector. A piece of bus bar from the choke to a button mica bypass forms a shunt inductor and a DC feed to the diode. The bias lead enters the UHF compartment via a button mica capacitor and an RF choke as shown.

Except for the pump-feed resonator and the UHF-lead trap, the entire one-port is composed of lumped circuit elements. The equivalent circuit of Fig. 5 includes all the elements which have been mentioned so far. Symbol L_2 designates the variable inductor formed by the movable contact fingers, C_p the gap capacitance across one end of the pump-feed resonator, and C_b the shunt capacitor formed between the stationary diode contact and its support.

The equivalent circuit of the diode cartridge has been included in Fig. 5. The non-linear junction capacitance C_j is in series with the bulk semiconductor resistance R_s. The shunt capacitance between the semiconductor block and contact spring is C_1. The inductance L_1 represents the magnetic field due to currents on the metal end pieces, and the package shunt capacitance is C_2. Both C2 and L_1 depend on the surrounding metal parts. In this diode mount, these surrounding metal parts have been carefully designed to enclose the diode as tightly as possible. This is to decrease L_1 and L_2 in order to obtain maximum bandwidth, as will be explained.

Using the simplest theory, the diode junction capacitance C_j is caused to vary sinusoidally at the radian frequency ω_p of the pump source. As far as the effect on the signal is concerned, the pump voltage and charge may be ignored and C_j may be replaced by the time-varying capacitance,

$$C_o + C\left(e^{j\omega_p t} + e^{-j\omega_p t}\right) \qquad (8)$$

where C_o is the average junction capacitance, $4C$ is the peak-to-peak variation. The time origin is taken at the forward peak of the pump waveform. The average capacitance can be included with all the other linear, non-varying circuit elements, as shown in Fig. 6.

The time-varying part of the capacitance to the left of the dotted line acts as a frequency translator. The admittance looking to the right from these terminals is Y_2 at idler radian frequency ω_2 and Y_3 at upper-sideband radian frequency ω_3. These admittances are respectively transformed into the admittances y_2 and y_3 which are seen looking to the left from the dotted terminals at signal radian frequency ω_1. The transformations are described by

$$y_2 = \frac{-\omega_1 \omega_2 C^2}{Y_2^*} \qquad (9a)$$

$$y_3 = \frac{\omega_1 \omega_3 C^2}{Y_3} \qquad (9b)$$

The admittance y_2 contains the negative conductance responsible for the amplifier gain. The asterisk in (9a) denotes complex conjugate. The choke in the UHF circuit lead makes it unnecessary to consider elements beyond capacitor C_b in computing Y_2 and Y_3.

It will be noted that although the circuit of Fig. 6 is not exactly a single mesh, only the small parasitic capacitances C_1 and C_2 prevent it

from being considered as such. It will turn out, however, that the diode package capacitance C_2 falls approximately at a voltage null for idler frequency ω_2. Moreover, the combination of C_o and R_s will be essentially reactive at idler frequency when a good-quality varactor is used. This permits C_1 to be lumped with C_o with a slight lowering of the value of R_s to maintain the same Q factor. These approximations result in the equivalent circuit of Fig. 7, where the reactance of C_p has also been neglected.

For maximum negative conductance, the admittance Y_2 should, according to equation (9a), be approximately real and of minimum value. This requirement is satisfied by parallel-resonating C_o with the other elements shown, and by not introducing dissipation other than R_s. Parallel resonance for Y_2 corresponds roughly to series resonance in the single mesh of Fig. 7.

For maximum bandwidth, the susceptance slope of Y_2 should be minimized. This corresponds to minimizing the reactance slope of the mesh. For any given R_s, maximum bandwidth means minimum $L_1 + L_2$. This is the reason for surrounding the diode as closely as possible with a close-fitting metal shell, since the inductance of any device approximating coaxial geometry depends on the diameter of the outer conductor.[2]

The idler-mesh bandwidth is independent of the choice of C_o and C_b. The largest available junction capacitance is used in order to minimize the signal-voltage swing at a given power level. This provides the widest dynamic range. Consideration of the maximum capacitance available in the different cutoff-frequency categories available from the diode manufacturer leads to the specification of 2.17 pf ± 15% as the junction capacitance at zero bias. The average junction capacitance when operating is assumed to have the nominal value of 1.9 pf. The stray capacitances C_1 and C_2 each appear to be about 0.12 pf for the pill package chosen.

The choice of C_b involves a compromise between two considerations. A large value of C_b increases the susceptance slope at signal frequency and narrows the bandwidth. A small value of C_b allows a large idler and pump voltage to appear across the capacitor and makes filtering in the UHF circuit lead more difficult. The value chosen, 1.8 pf, appears to be a good compromise.

If the capacitor C_2 is assumed to be at a voltage null at idler frequency, as already mentioned, the mesh capacitance is the series combination of 2.02 pf and 1.8 pf giving 0.95 pf. If the idler center frequency is chosen at 5502 Mc, the value of $L_1 + L_2$ required for resonance is 0.88 nh. The minimum inductance achievable in this diode mount is believed to be about 0.7 nh, so there is adequate margin between the nominal setting and the minimum.

The idler frequency quoted above corresponds to a pump frequency of 5930 Mc, which is near the low-frequency end of the tuning range of the VA220F klystron. This tube was chosen for its long life, fast delivery, adequate power output, and comparatively low price.

The value of R_s is about 2.0 ohms. If the idler circuit is simplified by lumping C_1 with C_o the corrected mesh resistance of Fig. 7 is

$$R_s' = R_s\left(\frac{C_o}{C_o + C_1}\right)^2 = 2.0\left(\frac{1.9}{2.02}\right)^2 = 1.8 \text{ ohms} \quad (10)$$

The mesh bandwidth is then

$$\frac{R_s'}{2\pi(L_1+L_2)} = \frac{1.8}{2\pi \cdot 0.88\text{nh}} = 325 \text{ Mc} \quad (11)$$

For a ± 22 Mc frequency change, the mesh phase angle changes by

$$\pm\frac{22}{325} \times 2 = \pm 0.135 \text{ radian} \quad (12)$$

The phase angle of Y_2 has approximately this same excursion. The mid-band value for Y_2 is approximately

$$G_2 = (\omega_{2o}C_o)^2 R_s, \quad (13)$$

where ω_{2o} is the center value of the idler radian frequency. The mid-band value of y_2 is approximately

$$-\frac{\omega_{1o}\omega_{2o}C^2}{G_2} = -\frac{\omega_{1o}}{\omega_{2o}}\left(\frac{C}{C_o}\right)^2 \cdot \frac{1}{R_s}$$

$$= -\frac{428}{5502} \cdot \frac{(0.2)^2}{2} = -1.553 \text{ m}\mho \quad (14)$$

assuming the capacitance ratio $C/C_o = 1/5$. The admittance y_2 will follow a constant-resistance circle if the idler mesh is slightly detuned to compensate for the variation of ω_1 in (9) over the frequency band and also to compensate for the fact that the conductance contributed by R_s is not constant over the idler band. The susceptance change in y_2 is roughly the value of expression (14) times that of (12), or ± 0.21 m℧.

The admittance y_2 appears at node B (defined in Fig. 5) in parallel with the node capacitance $C_o + C_1 + C_2 + C_b = 3.94$ pf. Over a ± 22 Mc band this node capacitance causes a ± 0.544 m℧ susceptance change. The inductive coupling network causes almost an additional ± 0.544 m℧ susceptance change. The admittance y_3 reflected from the upper-sideband domain according to (9b) contributes a small amount of negative susceptance which does not vary much with frequency. The total susceptance change at node B is + 1.29 m℧. The node conductance is almost constant at -1.55 m℧.

Use of Circulator for Broad-Banding

The equivalent circuit representing the impedance of the circulator is included in Fig. 5. The fixed resistor Ra can be considered to be the source. Its value was found by curve fitting to be 90 ohms. The capacitor C_a contributes a ± 1.58 m℧ susceptance change at node A over the frequency band. The two inductors of the circulator equivalent circuit, along with the inductive coupling circuit of the one-port add about ± 1.58 m℧ to the

susceptance change of node A.

Each node (A and B) is roughly resonant when the other one is short-circuited. Both node admittances have positive susceptance slope as a function of frequency. The admittance Y_L defined in Fig. 5 contains the self-susceptance of node A plus the inversion of the self-admittance of node B.

For 9.55 db gain at mid-band the mutual susceptance of the inductive network between node B and node A must be chosen such that the -1.55 m℧ node B conductance is transformed into -22.2 m℧ at node A. If this mutual susceptance were constant over the frequency band, the node B admittance locus would be transformed approximately into the constant resistance circle of Fig. 8.

The positive susceptance slope at the source node squeezes the ends of the locus together. The vectors representing numerator and denominator of the reflection coefficient at 406 Mc are shown dotted in Fig. 8. The ratio of these two vectors is 3.18, or 10.06 db, showing a 0.5 db rise in gain at the band edges.

The mutual susceptance of the inductive tuning network is, of course, not constant, since the mutual susceptance is inversely proportional to signal frequency. Some detuning of the diode contact node is required to avoid excessive skewing (loss of symmetry about horizontal axis) of the Fig. 8 loci. The internal circulator node must also be detuned. All told, there are three causes of skewing: variation of conductance in the Y_2 of Fig. 6, variation of ω_1 in expression (9a), and variation of the mutual susceptance between node B and node A. These three effects cancel each other to some extent. Detuning of the idler circuit and signal circuit nodes A and B removes tilt from the gain-vs-frequency curve. The over-all result is observed in adjusting the amplifier, so it is not necessary to specify exactly the various admittance loci.

Although the amplifier is adjusted for flat gain in the 44 Mc - wide band, the gain curve of the two-stage amplifier has been observed to be only 0.5 db down at 78 Mc width and 3 db down at 110 Mc width.

Without readjusting the inductance coupling network the gain curve tends to remain flat when diodes of different capacitance and cutoff frequency are inserted. The controls which are adjusted when changing diodes are idler inductance (L_2), diode bias, and pump amplitude. Retuning of the pump feed cavity is not usually required. Its loaded Q is only about 130. The diode bias control tends to compensate directly for changes in diode capacitance from unit to unit, while the idler inductance control takes up the rest of the variation as far as idler resonance is concerned.

Complete Dual-Channel Amplifier

Fig. 9 shows a block diagram of a dual-channel amplifier built for an airborne radar. A Type VA 220 F klystron is used to feed the four one-ports. The klystron power is split by a ring hybrid, using one output for each channel. Each of these outputs is then split in half by a second stage of ring hybrids to feed each stage. A strip-line assembly contains the three hybrids with terminations and four individual variable attenuators. Type RG-55/U coax with TNC connectors is used to feed the individual one-ports.

Fig. 10 shows a side view of the dual-channel amplifier. The circulator assembly for each channel is mounted on an aluminum side plate which obscures the view. Dotted lines have been added to the photograph to indicate the locations of the circulators.

Fig. 11 shows a close-up of the one-ports. These are mounted directly on the circulator jacks.

The isolator between the pump klystron and the power splitter was found necessary in order to adjust the pump level in one of the one-ports without varying the klystron output which would disturb the adjustments of the other one-ports. Klystron beam and repeller voltages are provided by an unregulated inverter-rectifier circuit fed from a regulated 28-volt DC supply.

The over-all noise figure is under 2.0 db when feeding a vacuum-tube post amplifier having a 7 db noise figure via 5 feet of RG-55/U cable. These measurements were made using a Kay Thermanode after comparing the impedance of the hot and room-temperature terminations.

Although the required flat gain can be achieved with many combinations of bias, idler tuning, and pump level, it was observed on one occasion that the back bias had to be less than 0.3 volts to keep the noise figure under 2.0 db. This observation was not pursued to determine the noise mechanism.

Diodes are inserted with the heat-sink end toward the idler tuner assembly. The radar pulse width is 6 μs, and at 6 watts peak incident power each one-port absorbs about 1 watt, all of which is presumed to be dissipated in the diode.

Acknowledgment

This work was a joint effort of the several individuals in the Hazeltine Research Corporation and the Hazeltine Electronics Division. The main contributors were T. Cafarella, G. Doncese, R. J. Farber, A. Grant, R. J. Keicher, D. Lambropoulos, R. A. Orlando, and A. Romeo. Special thanks are due to G. Doncese for the successful design of the UHF circuit before the role of the circulator impedance variation was fully understood. The basic one-port amplifier was a company-sponsored development. The dual-channel two-stage amplifier design was supported by the US Air Force under Contract AF33(600)-42704. The personal help of R. J. Farber and C. E. Dean in writing the paper is much appreciated.

REFERENCES

(1) L. A. Blackwell and K. L. Kotzebue, "Semiconductor-Diode Parametric Amplifiers," Prentice-Hall, Inc., Englewood Cliffs, N. J., 1961; see p. 6.

(2) W. B. Hauer, "Definition and Determination of the Series Inductance of Tunnel Diodes," IRE Transactions, ED-8, n. 6, pp. 470-475, November, 1961.

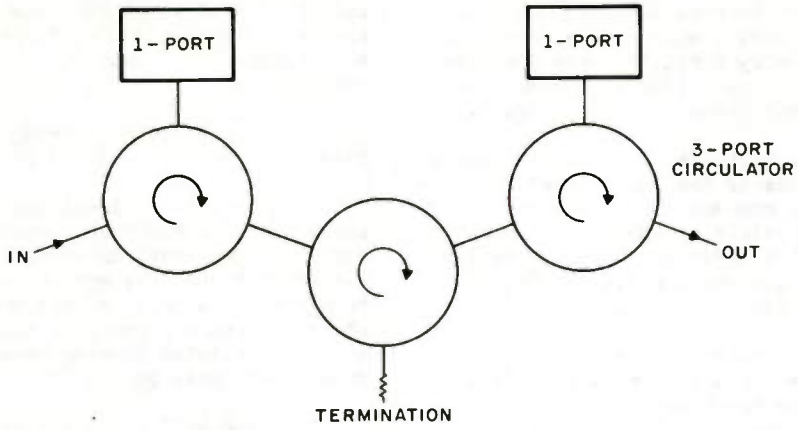


Fig. 1. Block diagram of amplifier.

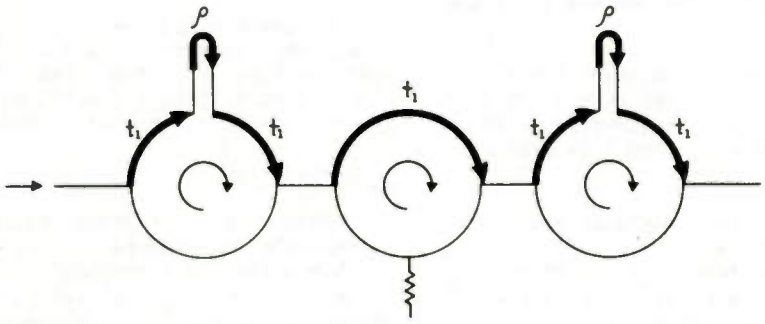


Fig. 2. Forward transmission path.

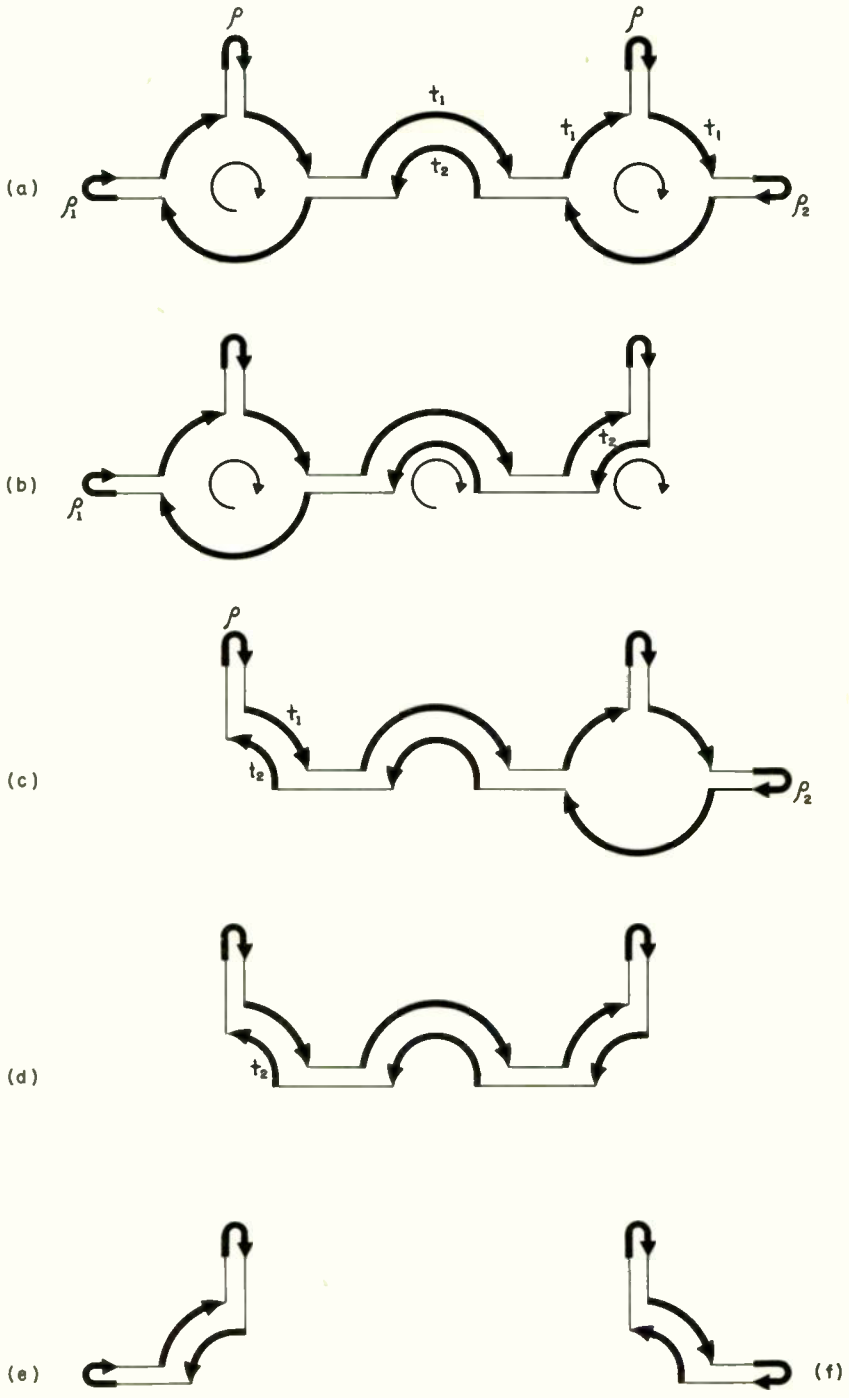


Fig. 3. Feedback paths.

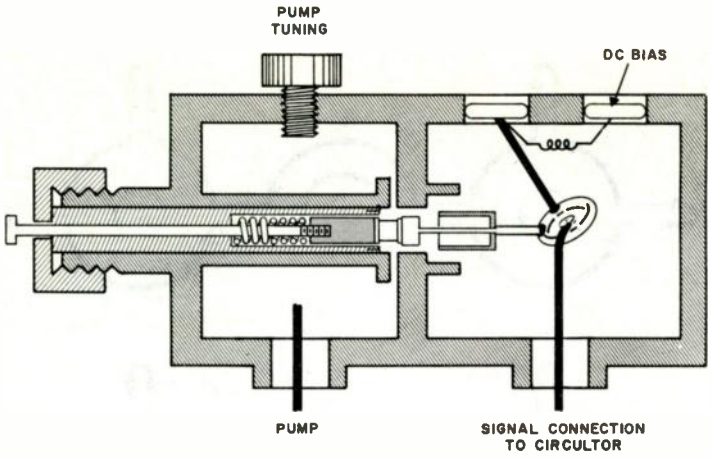


Fig. 4. Cross section of one-port.

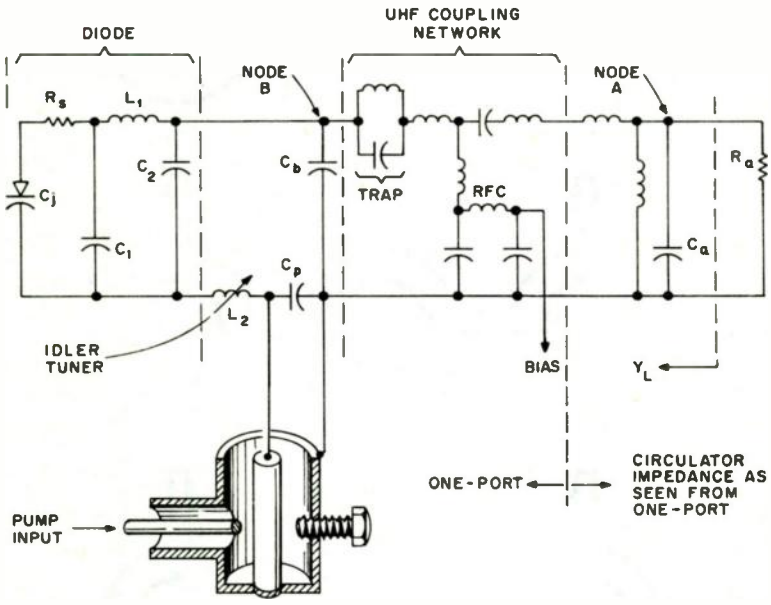


Fig. 5. Equivalent circuit of one-port and circulator.

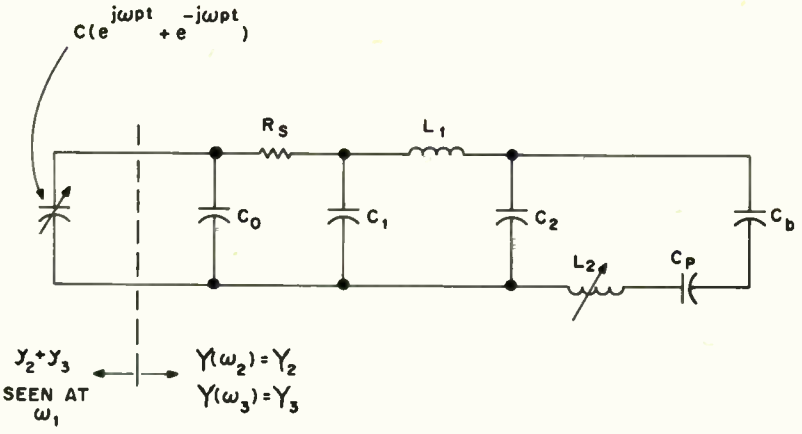


Fig. 6. Equivalent circuit for computation of negative conductance.

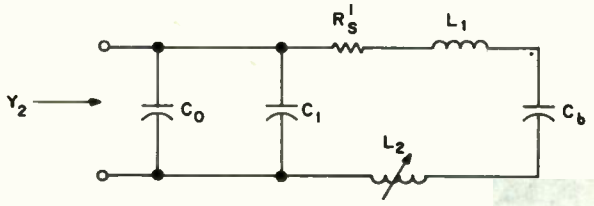


Fig. 7. Approximate idler equivalent circuit.

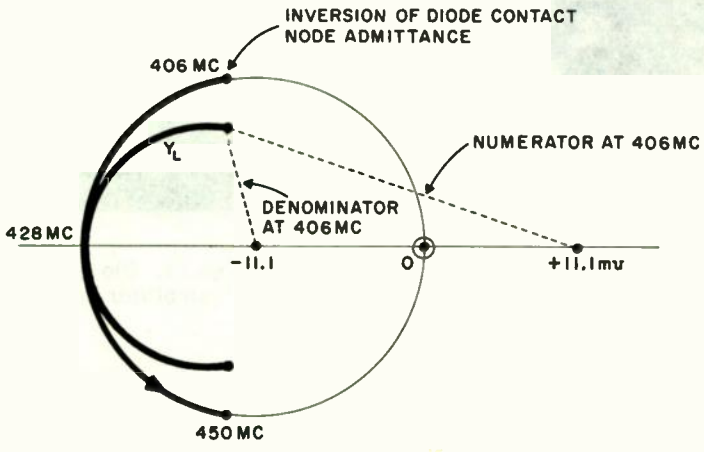


Fig. 8. Effect of circulator on admittance locus.

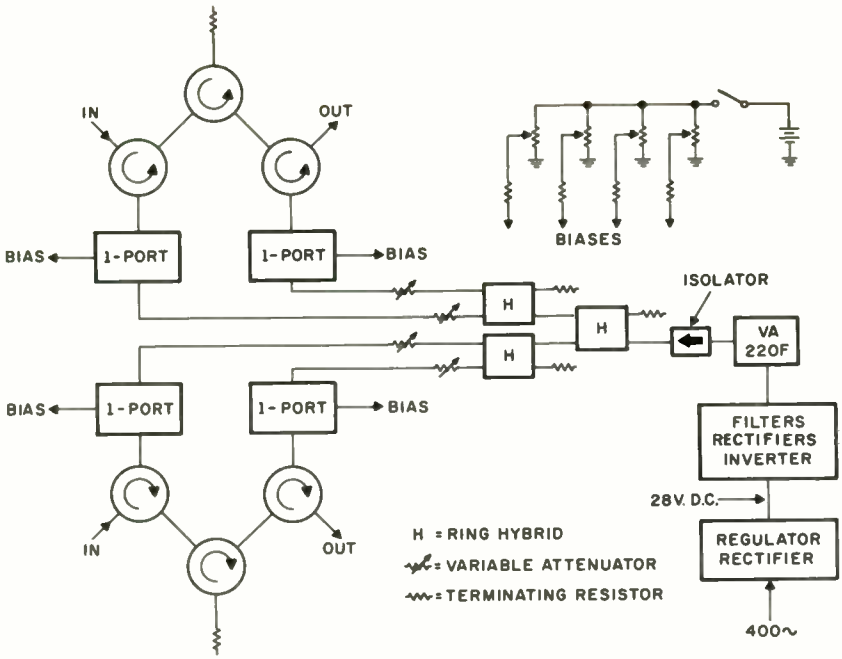


Fig. 9. Block diagram of complete two-channel amplifier.

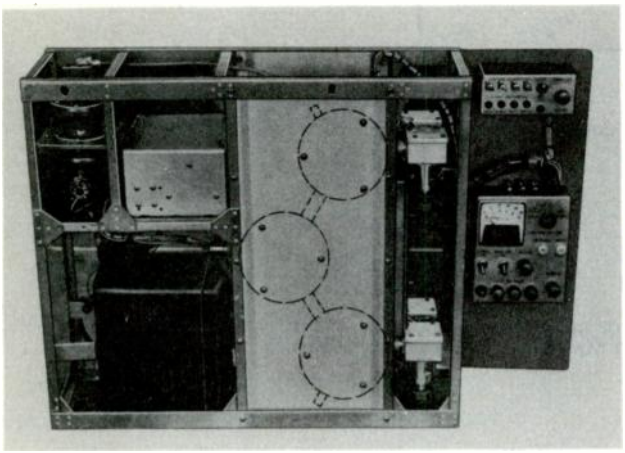


Fig. 10. Side view of dual-channel parametric amplifier.

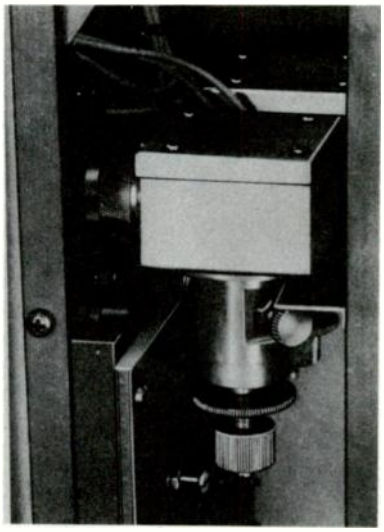


Fig. 11. Close-up view of parametric amplifier one-port.

SUPERCONDUCTING COAXIAL DELAY LINE

Peter K. Shizume and Eo Vaher
Air Armament Division
Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, L.I., New York

A superconducting delay line is described having a diameter of 0.036 inch. The loss of this line was measured to be 2.5 db per microsecond at 9 kmc and 3.5° K. Variation of line loss with frequency was found to be proportional to ω^2, so that for operation at 3 kmc a loss of 0.3 db per microsecond is expected. The line is capable of operation over the frequency from 0-10 kmc, and delay-line volumes of 14 cubic inches per microsecond are shown to be possible. A method for measuring the very low losses in the line using a three-inch sample of the line is given, and measured properties of some dielectrics and superconductors are presented for temperatures down to 3°K.

Introduction

The ability to store microwave signals for periods in excess of one microsecond has always been a problem to microwave engineers. Short delays of a few tenths microsecond have been attained using low-loss transmission lines; however, the associated loss and large volume of line required have limited the usefulness of such delays. More frequently, the technique has been to convert from microwave frequencies to megacycle frequencies where delay is accomplished in lumped constant or ultrasonic delay lines. These methods, although capable of giving delays of milliseconds in duration, suffer from limitations of narrow bandwidth and large conversion losses.

Recent developments in the generation of microwave phonons in quartz show promise of extending the frequency range of ultrasonic delay lines to the gigacycle range. This is made possible by cooling the quartz to temperatures below 20° K, where substantial reduction in attenuation of sonic waves is realized. The early works of H.E. Bommel and K. Dransfeld[1] and E.H. Jacobsen[2] on such an ultrasonic cryogenic line have demonstrated the feasibility of obtaining delays of several milliseconds at frequencies as high as 10 kmc. Factors that presently limit the usefulness of this technique are (1) the large transducer loss in the conversion from the electromagnetic to the ultrasonic mode of propagation and (2) the inherently narrow-band transducer techniques presently available. Conversion losses using cavity-type transducers used with this line have been reported to be of the order of 60 to 70 db, with bandwidths of only a few megacycles.

This paper describes the work done on another type of cryogenic delay line. This line uses the low-temperature, superconductive properties of metals to enable the realization of low-loss delays. The feasibility of obtaining (1) modest delays of the order of 10 microseconds with losses as low as a few tenths of a db per microsecond, (2) bandwidths of several octaves, and (3) small volume has been shown with this type of line.

The lack of practical engineering data on low-temperature properties of materials required this investigation to be primarily experimental. A coaxial configuration for the delay line was chosen, and the loss of various coaxial line configurations were measured using a resonant cavity technique. In this technique, measurements were made on three-inch sample sections of the line to determine the loss of the line. These measurements were also used to obtain the surface resistance and loss tangent of the materials composing the lines. The surface resistance of lead and niobium and the dielectric constant of Teflon and polystyrene were measured down to temperatures of 3° Kelvin. The frequency variation of the superconductor surface resistance was also measured and found to agree with elementary theory. A one-half microsecond length of line was later constructed and its loss was found to agree with the results as given by the resonant cavity method.

Low-Temperature Losses

The losses expected in materials at low temperature are of particular concern to this investigation. These losses will determine the ultimate reduction of size and attenuation of the delay line. Possible sources of loss are conduction loss resulting from finite resistance in the metal walls of the line, and displacement loss in the dielectric medium of the line. The conductor loss mechanism is significantly modified at low temperature from that at room temperature. There are two conduction processes peculiar to low temperatures which will determine the metal losses. One is the anomalous-skin effect, which occurs for all metals,

and the other is superconductivity, which is peculiar only to certain metals.

At temperatures below 30° Kelvin and frequencies in the X-band, high-conductivity metals exhibit losses that cannot be accounted for by the simple classical skin-effect relation. At these temperatures, the conductivity of the metal increases to the point where the skin depth, as determined by the classical description, becomes comparable to the increasing mean-free path of the electrons in the metal. When this condition prevails, the current in the metal varies from that predicted by Ohm's law. A kinetic theory of this effect has been developed by Reuter and Sondheimer[3] and is called the anomalous-skin effect. The result of their theory has shown that the anomalous behavior of metals at these low temperatures can be accounted for by the assumption of an effective conductivity proportional to $\omega^{3/2}$. This anomalous behavior is common to all good conductors at low temperatures.

The metals of primary interest to this investigation are those known as superconductors. A number of metals, alloys, and compounds have been found displaying this phenomenon, among them being lead, tin, and niobium - the metals examined in this work. The term superconductivity refers to the infinite d-c conductivity that these metals display at temperatures below their transition. According to the classical description, the a-c surface resistance is zero for a metal having infinite d-c conductivity. In a superconductor, however, in spite of the infinite d-c conductivity, a finite a-c resistance exists. This resistance is frequency dependent. For D.C., the superconductor exhibits zero resistance; however, as the frequency is increased, the resistance also increases, and, at microwave frequencies, a significant resistance can be detected. At infrared frequencies, the surface resistance of the superconductor increases to the point where it becomes equal to that of the metal in its normal conducting state.

London's macroscopic theory of superconductivity[4] gives a qualitatively accurate description of the conduction phenomena that takes place. In this theory, a two-fluid description of conduction is assumed. The currents in the superconductor are divided into normal current, j_n, and supercurrent, j_s. These are related to the total current j, by:

$$\bar{j} = \bar{j}_n + \bar{j}_s$$

$$= \sigma \bar{E} + \bar{j}_s$$

where σ is the d-c conductivity and $\bar{E}$ is the electric-field intensity within the metal. The supercarrier population is assumed zero at the transition temperature. As the temperature is lowered, the supercarrier population increases as more of the normal carriers are transformed into the super state.

An electric field impressed across the superconductor is assumed to accelerate the supercarriers in the same way that a force accelerates a body. Thus we have the equivalent of Newton's equation of motion,

$$\Lambda \frac{\partial}{\partial t} \bar{j}_s = \bar{E} \qquad (2)$$

where Λ is a constant characteristic of the superconductor and is, in general, a function of temperature. If this expression for $\bar{E}$ is substituted into Maxwell's equations, we have:

$$\nabla \times \left[\Lambda \frac{\partial}{\partial t} \bar{j}_s \right] = - \dot{\bar{B}} \qquad (3)$$

If the constants of integration are assumed zero, we obtain:

$$\nabla \times (\Lambda \bar{j}_s) = - \bar{B} \qquad (4)$$

Equations (1), (2), and (4) are the basis of London's macroscopic theory. Together with Maxwell's equations, these are sufficient to describe, in a qualitative manner, the electromagnetic behavior of superconductors. Using these equations and applying the results from Reuter and Sondheimer's theory to account for the behavior of the normal current, it can be shown[4] that the surface resistance will vary as ω^2 for frequencies where $\sigma \Lambda \omega << 1$. This condition prevails for superconductors at microwave frequencies.

The dielectric loss observed in nonpolar materials at room temperatures is generally due to polar impurities in the material. The polarizability of these ionic impurities are temperature dependent, and, at liquid-helium temperatures, they are expected to be largely " frozen in," so their contribution to the dielectric loss will be reduced. Therefore, substantial reductions in the loss tangent of certain dielectrics are to be expected at low temperatures.

Experimental Setup

Very little data is available on the properties of materials at liquid-helium temperatures. It was necessary, therefore, to generate this information by direct measurements in liquid helium. An exhaustive study to determine the optimum materials for the line was not intended. Instead, the investigation was limited to a few specific materials which were used in making up small sample-line sections for evaluation at liquid-helium temperatures.

Since the purpose of this investigation was to establish the degree of miniaturization possible, primary interest was focused on the measurement of line loss, since this ultimately controls the minimum practical line size. The measurement methods adopted were, therefore, based on the ease and accuracy with which this parameter could be measured, and, as a consequence, they did not lend themselves to precise evaluation of the individual properties of the materials composing the line. However, useful, if not precise, values for the low temperature properties of materials used in the line were obtained from these measurements.

The transmission-line configuration chosen for this investigation was the coaxial line. This choice was made on the basis of low dispersion, low loss per unit delay, compact size, broadband characteristics, and simple, rugged construction. The procedure used to evaluate the line was to construct short, three-inch samples of the line to be tested. These line segments were trimmed and capacitively coupled at both ends to form a coaxial cavity resonant at the desired measurement frequency. Figure 1 shows the configuration of the fixture used to contain the line sample, and Fig. 2 shows the measurement setup and the manner in which the experiment was contained for measurement in liquid helium.

The cavity-coupling coefficient could not be measured with the required accuracy because of thermal contraction of the feed lines and cavity parts that changed the coupling when the experiment was cooled to liquid-helium temperatures. Thus the cavity was loosely coupled, so the approximation $Q_L = Q_o$ could be assumed. In practice, the coupling was reduced to the point where a transmission loss in excess of 50 db was measured at resonance with the cavity in the superconducting state. For such loose coupling, Q_L could be assumed equal to Q_o with an error of less than one percent.

Initially, the measurements were made directly in the liquid helium; however, it was found that the presence of liquid helium in the line and coupling space resulted in modulation of the cavity resonant frequency. Apparently, this was due to the boiling of the liquid helium in and around the cavity. It was also noted that the cavity Q, measured while immersed in liquid helium, was consistently lower than the value when the cavity was lifted above the bath, which suggests that liquid helium has a measureable dielectric loss. The sample line was, therefore, isolated from the liquid by the chamber shown in Fig. 2. A copper compression gasket was used to make a liquid-helium-tight seal. The chamber was first evacuated, then filled with pure helium to a pressure of about 1 mm of Hg (when the experiment was cooled to 4.2° Kelvin) to allow for better thermal contact with the bath. Stable measurements of the cavity Q could be made with this arrangement. Typical cavity Q's of 10^6 were encountered in these measurements. The cavity Q was measured by the decay-decrement method.

The metals used in construction of the line samples were niobium and lead or lead-tin alloys. The dielectric used was Teflon. The metals were chosen on the basis of their high-transition temperatures, and Teflon was chosen because of its flexibility and ease in fabrication. Niobium was used for the center conductor, and the lead or lead-tin alloys were used for the outer conductor.

To measure the loss contributed by the various materials of the line, measurements were made using the specially constructed, large (0.028-inch diameter) coaxial cavity shown in Fig. 1. The dielectric loss was separated from the metal losses by measuring the cavity Q first when it was filled with Teflon, and again when the center conductor was bead-supported. Three thin beads were used for this purpose and they were positioned at the electric-field node positions in the cavity to minimize their effect.

The loss contributed by the niobium center conductor was separated from the lead outer conductor by measuring the loss of the cavity when a 0.062-inch diameter center conductor was used, and again when a 0.010-inch diameter center conductor was used. To use this method for separating the loss due to the inner conductor from those of the outer conductor, it must be assumed that the surface conditions of the different-diameter niobium center conductors are the same. In practice, this is not likely to be the case; however, the surface resistance values so obtained will be indicative of the average metal resistance.

Dielectric constants and loss tangents of some other dielectrics were measured in the cavity shown in Fig. 3. This cavity is an X-band waveguide cavity. It was formed by pressing from a solid billet of lead. The material to be

measured was machined to fit into the cavity, and the dielectric constant and loss tangent were determined from measurements of the resonant frequency and Q of the cavity.

Experimental Results

Figure 4 shows the equivalent loss per microsecond of a 0.036-inch diameter coaxial line. The center conductor of this line was a 0.010-inch diameter niobium wire; the dielectric was Teflon; and the outer conductor was 60-40 lead tin. The loss drops smoothly from 8 db per microsecond at 4.5° K to 1.9 db at 3.2° Kelvin.

In this same figure is shown the loss measured for a four-foot length of the same cable, showing the effect of a splice in the outer conductor of the line. The splice was made by overlapping one outer conductor and swaging it snugly down over the other. Fusion of the metal was obtained at the outer surface. These results indicate a significant increase in loss from 5 to 12 db per microsecond due to the electrical discontinuity at the splice. The volume of a 10-microsecond length of this line having an outer conductor wall thickness of 0.005 inches is about 140 cubic inches. This volume is capable of being contained in liquid-helium cryostate of the type commonly used in low-temperature work.

Measurements of the loss properties of lead, niobium, and Teflon were made using the 0.028-inch diameter cavity shown in Fig. 1. The lead outer jacket of the cavity was formed by casting lead in the form of a cylinder and boring a hole through the axis to a diameter of 0.028 inch. Figure 5 shows the variation in cavity Q for the case of a 0.062-inch diameter niobium center conductor and a solid Teflon dielectric. In Fig. 6 the same lead outer conductor is used as in Fig. 5, but the center conductors are supported by three Teflon beads placed at the electric-field node positions. Curve A of this figure shows the variation of Q when a 0.062-inch diameter center conductor is used, and curve B shows the variation when a 0.010-inch diameter center conductor is used. It will be noticed from these figures that the lead outer conductor becomes superconducting at 7.5°K. The normal transition temperature for pure lead is 7.22°K, and this increase of 0.28°K indicates a high degree of impurity in the lead used for the outer conductor.

The line with the 0.010-inch diameter center conductor is noticed to have a lower loss than the line with the 0.062-inch diameter center conductor at temperatures just above 7.5°K and below 4°K. This behavior can be explained on the basis of the minimum attenuation configuration of the line. At temperatures above 7.5°K, the niobium center conductor is superconducting (transition temperature of niobium, 8.9°K), whereas the outer conductor is still in the normal state. The ratio of surface resistance for the outer and center conductors at this temperature is therefore as much as 50:1. For such large differences in the resistance of the two conductors, the optimum ratio of outer to center conductor diameter for minimum loss is about 25. (The conventional outer to inner conductor ratio of 3.6 assumes equal resistance for the inner and outer conductors.) The line with the 0.010-inch diameter center conductor, with a ratio of outer to center conductor diameter of 28, approximates this minimum-loss configuration more closely than the line with the 0.062-inch diameter center, and explains why the loss of this line is lower at this temperature. As the temperature is reduced below the transition of the outer conductor, the resistance ratio of outer to inner conductor is reduced to near unity as the outer conductor becomes superconducting and the line with the 0.062-inch center becomes the lower loss configuration. At lower temperatures, however, the ratio of resistance of the outer to inner conductor begins to increase, due to the high residual resistance of the impure lead outer conductor, until below 4°K the loss of the 0.010-inch center conductor line once again falls below the 0.062-inch diameter center conductor line.

The results of measurements on the three line configurations discussed above were used to separate the dielectric and metal losses of the line. Figure 7 shows the temperature variation of the resistance of lead at 9 kmc. Results for two samples are shown in comparison with data obtained by Simon[5] on vacuum-cast, high-purity lead specimens. One sample is seen to approximate quite well the slope of Simon's result, whereas the other varies considerably from this. This later behavior is characteristic of rough or impure surfaces.

Figure 8 shows the temperature variation of the surface resistance of niobium. The results on two separate sample pairs are seen to agree in regard to slope; however, one sample pair is noticeably more lossy than the other. Variations in the surface resistance of niobium, however, were found to be smaller than in lead, and in part this is probably due to the fact that niobium is more resistant to the formation of oxides and other corrosion products than lead, so that day-to-day or sample-to-sample measurements are more consistent.

The loss tangent of three different samples of Teflon was measured using the coaxial cavity. The results of these measurements are tabulated in Table I. The measurements indicate no measurable change in the loss tangent over the

temperature range specified. Other measurements of dielectrics made using the waveguide cavity are also tabulated. These measurements show a definite temperature variation of the loss tangent. It will be noted also that a noticeably greater loss tangent was measured for Teflon in these measurements as compared with those obtained from the coaxial cavity. This difference might be due to the differences in the purity of the Teflon used in the waveguide cavity compared to that used in the coaxial cavity, since the samples were not cut from the same stock. The differences in the temperature variation of the loss tangent is not understood. The polystyrenes are seen to be considerably more lossy than Teflon, and here again a definite temperature variation is indicated.

The dielectric constant and loss tangent of liquid helium was measured. Due to technical difficulties, only one measurement at 4.2°K was possible. This measurement indicates that liquid helium has a small but significant loss at 9 kmc. This loss, however, might be due to condensed particles of other gases which are often noted to be present in liquid helium. The measured a-c dielectric constant of helium compares favorably with the results from Squire[6] for d-c measurements.

Figure 9 shows the frequency variation of the superconductor resistance. This result was obtained using the coaxial cavity by measuring the cavity Q at two successive resonant points. An approximate $\omega^{2.1}$ dependence of the surface resistance was found. The $R(\omega^2)$ curve predicted from theory is shown for comparison.

Conclusion

These experiments have shown the feasibility of using a superconductive coaxial delay line for obtaining delays of up to 10 microseconds. Volumes for such a line are approximately 14 cubic inches per microsecond of delay, and the maximum losses are 2.5 db per microsecond below 9 kmc and 3.5°K. Variation of line loss with frequency has been shown to approximate the ω^2 dependence predicted from elementary theory, and from this dependence, the expected attenuation at 3 kmc is only 0.3 db per microseconds. Bandwidths of several octaves are possible because of the inherent broadband character of the line. The ultimate bandwidth of the line is limited only by the increasing losses that result at higher frequencies. Some further improvement in insertion loss might be possible using other materials; however, it is not expected that these improvements will be substantial.

Because of the expense in maintaining large volumes at liquid-helium temperatures, the usefulness of this technique is limited to maximum delays of approximately 10 microseconds. Maximum operating frequencies are not expected to rise much higher than the X-band, because of the high losses at the greater frequencies. For obtaining delays up to 10 microseconds, the superconducting delay line has significent advantages in operating bandwidth and insertion loss over presently available techniques. A maximum insertion loss of about 30 db and useful bandwidths from 0 to 10 kmc are expected in such a line. The volume of the line (about 140 cubic inches for ten microseconds) makes it possible to contain the line in cryostats only somewhat larger than those presently being used in maser applications.

Acknowledgement

The author's wish to acknowledge the assistance and direction given by D. Kenyon, D. Shapiro, and the valuable laboratory assistance of W. Gerish, who was responsible for making many of the special apparatus used in this work.

References

1. H.E. Bommel and K. Dransfeld, Phys. Rev. LETT., 2, 298(1959).

2. E.H. Jacobsen, Phys. Rev. LETT.,2, 249 (1959).

3. E.H. Reuter and A.E.H. Sondheimer, Nature, 161, 394 (1948); Proc. Roy. Soc., London, A 195, 336 (1948).

4. F. London, Superfluids, Wiley and Son, Inc., New York (1950).

5. I. Simon, M.I.T. Res. Lab. of Electronics Tech. Rpt. 126 (1949).

6. C.F. Squire. Low-Temperature Physics, McGraw-Hill Book Co., New York (1953).

TABLE I

Material	Frequency (kmc)	T(°K)	ϵ_r	tan δ
Teflon (coaxial cavity)	9.0	2.8-4.2	~2	1.33×10^{-6}
		3-4.5		1.15×10^{-6}
		3-6.5		$.75 \times 10^{-6}$
Teflon (waveguide cavity)	8.609	4.2	2.03	3.03×10^{-6}
		3.1	2.03	2.33×10^{-6}
Cross-linked Styrene	8.842	4.2	1.910	2.94×10^{-5}
		2.5	1.910	1.94×10^{-5}
Polystyrene	8.846	4.2	1.907	4.31×10^{-5}
		2.5		3.12×10^{-5}
Liquid Helium (Squire)	9.004	4.2	1.05	$\sim 10^{-6}$
	D.C.	3.5	1.05	

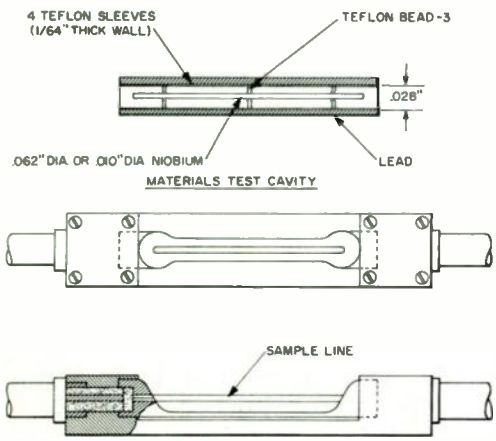


FIG. 1 COAXIAL-CAVITY CONFIGURATIONS

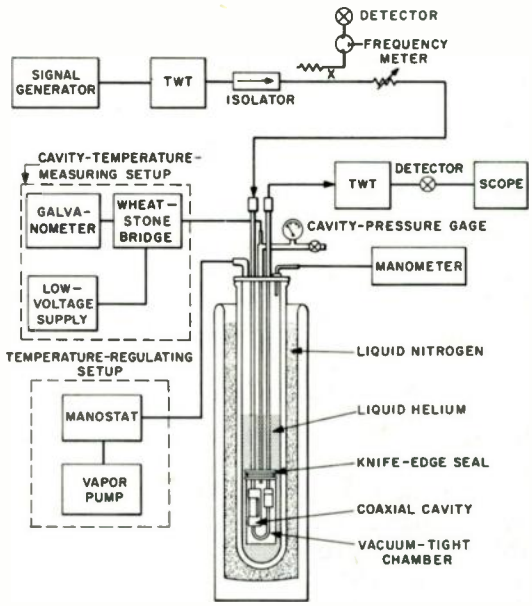


FIG. 2 COAXIAL-CAVITY SETUP

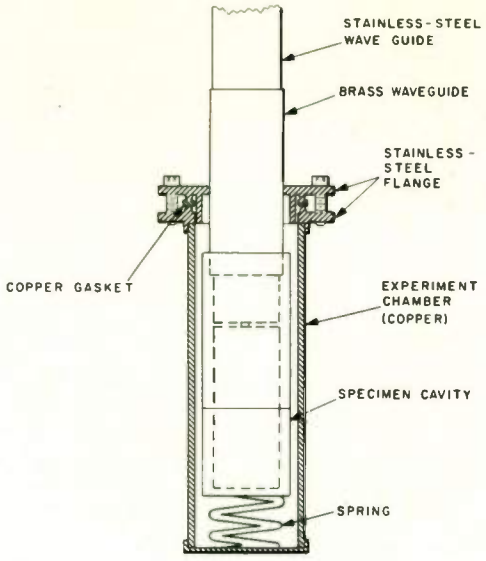


FIG. 3 DIELECTRIC-MEASUREMENT CAVITY

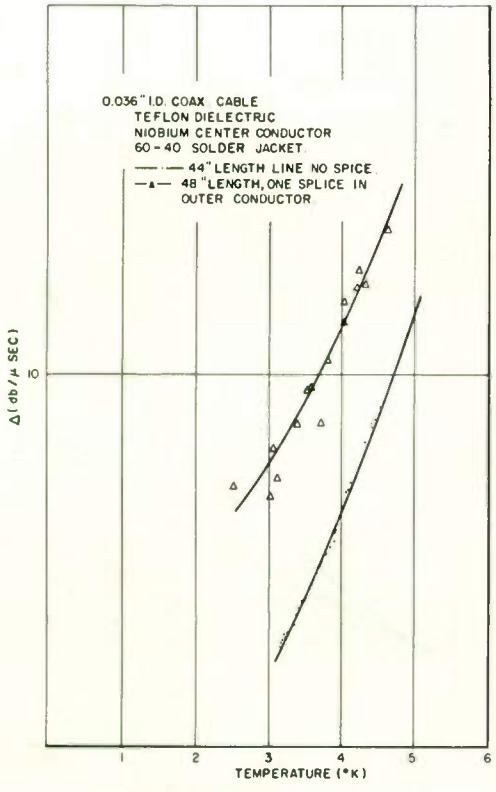


FIG. 4 LOSS VS. TEMPERATURE AT 9 KMC

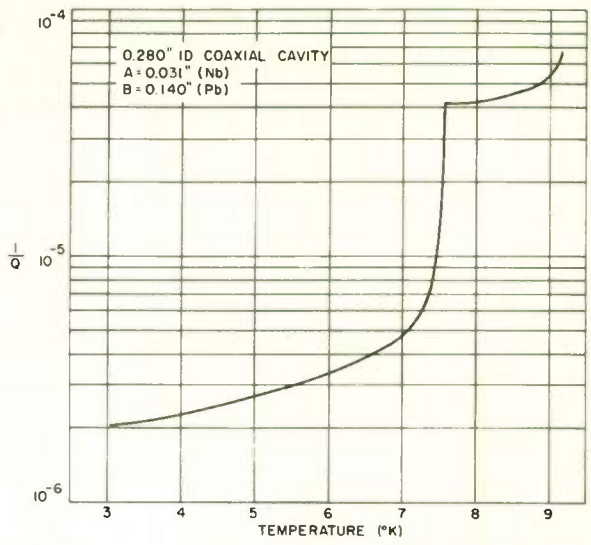


FIG. 5 CAVITY Q VS. TEMPERATURE WITH TEFLON DIELECTRIC

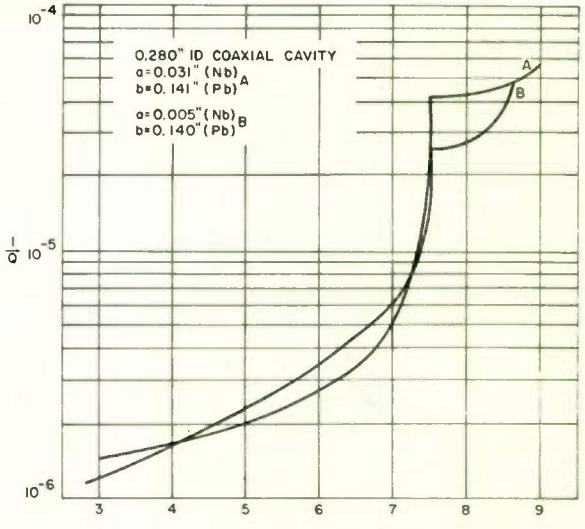


FIG. 6 CAVITY Q VS. TEMPERATURE WITH THREE-BEAD SUPPORT

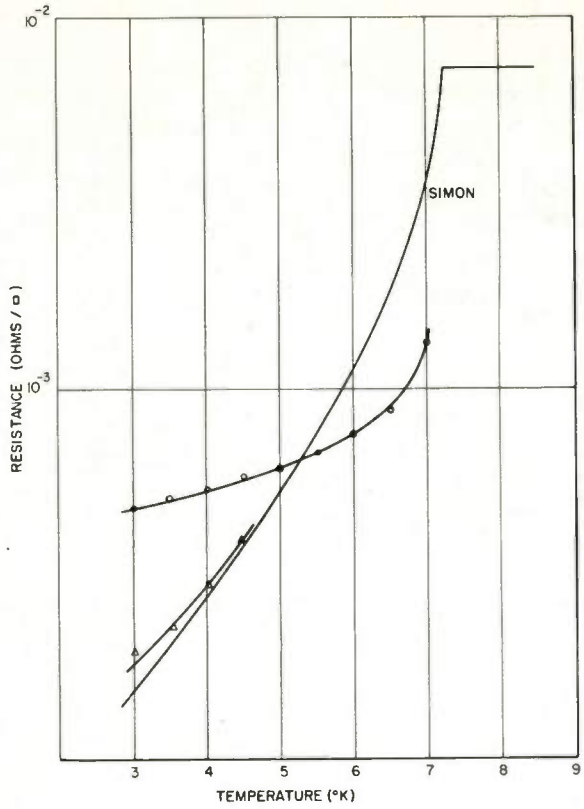


FIG. 7 SURFACE RESISTANCE OF LEAD VS. TEMPERATURE AT 9 KMC

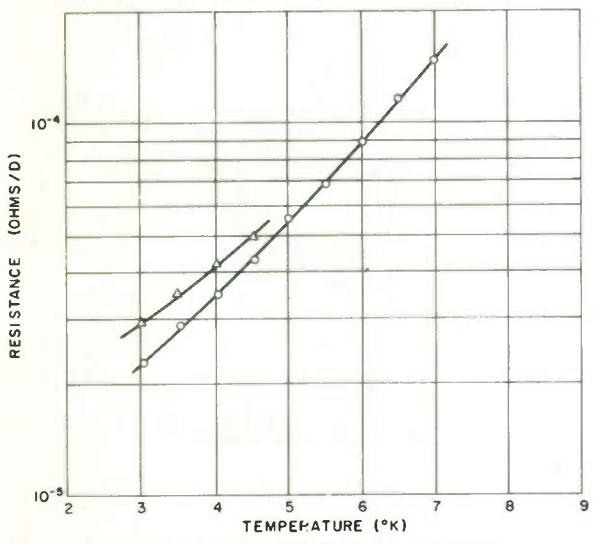


FIG. 8 SURFACE RESISTANCE OF NIOBIUM VS. TEMPERATURE AT 9KMC

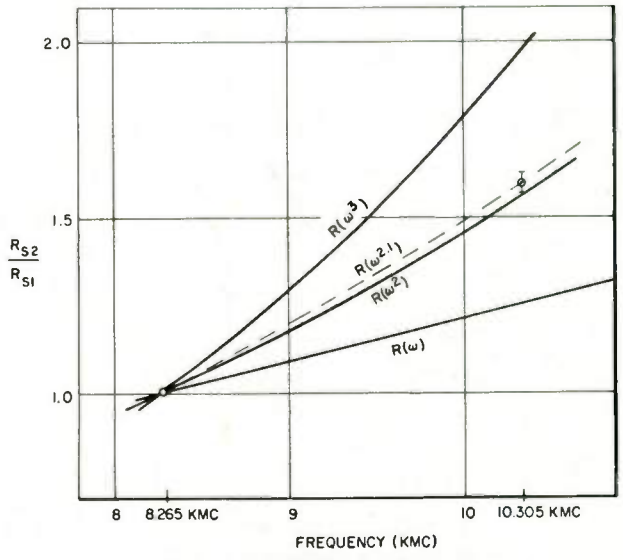


FIG. 9 FREQUENCY VARIATION OF SURFACE RESISTANCE

A WIDE-BAND MICROWAVE COMPRESSIVE RECEIVER

F. J. Mueller and R. L. Goodwin

Airborne Instruments Laboratory
A Division of Cutler-Hammer, Inc.
Deer Park, Long Island, New York

Summary

The compressive receiver, developed by W. D. White of AIL and others, is useful when both high-frequency resolution and a fast scan rate are desired. The receiver uses a network "matched" to an internally-generated FM wave of large bandwidth-duration derived from the external signal.

A practical receiver is described in which the network was realized as a tapped-delay-line filter (theoretical compression factor of 100). The scanning local oscillator was a voltage-tunable magnetron. The receiver scanned a selected 800-Mc-wide RF band at a 100 kc/sec rate, with a frequency resolution of 2 to 3 Mc. The compression factor obtained varied between 50 and 70.

I. Introduction

The compressive receiver, developed by White (reference 1) and others, is a new type of scanning superheterodyne designed especially for reconnaissance applications such as the interception of pulsed radar signals. By comparison, a conventional panoramic receiver is limited in its ability to scan rapidly with high sensitivity and good frequency resolution--a limitation imposed by its maximum scan rate, which is approximately equal to the square of the IF bandwidth (reference 2). Thus, a conventional panoramic receiver with a 10-Mc bandwidth is limited to a scan rate of about 100 Mc/μsec to obtain a frequency resolution of 10 Mc.

The compressive receiver uses a pulse-compression network in the IF section to time-compress the output of the IF amplifier. Since the time of occurrence of the output is proportional to its RF frequency (because of the receiver's scanning action), time compression of the IF amplifiers output is equivalent to an increase in frequency resolution. The increase is proportional to the square root of the compression factor of the compression network. Thus, the compressive receiver, scanning at 100 Mc/μsec and having a compression factor of 100, has a frequency resolution of 1 Mc rather than the 10 Mc possible with a conventional receiver with the same scan rate. It has the sensitivity of a 1-Mc receiver (rather than that of a 10-Mc receiver), and the image response is suppressed by a factor that is twice the compression factor; the image response is suppressed 23 db in a receiver with a compression factor of 100.

This paper describes the development of a wide-band compressive receiver that scans an 800-Mc RF band in 10 μsec. Major emphasis is given to the design and performance of the scanning local oscillator and the compression network.

II. Overall Receiver Description

The receiver is a double-conversion superheterodyne: the intercepted signals are converted to S-band by the first converter and then down-converted by a scanning second converter. Figure 1 is a block diagram of the compressive receiver. The first converter is single-ended and converts one of three, selected, 800-Mc-wide RF bands into the first IF band of 2,600 to 3,400 Mc. The first IF amplifier is a low-noise, S-band, traveling-wave-tube (TWT) amplifier. The local oscillator power for the second converter is obtained from a voltage-tunable magnetron (VTM) that is swept electronically from 2,900 to 3,800 Mc in 9 μsec with a retrace time of 1 μsec. Thus, the scanning-local-oscillator sweep rate is 100 Mc/μsec, and the sweep repetition rate is 100 kc.

The output of the second converter is a signal whose frequency varies with time (Figure 2). During the sweep period, the converter output frequency sweeps through the second IF band (300 to 400 Mc) at different times, depending on the frequency of the RF intercept. Since the slope frequency versus time of the sweeping signal is 100 Mc/μsec, and the second IF bandwidth is 100 Mc, a 1-μsec FM pulse is developed in the second IF system. This pulsed FM signal is amplified by two UHF TWT amplifiers, one preceding and one following the compression network.

The compression network (CN) in the second IF section reduces the 1-μsec FM pulse width by a factor of 100: thus, the output pulse width from the CN is 10 nsec (nanoseconds). The CN is a tapped-delay-line with a band-pass filter on each tap. The delay line consists of about 640 feet of coaxial cable and has twenty taps and twenty band-pass filters. The tapping and recombining function is performed by directional couplers.

Attenuators equalize the transmission losses through each tapping point to the output, and the coupler coefficients are varied from tap to tap. A shaping, or weighting, filter is also included in the second IF section; it shapes the frequency spectrum of the uncompressed pulse to provide

side-lobe reduction in the compressed pulse. A crystal detector and several wide-band (50-Mc) video amplifiers complete the receiver.

III. Scanning Local Oscillator

A. Introduction

The local oscillator (LO) in the scanning second converter must sweep 2,900 to 3,800 Mc in 9 μsec and return to 2,900 Mc in 1 μsec. Thus, the frequency sweep rate is 100 Mc/μsec, and the selected RF band is scanned at 100 kc/sec. The LO output power should be sufficient to drive a mixer, and the power variation across this band should be small. The prime requirement of this component (and of its associated sweep circuits) is a tuning rate with little deviation from the desired 100 Mc/μsec over the band, because, excessive deviations will result in poor pulse compression and low S/N ratio.

Two types of electronically tunable oscillators were considered: (1) backward-wave oscillator (BWO) and (2) voltage-tunable magnetron (VTM). Both types can be rapidly tuned at extremely high rates over bands approaching an octave and both can be obtained with permanent-magnet focusing structures, which minimize supply drain. The VTM was chosen, however, because its linear tuning curve permitted the use of a simpler sweep circuit. The General Electric Type GL-7398 VTM, which covers the range 2,200 to 3,850 Mc with an average power output of 5 watts, was used. The tuning sensitivity is about 2.6 Mc per volt. The VTM output frequency is determined primarily by its anode-to-cathode voltage but is a secondary function of anode current, load VSWR, and heater-voltage variations. Power output can be varied by changing the injection-electrode voltage (analogous to changing the voltage of the control grid in a conventional electron tube).

To determine the optimum operating conditions, tuning-voltage versus frequency was measured at 50-Mc increments, with injection-electrode voltage as a parameter. The results of these measurements showed that an injection-electrode voltage of +360 v, with respect to the cathode, gave the best tuning linearity over the desired bandwidth. Figure 3 shows the tuning curve of VTM Serial No. 1082-S and the corresponding slope-linearity curve, based on a "least squares," best straight-tuning voltage.

A Miller sweep circuit drives the VTM cathode with a negative-going waveform that drops from +142 v to -206 v in 9 μsec. The anode of the VTM was maintained constantly at about 1,400 v. To minimize incidental FM, a DC supply was used to energize the heater of the VTM.

B. Linearity Measurement Procedure

Final adjustment of the VTM sweep circuit, and subsequent measurement of the frequency-time linearity of the scanning LO, was obtained by using the test circuit shown in Figure 4. A time-mark generator (Tektronix 181) generated sync pulses at the 100 kc/sec system scan rate. Since these "time marks" were not of sufficient amplitude to synchronize the VTM sweep circuit, an additional pulse generator (DuMont 404) was used for this purpose.

The trigger pulse from the VTM sweep generator circuit was applied to the delay trigger (Time Base B) of the oscilloscope (Tektronix 545A). By means of the delay-time multiplier (Time Base A set to 0.1 μsec/cm), any 1-μsec portion of the VTM sweep can be examined.

One input to the dual-trace oscilloscope preamplifier is a 10-Mc sine wave from the time-mark generator; the other input is from the detector and frequency meter on the VTM RF output line. The detector is loaded with a 100-ohm resistor to decrease the width of the frequency-meter "pip" (increased video bandwidth).

To measure the scanning-LO linearity, the following procedure was used.

1. The delay-time multiplier was adjusted so that the first 1-μsec interval of the detected VTM output appeared on trace A. Ten sine waves appeared on trace B.
2. The frequencies corresponding to "pip" locations directly over successive positive peaks of the 10-Mc sine wave were noted.
3. The delay-time multiplier was adjusted to bring the next 1-μsec interval into position, and steps 1 and 2 were repeated.
4. This procedure was repeated until the whole 9-μsec sweep interval was examined.

The scanning-LO linearity-measurement procedure outlined above removes the effect of scope time-base errors and ensures that the actual frequency-versus-time of the VTM RF output is being determined. It is believed that each frequency was measured with an accuracy of ±2 Mc.

C. Scanning-LO Linearity Versus Time

The scanning LO converts a CW signal into a linearly frequency-modulated signal of 100 Mc deviation in 1 μsec. If the frequency deviation in 1 μsec differs greatly from 100 Mc, the system-output compressed-pulse width will increase, and the peak amplitude of the compressed pulse will decrease. Therefore, a reasonable criterion of scanning-LO performance is the deviation from the ideal 100-Mc/μsec rate over the LO range.

Scanning-LO linearity is defined by the following equation:

$$\text{Linearity in percent (at } f_x,\ t_x) = \frac{f_x - f_x - 100 \text{ Mc}}{100 \text{ Mc}} \times 100 \text{ percent}$$

where

f'_x = frequency at 1.0 μsec after t_x, in Mc

f_x = frequency at t_x, in Mc

t_x = time from beginning of sweep, in μsec.

The linearity versus time of the scanning LO (using VTM GL-7398 Serial No. 1082-S) is shown in Figure 5. For 60 percent of the sweep time, the linearity was ±2 percent. For 81 percent of the sweep time, the linearity was ±5 percent. This means that the scanning LO swept 100 ± 5 Mc/μsec for 81 percent of the total sweep time.

IV. Compression Network

A. Introduction

The purpose of the compression network (CN) is to transform the linearly frequency-modulated pulse from the second converter into a higher-amplitude (higher-peak-power), smaller-time-duration output. The CN gives the compressive receiver the properties of high sensitivity, good resolution, and high probability of intercept, which are difficult to obtain simultaneously in conventional scanning receivers.

The selection of the center frequency of the CN (and also of the associated second-IF amplifiers) was influenced by several factors:

1. The IF center frequency should not be too high to accomplish compression; a tapped-delay-line type of CN suffers from high attenuation at high frequencies.
2. The IF center frequency should not be too low to obtain good video detection; having an IF band that is close to the video band will result in poor video detection.
3. The IF center frequency should be compatible with readily available amplifiers of sufficient bandwidth.

On the basis of these considerations, the 300 to 400 Mc band was chosen as the frequency range of the CN of the second IF band. Components for the CN, as well as TWT amplifiers optimized for this band, proved to be readily available.

B. Formation of Compressed Pulse

The CN has a time-delay-versus-frequency characteristic that is the opposite of the signal waveform (that is, it is one class of "matched" filter). This time delay is composed of two parts: (1) a fixed time delay, t_f (independent of frequency), and (2) a linearly varying time delay that varies from zero μsec at 400 Mc to 1 μsec at 300 Mc. When the 300-Mc component of the input FM pulse is applied to the CN, it does not emerge until 1.0 + t_f μsec have elapsed. One μsec later, however, the 400-Mc component of the input FM pulse is applied, and this component passes through the CN in just t_f μsec. Thus, the 300-Mc and 400-Mc components of the input signal (as well as all the other components) emerge at the output simultaneously and add in-phase to a high-peak amplitude.

Idealized input and output waveforms associated with the CN are shown in Figure 6. Theoretically, the output pulse width, measured at -4 db points, is the reciprocal of the input bandwidth (100 Mc) or 10 nsec. The compression factor, given by the product of input signal bandwidth (100 Mc) and input signal pulse duration (1 μsec) is 100. This is the factor by which the output peak power is increased over the input peak power (for a lossless CN). The input and output pulse durations, and the input and output (resolution) bandwidths, are related to the compression factor by the equation

$$CF = \frac{T_i}{T_o} = \frac{B_i}{B_o}$$

where

T_i = input pulse duration,

T_o = output pulse duration,

B_i = input bandwidth,

B_o = output bandwidth.

Side lobes will also appear in the output, preceding and lagging the main response, because of the phase interference between the various components of the input waveform. In a network in which no attempt has been made to minimize these side lobes, they are of sin t/t form; the first side lobe is only 13.2 db below the peak response. In a tapped-delay-line-filter network, however, there is no output until the input FM pulse completely "fills" the line (because of the band-pass filters in each tap), and the side-lobe level is futher reduced.

Shaping or weighting of the input spectrum to the CN also tends to suppress the side lobes (references 3 and 4). These side lobes are undesirable because they may obscure weak signals adjacent in frequency to a stronger signal. In addition, the use of weighting aids in minimizing compressed-pulse-width broadening caused by amplitude and phase deviations from the ideal in both the scanning LO and the CN.

C. Compression Network Design

1. Network Configuration

The CN is a tapped-delay-line, in the form of a ladder, with lengths of coaxial cable and directional couplers forming the sides of the ladder. The cables provide the major amount of time delay, and the directional couplers sample and also recombine the input signal. The rungs of the ladder are composed of band-pass filters with equalizing attenuators on their input and output terminals. Figure 7 is a simplified diagram of the CN.

The spacings between taps and the center frequency of the band-pass filters vary in a quadratic manner rather than in a linear manner. On the basis of computations made by W. White of AIL, the minimum number of taps required for this type of CN to achieve a moderately low side-lobe level is given by the square root of the compression ratio. As a safety factor, twice this minimum was actually used; that is, 20 taps, rather than 10. Figure 8, a detailed schematic of the CN, shows the arrangement of the delay cables, directional couplers, filters, and equalizing attenuators. Figure 9 is a view of the rear of one of the three units in which the CN is packaged.

2. Electrical Design

The phase of the input signal versus time, with zero phase at 300 Mc and at zero time, is given by:

$$\theta = 2\pi F_o t + \frac{\mu t^2}{2}$$

$$= 2\pi \times 300 \times 10^6 \times t + \frac{2\pi}{2} \times 100 \times 10^{12} \times t^2 \text{ radians} \quad (1)$$

where μ is $2\pi \times 100 \times 10^{12}$ radians/sec

Also,

$$f = \left(\frac{1}{2\pi}\right)\omega = \frac{1}{2\pi}\frac{d\theta}{dt}$$

$$= 300 \times 10^6 + 100 \times 10^{12} \times t \text{ cycles} \quad (2)$$

The total phase shift in the input signal after 1 μsec is:

$$\theta_T = 2\pi \times 300 \times 10^6 \times 10^{-6} + \frac{2\pi}{2} \times 100 \times 10^{12} \times (10^{-6})^2$$

$$= 2\pi\ (350) \text{ radians} \quad (3)$$

If we space the delay-line taps at equal wavelengths, then the tap spacing is:

$$\theta_{SP} = \frac{2\pi(350)}{20} = 17.5 \times 2\pi \text{ radians} \quad (4)$$

Relative to the 300-Mc end of the network, the tap spacing is:

First: $1/2 \times 17.5\ (2\pi)$ radians from end

Second: $\frac{2+1}{2} \times 17.5\ (2\pi)$ radians from end (5)

Nth: $\frac{2n-1}{2} \times 17.5\ (2\pi)$ radians from end

Equation 1 can be solved for t_n with the result:

$$t_n = 3 + \sqrt{9 + .175(2n-1)} \text{ usec}, \quad (6)$$

$$1 \leq n \leq 20$$

The t_n values thus obtained are the total delay at each tap relative to the 300-Mc end of the network. The differential tap spacing, or spacing between each tap point, is the difference between t_n and $t_n - 1$.

NOTE

Although the delay-line-tap spacing is calculated from the 300-Mc end of the network (that is, 300 Mc occurs first in the signal waveform), in operation, the signal enters the 400-Mc end of the network.

The center frequencies of the band-pass filters are obtained by substituting the t_n values from equation 6 into equation 2.

Thus,

$f_1 = 302.9$ Mc, $f_2 = 308.6$ Mc, $f_3 = 314.2$ Mc, etc.

The filters should have the following characteristics:

1. Low insertion loss
2. Approximately linear phase (constant time delay) over their bandwidth
3. Filter crossover points not too far below the center-frequency response to pass the maximum amount of energy in the input pulse to the output
4. Moderate, rather than sharp, selectivity to minimize filter "ringing."

The band-pass filters were two-pole Butterworth types having about 7-Mc bandwidth at -3 db points. Adjacent filter crossovers were -1.0 db relative to center-frequency response. Table I lists the filter center frequencies, their bandwidths, and the filter tapping-point delay relative to the 300-Mc end of the network. Figure 10 shows the insertion-loss-versus-frequency plot of the filters. Type RG-213/U coaxial cables were used for the delay cables.

The proper length of each cable was calculated in the following manner. First, the required differential time delay was converted to an equivalent length of "air dielectric" transmission line by the following relation:

$$t_d(\text{nsec}) \quad \frac{11.788 \text{ inches}}{1 \text{ nsec}} = \underset{(\text{inches})}{\text{electrical length}} \quad (7)$$

Table I

Compression Network Parameters

Filter Center Frequency (Mc)	Filter Bandwidth (Mc)	Differential Time Delay (μsec)	Cumulative Time Delay (μsec)
302.9	8.0	0.02903	0.02903
308.6	7.9	0.05723	0.08626
314.2	7.8	0.05619	0.14245
319.8	7.7	0.05521	0.19766
325.2	7.6	0.05426	0.25192
330.5	7.4	0.05338	0.30530
335.8	7.3	0.05253	0.35783
341.0	7.1	0.05172	0.40955
346.0	7.0	0.05094	0.46049
351.1	7.0	0.05021	0.51070
356.0	6.9	0.04950	0.56020
360.9	6.8	0.04882	0.60902
365.7	6.7	0.04817	0.65719
370.5	6.6	0.04754	0.70473
375.2	6.6	0.04694	0.75167
379.8	6.5	0.04636	0.79803
384.4	6.4	0.04580	0.84383
388.9	6.3	0.04526	0.88909
393.4	6.2	0.0474	0.93383
397.8	6.1	0.04423	0.97806

NOTES

1. Cumulative time delay is in reference to the 300-Mc end of the network.
2. $f_n = 300 + 100t_n$ f_n in Mc, t_n in μsec
3. $t_n = -3 + \sqrt{9 + 0.175(2n-1)}$ μsec

$$1 \leq n \leq 20$$

Second, this length was divided by two, because of the differential time delay is due to both "sides" of the "ladder" (Figure 7). Third, the electrical length of the directional couplers was subtracted from the cable length arrived at in step 2. Finally, the electrical length of the right-angle connectors on each end of the cable was subtracted from the cable length arrived at in step 3. This value, divided by the square root of the cable's dielectric constant, is the length to which each cable was cut. Table II summarizes these calculations.

Table II

Compression Network Delay Cable Lengths

t_d, Differential Between Tap Points (nsec)	Delay Cable Lengths (inches) Exclusive of RF Connectors
57.2	214.1
56.2	210.2
55.2	206.3
54.3	202.7
53.4	199.2
52.5	195.7
51.7	192.5
50.9	189.4
50.2	186.7
49.5	183.9
48.8	181.2
48.2	178.8
47.5	176.1
46.9	173.7
46.4	171.8
45.8	169.4
45.3	167.4
44.8	165.5
44.2	163.1

NOTES

1. There are two cables at 214.1, 210.2, 206.3 inches, etc.
2. Cable length =

$$\frac{\left[t_d \times \frac{11.788 \text{ inches}}{\text{nsec}} \times 1/2\right] - L_{dc} - L_{conn}}{\sqrt{e}}$$

where

t_d = Differential delay in nsec

L_{dc} = Electrical length of directional coupler, 12.17 inches

L_{conn} = Electrical length of two right-angle RF connectors, 3.15 inches

$\sqrt{e}$ = 1.503--square root of dielectric constant, RG-213/U cable

Three methods of sampling and recombining the signal on the delay-line were investigated:

1. Probe couplers
2. Directional filters
3. Directional couplers

Probe couplers are small and inexpensive. However, tight couplings in excess of 15 db are not obtained without VSWR's that can combine to a large value when as many as 40 units are required. Directional filters have the advantage of being both filters and directional couplers combined and, in theory, are matched at all frequencies. However, in the 300 to 400 Mc range, their size opposed their use. Therefore, coaxial directional couplers having good coupling-versus-frequency characteristics, high directivity, and low primary-line VSWR were used.

Fixed attenuators are inserted in the rungs of the ladder between the band-pass filters and the coupling ports of the directional couplers. These attenuators serve two purposes:

1. Making the overall CN loss-versus-frequency characteristic flatter
2. Minimizing power reflection caused by off-tune filters; the high directivity of the directional couplers provides most of the isolation.

As mentioned in Section II, the scanning-LO-frequency sweep rate is 100 Mc/μsec. The actual deviation from this rate was about ±5 percent. Other investigators (reference 5) have shown that shaping or weighting of the input spectrum minimizes compressed pulse-width broadening due to these scanning deviations as well as to amplitude and phase nonlinearities inherent in systems covering such a broad RF bandwidth. A two-pole weighting filter centered at 350 Mc, with -3 db responses at 325 and 375 Mc and -12 db responses at about 300 and 400 Mc, was placed at the input to the CN.

D. Overall CN Characteristics

The amplitude-versus-frequency characteristic of the complete CN shows a loss that averages 53 db across the 300 to 400 Mc band (Figure 11). While this loss may seem excessive, it must be remembered that at the 300-Mc end of the CN (that is, with maximum amount of delay cable), the cable loss alone is about 24 db. Some measurements were also made to determine the phase-versus-frequency characteristic of the complete CN. Unfortunately, the lack of time forced abandonment of further work in that direction, and the phase-versus-frequency characteristic was never fully determined.

V. Overall Receiver Performance

The following compressive-receiver performance characteristics were measured:

1. Compressed pulse width (compression factor)
2. Side-lobe level of compressed pulse
3. Frequency resolution.

A. Compressed Pulse Width

CW signals, or pulsed carriers whose pulse duration is long compared with the system scan period, were applied to the first IF amplifier (in the 2,600 to 3,400 Mc range) to measure the compressed pulse width. Applying the test signal at this point removes the small amount of FM jitter contributed by the local oscillator in the first converter. This aids in stabilizing the compressed-pulse presentation on a monitor oscilloscope. The total jitter in the presentation was about 10 nsec. Compared with a sweep period of 10 μsec, the jitter was 1 part in 1,000.

The compressed pulse width was measured with the aid of a signal generator and a wide-band oscilloscope. The signal level from the signal generator was adjusted to provide a high S/N ratio (about 20 db). The oscilloscope gain was adjusted to provide a convenient display. The input power was then reduced 3 db, and the amplitude of the output was noted. The original input level was restored, and the pulse width was measured at the reference level previously noted. (This reference level was found to be between 0.65 and 0.70 of the peak amplitude.) These measurements were performed at 20-Mc increments over the band from 2.6 to 3.4 Gc. The compressed pulse width varied from 15 to 22 nsec (Figure 12). The compression factor (the ratio of 1 μsec to the compressed pulse width) thus varied from about 50 to 70, compared to a theoretical compression factor of 100. Figure 13 shows a typical compressed pulse.

B. Side-Lobe Level of Compressed Pulse

The side-lobe level was also measured while making the compressed-pulse-width measurements. The level of the highest side lobe was 18 to 20 db below the level of the peak amplitude of the compressed pulse. This is to be compared with the 13.2-db first side lobe of sin t/t.

C. Frequency Resolution

The frequency resolution was measured by applying signals from two signal generators to the first IF amplifier and observing the compressed pulse output. The frequency separation was then varied until the output pulses merged to the point of 3-db peak-to-valley ratio. The frequency resolution was found to be 2 to 3 Mc. Figure 14 shows the output indication for two signals that are 100, 50, 20, 10, 5, and 3 Mc apart, respectively.

VI. Conclusions

A wide-band microwave receiver that scans a selected 800-Mc-wide RF band at 100 Mc/μsec has been described with emphasis on the design of the compression network and scanning local oscillator. By application of pulse-compression techniques to the design of this receiver, a resolution of about 2 to 3 Mc was obtained with band coverage in 10 μsec. By comparison, a conventional receiver with a resolution of 2 Mc would be restricted to a scan rate of about 4 Mc/μsec and would require 200 μsec to cover the same band.

VII. References

1. W. D. White, "Electronic Circuits," U. S. Patent No. 2,882,395, 14 April 1959.
2. H. Batten, R. Jorgensen, A. MacNee, and W. Peterson, "The Response of a Panoramic Receiver to CW and Pulse Signals," Proc IRE, Vol 42, p 948-956, June 1954.
3. C. E. Cook, "Modification of Pulse Compression Waveforms," Proc NEC, p 1058-1067, 1958.
4. G. Ohman, "Getting High Range Resolution With Pulse Compression Radar," Electronics, p 53-57, 7 October 1960.
5. J. Klauder, A. Price, S. Darlington, and W. Albersheim, "The Theory and Design of Chirp Radars," B.S.T.J., Vol 39, p 745-808, July 1960.

Acknowledgment

The authors wish to acknowledge the helpful theoretical discussions with W. D. White, and the very able contributions of L. LoSasso, I. Maxwell, A. Williams, and G. Kanischak to the realization of the equipment.

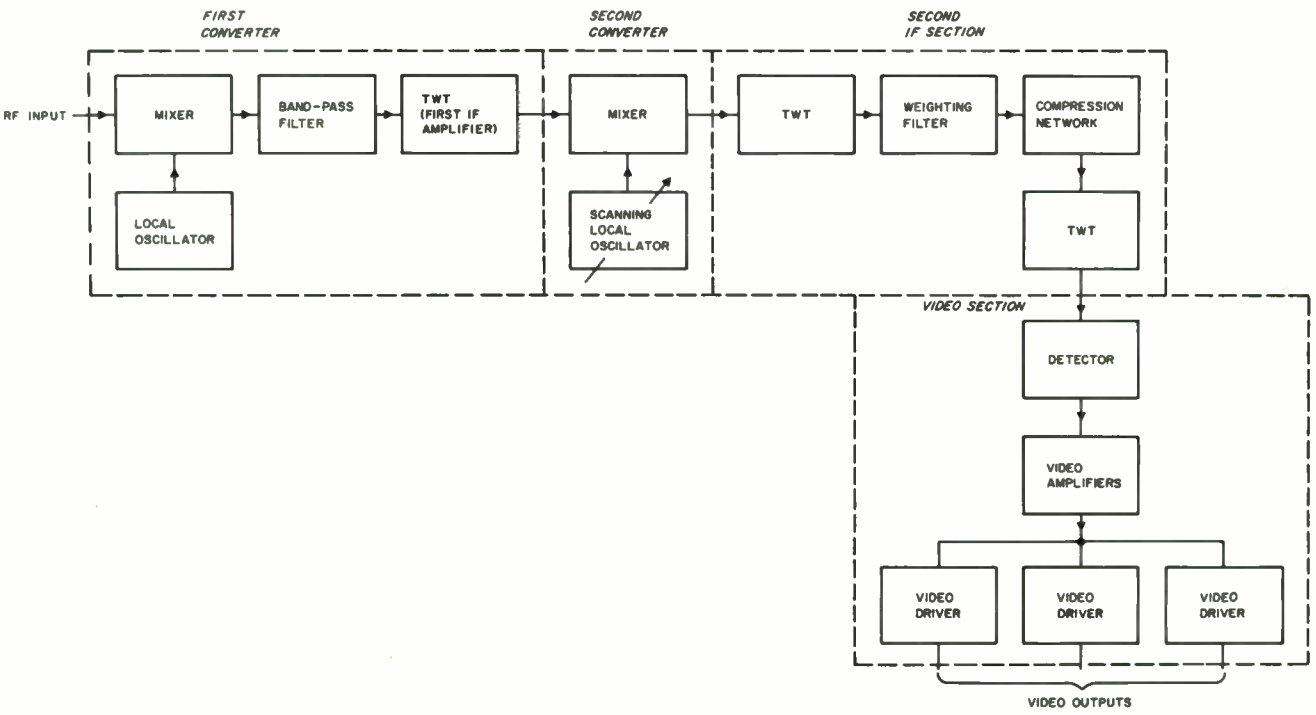


Fig. 1. Block diagram of compressive receiver.

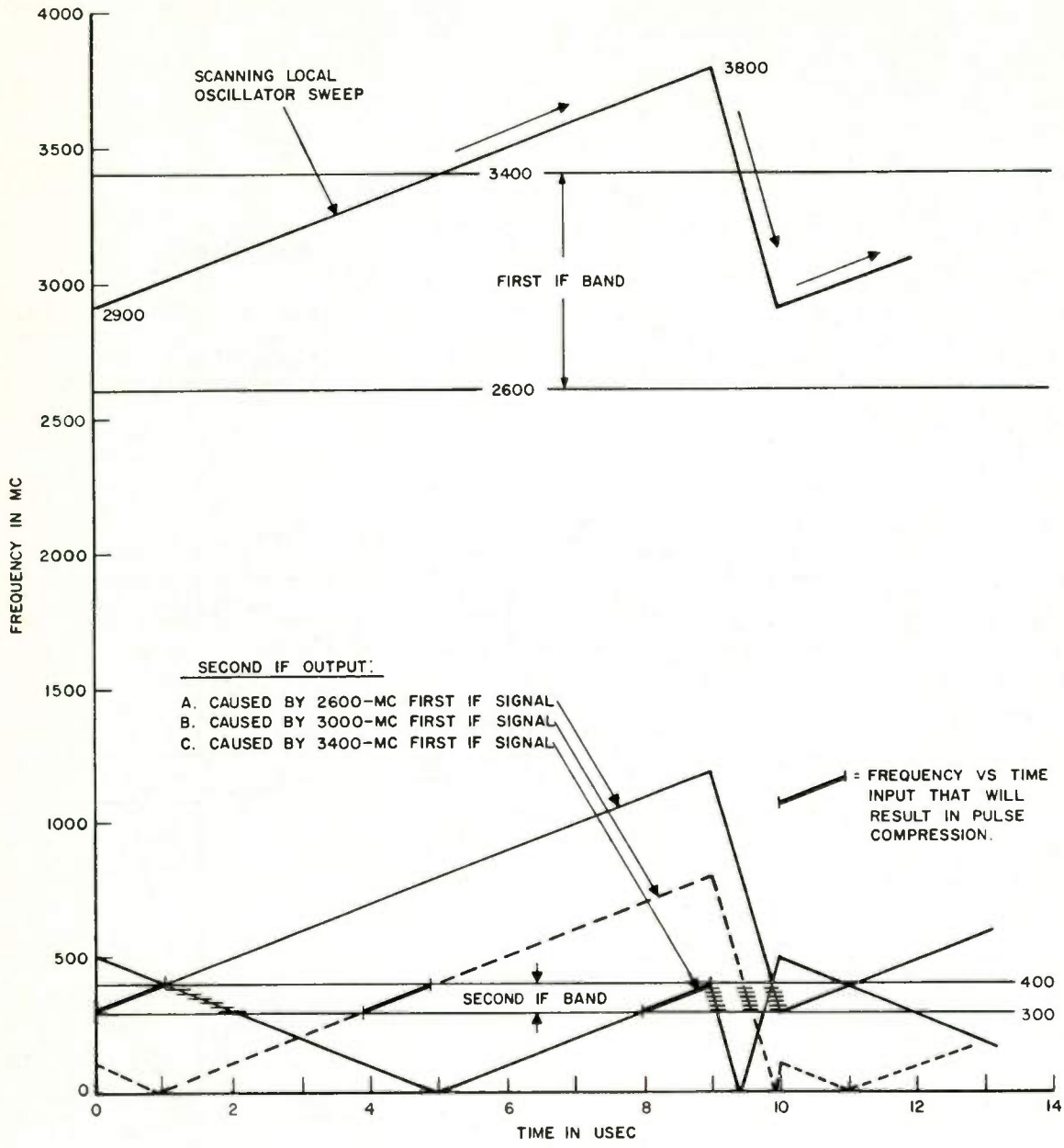


Fig. 2. Scanning action of compressive receiver.

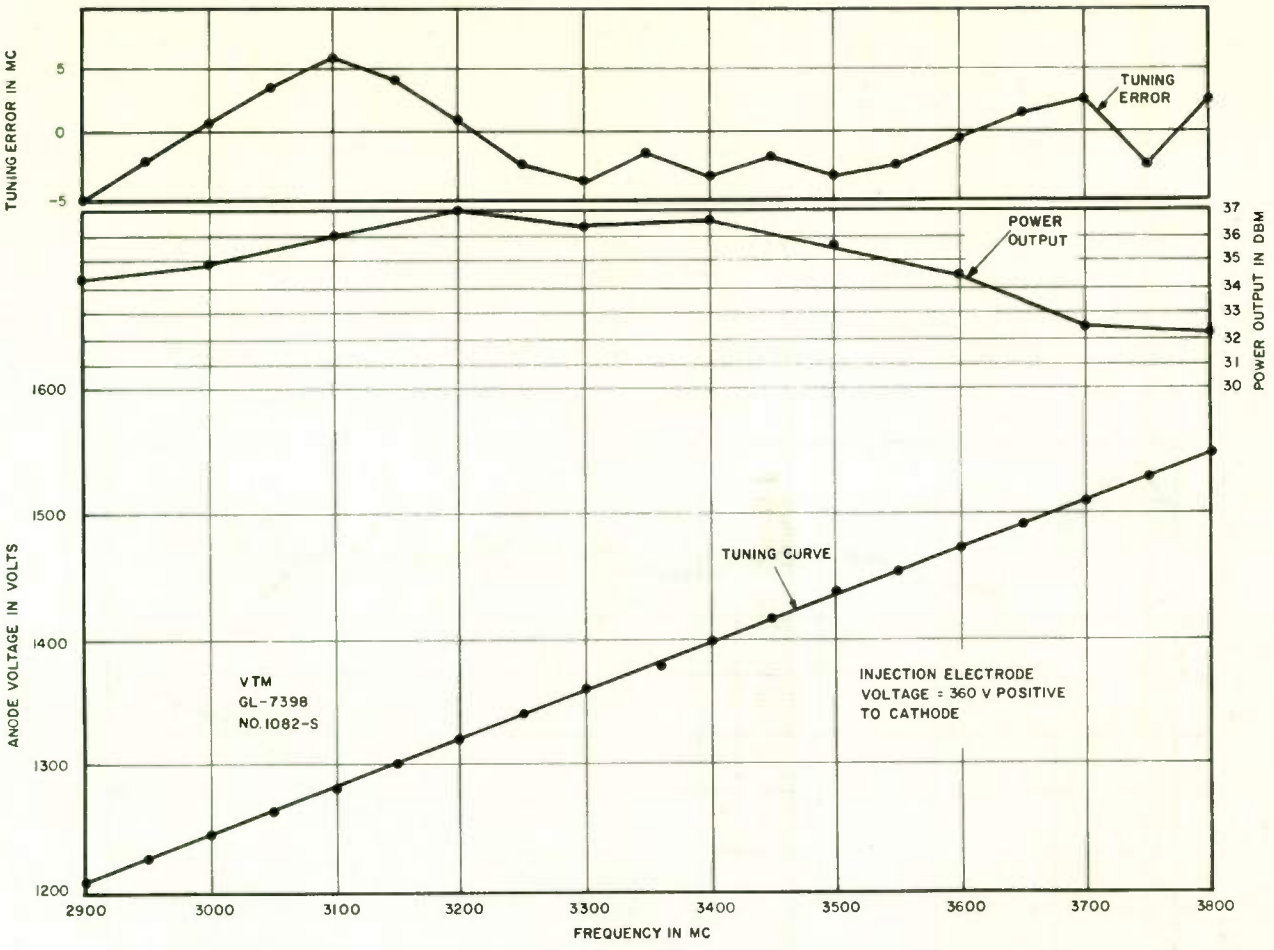


Fig. 3. Results of static tests on voltage-tunable magnetron.

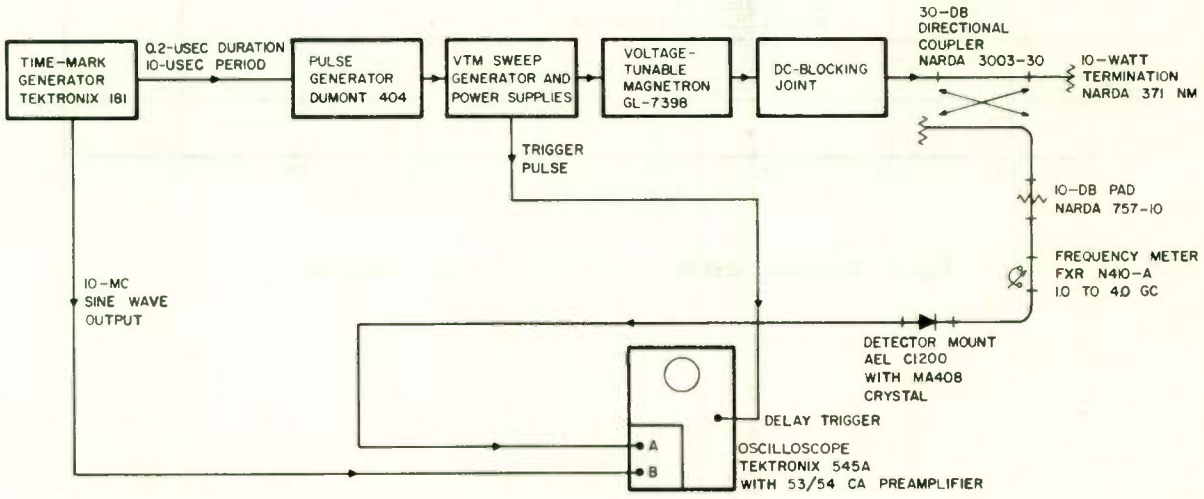


Fig. 4. Circuit for testing linearity of scanning local oscillator.

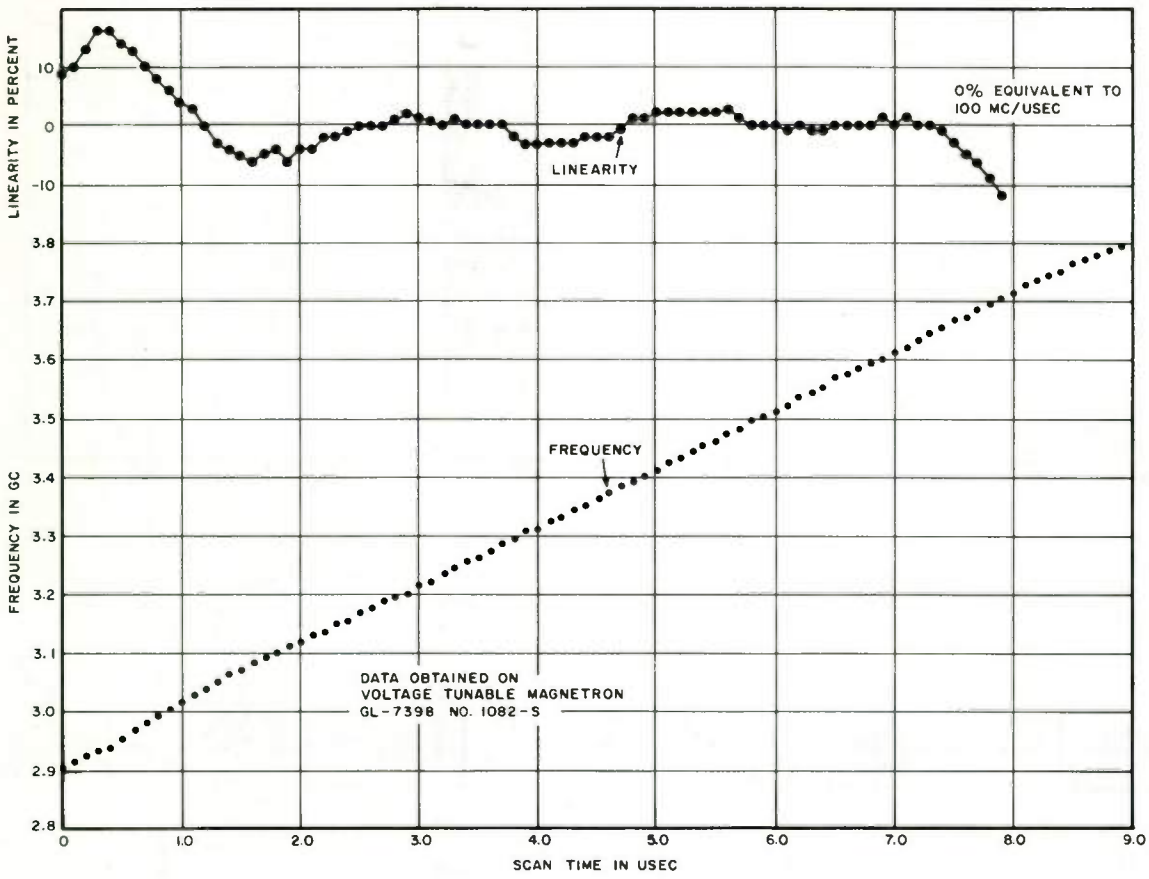
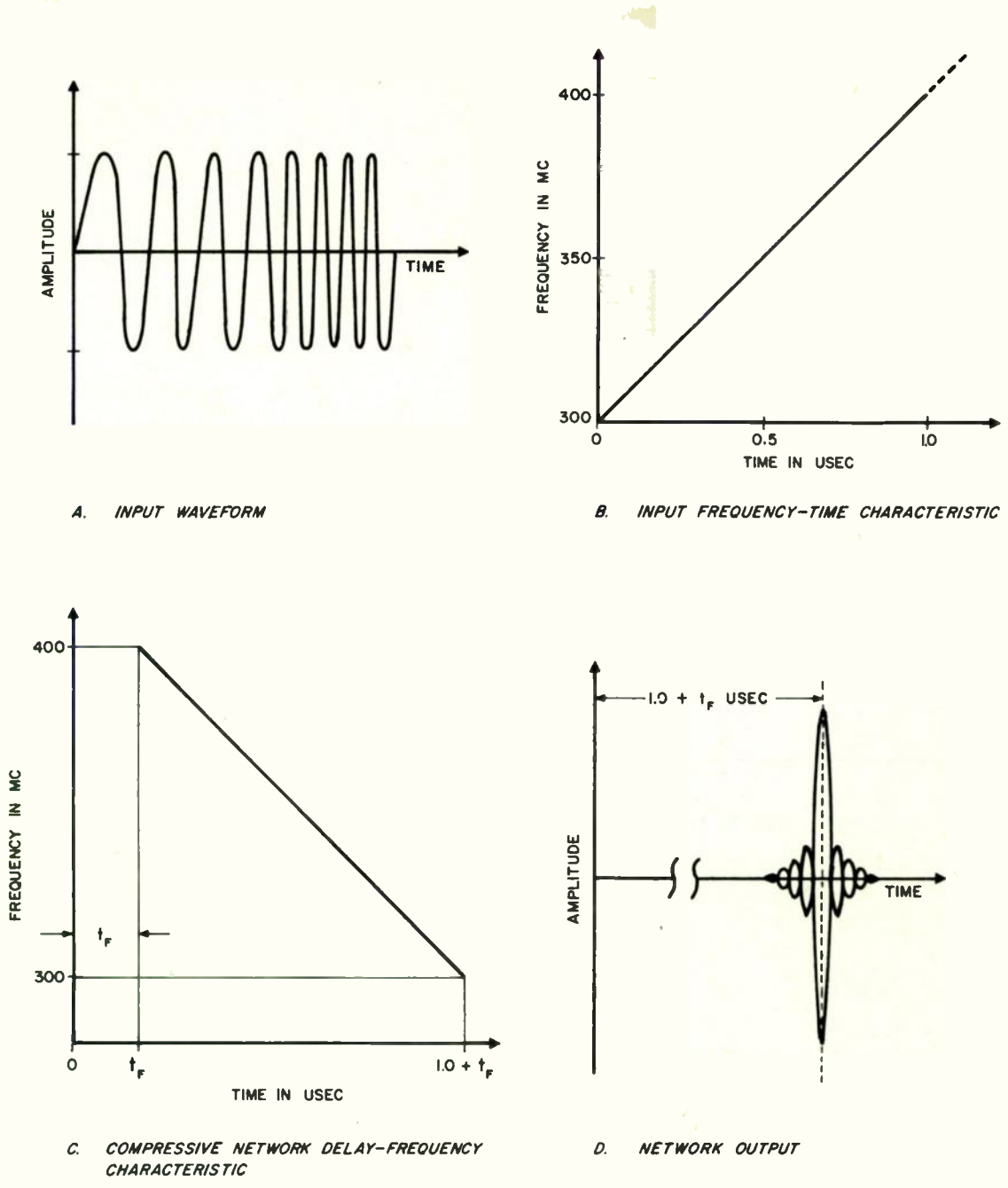


Fig. 5. Frequency and linearity of scanning local oscillator.



A. INPUT WAVEFORM

B. INPUT FREQUENCY-TIME CHARACTERISTIC

C. COMPRESSIVE NETWORK DELAY-FREQUENCY CHARACTERISTIC

D. NETWORK OUTPUT

Fig. 6. Idealized input and output waveforms of compression network.

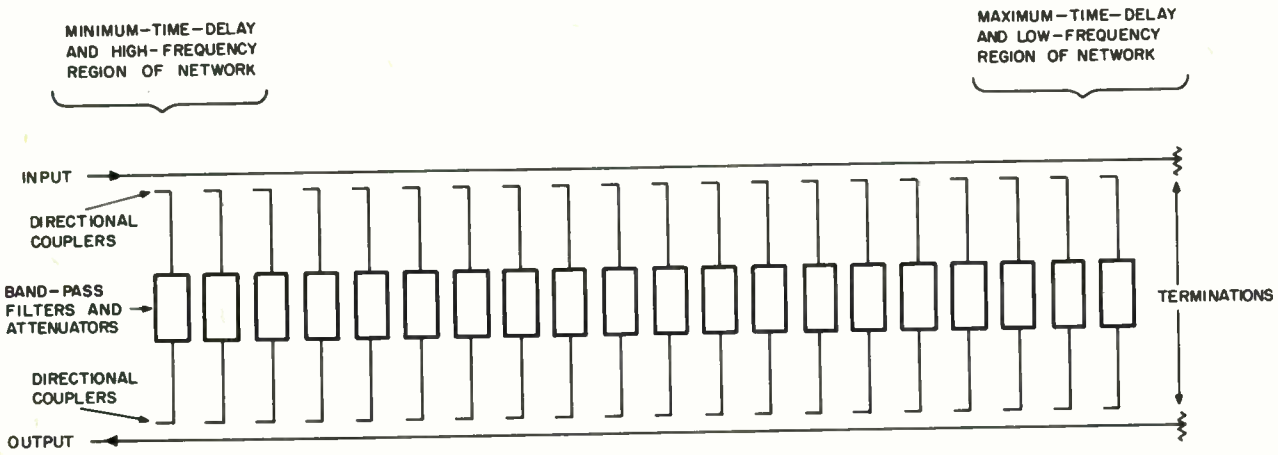


Fig. 7. Simplified diagram of compression network.

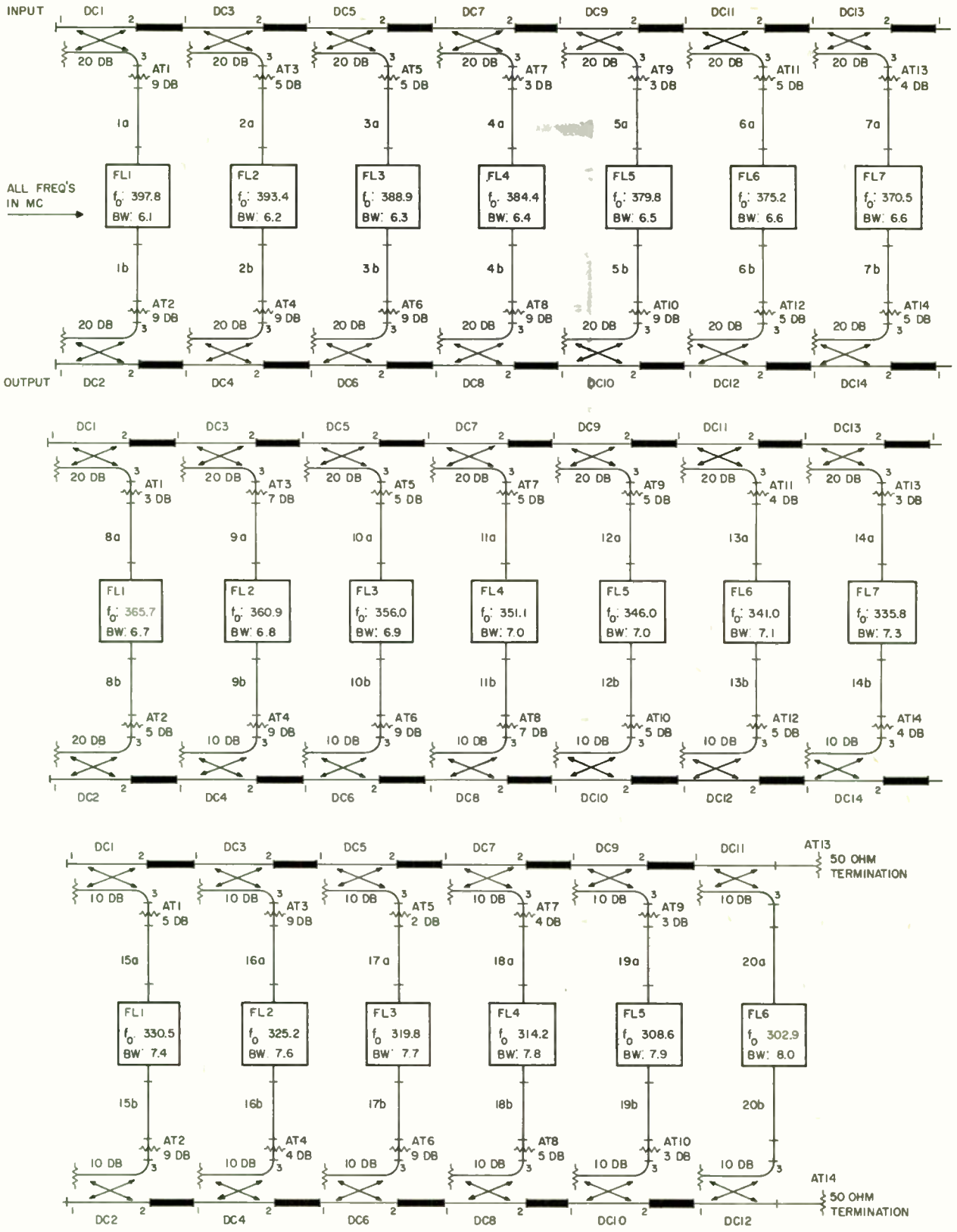


Fig. 8. Detailed schematic diagram of compression network.

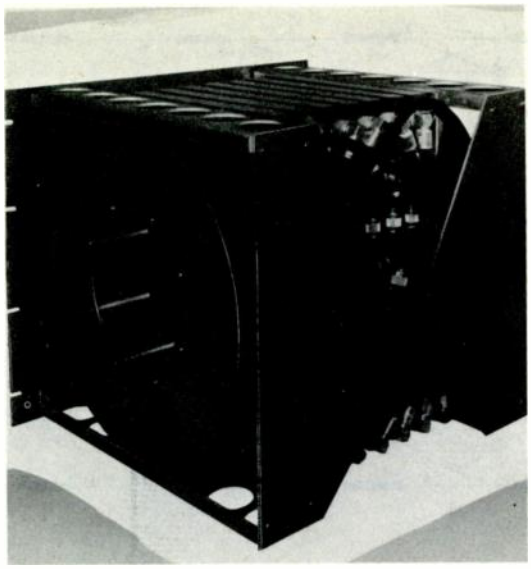


Fig. 9. Rear view of one compression network unit.

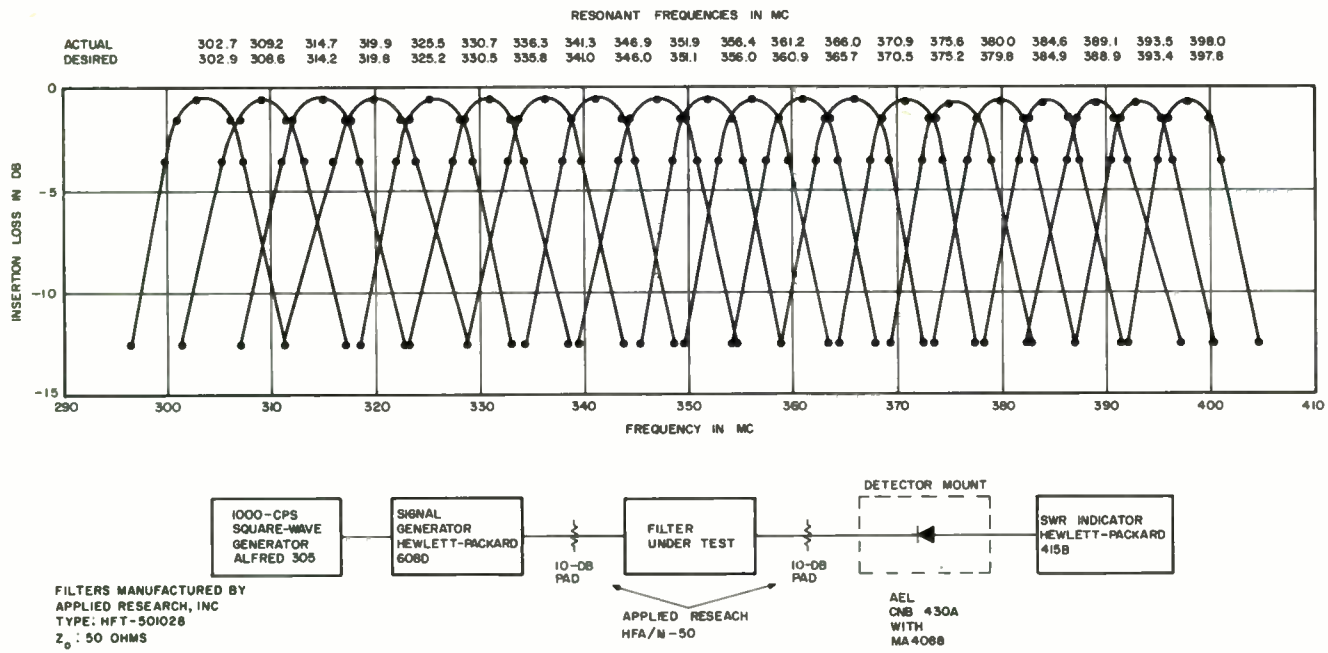


Fig. 10. Insertion loss vs frequency of compression network band-pass filters.

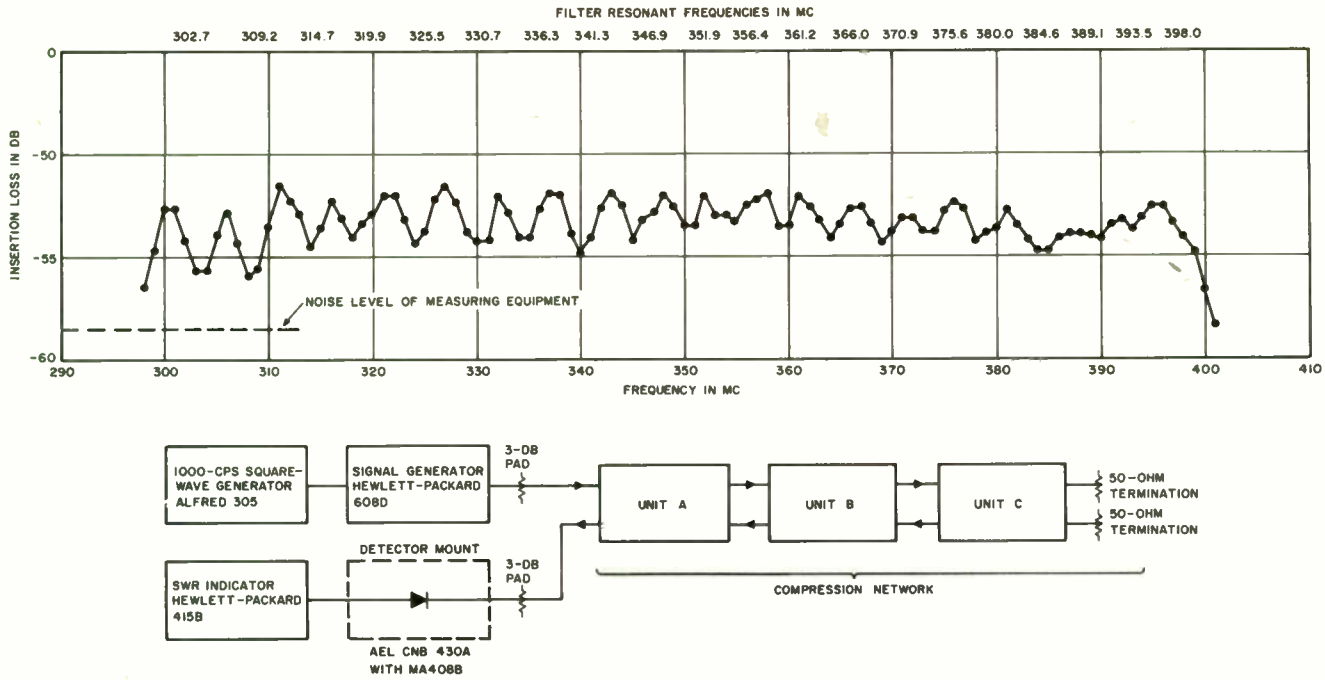


Fig. 11. Insertion loss vs frequency of complete compression network.

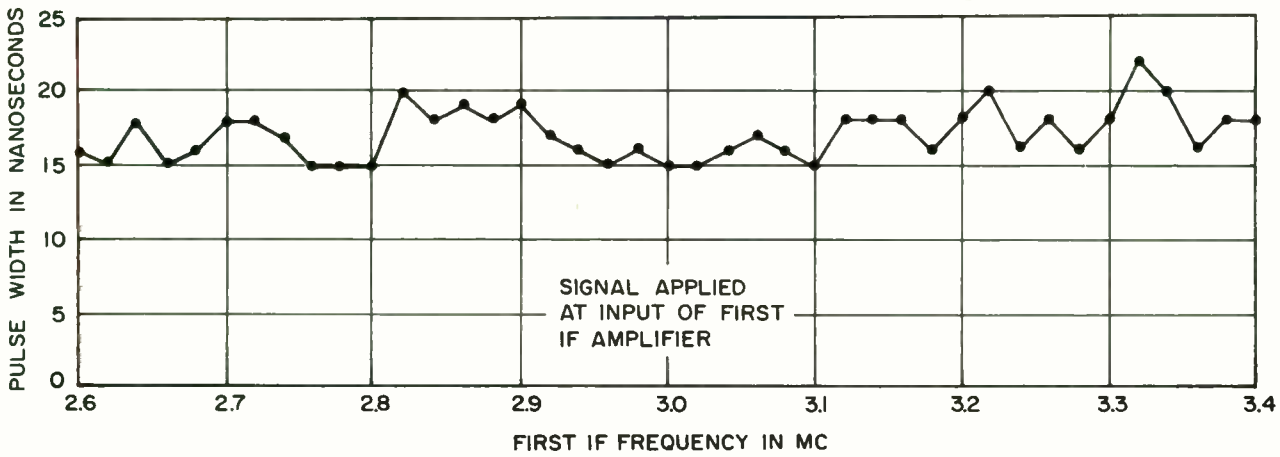
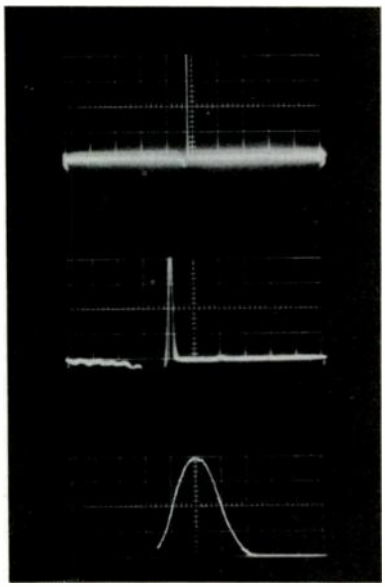


Fig. 12. Compressed pulse width vs frequency.



HORIZONTAL SCALE:
1.0 USEC/CM

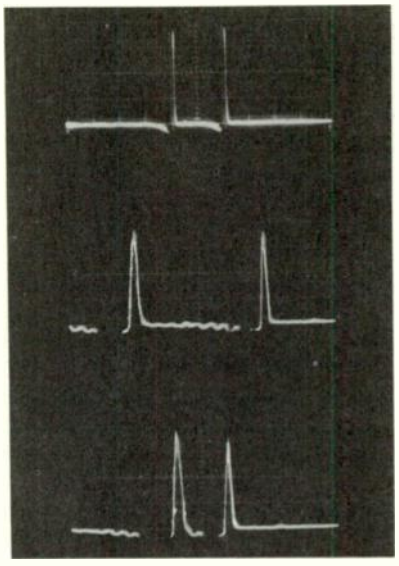
HORIZONTAL SCALE:
0.1 USEC/CM

HORIZONTAL SCALE:
0.01 USEC/CM

TIME INCREASES
FROM RIGHT TO LEFT

CW SIGNAL AT 3040 MC, COMPRESSED
PULSE WIDTH: 16 AT −3 DB POINTS

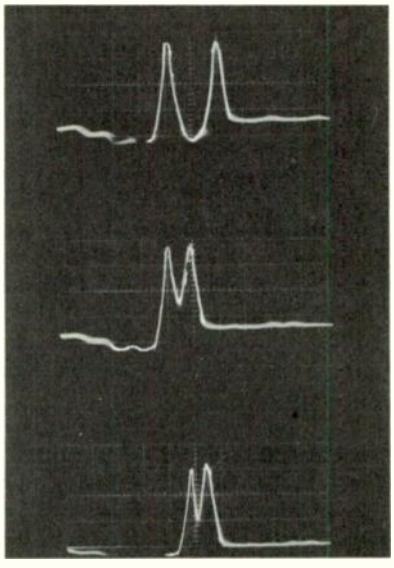
Fig. 13. Typical compressed pulse.



3300 MC, 3200 MC CW SIGNALS:
(HORIZONTAL SCALE 0.5 USEC/CM)

3300 MC, 3250 MC CW SIGNALS:
(HORIZONTAL SCALE 0.1 USEC/CM)

3300 MC, 3280 MC CW SIGNALS:
(HORIZONTAL SCALE 0.1 USEC/CM)



3300 MC, 3290 MC CW SIGNALS:
(HORIZONTAL SCALE 0.05 USEC/CM)

3300 MC, 3295 MC CW SIGNALS:
(HORIZONTAL SCALE 0.05 USEC/CM)

3300 MC, 3297 MC CW SIGNALS:
(HORIZONTAL SCALE 0.05 USEC/CM)

IN ALL PHOTOGRAPHS, TIME
INCREASES RIGHT TO LEFT

Fig. 14. Frequency resolution of compressive receiver.

A COMPACT UHF HIGH POWER FERRITE ISOLATOR

E. Wantuch and R. Poppe
Airtron Division of Litton Industries
Morris Plains, New Jersey

Summary

A compact UHF ferrite isolator was designed to cover the 406 to 450 Mc/s range. The isolator has been tested at power levels up to 3 MW peak and 5 KW average. This isolator has been designed in single-ridge waveguide with transitions to 3 1/8" diameter coaxial line at each end.

The ferrite, in the form of two strips, was mounted on the flat waveguide surface opposite the single-ridge near the edges of the ridge. This arrangement is advantageous for water cooling.

The ferrite used was a low saturation magnetization nickel aluminum ferrite capable of handling the high peak power without device deterioration.

This isolator possesses a front to back ratio of approximately 20 to 1 under high power conditions.

The overall length of this unit is 42" and its overall weight less than 100 pounds, which may be considered rather low at this low frequency range.

* * * *

One of the most challenging problems remaining in the ferrite device state of the art is the design of low frequency, high power isolators.

High power systems, except super power versions, in the frequency range below 1000Mc are generally designed in coaxial line rather than waveguide. High peak power breakdown problems are much aggravated by the use of coaxial transmission lines due to high electric field intensity at the center conductor surface.

Also, the design of ferrite isolators in coaxial transmission lines requires partial loading with (high) dielectric material to produce local circular polarized magnetic fields for nonreciprocal device operation. The use of such dielectric loading enhances the possibility of corona in any small air gaps left between the center conductor and the dielectric. The grinding of such large ceramic bodies to close tolerances also presents considerable difficulties.

In the frequency range of 406 to 450 Mc/s waveguide width of at least eighteen inches must be used for TE_{10} mode transmission. Previous designs of high power isolators in this frequency range have used coaxial transitions to reduced height waveguide of this or larger width. Waveguide height was reduced in order to lower the waveguide impedance and achieve better match to the 50 ohm coaxial line. In addition, reducing waveguide height also has the advantage of reducing magnet weight by shortening the air gap over which the proper magnetic field strength must be maintained.

The device described minimizes these design limitations by the use of single-ridged waveguide. A waveguide width of eleven inches was chosen with an aspect ratio of one-quarter. A ridge width of 6.6" was chosen, thus fixing the ridge gap at .66". The ridge edges were rounded with a one-quarter inch radius and the entire structure was fabricated of one-quarter inch aluminum plates.

The use of this transmission line permits the design of a rather simple transition to coaxial line.[1] This transition consists of a door knob with the inner conductor of the coaxial line fastened to the top of the ridge. The characteristic impedance of the ridged waveguide was designed to be 50 ohms. A low VSWR transition without elaborate matching structure was developed as is shown in Figure 1. This arrangement resulted in an input VSWR of less than 1.12 over

the 10% band-width of interest.

A low characteristic impedance transmission line also has the advantage of intensifying the r.f. magnetic field, thus resulting in stronger interaction with the ferrite material. In this manner more isolation per unit length may be achieved which is most desirable in the low microwave frequency range.

The choice of ferrite material is influenced by several factors. First, the saturation magnetization must be kept low in order to obtain good isolation to insertion loss ratio at this relatively low microwave frequency. The resonance linewidth must also be kept narrow for the same reason. The composition chosen must resist the onset of non-linear behavior, which is characteristic of many ferrite materials at high r.f. magnetic field intensity. These factors can be seen qualitatively in Figure 2.

The above parameters can best be met by a substituted nickel ferrite. The saturation magnetization of nickel ferrite may be lowered by the substitution of aluminum for iron.[2] This will result in a lowering of saturation magnetization with a compensation point obtained with an empirical formula of $NiAl_{.67}Fe_{1.33}O_4$. Compositions close to but below the compensation point result in a material having a high "g" factor which is undesirable. With aluminum substitutions of greater than .75, the "g" factor is less than 2. In the above empirical formula a more suitable material can be produced with a moderate sacrifice of Curie temperature. Figure 3 summarizes these relationships and indicates the composition used.

In order to maintain stable device operation, an efficient ferrite geometry had to be chosen from a heat-transfer standpoint. Fortunately, in single-ridged waveguide there are two circularly polarized magnetic field regions near the ridge edges. These have opposite senses of polarization; two ferrite slabs located in these regions may therefore be biased by means of a U-shaped permanent magnet. Proper direction of applied magnetic field may therefore be obtained with an efficient magnetic circuit. This arrangement is shown schematically in Figure 4.

The following device performance was obtained with this design approach:

Frequency range: 406 to 450 Mc/s
Isolation: 7 db
Minimum insertion loss: 0.3 db
Input VSWR: 1.20 max.
Power handling capability: 3 MW peak, 5 KW average into 2:1 load VSWR at 30 psia of dry air.

Overall dimensions of the device are shown in Figure 5, and a photograph of the unit is shown in Figure 6.

Cooling was provided by a water and glycol mixture at an inlet temperature of 50°C.

This device has been successfully flight tested in an existing radar system and has performed reliably under typical military conditions.

# # #

References

1. Very High-Frequency Techniques, Vol. II, McGraw-Hill Book Company, Inc.
2. E. W. Gorter, "Saturation Magnetization and Crystal Chemistry of Ferrimagnetic Oxides," Philips Research Report no. 9, Aug. Oct. Dec. 1954.

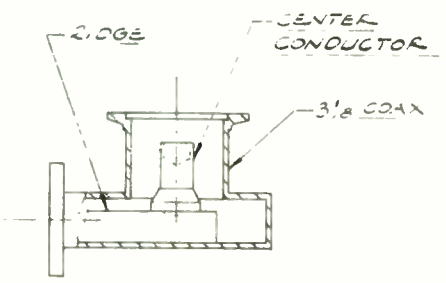


Fig. 1. Cross sectional view of transition.

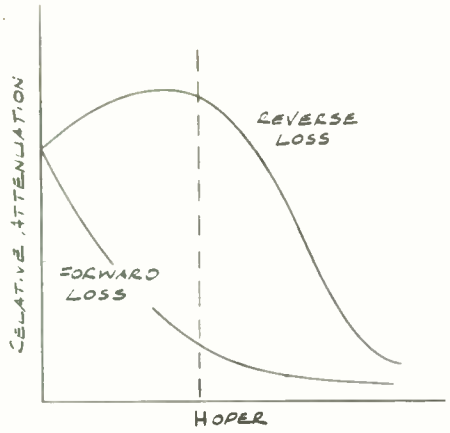


Fig. 2. Applied magnetic field—general isolation characteristics.

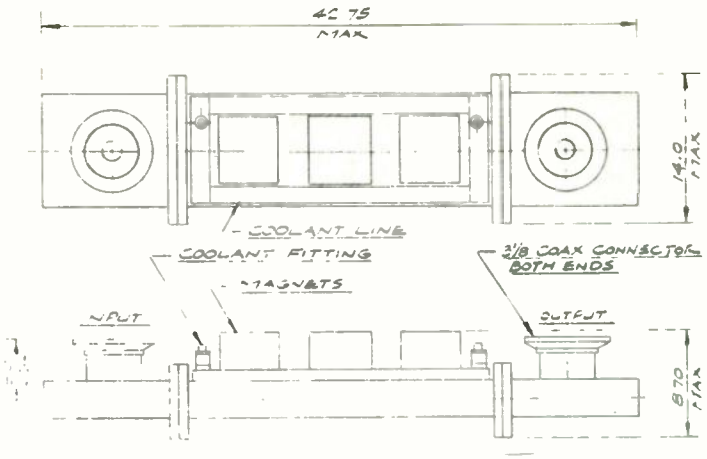


Fig. 5. High power isolator UHF.

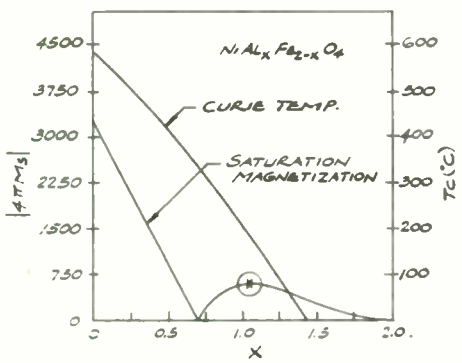


Fig. 3. Saturation magnetization and curie temperature vs. composition.

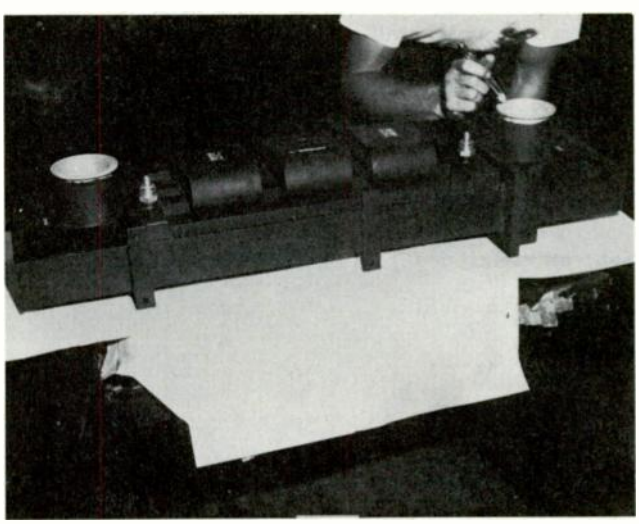


Fig. 6. Photograph of complete unit.

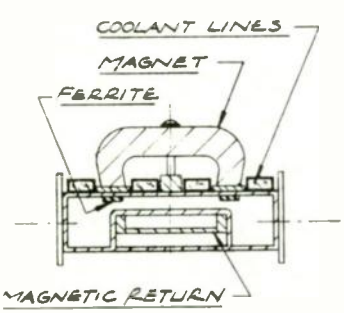


Fig. 4. Cross sectional view of ferrite section.

ANALYSIS OF RESONANT CAVITIES IN PARAMETRIC AMPLIFIERS AND FREQUENCY MULTIPLIERS

Y. KAITO
Nippon Electric Co., Ltd.
Kawasaki, Japan

Summary

Using two fundamental resonant circuits, the analysis on the composition of parametric amplifier or frequency multiplier having a varactor diode was carried out. One of the two fundamental circuits consists of a varactor diode, a transmission line and a parallel susceptance. The circuit using a quarter-wavelength resonator may be considered a variation of this circuit. The other circuit is composed of a varactor diode, a transmission line and a filter having different coupling coefficients. This resonant circuit can be conveniently used in frequency multipliers.

The results of these analyses are shown in the accompanying two design charts. 2 Gc parametric amplifiers, designed by making use of one of these charts, are being used in over-horizon communication systems in Japan and have gain of 16 dB, bandwidth of 15 Mc and noise figure of 2.5 dB. A frequency multiplier, which has been designed by utilizing the other design chart for doubling 0.5 Gc to 1 Gc, shows an efficiency of 50% at 500 mW output.

Introduction

The operating characteristics of microwave parametric amplifier with a varactor diode and frequency multiplier have already been discussed in the past by employing equivalent circuits. However there have been few discussions or papers on the subject of the composition of resonant circuits which become necessary when applying to actual circuits the results obtained through the use of equivalent circuits.

This paper discusses how to compose actual resonant circuits in order to realize a parametric amplifier or a frequency multiplier with desired performance. The optimum noise factor and large gain condition of parametric amplifier as well as the maximum efficiency of frequency multiplier depend on the cut-off frequency of diode, variation ratio of capacitance and frequency in use.

And the coupling coefficient between the diode and the external load which gives this optimum condition can be calculated by means of the equivalent circuit.

Accordingly in the analysis of a resonant circuit, the objective is to determine the electrical dimension of elements composing the circuit to get the desired coupling coefficient.

Equivalent circuits of parametric amplifier

Representation of performance by equivalent circuit

The relation between the constants of equivalent circuit and its operating characteristics will be briefly explained in the following, as it is necessary in the analysis of the resonant circuit of parametric amplifier. An example of the equivalent circuits of a circulator-type parametric amplifier or a lower sideband up-converter is shown in Fig.1. ω_1, ω_2 and ω_0 present the angular frequencies of signal, idler and pumping respectively. The parallel resonant circuits L_{10} C_{10} and L_{20} C_{20} in Fig. 1 are assumed to have the symbolic meaning of being resonant to ω_1 and ω_2 respectively while offering a short-circuit path to any other frequencies. G_{10} and G_{20} in the figure are likewise assumed to represent the external conductances of ω_1 and ω_2 while G_1 and G_2 are the conductances indicating the losses of the diode at ω_1 and ω_2 respectively. And it is further assumed that there is no loss in the circuit itself.

When ΔC is being pumped and changed as in the equation (1),

$$\Delta C = 2\, C_1 \cos \omega_0 t \tag{1}$$

As is widely known, the voltage reflection coefficient of $\omega_1(\Gamma)$, conversion gain (Conv. Gain) from ω_1 to ω_2 and noise figure (F) can be shown as the followings. And $|\Gamma|^2$ represents the power gain of a circulator-type parametric amplifier.[1,2]

$$\Gamma = \frac{(G_{10}/G_1 - 1)(G_{20}/G_2 + 1) + \omega_1\omega_2 C_1^2/G_1G_2}{(G_{10}/G_1 + 1)(G_{20}/G_2 + 1) - \omega_1\omega_2 C_1^2/G_1G_2} \tag{2}$$

$$\text{Conv. Gain} \doteqdot \frac{1}{1+G_2/G_{20}} \frac{\omega_2}{\omega_1} |\Gamma|^2$$

$$(\text{when } G_{10} \gg G_1) \tag{3}$$

$$F \doteqdot (1 + \omega_1/\omega_2)(1 + G_1/G_{10}) \tag{4}$$

In order to have Γ sufficiently large, the following equation must be realized.

$$(1 + G_{10}/G_1)(1 + G_{20}/G_2) \doteqdot \omega_1\omega_2 C_1^2/G_1G_2 \tag{5}$$

As can be seen from the equations (2) to (5), the gain and noise factor of a parametric amplifier

are determined by Gx/Gxo (X = 1,2) and $\omega_1 \omega_2 C_1^2/G_1G_2$.

Here, Gx/Gxo represents the coupling coefficient between the diode and the external load at signal or idler frequency, and when this external load is a matched transmission line, the Gx/Gxo stands for the input conductance normalized by its characteristic conductance at the time when the pumping power is not applied.

As to be described next, $\omega_1 \omega_2 C_1^2/G_1G_2$ is the value depending on the characteristics of diode and pumping condition; and when this value becomes the maximum, the noise factor can be made the minimum under the conditions of a definite gain and bandwidth.

Relationship between equivalent circuit and actual circuit

The equivalent circuit of a varactor diode may be represented as in Fig. 2 by a series element of a fixed resistance Rs and a variable capacitance Cb which changes according to its bias voltage.[3]

In order to be equal the actual resonator including a diode to the equivalent circuit such as shown in Fig. 1, when we look at the resonator from its external terminals, the input admittance must satisfy the following conditions: When the capacitance Cb of the diode is statically altered, the input conductance remains unchanged while only the input susceptance changes, and this change in susceptance must be plus and minus with 0 as its center whenever Cb is varied plus and minus with a certain value Co. In other words, the equivalent circuit shown in Fig. 3 has to be realized at ω_1 and ω_2.

Under this condition the resonant cavity mounted with a diode may be considered as a kind of impedance converter that transforms a diode impedance into an input admittance satisfying the above-mentioned conditions.

An example of an ideal impedance transformer satisfying the said conditions is shown in Fig. 4. F_1 and F_2 in the illustration indicate ideal filters that pass ω_1 and ω_2 only, blocking all other frequencies.

The transmission lines ℓ_1 and ℓ_2 have their lengths corresponding to the quarter wavelengths of ω_1 and ω_2 respectively, their characteristic impedances being taken as Zo. T_1 and T_2 are ideal transformers while B_1 and B_2 indicate parallel susceptances. When operating as an amplifier, the external conductances of signal and idler (G_{10} and G_{20} in Fig. 1) are to be connected at the terminals aa' and bb'. Also R-jX is to be the impedance normalized by Zo of Rs and Cb in Fig. 2.

When the diode is pumped at ω_o, the change in Cb takes place in the following manner:

$$Cb = Co + 2C_1' \cos\omega_o t + 2C_2' \cos 2\omega_o t \qquad (6)$$

When this change in capacitance is observed through ℓ_1 or ℓ_2, the normalized admittance becomes as follows:

$$\frac{Y}{Yo} = R - jX = \frac{Rs}{Zo} - j\frac{1}{Zo\,\omega_x Co}$$
$$\left(1 + \frac{2C_1'}{Co}\cos\omega_o t + \frac{2C_2'}{Co}\cos 2\omega_o t\right)^{-1} \qquad (7)$$

$$\text{Assuming } \frac{2C_1'}{Co},\ \frac{2C_2'}{Co} \ll 1 \qquad (8)$$

$$\frac{Y}{Yo} \doteqdot \frac{Rs}{Zo} - j\frac{1}{Zo\,\omega_x Co^2}(Co - 2C_1'\cos\omega_o t)$$

Next, the admittance is transformed by T_1 or T_2 and the fixed portion of the transformed susceptance, which does not change at the frequency of ω_o, is canceled out by B_1 or B_2. The input admittance as seen from aa' and bb' can be represented in the following manner:

$$Yin(X) \doteqdot Rs\,\omega_x^2 Co^2 + j\omega_x 2C_1' \cos\omega_o t \quad (X=1,2)$$

By comparing this with the equation (1) and the equivalent circuit in Fig. 3, it becomes apparent that it can be denoted as follows:

$$C_1 \doteqdot C_1' \qquad (9)$$

$$Gx \doteqdot Rs\,\omega_x^2 Co^2 \qquad (10)$$

The equations (9) and (10) will be realized under the condition of the equation (8), and it is not imperative that the angular cut-off frequency ω_c corresponding to Co defined by the equation (11) has to be far greater than ω_x (X = 1,2).

$$\omega_c = \frac{1}{CoRs} \qquad (11)$$

Now, $\omega_1 \omega_2 C_1^2/G_1G_2$, a factor determining the performance characteristic of parametric amplifier, may be represented in the following manner by using the equations (10) and (11).

$$\frac{\omega_1 \omega_2 C_1^2}{G_1G_2} \doteqdot \frac{\omega_c^2 \delta^2}{\omega_1 \omega_2} = \frac{M^2}{\omega_1 \omega_2} \qquad (12)$$

Here, δ and M are to be defined as follows:

$$\delta \equiv \frac{C_1}{Co}, \qquad M \equiv \delta\,\omega_c \qquad (13)$$

The lefthand side of the equation (12) is proportional to the product of the ratio of the absolute values of input susceptance to input conductance at ω_1 and the similar ratio at ω_2. Accordingly, even if the impedance transformation ratio of T_1 and T_2 is changed in order to select the value of Gx/Gxo as one pleases, the value of

$\omega_1\omega_2C_1^2/G_1G_2$ remains unaffected or fixed. In other words, this value is dependent on the figure of merit M of diode and the frequency used.[4,5]

Fig. 4 shows an ideal circuit which transforms a varactor diode, shown as a series element of variable capacitance and resistance, into a parallel circuit of variable capacitance and resistance, and in which a coupling coefficient Gx/Gxo can be chosen at any desired value. However, in materializing this circuit in the microwave region, there are the problem of too many components, practical way of realizing an ideal transformer, etc. When an actual circuit satisfies the following conditions, it becomes same to the equivalent circuit of Fig. 4 or that of Fig. 1. The conditions mentioned below serve as a basis in the analysis of fundamental resonant circuits.

Conditions: When a varactor diode in no pumping power or cold state is observed through a some circuit by statically varying its bias voltage,

1. Input conductance remains fixed regardless of the variations in the capacitance of diode.
2. Input susceptance becomes zero when the capacitance of diode becomes Co.
3. Under the conditions of 1 and 2, the normalized input conductance (Gx/Gxo ≡ GIN) can be taken at any value.

Analysis of fundamental resonant circuit

Deferring to the next chapter the method of realizing frequency selective circuits such as F_1 and F_2 shown in Fig. 4, the analysis of a circuit transforming the diode impedance into parallel admittance will be described below.

Transmission line resonator using a susceptance

A fundamental resonant circuit as shown in Fig. 5 which consists of a coaxial line terminated by a diode and a susceptance shunted the line will be analyzed. Fig. 5 (a) may be redrawn like Fig. 5 (b).

The symbols or notation in the figure have the following meanings.

$R = \frac{Rs}{Zo}$: Series resistance of diode normalized by the characteristic impedance of transmission line.

$-jX = -j\frac{1}{\omega CbZo}$: Reactance of diode normalized by the characteristic impedance of transmission line.

$\theta = 2\pi\frac{\ell}{\lambda}$: Electrical length between diode and jB (radian)

jB : Shunt susceptance normalized by the characteristic admittance of transmission line.

φ : Electrical length between jB and the reference plane AA' for making Fig. 5 (c).

Gin : Input conductance seen from the reference plane AA' normalized by the characteristic admittance of the line (corresponds to Gx/Gxo in Fig. 3).

Fig. 6 is redrawn for the purpose of computing the input admittance as seen from AA' in Fig. 5 (b). The following relationship exists among E_1, I_1, E_2 and I_2 of the circuit in Fig. 6.

$$\begin{bmatrix}E_1\\I_1\end{bmatrix} = \begin{bmatrix}AB\\CD\end{bmatrix}\begin{bmatrix}E_2\\I_2\end{bmatrix} = \begin{bmatrix}1 & R-jX\\0 & 1\end{bmatrix}\begin{bmatrix}\cos\theta & j\sin\theta\\ j\sin\theta & \cos\theta\end{bmatrix}$$

$$\begin{bmatrix}1 & 0\\ jB & 1\end{bmatrix}\begin{bmatrix}\cos\varphi & j\sin\varphi\\ j\sin\varphi & \cos\varphi\end{bmatrix}\begin{bmatrix}E_2\\I_2\end{bmatrix} \qquad (14)$$

The input admittance of Fig. 5 may be shown as $-I_2/E_2$ when the left side of Fig. 6 is short-circuited or as A/B of the equation (14). After calculations, the equation (15) will be obtained from the equation (14),

By taking,

$$\frac{A}{B} \equiv \frac{\overline{AA}}{\overline{BB}} \qquad (15)$$

$$\begin{aligned}\overline{AA} = R + j\Big[&(R^2 - 1 - 2BX + X^2)\frac{\sin(2\varphi + 2\theta)}{2}\\ &+ (R^2B + BX^2)\cos^2\theta\left(\cos 2\varphi - \frac{B\sin\varphi}{2}\right)\\ &+ (B + B^2X - R^2B - 2X - BX^2)\sin 2\varphi\frac{\sin 2\theta}{2}\\ &- \frac{B^2}{2}\sin 2\varphi\sin^2\theta + X\cos 2\varphi\cos 2\theta\\ &+ B\cos 2\varphi\sin^2\theta\Big]\end{aligned}$$

$$\begin{aligned}\overline{BB} = \Big[&x\left\{\cos(\theta+\varphi) - B\sin\varphi\cos\theta\right\}\\ &+ B\sin\varphi\sin\theta\\ &- \sin(\theta+\varphi)\Big]^2 + R^2\\ &\left[\cos(\varphi+\theta) - B\sin\varphi\cos\theta\right]^2\end{aligned}$$

From the condition that the input conductance in the equation (15) remains unaffected irrespective of changes in X, the term having X in the denominator has to be zero. From this the equation (16) results:

$$\cot\varphi = \tan\theta + B \qquad (16)$$

Also assuming a resonant condition is reached at $x = x_o$ corresponding to Cb = Co, then the imaginary part must become zero. Out of these stipulations after considerably long calculations the equation (17), convenient for design purpose, is obtained.

$$2\theta = \cos^{-1} \frac{-(2X_o + B^2X_o)}{|B|\sqrt{(x_o^2+1)(B^2+4)}}$$

$$+ \cos^{-1} \frac{2B + B^2X_o}{|B|\sqrt{(x_o^2+1)(B^2+4)}} \qquad (17)$$

In the equation (17) the following assumptions are to be made: that B takes the limits of $4X_o^2 \leq B^2 (B^2 + 4)$, and if the angle of the first term has the limits of $\pi/2 \sim \pi$, the second term takes the limits of $0 \sim \pi/2$ for $2X_o > B$ and the limits of $0 \sim -\pi/2$ for $2X_o < B$ when $B > 0$, while the second term has the limits of $-\pi/2 \sim -\pi$ for $|B| < 2/X_o$ and the limits of $0 \sim -\pi/2$ for $|B| > 2/X_o$ when $B < 0$.

From the equations (15), (16) and (17), the equation (18) for the input conductance Gin will be obtained.

$$\frac{Gin}{R} = \frac{2 + B^2 + \sqrt{B^2(B^2+4) - 4X_o^2}}{2(X_o^2 + 1)} \qquad (18)$$

The results of calculations obtained from the equations (16), (17) and (18) are shown in Fig. 7.

It is apparent that by applying the characteristics R and X_o of a diode and the desired Gin = Gx/Gxo in Fig. 7, B, θ and φ required for the composition of a resonant circuit may be obtained. By the way, since the calculations or computations needed to arrive at the curves in Fig. 7 are quite complicated, for the check of Fig. 7 it is convenient to make use of the Smith chart.

Transmission line resonator using a filter

Just as in the previous case, the analysis of this circuit will be carried out by using a simplified circuit as shown in Fig. 8 (a). Fig. 8 (a) shows the circuit which consists of a coaxial line terminated by a diode and has a filter having its input and output coupling coefficients differing from each other connected to the lines. To make the explanation simpler, it is assumed that the resonant frequency of the filter is identical to the frequency of the resonant circuit. Fig. 8 (a) can be redrawn like Fig. 8 (b).

Of the symbols in the illustration, R, -jX and Gin have the same definitions as before, and the rest has the following meaning:

$\theta \doteq \frac{2\pi\ell}{\lambda}$: Electric length to the point where the filter can be considered as an ideal transformer looking from the diode. (radian)

k : Impedance transformation ratio when the filter is considered as an ideal transformer at the resonant frequency.

φ : Electrical length from the point where the filter can be considered as an ideal transformer to the reference plane AA' where it is replaced with one like Fig. 8 (c).

The input admittance, observing the diode side from the reference plane AA' in Fig. 8 (b), can be obtained by computing A/B of the equation (19) similarly as before.

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} 1 & R-jX \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & j\sin\theta \\ j\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{k}} & 0 \\ 0 & \sqrt{k} \end{bmatrix} \begin{bmatrix} \cos\varphi & j\sin\varphi \\ j\sin\varphi & \cos\varphi \end{bmatrix} \qquad (19)$$

Same as in the previous case, from the two similar stipulations that the input conductance unalters even if X is changed and that the input susceptance at resonance or $X = X_o$ is zero, the following equation is obtained.

$$k \cot\varphi = \tan\theta \qquad (20)$$

$$\cos 2\varphi = \mp \sqrt{1 - \frac{4X_o^2}{(1/k - k)^2}} \qquad 1 > k \qquad (21)$$

$$\cos 2\varphi = \pm \sqrt{1 - \frac{4X_o^2}{(1/k - k)^2}} \qquad 1 < k$$

$$\frac{R}{Gin} = \frac{k + 1/k \pm \sqrt{(k - 1/k)^2 - 4X_o^2}}{2} \qquad (22)$$

The results obtained from the euqations (20), (21) and (22) are shown in Fig. 9. In this chart, the lines corresponds to lines and dotted lines to dotted lines. By using this chart, θ, k and φ will be got at once from R and X_o of a diode, and desired Gin.

Now the case where the resonant frequency of the filter is somewhat away from the frequency of a resonant circuit will be considered. In this case the filter is to be equivalently represented by the combination of an ideal transformer with turn ratio of $\sqrt{K}$: 1 and a parallel susceptance of $\pm$ jb, (when the resonant frequency of filter is higher than the frequency of the circuit, the notation becomes -jb).

When this circuit is operated as a parametric amplifier, a characteristic admittance is to be connected across the external terminals. In other words, since an external conductance 1 is to be connected across AA' in Fig. 8 (b), if we consider an excess line $\mp\Delta\theta$ added to the detuned filter, as Fig. 10 (a) and the admittance seen from BB', we shall find the circuit becomes like Fig. 10 (b) in which the susceptance has been canceled out, leaving the ideal transformer having

ratio of $\sqrt{K'}$: 1. In this case if $K \ll 1$ and moreover if the frequency is only slightly out of resonance, then $K' \doteqdot K$. This means that in Fig. 8 without any change in ℓ, θ can be equivalently altered. For example, if the resonant frequency of a filter is shifted toward the higher direction, the equivalent θ becomes shorter, resulting in the higher resonant frequency of this system.

This fact indicates that the resonant frequency of the system can be changed without altering the equivalent short plane of the filter at the frequency considerably away the resonant frequency of this filter, and from practical standpoint this is a very convenient characteristic of the resonant circuit.

A quarter-wavelength resonant circuit

As shown in Fig. 11, when a quarter-wavelength resonator is terminated by a conductance Ga which is much smaller than the characteristic admittance Y_o, the normalized input admittance Gin seen from the terminal placed at θ_1, from shorting end is expressed as follows[6];

$$Gin \doteqdot \frac{Ga}{Yo \sin^2\theta_1} \tag{23}$$

Accordingly, Gin can be taken at any value by shifting the position of taps. If the relationship of the equation (24) exists between the operating frequency ωx and ωc as defined by the equation (11), the diode may be approximated to parallel elements consisting of Gx, Co and ΔC.

$$\omega x \ll \omega c \tag{24}$$

Therefore within the range where the equations (25) and (26) hold true, instead of Ga in Fig. 11, a diode may be placed to form a resonant circuit, the normalized input conductance taking the form of expression (27):

$$Gx \ll Yo \tag{25}$$

$$\omega Co \ll Yo \tag{26}$$

$$Gin \doteqdot \frac{Rs\,\omega x^2 Co^2}{Yo} \frac{1}{\sin^2\theta_1} = \frac{R}{X_o^2} \frac{1}{\sin^2\theta_1} \tag{27}$$

For the range where the expression (24) to (26) do not hold true, the errors resulting from the equation (27) become considerable. Whenever this happens, since the resonant circuit can be restated like that in Fig. 12, this can be taken as a modification of Fig. 5 (b). By substituting $\theta = \theta_2 - \theta_1$, and $B = \cot\theta_1$ in Fig. 7, the resonant conditions or Gin can be got. Furthermore, within the range where the expressions (24) to (26) are applicable, the equation (27) coincides remarkably well with the results obtained from Fig. 7.

Example of applications to parametric amplifier

Composition of resonant circuits

The structure of the resonator in a 2 Gc circulator-type parametric amplifier is shown in Fig. 13. The frequency ranges of signal (ω_1), pumping (ω_0) and idler (ω_2) are 1.7 - 2.3 Gc, 7.75 - 8.1 Gc and 6.17 - 5.75 Gc respectively. The signal frequency (ω_1) enters from a 50 ohm coaxial line and after passing through two capacitive metallic cylinders and a low-pass filter, is impressed onto the varactor diode MA-450E located within a 7 Gc band low-impedance waveguide. The resonant circuit of the signal frequency is composed of the varactor diode, two capacitive metallic cylinders and the lines connecting them, its equivalent circuit becoming like the one shown in Fig. 5. By making the structure of the resonant circuit like the one as stated above, θ in Fig. 5 can be made as desired by changing the distance between the diode and the metal cylinders while B in Fig. 5 can be chosen likewise by altering the distance between the two cylinders.

Now assuming the average capacitance of diode Co = 1 p F, series resistance Rs = 5 ohms and the lead inductance of diode L = 2 mμH, the reactance of diode at the frequency of 2 Gc becomes $j\left(\frac{1}{\omega Co} - \omega L\right) = j\,55$ ohm. Accordingly, the impedance of the diode normalized with the characteristic impedance of 50 ohm coaxial line becomes $X_o = 1.1 \quad R = 0.1$

Now in order to get the normalized input admittance of this resonator Gin = 0.2 at the input signal frequency, by substituting Gin/R = 0.2/0.1 = 2 and $X_o = 1.1$ in Fig. 7, $|B| = 1.8$ and $\theta = 75^\circ + n \times 180^\circ$ will be obtained (where n = 0,1,2.. and B>0). These values of B and θ are the values when the uniform 50 ohm coaxial line is used. However, in the resonator designed like that shown in Fig. 13, it is necessary to get the actual values of B and θ from Fig. 7 by adding to the above mentioned value of X_o the series reactance produced due to the discontinuity by the waveguide. Moreover to determine the actual resonant length from θ found in this manner, the equivalent length of low-pass filter must be taken into consideration.

Next, the resonant circuit of the idler frequency (ω_2) is composed of the section between the diode and the inductive window provided in the 7 Gc low-impedance waveguide. The shorting plate behind the diode is located at approximately a quarter wavelength of ω_2, and it does not affect the idler circuit much and is mainly used for matching pumping power. Therefore its equivalent circuit, just as before, is represented by the one shown in Fig. 5 (b). The minute adjustments of and B are carried out by means of a tuning screw set between the inductive window and the diode as well as by a screw set on the inductive window. In designing this circuit because of the obscure characteristic impedance of waveguide, there are

many factors which have to be determined experimentally, nevertheless the design chart shown in Fig. 7 is very useful in conducting such experiments.

For instance, when experimentally realizing Gin by making use of a capacitive stub with variable position, if B of the stub as well as its resonant length ℓ are actually measured, the values of B and ℓ which give the corresponding the same Gin by replacing the stub with an inductive window can be easily found from the chart of Fig. 7.

The pumping power (ω_0) is impressed to the diode through 10 Gc band waveguide. Since this 10 Gc waveguide is in the cut-off range against ω_1 and ω_2, there is no leakage.

Performance characteristics

Fig. 14 shows the input admittances of signal and idler circuits in no-pumping state for changes in the bias voltage of diode. These input admittances were obtained when getting the performance characteristic to be described below.

Although the fixed bias voltage under operating condition is -0.7 volt, as can be seen from the chart the bias voltage is approximately around -0.3 volt for the zero value of input susceptance. This indicates that the bias voltage corresponding to Co in the equation (6) is around -0.3 volt. Furthermore it is needless to say that the Co which gives the maximum figure of merit M depends on the variety of diode and coefficient of pumping (δ), but it is essential to make note of the fact it is also dependent on the resonant circuit of pumping frequency.

In other words, even though the bias voltage in pumping state varies at the frequency of ω_0, that the wave forms of pumping voltage gets the sine wave or the distorted form containing harmonics depends on the selection of the pumping resonant circuits. Therefore, the same diode and the same fixed bias voltages do not result the same values of Co and C_1. The 2 Gc parametric amplifier was composed of the Fig. 13 resonator adjusted as Fig. 14, 2 Gc circulators with the forward loss of about 0.4 db and the reverse loss of over 40 dB, and the 2 Gc receiver with the noise figure of about 8 dB. We have got the following performance; the overall noise figure 2.5 dB (1.8 dB for the parametric amplifier only), gain of parametric amplifier 16 dB, bandwidth 15 Mc, and pumping power approximately 30 mW.

The calculated value of noise figure arrived at by substituting in the equation (4) the numerical values obtained from the chart of Fig. 14, $G_1/G_{10} = 0.17$, $G_2/G_{20} = 1.5$ and $\omega_1/\omega_2 = 0.286$, coincides quite well with the actually measured value. Also by putting the afore-mentioned values G_1/G_{10} and G_2/G_{20} in the equations (5) and (12), the figure of merit M for the diode MA-450E is found to be 11.1 Gc.

Stability Figs. 15 and 16 are examples showing the fact that the characteristics of this parametric amplifier is stable against pumping variation, Fig. 15 giving the changes in gain and amplitude characteristics due to the variations in pumping power while Fig. 16 giving the changes in the similar characteristics owing to variations in pumping frequency.

In order to utilize a parametric amplifier as practical device, its gain sensitivity against the changes in frequency and power of pumping have to be dull as shown in Figs. 15 and 16. Now why such characteristics can be got will be investigated. The broad-banding against the pumping frequency is to be accomplished by making the bandwidths of both pumping and idler circuits broad. And the saturation characteristic of the gain against pumping power as shown in Fig. 15 can be considered to be due to the following two causes.

To begin with, the first cause will be explained. Along with an increase in the pumping power, the variation ratio of capacitance increases simultaneously augmenting Co. This increase in Co makes the tuning of both signal and idler circuits detuned toward the lower frequency. The shifting of the resonance of idler circuit toward the lower side means the increase in the gain at the somewhat toward the lower side from the center idler frequency, and this is equivalent to an increase in its gain at the higher side from the center frequency at the signal. Therefore with an increase in the pumping power, apparently the detuned signal and idler circuits produces saturation characteristic without any change in the frequency response of the gain.

The second cause to be considered is that when the pumping power is increased, the variation ratio of capacitance increases. However because of the subsequent increase in Co, the cut-off frequency ω_c gets lower with the consequence that M, the product of δ and ω_c, does not become so great that the gain remains unchanged.

The saturation characteristic of gain against pumping power, it seems, is caused by the composite actions of foregoing two causes, and this characteristic was positively verified by testing numerous varactor diodes MA-450E. Many parametric amplifiers stabilized in this manner are already being employed in Japan in the over-horizon communication systems having up to 60 channels for telephone. By the by, the following fact is given for the reader's information that a similar saturation characteristic can be produced when the amplitude characteristic is made flat by adding another cavity resonator to the idler circuit.

The stabilities of a parametric amplifier used in the over-horizon television transmission circuit used 900 Mc band are shown in Figs. 17 and 18. The structure of the cavity resonator of this parametric amplifier resembles that illustrated in Fig. 13, only an additional filter being provided on the idler. And the noise figure of the

amplifier is about 2.0 dB at the point of its maximum gain.

Example of applications to frequency multiplier

Requirement for the composition of circuit

The maximum efficiency (η) of frequency multiplier using a varactor diode depends on the cut-off frequency at the equivalent bias point of the diode, input and output frequencies and variation ratios of a capacitance. Furthermore this maximum efficiency is obtained when not only the input circuit is matched at the operating condition but also the coupling coefficients of input and output sides are equal. Assuming the equivalent circuit of frequency multiplier being alike to that of Fig. 1, there exists the following relationship between this efficiency (η) determined by the characteristic of diode and the coupling coefficients of circuit.[7,8,9]

$$G_1/G_{10} = G_2/G_{20} = G/Go \qquad (28)$$

$$\eta = \frac{1 - G/Go}{1 + G/Go} \qquad (29)$$

Moreover, in order to realize the maximum efficiency it is important to compose the resonant circuit such as the undesirable frequencies are not to be consumed in the series resistance Rs of the diode. For this reason, when the input and output circuits are observed from the diode at the undesired frequencies, they must be reactive, not possessing any resistive component, and in addition they must be far detuned from their resonant conditions.

Composition of resonant circuit

Fig. 19 illustrates the structure of resonator unit used in doubling 482 Mc (ω_1) to 964 Mc (ω_2). ω_1 passes through a re-entrant cylindrical filter with variable input coupling coefficient and is applied to the varactor diode placed in series between the input filter and another re-entrant cylindrical filter likewise with variable output coupling coefficient, that is nearly resonant at ω_2.

ω_2 generated by the diode passes through the output filter and comes out at the output terminal. All coaxial lines are 50 ohm.

The equivalent circuits for ω_1, ω_2 and undesired harmonic frequencies become respectively as shown at (a), (b) and (c) in Fig. 20. In these illustrations, ℓ_3 and ℓ_5 are the distance from either AA' or BB' to the point where the filters may be regarded as ideal transformers; and these can be found through the measurements of the admittance of the circuit from AA' or BB' with a dummy load attached at the external terminals by varying frequencies near ω_1 or ω_2 respectively. Also the impedance transformation ratios k_1 and k_2 can be obtained from the Smith chart at the same time. The symbol ℓ_4 represents the distance of the output filter at ω_1 from BB' to its equivalent short while ℓ_6 denotes the same of the input filter at ω_2 from AA' to its equivalent short plane. These values may be measured similarly. Besides, even if not only the input and output coupling coefficients K_1 and K_2 of the filters but also the tuning of all filters are adjusted for matching, there will be hardly any change in the equivalent shorting lengths ℓ_4 and ℓ_6.

The input conductance under cold state satisfying the equation (29), or the coupling coefficient between the diode and its external load, may be materialized by means of the graph of Fig. 9.

In case of ω_1 for instance, X_o in Fig. 9 has to be taken as shown below:

$$|X_o| = \left| -jX_o' + j\tan 2\pi\frac{\ell_2 + \ell_4}{\lambda_1} \right| \qquad (30)$$

Where, X_o' being the impedance of diode at the average capacitance. By use of the equation (30) and the graphs in Fig. 9, K and θ necessary in getting the desired coupling coefficient will be obtained. Thereby we can determine ℓ_1, but since the effects of undesired harmonic have to be taken into consideration, ℓ_1 must not be decided upon by its matching requirements alone. However even though the value of ℓ_1 is different from the afore-mentioned matching conditions, the θ that fulfils the resonant conditions can be had by detuning the input filter as stated previously. This was mentioned during the analysis of the fundamental resonant circuits. The same thing can be said for the determination of ℓ_2.

As the input and output filters serve as short-circuit paths for those frequencies excepting ω_1 and ω_2, the equivalent circuit for the undesired harmonics becomes like the one shown in Fig. 20 (c). Hence by properly selecting either ℓ_1 or ℓ_2 or the both of them, the circuit including a diode can be brought to a state far beyond the resonant conditions at undesired harmonics.

Performance

Fig. 21 shows the variations in 964 Mc output power against the changes in L under the conditions that with 482 Mc input power kept constant at 1W and also with ℓ_2 in Fig. 19 held fixed, the matching of the input and output circuits was being perfectly maintained. The matching of input and output was conducted both by turning the coupling loops and by detuning the filters. That the conversion efficiency becomes the maximum when $L = \ell_1 + \ell_2 = 130$ mm is due, it seems, to the fact that the consumption at the diode of the undesired harmonics component becomes the minimum. The output at this time was 500 mW, the conversion efficiency being 50%. And the fixed bias voltage was -1.2 V while the current

was + 1.1 mA. And MA-4287, 30 volt varactor, was used.

The admittances measured at a low level from the input and output terminals are shown in Fig. 22. The results obtained from the chart in Fig. 22 coincide remarkably well with compulations using Fig. 9.

Conclusions

The analyses of resonant cavities used in parametric amplifiers and frequency multipliers can be carried out under the two conditions; that if the bias voltage of a varactor diode is statically changed, only the input susceptance varies while the input conductance as seen from external terminals remains unaltered and that the input susceptance becomes zero at a certain impedance of the diode.

There are two fundamental resonant circuits analyzed according to these conditions, and the results obtained are represented in the accompanying design charts.

In a parametric amplifier or a frequency multiplier, if the characteristics of a diode as well as the operating frequency are given, the input conductance at cold sate that gives the optimum performance is determined. Therefore the electrical dimension of the component in a resonant circuit can be found by applying to these design charts the above-mentioned input conductance and the average capacitance and the series resistance of the diode. But in practical design, in order to make an actual circuit approach the fundamental circuit, both experiments and calculations are required to some extent.

The circuit composition methods for a 2 Gc parametric amplifier with noise factor of 2.5 dB, bandwidth of 15 Mc and gain of 16 dB and also for a frequency multiplier, which has efficiency of 50% and doubles 482 Mc to 964 Mc, are shown as examples of practical applications.

Acknowledgement

In concluding this paper I wish to thank Dr. M. Morita who gave instructive advices in carrying out the research, and also my special thanks are due Mr. K. Hata and Mr. K. Hayakawa who cooperated in conducting the experiments and investigations.

References

1. H. Heffner and G. Wade, "Gain, bandwidth, and noise characteristics of the variable-parameter amplifier" Journal of applied physics 29.9. P.1321-1331 (Sep. 1958)

2. Y. Kaito and K. Hata, "2000 Mc band parametric amplifier" Convention record of the I.E.C.E of Japan No. 1208 (July, 1960)

3. A. Uhlier Jr. "The potential of semiconductor diodes in high-frequency communications" Pro. I.R.E. 46. P.1099-1115 (June, 1958)

4. K.L. Kotzebue, "Optimum noise performance of parametric amplifier" Pro. I.R.E. 48. P.1324-1325 (July 1960)

5. K. Siegel "Comparative figure of merit for available varactor diodes" Pro. I.R.E. 49, P.809-810 (April 1961)

6. "Reference data for radio engineers" 4th edition P.576, International Telephone and Telegraph Corporation.

7. D.B. Leeson, S. Weinreb "Frequency multiplication with nonlinear capacitors - circuits analysis" Pro. I.R.E. 47, P.2076-2084. (Dec. 1959)

8. K. M. Johnson "Large signal analysis of a parametric harmonic generator" IRE Trans. MTT 8, P.525-532 (Sep. 1960)

9. T.M. Hyltin, K.L. Kotzebue "Solid-state microwave source from reactance diode harmonic generators" IRE Trans. MTT 9, P.73-78 (Jan. 1961)

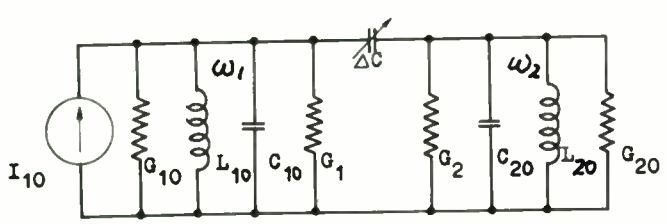


Fig. 1. An equivalent circuit of a parametric amplifier.

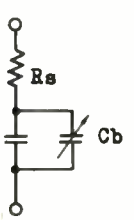


Fig. 2. An equivalent circuit of a varactor diode.

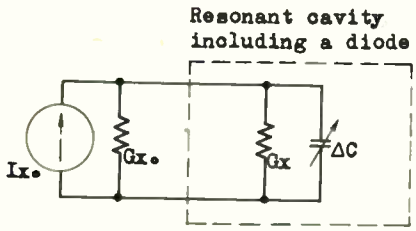


Fig. 3. An explanation of Fig. 1.

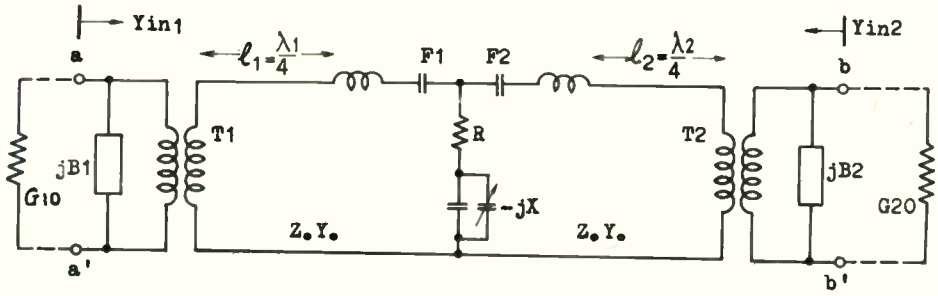


Fig. 4. An ideal circuit showing the relation between Fig. 1 and Fig. 2.

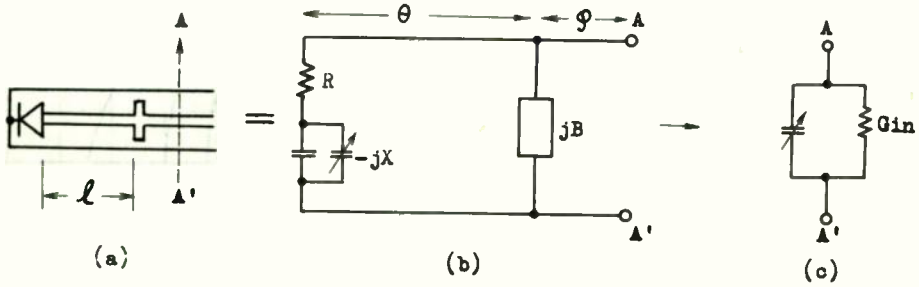


Fig. 5. A transmission line resonator using a susceptance.

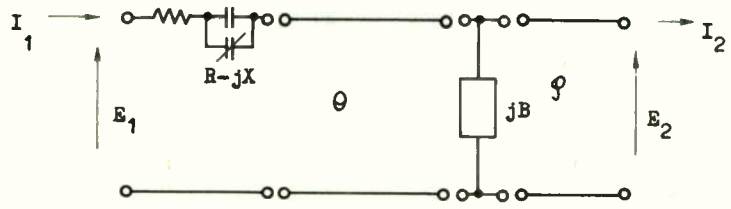


Fig. 6. The circuit for the calculation of input admittance in Fig. 5.

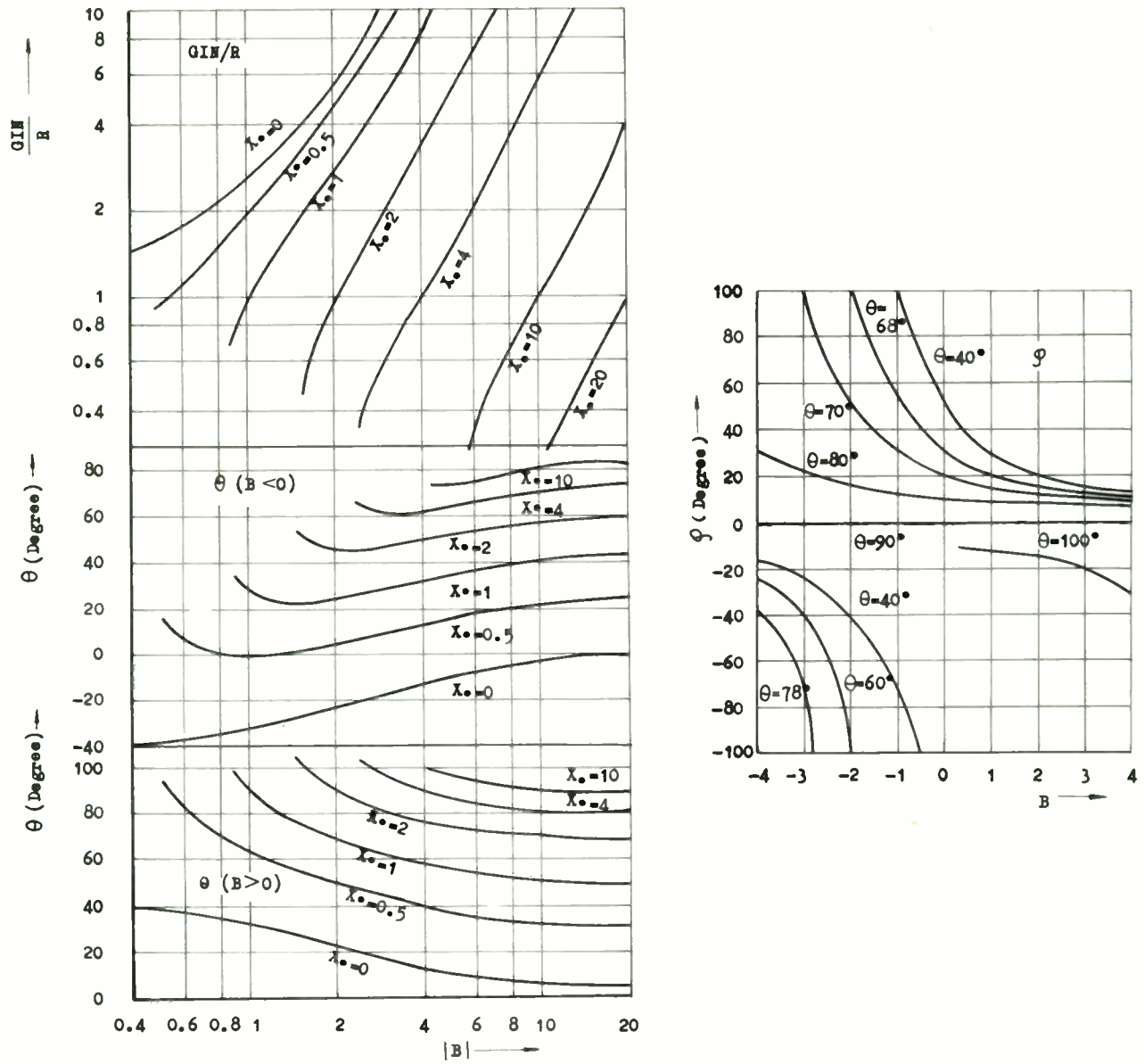


Fig. 7. Design chart of the transmission line resonator using a susceptance.

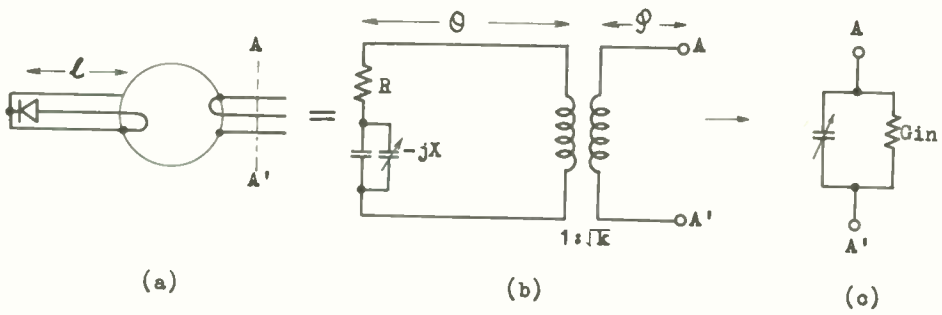


Fig. 8. A transmission line resonator using a filter.

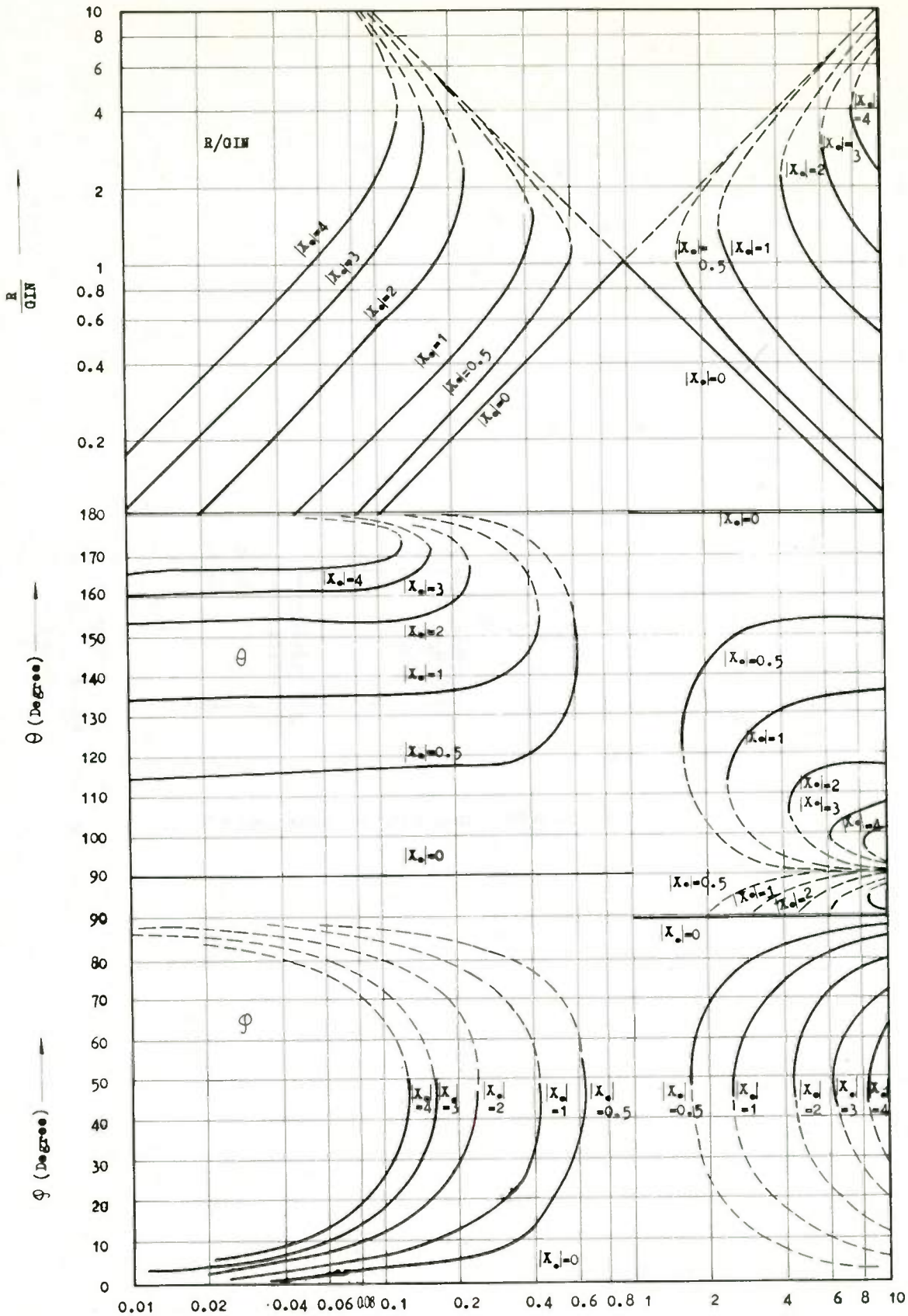


Fig. 9. The design chart of the transmission line resonator using a filter.

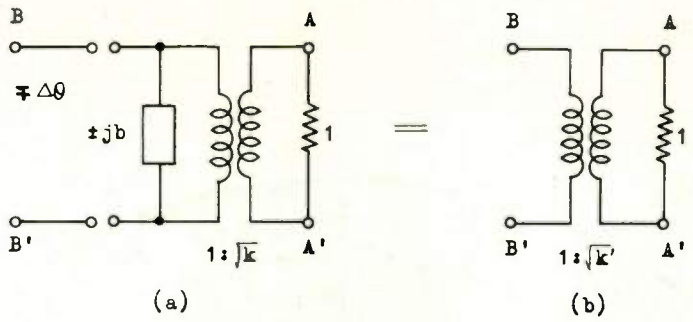


Fig. 10. An equivalent circuit of a detuned filter.

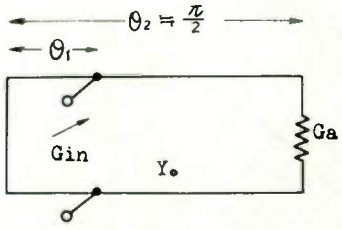


Fig. 11. An impedance transformer using a quarter wavelength resonator.

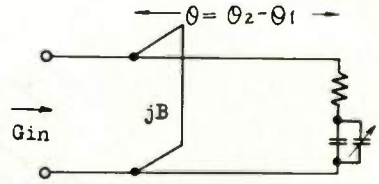


Fig. 12. A system using a quarter wavelength resonator.

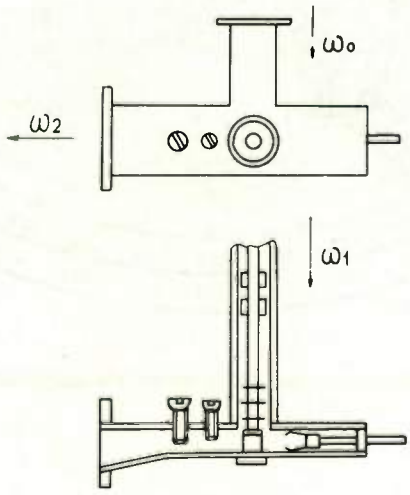


Fig. 13. The resonator of the 2GC parametric amplifier.

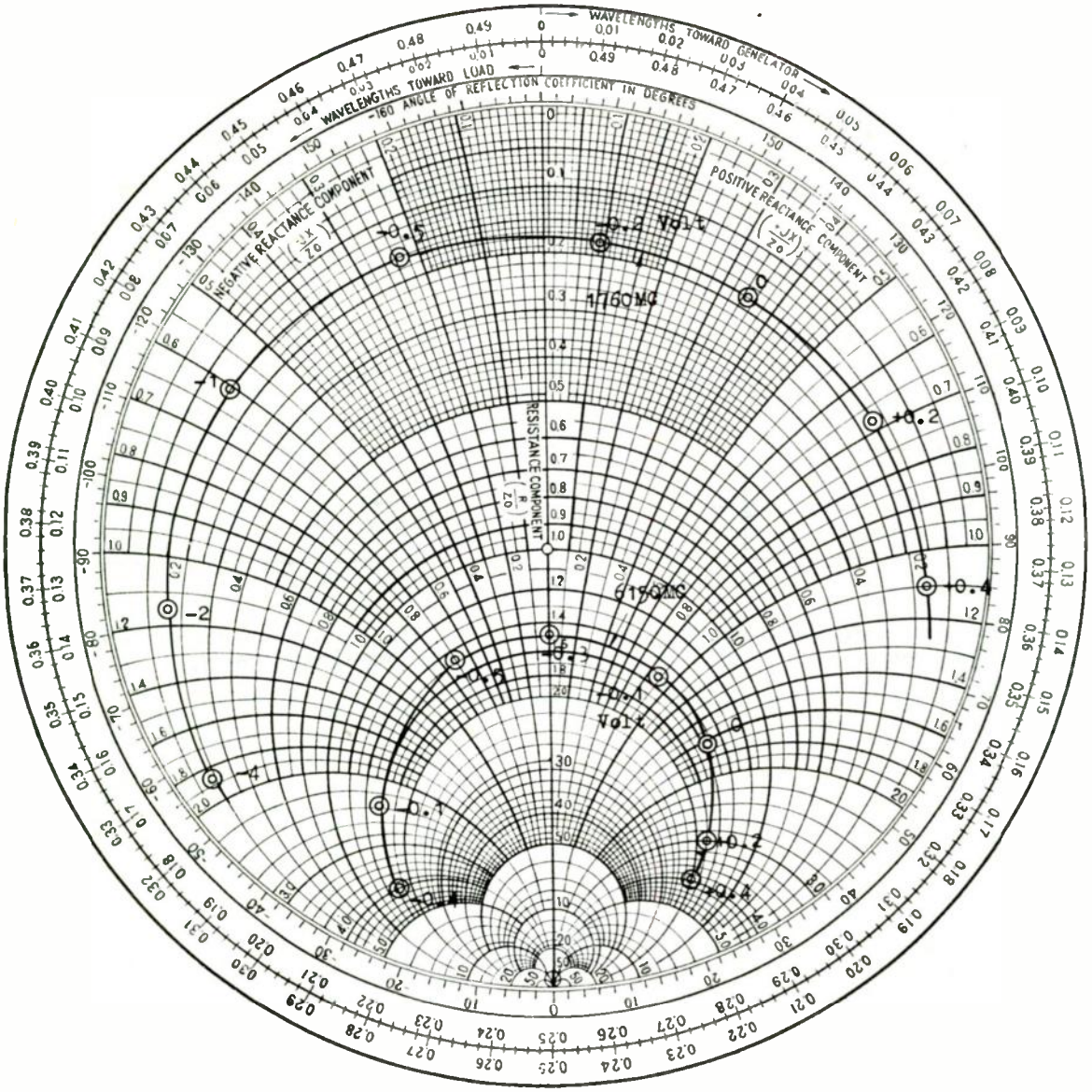


Fig. 14. The input admittance of the 2GC parametric amplifier in no pumping power.

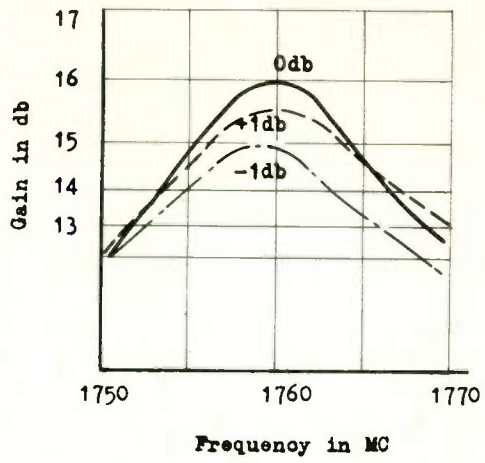


Fig. 15. The variation of amplitude response due to the pumping power in the 2GC parametric amplifier.

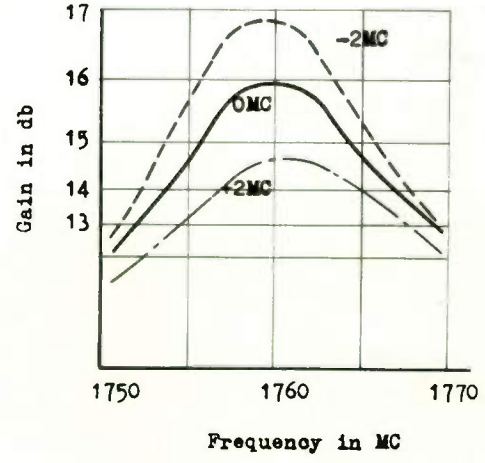


Fig. 16. The variation of amplitude response due to the pumping frequency in the 2GC parametric amplifier.

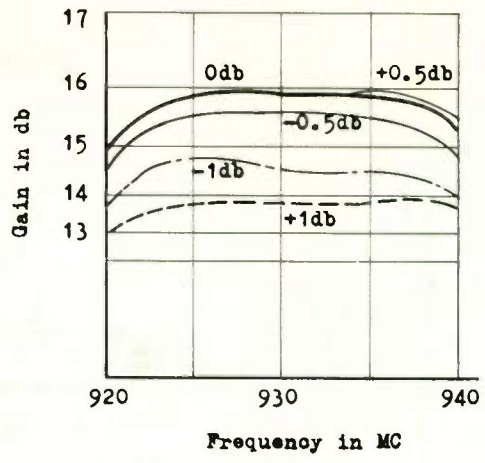


Fig. 17. The variation of amplitude response due to the pumping power in the 900 Mc parametric amplifier.

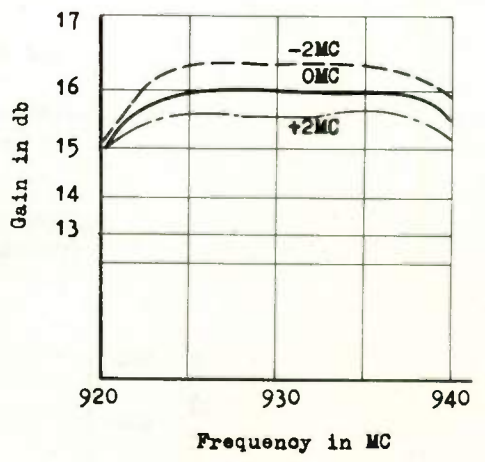


Fig. 18. The variation of amplitude response due to the pumping frequency in the 900 Mc parametric amplifier.

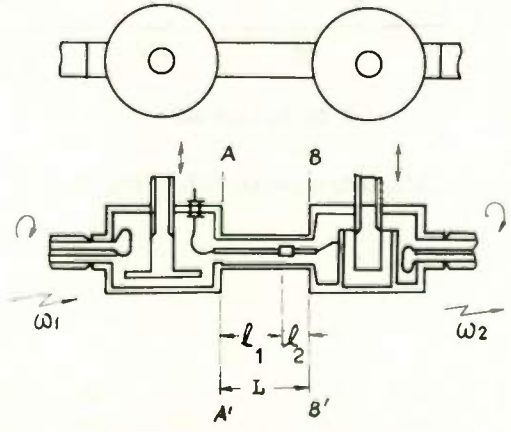


Fig. 19. The resonator of the frequency multiplier.

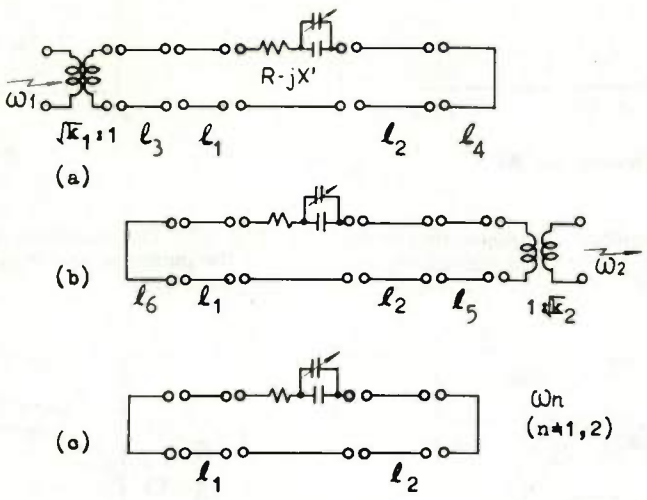


Fig. 20. The equivalent circuit of the frequency multiplier.

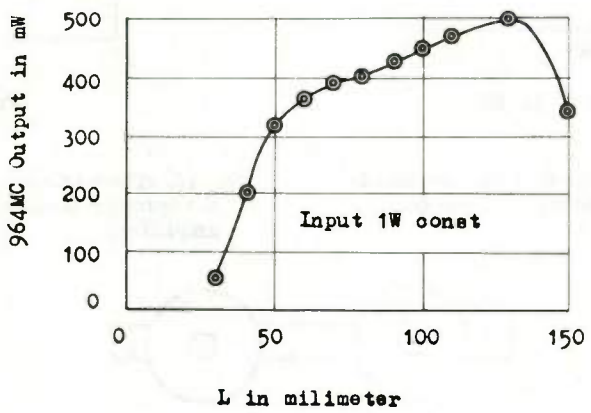


Fig. 21. Output power VSL in Fig. 19.

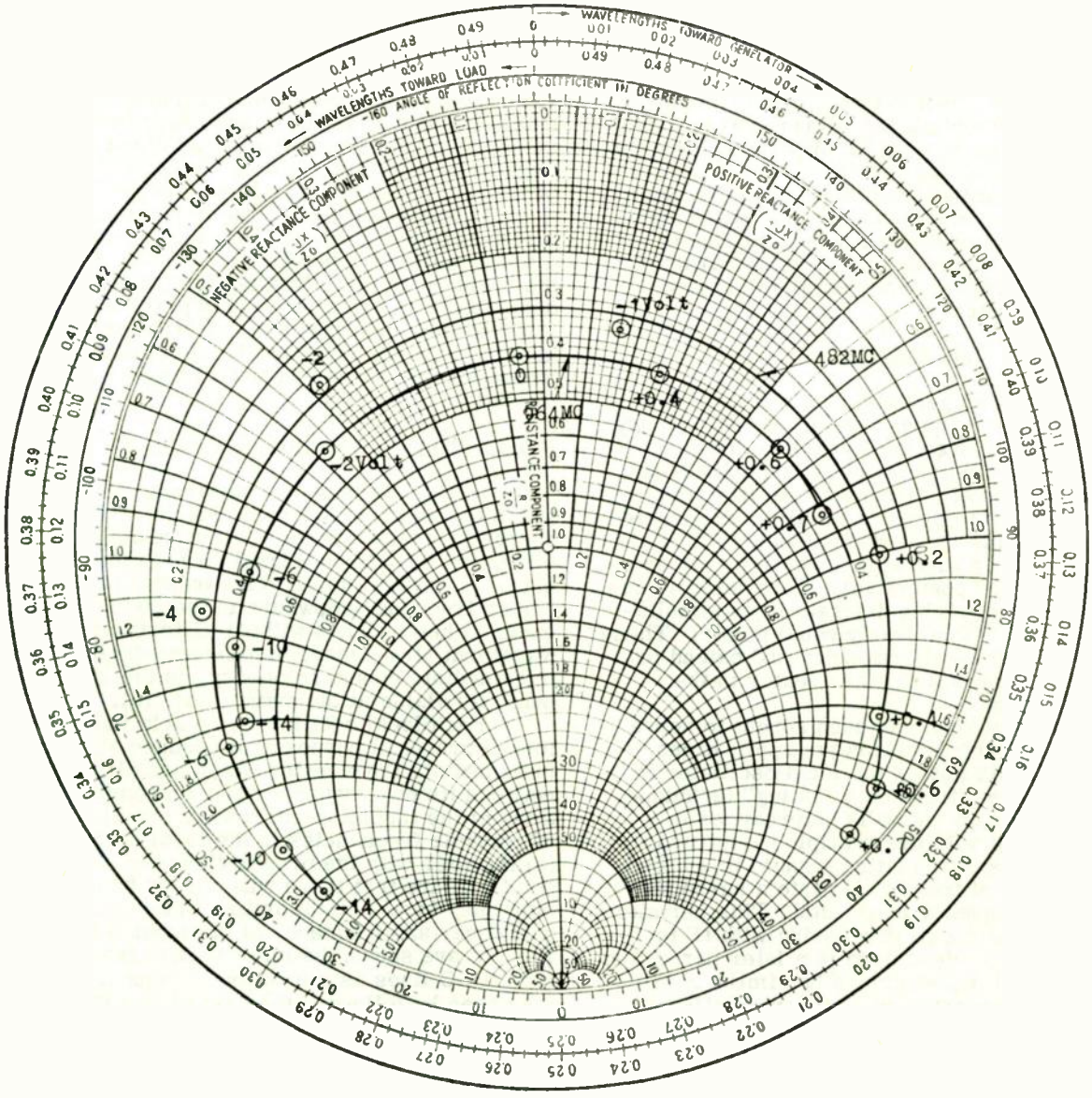


Fig. 22. The input admittance of the frequency multiplier in small signal condition.

MULTIPLE HARMONIC LOCAL OSCILLATOR SOURCE

Josef Gartner
Seymour Sutkin
Air Armament Division
Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, L.I., New York

A method is described of generating a microwave signal whose power spectrum consists of a large group of stable discrete lines of nearly equal power, equispaced in frequency, and extending over an octave or more in the microwave frequency region. The technique uses a low-frequency crystal-controlled oscillator feeding a varactor diode mounted in a modified microwave crystal holder to produce the desired spectrum and output levels. Design considerations which lead to optimum transfer of low-frequency power to the varactor and maximum broadband harmonic microwave power are presented.

Introduction

The harmonic generation of a microwave signal of a single frequency from a low-frequency generator has received considerable attention elsewhere.[1,2] A less commonly described, but important version of harmonic generation is one which is required to produce multiple frequencies in its output. This paper describes a method of generating a microwave signal whose power spectrum (i.e., power content versus frequency) consists of a group of stable discrete lines simultaneously present and of nearly equal power. These lines are equispaced in frequency and extend over a frequency range of 2 to 1 or greater in the microwave region.

Such a waveform has been sought to serve in a number of recent microwave applications, one of which is its use as the local oscillator in a broadbanding or "frequency folding" technique. In this case an input signal is heterodyned in a single component with the multiple harmonic local oscillator source, translating signals in a broad microwave band of frequencies into a narrower intermediate frequency bandwidth. This permits amplification and signal processing with IF techniques. As will be described in a later paragraph, the local oscillator source can be similarly utilized in an up-converter as well, translating a single IF signal into higher center frequencies extending over a broad portion of the microwave spectrum. In addition to this application, a multiple harmonic waveform at microwave frequencies can be used in communication systems. For example, the required stability of one part in 10^6 exhibited by the discrete frequencies of the local oscillator source permit frequency coding of a communications transmitter chain in a prearranged way by following the harmonic generator with a narrow-band electronically tunable amplifier (which exists in current microwave technology).

Because of the relationship that exists between the frequency description of such a frequency group spectrum and its time-waveform description as a narrow pulse train, a multiple harmonic signal might also form the basis of an extraordinary microwave short-pulse generator. These, as well as other applications, including stable frequency markers and other frequency measuring devices, all indicate the value of a convenient multiple harmonic microwave source of useable power level and low noise content.

It was in support of the first application mentioned above, its use as a local oscillator in a broadband frequency conversion circuit, that the varactor multiple harmonic circuit to be described was examined. It is perhaps worthwhile to first describe the operation of a mixer using a multiple harmonic local oscillator and to indicate some of the characteristics required of the latter.

Broadband Conversion by Frequency Folding

In Fig. 1a is shown an arrangement in which the multiple harmonic local oscillator source (a) feeds a broadband microwave crystal mixer (b). The input signal, (c), in this case a signal of several megacycles bandwidth with center frequency anywhere in the 2000 - 4000 Mc band, is heterodyned without distortion of its waveform, frequency, or phase modulation to an intermediate center frequency, in the 20 - 30 Mc band, available at the output port, (d).

Each local oscillator line acts independently for all practical purposes to produce 20 - 30 Mc intermediate frequencies which are differences between its own frequency and appropriate portions of the incoming microwave signal. The use of multiple frequencies in the local oscillator at (a) together with the proper choice of IF band limits insure that there will always be one, and only one, local oscillator line in the mixer which will cover any frequency of the input band. In Fig. 1c are shown the intermediate frequencies for all LO lines as a function of input frequency. It is seen by following the dashed line RST that a signal at 2048 Mc, for example, is converted to 28 Mc which is the first order difference between 2048 Mc and the LO line at 2020 Mc. The heavy broken line in Fig. 1c shows how the r-f band from 2000 to 4000 Mc is "folded" into an IF bandwidth between 20 and 30 Mc. In order to reduce direct leakage of the local oscillator intermodulation products into the output, a balanced mixer configuration may be used. It is found best to operate with an input signal exceeding the LO power per line by at least 10 db. It is possible to obtain IF signals in this manner which are less than 10 db below the local oscillator input power per line.

In Fig. 1b is shown the arrangement in

which the multiple harmonic local oscillator source supplies an up-converter modulator in which an IF signal (20 - 30 Mc) is simultaneously translated to a large array of center frequencies in the 2000 - 4000 Mc band. This reverse of the previously described down-conversion process can be seen, from the typical process LMN in Fig. 1c, to produce many output frequencies for the one input IF frequency.

Both types of conversion have been tested successfully for applications including an exact frequency measurement and analysis device using the down-converter, and a highly stable repeater with the combination of down and up-converters. A stable multiple harmonic local oscillator source was needed here with a separation frequency of 20 Mc in the frequency range of 2000 - 4000 Mc. Since (1) efficiency was not of the first importance and (2) high power output per line was not as important as frequency stability of all the lines (1 part in 10^6) and uniformity of output ($\pm$ 3 db unequalized), then only -40 dbm/line was accepted as a goal, although considerations leading to more efficient operation are useful.

Among alternative approaches considered to implement such a signal source were frequency modulation of a CW oscillator, generation of ultra short pulses by amplitude modulation of a CW source and use of a nonlinear semiconductor element to be driven by the desired separation frequency and if necessary, some of its multiples. Primary consideration was given to the last approach because of its compactness and simplicity.

Nonlinear resistance, i.e., conventional microwave diode elements were tried first. These produced fair results but very little output power per line. The use of mixed frequencies to drive these elements and to boost output power led to several drawbacks, one being a pronounced and unacceptable variation of power per line over the output spectrum. In addition to its expense, the use of a second drive source also led to stability problems unless it was phase-locked to the fundamental oscillator. Attempts to increase this output power led to the use of the relatively less lossy and higher voltage capability varactor elements for operation in a broadband circuit to be described below.

Background and Operation of the Multiple Harmonic Generator

Varactor diodes are p-n junctions designed to act as nonlinear capacitors. They are low resistance, low capacitance devices and are capable of operation at extremely high frequencies. Unlike the previous microwave diode types, varactors can handle higher input power, and some types are made with substantial reverse breakdown voltages. Since the resistive loss of the varactor under comparable conditions is lower, it is therefore reasonable to expect that it should be a more efficient harmonic generator than the previously used nonlinear resistive diode.

Varactor harmonic generators of low orders of multiplication involving a single input frequency and a single output frequency have been developed with very high efficiencies when both input and output circuits were tuned. Theoretically, with the proper shape of the nonlinear characteristic, unity efficiency of conversion with a nonlinear capacitor is possible. Of particular interest here was the goal of generating a high order spectrum of harmonics for use as a local oscillator source. There is evidence that in the case of broadband frequency multiplication, where all harmonics are generated and no attempt is made to economize by reflecting the unwanted ones, the maximum attainable efficiency drops off at $1/n^2$ for ideal nonlinear capacitors and $1/n^4$ for nonlinear resistors, n being the order of the harmonic.[3]

The varactor (Fig. 2) considered here is the graded p-n junction type having a behavior at high frequencies described as being equivalent to a small resistance, R_s, in series with a depletion layer capacitance, C, whose value is a function of the voltage across the unit. The incremental capacitance is given by:

$$C = \frac{dq}{dV} = \frac{C_o}{\sqrt[3]{1 - V/\varphi}} \qquad (1)$$

where V is the applied voltage across the varactor, considered positive when applied so as to induce forward conduction, q is the charge stored in the varactor, φ is about 0.5 volts for silicon diodes at room temperature and C_o is the capacitance defined at V = 0 bias. For positive (conducting) bias voltages approaching φ, the value of C increases without limit, indicating the ability of the varactor to conduct a current that is limited only by the external circuit impedance and the applied voltage. For negative bias the capacitance decreases to a value, C_{min}, for a reverse voltage just short of breakdown at V_b volts.

When an ideal varactor is put in series with a voltage generator and a series resistance, part of which may be considered the load resistance and part the internal resistance of the generator, the current exhibits a waveform similar to that indicated in Fig. 3b. During most of the interval from a to c the varactor is biased for forward conduction and, neglecting φ and the small internal series resistance, the current through the circuit is given by:

$$i(t) = \frac{E \sin\omega t - V_E}{R_g + R_\ell} \qquad (2)$$

The charge in the capacitor increases between times a and b, at which latter point the net impressed voltage becomes negative. From b to c the stored charge in the varactor decreases to zero as the varactor voltage decreases slightly from φ to zero volts. The point c at which the discharge current changes rapidly to zero is determined by the requirement that the entire charge accepted by the varactor, i.e., numerically the area under the positive current curve from a to b, must equal the charge returned to the circuit, or the area above the negative current curve from

b to c. The average current through the varactor is ideally zero. The voltage V across the varactor itself is then one which follows the impressed voltage except for the time interval from a to c, during which it stays between zero and φ volts.

By comparison, a nonlinear resistance produces only a positive pulse of current, similar to the fractional sine wave shown between a and b in Fig. 3b, its amplitude limited by its own forward resistance as well as the circuit resistance. Actual varactors also rectify current to some extent because of recombination effects, and exhibit substantial charge storage capability only at high frequencies, i.e., above several megacycles.

Leenov and Uhlir have compared the saturation conversion loss of ideal nonlinear resistances and capacitances used as broadband harmonic generators by obtaining the Fourier coefficients of the two kinds of current wave forms characteristic of these elements. It was found that the nonlinear capacitor was superior, especially at the higher harmonics. An untuned, broadband, harmonic generator using varactors and demonstrating this behavior may be obtained by approaching the series circuit suggested above and arranging the d-c bias and a-c drive voltage for maximum harmonic output.

Technique and Experimental Results

The technique described here utilizes a single crystal controlled low frequency oscillator and a single nonlinear element held in a specially constructed coaxial holder and connected as shown in Fig. 4a. The varactor originally used as the nonlinear element was the Microwave Associates 460A which had a power rating of 0.3 watts and a φ of approximately 0.5 volts. The capacitance at zero bias was specified between 2.4 and 6.0 $\mu\mu$f and the breakdown voltage was V_b = -6 volts. The indicated cutoff frequency of 30 kmc assured that the capacitive reactance at minimum capacitance was large compared to the series resistance at the frequencies of interest in this harmonic generator.

The physical outline of the varactor element was identical to that of the conventional microwave crystal cartridge. The varactor was mounted as part of the center conductor in a coaxial holder, Fig. 4b, containing an inductive D.C. return on the microwave output side, and a shunt capacitance from center conductor to ground on the low frequency input side of the varactor. The object of this configuration was to approach the circuit described in the preceding section, in which the high harmonics of the resultant series current waveform passed thru the 50-ohm microwave load to produce the required uniform array of microwave frequencies.

A crystal controlled 20 Mc oscillator of the Pierce configuration was used to supply the fundamental drive power. The output of the oscillator was fed to a 6AN5 power amplifier stage which delivered 20Mc to the varactor thru an impedance transformer. This circuit is shown in Fig. 4c. The capacitor built into the varactor holder was important in the efficient transfer of 20 Mc power from the source because it was used as part of the tuning capacity for the 6AN5 output transformer. It was parallel-resonated with the output tap of the impedance transformer to maximize the voltage across the varactor. A variable capacitor was available to trim the circuit for this purpose.

The frequency stability of the 20 Mc oscillator was controlled by the stability of the crystal. Long term frequency stability of 0.0001 per cent was obtainable with the use of a temperature-controlled crystal oven.

The d-c return at the microwave side of the varactor holder, a fine wire placed across the dielectric portion of the coaxial line, provided a low impedance to ground to complete the low frequency current path while offering a high impedance of several hundred ohms to the desired output frequencies. In this way maximum current pulses could be delivered to the varactor from the input generator while the higher harmonics of current could be conveniently absorbed in the load.

By virtue of the shunt capacitor (approximately 40$\mu\mu$f) across the coaxial line, the varactor was located near a broadband voltage null in the microwave line, the object here being to minimize frequency sensitive reflections and standing waves in the coaxial line which would enhance undesirable variations of output power with frequency. To accomplish a uniform output, the holder was designed so that the capacity at the low frequency input side of the varactor was made as large as would be consistent with the input drive requirements. It was found that this capacitor should be physically located cylindrically about the broad base of the varactor cartridge as shown in Fig. 4b for the smallest fine structure variations of the output frequencies.

In the 2-4 kmc band the circuit impedance seen by the varactor was largely determined by the microwave load, while at 20 Mc the plate resistance of the 6AN5, as seen through the matching transformer, dominated. The 6AN5 amplifier tube can deliver about 1 watt at 20 Mc. The plate circuit was tuned to resonance and the transformer output circuit was tuned for maximum voltage with the varactor and coaxial holder attached. The varactor drive was kept at a low level when the tuning adjustment was made so as not to load the tuned circuit.

The use of an externally applied d-c bias voltage across the varactor was considered and attempts were made to provide such a bias in order to optimize power output at the frequencies of interest. Initial experiments indicated that output power at the fiftieth harmonic could be greatly affected by the d-c bias for a given generator peak-to-peak voltage, but that a broad maximum output existed for high drive levels for a range of negative bias voltages close to zero, see Fig. 5. It was further found that the self-bias, established by the partial rectification in the varactor of the applied 20 Mc voltage produced a bias voltage which was consistent with maximum output power, see Fig. 6. A comparison of Figs. 5 and 6 show that externally applied d-c bias was not necessary and self-bias was ultimately used.

A measurement of the harmonic power per line absorbed into a matched 50-ohm line was made by a substitution method using an r-f attenuator and narrowband receiver. The result in Fig. 7 shows a uniformity of harmonic power per line within ± 3db in the 2000 - 4000 Mc range at an average level of -40 dbm. Some varactor selection was necessary since the average output power level could vary by several db among different varactors.

Since this device generated harmonic power at frequencies below 2000 Mc and above 4000 Mc as well, it was necessary to use filtering in the local oscillator application described earlier.

In all cases when compared in the identical circuit with nonlinear resistance diodes such as the 1N21, the varactors were seen to produce between 15 db and 30 db more saturated output power than the resistance diode. The 1N21 output fell off at a faster rate as the harmonic number increased. The result is shown in Fig. 8.

Measurements were repeated for a number of different varactor types in the same circuit as described above. Measurements of harmonic output power versus drive for a 30 volt breakdown series MA4283 varactor are shown in Fig. 9. These varactors saturate at higher levels of output power than the MA460A, a 6-volt breakdown unit. With increasingly higher power varactors it therefore appears possible to build a multiple local oscillator source which will be capable of delivering more power per harmonic and over a frequency spectrum limited by the cutoff frequency of the varactor used.

Conclusion

A low-power, multiple harmonic, local oscillator source for use in a broadband frequency-folding technique has been constructed and found to have met the electrical requirements set for it. Varactor elements which are coming into increasing use are well suited as a basis of such a device and are superior to nonlinear resistance diodes. A highly stable, low noise, uniformly spaced line spectrum of nearly constant output power per line is easily attainable at levels of -40 dbm and higher over a broad range of microwave frequencies, by the use of a varactor held in a simple coaxial holder and driven by a low-frequency crystal-controlled oscillator.

References

[1] T. M. Hyltin and K. L. Kotzebue, "A Solid State Microwave Source from Reactance Diode Harmonic Generators," I.R.E. Transactions on Microwave Theory and Techniques, pp. 73-78, January 1961.

[2] D. Leeson and S. Weinreb, "Frequency Multiplication with Nonlinear Capacitors - A Circuit Analysis," Proceedings of the I.R.E., Vol. 47, pp. 2076-2084, December 1959.

[3] D. Leenov and A. Uhlir, Jr., "Generation of Harmonics and Subharmonics at Microwave Frequencies with Junction Diodes," Proceedings of the I.R.E., Vol. 47, pp. 1724-1729, October 1959.

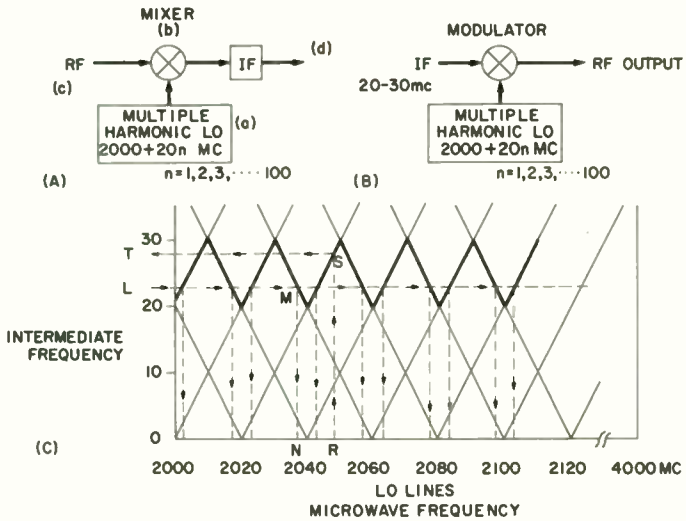
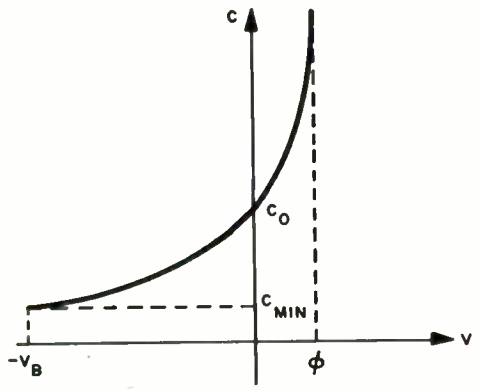
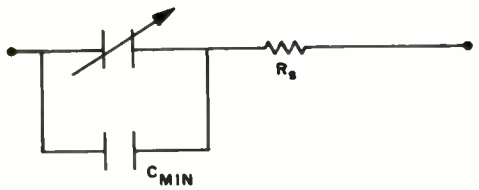


FIG. 1 BROADBAND UP AND DOWN CONVERSION USING THE MULTIPLE HARMONIC LOCAL OSCILLATOR

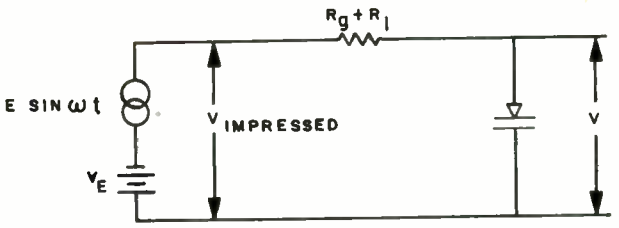


(A) CAPACITANCE VS. VOLTAGE

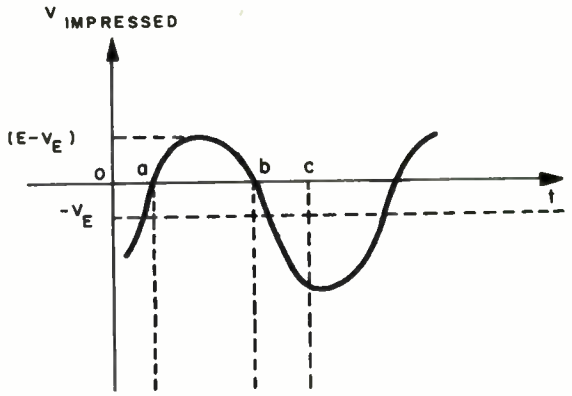


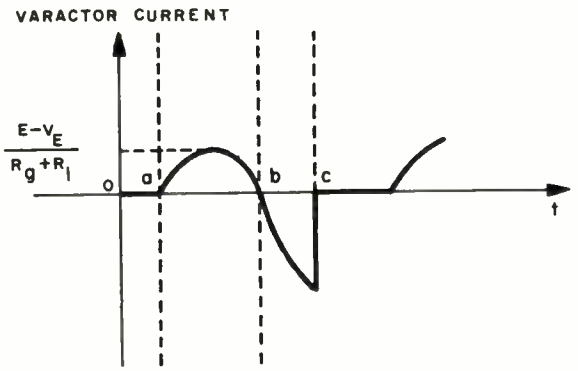
(B) HIGH-FREQUENCY EQUIVALENT CIRCUIT

FIG. 2 VARACTOR CHARACTERISTICS



(A) SERIES CIRCUIT





(B) VOLTAGE AND CURRENT WAVEFORMS ASSUMING $\phi=0$

FIG. 3 VARACTOR CURRENT WAVEFORM

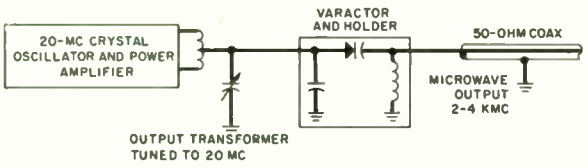


FIG. 4(A) MULTIPLE HARMONIC GENERATOR

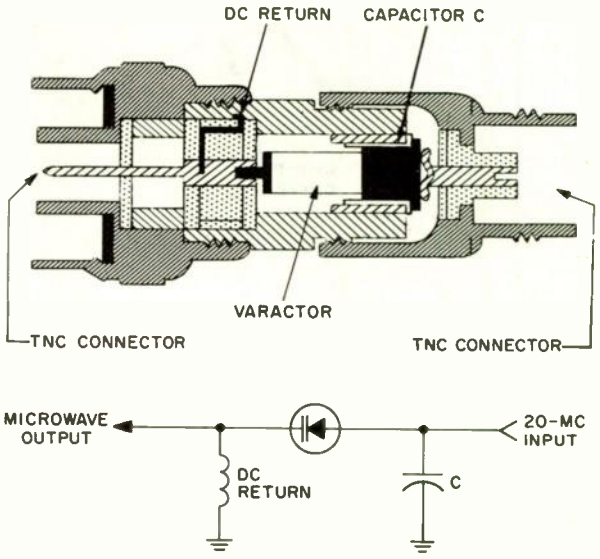


FIG. 4(B) VARACTOR HOLDER

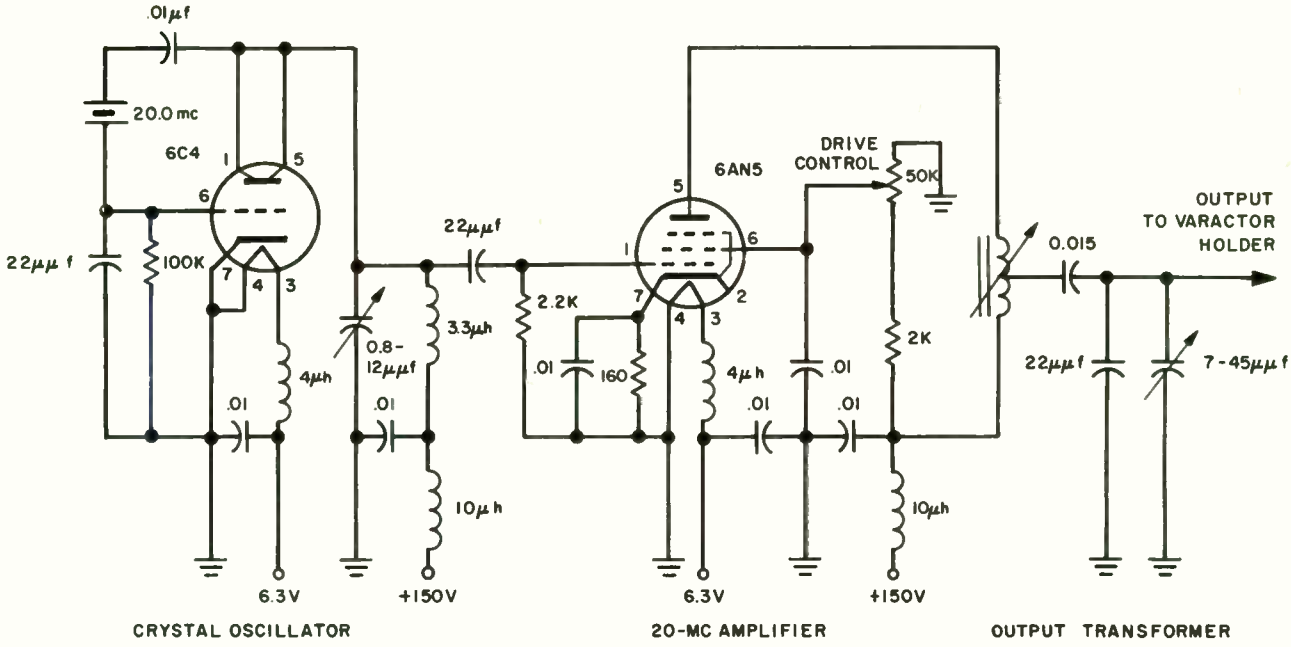


FIG. 4(C) 20-MC CRYSTAL OSCILLATOR AND POWER AMPLIFIER

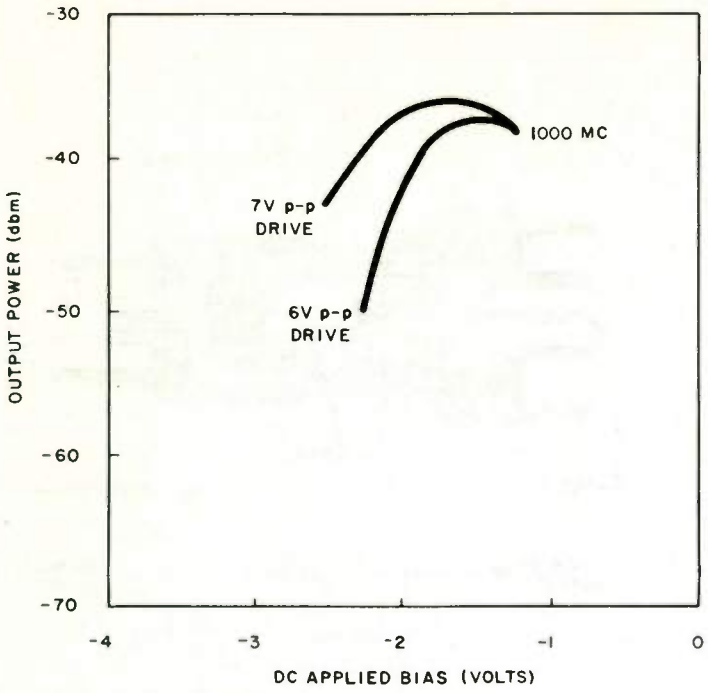


FIG. 5 OUTPUT POWER VS. BIAS AND DRIVE

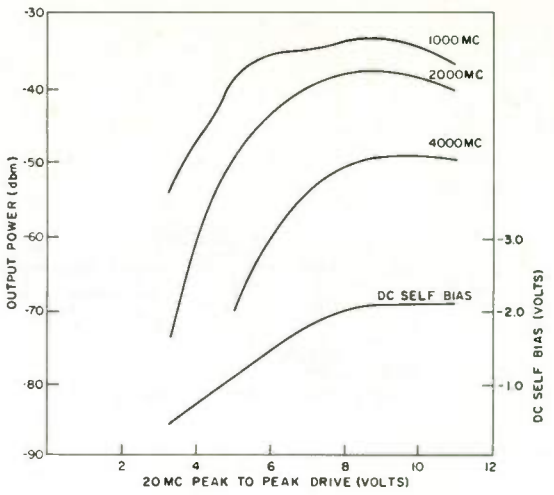


FIG. 6 OUTPUT POWER AND SELF BIAS VS. INPUT VOLTAGE

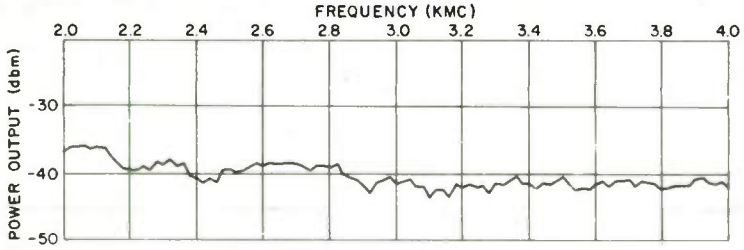


FIG. 7 VARACTOR POWER OUTPUT VS. FREQUENCY
LINE SPECTRUM SPACED 20MC APART

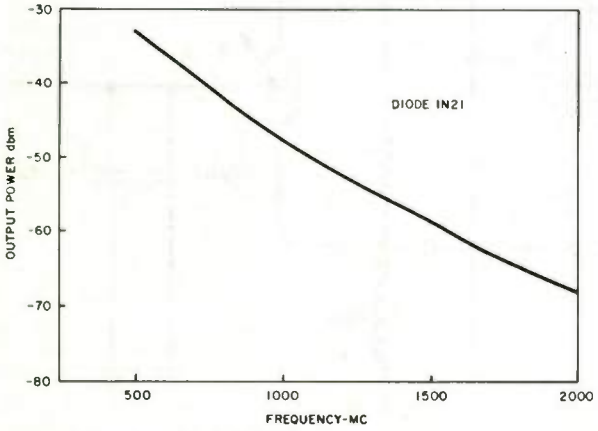
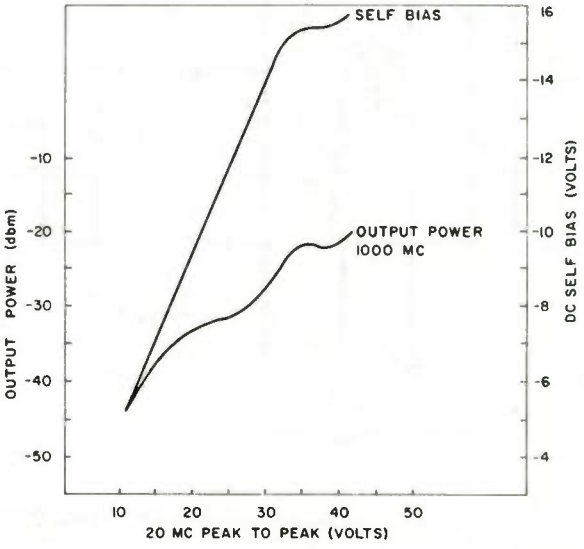


FIG. 8 HARMONIC OUTPUT POWER VS FREQUENCY



IG. 9 OUTPUT POWER VS. DRIVE AND SELF BIAS

SYNTHETIC RUBY FOR MASER APPLICATION

R. L. Hutcheson
Engineering Supervisor
Linde Company
Division of Union Carbide Corporation
4120 Kennedy Avenue
East Chicago, Indiana

Abstract

Synthetic ruby, which is single crystal aluminum oxide doped with varying small percentages of chrome oxide, is used for both optical and microwave maser. This material is primarily produced by the flame fusion (Verneuil) crystal growth technique. Recent developments in growth processes have produced materials for optical maser application in which threshold powers have been lowered by a factor of five and that have linewidths as low as 0.08 cm^{-1}. In addition, modified crystal geometries have been produced which lower the required threshold power by a factor of fifty compared to the original rods tested. This paper describes the flame fusion growth technique for synthetic ruby with primary emphasis on optical maser (laser) material. Included are physical properties of the material pertinent to optical maser application.

Introduction

Synthetic ruby is single crystal aluminum oxide in which a small percentage of chromium ions have been substituted for the aluminum ions. This material has found application in both optical and microwave MASERS*. Townes,[1] Bloemberger[2] and others first suggested the use of solid state materials for microwave masers in 1956. Geusic,[3] Kikuchi[4] and others[5, 6] investigated the paramagnetic properties of ruby for microwave masers. Schawlow and Townes[7] further proposed the application of solid state devices for stimulated emission at a higher frequency. In 1959, Maiman[8] observed this phenomenon in ruby at a wave length of 6943A°. Collins, Schawlow, Nelson[9] and others at Bell Laboratories confirmed the coherence and directionality of the stimulated light emission.

*Microwave amplification by stimulated emission of radiation now used to refer to all types of amplification by stimulated emission.

Ruby has been grown successfully by several techniques including flame fusion, hydrothermal, flux, melts and vapor deposition. The flame fusion process is the predominant growth procedure for ruby being used commercially both in the United States and Europe with an annual world-wide production of about sixty tons. The process invented by Verneuil[10] in 1904 is described mechanically in Figure 1. Powder of the proper particle size and composition is contained in the hopper (A). This powder is dispensed through a mechanically tapped screen (B) into a cone. From the cone, oxygen gas carries the powder through a post mixed oxygen-hydrogen burner (C). The powder is heated in the flame to a near molten state and then deposited on the molten cap of a seed crystal (D). The furnace chamber (E) is designed for a continuous heat leak from the seed to the furnace and atmosphere so that there is a continuous solidification of the material from the melt onto to seed. The crystal is then lowered to maintain a constant distance between the molten cap and the burner tip. In the case of both white sapphire and ruby, growth rates of from 4 to 400 grams/hour have been successfully used.

Figure 2 shows some of the sapphire and ruby crystals grown by the flame fusion technique and modifications thereof. Prior to 1954 the largest crystals available are as shown in area A and the rods as shown in area B. In 1954, the Drost-Kebler[11] disk growth process (Figure 3a) was introduced, thereby allowing the growth of crystals in disk form shown in area C of Figure 2. In 1957, the Kebler-Hutcheson[12] dome growth process (Figure 3b) was introduced allowing the growth of dome shaped crystals up to 5 inches in diameter and weighing 3 pounds, as shown in area D of Figure 2. In 1961, the Linde Company

modified the conventional flame fusion growth process, as shown in Figure 3 c, and was thus successful in the growth of crystals in the conventional carrot shaped form that were approximately three times as large as those grown previously. The larger carrot shaped crystals or boules are as shown in area E of Figure 2.

The above described crystals were intended primarily for such uses as infrared windows, pressure cells, wear applications, high temperature furnace components and electronic applications. They are not high perfection crystals and contain such defects as lineage or low angle grain boundaries, twinning, slip planes, microscopic bubbles, and some powder inclusions. It is the purpose of this paper to describe some of the recent work on flame fusion ruby to improve its properties for maser application.

Discussion

The flame fusion crystal growth process has several advantages over other techniques for the growth of high melting point crystals such as ruby. Among these advantages are: (1) the process requires no crucible since the material holds its own melt; (2) the equipment is basically simple in design and consists only of a powder dispenser, a burner, a furnace chamber, a lowering mechanism and a supply of clean oxygen and hydrogen; (3) the source of heat energy is clean and thus does not contaminate the material being grown; (4) the growth rates for commercial grade crystals are fairly high and thus economical; (5) the process allows fairly close control of the resulting crystal geometry; and (6) large crystals can be grown. The disadvantages of the conventional flame fusion process are: (1) there is a relatively high thermal gradient in the crystal during growth and consequently the crystal is subjected to high thermal stresses; (2) only crystals resistant to oxidation or reduction at their melting point can be grown; (3) for the alumina lattice only doping materials with a low vapor pressure at 2000°C can be used; (4) high purity powders must be used; (5) powder feed must be relatively uniform and accurate; and (6) doping material burnout is nonuniform.

The predominant defects in ruby which affect maser action as reported by Schawlow,[13] Collins[14] and others are strain crystal inhomogeneities, particularly lineage, bubbles, and nonuniform chromium concentration, and C-axis wander. The strain in crystals comes from three sources. Lattice strain results from the introduction into the lattice of larger chromium ions in the order of 10^{17} to 10^{20} ions/cm^3 and other impurities in the order of 10^{18} ions per cubic centimeter. Structural strain is the result of the dislocations and voids occurring during the growth process as evidenced by the etch pit density of 10^6/ sq. cm on the basal planes of conventional material. Thermal strain is the result of the thermal gradients in the growth process. Lattice strain is inherent in crystals and can be lowered only by reducing the total impurities in the crystal. Structure defects can be reduced by controlling the growth direction and the growth conditions. Thermal strains can be reduced by modifying the growth procedures and annealing techniques.

Figure 4 shows a growth position for growing ruby disks at furnace temperatures of about 1900°C. By using this furnace and a modified burner, crystals can be grown with total thermal gradients in the crystal during growth not exceeding 150°C. This situation compares to that of 400 to 600°C in the conventional flame fusion process. The furnaces used are constructed of high purity alumina refractory with spherical shaped chambers to obtain near black body conditions on the furnace walls. The burner is of the same basic tri-cone design as is used for conventional growth; however, the water cooling of the tip has been improved to allow for operation at higher temperatures. Figure 5 shows the modified conventional growth station with the same modifications.

Figure 6 shows two pieces of ruby between crossed polaroids as follows: (a) grown by the conventional process and cut perpendicular to the optic axis and (b) a slab from a disk boule cut perpendicular to the optic axis. As can be observed, not only is the strain pattern considerably less predominant in (b) but also the crystal inhomogeneities are lessened.

The predominant crystal inhomogeneity in ruby is lineage which can be defined as low angle grain boundaries in which the C-axis wander does not exceed five degrees. Lineage originates on the basal plane and after it is once formed, fans out from this plane. Figure 7 shows a disk with the C-axis in the plane of the disk face and horizontal in the plane of crossed polaroids. Lineage in the boule occupies the area indicated by the arrows.

The degree of lineage in ruby is dependent on growth rate. For 90° orientation disk, that is a disk with the C-axis in the plane of the faces, lineage increases as the growth rate is increased from 1/2 linear inch per hour until the piece shows complete polycrystallinity at a growth rate of 1 1/2 linear inches per hour. At growth rates of from 1/4 to 1/2 linear inch per hour, lineage formation appears to be independent of growth rate. Slower growth rates have been investigated and show no definite results. Investigation of growth rates less than 1/4 linear inch per hour is continuing.

The predominant source of bubbles in ruby crystals is impurities which exist in the growth powder and boil at near the melting point of the ruby and thus form gas bubbles which are entrapped in the solid. One of the worst offenders is silica which causes bubbles that connect to form a series of small feather shaped patterns. Higher growth rates and higher growth temperatures both result in higher bubble formation, when excess impurities exist in the growth powder. To date, no material that is completely bubble free has been produced.

Color banding or variation in chromium ion concentration is caused by fluctuation of the thermal input to the crystal. As the solubility of chromium ions in an alumina solution is temperature dependent, thus very minute fluctuations in gas flows into the furnace chamber will cause changes in the melt temperature and consequently in the chromium ions concentration in the melt. In Figure 8 the upper crystal is heavily banded 90° orientation modified conventional boule. During the growth of this crystal, a variation of approximately 1/2% in the thermal input occurred. This variation was caused by nonuniform gas flows. The lower boule was grown with fluctuation of less than .2% in the thermal input with the exception of point A where a relatively large gas pressure fluctuation occurred.

Disregarding the lineage effect, C-axis wander in ruby takes two forms. Rotation of the C-axis is relatively minor being normally only 1° to 2° in an 18 inch length. Plane drift, i.e., movement in the plane formed by the C-axis and growth axis is more predominant and is orientation dependent. For 0° orientation, i.e., the C-axis is parallel to the growth axis, the normal drift is about 3° toward 90° per 10 inches of length. For 60° orientation the drift is about 2° toward 90° per 10 inches of length. For 90° orientation, the axis is approximately stable. In 90° disk boule the C-axis wander is virtually non-existent in the areas of the crystal which are apparently lineage free.

To date, all ruby crystals used for maser application that have been supplied by Linde are annealed. The annealing is performed in a special furnace separate from the growth furnaces (Figure 9). The crystals are held at 1920°C for a period of from 24 to 48 hours and then slowly cooled. Prior to annealing, the crystals are stressed so excessively that they can not be fabricated. Annealing reduces the stresses sufficiently to allow diamond sawing and grinding to be performed without cracking the material in question. There is evidence of impurity diffusion during annealing since lineage boundaries become more evident after annealing, appearing to concentrate the impurity ions on the sub-grain boundaries. However, at present there is no evidence of chromium ion diffusion in the alumina lattice at annealing temperatures. An extensive study on chromium diffusion in the alumina lattice at temperatures above 1920°C is now being made at the Linde Company, Speedway Laboratories.

Results

For maser application, Linde supplies three different types of ruby. Most common is 90° orientation disks supplied in Cr_2O_3 concentrations from 0.0005% to .5% by weight.

From this material, rods of any orientation relative to the C-axis can be fabricated as shown in Figure 10. The size limit for rods cut from disks is 4 1/2 inches long by 1/2 inch diameter. While lineage in rods longer than 1 inch can not normally be eliminated, the effect is minimized as the lineage boundaries are formed nearly planer with the C-axis. As proof that lineage is the predominant factor in optical maser filamenting, Dr. R. J. Collins[15] at Bell Telephone Laboratories selected a disk from which he could cut two rods at 90° to one other, one parallel to the C-axis, in lineage free areas of the disk. Both of these rods mased uniformly over the full cross sectional area of the crystal.

Two variations of the disk process are the Schawlow overlay[16] shown in Figure 11,which consists of a ruby core cut from disk material of either 0° or 90° orientation with a white sapphire sheath grown over it and the continuous wave ruby, recently announced by Nelson and Boyle[17] of Bell Telephone Laboratories, cut from a disk grown on a large diameter square white sapphire seed (Figure 12).

For large optical masers and microwave masers, Linde currently supplies material grown by the modified conventional technique in either 60° or 90° orientation. Both types of these boules are being used for some of the higher power optical maser work. In addition, material of this type is employed in the larger microwave cavities utilizing ruby masers operating in both the 5.6 KMC and 1.35 KMC regions.

Table I is the result of a joint investigation by Trion Instruments, Inc. of Ann Arbor, Michigan and the Linde Company on various types of ruby. As all the rods are of the same size, the same approximate Cr_2O_3 content, and all were mased in the same cavity, the threshold can be assumed to be related to the quality of the crystal. The rods cut from 60° split modified conventional boule show the lowest threshold with the annealed split boule having a slightly lower value. This result can be explained by the fact that until annealing techniques were developed all ruby boules were split on the plane formed by the growth axis and the C-axis of the crystal to relieve growth strains prior to fabrication. As would be expected annealing further reduces these stresses. However, this material is available only in the 60° orientation whereas most applications require either 0° or 90° material. Of the rods cut from whole annealed material the 0° orientation rod cut from disk boule containing light lineage shows the lowest threshold. Rods 1 and 2 were cut from the same disk; however, it cannot be assumed that 0° mases at a lower threshold than 60° as a rod fabricated at 60° to the C-axis from disk necessarily contains more lineage than a 0° rod cut from the same boule. The predominant effect of lineage on threshold can be seen by the comparison of rods 1 and 3. The relationship between strain in the crystal and lineage content is not known; however, these data indicate there is a relation. Rods 4, 5, and 6 were fabricated from the same crystal. These results indicate the seed end of the crystal is more highly strained than the cap end. This is because during growth start-up the seed end of the crystal is subjected to higher thermal gradients. The increase in threshold caused by reannealing rod 6 after fabrication is probably because of the recrystallization of the rod surface during annealing resulting in decreased optical coupling efficiency. Repeat tests will be necessary to verify this effect. The rod cut from conventional material shows the highest threshold of all the rods, as expected.

The beam pattern and beam divergence can be related to the crystal inhomogeneities and end conditions. Irregular shapes in the pattern indicate nonuniform masing with filaments occurring only in sections of the rods. Beam divergence measures the internal scattering as well as the lens effect of the ends. Additional work on these two factors to more closely coordinate the data to crystal quality is planned.

Thresholds on ruby rods made from disks cut 0° orientation by 1/4 inch diameter by 2 inches long have reported threshold powers as low as 240 joules. These measurements were made in cavities with improved optical coupling. The Schawlow overlay[16] described earlier has a threshold of 48 joules. The improvement of threshold in the overlay is because of

the focusing of the light energy into the ruby core by the white sapphire sheath. The Nelson-Boyle C-W ruby maser[17] has a continuous input of 850 watts from a Hg-dc lamp.

Another tool for measuring crystal quality is the fluorescence linewidth. Whereas conventional material has a normal linewidth of 0.3 cm^{-1}, both disk and modified conventional have normal linewidths near 0.12 cm^{-1}. Linewidths of some disk material have been reported as low as 0.08 cm^{-1}.[18] The above measurements assume a constant temperature of 77°K.

For microwave application, chromium ion content and distribution is critical as shown in Figure 13. For maximum performance, Cr_2O_3 concentration in the maser must be held to with 10% of the base value. Lineage causing C-axis wander is critical for MW application. For optimum performance, the C-axis must not vary more than one degree throughout the crystal.

In conclusion, the author wishes to thank Dr. R. J. Collins and Mr. E. Schultz-Dubois of Bell Telephone Laboratories for the helpful discussion on the operation of maser devices. In addition, the author thanks the Trion Instruments, Inc. for the data on optical maser rod performance. The author also wishes to thank Mr. B. N. Callihan of the Linde Company for his helpful discussions in preparing this paper.

References

1. Combrisson, Honig, Townes, Comptes, Rendes; 242,2451 (1956)

2. N. Bloemberger, Physical Review 104,325 (1956)

3. J. E. Geusic, Physical Review 102,1252 (1956)

4. Kikuchi, Lambe, Makov, Terhune; J. Appl. Physics 30,1061 (1959)

5. Scovil, Ferber, Seidel; Physical Review 105,762 (1957)

6. Townes, Alsop, Giordmoine, Mayer; Astron. Jour. 63,301 (1958)

7. Schawlow, Townes; Physical Review 112,1940 (1956)

8. Maiman, Physical Review Letters (6-1960)

9. Collins, Nelson, Schawlow, Bend, Garrett, Kaiser; Phy. Rev. Let. (10-1960)

10. Verneuil, U.S. Patent #1,004,505 (1911)

11. Drost, Kebler and Dutchess; U.S. Patent #2852890 (9-1958)

12. Kebler, Dutchess, Hutcheson; U.S. Patent #2962838 (12-1960)

13. Schawlow, Advances in Quantum Electronics, Edited by Singer, Published 1961, Columbia University Press.

14. Collins and Nelson, Presented to Proceedings of International Commission on Optics, London (6-1961)

15. Collins, Private Communication

16. Schawlow, Presented to Proceedings of International Commission on Optics, London (6-1961)

17. Nelson and Boyle, Presented Am. Physical Society (1-1962)

18. Schawlow, Presented to OSA (3-1961)

TABLE I

Rods: 1/4" Diameter by 2" Long, 0.05% Cr_2O_3 by Weight
Ends Polished Flat to Less Than 1 Wave of Helium Light and
Parallel to Within One Minute*

	Material	Threshold[1] Joules	Beam Pattern[2] Dia., Inches	Beam Divergence[3][4] Milliradians
1.	0° Disk Light Lineage	1207	5/8	80
2.	60° Disk Light Lineage	1314	5/8	100
3.	0° Disk Heavy Lineage	1560	Irregular 3/8	100
4.	60° Modified Cap End	1278	3/8	60
5.	60° Modified Seed End	1332	Not Available	80
6.	Same as 4. Reannealed after Fabrication	1369	1/2	80
7.	60° Modified Unannealed Split Boule	1173	3/8	80
8.	60° Modified Annealed Split Boule	1139	Not Full 3/8	100
9.	60° Conventional Annealed	1580	Light 1/2	140

(1) Threshold measured at 150°K transmission ends unsilvered.
(2) Beam pattern photographed at 36" from rod end and at 3200 joules input.
(3) Beam divergence determined by the burn on polaroid film at a distance of 8" from the face and at 125% threshold output, 2025 joules input.
(4) All measurements made by the Trion Instruments, Inc., Ann Arbor, Michigan.

*Most recent fabrication tolerances are: rod ends flat to less than 1/10 wave of sodium light and ends parallel to within 10 seconds.

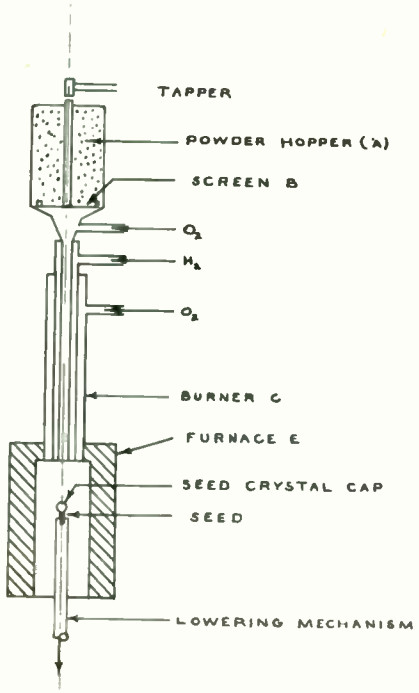


Fig. 1. Flame fusion crystal growth setup.

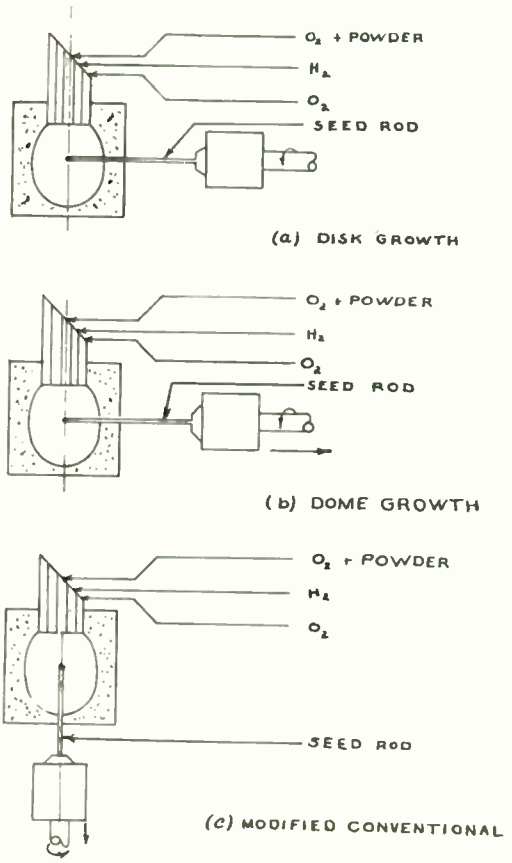


Fig. 3. Modifications of the flame fusion crystal growth process. (a) Disk. (b) Dome. (c) Modified conventional.

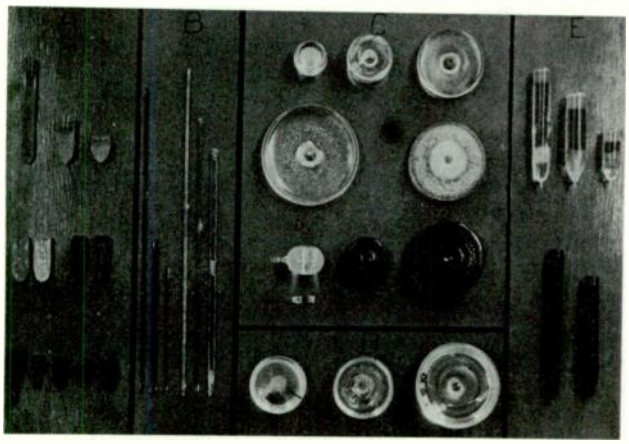


Fig. 2. Sapphire and ruby crystal produced by the flame fusion process.

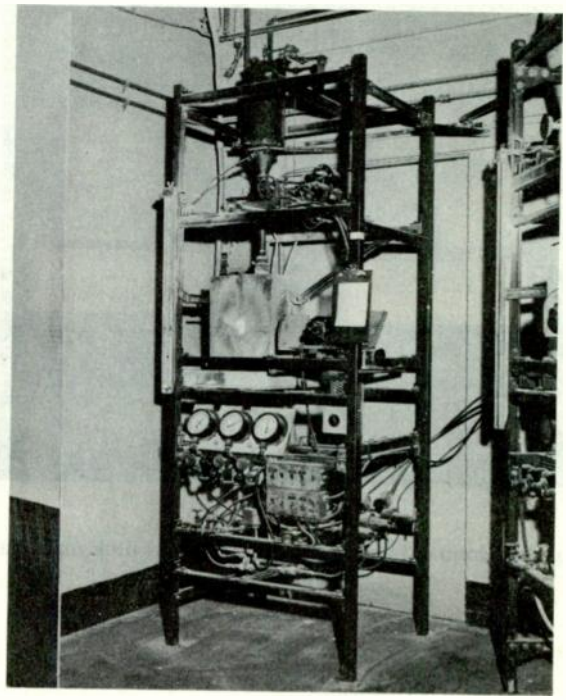


Fig. 4. Disk growth machine.

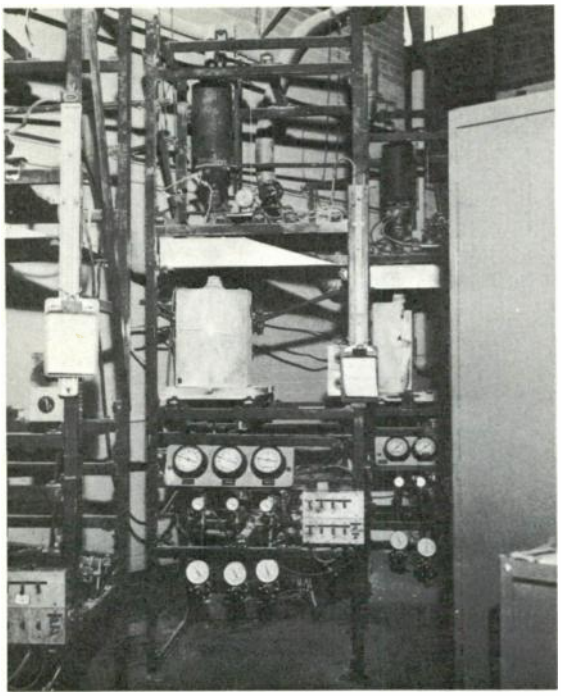


Fig. 5. Modified conventional growth machine.

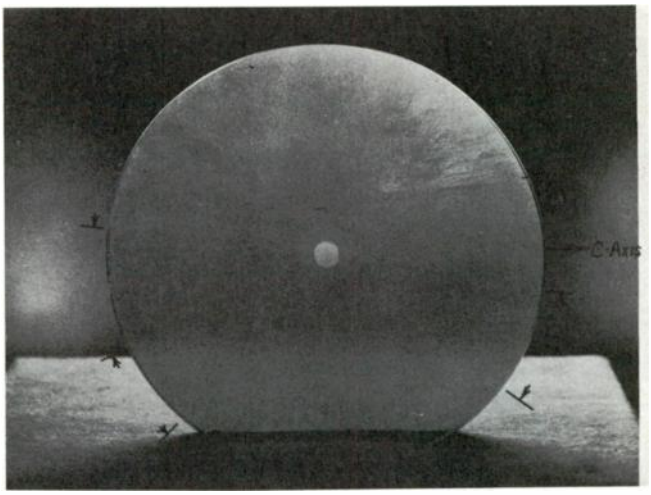


Fig. 7. 90° Orientation disk showing lineage.

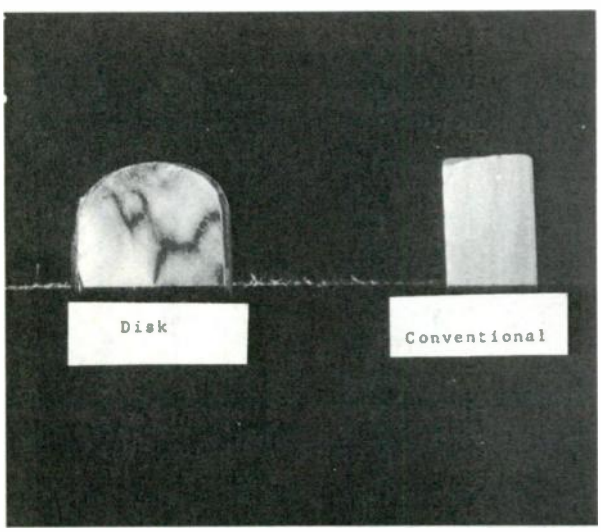


Fig. 6. Comparison of conventional and disk material.

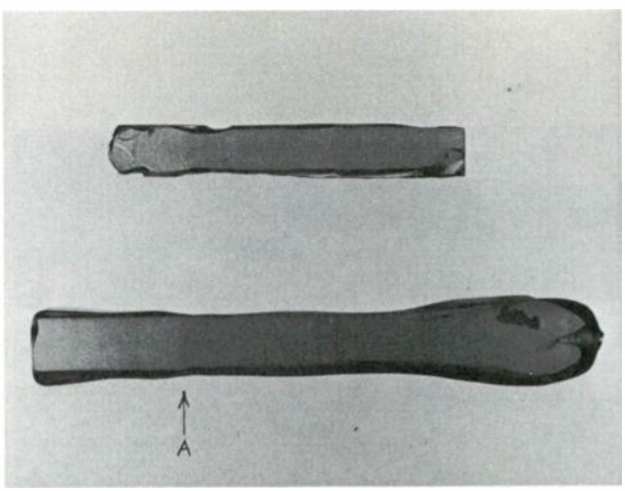


Fig. 8. Color banding in ruby. Upper crystal heavily banded; lower crystal lightly banded.

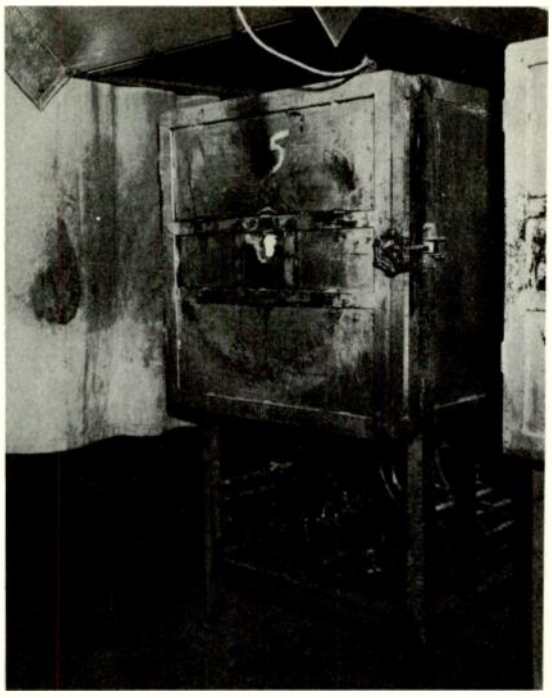


Fig. 9. Annealing furnace.

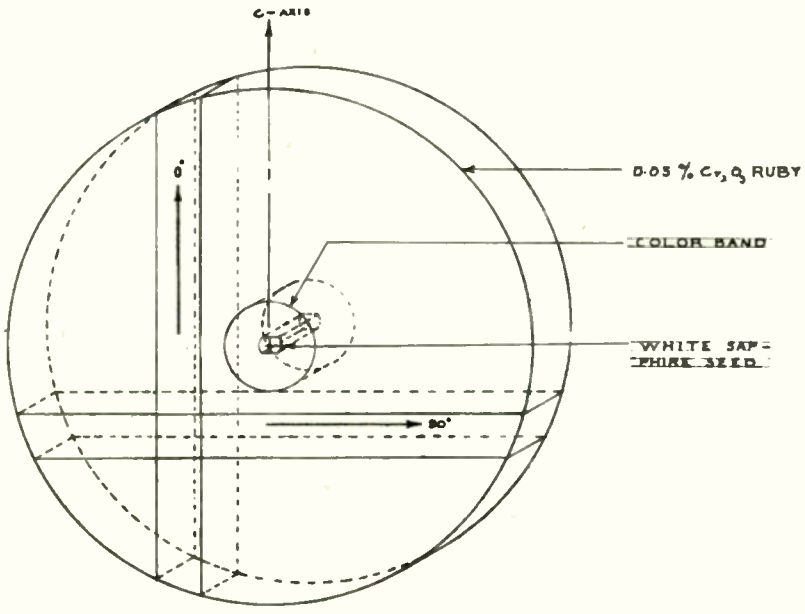


Fig. 10. Fabrication of 90° disk ruby.

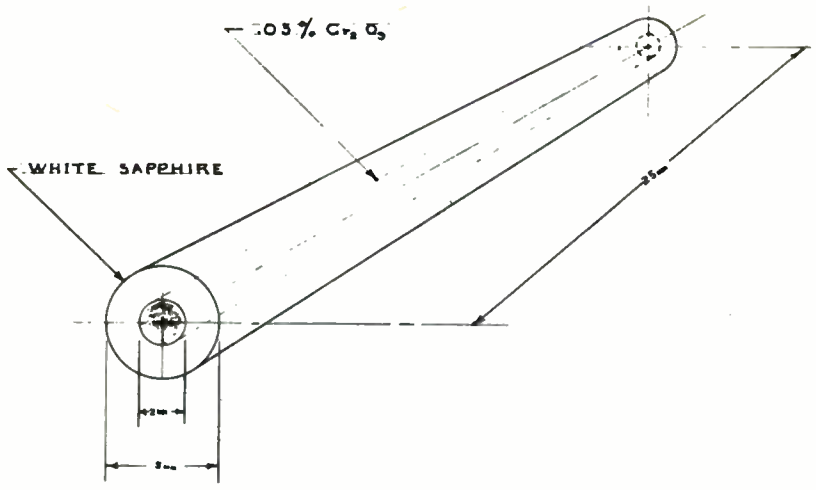


Fig. 11. Schawlow overlay.

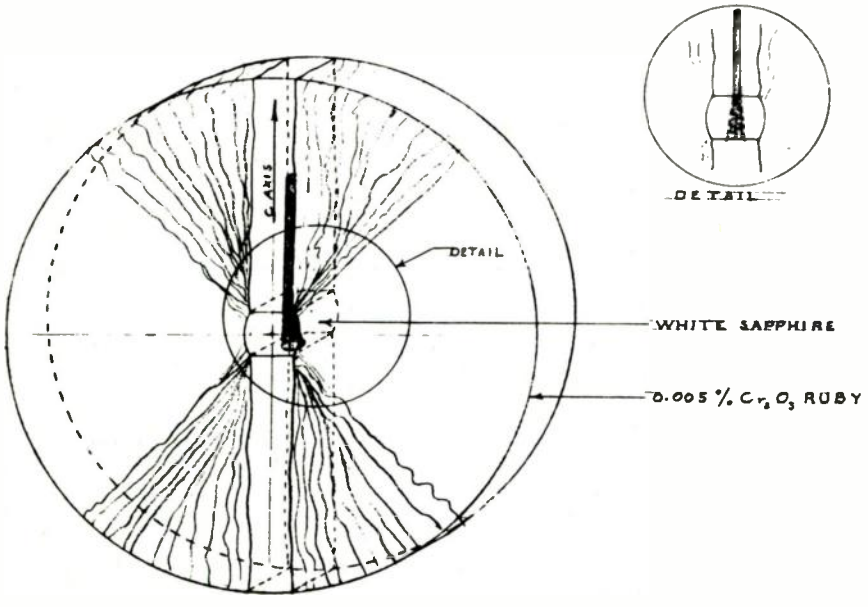
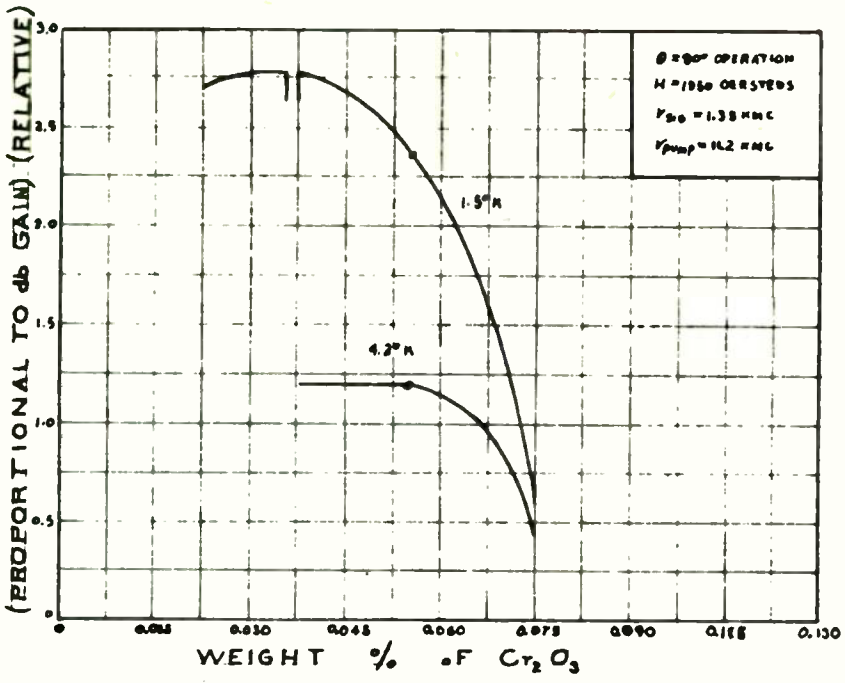


Fig. 12. Nelson-Boyle C-W ruby maser disk.



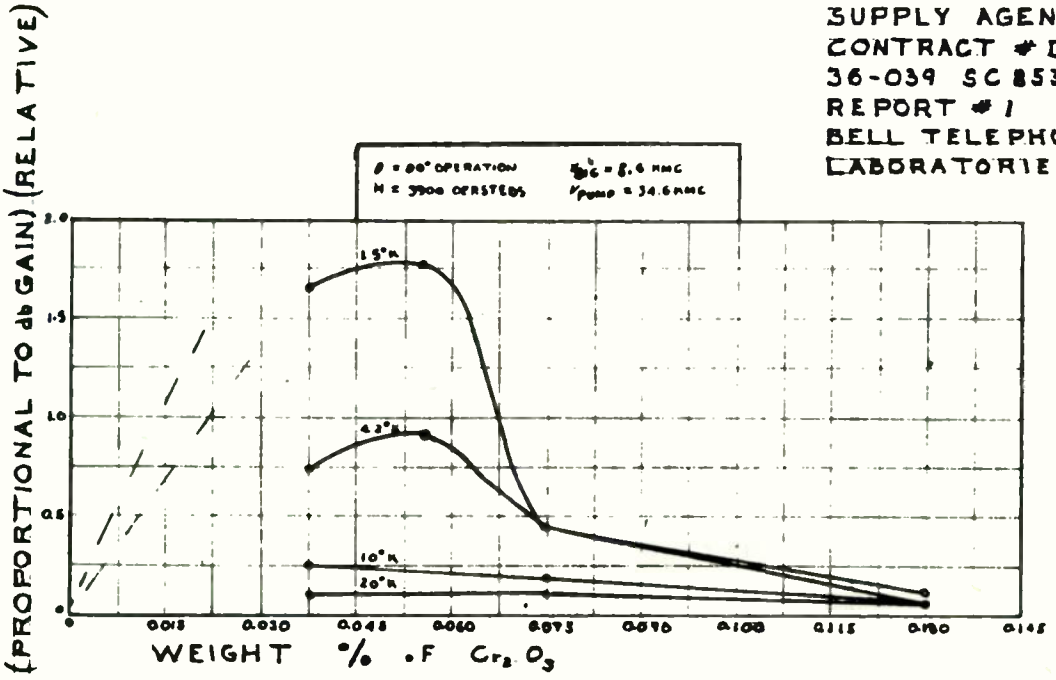
REFERENCE

U.S. ARMY SIGNAL SUPPLY AGENCY CONTRACT # DA 36-039 SC 85357 REPORT # 2 BELL TELEPHONE LABORATORIES

L BAND 1.35 KMC

REFERENCE

U.S. ARMY SIGNAL SUPPLY AGENCY CONTRACT # DA 36-039 SC 85357 REPORT # 1 BELL TELEPHONE LABORATORIES



C BAND 5.6 KMC

Fig. 13. Cr_2O_3 Concentration vs gain for microwave masers.

MICROWAVE MODULATION OF LIGHT

Warren M. Macek, Robert Kroeger, J. Raymond Schneider
Air Armament Division
Sperry Gyroscope Company
Division of Sperry Rand Corporation
Great Neck, L.I., New York

The requirements for microwave light beam modulators and detectors have become particularly significant of late with the increased emphasis on space systems, and the successful demonstration of optical masers. Numerous techniques are readily available to permit intensity modulation of light at low frequencies. Moreover the need to extend the modulation frequency into the microwave region is apparent in practical Doppler radar and wideband communication applications.

Although there are many physical effects which might seemingly be utilized to achieve high-frequency modulation, only a few can be practically applied to this problem.

This paper contains in the introduction, a discussion of the advantages inherent in utilizing modulated light beams for space communications and navigation. In addition, light source requirements for this type of modulator will be reviewed. The need to extend the modulating frequencies beyond the 100-Mc region will then be mentioned including the specific applicability of the Pockels effect. Chief emphasis will be devoted to the application of the Pockels effect to an electro-optical microwave light beam modulator.

Following a basic fundamental review of the modulator technique, the resulting equations will then be applied to a practical device operating in the region from 200-2000 Mc. This modulator has resulted from an extensive research and development program conducted during the past two years. The program will be briefly reviewed, emphasizing the laboratory measurements and modulator parameters. The pertinent specifications and recorded data applicable to one specific modulator model will be outlined and discussed. This unit operates at 870 Mc C.W. with a modulating index of 30 per cent. The detection problems at these frequencies, as well as the numerous methods employed to quantitatively evaluate this device, will be reviewed.

Introduction

The current technological revolution is exemplified by recent developments in the field of communications. The transmission of information by means of electromagnetic waves has played a major role in this revolution. Initially it was only necessary to transmit information to as many people as possible. For this purpose ordinary radio broadcasting using the lower radio frequencies, from 100 kc to 1 Mc was adequate. However, as the broadcast spectrum became crowded and the need for long-range point-to-point communicating arose, a move to higher frequencies was inevitable. However, a selective propagation of electromagnetic energy requires the linear dimensions of the radiating structure to be of the order of magnitude of the wavelength. It therefore requires an antenna a hundred million times larger in area to obtain the same radiation patterns at 1 Mc as that obtainable at 10 kmc.

As a result of the general need for more communicating channels and, in particular, for channel frequencies suitable for highly directive propagation, an extensive effort was initiated to expand the useable communication spectrum. This effort has resulted in the present consideration of the application of millimeter waves. However, their widespread use in communications systems does not appear to be imminent. Needless to say, the ever-growing need to transmit more information, especially in the light of current space exploration requirements, can of course be partially satisfied by a more efficient utilization of the existing radio spectrum. However, a rigid upper limit to the total amount of information that can be transmitted arises from the bandwidth limitation of about 30 kmc.

Recent developments in the field of optics, namely the optical maser, will soon permit a portion of the electromagnetic spectrum, which has been largely overlooked in the past, to be used for transmitting intelligence. The visible near infrared spectrum covers a frequency range from about 10^{14} cps to 8×10^{14} cps (0.4 microns to 3.0 microns) a total bandwidth of approximately 800 millon Mc. Also, because of the extremely small wavelengths involved, highly efficient directional radiators and guiding structures of modest size are possible. Incoherent light sources with 10 per cent efficiencies have already been realized and current optical maser estimates of 2 to 3 per cent are now reported as possible. In addition, photomultiplier light detectors having an S-20 characteristic with 2-3 per cent quantum

efficiencies and a conservative NEI of (10^{-13}) watts at a ruby laser wavelength of 6943 A are currently available*. Communications over extensive ranges is therefore readily possible. There are however, a few principle factors limiting the practical realization of the potential of light as an information carrier.

Limitations of Light Beams as Carriers

The applicability of light beams for long-range communications within the earth's atmosphere is severely limited due to the sunlight and other sources of noise at light frequencies. In addition, atmospheric scattering of light and susceptibility to weather conditions make it extremely difficult to achieve reasonably acceptable signal-to-noise ratios in the transmission of intelligence by light carriers.

Also, light sources are generally incoherent as compared to radio-frequency sources. The latter are generally produced by a coherent motion of electrons in an electrical conductor such that c-w single frequencies can be readily obtained. Light however is normally generated by uncorrelated individual electronic transitions. A single photon results from each transition to which is normally associated a wavetrain of specific frequency but of finite duration. Because of the frequency spread resulting from this frequency discontinuity and due to the incoherent nature of the light beam, normal methods and techniques employed at radio frequencies are not totally applicable at light frequencies.

And finally, light beams until fairly recently were almost impossible to frequency, or amplitude, modulate at high radio frequencies thereby severely limiting their potentially available wide bandwidths.

Some of these arguments are still valid to a degree. For example, although the noise and scattering limitation will of necessity limit the application of light for earth-bound purposes, nevertheless, for point-to-point communications using highly directional devices (and narrow optical filters) to minimize noise, light beams are quite attractive for fair-weather modest ranges within the atmosphere and especially for outer space exploitation. In regard to the nature of the light source, the coherency argument is no longer applicable since the recent discovery of optical masers, although specialized tooling techniques are still a necessity. Optical masers now make it practically feasible to anticipate spectral line sources having about a hundredth of an Angstrom spread with a peak pulse power in the multimegawatt range, and nominally radiated into beams of milliradian width from transmitter objectives of only a few square inches. In addition, it should also be noted that light heterodyning has already been accomplished by the application of the nonlinear detection characteristics of photoelectric devices. Beat frequencies resulting from the superpositioning of two laser signals onto a photocathode have been detected and used to determine the spectral content of the laser sources. The application of such techniques to the extrapolation of the Doppler frequency from a laser return signal is of course quite apparent. Coherency as such, for its own sake, is no longer a relevant argument against the use of light.

An important advantageous feature of light as a carrier is the permissible modulation frequency range. The highest modulating frequency must, in general, be no greater than a fraction of the carrier frequency. For example, a bandwidth of a few hundred megacycles might be associated with a 3000-Mc carrier. Using light as a carrier, modulation frequencies as high as the infrared region are potentially available. Consequently, considerably more information can be sent on a single carrier at light frequencies than at radio frequencies. However, the modulation of light over broad frequency bands poses many difficult problems. We shall now review a few of the many possible modulating schemes and in the process indicate one in particular which we find is best suited to a practical microwave modulator commensurate with current systems requirements.

Modulation of Light

The modulation of light over broad frequency bands poses many difficult problems. Microwave light beam modulators and detectors have been among the primary missing components required for optical communication systems. Numerous techniques are readily available to permit coherent intensity modulation of light beams at low frequencies. However, for communications, when the maximum amount of information must be transmitted, and for many applications, such as optical velocimeters, modulating frequencies of necessity must be extended out into and beyond the microwave region.

---

*It has also been recently announced that a high gain microwave photomultiplier detector, having a distributed multiplier structure using a dynamic crossed-field electron multiplication process with a frequency response of 3 kmc is currently operating with an NEI of 10^{-11} watts.[1]

Many physical effects can seemingly be utilized to realize this transition; however with further consideration only a few are practically applicable. The modulation of light can be generally divided into two classes. Either the light beam emitted from the source or the light source itself may be modulated. The latter may be accomplished by many techniques. For example:

1. The modulation of the power sustaining a gaseous discharge.

2. Modulation of an electron beam exciting fluorescence.

3. A spectral line generation or line splitting by application of high electric or magnetic fields to a gaseous discharge, namely, the Stark or Zeeman effects.

In addition, a steady light beam may also be modulated by many techniques as for example:

1. Mechanical shuttering.

2. Variation of gas absorption frequencies by electric or magnetic fields.

3. A rotation of the polarization of a light beam by magnetic fields applied to transparent optical media, namely, the Faraday effect.

4. The eliptical polarization of plane polarized light by electric fields applied either transverse or longitudinally to the light propagation in specific optical media, namely, the Kerr or Pockels effect.

A discussion of the relative merits of each of these techniques is beyond the scope and intent of this paper. Some of these methods are restricted to low frequencies while others are capable of high-frequency response, but limited to laboratory use. Also, while gaseous media have been mentioned, these same effects have, of course, been observed in solids and liquids and for completeness these media should also be considered. We will now discuss a few techniques applicable to high-frequency light modulation. Primarily however, we will specifically be concerned with recent developments in the application of the Pockels effect to the microwave modulation of light beams.

Paramagnetic crystals may be used to modulate light at microwave frequencies making use of their mechanical, electrical, or magnetic properties.[2] For example, the elasto-optical effect may in principle be modulated by ultrasonic waves at microwave frequencies since such waves can be readily generated. However, it is difficult to obtain a well-defined acoustical mode essential for an optical modulation and as such this technique does not offer much primise of reduction to a practical device. The mageto-optical (Faraday) effect, which describes the change in propagation constant of right and left circularly polarized light, can however be utilized. The rotation of the plane of polarization per unit length is proportional to the component of magnetization parallel to the direction of light propagation. As a result, a time-varying component of magnetization will cause a modulation of the plane of polarization. If a constant magnetic field perpendicular to the direction of light propagation is applied to a paramagnetic crystal containing rare earth ions; and if a microwave field parallel to the light propagation near a magnetic resonance frequency is also applied, then a large microwave component of magnetization will be created. As a result, the plane of polarization can be rotated by about 1 degree per esu, or 1 degree per gauss-centimeter, in a typical rare earth salt at about 2 degrees K. The requirement of a LHe temperature however, makes it complex from a practical point of view. The analogous effect in transparent ferromagnetic insulators could, however, be used at room temperatures; but the effect would be in the best available garnet, an order of magnitude smaller than in a paramagnetic salt at the low temperature. However, the microwave power consumption in the latter case is quite small and the effect could be readily extended to higher frequencies in the millimeter range.

Although the magnetic effects are capable of exploitation, the best technique to date is the electro-optical effect. Kerr cells have already been used as extremely fast light shutters to produce pulses of 0.3 nanosecond duration.[3] The effect is based upon a statistical preferential alignment of dipolar molecules in an electric field. The index of refraction of the optical media is a function of the average orientation of the molecules. Therefore, for partial alignment, the propagation constant depends on the polarization direction. With dipolar liquids however, the upper frequency limit is determined by the relaxation effects and losses occurring when the inverse frequency becomes comparable to the tumbling time of the molecules, namely, to the characteristic correlation time for Brownian motion.

With solids, the dipoles are usually completely frozen in, or the characteristic time for motion of the dipoles is much longer than in liquids. A striking exception however is presented by the motion of the hydrogen atoms in the dihydrogen phosphates. The displacement can follow rapid variations of an electric field well into the microwave range beyond 10^{10} cps. The optical index of refraction is slightly dependent on the position of

the hydrogen atoms and, therefore, on the angle between the polarization vector and the applied electric field. Consequently, an electrical modulation of the optical propagation tensor is possible. The rotation of the plane of polarization is about 0.5 degree per esu which is accidentally of the same order of magnitude as the Faraday effect. However, the low temperature and constant magnetic fields are not required and the necessary electric fields can be simply generated using standard microwave cavity resonator techniques and medium power sources. We shall now investigate this latter effect more fully.

The Electro-Optical Effect in Crystalline Materials

The variation of the refractive properties of an optical medium by the application of an electric field is called the "electro-optic effect." In a liquid medium with the field normal to the light propagation, the effect is designated the "Kerr effect." In a crystalline medium with the field parallel to the light propagation, it is called the "Pockels effect." The principle practical interest in these effects is associated with their applicability to producing an electrically controlled birefringence in either an electroded liquid Kerr cell, or an electroded crystal section. As a result, if the electric fields are properly oriented relative to the medium and the light propagation, an electrically variable phase difference may be produced between orthogonally polarized interfering light beams thereby resulting in a possible light beam intensity modulation if the cells are positioned between crossed polaroids. In other words, the plane polarized light is changed to elliptically polarized light before passing through an analyzer (see Figs. 1 and 2).

All piezoelectric crystals to some extent show electro-optical effects with polarized light. The use of a device incorporating Pockels effect offers a decided advantage over the ordinary liquid Kerr cell in that the induced retardation is proportional to the applied voltage rather than the square of the voltage. As a result, a linear response is available without necessitating large biasing voltages. This results from the fact that in liquids the molecules are all free to rotate. Thus in averaging over-all orientations of the molecules, the linear terms drop out. In a solid crystal, however, the molecules are all held in fixed positions, and assuming no inversion symmetry, the linear terms will not average out. However, a linear response can be obtained in the liquid cells if a large biasing field is applied. This field effectively serves the same purpose for the liquids molecules as the internal crystalline field does for the solids molecules; i.e., it destroys the parity of the ground state. There are, in addition, many other features attesting to the practical applicability of the Pockels cell over the Kerr cell for modulating light at microwave frequencies Typical examples might be the extreme purity required of the nitrobenzene for the Kerr cell, the increased cavity complexity to contain the liquid, the extended cell lengths (limited by the modulating-wave length) required to produce comparable retardations using reasonable field strengths, etc.

It should also be pointed out that the electro-optical effect of a crystal is dependent upon its mechanical condition resulting from the electro-optical effect of the piezoelectric deformation which contributes to the normal effect if the crystal is free to move. As a result, the crystals birefringence due to these photoelastic effects must be distinguished from the direct electro-optical effect since with high frequencies (above 40 kc) the crystals are essentially clamped due to inertia effects.

The dihydrogen arsenates and phosphates of ammonium and potassium are available as large single crystals, and are of particular interest in so far as their electro-optical properties are concerned. In crystals of this symmetry class ($V_d = \bar{4}2m$), electro-optical retardation can be obtained free of overlapping natural retardation if both the electric field and the light propagation are parallel to the crystal's C-axis. Thus a Z-cut (001) basal section can be most usefully employed as a variable retardation plate. The chief advantage of these crystals is their sensitivity to low voltages and their capability of responding to high frequencies with a low loss tangent angle (see Fig. 3 and Tables 1 and 2). For example, KDP and ADP has already been utilized at microwave frequencies and their exact upper limits are as yet not definitely determined.

The Retardation Effect

The theoretical and experimental aspects of the electro-optical effect in tetragonal scalenohedral crystals of the type KH_2PO_4 (KDP) have been discussed in considerable detail in the literature.[4,5] It was shown by Billings that the equation of the Fresnel index ellipsoid for crystals of tetragonal scalenohedral symmetry under arbitrary electric fields can be written:

$$o^2 (x^2 + y^2) + e^2(z^2) + 2r_{63} E_z xy$$

$$+2r_{41}(E_x yz + E_y zx) = 1$$

where the xyz coordinates are taken as the standard crystallographic axes, o and e are the

reciprocals of the ordinary and extraordinary refractive indices n_o and n_e, respectively, of the crystal in the absence of an electric field, and r_{63} and r_{41} are the two independent electro-optical coefficients describing the changes in the index ellipsoid when the electric field is applied. The r_{63} coefficient thus describes the effect of a field parallel to the optic axis (the crystal C or Z axis), and r_{41} the effect of a field perpendicular to this axis. Under the electric field, the crystal becomes biaxial and the three axes of the index ellipsoid no longer coincide with the standard xyz axes. The new principle axes and refractive indices may however be determined purely algebraically by considering the effect to be a rotation of the coordinate axes to coincide with the principal axes of the ellipsoid defined by Equation 1.

Thus for a field in the Z-direction (Ex = Ey = 0), the axes of the ellipsoid can be denoted by (x' y' z'). Thus z' coincides with z, while x' and y' are rotated by 45 degrees from x and y. The values of the three indices of refraction are then given by:

$$n_{x'} = n_o + \frac{1}{2} n_o^3 r_{63} E_Z$$

$$n_{y'} = n_o - \frac{1}{2} n_o^3 r_{63} E_Z \qquad (2)$$

$$n_{z'} = n_e$$

The retardation through a basal section of the crystal is then given by:

$$\Gamma = \frac{d}{\lambda}(n_{x'} - n_{y'}) = \frac{n_o^3 r_{63} V}{\lambda} \qquad (3)$$

Values found for r_{63} by R. Carpenter, et al, are 2.54×10^{-7} cm/stat volt for ADP and 3.15×10^{-7} cm/stat volt for KDP. It should be noted that the retardation is a function of the sign of the electric field. Therefore, after reflection, if the field essentially remains fixed, the phase difference of the orthogonally polarized light beams will be additively retarded. As a result, if the cell produces a phase difference of 45 degrees for a single pass, then after reflection the total difference will be 90 degrees (i.e., 1/4 wave plate or plane polarized light converted to circularly polarized light). On the other hand, if the field has reversed, then the phase retardation will be cancelled and the light will be reflected essentially the same as it was incident. This technique can be specifically applied to demonstrating the attainment of microwave modulation or, still better, practically doubling the over all effect as we will describe more fully a little later on.

Electrical-Optical Light Modulation

General Principles

Tetragonal crystals, such as KDP and ADP, have a unique crystalline direction called the optic axis along which light rays propagate with the same velocity regardless of their polarization. Such crystals are called uniaxial. In general, they possess two principle indices of refraction and as a result, light traveling in any direction other than the optic axis consists of rays with two distinct velocities for two specific polarizations of the light. If a uniaxial crystal such as KDP is polished and placed in an electric field which is parallel to its C-axis, namely its optic axis, then for light propagating along the optic axis the uniaxial crystal becomes biaxial resulting from the electro-optical effect of KDP. Therefore, whereas with no field the same propagating velocity existed for all polarized rays propagating along the optic axis, with the field applied two specific velocities exist for two specific light-ray polarizations. As a result, for plane polarized light initially aligned at 45 degrees to these two specific polarization directions, an electrically variable retardation plate results. If an analyzer is then added, crossed to the polarizer, an intensity modulation of the light beam will result. In the absence of an electric field no light is transmitted, while for the half-wave voltage the light is plane polarized at 90 degrees to the polarizer, and is therefore parallel to the analyzer with the resulting transmitted light at a peak intensity.

In general, therefore, the net E-field causes eliptically polarized light to emerge from the crystal resulting from the phase delay introduced between the two specific light polarizations as derived from the incident plane polarized light by the crystal. This effect is typical for birefringent crystals and is most notably applied in designing wave plates. However, due to the alternating electric field, the eccentricity of the resulting elliptically polarized light will change and consequently the analyzer passes a periodically changing light intensity.

Plots of the relative transmitted light intensity versus the applied voltage across the crystal, (the electric field in the Z-direction), are shown in Fig. 4 for the case of the crossed polarizer-analyzer and two conditions of bias. If $\emptyset$ represents the crystal retardation, I_o the maximum transmitted light intensity, V_o the corresponding voltage required for 100 percent modulation of

the light, and I and V the instantaneous intensity and voltage; then the relative light transmission is given by:

$$I = I_o \sin^2\left[\frac{\phi}{2}\right] \text{ or } I = I_o \sin^2\left[\frac{\pi}{2}\frac{V}{V_o}\right]$$

The operating conditions depicted in the second figure (biased condition) may be obtained by employing either (1) a d-c bias across the crystal or (2) a quarterwave retardation plate placed between the crystal and either the polarizer or analyzer, with its principal axis of refraction parallel to the birefringenet axis of the crystal. Obviously the latter method is normally preferred.

These results assume that the collimated light from the incoherent source is incident normally upon the crystal and parallel to its optic axis. However, off-axis rays suffer a natural retardation due to the crystal symmetry thereby reducing the transmission contrast ratio, namely complete extinction is not possible. All light sources in addition to having a finite but significant spectral spread also have an associated source size. The former can be compensated for by filtering. The latter results in a limit to the degree of collimation in keeping with a required aperture size and therefore a maximum peak intensity since to further collimate necessitates a drop in available power. This departure from perfect collimation as well as the finite spectral spread must therefore be accounted for in optimizing the modulating index of an EOLM. The angular aperture of the EOLM is, therefore, limited by the permissible contrast ratio; and for a contrast ratio of 100 to 1, for a crystal thickness of 1/2 inch and light centered at 5000A, the acceptable divergence is approximately 2 degrees. The angular aperture is inversely proportional to the square root of the crystal thickness and directly proportional to the square root of the wavelength for a constant contrast ratio. However, larger cones of light can be utilized for more flux at the expense of the contrast ratio.

In this regard, optical masers are especially advantageous, not only for their narrow spectrums and small beam divergences, but because they can be recollimated to extremely small apertures by simple lenses. They may also be designed so as to produce plane polarized light, using 90-degree oriented ruby crystals, and as such the 50 per cent intensity loss suffered in the polarizer can thereby be avoided. The application of EOLM's to modulating laser beams is therefore readily apparent. However, they may also be readily incorporated into a laser for a feedback control, or as a laser cavity Q switching technique. These latter techniques hold considerable promise for significantly increasing the available peak powers from pulsed solid state lasers.

Figure 5 is a diagram of a typical EOLM and detector arrangement with the polarization orientation shown stage by stage. The polarizers can consist of ordinary commercial grade sheet polarizers, Nicol or Glan-Thompson polarizing prisms or such other sophisticated units such as multi-thin-film polarizers. The specific choice is dictated essentially by the required ratio of maximum to minimum light intensity, namely, the crossed relative to the parallel arrangement of the polarizer-analyzer combination, as well as the permissible linear cell aperture. For example, sheet polarizers having transmissions from 22 to 38 per cent with associated extinctions of 0.0005 to 0.05 per cent are currently available at modest cost and applicable over essentially the complete visible region with no aperture limitations. On the other hand, prism polarizers with considerably better extinction ratios are also available with improved transmission characteristics, but limited in aperture by the exponential increase in cost due to the limited crystal sizes currently available. The thin-film technology holds some promise of compromising these two extremes although their widespread application is not extensive due to lack of commercial sources and the attendant high cost of making them individually in small quantities.

Electrodes

Two types of electrodes are essentially available for an electro-optical light modulator based on the Pockels effect. In general, they must be transparent and conducting, and a means of protecting the polished hygroscopic crystal surface is required. The latter can be best accomplished by coating the crystals with an anti-reflection coating such as MgO or lithium fluoride. A possible electrode might be a laminating glass with a thin conductive layer, such as Nesa glass. However, although this electrode has a high transparency and provides a uniform field, the high electrode resistance restricts the frequency response to about the 100 kc range. These electrodes also cannot be used for d-c or unilateral pulses because of the deterioration of the laminating layer due to electrolysis. The most effective electrodes are therefore either evaporated gold or metallic structures in the form of grids or rings. Since the opaque rings or grids will obscure from 20 to 30 per cent of the optical aperture, the proposed configuration employs the rings shaped as washers with the inner diameter the same as the crystal thickness, and the outer diameter conveniently set equal to or slightly less than the diameter of the cylindrical

KDP crystal. Therefore, the fringing of the electric field between the electrodes provides a field in the transparent region. This requires that the spacing between the electrodes (the inner ring diameter) not exceed the crystal thickness. Thus grids can be used on thin crystals, and rings, as herein proposed, on thick crystals. As such, the effective voltage will vary from 100 per cent of the applied voltage at the edge of the electrodes to approximately 70 per cent in the center of the transparent region. Therefore, the retardation will not be quite uniform over the aperture and on the average about 20 per cent more than the theoretically calculated voltage is required for a given response. These effects will not be intolerable, however, if the modulator contrast ratio is not especially sensitive to or restricted by the specific retardation. It should also be noted that if it is desired to keep the peak voltages to a minimum, another proven technique would be to connect two electroded crystals, properly oriented and laminated together, electrically in parallel. The optical effects would thereby be additive and require half the voltage but have twice the capacity of a single cell. These techniques, however, would not be amenable to a simple cavity design.

Transmission Losses

From available data, the absorption losses of ADP and KDP in the visible region do not appear to be foreboding. The specific measurement of the transmissivity of these crystals is to some extent complicated by the fact that the quality of the crystals surface is especially susceptible to variation from one sample to the next. This results from the basic hygroscopic nature of the crystal and can be prevented by careful polishing techniques and overcoating. However, allowing for this, as well as possible inhomogeneities of the material, it does not appear unreasonable to expect the absorption losses to be less than 3-4 per cent for a 1 cm thickness. This is based on the fact that a KDP sample 0.580 inch long had a measured transmittance of 87.5 per cent. Allowing 8 per cent for reflection losses, the absorption and scattering loss is therefore less than 4.5 per cent. Admittedly the reflection losses can be compensated for by overcoating and a minimum crystal thickness is desirable. The required thickness depends directly on the requirement to minimize cell capacity and to maintain an adequate contrast ratio across the cell aperture. The crystal material itself poses no problem since it can be cut to 50 mils before spontaneous fracturing occurs due to its fragility, and will in addition support 25 KV/mm. The field intensity problem can therefore be alleviated by potting techniques. The primary limitation is essentially the power dissipation and cavity design configuration.

Cavity Design

Since it was necessary to apply several thousand volts peak to peak across the crystal at microwave frequencies in order to achieve an effective light-beam modulation index, it was necessary to employ standard microwave techniques as opposed to conventional low-frequency procedures. As a result, a high-Q re-entrant type capacitive-loaded coaxial-line resonator was employed to develop the necessary electric fields along the desired crystal axis. The electro-optic crystal was employed, in the form of a cylinder, as the dielectric medium of the cavity. Thus the cavity concentrates an extremely high electric field in the limited region in which the crystal is placed. In addition, the requirements that the electric field and optical path through the crystal be parallel to the Z, or optic C, axis of the crystal are thus quite easily fulfilled. A cross section of a laboratory model cavity resonator is shown in Fig. 6. Its resonant frequency can be determined with reasonably good accuracy by setting the susceptance of the capacitance, formed by the crystal, equal to the inductive reactance of the coaxial transmission line, formed by the outer and center conductors. In addition, although the loss tangent of the potassium and ammonium dihydrogen phosphates are low (see Table 2) for high voltage at microwave frequencies, the crystal loss can be excessive; and as a result produce unwanted thermo-optic effects in the crystal. In extreme cases the crystal might even be permanently damaged. Therefore, an extremely low crystal capacitance and/or cooling is mandatory.

Due to the necessity of passing the collimated light beam through the cavity and also the crystal cylinder, the following equations can only approximate the actual characteristics of the particular physical configuration employed. The cavity's resonant wavelength, shunt impedance and unloaded Q-Q_o- of the cavity can therefore be approximated by:

$$\lambda_o = 2\pi \left[r_1 h \left[\frac{r_1}{2t'} + \frac{2}{\pi} \ln \frac{elm}{t'} \right] \ln \frac{r_2}{r_1} \right]^{1/2}$$

$$R_{sh} = \frac{2\eta_o h^2}{\lambda_o \delta'} \cdot \frac{\left[\ln \frac{r_2}{r_1} \right]^2}{\left[\frac{h-t'}{r_1} + \frac{h}{r_2} + 2 \ln \frac{r_2}{r_1} \right]}$$

$$Q_o = \frac{R_{sh}}{\omega L} = \frac{2h}{\delta'} \cdot \frac{\ln \frac{r_2}{r_1}}{\left(\frac{h-t'}{r_1} + \frac{h}{r_2} + 2\ln\frac{r_2}{r_1}\right)}$$

The above equations assume the crystal to be lossless. Taking into account the radio-frequency losses (tan δ) in the crystal, these equations have to be modified by paralleling the shunt resistance (R_{sh}) with the equivalent dielectric resistance (R_δ) where:

$$R_\delta = \frac{1}{\omega c \tan \delta}$$

Modified equations are as follows:

$$R'_{sh} = \frac{R_{sh}R_\delta}{R_{sh}R_\delta} \quad ; \qquad Q_o' = \frac{R'_{sh}}{\omega L}$$

where

L = cavity inductance $= (\mu_o h/2\pi) \ln r_2/r_1$

μ_o = permeability of free space = $4\pi \times 10^{-7}$ henry per meter

lm = $1/2 \left[(r_2-r_1)^2 + h^2\right]^{1/2}$

e = 2.718...

η_o = free space impedance = 120 π ohms

t' = thickness of an equivalent air capacitor = t/ϵ_r

t = actual thickness of the crystal

ϵ_r = relative dielectric constant of the crystal

r_1 = cavity re-entrant post radius (also crystal radius)

r_2 = radius of the cavity

h = height of the cavity

δ' = skin depth of the cavity

tan δ = loss tangent of the crystal

ω = resonant frequency of the cavity

Power dissipated in the crystal is only due to the loss tangent; and it can be expressed as follows: $P = \omega CV^2 \tan \delta$, where C = crystal capacitance, V=voltage across the crystal, or more familiarly as: $P = 2\pi \Delta f CV^2$ since $\tan \delta \simeq 1/Q = 2\pi \Delta f/\omega$ (if the material losses of the cavity are assumed to be negligible.)

In addition, since the cavity is coupled to a transmission line, λ_o and Q'_o are somewhat altered from the values given by the equations. For the matched conditions the Q of the cavity will be half of Q'_o.

The regions through which the light beam passes on either end of the crystal may be treated as circular waveguide segments. In order to limit microwave energy losses in these regions, their dimensions must be such that the critical cut-off frequency occurs below the resonant frequency of the cavity. As can be seen from Fig. 6 the polaroid holders and the long axial path length on one side of the crystal in combination with a relatively small cross-sectional optical path opening provides adequate attenuation.

A cavity EOLM with a quarter-inch aperture and a resonant frequency of approximately 900 Mc has been designed and operated successfully. The loaded Q measures about 1000. The best modulation index, with 8 watts of r-f power input, was measured to be 30 per cent using ADP, and 40 per cent using KDP.

In one application where a much larger aperture was required, another type of cavity was designed, where the optical aperture was increased from 1/4 inch to 3/4 inch. Provisions were also made for liquid, or gas, cooling of the cavity in order to enhance the electro-optical effect and improve the stability of the resonant frequency. Due to such a large aperture (3/4 inch), it was necessary to install vanes across the aperture on either side of the crystal. These vanes help to distribute the electric field more uniformly and axially across the crystal and still transmit about 90 per cent of the light. An exploded view of the EOLM is shown in Fig. 7.

The resonant frequency and the loaded Q of this cavity are 600 Mc and 1000 respectively. With 8 watts of r-f power input, the modulation index was measured to be 18 per cent.

Due to difficulties in detecting the light signal over the previously mentioned frequency range it was necessary, for one recent application, to make an EOLM with an operating

frequency equal to about 200 Mc. At this frequency the use of a cavity for the EOLM is impractical due to its size being prohibitively large. Hence a 200-Mc oscillator was developed using two 8025A vacuum tubes in push-pull. For the resonant network in the plate circuit, 3/8-inch silver plated copper tubing is being used as a short-circuited transmission line with the ADP or KDP crystal acting as the capacitive element at the open end of the line. The 200-Mc oscillator has two significant advantages over the re-entrant cavities, in that it is simpler to build and it always has the crystal in the maximum electric field, due to the crystal being part of the resonating circuit itself. Whereas with the re-entrant cavities, when the resonant frequency drifts, power coupled into the cavity drops, unless the r-f oscillator frequency is changed to "follow" the resonant frequency of the EOLM cavity.

With 850 volts applied to the tubes and drawing about 45 ma of plate current, the modulation index was measured to be 36 per cent. Future plans are to increase the plate voltage and also to apply cooling to the crystal. The expected modulation index should be about 60 per cent.

In cooling the ADP crystal, one must be careful to avoid thermal transients and minimize thermal gradients through the crystal. It has been found that thermal transients affect the electro-optical effect considerably.

Measurement Techniques

Concurrent with the design and development of a microwave light modulator, it is necessary to devise methods whereby the degree of modulation can be measured. Calculated values are appropriate and necessary but monitored results are often more expedient and obvious to everyone concerned. We shall now discuss a simple and direct technique to accomplish this, assuming photomultipliers with adequate frequency response are not readily available. The light passing through the EOLM is reflected from a mirror back into the modulator, where, by using a beam splitter (half-silvered mirror) in front of the cavity, a portion of the reflected light reaches a photomultiplier. If the mirror is properly positioned relative to the modulator, then light initially passed through the shutter while it was open can be returned during this same phase and thereby reach the detector. In so doing we must note that the phase retardation of the detected light will, as a result, be doubled over that ordinarily resulting from a single pass. If the mirror is then moved to another position, a quarter of a wavelength away in terms of the modulating frequency, then the light initially passed by the open shutter will now return out of phase, while the shutter is closed, so that no light will reach the detector. In other words the retardation is cancelled. On the other hand, if the reflecting mirror is uniform formly translated, then the photocurrent will vary from a maximum to a minimum with the relative intensities indicative of the degree of modulation and the rate-of-change of the signal corresponding to the translation velocity. Similar measurements could also be used to determine the actual modulating frequency by noting the separation between the peaks and nulls.

A simpler method to continuously indicate the degree of modulation is to modulate the microwave modulating frequency with a much lower frequency to which the photomultiplier can respond and which can be easily amplified and displayed on a scope. This technique affords a simple method for aligning polarizers, KDP crystal, as well as optical biasing wave plates. It also permits us to estimate the degree of modulation as before, by rms measurements, but with better sensitivity since the noise resulting from the detectors shot noise and the light transmitted by the knee of the response curve (poor extinction of crossed polaroids) can effectively be eliminated by narrowband amplifiers tuned to the low frequency as compared to the former d-c measurements.

For practical applications however, a detector capable of responding to microwave frequencies is of course mandatory. Heterodyning techniques using photocathodes or photomultiplier dynode mixing schemes have been demonstrated, as well as a few recent photodetector techniques with microwave response capabilities. A standard 1P28 photomultiplier tube was successfully used as a heterodyne detector at 870 Mc by injecting a low impedance 870-Mc signal on a suitable pair of dynodes. The signal was then conventionally detected by a mixing scheme whereby the injected (L.O.) signal was offset from the received signal and the low-frequency difference extracted and amplified. However, all these detection techniques will require further refinement before they can be considered for immediate application.

Another method of detecting microwave modulated light is to use a second EOLM as a demodulator (see Fig. 8). We, therefore, essentially utilize two electro-optical units as a modulator-demodulator combination as we did in the reflector detection scheme. The applicability of such a configuration, namely two separate electro-optical devices, to a wideband communications system is depicted in Fig. 9. In keeping with the finite bandwidth limitations of must modulators, the EOLM's being no exception, a wide bandwidth capability would therefore result from essentially

packaging a number of electro-optical modulators and demodulators into a single system. Each is assigned a specific channel frequency and bandwidth and, if need be, utilizes a single source and transmitter by beam splitting and reassembling techniques such as with fiber optics. However, many sources could also be used in conjunction with a single transmitter. The receiver is similarly designed such that each demodulator in conjunction with a channel amplifier and filter will in essence select out their particular channel (by means of a straightforward superheterodyne arrangement as shown). Similar techniques to achieve high information rates could also be accomplished with other detection schemes employing essentially the same modulating technique.

Temperature Effects And The Advantages Of Other Materials

Effects of temperature upon the electro-optical coefficients of ADP and KDP are now being studied intensely. Not only is it expected that by cooling these crystals a considerable increase in their electro-optical coefficients will result, but in addition, the heat dissipation problem will be considerably enhanced with thermal gradients and localized heating kept to a minimum. Thus degradation of the modulation effect in EOLM applications resulting from severe thermal gradients can be avoided and a reduction by a few orders of magnitude of the required half-wave retardation voltages might result. As a result, tests have been made utilizing dry ice as well as dry N_2 gas precooled by immersion in L N_2, to lower the crystal temperature to approximately -40 degree C. The results to date have been as anticipated[8] and the current test work was initiated. A helium Dewar has been modified, Fig. 10, and L N_2 utilized to effectively lower the crystals temperature close to its Curie point (for KDP, 123 degrees K), at which point the maximum effect is anticipated (Fig. 11). These results have been encouraging and are presently limited by thermal gradients introduced along the optic axis of the crystal resulting from the asymmetrical design of the cold sink. Typical results achieved so far indicate a 5 to 1 reduction in the half-wave retardation voltage of KDP. A symmetrical cold sink is presently being designed and results comparable to those indicated in Fig. 11 are expected.

This investigation is significant and warrants further study in the future in so far as adopting these cooling techniques to high-frequency electro-optical components. Preliminary results have shown that the dielectric constant, as well as the tangent loss angle, will increase slightly with cooling at high frequencies. The extent of the loss tangent increase, however, is not expected to limit the applicability of cooling to increasing the electro-optical coefficients. It is however, obvious that further study must be devoted to, the effects of temperature on the materials loss tangent and dielectric constant at high frequencies. Nevertheless, it is conservatively estimated from available data that by cooling the KDP crystals to the vicinity of their Curie point (300 degrees K to approximately 123 degrees K), an improvement factor of at least 20 to 30 can be practically realized.

In addition to cooling KDP, the practical feasibility of effectively raising the crystals Curie point from 123 degrees K to 213 degrees K by utilizing deuterated KDP has already been indicated[8] and these crystals are now available in limited quantities. As a result, the anticipated improvements resulting from cooling KDP can be realized by cooling the deuterated KDP with a considerably less severe cooling requirement. These potassium dideuterium phosphate crystals are also noteworthy since their electro-optical coefficients are approximately three times better than KDP at room temperature, (see Fig. 11). It is therefore expected that by cooling the deuterated KDP from 300 degrees K to approximately 225 degrees K, a conservative improvement factor of at least 2 to 3 resulting from deuteration and 20 to 30 resulting from cooling can be realized.

Other materials, such as the arsenates previously mentioned in another section, as well as zinc sulphide should also be investigated. These advantages are not as obvious, however substitution of cubic crystals for tetragonal can be advantageous, especially for off-axis rays which suffer a natural birefringence in tetragonal crystals and of more importance the applicability of multireflection techniques (using off-axis rays) to considerably enhance the electro-optical effect without increased power or complexity. The zinc sulphide problem to date is that they are not available as synthetically grown large single crystals. Further research, however, may soon overcome these problems and their applicability to electro-optical components should not be discounted based on the fact that they are only naturally available for analysis in limited quantities of extremely small single crystal size.

Conclusion

The current need for wideband communications to permit high information rates has resulted in a critical requirement for microwave light modulators and detectors to fully exploit the potential of current state-of-the-art developments in regards to light sources. These requirements are especially notable today as a result of the increased emphasis on space systems and the successful

demonstration of optical masers permitting an accurate control of the direction of high-power coherent light beams.

The infrared/optics group at Sperry has been actively conducting advanced research and development in these immediate areas for the past two years and was able to successfully demonstrate, for the first time, a c-w microwave light beam modulator in June 1960. This unit operated at 900 Mc using electro-optical crystals and cavity resonator techniques.

This paper has attempted to review some of the pertinent background material necessary for a fundamental and complete understanding of these modulation techniques in addition to describing some specific design concepts and basic measurements. A few promising areas of study warranting further consideration have also been mentioned although we have been far from exhaustive in this area for obvious reasons.

As a final note we would like to point out that with a microwave modulation of light, the old arguments as to the basic nature of light can again be considered. Since the duration of a single photon wavetrain, as determined by spectral line widths, is normally around a nanosecond for many sources excluding gas lasers, and since our microwave modulating period is of the same order of magnitude, we are essentially capable, with full modulation, of splitting a photon if the photon's energy can be thought of as being distributed over the wave. This concept is naturally completely unacceptable to physicists. Nevertheless, we are grateful in having been able to contribute to the continued research necessary to clarify this and other philosophical questions relevant to the concept of the basic nature of light. We therefore leave it to you to decide - did we or did we not split a photon?

References

1. O.L. Gaddy and D.F. Holshouser, Proc. I.R.E., Vol. 50 (1962), p. 207.

2. H. Bloembergen, 1960 NEREM, Boston, Massachusetts.

3. D.F. Holshouser and H.M. Foerster, "Time Element in Photoelectric Emission," (ASTIA Document No. 158-274, Electrical Engineering Research Laboratory, University of Illinois).

4. B.H. Billings, J. Opt. Soc. Am., Vol. 39 (1949) pp. 797, 802.

5. R.D. Carpenter, J. Opt. Soc. Am., Vol. 40 (1950), p. 225.

6. Amer. Inst. of Physics Hd. Bk. (New York, GcGraw Hill Book Company, 1957).

7. A. Von Hippel, Tables of Dielectric Materials, Vol. IV (January 1953, M.I.T.).

8. B. Zwicker and P. Scherrer, Helv. Phys. Acta., Vol. 17, p. 346.

Table I

ELECTRO-OPTICAL COEFFICIENTS
AND 1/2 WAVE RETARDATION VOLTAGES
FOR SELECTED CRYSTALLINE MATERIALS

Crystal	r_{41} (10^8 cm/stat volt)	r_{63} (10^8 cm/stat volt)	$V\lambda/2$
Potassium Dihydrogen Phosphate (KDP)	25.8	31.5	7.5 KV
Potassium Dihydrogen Arsenate	-	39	6.2 KV
Potassium Dideuterium Phosphate	-	70	3.4 KV
Ammonium Dihydrogen Phosphate (ADP)	62.5	25.4	9.6 KV
Ammonium Dihydrogen Arsenate	-	19	13.0 KV
Rubidium Dihydrogen Phosphate	-	33	7.3 KV
Zinc Sulphate (cubic symmetry)	5.0 6.4	- -	12.4 KV 9.3 KV

Ref: A.I.P. Handbook[6]

Table II

ADP, KDP LOSS TANGENTS & DIELECTRIC CONSTANTS VS. FREQUENCY

Material	Field Orientation	Temperature	Frequency	10^2	10^3	10^4	10^5	10^6	10^7	10^8	3×10^8	3×10^9	10^{10}
ADP	⊥ optical axis	25°C	ϵ'/ϵ_o	56.4	56.0	55.9	55.9	55.9	55.9	55.9	55.9	-	-
			tan δ	400	46	4.6	<5	<5	<5	<5	<10	-	-
	|| optical axis	25°C	ϵ'/ϵ_o	16.4	16.0	15.4	14.7	14.3	14.3	-	14.3	-	13.7
			tan δ	2400	240	70	70	60	10	-	5	-	50
KDP	⊥ optical axis	25°C	ϵ'/ϵ_o	44.5	44.3	44.3	44.3	44.3	44.3	44.3	-	-	-
			tan δ	98	15	<5	<5	<5	<5	<5	-	-	-
	|| optical axis	25°C	ϵ'/ϵ_o	21.4	20.7	20.5	20.3	20.2	20.2	20.2	-	-	-
			tan δ	170	24	<20	<5	<5	<5	<5	-	-	-

NOTES:
1) Values for tan δ must be multiplied by 10^{-4}
2) Frequency in c/s
3) ϵ'/ϵ_o is dielectric constant or permittivity relative to vacuum (also designated K, ϵ, etc.)
4) Tan δ is dielectric loss tangent or dissipation factor
5) Samples were dried over phosphorous pentoxide
6) Nominal acc. ϵ'/ϵ_o = ±2%
7) Nominal acc. tan δ = ±5%

Ref: A Von Hippel, Tables of Dielectric Material, Vol. 4, Jan, 1953

$$\phi = 2\pi\Gamma = 2\pi\frac{l}{\lambda}(n_{\parallel}-n_{\perp}) = 2\pi lBE^2$$

l = LENGTH OF CELL
E = ELECTRIC FIELD
B = KEN CONSTANT
ϕ = PHASE DIFFERENCE
Γ = RETARDATION
$n_{\parallel}-n_{\perp}$ = BIREFRINGENCE

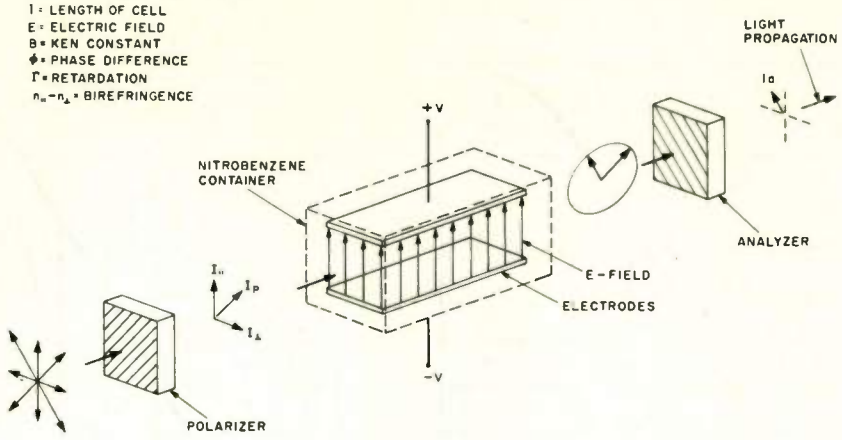


FIG. 1 KERR CELL MODULATOR

$$\phi = 2\pi\Gamma = \frac{2\pi d}{\lambda}(n_{x'}-n_{y'}) = \frac{n_o^3 r_{63} E_z d}{\lambda}$$

ϕ = PHASE DIFFERENCE
Γ = RETARDATION
n_o = ORDINARY REFRACTIVE INDEX OF CRYSTAL
Y_{63} = ELECTRIC OPTICAL COEFFICIENT
E_3 = ELECTRIC FIELD IN Z-DIRECTION
d = CRYSTAL THICKNESS
λ = WAVELENGTH OF LIGHT

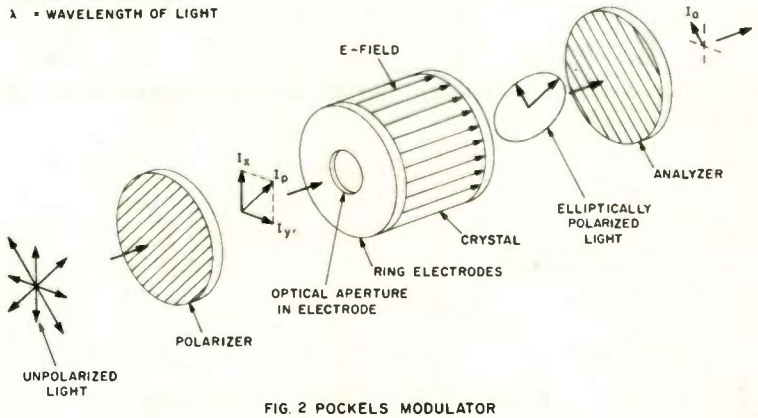


FIG. 2 POCKELS MODULATOR

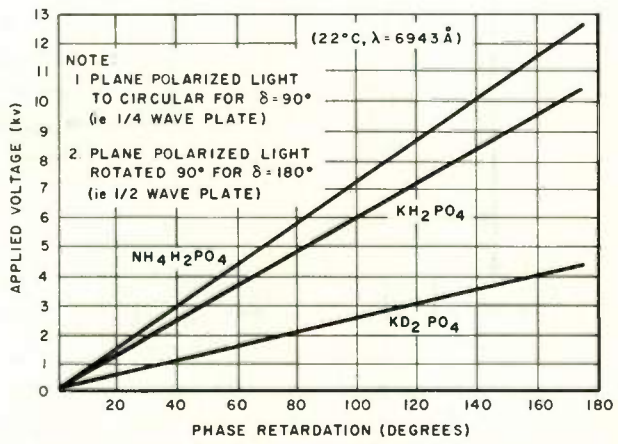


FIG 3 PHASE RETARDATION VS. APPLIED VOLTAGE

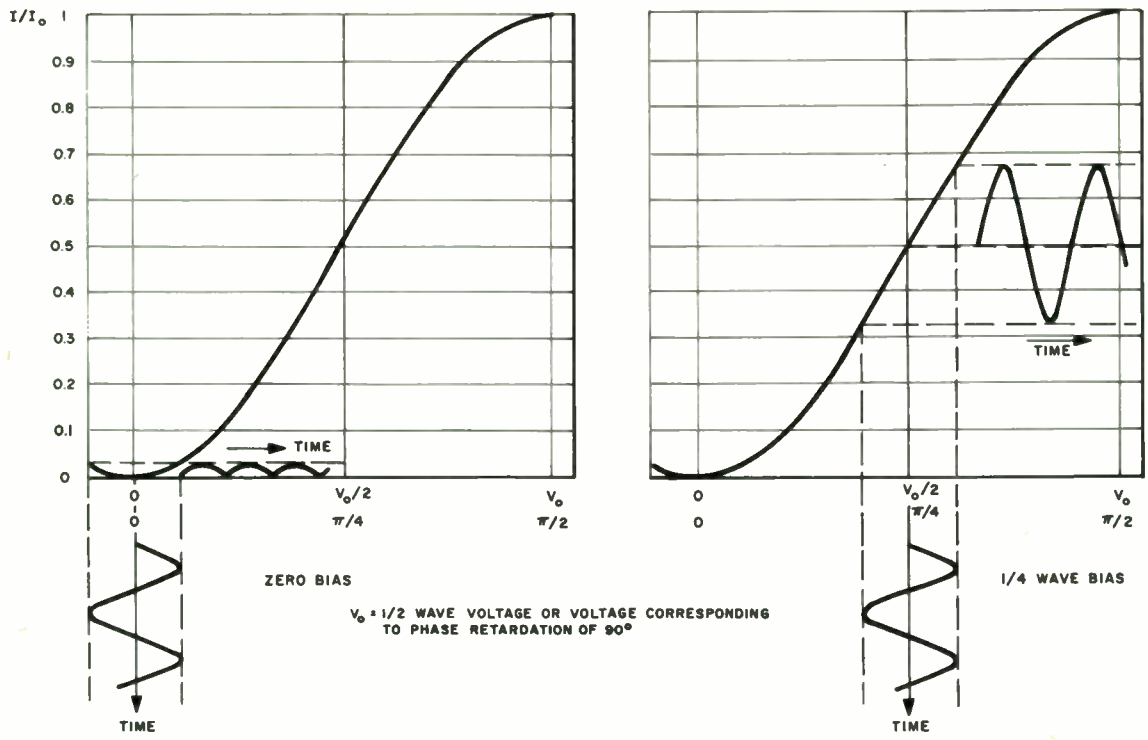


FIG. 4 RELATIVE INTENSITY VS. APPLIED VOLTAGE WITH AND WITHOUT OPTICAL BIASING

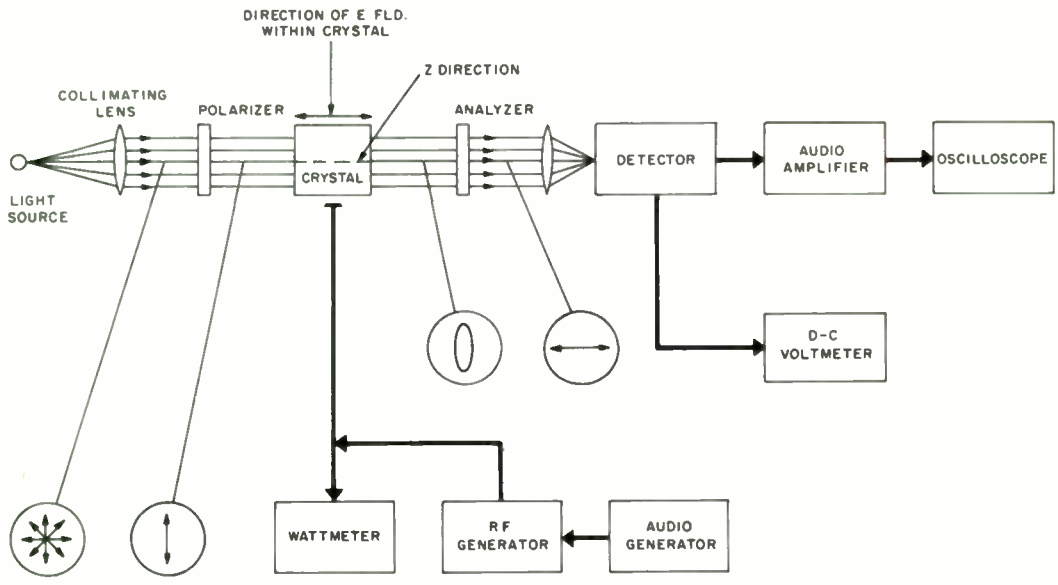


FIG. 5A EOLM BASIC TEST SETUP

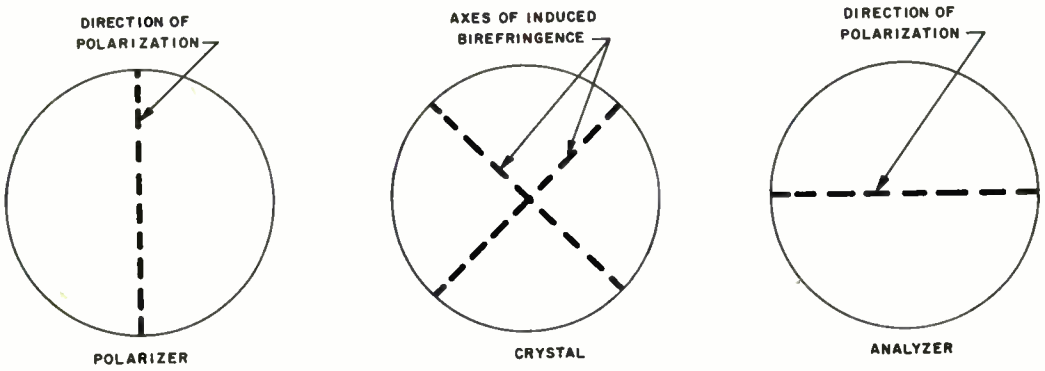


FIG 5B ORIENTATION OF PLANES OF POLARIZATION

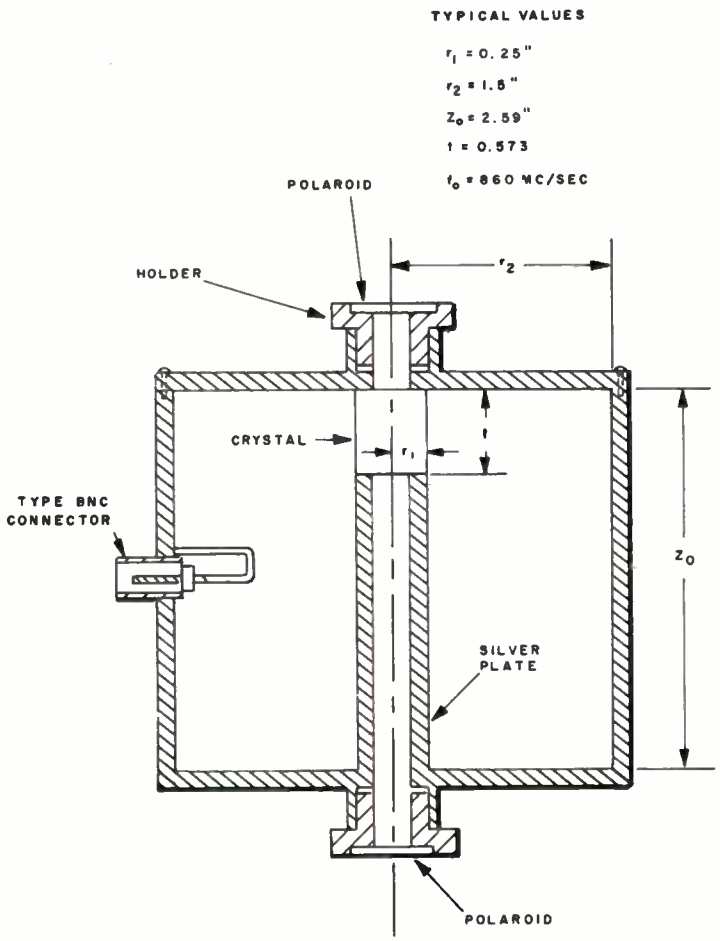


FIG.6 CUTAWAY VIEW OF COAXIAL RE-ENTRANT CAVITY

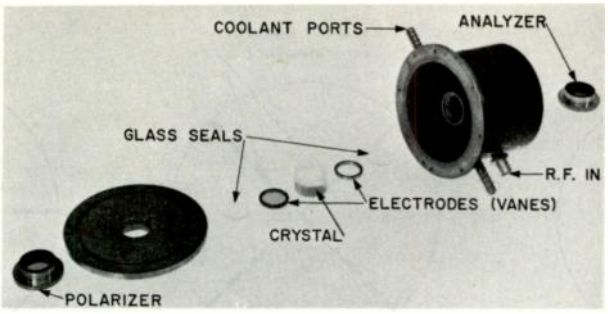


FIG. 7 600-MC LIGHT BEAM MODULATOR
WITH 3/4-INCH APERTURE

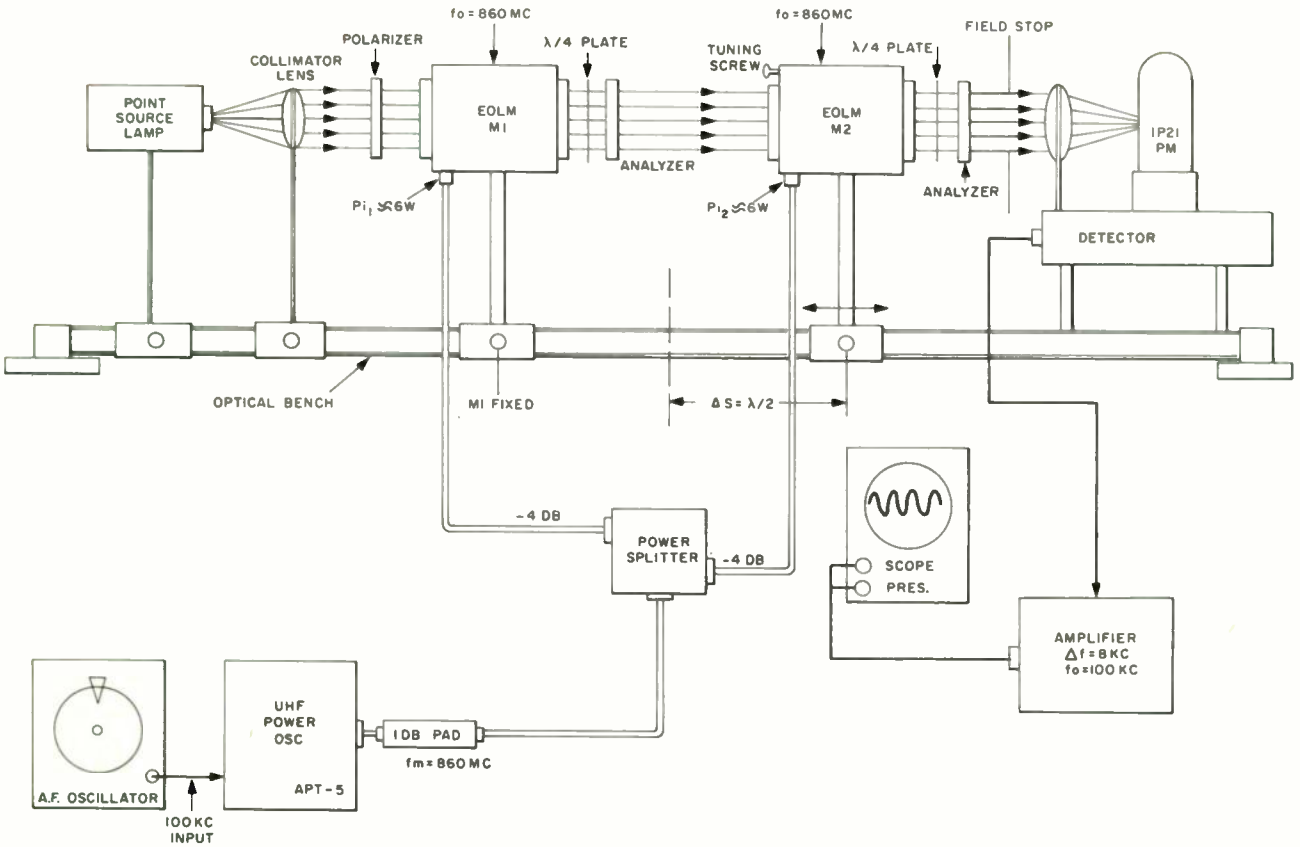


FIG. 8 EOLM MODULATOR-DEMODULATOR PHASE MODULATION TEST

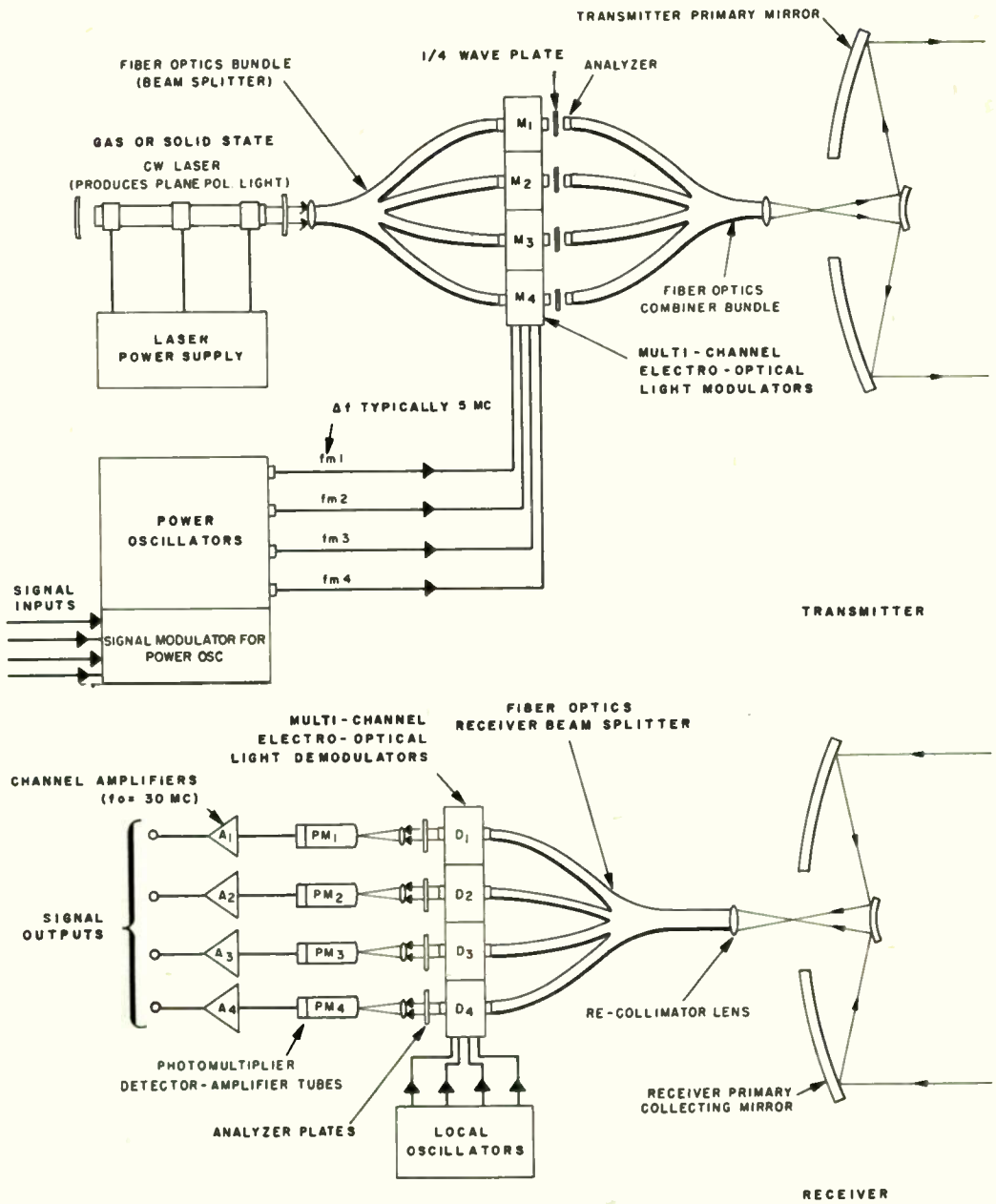


FIG. 9 TYPICAL MULTICHANNEL WIDE-BAND ELECTRO-OPTICAL MODULATOR-DEMODULATOR COMMUNICATION SYSTEM

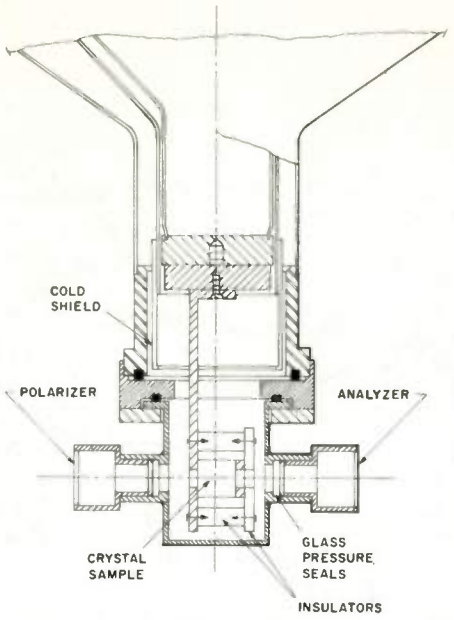
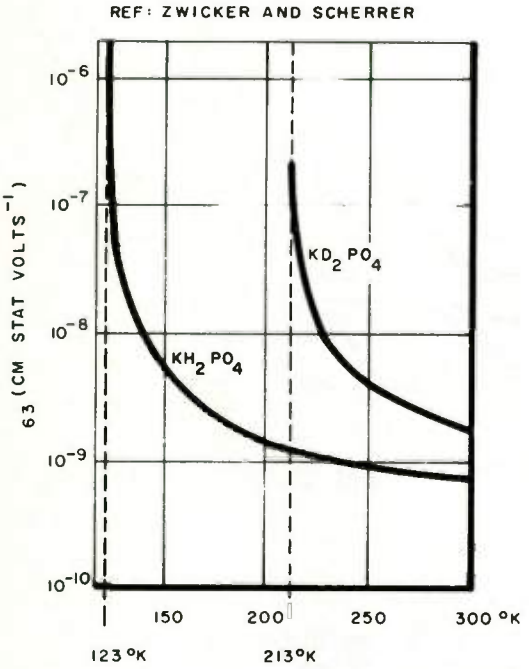
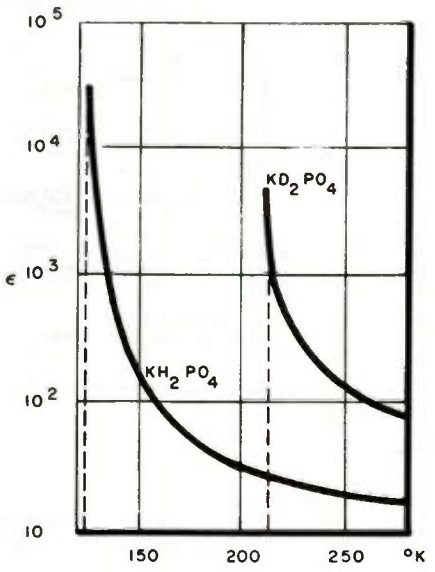


FIG. 10 DEWAR CONFIGURATION FOR DC LOW-TEMPERATURE TESTING OF CRYSTAL SAMPLES



(A) ELECTRO-OPTICAL COEFFICIENTS VS. TEMPERATURE



(B) DIELECTRIC CONSTANT VS. TEMPERATURE

FIG. 11 EFFECTS OF COOLING KDP AND DEUTERATED KDP

PRACTICAL SOLID STATE THREE DIMENSIONAL (3-D) DISPLAY

John L. Coddington And Robert J. Schipper
Avco Electronics and Ordnance Division
Cincinnati 41, Ohio

Summary

This paper presents a fresh and unique approach for generating a practical three dimensional display. The feasibility of this concept has been substantiated by a demonstrational model which offers a simple approach to dynamically presenting multi-parameters in their true perspective for fast evaluation and decisions concerning complex situations.

An electroluminescent (EL) cross-grid, non-suppressed, matrix panel is the heart of the "3-D" display. When both a vertical and a horizontal line is energized (vertical for range and horizontal for height), their intersection produces a bright spot of light in two dimensions. As the panel is rotated in azimuth about a vertical axis, at a speed of 20 cycles per second above flicker rate, the third dimension is obtained. Electronically pulse gating the EL panel displays floating spots of light representing data or aircraft in three dimensions. These spots appear to be floating within the non-evacuated enclosure.

The paper describes the following: the display design considerations, discussion of the working model, system operation, advantages and limitations, new displays, and concludes with an example of a technique to incorporate this display in the Naval Tactical Defense System (NTDS) or Semi-Automatic Ground Environment (SAGE) System.

Introduction

In the past the three dimensional (3-D) display has been considered more as a novelty or toy than a truly new and useful tool. This situation was undoubtedly the result of two basic factors; the numerous "3-D" display limitations, and the lack of systems that required utilization of such a display. Recent advances in volumetric and multi-parameter systems has given impetus to the generation of a practical "3-D" display.

"3-D" Display Considerations

General

A display system can only be as useful as the design objective and specifications designated. It is quite difficult to state a single overall objective therefore the consideration of many areas in the overall design is necessitated to insure a good design. Any new design idea presenting parameters never before truly employed, limits the extent of specifying some of the normal criteria along with new criteria until certain concepts are proven and are available for extensive study. The Avco "3-D", display falls into this catagory of displays. The initial objectives of our display were as follows:

Design. The prime consideration was to eliminate as many of the major disadvantages of other "3-D" displays as possible while retaining as simple and practical a design as possible.

Decisions. Almost without exception the most general objective of any visual display is to enhance human decision making. These decisions encompass decisions and data from a Controller vectoring the landing of an aircraft to a Commander deciding a tactical decision. To insure an accurate decision, a display must show an element of understanding or realism. This was a primary factor in the design.

(Complex Situations — Equally as important a factor in design is the presentation of complex situations. Presenting data such as aircraft in their true perspective — range, height, and azimuth — is essential to obtaining a useful display.)

Speed and Accuracy. Rapid reaction time, accompanied with high accuracy for multi-dimension displays are incompatible in the human mind where rapid important decisions are required. The speed and accuracy of response for decisions will be enhanced if a human is able to evaluate the answer in a binary type form (yes or no). This display permits this judgement without unnecessary degradation.

Improved System Operation. System operation is generally improved if data is combined in one display rather than separate consoles such as Plan Position Indicators and Range Height Indicators. This area was to be more fully investigated and incorporated if possible.

Other. Many other areas such as coding, control, access of information, human factors, etc., are to be considered later in a full evaluation which is not completed at this time.

Electroluminescence

Since Electroluminescence (EL) is the basic "heart" of the system it is logical to present here a cursory review of this phenomenon. EL is the direct conver-

sion of electricity into light. This process eliminates all intermediate steps normally connected with electricity-to-light conversion. Electrons within the EL phosphor are excited to a higher state energy level, generating a photon type of light. This light is usually referred to as a solid state transfer of energy. The visible light is dependent upon photon emission which is a direct result of an atom entering into and then returning from a highly excited state. Visible light is resultant only when the electron within the atom is in motion to or from its higher energy level.

Model Discussion

Figure No. 1, Demonstration Model, illustrates the practical hardware of the program. This is a feasibility model and not a final type which would later be designed to a particular application. This model was designed as a self-contained simulator, rather than being dependent on other equipment for its input data information. There are three individual Operator Control Panels to present each operator with individual targets for positioning in three-dimensions anywhere within the non-evacuated enclosure area. The parameters for this display are: Range 100 miles in 5 mile increments, Height 30,000 feet in 500 foot increments, and Azimuth 1.4 degree increments. These parameters (Range, Height and Azimuth) are operator selected by appropriate switches on the Operator Control Panel.

Figure No. 2, Layer Structure of Electroluminescent Panel, displays a commercial type EL panel. The glass base sheet is primarily for mechanical rigidity, and also acts as the foundation surface for the transparent conductive film. This film acts as one plate of a capacitor. A ceramic dielectric with an EL phosphor is next, and then another layer of transparent conductive film which acts as the other plate of the capacitor. Electrical connections are then attached to each film. This effectively forms an electronic capacitor.

Figure No. 3, EL Panel Cross-Grid Pattern, displays the type of panel construction employed. The basic difference is in the conductive strips in the "x-y" directions. If 1 "x" and 1 "y" strip is energized each line will light, but their intersection will light very much brighter depending on the signal: frequency, voltage, phosphor, waveform, and duty cycle.

Figure No. 4, The Electroluminescent (EL) Panel is a cross-grid, non-suppressed, matrix type. It is comprised of vertical lines extending from the horizontal axis (representing Range), and horizontal lines extending from the vertical axis (representing Height). Since non-suppression (a light vertical and horizontal line with a bright dot at the intersection) was utilized, the two halves of the panel were divided into odd and even height lines. The EL panel is 10 inches in diameter and 11 inches in height utilizing a 1 inch vertical center area for electrical connection for height information. The Range lines are connected across the bottom of the panel for electrical range insertion. The spot to line ratio, voltage wise, is 15:1. By the use of grid suppression techniques, a contrast ratio of 1000 through 5000 to 1 can be obtained.

Mechanical Considerations. The EL panel is attached to a split-shaft for rigid mounting, support, and positioning. The split-shaft in turn is attached to a bearing housing unit located at the top of the viewing dome, and to a coupler at the opposite end. The opposite end of the coupler is attached to a slip-ring assembly (100 rings) shaft. The opposite end of the slip-ring assembly shaft is attached to a shaft-angle position encoder. The entire unit is rigidly supported and mounted in a vertical plane. A rigidly supported motor operates at approximately 1200 revolutions per minute, or 20 per second, rotating the split-shaft, EL panel, slip-rings, and angle encoder.

Circuit Design and Power Requirements. The circuit design is entirely solid-state and utilizes 500 milliamperes at -30 volts, and 50 milliamperes at +30 volts or an average of 16.5 watts. There are four chassis of which three are identical comparators, and one containing the oscillator and encoder amplifiers, along with two power supplies (+30 and -30 volts). The Range, Height, and Azimuth switch-positions on the Operator Control Panels make up the fixed memory for the display.

System Operations

The electronic operation of the "3-D" display proceeds as follows:

The Power Switch is placed in the "ON" position allowing 60 cycle 115 volts to energize the motor and power supplies. Figure No. 5, System Block Diagram, will aid in understanding the discussion of system operation. The EL panel, slip-rings, and azimuth encoder begin to rotate when power is applied. Selection of a target's Range, Height, and Azimuth is made from each of the Operator Control Panels with corresponding EL Parameter readouts.

The Interrogate Oscillator, operating at 50 kilocycles (KC) sends an Interrogate Signal to the 2^8 bit Gray-Code Shaft Encoder. Each of the encoder's eight tracks, depending on angular position, sends its coded reply (0,1) to its corresponding track amplifiers. The amplifiers send a 0 volt or -30 volt to the three Azimuth Comparator Chassis. The Azimuth Switches on the Operator Control Panels have set fixed 0 volt and -30 volt levels into this chassis. When the encoder output is identical with that of the Azimuth Selector Switches, a Compared Gate is generated. The length of the Compared Gate depends on the speed of panel rotation and azimuth resolution. Our gate is

generated for approximately 195 microseconds.

The 50 KC signal, besides being sent to interrogate the encoder, is amplified and sent to Range and Height "and" circuits in the Azimuth Comparator Chassis. The Compared Gate "enables" the Range and Height "and" Circuits allowing the 50 KC signal to pass to the Range and Height output amplifiers for this 195 microsecond period. The output amplifiers are operated in push-pull with an output signal peak-to-peak of 350 volts. The output push-pull signal is then sent to the Range and Height Selector Switch whose position determine which brush-block of the slip-ring assembly accepts the signal. The 50 KC signal is then sent through the slip-rings to the correct height and range line for a period of 195 microseconds. The EL panel illuminates displaying the target in three dimensions floating within the space enclosure.

System Advantages

Figure No. 6 illustrated the inherent advantages of this type of system namely:

- Full 360° viewing angle
- Completely solid state design
- Multiple colors for target identification
- Low power requirements
- Simplicity of design eliminating:
  - Special Optics
  - Special Viewing Lenses
  - Special Projection Equipment
  - Cathode-Ray Tubes (Normal and Storage)

System Limitations

EL Panel. This display is presently limited in display volume to 30 inches in diameter by 20 inches in height. This is primarily due to panel manufacturing equipment limitations. The display must be viewed in subdued lighting because of the relatively low level of light output. At present 15 lines per inch horizontally and 15 lines per inch vertically are practical quantities for resolution. For a display with a range of 300 miles radius, a range resolution of 1.3 miles is possible. With a height display covering 100,000 feet, a resolution of 333 foot increments is possible, and an azimuth resolution of 0.35 degrees. These figures can be improved by the use of expanded range and height techniques.

Mechanical Problems. The primary mechanical problem is slip-ring life. The present state of the art in slip-ring design indicates approximately 1000 hours operating time before maintenance is necessary. Normally only the brush-blocks require replacement at the 1000 hour interval. We are presently investigating techniques to eliminate the use of the slip-ring assembly.

Possible Applications

The advantages presented by this type of display make the following types of usage more of a reality than has ever been possible with earlier systems.

The Control Concept seen in Figure No. 7, illustrates an Air-Traffic Control situation with the runway or landing strip located at the center of the display. The illustration shows only three targets, but the actual number that could be displayed would be limited only by the amount of information available or by the number of aircraft that human engineers determine that a Controller can effectively handle. It should be noted that not only is there a volumetric display, but also a PPI display to which the operator is normally accustomed. It is also possible to add other target identification techniques such as bands of color, denoting height for different altitudes, beacon codes, future or past-target-position tracks, etc. The entire display in each case is generated solely by one rotating EL panel.

The SURVEILLANCE Concept seen in Figure 8, illustrates a military volumetric or "3-D" type of display. Again, the target information data that can be displayed is quite extensive.

The SATELLITE TRACKING Concept seen in Figure 9, illustrates the true positions of objects in space. This type of display could be changed to display various planets of the universe with any type of modified space relationship.

The COMMAND CENTRAL Concept seen in Figure 10 illustrates a tactical command decision-type display. This display illustrates data such as ships on the ocean, aircraft above the surface of the ocean. This display can present all of the target identification techniques as previously noted, and all with the one rotating EL panel.

Practical Military Application

This system, due to its demonstrated flexibility, makes itself readily useful in systems concepts presently in use, such as the Semi-Automatic Ground Environment (SAGE) or the Naval Tactical Defense System (NTDS). The following is a discussion of a "3-D" System built for use with such strategic planning systems.

Figure No. 11, SIMPLIFIED SYSTEM BLOCK DIAGRAM, illustrates functional logic of the system. The basic elements of the "3-D", display system are identical to those discussed earlier. The major differences lie in the sections concerning Memory, Information Access, and Information Selection. A discussion of these sections follows.

Memory. A nonvolatile (nondestructive) type of memory system is utilized in a coincident current selection principle. The three basic reasons for the selection of a nonvolatile system are: (1) a computer or data processing device will require the least modification to allow operation with the "3-D", (2) if a power or intermittant failure occurs there is the assurance of no lost information, and (3) a unique application for flexibility and reliability of design is available.

Information Access. Processed data either from a volumetric radar, or from a computer may be accepted. The information is inserted into the Input Buffer directly from the Computer where it has been changed from analogue to binary.

Information Selection. Either the Radar or the Computer initiates system operation by informing the Control Section with a Start Signal. The information from the Input Buffer is inserted into the appropriate Range, Height, and Azimuth Registers or Matrixes by the access gate. The Azimuth Encoder information is amplified and compared with the Azimuth Memory information. A Compared Trigger is generated upon proper comparison which in turn selects the Range and Height information at this azimuth. A Channel-Selector Translator then selects the correct Range and Height Line from the selected memory information. A gated AC Signal is then allowed to pass down these lines to the EL Panel via the slip-ring and brush-block assembly. The Compared Trigger also informs the Control Section that the display is in operation, and to allow no new information during this gating period. After the gating time a resume signal is sent to the Computer to inform it that new information can be accepted.

Conclusion

A "3-D" display system with advantages outweighing the disadvantages is now feasible. Further work in refinement of this system promises to result in a very useful tool for the system designer, who, in this present rapidly expanding technology must think in terms of more and more information handling capability while still retaining the human understanding necessary for decision making. This display is a step in that direction.

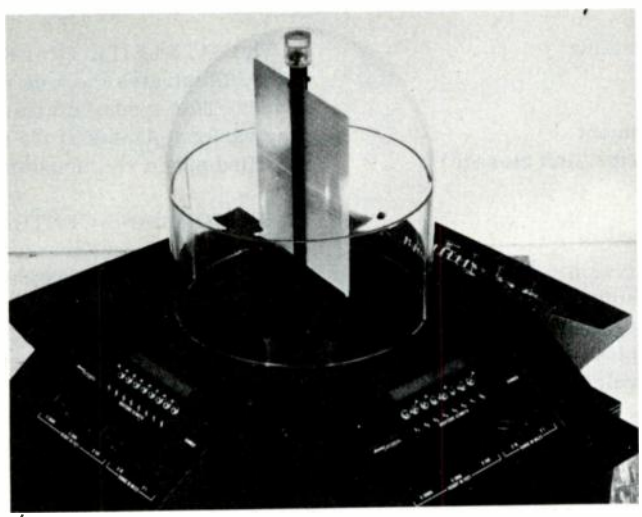


Fig. 1.

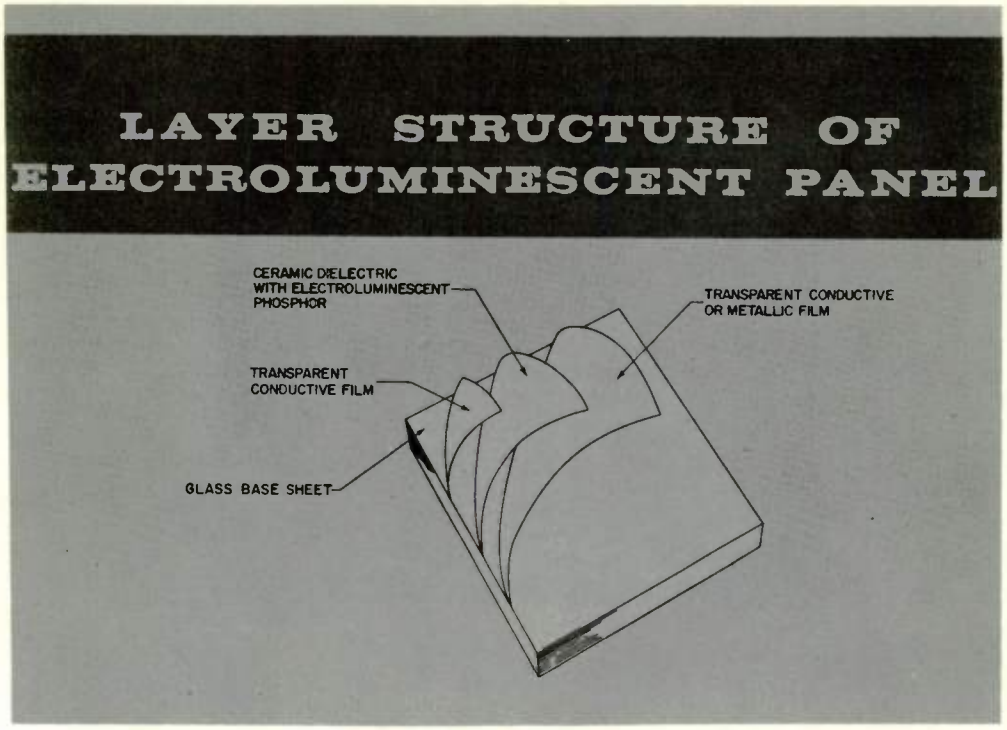

LAYER STRUCTURE OF
ELECTROLUMINESCENT PANEL
CERAMIC DIELECTRIC
WITH ELECTROLUMINESCENT
PHOSPHOR
TRANSPARENT CONDUCTIVE
OR METALLIC FILM
TRANSPARENT
CONDUCTIVE FILM
GLASS BASE SHEET

Fig. 2.

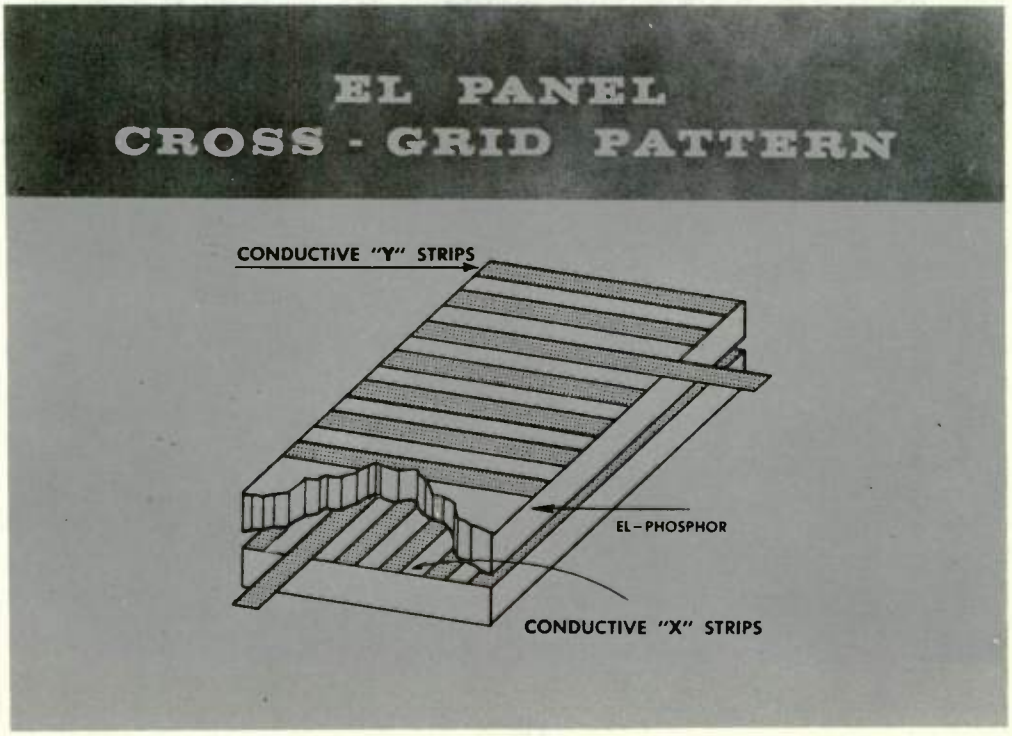

EL PANEL
CROSS - GRID PATTERN
CONDUCTIVE "Y" STRIPS
EL-PHOSPHOR
CONDUCTIVE "X" STRIPS

Fig. 3.

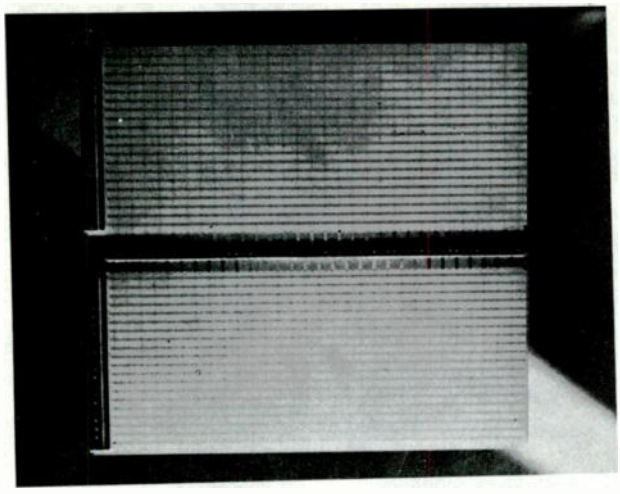


Fig. 4.

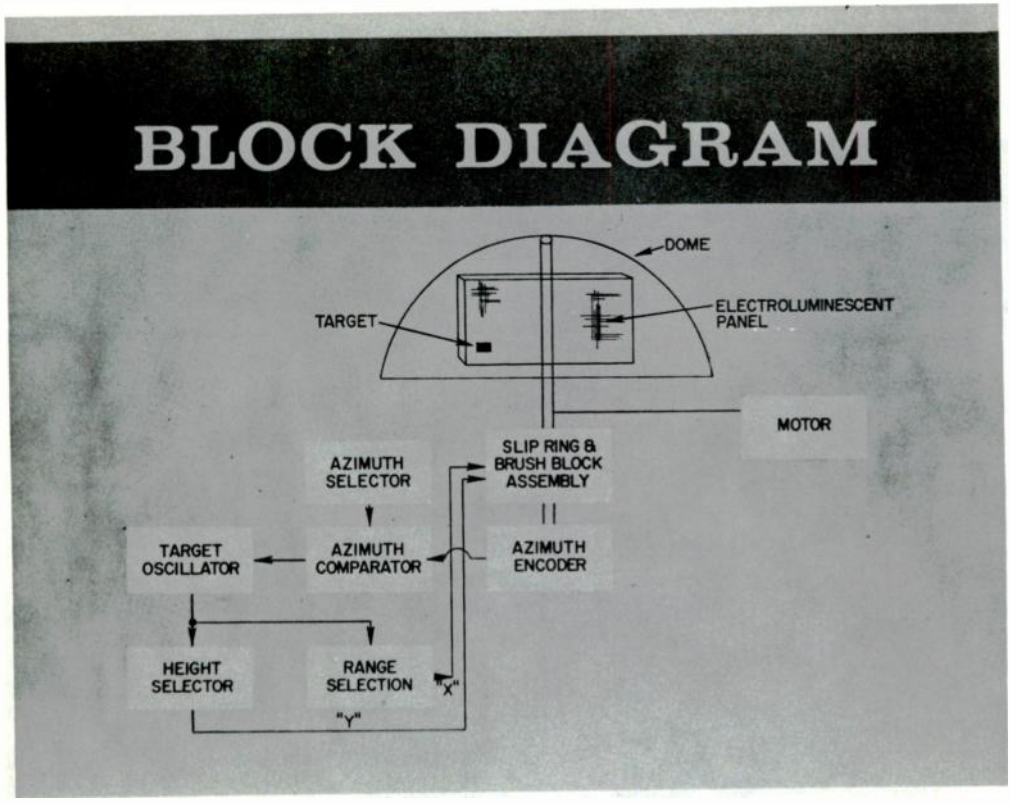

BLOCK DIAGRAM
DOME
TARGET
ELECTROLUMINESCENT PANEL
MOTOR
SLIP RING & BRUSH BLOCK ASSEMBLY
AZIMUTH SELECTOR
TARGET OSCILLATOR
AZIMUTH COMPARATOR
AZIMUTH ENCODER
HEIGHT SELECTOR
RANGE SELECTION
"X"
"Y"

Fig. 5.

Fig. 6.

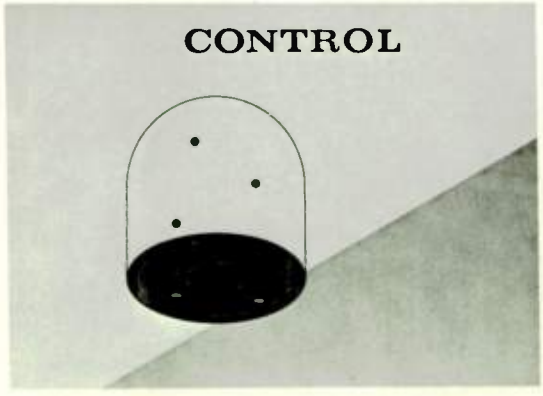


Fig. 7.

Fig. 9.

Fig. 8.

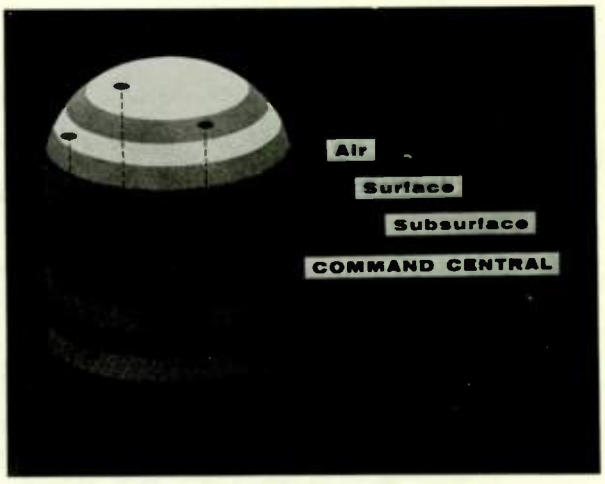


Fig. 10.

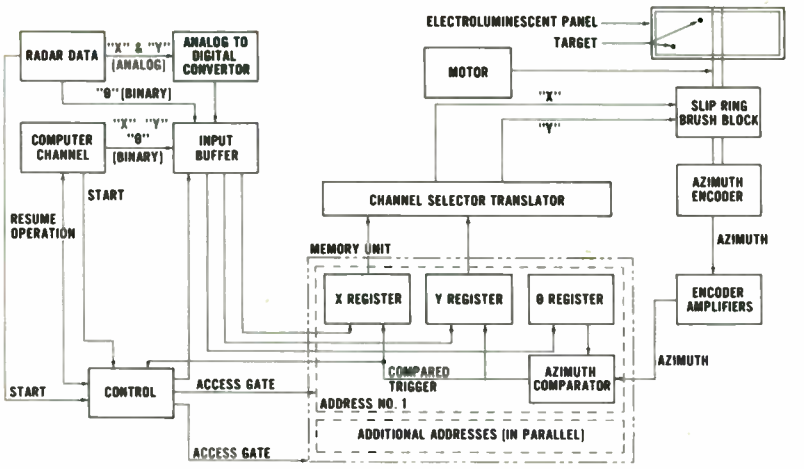

ELECTROLUMINESCENT PANEL
TARGET
RADAR DATA
"X" & "Y" (ANALOG)
ANALOG TO DIGITAL CONVERTOR
MOTOR
"θ" (BINARY)
"X"
"Y"
SLIP RING BRUSH BLOCK
COMPUTER CHANNEL
"X" "Y" "θ" (BINARY)
INPUT BUFFER
AZIMUTH ENCODER
START
CHANNEL SELECTOR TRANSLATOR
RESUME OPERATION
AZIMUTH
MEMORY UNIT
X REGISTER
Y REGISTER
θ REGISTER
ENCODER AMPLIFIERS
AZIMUTH
START
CONTROL
ACCESS GATE
COMPARED TRIGGER
AZIMUTH COMPARATOR
ADDRESS NO. 1
ADDITIONAL ADDRESSES (IN PARALLEL)
ACCESS GATE

Fig. 11.

THE ELECTROSTATIC VIDICON AND METHODS OF EVALUATION

R. J. Doyle
Westinghouse Electric Corporation
Elmira, New York

The electrostatic vidicon, developed by Westinghouse, is an image tube sensitive to visible and near visible electromagnetic radiation, that features resolution capabilities similar to those of the electromagnetic types. The tube also incorporates high deflection sensitivity, low power consumption, and ruggedization.

During this electrostatic vidicon development program specialized evaluation techniques were devised and used.

Introduction

The vidicon is an electronic transducer that converts electromagnetic radiation of the visible and near-visible range into electrical signals. These video signals can then be transmitted over wire or carrier-type circuits and the video information viewed on television or similar type displays.

Today the vast majority of vidicons in use are of the electromagnetic type which require bulky, power consuming external focus and deflection apparatus. The electrostatic vidicon, however, is completely self-contained in that all the elements required to control the electron beam are within the tube's one inch diameter envelope. Thus, this electrostatic tube finds many applications where size, weight, and power consumption must be kept to a minimum.

Tube Parameters and Operation

The electrostatic vidicon is basically a miniature cathode-ray tube with a photoconductor in place of the phosphor screen, see Figure 1. In fact, tubes have been built using the same geometry and electron optics with a phosphor output; such tubes are very useful for certain aspects of design evaluation and in the early stages of camera chain breadboarding.

Beam Generation. The electron beam in a vidicon is generated from a triode structure similar to that found in most cathode-ray guns.[1] However, after the beam passes into the G2 region it is mechanically masked by a tiny aperture - normally between .0005 and .005 inches in diameter. In the low velocity gun this brute-force reduction in beam diameter creates a small spot which is the object of the main focus lens. This technique is employed in the triode structure to aid in minimizing the diameter of the spot that ultimately lands on the photoconductor.

Focusing. Focusing of the electron beam in the electrostatic vidicon is accomplished by a straight cylinder type lens.[2] The lens consists of three electrodes: the grid 2 cylinder, the first cylinder of the grid 4 group, and the grid 3 cylinder which is adjusted for optimum focusing action. The grid 3 is normally operated at about 25% of the grid 4 voltage.

Deflection.[3] The beam after it emerges from the focus lens enters the deflection plates where deflection is accomplished by two consecutive sets of parallel plates located at right angles to each other and separated by an aperture. This aperture electrode between the two sets of plates minimizes the capacitance, and thus the crosstalk, between the horizontal and vertical sets of deflection plates.

The deflection plates themselves are so designed and located as to provide maximum deflection sensitivity and yet maintain minimum mechanical size. And since the ratio of: (1) the distance from the center of vertical deflection to the photoconductor, to (2) the distance from the center of horizontal deflection to the photoconductor, is nearly unity, the extensive defocusing and distortion effects theoretically attributed to the sequential deflection system are undetectable in the operation of the tube.

Collimation. Subsequent to deflection the electron beam is collimated to insure its landing at all points orthogonal to the photoconductor. Without this lens the beam would strike the photoconductor at angles increasingly less than 90° as it is deflected from the center to the edge of the faceplate. If this were to occur, the beam's velocity component normal to the photoconductor surface would decrease causing shading and resolution degradation from the center to edge of the generated image.

Photoconductor. The photoconductor is a thin semiconductor material, sensitive to visual and near-visual electromagnetic radiation, that is deposited on a transparent conductive lamina on the inside of the faceplate. These materials in combination with the electron beam act like an array of elemental capacitors and resistors whose resistivity varies as a function of the amount of incident radiation.

Initial scanning of the photoconductor with the low velocity electron beam stabilizes

its back surface at cathode potential. Thus, by applying a positive voltage to the conductive lamina a potential gradient is set up across the photoconductor. Where radiation falls on the target a more conductive path is established; a current flows through the photoconductor and the potential of the illuminated photoconductor elements rises towards the conductive lamina potential. When the beam rescans the photoconductor it restores the elements to cathode potential, thus giving rise to a signal current.[4]

Design Objectives

The objectives of this development program were simply to develop a tube with the same transducing capabilities as the electromagnetic vidicon, while retaining the inherent advantages of electrostatic type tubes. Ruggedization and low heater power were also to be incorporated for missile and airborne applications.

Resolution. Resolution is a measure of the number of elements or bits of information a device can resolve, and must be qualified by an indication of the relative contrast between bits. In the normal mode of operation, the electromagnetic vidicon is capable of resolving 600 TV lines, at 10% response factor, in the center of the tube.[5] The electrostatic vidicon should be able to do the same.

Deflection Sensitivity. The deflection sensitivity of the electrostatic vidicon is one of the basic advantages this type has over the electromagnetic. Therefore, the deflection system should have maximum sensitivity, minimum defocusing and distortion effects, and yet be a simplified type of mechanical structure in the tube.

Power Consumption. The electrical power needed to operate an electrostatic tube is inherently much less than that required to run an electromagnetic type. To further enhance this inherent advantage, a low power heater and compatible cathode were to be designed.

Ruggedization. Since the electrostatic vidicon obviously lends itself to use in aircraft and satellite video systems it must be ruggedized to the point where it will withstand aircraft vibration, and the environmental conditions associated with rocket blast-off, without degradation in its performance.

Measuring Techniques

As the electrostatic vidicon development program progressed, it became necessary to establish thorough testing procedures so that the relative performance of each tube design could be accurately evaluated.

Resolution. Resolution,[6] as was mentioned under "Design Objectives," is a measure of the ability of a device to resolve picture detail. The most common method for determining the resolution of a vidicon involves the imaging of the RETMA test pattern on the photoconductor and determining from a television monitor the point where the black and white bars in the resolution wedge are no longer distinguishable. This method is fine for setting up camera chains and for getting a rough idea about the performance of a tube, however, for the laboratory evaluation of electrostatic vidicons, where slight differences between tubes must be detected, the technique was not found adequate.

Therefore, a new type of test pattern was devised (Figure 4) which consists of nine line-groups; each of these groups contains a black and white bar to represent 100% response factor, followed by ten sets of four black and three white lines which represent 100 to 1000 TV lines, in 100-line increments.

This new chart makes it possible to obtain data for a complete square-wave aperture response curve with one oscilloscope presentation. Figure 5 is typical of the type of presentation that is obtained from an oscilloscope that is fed the video signal and set to select one horizontal scan line when an image from the "line selector" pattern is being transduced.

The procedure for obtaining the aperture response curve is as follows:

1. The tube is set up under the desired test conditions, in this case optimum overall performance. (Figure 2 shows the RETMA test pattern image for a basis of comparison, while Figure 3 was obtained under the same operating conditions with "aperture compensation" added.)

2. The RETMA test pattern is replaced in the light box with the "line selector" pattern (Figure 4).

3. The video signal from a selected horizontal line is presented on an oscilloscope and a photograph is taken. See Figure 5.

4. The photograph is measured for the relative response of the line-sets to that of the black and white bars, and a curve is plotted. See Figure 6.

From this procedure outline, and associated figures, one can equate the square-wave aperture response curve to the familiar presentation of the RETMA test pattern.

The square-wave aperture response curve,

now so easily obtained, proved to be a most useful tool for the detailed study of factors affecting resolution.

Parameters and Their Effects on Resolution. In reality probably everything within the vidicon effects the resolution to some extent, however, some effects are more prominent than others and are described below.

Grid 2 Voltage. In the electrostatic vidicon, the Grid 2 (G_2) is a cylindrical electrode containing apertures which help control the final diameter of the electron beam. The G_2 serves several functions, two of which are to accelerate the electrons toward the photoconductor, and to act as part of the focusing lens system.

By varying the G_2 voltage independent of the other electrode voltages, and measuring the aperture response at discrete points after the beam had been refocused, it was found that as the G_2 voltage was lowered the resolution, in most designs, tended to increase. Of course lowering the G_2 voltage also reduced the strength of the field on the cathode space charge and at voltages which were very low the beam was often insufficient to charge the back surface of the photoconductor.

From this type of test it has been found that for a best compromise the G_2 should be operated at between 200 and 250 volts.

Focus Voltage. As would be expected, the focus voltage has a marked degree of control on the resolving power of the tube. Of interest, however, is just how much control.

The sensitivity of the focus electrode (G_3) was determined in the following manner.

The tube under test was set up in the camera chain with an image of the "line selector" test pattern focused on the photoconductor. Then data for a series of aperture response curves were taken as a function of G_3 voltage, using the procedure previously outlined. Figure 7 is a typical set of curves from such an experiment except that for clarity, the curves for 71, 72, 73, and 74 volts were omitted since they lie in approximately the same positions as the curves shown for 69, 68, 67, and 66 volts.

At the point of 50% response factor (indicated by the dash line) a measurement of the number of TV lines resolved was made and the data were then plotted as a function of the G_3 voltage (see Figure 8). In this graph the data from below and above optimum focus voltage are both included.

From the slopes of this curve are measured the change in resolution, at 50% response factor, in TV lines per volt. As measured from Figure 8, the focus sensitivity over the straight parts of the curve is 62 TV lines per volt. Other detailed studies of the curve within the area of $\pm$ 1 volt of optimum focus, have revealed that the sensitivity curve is relatively flat at the top.

Grid 4 and D.C. Deflection Plate Voltage. The G_4 voltage is internally connected to three electrodes: 1) the entrance aperture in front of the horizontal plates which forms part of the focus lens, 2) the crosstalk reducing aperture between the two sets of deflection plates, 3) the post-accelerator which follows the vertical set of deflection plates and forms part of the collimating lens. The deflection plate D.C. voltage is set at the same value as the G_4 voltage except for 2 or 3 volts difference between the sets of plates for astigmatism correction.

Here again, as in the focus sensitivity experiments, the G_4 and deflection plate voltages were varied together, and at discrete voltages aperture response curves were plotted. The data plotted for the most recent designs show some increase in resolving power above 250 volts, however, not enough to recommend operation much above 300 volts.

Supply Voltage. In many electrostatic vidicon cameras the voltages for all the positive electrodes are developed from a single power supply through a series of voltage divider networks, and in the course of designing such cameras the question arises as to how much regulation is required in this main power supply. Thus a series of aperture response curves were plotted as a function of the change in the main supply voltage. The filament, Grid 1, and photoconductor were supplied from independent sources.

Before analyzing the results presented note that the G_1 voltage plays a major role in determining the shape of the resulting curve, and that the data for the curve shown were taken under conditions of fixed G1, e.g., the G1 was set so that when the power supply was lowered by 50 volts there was still sufficient beam current to charge the back surface of the photoconductor.

For the experiment the power supply was set at 500 volts, the focus optimized, and then the power supply voltage was adjusted and data, as in previous tests, were taken for aperture response curves, then at 50% response factor on each curve the number of TV lines resolved was measured and plotted as a function of the change in power supply voltage (see Figure 9). It is felt that these results are the most significant to the camera designer, when for example, the tube will be operating in a satellite or moon probe where a minimum of control will be available. Figure 9 illustrates that even with fixed G_1 voltage the tube suffers relatively little resolution degradation over a 30 volt change in

power supply voltage.

When the same experimental procedure is followed, but the G_1 voltage is adjusted at each power supply voltage setting for optimum beam landing at the photoconductor, the curve continues to rise beyond 500 volts as would be expected from basic principles of electron optics.

Deflection Defocusing. From electron optic theory it is known that beam defocusing occurs when the distance from focus lens to the image plane changes without equivalent lens refocusing. Such a change takes place in the electrostatic vidicon when the beam scans from the center to the edge of the raster area.

To investigate this deflection defocusing a series of aperture response curves were plotted using the nine line groups on the "line selector" test pattern. Figure 6 indicates a typical change in response from the center to corner of the raster area with no refocusing. This center to corner degradation can, to some degree, be relieved by dynamic focusing.

Deflection Sensitivity and Linearity. Deflection sensitivity is a measure of the voltage change required on each deflection plate to generate a specified raster size. In the one inch vidicons the standard raster is 1/2 x 3/8 of an inch.

Two methods have been employed to measure the deflection sensitivity in the electrostatic vidicon. The first method uses the electron gun in a standard vidicon structure. Here a collimated light source is projected through a 1/2 x 3/8 inch rectangular opening in an opaque template that is centered over the faceplate of the tube. Then from a condition of overscan, in the normal vidicon operating mode, the raster size is decreased until, as observed on a monitor, it just fills the illuminated area. Then with an oscilloscope the peak-to-peak sawtooth voltage being fed to the deflection plates is measured.

The second method requires the building of a special tube with a TV-type phosphor coated faceplate. With this tube, containing the electrostatic vidicon gun set at the desired operating voltages, the beam is statically deflected across the phosphor and the voltage required to move a specified distance is measured.

Figure 10, indicates the peak-to-peak sawtooth voltage required on each deflection plate, in push-pull operation, as a function of the D.C. potential on the deflection plates. The horizontal plates, although slightly more sensitive than the vertical plates, require more deflection voltage because of the four to three ratio between horizontal and vertical scanning distances.

Within the tube itself the linearity of deflection has been studied using the same phosphor output tubes used for the second method of measuring sensitivity with a calibrated graticule affixed to their faceplates. Then the beam was focused and indexed across the phosphor from one intersection to another; each time recording the deflection plate voltages. The maximum variation found in the voltage required to cover the standard raster area in any of the recent dip-beaded type gun structures was well within tolerable limits.

Vibration Effects. Figure 1 shows the WX-4871 ruggedized, low heater power electrostatic vidicon that has been thoroughly evaluated in the X, Y, and Z axes for the effects of vibration of its performance.

In the non-operating evaluation procedure the tube is vibrated at a particular acceleration level while the frequency is automatically cycled within a specified range. Under these conditions the tube is examined with a synchronized stroboscope flashing at a rate one cycle per second less than the frequency at which the tube is moving. Thus, any resonances in the tube structure appear as slow motion movements and their amplitudes can be measured.

Operating vibration tests are performed using special holding fixtures. A test pattern is affixed to the faceplate of the tube while a collimated beam of light floods it and generates an image on the photoconductor; this eliminates the complications of a vibrating lens system. Using this technique, it is possible to view on a monitor any degradation in tube performance due to vibrations similar to those that take place when the tube is operating in an aircraft or rocket.

Test Results

The test results from the latest electrostatic vidicon designs indicate a marked improvement since the beginning of the development program—the various illustrations used in the text of this paper are typical of the performance of these tubes.

Resolution. The resolution (better defined as aperture response) of the present electrostatic vidicon designs is, in many instances, equal to that of the standard one inch electromagnetic types. However, the development program continues to further improve the electron-optics, making use of gun structures accurately controlled by means of the dip-beaded construction technique. This method of assembly, commonly used in cathode-ray gun construction, insures reproducibility and rigid placement of the many elements in the tube.

Deflection Sensitivity. Deflection sensitivity has been maximized to the point where deflection

is easily achieved with simple transistor drive circuitry while deflection defocusing and structure size have been kept to a minimum.

Power Consumption. The power consumption of the low power electrostatic vidicons, with a .945 watt heater, is only a fraction of that required by the standard magnetic types, and although work continues to develop even more efficient cathodes, it is felt that this tube fulfills the present-day low power requirements. It is realized that heater-cathode structures exist which operate at lower heater power, however, the reliability and degree of ruggedization possible using such structures is not always satisfactory for "space-age" environments.

Ruggedization. The ruggedized types of electrostatic vidicons are all similar in appearance to the WX-4871 shown in Figure 1, and their net weight is less than three ounces.

Two ruggedization requirements have been imposed on the tube, one operating and the other nonoperating. Nonoperating tests apply under circumstances where, for example, the tube will be required to survive the shock and vibration of rocket blast-off and then be turned on when the satellite is in orbit. The operating tests apply where the tube is required to be on during blast-off or severe environmental conditions.

All the tubes similar to the WX-4871 with their electrodes embedded in the multiform glass beads have had little trouble meeting the imposed requirements. One problem was encountered, in operating tests, from spurious signals being generated by the movement of the thin G_5 mesh electrode. This trouble has been eliminated to a great extent but has set the ultimate vibration and shock levels that the tube can withstand while operating. Work is continuing to eliminate this final barrier.

Conclusions

Certainly the objectives to develop an electrostatic vidicon with low power consumption, high deflection sensitivity, etc. have been met and incorporated into tubes now in pilot production. Yet, work continues to develop new and more improved tubes to meet the ever increasing demands of defense and space-exploration. And, although not mentioned in this paper, an extensive program is being carried on to develop improved photoconductors.

It should also be mentioned that the same electron guns developed and used in this electrostatic vidicon program are being successfully incorporated into the low velocity scanning sections of the all electrostatic scan-converter, ebicon, and uvicon.

Acknowledgement

The author is much indebted to Mr. C. Macdonald, Dr. H. Moss, and Mr. R. Simms of the Westinghouse Electronic Tube Division, who read the original manuscript and suggested significant improvements.

References

1. O. Klemperer, Electron Optics (Cambridge, The University Press, 1953), pp 264-269.
2. J. R. Pierce, Theory and Design of Electron Beams (D. Van Nostrand Co. Inc., 1949) pp 72-75.
3. I. G. Maloff and D. W. Epstein, Electron Optics in Television, (McGraw Hill Book Co., Inc., 1938) pp 190-210.
4. H. G. Lubszynski, S. Taylor, and J. Wardley, "Some Aspects of Vidicon Performance," Journal of the British IRE, Vol. 20, May, 1960, pp 323-334.
5. Martin Rome, "Recent Advances in Vidicons," IRE Transactions, Vol. BC-7, No. 3, August, 1961, pp 12-18.
6. "IRE Standards on Video Techniques: Measurement of Resolution of Camera Systems," Proceedings of the IRE, Vol. 49, March, 1961, pp 599-602.

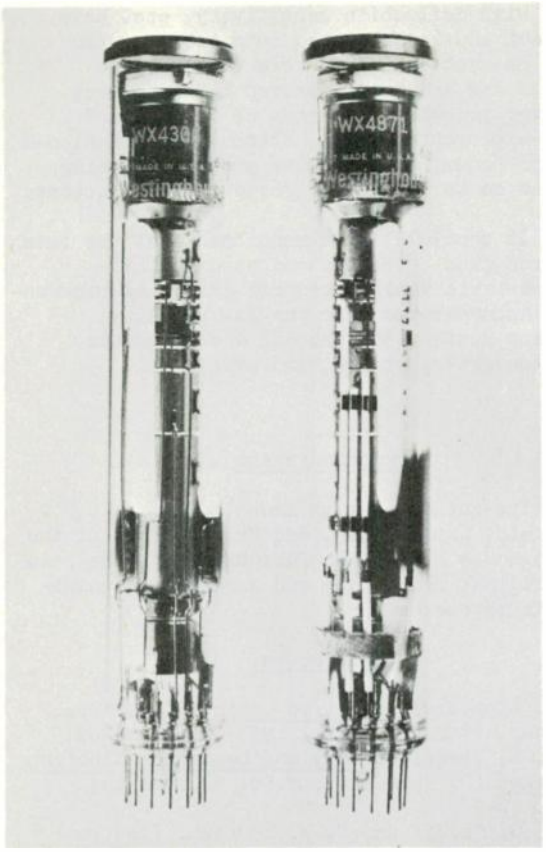


Fig. 1. WX 4306 Industrial type electrostatic vidicon and WX 4871 ruggedized type electrostatic vidicon.

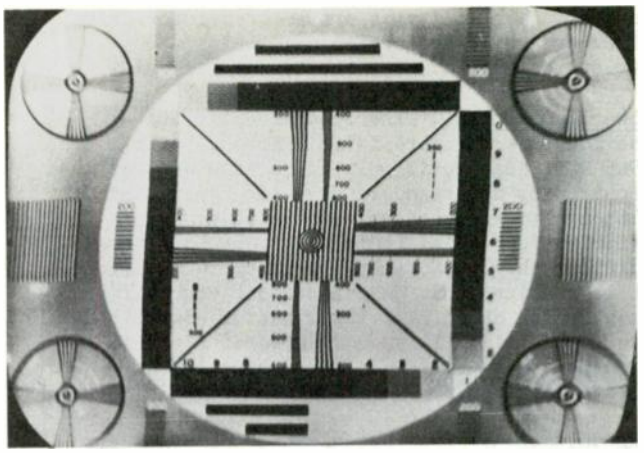


Fig. 3. Monitor display of RETMA test pattern (with aperture compensation).

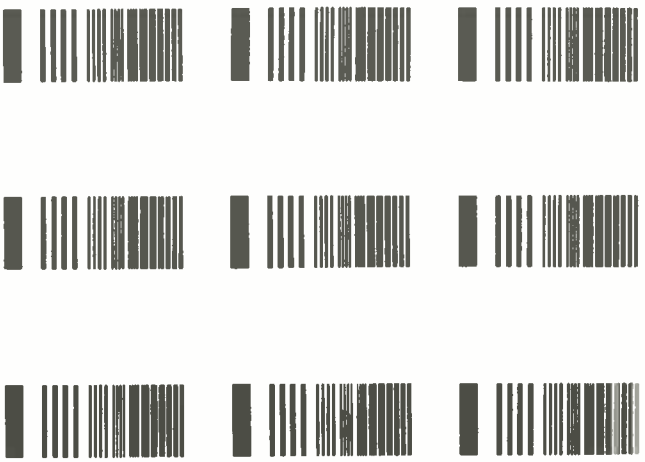


Fig. 4. "Line-selector" test pattern.

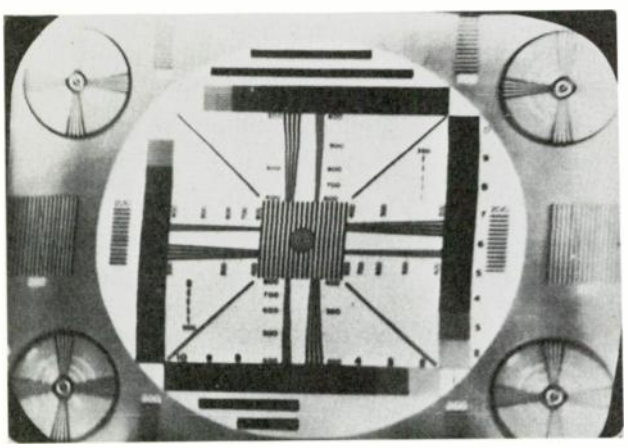


Fig. 2. Monitor display of RETMA test pattern (without aperture compensation).

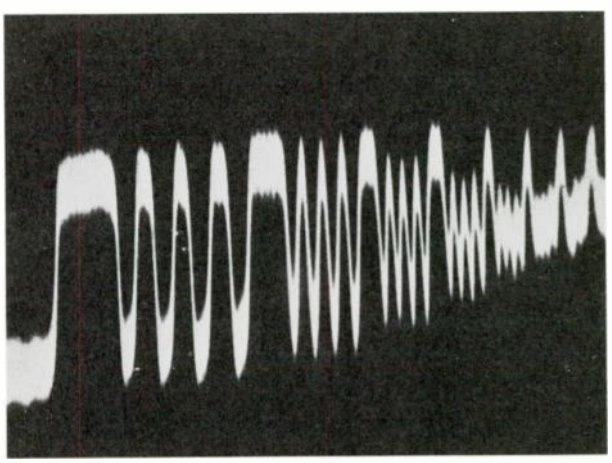


Fig. 5. Oscilloscope presentation of video signal from selected horizontal scan line (Fig. 4 being imaged).

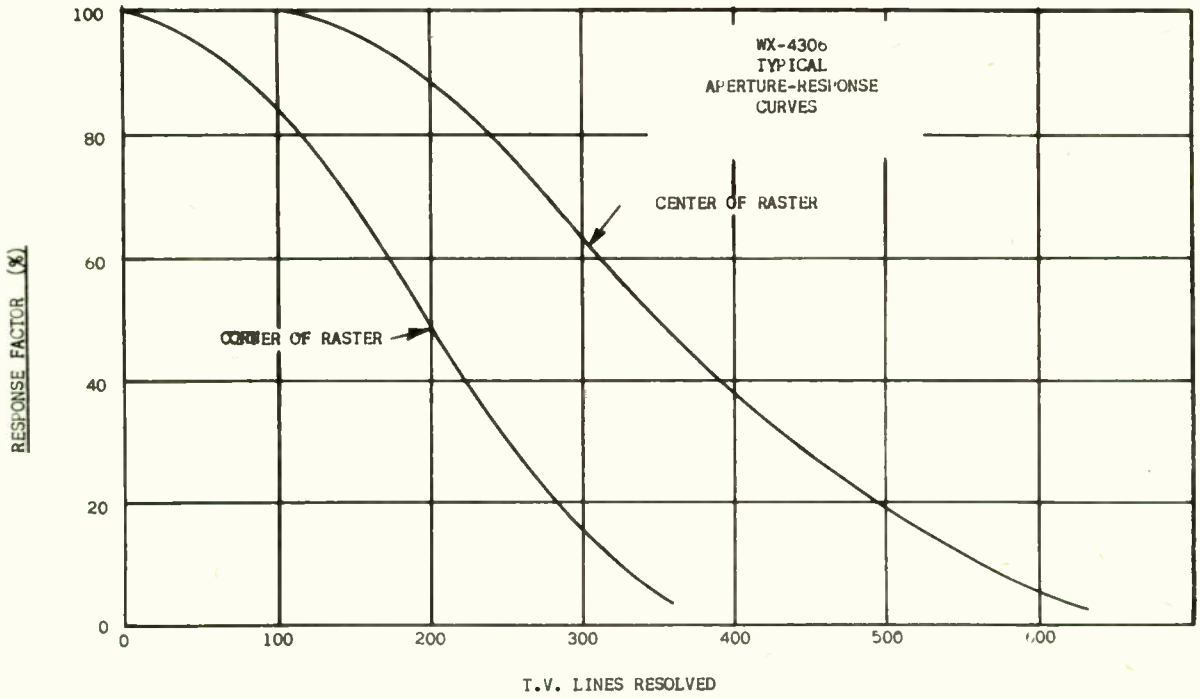


Fig. 6. Typical square wave aperture—response curves for electrostatic vidicon.

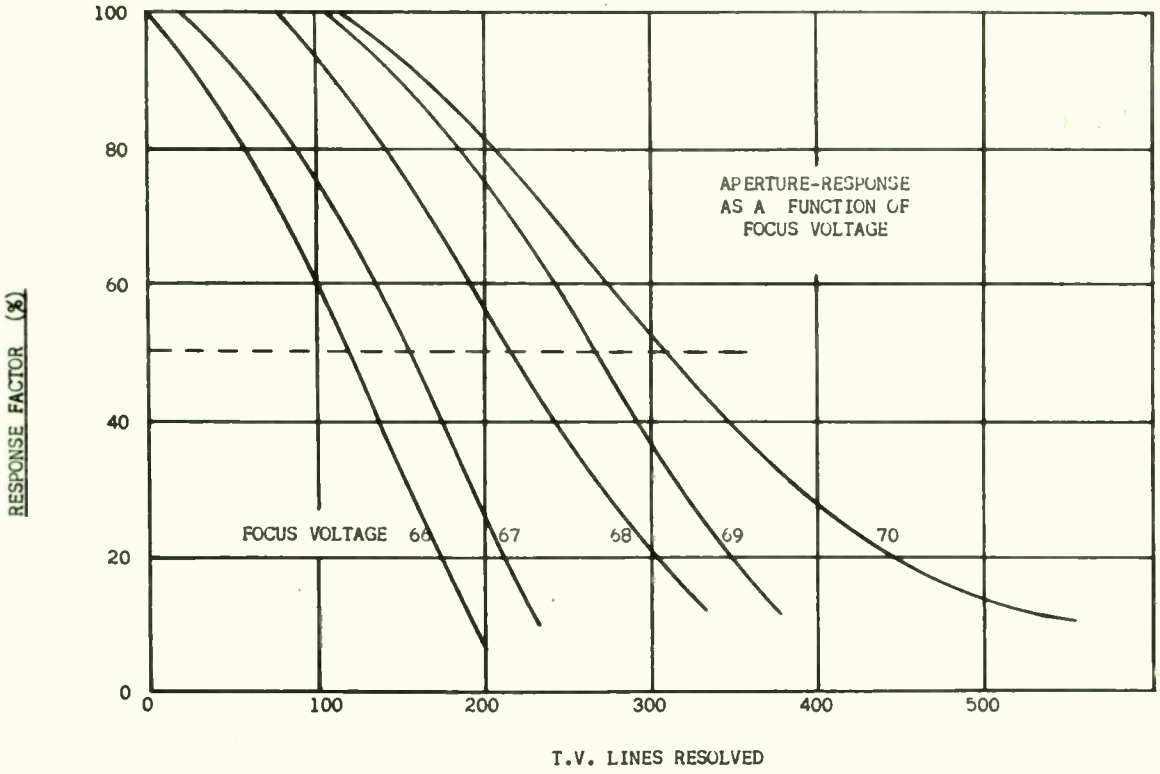


Fig. 7. Square wave aperture response as a function of focus electrode voltage.

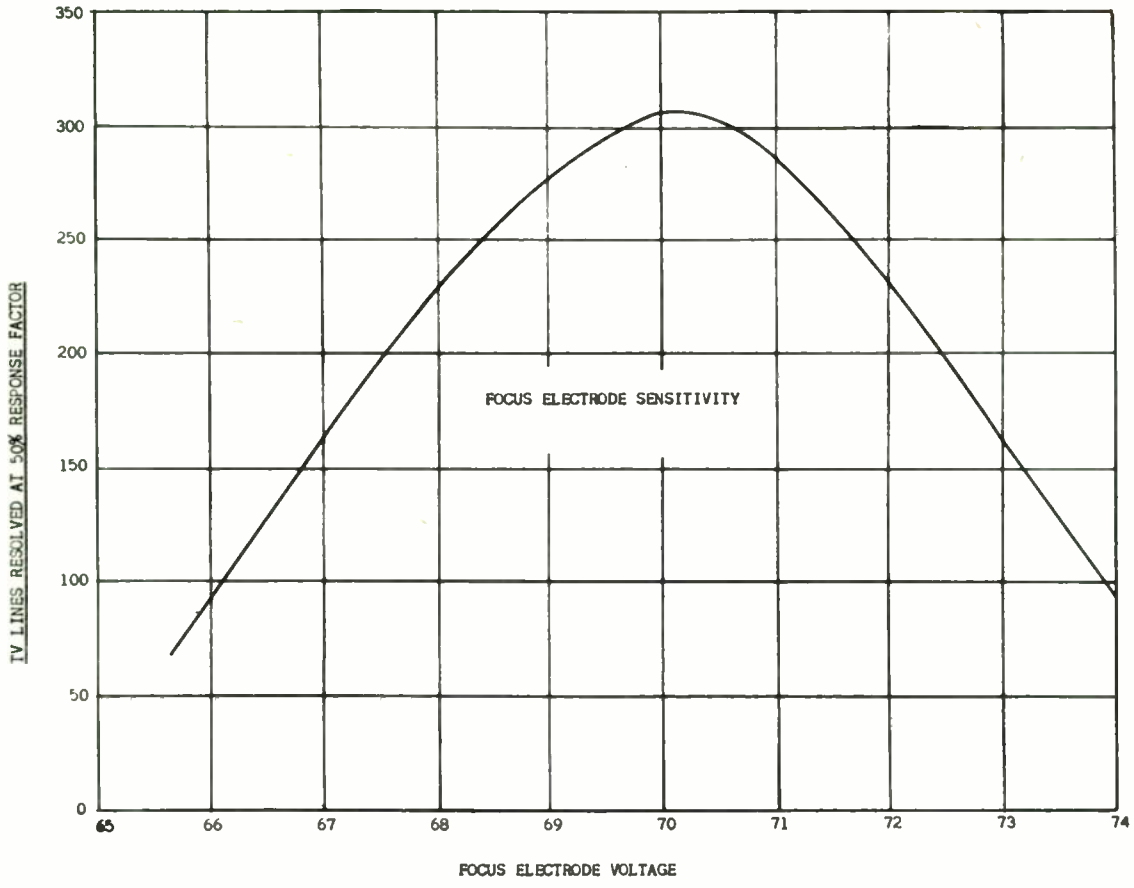


Fig. 8. Focus electrode sensitivity.

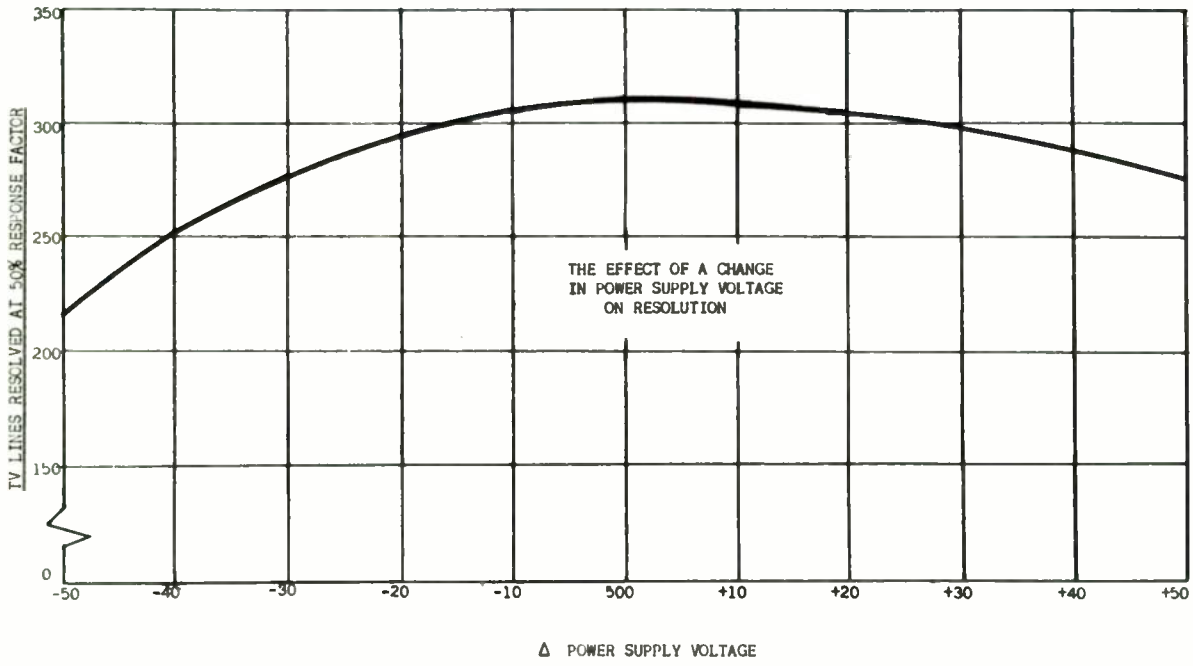


Fig. 9. The effect of a change in power supply voltage on resolution.

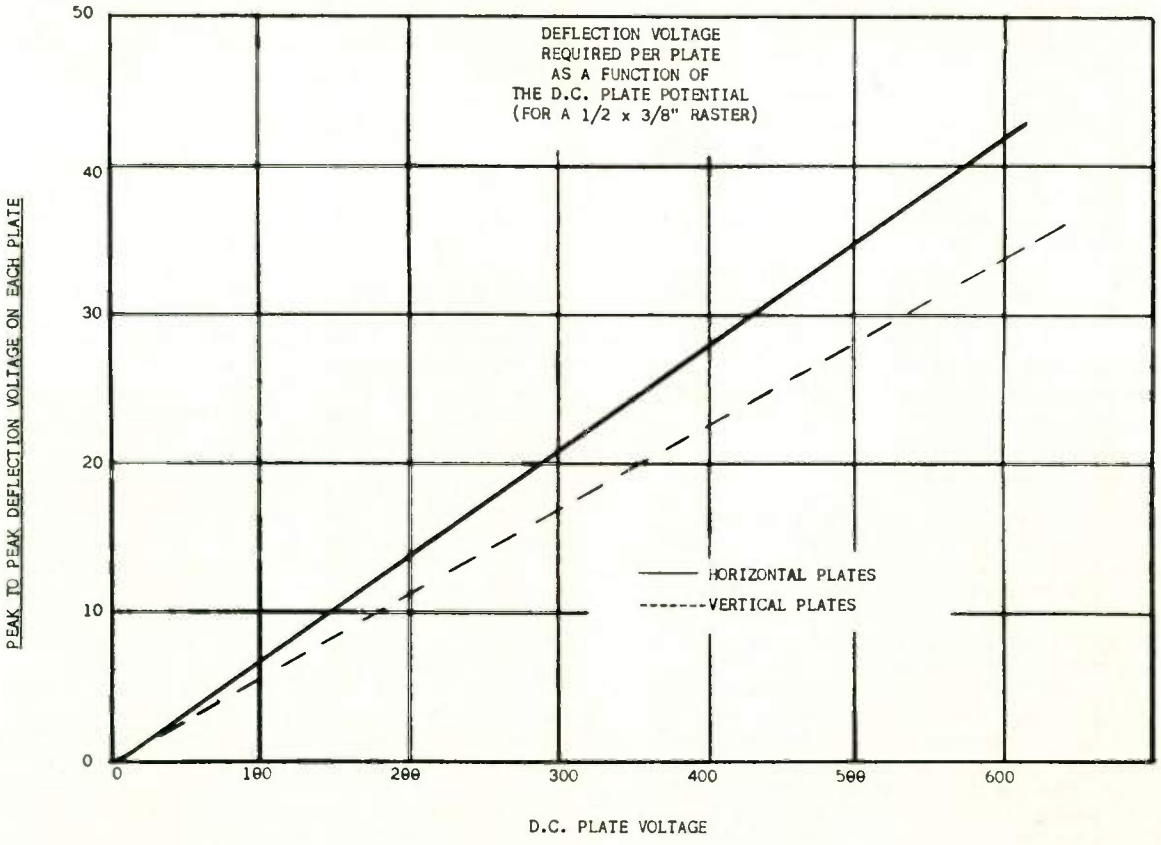


Fig. 10. Deflection plate sensitivity.

NEW DEVELOPMENTS IN ULTRA-FAST WARM-UP PLANAR TUBES

J. M. Connelly and D. D. Mickey
General Electric Company
Owensboro, Kentucky

Abstract

Present day military weapons systems require the reliability and the high frequency performance made possible by using modern ceramic planar tubes, but in addition, these weapon systems need to possess the ability to perform their mission with a few seconds notice.

As a result of this requirement, a new heater-cathode structure has been designed that enables high performance planar tubes to be produced that have a warm-up time of less than three seconds. This paper discusses the methods used to analyze the thermal problems and obtain solutions leading to the achievement of this ultra fast warm-up time. The performance characteristics of planar tubes using this new heater-cathode structure will also be disclosed.

This new heater-cathode structure is adaptable to other types of tubes, such as cathode-ray tubes and traveling wave tubes.

Introduction

There are a number of important military applications in which the ability of electronic systems to achieve full operation in a short period of time is a prime requirement. Many of these systems also require the superior high frequency performance and the ability to withstand severe environments made possible by use of modern planar ceramic tubes. A notable example of an application where fast operational warm-up is vitally necessary is in guided missiles intended for interception purposes. In addition to this specific use where fast availability of stand-by systems is essential, there is a wider area of interest in both industrial and military applications where fast warm-up of electronic systems is a desirable feature.

In recognition of the need for fast warm-up high frequency tubes, development programs at General Electric were accelerated to reduce drastically the warm-up time of planar tubes incorporated in or planned for use in these systems. The immediate objective of this program was a heater-cathode structure capable of delivering 90% of the steady-state current of approximately ten milliamperes in three to five seconds. The ultimate objective is incorporation of this fast heating cathode in a family of high frequency ceramic triodes. Realization of these objectives requires a reduction of warm-up time from approximately 20 seconds to three seconds.

This program resulted in the development of a unipotential heater-cathode structure having the ability to warm up in less than three seconds. This heater-cathode also possesses superior electron emission characteristics, and tubes employing this structure should provide reliable operation over long periods of time.

Methods of Approach

The first step in this program was a comprehensive thermal analysis of several proposed structures. In order to evaluate economically and accurately the effects on warm-up time of varying parameters such as thermal mass, emissivity and thermal conductivity, a thermal analogy was established on an analog computer. The accuracy of this study was greatly improved by determining boundary conditions in the analog circuit using data derived from prototype tubes having a number of features in common with the models being studied.

Two basically different types of heater-cathode structures were studied. The first of these depends primarily on radiation transfer of energy from heater to cathode. This represents a conventional approach since practically all tubes having electrically isolated cathodes utilize this type of heat transfer. The second type structure studied is based on conduction transfer of heat energy from heater to cathode. In this case it is necessary to physically bond heater and cathode in order for the conduction mode to predominate.

In order to present some idea as to the size of the tubes being considered in this study, a photograph of a family of prototype ceramic tubes is shown in Figure 1.

A cut-away drawing of a radiation transfer heater-cathode structure having reduced thermal mass but otherwise similar to those in the preceding tubes is presented in Figure 2. The various portions indicated were considered as separate reference points, or nodes, in the computer study. Actually eighteen nodes were considered in this study; for simplicity, only four typical nodes are shown in the drawing. Each of these nodes were receiving and losing energy in a manner illustrated in Figure 3. These relations can be expressed mathematically and were simulated in an analog computer by choosing appropriate analog quantities and establishing the necessary number of electrical

circuits corresponding to these nodes. It is of interest to note that this study represents the first known example where thyrites were used extensively to simulate the 4th power relationships involved in radiation type heat transfer.

Similar studies were made using models having bonded heater-cathodes. A cross-sectional drawing of this type structure is shown in Figure 4. Again nodal points of interest were chosen and the proper analog constants for thermal masses, thermal conductivities, heat sources and heat losses were applied to the computer circuits.

Results of Theoretical Studies

Typical results of a radiation model structure obtained using the analog computer are shown in Figure 5. The exponential nature of input power with a constant applied heater voltage is seen in the upper curve. This non-linear power input results from the large change in resistivity of tungsten heater wire with temperature and permits a faster warm-up than would be possible with a material having a lower temperature coefficient of resistivity. Predicted instantaneous values of heater and cathode temperatures are also shown in Figure 5. It is seen that steady-state cathode temperature is not attained until approximately ten seconds, although equilibrium heater temperature is reached in about four seconds. This time lag is caused primarily by the thermal mass of cathode components including the radiation shield that surrounds the heater. The thermal emissivities of heater, cathode and radiation shield are also factors in this time lag. By optimizing heater and cathode emissivities, decreasing thermal masses and increasing conduction losses along the cathode support it was predicted that warm-up time of radiation heated cathodes could be decreased to about five seconds. Other techniques requiring higher steady-state power input can be used to reduce this time to approximately three seconds.

Results obtained from studies of conduction transfer are shown in Figure 6 where predicted input power, heater temperature and cathode temperatures are plotted versus time. In this case 90% warm-up time should be effected within the desired three seconds.

Based on the results obtained in these studies, the major portion of subsequent developmental work was directed toward perfecting the conduction transfer structure.

Fabrication Techniques

Earlier work performed at the General Electric Research Laboratory had demonstrated that small planar tubes having bonded heater-cathodes would be practical to make. Consequently, a number of heater-cathodes containing various material combinations were evaluated in diodes. A ceramic diode was chosen as a test vehicle for these various cathodes because of its simplicity of construction. A number of these bonded units are shown in the photograph presented as Figure 7. The coil heater is bonded to a cathode cup and electrically insulated from it by means of a thin layer of ceramic which is applied to the desired areas and allowed to dry. Bonding between coil, ceramic and cup is effected at high temperature in a protective atmosphere. Another high temperature firing is used to braze heater legs to each terminal of the heater coil. This sub-assembly then is incorporated into the complete tube following procedures previously established at General Electric. A cylindrical cathode support is welded to the bonded cathode assembly, and in turn this support is welded to a cathode ring and ceramic sub-assembly as shown in Figure 8. The cathode in this photograph has emissive coating applied on its planar surface. The relationship of the cathode to other tube elements can be seen in the cross-sectional drawing of Figure 9. Although a radiation heater is actually shown in this drawing; the same comments apply to diodes having conduction heaters. Positioning of the cathode is such that the emissive surface lies flush with a ceramic reference surface and by this means the cathode is accurately located with respect to other tube elements. Metal parts are shaded and ceramic parts are dotted. Metal portions of the external envelope are titanium. This metal provides very desirable gettering action at the high temperatures1 (1000 C) employed to seal metal and ceramic parts together. It also is the source of the titanium used in the nickel-titanium metal to ceramic seal. The same basic materials and methods are used in the fabrication of planar ceramic triodes.

Results of Experimental Program

The degree of improvement in warm-up effected by these techniques is illustrated in Figure 10 where plate current is plotted versus time for three tube designs. Curve A represents a typical response curve of a prototype utilizing radiation transfer from heater to cathode. In the design of this tube type warm-up time was not a primary consideration;

time to achieve 90% steady-state plate current is approximately 23 seconds. Curve B gives results obtained with a similar tube containing a radiation transfer heater-cathode whose thermal mass has been reduced to a practical minimum. In this case the time required to achieve 90% warm-up was eight seconds. Finally, Curve C shows the warm-up attainable when the heater is physically bonded to the cathode and both are of minimum mass. A warm-up time of less than three seconds is realized; this value closely agrees with predicted values obtained during the study phase. Warm-up time as a function of ambient temperature is shown in Figure 11. The dotted curve was obtained during the computer study; the two measured values indicated again demonstrates the close correlation between predicted and actual values.

In addition to studying the effect on warm-up of varying such parameters as emissivity and thermal mass of components, the idea of using a series connected, non-linear resistor to improve warm-up time was explored. By choosing an appropriate value resistor having a longer thermal time constant than the heater of the tube and a positive temperature coefficient, it is possible to effect a substantial reduction in warm-up time. This technique is applicable to either conduction or radiation type heater-cathodes.

If a constant voltage source supplies power to the combination heater and non-linear or ballast resistor as shown in Figure 12, the instantaneous power in the heater can be varied over rather wide limits by varying the resistance or time constant of the ballast. The peak heater power can be adjusted by changing the electrical resistance of the ballast and the time for which higher than normal heater power is applied is determined by the thermal time constant of the ballast. This circuit can be designed to operate in several different ways. For example, it might be considered desirable to limit heater temperature to its normal steady-state value, or it may be that a small "overshoot" in cathode temperature for a short time is desired. Thus, considerable design flexibility is possible. The penalty involved in the use of such an arrangement is some increase in steady-state power requirements. However, a properly designed combination should give more reliable operation in a more compact package than programmed heater supplies involving some devices that are commonly used, such as timers and relays.

The effect on plate current of a ballast in series with the heater of a bonded heater-cathode is shown in Figure 13, where plate current versus time curves are plotted when the tube is operated with and without a ballast. Warm-up time, when the ballast is used, is reduced to 1.3 seconds. Steady-state cathode temperatures are approximately the same in all examples so that a true measure of performance improvement is shown.

The low mass, integrally bonded type construction that gives this heater-cathode its fast warm-up characteristics also results in an extremely sturdy unit demonstrating excellent resistance to mechanical shocks and vibration. Sample quantities of tubes were subjected to 35G swept frequency fatigue tests according to established procedures for reliable, military type tubes. The frequency of vibration was swept from 70 to 2000 cycles per second and returned with 10 minutes per sweep. This was done for one hour in each of three mutually perpendicular planes.

This type structure also possesses excellent stability during life. Life tests were conducted with heater voltage cycled (on for ten minutes and off for five minutes). Results are shown in Figures 14 and 15 where heater current and tube drop of a group of diodes are plotted versus time. The slight decrease of tube drop during life testing is a result of an increase in emission.

An increase in reliability should be effected by use of this type heater-cathode by virtue of the exceptionally low heater temperature required. A heater wire temperature of only 800°C is required here for proper operation whereas in conventional radiation transfer heater-cathodes heater temperatures in the order of 1200-1300°C are used. This reduction in heater temperature has several important advantages. The cooler heater is subjected to less mechanical stress during cycling. The chemical reaction rates between the heater wire and the insulating heater coating are decreased so that heater-cathode leakage and the probability of heater-cathode shorts are reduced. In addition, the lower heater temperature reduces any tendency for material to be evolved from the heater that might poison the cathode. This provides a higher and more stable emission level from the cathode.

Triodes

A few triodes incorporating bonded heater-cathodes have also been made and evaluated to a limited extent for the following characteristics:

Swept Frequency Vibration Output
Noise Figure (450 Mc)
Oscillator Power Output (450 Mc and 1.2 Gc)

Vibration

Further evidence of the rigidity of this bonded cathode is given in Figure 16 where swept frequency vibration output characteristics of a triode are shown. These curves were obtained by subjecting a triode to a variable frequency vibration at an acceleration level of 10G. The frequency of applied vibration was varied from 100 to 7000 cycles per second. The gain of the particular tube and circuit used was about 50 and the amplifier in the test equipment used to amplify microphonic output was essentially flat from 20 cps to 50 KC. Thus, any harmonics being generated within the tube would be amplified and be present in the output. Maximum output over the entire range is about six millivolts peak to peak; this is a very low level for a high performance triode.

High Frequency Performance

When a typical triode was tested as a small signal amplifier at 450 Mc, noise figure and gain measurements were approximately 5 db and 14.5 db respectively. These measurements were made with impedances matched for maximum power transfer with a 7 Mc bandwidth. Performance as an oscillator at 450 Mc is shown in Figure 17 where oscillator power is plotted as a function of total cathode current at various plate voltages. Oscillator data obtained at 1.2 Gc and presented in a similar manner as before is shown in Figure 18. It is seen that approximately 500 milliwatts at 450 Mc and 350 milliwatts at 1.2 Gc are available at a cathode current of 12 milliamperes. It should be noted that these triodes were not optimized for high frequency performance since existing parts and assembly fixtures were used to a large extent. For this reason, these data should be considered tentative and subject to slight improvement. Tests at 1.2 Gc were made in a grid separation planar oscillator designed for production testing tubes for radio-frequency characteristics. This unit is shown in Figure 19. It is electrically similar to a coaxial type circuit with the added feature that tubes being tested are readily accessible.

Other circuits using tubes to which this heater-cathode can be adapted are shown in Figures 20 and 21. In Figure 20[3] a laboratory model grounded grid amplifier designed for 2.7 Gc operation is presented. The small unit to the left is an adjustable series output capacitor removed from the main unit here in order to show more circuit details. Similar tubes have been operated as high as 10 Gc in appropriate cavities. The fast warm-up heater-cathodes described in this paper can be incorporated into tubes designed for frequencies ranging from audio to microwave. For example, the two stage amplifier in Figure 21 was designed for use as a low level audio amplifier featuring low microphonism and high input impedance.

Tubes shown in Figures 20 and 21 and discussed herein are types 7077 and 7625. These types resulted from earlier General Electric development programs.

Other Applications

In addition to the various applications in which fast warm-up planar diodes and triodes are needed, this planar heater-cathode lends itself ideally to use in other electron devices such as cathode ray tubes, traveling-wave tubes and parametric amplifiers.

Conclusions

A fast warm-up heater-cathode utilizing conduction transfer of energy from heater to cathode has been developed that will attain essentially steady-state operation within three seconds. The addition of a series non-linear resistor further reduces warm-up time to less than 1.5 seconds. Low operating temperatures permit excellent electrical insulation between heater and cathode and offer promise of high reliability in continuous or cycling type operations. The exceptional rigidity of the bonded structure provides a high degree of resistance to externally applied mechanical vibrations.

The planar nature of the cathode makes it readily adaptable to existing planar ceramic receiving tubes and other thermionic devices requiring this type cathode.

Acknowledgments

The most recent phases of this program were sponsored by Sandia Corporation. The permission granted by Sandia to publish this information at this time is acknowledged and appreciated.

The authors gratefully acknowledge the fundamental contribution of Mr. A. J. Kling of the General Electric Research Laboratory in developing many of the necessary techniques involved in this project, and the importance of contributions by Messrs. G. W. Hume, D. W. Leiby and F. W. Staub of the General Electric General Engineering Laboratory for their work in analyzing theoretical aspects of thermal and mechanical problems. Valuable assistance in assembly and evaluation of tubes was rendered by Messrs. J. W. Hall and W. F. Joest of the General Electric Receiving Tube Department.

Bibliography

1. V. L. Stout and M. D. Gibbons, JAP 26 1438 (1955).

2. J. E. Beggs, "Sealing Metal and Ceramic Parts by Forming Reactive Alloys", Transactions-Professional Group Electron Devices, IRE, October, 1956.

3. J. W. Rush, "Application Techniques for Socketless Vacuum Tubes", Electronic Design, October 25, 1961.

MODERN GENERAL ELECTRIC

CERAMIC TUBES

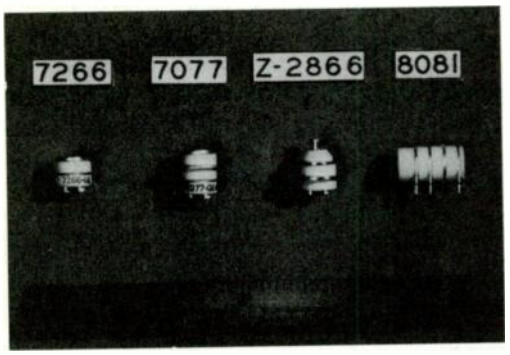
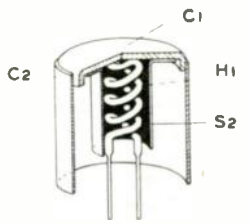


Fig. 1. Modern general electric ceramic tubes.

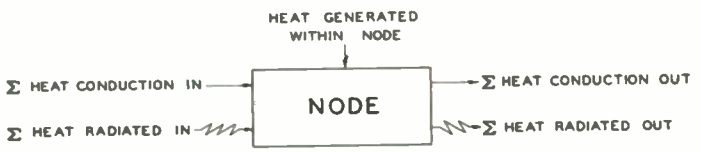
PLANAR CATHODE
HEATED BY RADIATION HEATER



LETTERS INDICATE
TYPICAL NODES

Fig. 2. Cut-away drawing of radiation heater-cathodes.

PORTION OF HEATER-CATHODE SYSTEM
RECEIVING AND LOSING ENERGY



QUANTITY	ANALOG
HEAT FLOW	= CURRENT
TEMPERATURE	= VOLTAGE
THERMAL RESISTANCE	= ELECTRICAL RESISTANCE
THERMAL CAPACITY	= ELECTRICAL CAPACITANCE
TIME	= TIME

Fig. 3. Nodal drawing and analog quantities.

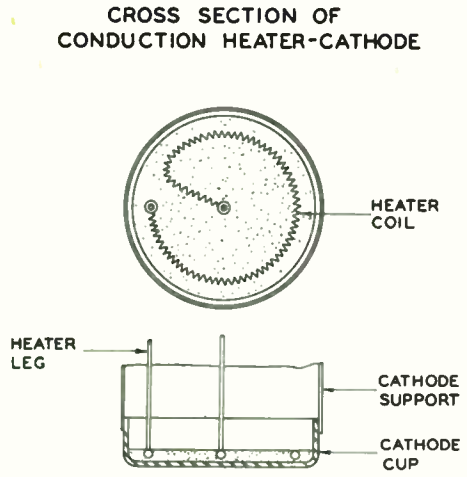


Fig. 4. Cross-sectional drawing of conduction heater-cathode.

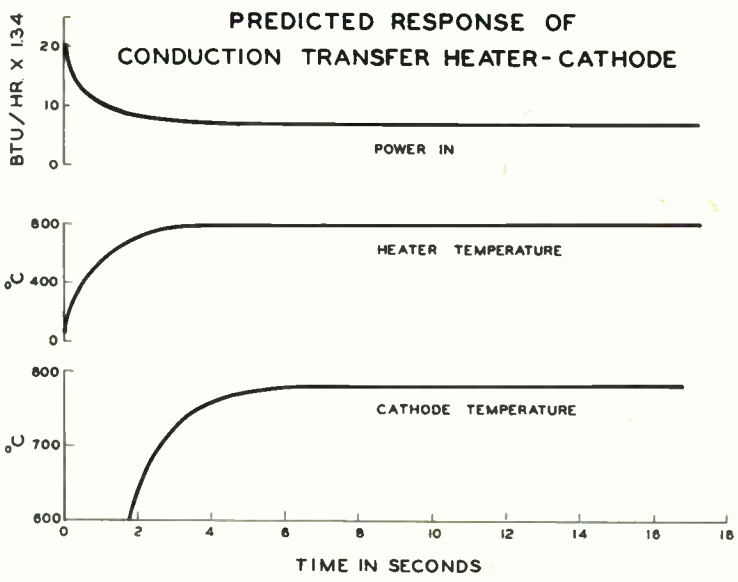


Fig. 6. Predicted response for conduction heater-cathode.

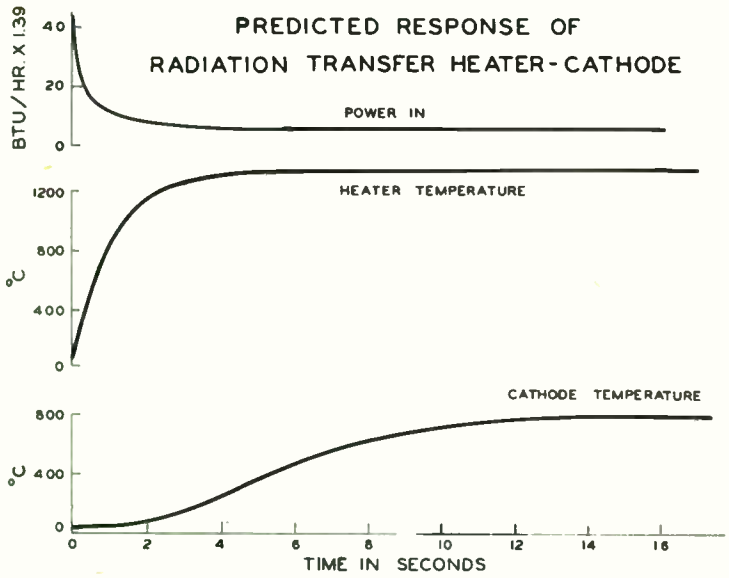


Fig. 5. Predicted response for radiation heater-cathode.

CONDUCTION TRANSFER HEATER - CATHODES

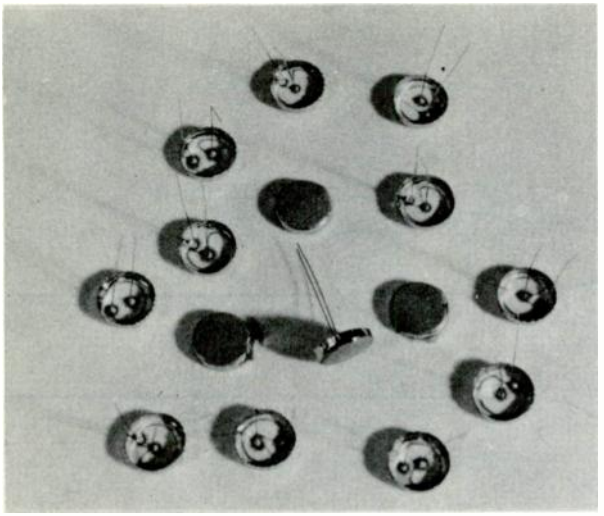


Fig. 7. Photograph of conduction transfer heater-cathode.

COATED CATHODE SUB-ASSEMBLY

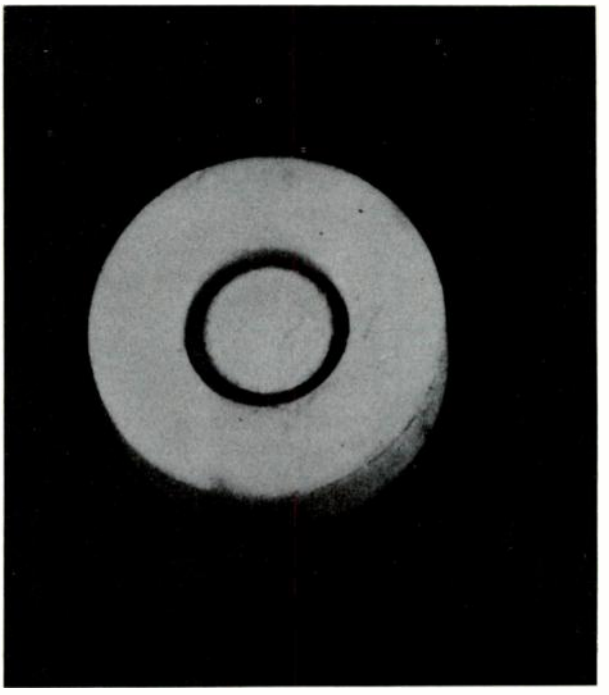


Fig. 8. Coated cathode sub/assembly.

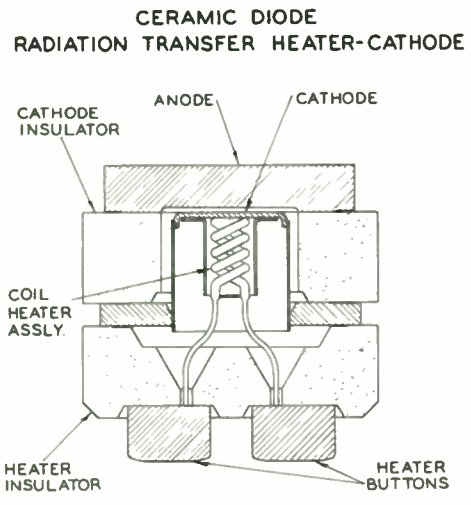


Fig. 9. Cross-sectional drawing of diode.

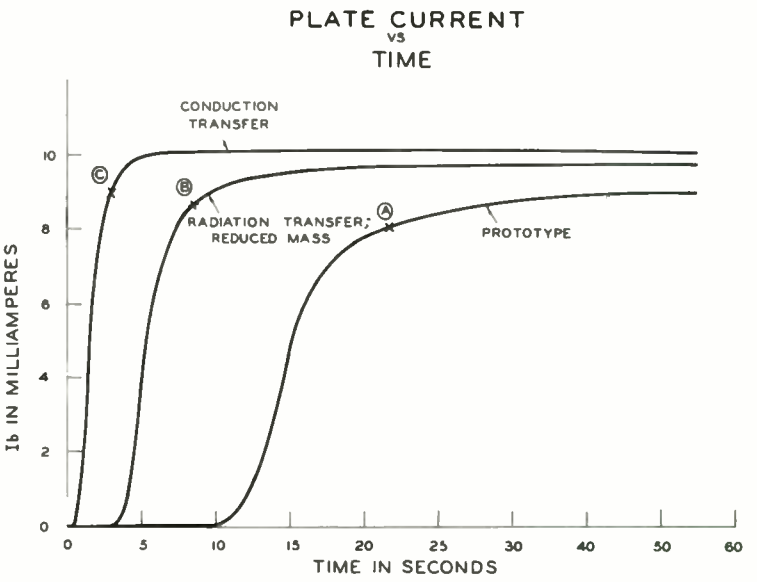


Fig. 10. Plate current versus time; prototype; radiation model; conduction heater-cathode.

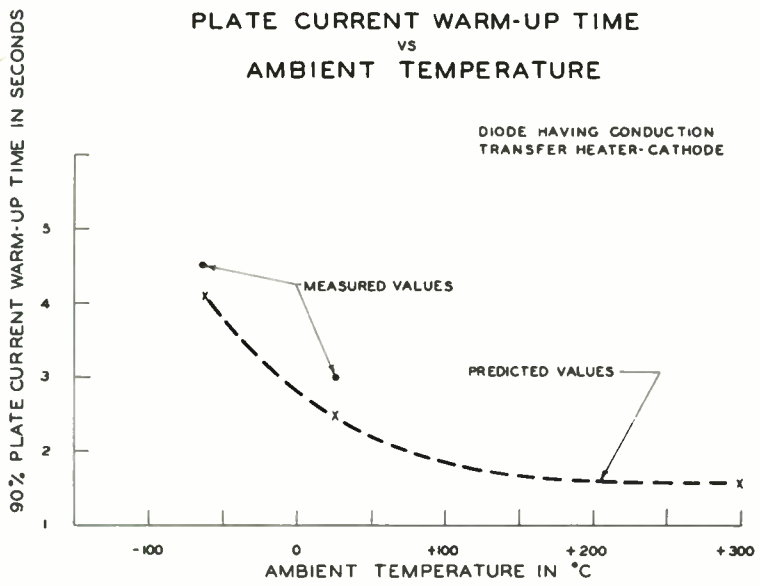


Fig. 11. Warm-up time versus ambient; conduction heater-cathode.

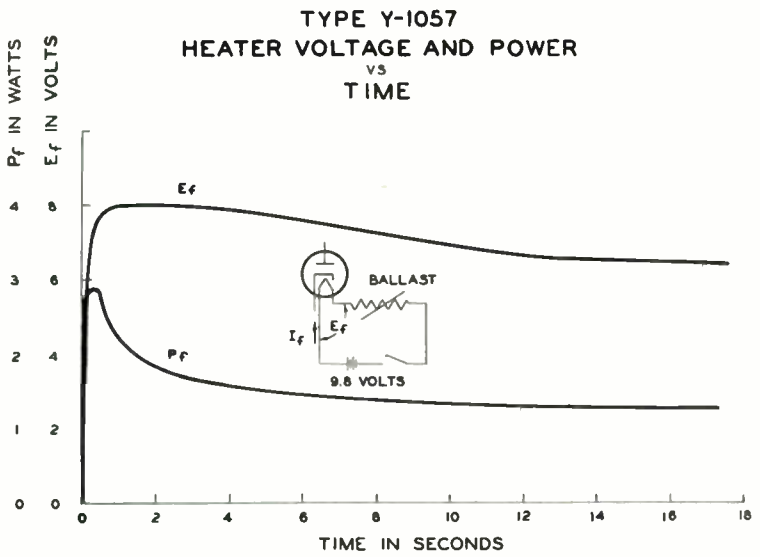


Fig. 12. Heater voltage and heater power vs. time (conduction heater-cathode with ballast).

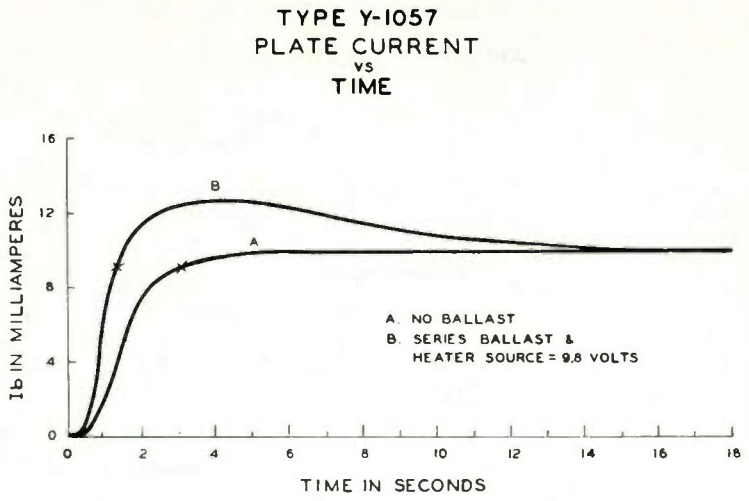


Fig. 13. Plate current versus time (conduction heater-cathode with and without ballast).

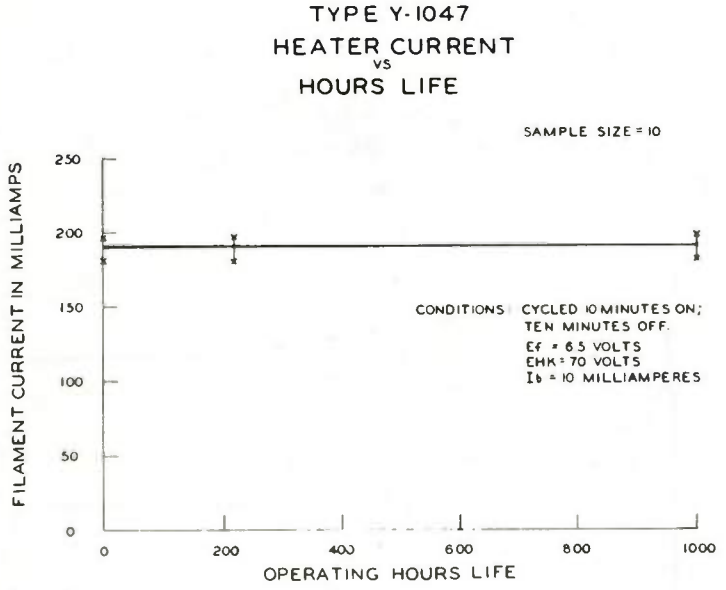


Fig. 14. Heater current during life test.

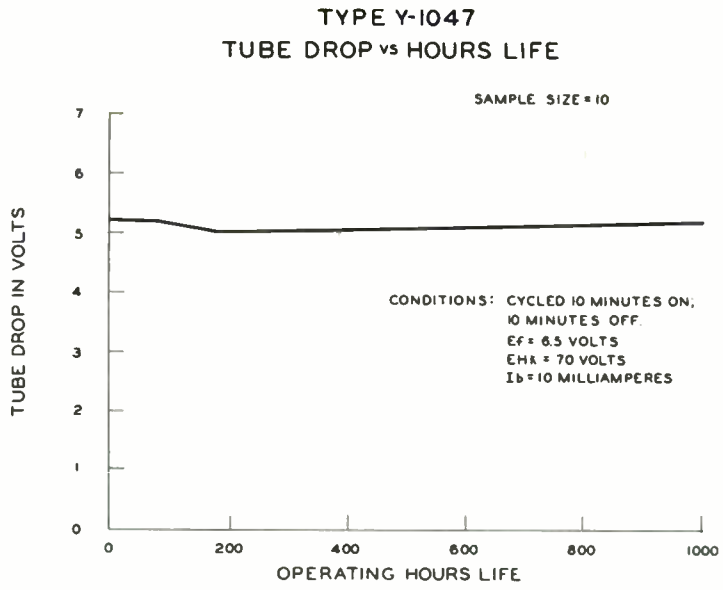


Fig. 15. Tube drop during life test.

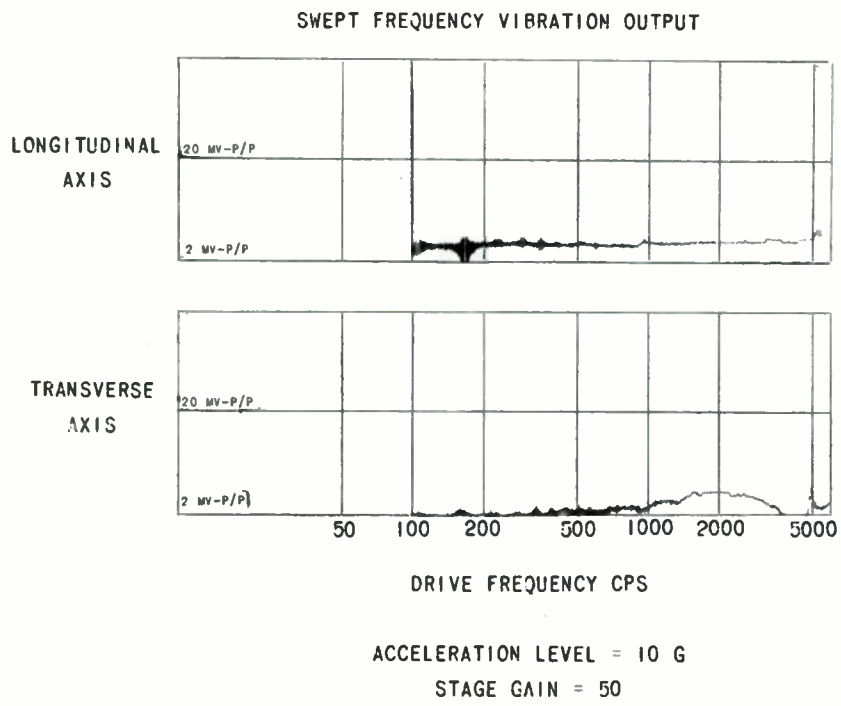


Fig. 16. Swept frequency vibration output.

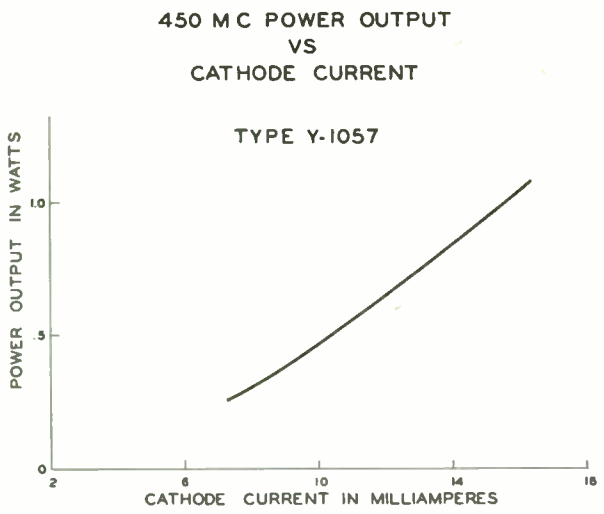


Fig. 17. 450 MC power output versus cathode current.

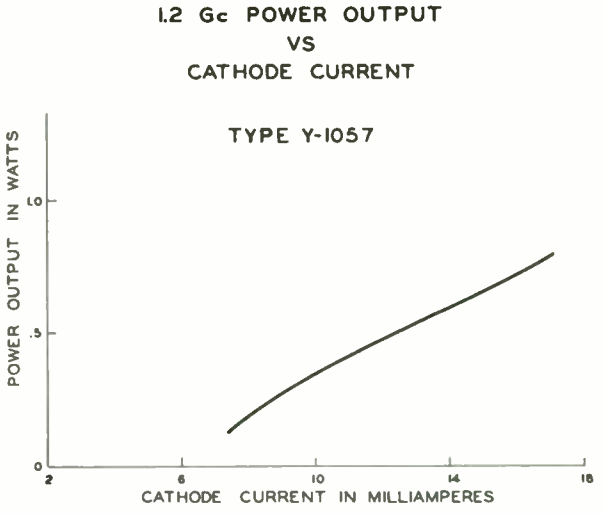


Fig. 18. 1.2 Gc power output versus cathode current.

1.2 Gc PLANAR OSCILLATOR

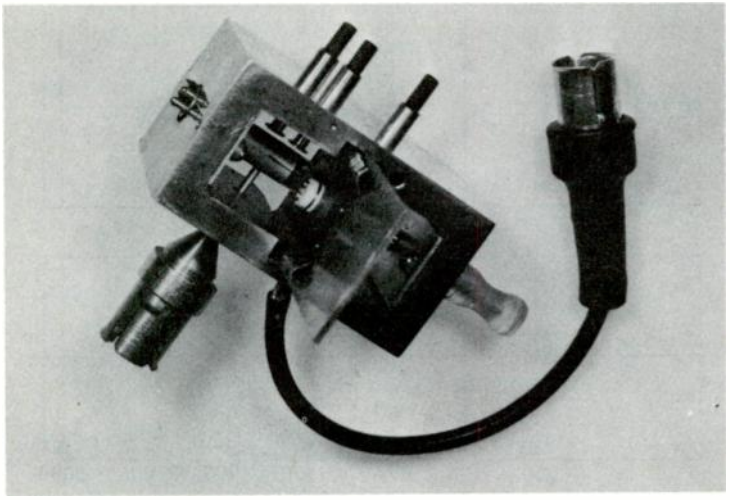


Fig. 19. 1.2 Gc planar oscillator.

2.7 Gc AMPLIFIER

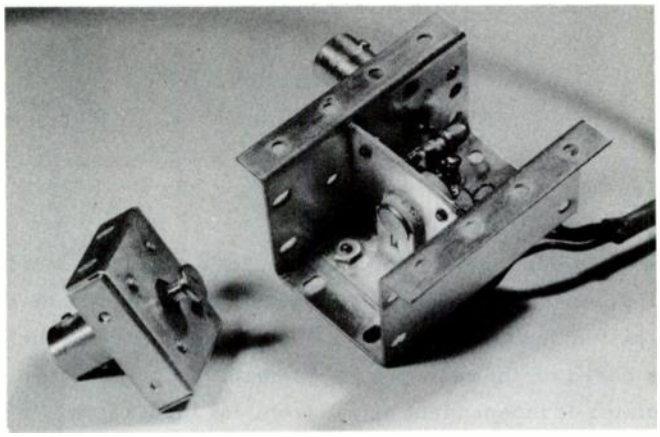


Fig. 20. 2.7 Gc amplifier.

TWO STAGE AUDIO AMPLIFIER

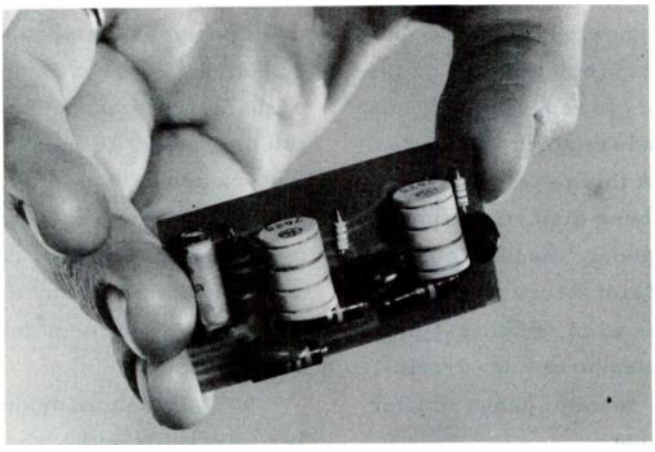


Fig. 21. 2-Stage low noise audio amplifier.

A HIGH ACCURACY TECHNIQUE FOR THE MEASUREMENT OF MICROWAVE FREQUENCY AND PHASE MODULATION

Samuel A. Rosen,
Rantec Corporation,
Calabasas, California

Summary

This paper describes a technique for the measurement of the frequency modulation or phase modulation produced by an active microwave device. The system, which is based on the use of a ferrite single sideband modulator is independent of both source frequency and source amplitude variations. Two such systems have been constructed, one for the measurement of the frequency modulation produced by a conically-scanning antenna, and the other for the measurement of peak dynamic phase shift vs. dc bias current of a ferrite phase shifter. Frequency deviations of less than one cycle-per-second in 10,000 Mc have been accurately measured.

Introduction

With the increased use of microwave systems which require the precise knowledge of frequency and phase, greater emphasis is being placed on the ability of system components to reproduce signals with a minimum of phase- and frequency distortion. The equipment described in this paper was developed in order to measure these distortions; particularly the amount of frequency modulation produced by a ferrite-scanned slot-array antenna. In an extension of the principles used, another equipment is discussed which measures the dynamic phase shift produced by a ferrite phase shifter when driven by an a-c current.

Measurements of this type formerly required the use of a highly stable microwave source. AFC systems with cavity references, or phase-locked local-oscillator techniques were required. The equipment was costly and the accuracy of the information obtained was several orders of magnitude less than that which can be obtained by the system described in these pages.

The ability of the single sideband modulator to supply an output which is displaced in frequency from its input allows measurements of this type to be free of the previous requirements for signal source frequency stability.

Ferrite Single Sideband Modulator

Before entering into the description of the equipment, a brief discussion of the ferrite single sideband modulator is in order; since it is upon its operation that this equipment is based.

As the name implies, the single sideband modulator is a device that produces a single frequency output, displaced from the input carrier by an amount equal to the modulating frequency. Figure 1 portrays the modulator operation. The input is split into two components of equal amplitude but of quadrature phase. Each component is then supplied to a ferrite balanced modulator. The modulators are driven in quadrature phase at the desired offset frequency and their outputs are combined in a magic-Tee. The upper sideband may then be obtained from the sum arm of the magic-Tee, and the lower sideband from the difference arm.

The balanced modulator consists of a circular waveguide with a shorting plate at one end and an orthogonal mode junction at the other end. A ferrite rod is placed along the waveguide axis with a solenoid wound around the periphery of the guide. Energy entering one of the orthogonal ports will be reflected by the short and leaves the same port. If a direct current is supplied to the solenoid, the

polarization of the energy will be rotated and an output will be obtained from the other port that is porportional to the sine of the rotation angle. When an alternating current is supplied, maximum output is obtained at both the postive and negative current-peaks and the output phases are displaced by 180^{o}. The resultant spectrum is that of a balanced modulator and contains two-sidebands and no carrier. With this in mind, a brief analysis of the single sideband modulator may be made.[1] [2]

$$\text{Input} = E_{in}\, e^{j\omega_o t}$$

$$\text{Input A} = 1/2\, E_{in} e^{j\omega_o t}\, e^{j\phi}$$

$$\text{Input B} = 1/2\, E_{in} e^{j\omega_o t}$$

$$\text{Output A} = 1/2\, E_{in} e^{j\omega_o t}\, e^{j\phi} \sin \theta_A t$$

$$\text{Output B} = 1/2\, E_{in} e^{j\omega_o t} \sin \theta_B t$$

where

ω_o = carrier frequency

θ = rotation angle

Then

E_{out} = (the output from either port)

$$E_{out} = 1/2\, E_{in} e^{j\omega_o t} (\sin \theta_B t \pm e^{j\phi} \sin \theta_A t)$$

Let

$$\phi = \pi/2$$

and

$$\omega_A = \omega_B + \pi/2$$

then

$$E_{out} = 1/2\, E_{in}\, e^{j\omega_o t} \left[\sin \theta t \pm (\cos \pi/2 + j \sin \pi/2) \cos \theta t \right]$$

since

$$\cos \pi/2 = 0$$

and

$$\sin \pi/2 = 1$$

$$E_{out} = 1/2\, E_{in}\, e^{j\omega_o t} (\sin \theta t \pm j \cos \theta t)$$

$$= 1/2\, E_{in}\, e^{j\omega_o t}\, e^{\pm j\theta t}$$

$$= 1/2\, E_{in}\, e^{jt(\omega_o \pm \theta)}$$

Since θ (the rotation angle) varies sinusoidally with modulation, the output frequency is translated by an amount equal to the modulating frequency. The sum arm will contain the upper sideband and the difference arm the lower sideband. Figure 2 is a photograph of a single sideband modulator similar to the one used in the equipment described.

System Theory

The basic system operation is shown in the block diagram of Figure 3. The output of the microwave signal source, f_c is split into two components. One component is connected to a ferrite single sideband modulator and the other to the local oscillator arm of a crystal mixer. The single sideband modulator is driven by a two-phase driver at the modulating frequency, f_m and the resultant output $(f_c + f_m)$ is supplied as the input signal to the device being tested. The output from the device is then connected to the signal arm of the crystal mixer.

Since the local oscillator frequency is f_c, and the signal frequency is $(f_c + f_m)$, the resultant I-F output from the mixer will be at f_m. This I-F frequency is completely independent of the carrier frequency and its stability is determined solely by the modulating frequency.

The I-F frequency is then amplified, amplitude-limited, and supplied to a frequency discriminator. If a frequency shift is introduced by the device being tested, there will be a change in the signal frequency supplied to the mixer. The new frequency will be $(f_c + f_m \pm \Delta f)$, which when combined with the local oscillator signal, f_c, will result in a new I-F frequency of $(f_m \pm \Delta f)$. An output voltage proportional to Δf will be produced by the frequency discriminator. In the case of frequency modulation, an a-c signal is obtained. The frequency components can be measured, and peak deviation obtained with a peak-reading voltmeter.

System Requirements for Frequency Modulation Measurements

The system described is used to measure the peak dynamic frequency deviation vs. azimuth rotation angle as produced by a conically scanning slot antenna. The deviation range is 10 cps, and the measurement error requirement no greater than 10% of the deviation range. The X-band antenna carrier frequency results in a required effective stability of 1 part in 10^{10}. In the previous discussion of system theory, it was stated that the accuracy is dependent on the modulator frequency stability. However, in practice, one must add to this several other factors.

The output from a linear discriminator can be written as follows:

$$E_{out} = (k)\ (\pm \Delta f)$$

where

$$k_f = k_1 \frac{E_{in}}{1/2 \cdot \beta}$$

E_{in} = the input to the discriminator

β = the bandwidth

and

k_1 = the gain factor of the discriminator

$$E_{out} = k_1 \frac{E_{in}}{1/2 \cdot \beta}\ (\pm \Delta f)$$

Therefore, the maximum percent amplitude modulation for a given accuracy would be:

$$\frac{\text{Max error in cps}}{1/2 \text{ Bandwidth}} \times 100$$

Amplitude modulation therefore can be a severe problem when dynamic measurements are made. In this equipment, a bandwidth of 1200 cps is used, so that for an error of 0.6 cps the maximum percentage of modulation which can be tolerated is 0.1%. There is 10% amplitude modulation produced by the antenna at the extremes of the measurment, which must be reduced by amplitude limiting and clipping. Another factor which can affect dynamic measurement-accuracy is frequency-dependent phase shift in the I-F amplifier. In addition, when fixed offset frequencies are to be measured, the thermal-stability of the discriminator becomes decidedly important.

Circuit Design

Figure 4 is a block diagram of the final system configuration. The deviation measurement equipment is part of a complete test facility for the antenna, so that other considerations determined the choice of 455 kc as the I-F frequency. The single sideband modulator is therefore driven by a two-phase 455 kc crystal controlled driver. Frequency stability has been obtained by placing the crystal in a thermostatically controlled oven whose temperature is held to within 0.2°C.

For a useable output, at one cycle deviation, a steep discriminator slope is required. To obtain a steep slope and permit a sufficiently wide passband, a discriminator center frequency of 15 kc was chosen. The conversion is performed by another mixer and oscillator. The oscillator is crystal controlled and operates at a frequency of 440 kc. The crystal is contained in an oven similar to the one employed in the single sideband modulator driver, since its stability also affects the accuracy of the system. Utilization of this second conversion yields a simple, but effective technique for system calibration. By purposely changing the frequency of this oscillator, an output from the discriminator is obtained. In order to obtain a frequency shift of 10 cps, a front panel switch is used which changes the capacity across the crystal.

The input signal to the 15-kc amplifier is amplitude limited by standard techniques. Several stages of limiting are required to reduce the percentage of amplitude modulation to 0.1%. The limiter output connects to the frequency discriminator which is a Foster-Seeley type. Ratio detectors and solid state frequency discriminators are presently being investigated to determine whether a less amplitude-sensitive discriminator can be obtained.

A peak reading voltmeter is connected to the discriminator through a band-pass filter to meas-

ure peak deviation at the antenna scan frequency. A voltage proportional to the meter reading is supplied to the Y-axis of an X-Y recorder. X-axis information is obtained from a data-take off potentiometer on the antenna mount. A typical curve of peak deviation vs. azimuth look-angle is shown in Figure 5.

The frequency modulation measuring equipment is completely self-contained in a rack type of enclosure, and is used in conjunction with a microwave anechoic chamber. Figure 6 is a photograph of the complete antenna test equipment with the portions which are utilized for deviation measurements highlighted. Waveguide runs are utilized between the receiving horn and mixer and between the single sideband modulator and antenna mount, thereby assuring maximum input signal for good limiting in the I-F strip.

Phase Modulation Measurement System

When a carrier is phase modulated, its instantaneous frequency will change. The maximum frequency excursion which occurs is:

$$\Delta f = -f_m \Delta\theta$$

where

f_m = the modulating frequency

$\Delta\theta$ = the peak phase shift in radians.

If the modulating frequency is held constant, the frequency excursion is directly proportional to the peak phase shift. Since this direct relationship exists between peak phase shift, and the frequency deviation it produces on a carrier, a system similar to the one previously described can be used to measure dynamic phase shift.

Figure 7 is a block diagram of a system which has been developed to measure the peak phase shift of a ferrite phase shifter. The microwave section is basically the same as that used in the frequency deviation system, The I-F frequency is 450 kc with a bandwidth of over 2 kc. A broader bandwidth is used since the high index of modulation produced by the phase shifter results in a broad spectrum. The signal is amplitude-limited in the latter stages of the I-F strip. The discriminator output is a sine wave whose frequency is the same as that of the modulation, and whose amplitude is proportional to the peak phase shift. The sine wave is amplified, filtered to remove noise, and peak-rectified to produce a d-c level proportional to the discriminator output.

The phase shifters, whose characteristics are measured by the equipment shown in Figure 8, are used in the slot array previously mentioned in this paper. For symmetrical scanning, the phase shifters must be matched in groupings of four. A constant-amplitude a-c voltage is impressed on the drive coil of the phase shifter. Concurrently, a bias winding on the phase shifter is supplied with a d-c current which varies linearly from zero to a pre-set maximum and back to zero again. As the current changes, an X-Y recorder plots a curve of dynamic phase shift as a function of dc bias current. Figure 9 is a representative curve of one of the phase shifters tested.

System Errors

System errors can be produced by insufficient amplitude limiting, non-linear phase shift over the pass band of the I-F, and frequency drift of the single sideband modulator driver and discriminator. With careful design and control, these errors can be minimized.

The output of a practical single sideband modulator consists of the desired sideband plus a small amount of the undesired sideband, the carrier and sideband harmonics. These spurious outputs, particularly the unused sideband, can affect the accuracy of the system. For a fixed microwave frequency, however, the carrier and undesired sideband can be reduced by careful adjustment of the bias magnets on the modulator (ref. Figure 2) and the phase and amplitude balance of the driver, so that they are at least 30 db below the desired sideband.

The harmonics can be reduced by tuning the modulator coils to the drive frequency and utilizing low level drive[1]. Their effects however are second

order, and since they normally fall outside of the I-F pass band these may be neglected for most measurements.

Conclusions

Two sets of test equipment were developed and constructed based on the techniques described. Measurement accuracies were obtained which were several orders of magnitude greater than heretofore possible. The primary reason for the increased accuracy was the phase coherent frequency translational properties of the microwave single sideband modulator. The use of the single sideband modulator in microwave instrumentation opens new fields of measurement that were previously untouched.

Acknowledgements

The author wishes to thank Mr. Alvin Clavin for his development efforts on the Microwave Ferrite Single Sideband Modulator, since without it, this equipment could not have been developed. In addition, the author is indebted to Messrs. Del Rairigh and Leon Traister for their assistance in the design of the electronic chassis. Acknowledgement is also due Mr. Louis A. Kurtz for the development of the antenna for which this test equipment was constructed.

The work described in this paper was performed at Rantec Corporation with the support of General Dynamics/ Pomona, California.

References

(1) Theory and Applications of Microwave Single Sideband Modulator: Alvin Clavin, Rantec Report No. 110, 3 April 1961

(2) J. Cacheris "Microwave Single Sideband Modulator Using Ferrites," Proc. I.R.E. Vol 42, pp 1242-1247; August, 1954

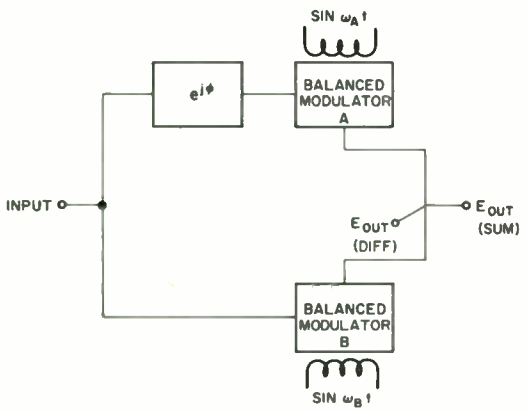


FIGURE 1. SINGLE-SIDEBAND MODULATOR BLOCK DIAGRAM

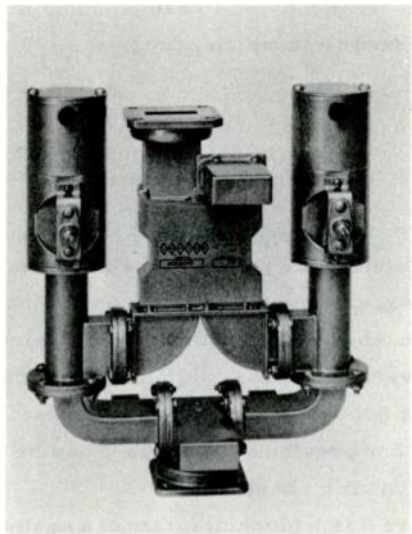


FIGURE 2. X-BAND SINGLE-SIDEBAND MODULATOR

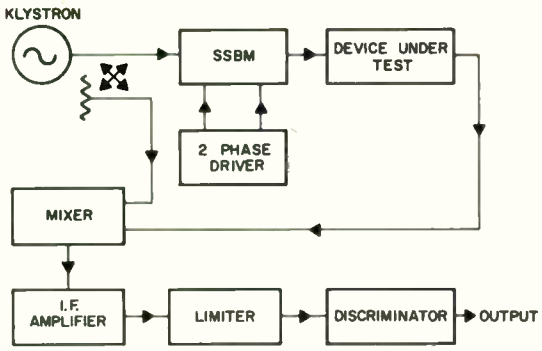


FIGURE 3. FREQUENCY DEVIATION MEASURING SYSTEM BASIC BLOCK DIAGRAM

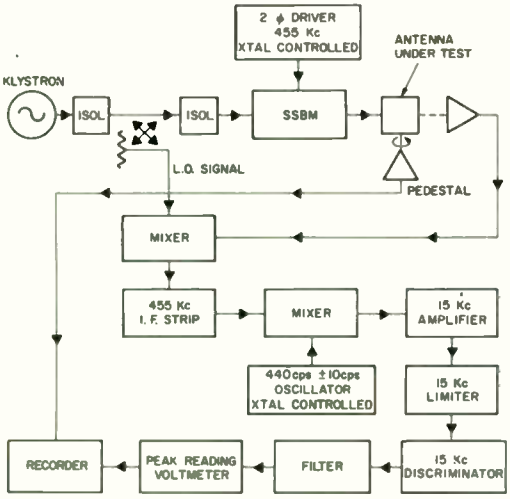


FIGURE 4. FREQUENCY DEVIATION MEASURING SYSTEM COMPLETE BLOCK DIAGRAM

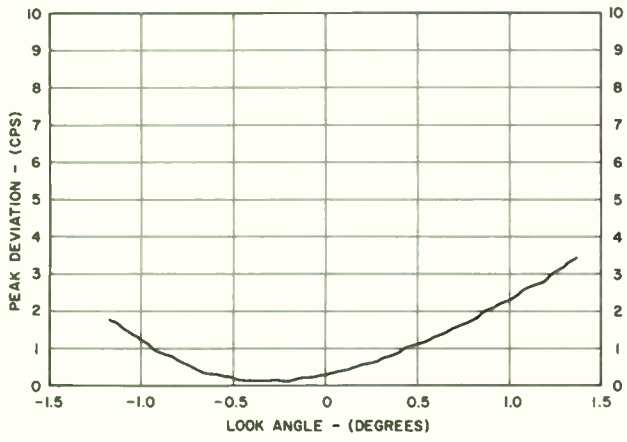


FIGURE 5. TYPICAL CURVE OF PEAK FREQUENCY DEVIATION Vs. AZIMUTH LOOK ANGLE FOR SLOT ARRAY

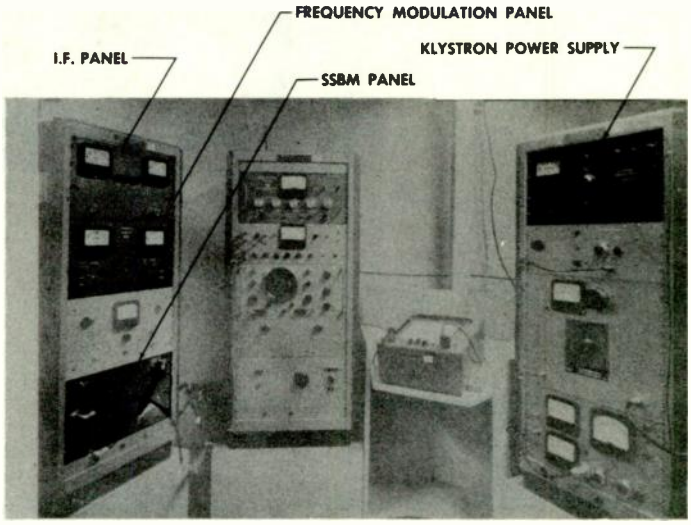


FIGURE 6. FREQUENCY DEVIATION MEASUREMENT EQUIPMENT

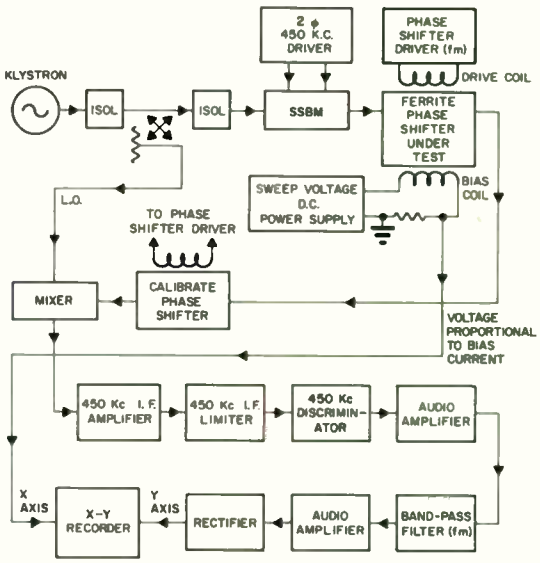


FIGURE 7. DYNAMIC PHASE-SHIFT MEASUREMENT EQUIPMENT BLOCK DIAGRAM

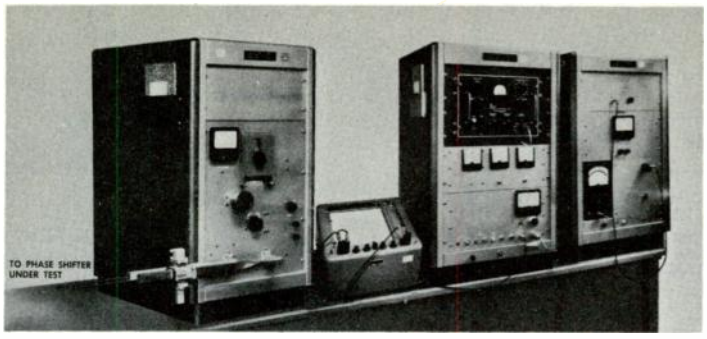


FIGURE 8. DYNAMIC PHASE-SHIFT MEASUREMENT EQUIPMENT

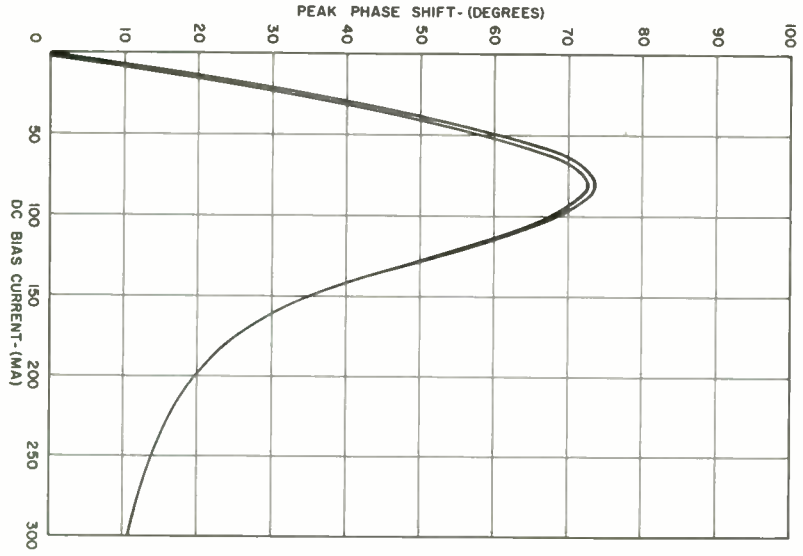


FIGURE 9. TYPICAL CURVE OF DYNAMIC PHASE SHIFT Vs.
DC BIAS CURRENT FOR A FERRITE PHASE SHIFTER

ANALYSIS OF ROTATION ERRORS OF A WAVEGUIDE ROTARY VANE ATTENUATOR

Wilbur Larson
National Bureau of Standards
Radio Standards Laboratory
Boulder, Colorado

Summary

In the Electronic Calibration Center of the NBS Radio Standards Laboratory, Boulder, Colorado, an analysis of calibration data obtained from measurements made on a rotary vane type attenuator reveals that the vane alignment produces an inherent vane angle error. It is found that the standard deviation of the measured values of attenuation is correlated with the vane angle error. Approximately two-thirds of the attenuators calibrated were in error by less than 50 percent of the manufacturer's stated specifications of accuracy.

The treatment of the errors of a rotary vane attenuator is extended to cover the vane angle error, resettability error, and mismatch error. The method of analysis is applicable to all attenuation values and vane angle errors of this sample. Graphs are used to estimate the limits of these errors. Several representative cases are given using the deviation from the "estimated vane angle error" to establish limits in the correction of vane angle error.

A modified rotary vane attenuator with a resolution equal to 30 times the resolution of the attenuators comprising the sample was used to verify calculated vane angle error at various values of attenuation. Inherent vane angle error can be reduced by increasing the dial resolution and by a technique of precise vane alignment at zero dial setting.

Introduction

The rotary vane attenuator which is commonly used as an interlaboratory attenuation standard by industrial and governmental laboratories, is characterized by a simple relationship between the angle of rotation of the vane from a zero reference position and the resulting attenuation.[1] The expression stating this relationship may be written as

$$A = 20 \log_{10} \cos^2 \theta \qquad (1)$$

where θ is the angle of the rotating section relative to the stationary end sections and A is the resulting attenuation in decibels. The attenuation of devices of this type is nearly independent of frequency throughout the range of operation. The rotating center section in most commercially available rotary vane attenuators is coupled to a direct-reading dial marked in decibels with gears having a ratio in the range of 12 to 24.

Data has been recorded on attenuators calibrated in the Electronic Calibration Center of the NBS Radio Standards Laboratory, Boulder, Colorado. The analysis of calibration data presented in this paper suggests methods to improve the rotary vane attenuator for use as an interlaboratory standard. The results from more than one hundred calibrations performed on several types of waveguide rotary vane attenuators are given. These attenuators were calibrated in the frequency range of 4.0 to 18.0 Gc.

Error Analysis

When an interlaboratory standard is calibrated in the Center the value assigned to the standard is reported to be correct within certain limits. Errors that determine these limits are both systematic and random. The systematic error is always of the same sign, while the random error present in the measurement can be positive or negative with equal probability.

The rotary vane attenuator has a systematic error which is inherent and is caused by the incorrect vane alignment of the attenuator. This inherent error can be positive in some cases and negative in others. The random error considered here is an error caused by the operator. Both of these errors must be considered in setting the rotary vane attenuator to a given value of attenuation. A positive or negative inherent error is determined by the direction of rotation of the center vane from perfect alignment with the stationary sections. The operator error can be either positive or negative because the direction of rotation is random in making a dial setting.

The attenuation difference of a variable attenuator can be measured by moving the dial from zero or reference position to another position that is marked on the scale. In making an attenuation difference measurement of a rotary vane attenuator one may write the initial

setting as

$$A_i' = A_i \pm \epsilon_{io} + \epsilon_{iI} \qquad (2)$$

where A_i is the calculated attenuation at the initial setting, ϵ_{io} is the error made by the operator at the initial setting, and ϵ_{iI} is the inherent error at the initial setting.

The errors at the initial setting may be written as

$$\epsilon_{io} + \epsilon_{iI} = 40 \log_{10} \cos(\theta_i \pm \theta_{io} + \theta_{iI}) - 40 \log_{10} \cos \theta_i$$

$$= 40 \log_{10} \frac{\cos(\theta_i \pm \theta_{io} + \theta_{iI})}{\cos \theta_i} \qquad (3)$$

where θ_i is the vane angle at the initial setting to obtain the calculated attenuation, θ_{io} is the vane angle error made by the operator at the initial setting, and θ_{iI} is the inherent vane angle error.

In a similar manner, the final setting may be written as

$$A_f' = A_f \pm \epsilon_{fo} + \epsilon_{fI} \qquad (4)$$

where A_f is the calculated attenuation at the final setting, ϵ_{fo} is the error made by the operator at the final setting, and ϵ_{fI} is the inherent error at the final setting.

The errors at the final setting may be written as

$$\epsilon_{fo} + \epsilon_{fI} = 40 \log_{10} \frac{\cos(\theta_f \pm \theta_{fo} + \theta_{fI})}{\cos \theta_f} \qquad (5)$$

where θ_f is the vane angle at the final setting to obtain the calculated attenuation, θ_{fo} is the vane angle error made by the operator at the final setting, and θ_{fI} is the inherent vane angle error.

The error due to mismatch must be considered in making an attenuation measurement because the attenuator is inserted in an imperfectly matched system. This mismatch error may be written as[2,3]

$$\epsilon_m = 20 \log_{10} \frac{(1 - {}^{(f)}\Gamma_i \Gamma_G)(1 - {}^{(f)}S_{22}\Gamma_L)}{(1 - {}^{(i)}\Gamma_i \Gamma_G)(1 - {}^{(i)}S_{22}\Gamma_L)} \qquad (6)$$

where the front superscripts, $^{(i)}$, and $^{(f)}$, refer to the initial and final values respectively, and where Γ_i is the input voltage reflection coefficient of the attenuator when terminated with a load (detector system) having a reflection coefficient Γ_L. The quantity Γ_G is the reflection coefficient of the generator and S_{22} is an element of the scattering matrix of the attenuator.

The attenuation difference measurements of this sample were made with the initial dial setting at zero decibels. Therefore, the rotation errors at the initial setting are very small as shown in Equation (3). However, the error could be significant with the initial dial setting at zero decibels for a very large inherent vane angle error. If the initial setting is made at a high value of attenuation, the errors at the initial setting will become significant since θ_i will be large.

Graphical Presentation of Data

To compare measured values of inherent vane angle error with calculated values, a rotary vane attenuator was modified to permit precise determination of the angle of rotation of the vane by increasing the gear (coupling the rotary section to the dial mechanism) ratio to 360. With this modified attenuator, the inherent vane angle error was adjusted to a desired value and attenuation difference measurements were made at several values of attenuation.

The solid-line curves of Figure 1 represent the calculated deviation in attenuation for vane angle errors of +5' and -5'. These values are obtained from Equation (5) with $\theta_{fo} = 0$. Table 1 lists the calculated deviation in attenuation, the measured deviation in attenuation, and the difference between the calculated and measured results for vane angle errors of +5' and -5' at the dial settings of 10, 20, 30, and 40 db. The measured values of deviation in attenuation from Table 1 are plotted in Figure 1.

The dashed-line curves of Figure 1 represent the calculated deviation in attenuation for operator errors of ± 30", estimated to be the maximum limits of resettability by the operator for the modified attenuator.

The broken-line curves of Figure 1 represent the calculated deviation in attenuation for the combined operator and mismatch errors. The mismatch error was calculated from Equation (6) using values of $^{(i)}\Gamma_i$, $^{(f)}\Gamma_i$, Γ_L, Γ_G, $^{(i)}S_{22}$, and $^{(f)}S_{22}$ obtained from the usual relation between Γ and VSWR and to VSWR values shown in Figure 1. The mismatch error was calculated with the phase of the reflection coefficients chosen to give the maximum mismatch error.

In this analysis of the performance of rotary vane attenuators, data were selected from calibrations performed on 52 attenuators. This included 114 calibrations, each consisting of the measurement of the attenuation difference for several values of attenuation settings at a given frequency. This sample had a frequency coverage from 4.0 to 18.0 Gc in waveguide sizes WR 187, WR 137, WR 112, WR 90, and WR 62. For most of the calibrations performed in this sample, the measured values of attenuation difference did not correspond to the dial settings of the attenuators. This lack of correspondence caused by inherent vane angle error, will increase at the larger values of attenuation difference as indicated in Equation 5. If the vane angle error is large, lack of correspondence between measured values of attenuation difference and dial settings will become significant even at the small values of attenuation difference.

Since the values of attenuation difference of 10, 20, 30, and 40 db were common to all calibrations in the sample, these values were used in computing the arithmetic mean and standard deviations recorded in Table II. A vane angle error corresponding to a deviation in attenuation equal to the standard deviation of the sample at a 30 db setting was calculated to be approximately 5.5'. The curve shown in Figure 2 was plotted from calculations obtained using Equation (5) with $\theta_{fI} = 5.5'$ and $\theta_{fo} = 0$. The standard deviations for each value of attenuation difference of the sample also are plotted in Figure 2. It will be noted that an increase in attenuation due to a vane angle error of + 5.5' passes approximately through the standard deviation of each value of attenuation difference determined for this sample. The greatest departure from the curve occurs at 20 and 40 db, and is 0.002 db.

Shown in Figure 3 are graphs of the arithmetic mean and the standard deviation of this sample. Approximately 32% of the results from individual calibrations differ from the nominal setting by an amount greater than σ, 4.6% for greater than 2σ, and 0.27% for greater than 3σ. Or, in other words approximately 68% of the nominal dial readings are accurate to $\pm$ 0.7%, 95.4% are accurate to 1.4% and 99.7% are accurate to $\pm$ 2.1%. Results from Figure 3 indicated that two-thirds of the attenuators of the sample were in error by less than 50 percent of the value of the manufacturers' stated specification of accuracy.

Let us consider the measurements of one attenuator in this sample and estimate the vane angle error from these measurements. The difference between the attenuator dial readings and the measured value in decibels was used to determine the estimated vane angle error. The results for one attenuator in this sample with an average estimated vane angle of - 2.8' are shown in Table III.

The estimated vane angle error is determined from the dial error in column 3 with the use of Equation (5), and is listed in column 4 of the table. The average estimated vane angle error of column 4 is - 2.8'. The deviation from this average is listed in column 5. If a correction is made for the - 2.8' estimated vane angle error, the remaining error in decibels, listed in column 6 of the table, results. This shows that for a correction of +2.8 minutes in setting the dial, the greatest error would occur at 30 db. The error at this value is 0.019 db or 0.06% of the attenuation value measured.

In Figure 4, the graph for the rotary vane attenuator with a vane angle error of - 2.8' is plotted in a manner similar to Figure 1, with the operator and mismatch error included. The measured values of attenuation difference fall within the limits of operator error except at 10 db. However, when the limits of mismatch error are included, all values fall within limits of total error.

Data selected from one attenuator calibrated at three different frequencies is summarized in Table IV and presented graphically in Figure 5. This attenuator was calibrated at 9.0, 9.8, and 11.2 Gc, designated as A, B, and C, respectively, in the table and the graph. The average estimated vane angle error is +5.9', at A, +5.7' at B, and +5.5' at C. The maximum error in decibels for these vane angle corrections is 0.016 db for A, 0.015 db for B, and 0.011 db for C. The error at frequencies A and B is 0.06% of the attenuation value measured, and the error at frequency C is 0.03%.

Deviations from the estimated vane angle error of any given measurement are in the same direction for each frequency, except for one point, 35 db, at the 9,000 Mc measurement. Deviations from the estimated vane angle error curve could be caused by a warp in the vanes or eccentricity during rotation of the center section.

The estimated vane angle error was determined for each of the attenuators in the sample. Deviations from the estimated vane angle error were recorded at the 10, 20, 30, and 40 db dial settings. The deviation averaged for all the attenuators was approximately 1.9' at 10 db, 2.1' at 20 db, 1.2' at 30 db, and 1.4' at 40 db. These angles correspond to dial readings in error by 0.014 db at 10 db, 0.031 db at 20 db, 0.033 db at 30 db and 0.069 db at 40 db.

Thus, if a correction equal to the estimated vane angle error were made at the respective attenuation values, all values of the dial readings would be in error by less than 0.2 percent of the value measured.

The positioning of the vane at a dial setting of zero can be aligned with a precision attenuation measurement system so that the vane angle error is very small. The modified rotary vane attenuator referred to above was adjusted so that the differences between dial settings and measured values were within 0.1 percent throughout the range 0 - 40 db. The modified attenuator was aligned several times by this technique. One adjustment resulted in an average estimated vane angle error of approximately + 1'10" and another in approximately + 12". Figure 6 shows a curve of vane angle error of + 12" and the measured values of deviation in attenuation from the dial setting for the modified attenuator. These measured values of attenuation difference agree within 0.1 percent of the dial settings, at 35 and 40 db and within 0.05 percent at the lower values of attenuation. Details of this method of vane angle alignment will be presented in a forthcoming paper.

Conclusions

It has been shown the inherent error of the rotary vane attenuator is correlated with the standard deviation of the measured values of attenuation difference. The accuracy of the dial settings of the rotary vane attenuator can be improved with a correction of vane alignment equal to the "estimated vane angle error". With present rotary vane attenuators, the technique of correcting vane angle alignment can bring dial settings in agreement with precisely measured values of attenuation difference to within 0.2 percent.

Acknowledgements

The author extends his appreciation to R. E. Larson, and W. F. Snyder for suggestions which have helped in formulating the ideas presented in this paper. R. L. Ehret and G. J. Counas assisted in the measurements and calculations.

References

1. Southworth, G. C., "Principles and Applications of Waveguide Transmission", p. 374, D. VanNostrad, New York.
2. Beatty, R. W., "Mismatch Errors in the Measurement of Ultrahigh-Frequency and Microwave Variable Attenuators", J. Research National Bureau of Standards, Vol. 52, pp. 7-9, January, 1954.
3. Schafer, G. E. and Rumfelt, A. Y., "Mismatch Errors in Cascade-Connected Variable Attenuators", IRE Trans. on Microwave Theory and Techniques, Vol. MTT-7, pp. 447-453, October, 1959.
4. McNish, A. G., and Cameron, J. M., "Propagation of Error in a Chain of Standards", IRE Trans. on Instrumentation, Vol. I-9, pp. 101 - 104, September, 1960.

TABLE 1

Comparison of Calculated Deviation in Attenuation and Measured Deviation in Attenuation for a Modified Rotary Vane Attenuator with an Induced Vane Angle Error of +5 and -5 minutes

Dial Setting	VANE ANGLE ERRORS + 5'				VANE ANGLE ERROR 5'			
	Calculated Deviation in	Measured Deviation in		Difference in	Calculated Deviation in	Measured Deviation in		Difference in
	DECIBELS				DECIBELS			
10	0.037	a	0.040	+ 0.003	0.037	a'	0.036	- 0.001
20	0.076	b	0.074	- 0.002	0.076	b'	0.071	- 0.005
30	0.139	c	0.145	+ 0.006	0.139	c'	0.138	- 0.001
40	0.250	d	0.247	- 0.003	0.247	d'	0.248	+ 0.001

TABLE II

Arithmetic Mean and Standard Deviation in Decibels for a Sample of 114 Rotary Vane Attenuator Calibrations

Dial Setting	Arithmetic Mean	Standard Deviation		
		1σ	2σ	3σ
10	10.014	±0.041	±0.082	±0.123
20	19.985	±0.085	±0.170	±0.255
30	30.014	±0.154	±0.308	±0.462
40	40.026	±0.279	± .558	±0.837

TABLE III

Rotary Vane Attenuator Calibrated at 11,200 Mc

Dial Setting in DB	Measured DB	Dial Error in DB	Estimated Vane Angle Error	Deviation From Average Estimated Vane Angle Error	Error in DB with Estimated Vane Angle Correction
10	9.992	.008	- 1.1'	+ 1.7'	012
15	14.980	.020	- 2.0'	+ 0.8'	.009
20	19.946	.054	- 3.6'	- 0.8'	.012
25	24.931	.069	- 3.3'	- 0.5'	.011
30	29.902	.098	- 3.5'	- 0.7'	.019
35	34.904	.095	- 2.7'	+ 0.1'	.004
40	39.867	.133	- 2.7'	+ 0.1'	.005

AVERAGE - 2.8'

TABLE IV

Rotary Vane Attenuator Calibrated at Three Frequencies

A-9,000 Mc, B-9,800 Mc and C-11,200 Mc

Dial Setting in DB	Dial Error in DB			Estimated Vane Angle Error			Deviation from Average Estimated Vane Angle Error			Error in DB with Estimated Vane Angle Correction		
	A	B	C	A	B	C	A	B	C	A	B	C
10	.047	.046	.052	6.35'	6.20'	7.00'	+.46'	+.44'	+1.49'	.004	.003	.010
15	.056	.061	.059	5.10'	5.55'	5.35'	-.79'	-.21'	-.16'	.009	.003	.002
20	.085	.085	.079	5.65'	5.65'	5.25'	-.24'	-.11'	-.26'	.004	.002	.004
25	.140	.137	.121	6.65'	6.50'	5.75'	+.76'	+.74'	+.24'	.016	.015	.005
30	.154	.147	.128	5.50'	5.33'	4.60'	-.38'	-.43'	-.11'	.010	.013	.003
35	.218	.197	.179	6.21'	5.62'	5.12'	+.32'	-.14'	-.29'	.012	.005	.011
40	.290	.274	.275	5.81'	5.50'	5.50'	-.09'	-.26'	-.01'	.005	.013	.001
				5.9'	5.7'	5.5'						
				AVERAGE								

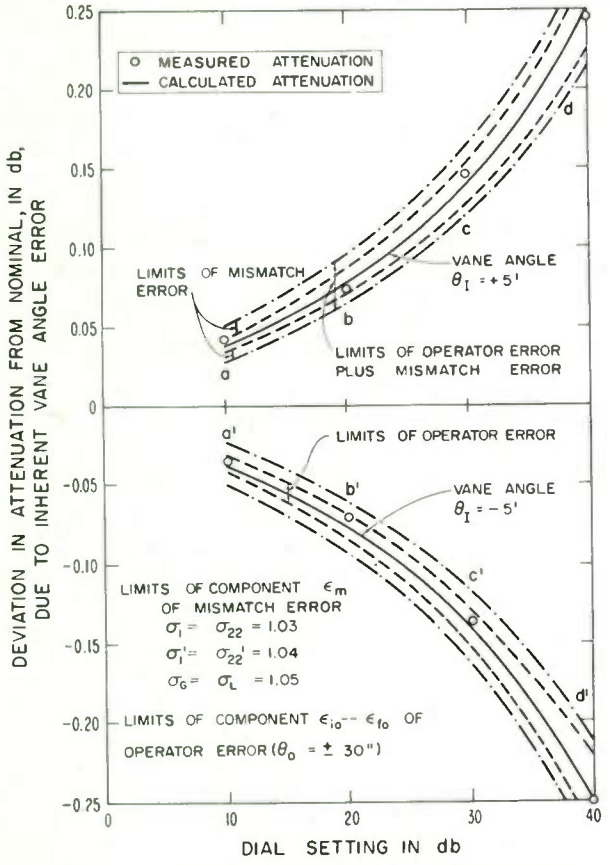


Fig. 1. Comparison of calculated deviation in attenuation and measured deviation in attenuation for a modified rotary vane attenuator with an induced vane angle error of + 5 and - 5 minutes.

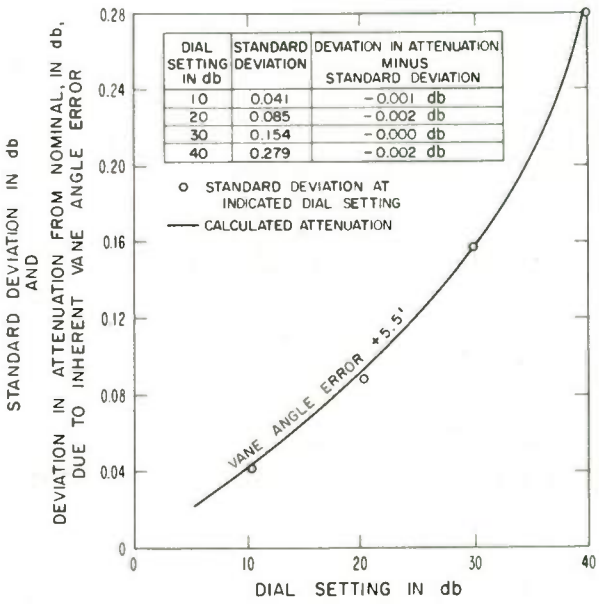


Fig. 2. Calculated deviation in attenuation for a rotary vane attenuator with a vane angle error of 5.5 minutes and the standard deviation for a population of 114 rotary vane attenuator calibrations at 10, 20, 30, and 40 db dial settings.

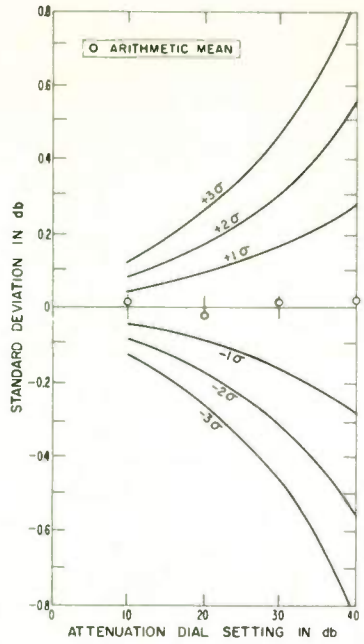


Fig. 3. Arithmetic mean and standard deviation for a population of 114 rotary vane attenuator calibrations at 10, 20, 30, and 40 db dial settings.

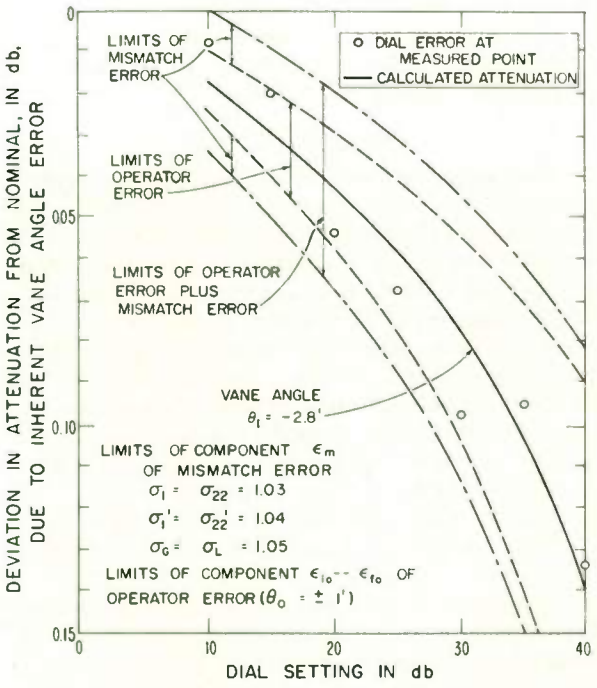


Fig. 4. Rotary vane attenuator with estimated vane angle error of -2.8 minutes determined from calibrated points measured at 11, 200 Mc. (IF substitution method).

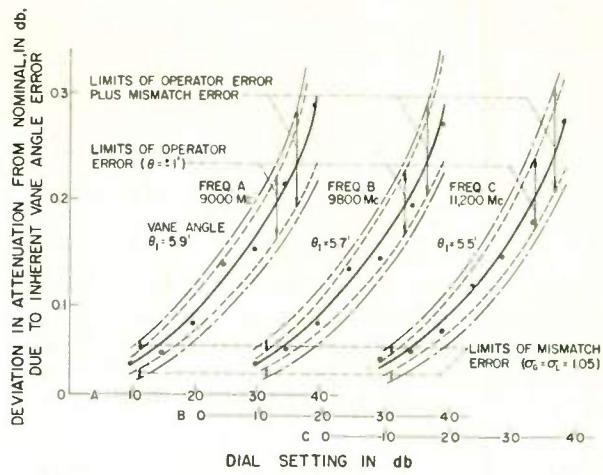


Fig. 5. Rotary vane attenuator calibrated at three frequencies A-9,000 Mc, B-9,800 Mc, and C-11,200 Mc, with an estimated vane angle error at A, 5.9 minutes, at B, 5.7 minutes, and at C, 5.5 minutes.

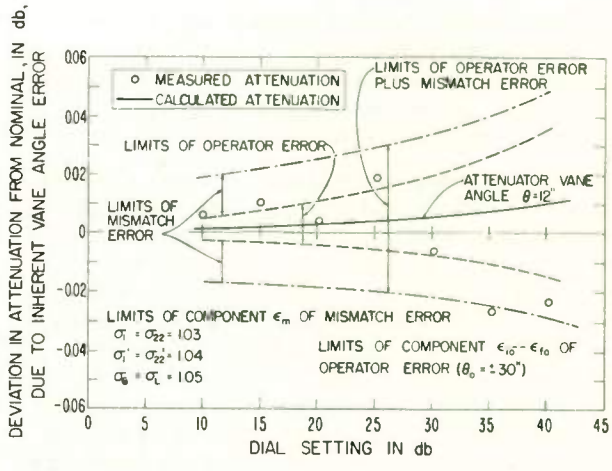


Fig. 6. A modified rotary vane attenuator with estimated vane angle error of 12 seconds determined from calibrated points measured at 9390 Mc. (IF substitution method).

MEASUREMENT OF EFFECTIVE TEMPERATURES OF MICROWAVE NOISE SOURCES

J. S. Wells, W. C. Daywitt, and C. K. S. Miller
National Bureau of Standards
Boulder, Colorado

Summary

The need for the calibration of gas discharge noise sources has existed for some time. In an effort to accommodate this need, work has been completed on a comparison system in X-band that uses a heated resistive wedge as a standard source of noise power.

The reference standard is based on Nyquist's Theorem and consists of a high temperature waveguide terminated by a suitably matched resistive element. The waveguide itself is one continuous piece with a heat sink on the flanged end. Two standards are in use; one of platinum - 13 percent rhodium waveguide and the other of gold waveguide.

The comparison system is a modified radiometer of the type used by Dicke, the principal modification being the use of one arm for comparison purposes and the other arm for a reference. The powers from an unknown noise source and a standard noise source are compared through a single arm, and an indication of the ratio of the power levels is read as an attenuation difference on a precisely calibrated variable attenuator. The noise temperature of the unknown source is then determined from the attenuation difference, the temperature of the attenuator, and the effective noise temperature of the standard.

An error analysis includes sources of errors in the comparison system and in the standard source. The errors due to the comparison system consist of: mismatch and calibration errors associated with the precision variable attenuator; chart resolution uncertainty; switch uncertainty; and the ambient temperature uncertainty of the attenuator. For the standard source the errors include the mismatch factor of the standard (and also the corresponding one for the unknown source), errors due to the uncertainty in the temperature of the hot load, and the uncertainties in the calculation of the effective noise temperature.

All of these uncertainties will give a total error of 0.07 db in the excess noise ratio of a suitable unknown noise source to be calibrated.

The evidence of good repeatability in a set of measurements is presented along with the determination of the excess noise ratio of a commonly used, terminated, tube-in-mount noise source.

Introduction

Recent work in microwave noise measurement techniques in the Radio Standards Laboratory has led to the completion of a hot load standard noise source and an improved microwave radiometer. These are now incorporated in a microwave noise calibration system.

This report includes a description of recent improvements in the radiometer, the reference standard noise source, and an analysis of possible errors involved in making a measurement. The results of experiments to determine the effective noise temperature of one type of commonly used argon noise source with a nominal pressure-radius product of 11.9 mm Hg-cm are presented. This source is used as a working standard for the calibration service in the frequency range 8.2 to 12.4 Gc.

Radiometer

The comparison system consists of a modified radiometer of the type developed by Dicke[1]. The principal modification of the Dicke radiometer is the use of one arm for comparison purposes and the other for establishing a reference noise level. A block diagram of the system is shown in Figure 1. The combining junction shown in this figure may be either of two configurations. When two noise sources of approximately equal power are being compared, a magic tee junction, Figure 2, is used. When two sources of very different power levels are compared, a directional coupler assembly is used, Figure 3. A photograph of the comparison system is shown in Figure 4. An earlier version of this system and its operation have been previously described[2]. Therefore, a detailed description of its components and operation will not be given.

In brief, the power from an unknown noise source and the power from a standard noise source are compared through a single arm of the radiometer, Figure 1, and the ratio of the power levels is determined from an attenuation

difference read on a precisely calibrated variable attenuator. The following procedure is used to determine the attenuation difference. The unknown source is connected to the radiometer input and the precision attenuator is set to one of the calibrated points in the vicinity of 10.6 db. A null output of the radiometer is then obtained and is achieved by readjusting the balance attenuator, which remains fixed during the remainder of the measurement. The hot load standard is then substituted for the unknown source and the precision attenuator readjusted to null output. The attenuation difference is then the difference between the attenuation readings on the precision attenuator at the two null conditions. The temperature of the unknown source is then determined from the attenuation difference, the temperature of the attenuator, and the effective temperature of the standard.

Reference Standard

The reference standard is based on Nyquist's Theorem and in practice consists of a high temperature waveguide terminated by a suitably matched resistive element. The waveguide itself is of one continuous piece with a heat sink in the form of a water jacket enclosing the flanged end. This design eliminates two flange joints and the resulting reflections which had existed in the previous standard[2], and since flange joints lack exact reproducibility, the new design permits greater uniformity. The use of a homogeneous material for the entire length of waveguide also simplifies the calculation of attenuation as a function of temperature.

The choice of materials for a high temperature waveguide is somewhat limited. Magnetic materials such as inconel and nickel were eliminated due to some uncertainty regarding their behavior in the vicinity of the Curie point. The materials given further consideration were restricted to the noble metals because of their relative stability and resistance to oxidation and corrosion. Most of these were eliminated for various reasons. Silver has too low a melting point; palladium and rhodium oxidize at relatively low temperatures; ruthenium, rhodium, and iridium are difficult to work, and osmium produces poisonous fumes when heated[3].

It was finally decided to utilize waveguides of two different materials, as shown in Figure 5. One is of gold and the other is of a platinum-13% rhodium alloy. Both gold and platinum are too soft for repeated use at high temperatures. The platinum-rhodium alloy is quite satisfactory mechanically and will maintain its shape over a long period; however, the uncertainty in its attenuation is larger than that of a gold waveguide. The attenuation of the gold waveguide is calcuable, but its softness and lower melting point are disadvantages. The gold waveguide was terminated by a silicon-carbide load, and the platinum-rhodium alloy waveguide was terminated by a zinc-titanate load. It was necessary to use something other than a silicon-carbide load in the platinum alloy waveguide, since platinum and silicon combine to form a eutectic which melts around 800° C[4].

The temperature controller and oven used to heat the waveguide and its resistive termination are shown in Figure 6. The temperature controller was designed by E. Campbell of the Bureau Laboratories and was found to be very effective as can be seen on strip 4 of Figure 10. A horizontal orientation of the oven was found to be more desirable than a vertical orientation for several reasons. Possible cooling of the load by convection currents is reduced. In addition the thermocouples which measure the temperature of the termination are brought out the rear of the oven and thus the possibility of error due to kinking is minimized. With the oven in the horizontal position the flange of the output terminal of the standard may be connected directly to a reflectometer for a more accurate measurement of the reflection coefficient. The load may be positioned within the waveguide for optimum performance by use of a ceramic rod inserted from the rear of the oven. The desired temperature distribution is achieved by controlling the current through three separate heater coils wound on a single core inside the oven. While in theory it would seem desirable to servo all three coils, proportional control of the current in the center coil has proved satisfactory. The sensing thermal junction is located longitudinally in the proximity of the hot load. The current through the auxiliary coils on either side of the central coil is adjusted manually to minimize the gradient along the length of the load. This gradient is measured by thermocouples in contact with the face of the load. The load itself contains a thermocouple that is used to monitor its temperature.

A combination water flow and temperature indicator is attached to the front of the oven. It was found necessary during the initial heating to keep the jacketed section of the waveguide considerably above room temperature to prevent water from condensing on the interior of the waveguide. Some earlier erroneous results are attributable to such condensation. Another source of error during the developmental stage was power absorption by an inorganic growth which formed in the inconel waveguide then in use, as pictured in Figure 7. The bulk of this material was lost in the process of measuring its power absorption. Hence a

proposed chemical analysis could not be attempted. Later efforts to duplicate this growth were unsuccessful.

Calculation of Output Temperature

General Analysis

The noise power entering the radiometer from the standard may be written as

$$P = kT_s B \frac{(1-|\Gamma_s|^2)(1-|\Gamma'|^2)}{|1-\Gamma'\Gamma_s|^2} ,$$

where the reflection coefficients, Γ, are defined in Figure 11 (Appendix 1), T_s is the effective temperature of the standard, k is Boltzmann's constant, and B the bandwidth of the system. T_s will be less than the measured value of the load temperature, due to the attenuation of the noise power by the walls of the high temperature waveguide. However, the resistivity of the walls also contributes to the noise output, and this must be taken into account.

The analysis used here to determine the output temperature of the standard is essentially that outlined by Sees[5]. His expression can be written

$$T_s = 10^{-A_g'/10}\left[T_m + \int_0^b a(z) T_g(z)\, e^{\int_0^z a(z')dz'} dz\right], \quad (1)$$

where:

T_s = the output noise temperature in °K

$T_g(z)$ = the temperature of the waveguide in °K

T_m = the measured temperature of the load in °K

A_g' = the total attenuation of the waveguide in decibels.

z = the axial distance from the tip of the load (for convenience the reference point was taken to coincide with the tip of the load) to the waveguide region under consideration in inches

b = the length of the waveguide from the tip of the load to the output flange in inches

a(z) = the attenuation in nepers per unit length, $= \left(\frac{a_o}{\rho_o^{1/2}}\right)\left(\rho\right)^{1/2}$,

where ρ is the resistivity of the waveguide. The subscript "$_o$" refers to 0°C. The quantities a_o and ρ_o are constants, depending on the size and composition of the waveguide.

Equation (1) is obtained from

$$T_s = T_m\, e^{-\int_0^b a(z')dz'} + \int_0^b a(z) T_g(z)\, e^{-\int_z^b a(z')dz'} dz$$

The physical interpretation of Equation (1) is made evident with this alternate form; the first term in the alternate expression is proportional to the power generated by the hot load and attenuated from 0 to b, and the second term is proportional to the sum of the power generated in each infinitesimal segment dz of waveguide at z and attenuated from z to b.

To make practical use of Equation (1), it is convenient to make several approximations. With the section of the waveguide between the hot load and the output flange considered to be divided into sufficiently small segments, the quantities 'a' and T_g can be considered constant over each segment. Then when Equation (1) is applied to each segment. the output of the kth segment may be expressed as

$$T_{k+1} = 10^{-A_k'/10}\left[T_k + \int_{z_k}^{z_{k+1}} a_k T_{gk}\, e^{\int_{z_k}^{z} a(z')dz'} dz\right], \quad (2)$$

which becomes

$$T_{k+1} = T_k - 0.23 A_k' (T_k - T_{gk}) , \quad (3)$$

provided that second and higher order terms in A_k, the total attenuation of the kth section, are discarded and the average values of a and T_g are taken. A_k' is related to a_k by $A_k' = 4.34\, a_k\, \Delta z_k$, where $4.34\, a_k$ is the average attenuation in decibels per unit length for the kth segment and Δz_k is the length of this segment.

To apply Equation (3), the temperature distribution along the waveguide must be measured and a(z) calculated as a function of this distribution. The expression for the attenuation per unit length (db/in) along the waveguide is given in Appendix 2 as

$$C_1 (\rho/f)^{1/2} \frac{(f^2 + C_2)}{(f^2 - C_3)^{1/2}} , \quad (4)$$

where f is the frequency in Gc, ρ is the resistivity in $\mu\Omega$ - cm. Equation (4) is used to calculate A'_k for each segment. Roeser and Wensel's[6] values for resistivity as a function of temperature for platinum and platinum-rhodium alloys appear in Table I. Values for resistivity as a function of temperature for gold are taken from Northrup's work[7].

The uncertainty in the output temperature of each section, δT_{k+1}, is found by taking the total differential of Equation (3) and is

$$\delta T_{k+1} = |\delta T_k| + 0.23(T_k - T_{gk})|\delta A_k'| + 0.23 A_k' |\delta(T_k - T_{gk})| . \quad (5)$$

The total uncertainty in T_s is then found from successive applications of Equation (5).

Sample Calculation

The analysis in the previous section will now be used to calculate the correction to the final output noise temperature of the hot load standard and also the accompanying uncertainty. Figure 8 shows a typical temperature distribution for a gold waveguide as measured by temporarily positioning a thermocouple at integral values of z inside the waveguide and reading the thermal emf after equilibrium had been attained. The tip of the load is located at z = 0. When the system is in steady state, the gradient can be held to within 1°C per load length. The temperature at z = 11 inches is dependent upon the temperature and rate of flow of water through the water jacket. Also indicated in Figure 8 are the assigned values of T_{gk} and A'_k for each segment.

T_s and δT_s are calculated from Equations (3) and (5) and Figure 8. The details of the calculation are shown in Table II. This calculation shows that T_s, the temperature of the reference standard, is reduced by 5.9°C; and that the uncertainty is T_s, δT_s, is 4.1°C.

Error Analysis

General Analysis

Several factors limit the accuracy to which the noise temperature of an unknown source can be measured. These are:

(1) Possible uncertainty in the power ratio. Two factors contribute to this uncertainty. One is due to the difficulty in recognizing the proper adjustment to return the system to a null after changing radiometer inputs. The other is due to the inability to evaluate one of the small terms in the first order balance equation.

A comparison of two noise sources by the radiometer requires the following adjustments: One noise source is attached to the input of the radiometer and the system is adjusted to a null by observing a level on a strip chart recorder. On replacing the first noise source by the second the precision attenuator is then readjusted to obtain the previous null. Obtaining these two nulls is made difficult by fluctuations in the recording and by the long integrating times needed for averaging.

The complete balance equation as given in Equation (11)of Appendix 1 is

$$10\log_{10}\frac{(T_{re}-T_a)}{(T_s-T_a)}\frac{1}{(1-|\Gamma_s|^2)} = \Delta A + 20\log_{10}\frac{|1-C_n\Gamma_n|}{|1-C_n\Gamma_s|} + 20\log_{10}\frac{|1-I_{22}A_n|}{|1-I_{22}A'_n|}\frac{|1-\Gamma_2\Gamma_L|}{|1-\Gamma'_2\Gamma_L|} + 10\log_{10}(1+\epsilon) \qquad (6a)$$

where the quantities are defined in the appendix.

Additional sources of noise power beside those being compared are generated in the attenuator section and receiver section[8] (as defined in Figure 11)of the radiometer. Portions of these two powers appear in the complete balance equation because of incomplete cancellation due to Γ_2 changing when the attenuation setting is changed. The uncertainty caused by this effect is incorporated in the term $10\log_{10}(1+\epsilon)$ of Equation (6a).

(2) System uncertainty. One might question how closely an indication of equality relates the effective noise temperatures of two sources. The variables here are the attenuator error, which includes mismatch as well as resolution and resettability, and the impedance matches at the input of the radiometer and the output of the noise power generators.

From Equation (11) of Appendix 1 it is apparent that the contribution to the system uncertainty due to the attenuator is

$$\frac{|1-I_{22}A_n|^2}{|1-I_{22}A'_n|^2}\frac{|1-\Gamma_2\Gamma_L|^2}{|1-\Gamma'_2\Gamma_L|^2}.$$

The corresponding contribution from the input mismatch is

$$\frac{|1-C_n\Gamma_n|^2}{|1-C'_n\Gamma_s|^2}.$$

All of the contributions to the uncertainties given in the first two parts of this section are due to the unknown phase of the pertinent reflection coefficients.

(3) The uncertainty in the noise temperature at the output terminal of the standard source. This depends on three factors: the accuracy with which temperatures can be measured, the accuracy to which the attenuation or resisitivity of the high temperature waveguide is known, and finally, on the approximations made.

If the last three terms of Equation (6a) are

included as a composite attenuation error, then the balance equation becomes, to first order,

$$\Delta A = 10 \log_{10} \frac{kB(T_{ne} - T_a)}{(T_s - T_a)(1 - |\Gamma_s|^2)kB}$$

$$= 4.34 \log_e \frac{(T_{ne} - T_a)}{(T_s - T_a)(1 - |\Gamma_s|^2)} \quad . \qquad (6b)$$

The quantities in the above equations are defined as follows:

- ΔA = the attenuation difference in db, as read on the precision attenuator
- k = Boltzmann's constant
- B = the bandwidth of the system
- T_a = the temperature of the attenuator in °K
- T_s = the hot load thermocouple temperature modified by the effective attenuation of of the guide in °K
- T_{ne} = the measured effective temperature of the noise tube and mount in °K. (The mount includes a termination at room temperature at one end).

By differentiating (6b) and rearranging, one obtains

$$\delta T_{ne} = 0.23(T_{ne} - T_a)\delta(\Delta A) + \frac{(T_{ne} - T_a)}{(T_s - T_a)} \delta T_s - \left[\frac{T_{ne} - T_a}{T_s - T_a} - 1\right] \delta T_a \qquad (7)$$

since the small uncertainty in $(1 - |\Gamma_s|^2)$ is negligible compared to the terms retained.

The quantities $\delta(\Delta A)$, δT_a and δT_s are the uncertainties in the first order balance equation. From Equation (7) the maximum uncertainty is

$$|\delta T_{ne}| \leq 0.23 (T_{ne} - T_a)|\delta(\Delta A)| + \frac{(T_{ne} - T_a)}{(T_s - T_a)} |\delta T_s| + \left[\frac{T_{ne} - T_a}{T_s - T_a} - 1\right] |\delta T_a| \quad . \qquad (8)$$

The quantities $\delta(\Delta A)$, δT_a and δT_s are discussed separately.

Associated Attenuation Uncertainties

The expression giving ΔA is found in Appendix 1. The uncertainty, $\delta(\Delta A)$, due to the radiometer, is composed of the last three terms of Equation (6a): a precision attenuator mismatch uncertainty $\delta(\Delta A)_\epsilon$, a source mismatch uncertainty $\delta(\Delta A)_m$, and a balance uncertainty $\delta(\Delta A)_u$. In addition to these three terms a fourth term and optional fifth term must be included in $\delta(\Delta A)$. The fourth term is due to the uncertainty in the calibration of the attenuator, $\delta(\Delta A)_c$, and the fifth term is due to the switch uncertainty, $\delta(\Delta A)_s$.

a. Precision Attenuator Mismatch Uncertainty. The maximum value of precision attenuator mismatch uncertainty is given by

$$\delta(\Delta A)_\epsilon = 20 \log_{10} \frac{(1 \pm |\Gamma_{22} A_{11}|)(1 \pm |\Gamma_L' \Gamma_L|)}{(1 \mp |\Gamma_{22} A_{11}'|)(1 \mp |\Gamma_L' \Gamma_L|)} \quad .$$

When the equality $\log_{10} x = 0.434 \log_e x$ and the approximation $\log_e (1 + x) \approx x$ are inserted into this expression, it becomes

$$\delta(\Delta A)_\epsilon \approx 8.68 \left[|\Gamma_{22}|(|A_{11}| + |A_{11}'|) + |\Gamma_L|(|\Gamma_L'| + |\Gamma_L'|)\right]. \qquad (9)$$

Multiple-stub tuners are used to reduce the reflection coefficients $|\Gamma_L|$ and $|\Gamma_{22}|$ to as small a value as possible in order to minimize $\delta(\Delta A)$. The uncertainty in measuring these parameters is small compared to their absolute value and may be ignored. It was found necessary to use two isolators in cascade following the tuners to insure that the measured reflections remained the same when the units were connected to the radiometer.

The reference planes, at which the magnitudes of the reflection coefficients are measured, are indicated in Figure 11 and discussed in Appendix 1. $|A_{11}|$, the reflection coefficient at the input of the attenuator, is measured with Γ_L replaced by a matched load ($\Gamma_L \leq 0.001$). Typical curves of reflection coefficient as a function of frequency are presented in Figure 9, with the IF signal and image frequency responses indicated by cross hatching. The primes of Equation (9) refer to the values of the quantities obtained with the attenuator set at approximately zero db, as is the case when the hot load is supplying the power at the radiometer input. Using Figure 9 one can calculate $\delta(\Delta A)_\epsilon$ with the result

$$\delta(\Delta A)_\epsilon \leq 0.016 \text{ db} \quad .$$

b. Source Mismatch Uncertainty

The second factor in the expression for ΔA may be considered to be an uncertainty due to multiple reflections resulting from two mismatched sources supplying noise power to a mismatched input terminal at different times. This

source mismatch factor may be written

$$\delta(\Delta A)_m \leq 20 \log_{10} \frac{(1 \pm |C_n \Gamma_n|)}{(1 \mp |C_n' \Gamma_s|)}$$

or,

$$\delta(\Delta A)_m \approx 8.68 \left[|C_n \Gamma_n| + |C_n' \Gamma_s| \right].$$

A typical value is

$$\delta(\Delta A)_m \leq 0.005 db.$$

c. Balance Uncertainty

Using typical values of the parameters in Equation (10) of Appendix 1 to calculate ϵ, one finds that

$$10 \log_{10}(1+\epsilon) \approx 0.006 db.$$

The experimental resolution of the radiometer is found by changing the precision attenuator setting a small amount and noting the resulting deflection on the recorder. Deflections less than this resolution must be included as a balance uncertainty. With a power level corresponding to a hot load temperature of 1000°C at the radiometer input, unbalances of less than 0.005 db are discernible at the output, and unbalances of 0.002 have been detected. The limit of the resolution is certainly less than 0.01 db.

Adding these two uncertainties the total balance undertainty is

$$\delta(\Delta A)_b \leq 0.016 db.$$

d. Attenuator Calibration Uncertainty

The limits of resolution and repeatability were established by means of observations made during the precise calibration of the precision attenuator and are estimated to be ± 0.002 db. The calculated mismatch error was less than 0.005 db. Therefore, in summing both contributions, the uncertainty in the precision attenuator calibration is

$$\delta(\Delta A)_c \leq 0.007 db.$$

The requirement for keeping the error in the excess noise ratio to within 0.1 db necessitated a more precise calibration of the attenuator than was commonly available. The method decided upon was a subcarrier technique[9]. In order to assure the accuracy of the attenuator, it is calibrated periodically. The need for assurance was pointed up when it was noted that a previous calibration had changed by 0.04 db at the higher levels of attenuation during an 18 month period.

e. Switch Uncertainty

Another possible error (due to the switch) should be mentioned when discussing attenuation uncertainties. The system is sufficiently stable that the reference standard and the unknown source may be attached alternately to the same input terminal and may be compared without changing the position of the switch. Thus, in this case, the error due to switching is eliminated.

Comparison with the working standard (Figure 1), however, depends on the repeatability of the switch. Both reflection coefficient measurements on the switch and tests of the switch in the system indicate that the switch uncertainty is less than 0.002 db, or

$$\delta(\Delta A)_s \leq 0.002 db.$$

The maximum associated attenuation uncertainty is then obtained by adding the uncertainties of a, b, c, d, and e. The result is

$$\delta(\Delta A) \leq 0.046 db.$$

Uncertainty in the Temperature of the Radiometer

δT_a is the uncertainty in measuring the ambient temperature of the attenuator section of the radiometer. The room is temperature controlled, and T_a does not vary appreciably. A thermocouple was taped to the side of the resistive element in an attenuator (which had been substituted for the precision attenuator) to measure heating due to radiation when the thermal noise standard is connected to the radiometer. An increase in temperature of less than one degree was observed 30 minutes after noise power from the reference standard had been applied. This is the normal time duration for a measurement. The waveguide switch apparently absorbs and reflects a large portion of the infrared energy from the hot load as the reading of a thermometer directly in front of the terminal of the hot load standard rose about 40°C during this same time.

As may be seen by examining the Equation (8), a δT_a of 1°C in the equation causes an uncertainty which is small compared to that contributed by the other sources. Hence a conservative estimate is

$$\delta T_a \leq 2^\circ .$$

Uncertainty in the Temperature of the Standard

It is convenient to define δT_s as $\delta T_m + \Sigma\delta T_k$, where δT_m is the uncertainty in the measurement of the load temperature and is due to a small inaccuracy in the calibration of the thermocouple and to its physical placement at a small distance from the point of interest. Error due to the latter is reduced by minimizing the temperature gradient across the load and waiting sufficiently long to ensure thermal equilibrium. The quantity $\delta T_m / T_m$ can be as much as ± 0.25 percent. However, precautions are taken to minimize this uncertainty by annealing the platinum thermocouple at approximately 1450°C for one hour. The utmost care is exercised to avoid contamination in construction of the thermocouples. Insulation and protection tubes of high purity alumina (better than 99.0 percent) are used for greater stability and accuracy. The portion of the thermocouple not enclosed in protective ceramic tubing is prevented from kinking by enclosing it in thick walled plastic tubing.

The $\Sigma\delta T_k$ contribution to the hot load standard uncertainty represents uncertainties arising from the waveguide portion of the hot load and have already been discussed in "Calculation of Output Temperature". The addition of both of these contributions leads to

$$\delta T_s \leq 4.1^\circ .$$

Maximum Total Uncertainty

The maximum total uncertainty given by Equation (8) will now be calculated. For representative values it is assumed that T_s is 962°C, that T_a is 23°C, and that T_{ne} is 10,796°K. Then $(T_{ne} - T_a)/(T_s - T_a)$ is 11.23 and from Table II δT_s is found to be 4.1°C. $\delta(\Delta A)$ is 0.046 db and δT_a is 2°C, where the values are found on this and the preceding page. Substituting these values into Equation (8) gives

$$|\delta T_{ne}| \leq (0.23)(10{,}500) + (2)(10.23) + (4.1)(11.23) \approx 178^\circ C .$$

The excess noise ratio, ENR, is defined to be

$$ENR = 10 \log_{10} \left[\frac{T_{ne} - T_o}{T_o} \right],$$

where $T_o \equiv 290^\circ K$.

From this equation it is easily found that

$$\delta(ENR) = 10 \log_{10} \left[1 + \frac{\delta T_{ne}}{T_{ne} - T_o} \right]$$

Substituting the above values into this equation their results

$$\delta(ENR) \leq 0.070 \text{ db} .$$

These results represent the maximum total uncertainty in the measurement of effective noise temperature and excess noise ratio of a waveguide noise source given in this analysis.

Experimental Results

The performance of the comparison system may be determined by examining Figure 10 which contains recordings of the radiometer output and a temperature recording made simultaneously with part of Strip No. 3. Corresponding points in time are indicated by pairs of markers designated X, Y, and Z. The stability and sensitivity at an excess noise temperature of 10,500°K are indicated on Strip No. 1. The distance from the bottom of the chart to the point '1a' represents six hours. The power level was changed 0.02 db at '1a' and returned to its original level at '1b'. A magic tee was the combining junction for the result indicated here.

The power level in Strip No. 2 corresponds to an excess temperature of 940°K at the radiometer input which is the nominal operating temperature of the hot load. The increase in fluctuation of the amplitude is due to the determination of the signal-to-noise ratio at the lower signal input temperature and the consequent increase in gain required to maintain the same nominal deflection sensitivity. Strip No. 2 shows changes of 0.02 db at '2a' and 0.04 db in the opposite direction at '2b'. Here the magic tee has been replaced by the directional coupler assembly.

Strips Nos. 3 and 4 are indicative of the types of records made during an actual measurement. At the beginning of the measurement the precision attenuator was adjusted to 10.62 db. A small unbalance was noted and the balancing attenuator was used to obtain a null output as indicated at '3a'. After sufficient time to establish the reference level, the recorder input was short circuited and waveguide source No. 1 was replaced by the hot load standard. The precision attenuator was changed from 10.62 db and the output was recorded again at '3b'. Note that the temperature indicated on Strip No. 4 increases until point Y. At point Y an unbalance is detectable on Strip No. 3 and the attenuator repositioned to 0.10 db, and the system uninterrupted for 30 minutes, during which time the temperature is

stable as indicated. At '3d', the waveguide noise source No. 1, is substituted for the hot load standard and attenuator returned to 10.62 db. Flange reproducibility and attenuator resettability appear good. At '3e', waveguide noise source No. 2 is attached to the system and no other adjustments made.

The results of two sets of 10 consecutive runs such as illustrated on Strip No. 3 are shown in Table III to indicate the system repeatability and self consistency. One set was taken using the gold and silicon-carbide model of the standard noise source. The other set was obtained using the platinum alloy and zinc-titanate model. The center calibration frequency was 9.8 Gc and the average excess noise ratios were 15.59 db for the gold combination and 15.58 db for the platinum combination. In either case the uncertainty was less than 0.1 db.

To obtain a better idea of the reproducibility of gas-tube noise sources, several different comparisons were made. For this purpose, the 10 db coupler assembly was replaced by the magic tee assembly, and the radiometer was operated at a level corresponding to an excess temperature of 10,500°K at its input. Strip No. 1 in Figure 10 indicates the stability and sensitivity at this level.

Four each of two different specially constructed noise tubes, one type using a heated cathode, the other operating with a cold cathode, were compared in the same mount, which remained fixed to the comparison system. The spread in excess noise ratio was less than 0.02 db for both sets. Two tubes with nearly identical excess temperatures were then put in different mounts, and the tube-in-mount units were compared. The output of these units differed by as much as 0.05 db. This seems to suggest that the repeatability among tubes is better than that for mounts. Factors, such as a flaw in the glass envelope and a deviation from nominal mount insertion angle were found to cause even larger differences.

Conclusion

Because of the very good agreement between measurements made with the hot load noise sources it was decided that the outputs from both sources were sufficiently calculable, and that either could be used as a reference noise source. From experimental results discussed it is apparent that the system has sufficient sensitivity for the purpose of comparing sources, and the above analysis shows that the system is capable of the desired accuracy. It was concluded from these considerations that a calibration service for x-band waveguide noise sources could be offered.

Due to small differences in output power of of one tube in different mounts it was concluded that for a calibration service to be offered with the desired accuracy it would be mandatory for the units submitted for calibration to consist of a tube-in-mount.

Consequently, a calibration service is being offered at the selected frequencies of 9.0 Gc, 9.8 Gc, and 11.2 Gc to the accuracy of ± 250°K of the effective noise temperature and ± 0.1 db of the excess noise ratio.

Acknowledgements

The authors wish to acknowledge A. J. Estin for his contributions concerning the analysis of the reference standard during an earlier collaboration. The fruitful discussions with G. F. Engen which have helped further the state of the art and his suggestions leading to a more compact error expression are greatly appreciated. Thanks are due to H. W. Lance for suggestions regarding the paper and to R. E. Larson for his interest and support of the project. The technical services of W. Foote and E. Campbell were indispensible in the completion of the present phase of the work.

Appendix 1

Derivation of Balance Equation

Figure 11 depicts how the radiometer is considered to be divided for the purpose of this analysis. With reference to Figure 1; the input terminal of the radiometer is coincidental with the input plane of Figure 11, and the terminal surface between the output of the precision attenuator and the following tuner is coincidental with the comparison plane of Figure 11. The attenuator section is considered to be at constant temperature, T_a, in degrees Kelvin, with an efficiency η from the input to the output of the receiver section. The input portion of the receiver section is assumed to be at a constant temperature, T_a, with an efficiency ξ from the output to the input and the output portion at a constant temperature, T_r, in degrees Kelvin. The input portion of the receiver section consists of the tuner, and isolator immediately following the comparison plane. All Γ's appearing in Figure 11 are reflection coefficients.

When used for the measurement described in this paper, the radiometer may be considered as a device which indicates the equality of noise power propagating toward the comparison plane under two different conditions. This noise power is composed of power originating in a noise generator at temperature T which is transmitted by the network, $P_a(T)$, power originating to the right of the comparison plane (Figure 11) which

is reflected back into the radiometer by impedance discontinuities denoted by either Γ_2 or Γ_2'. Let P denote power originating to the left of the comparison plane and p denote power originating to the right of this plane. The subscripts $_i$, $_r$, and $_d$ refer to the incident, reflected and delivered power respectively. The primed variables refer to the situation in which the thermal noise standard is connected to the input terminal of the radiometer; the same variables without primes refer to the situation in which the waveguide noise source is connected to the input terminal. When one noise source is compared to another and the system is returned to a null, the power entering the receiver section in either case must be the same. Since noise power from independent sources can be added linearly, we get

$$P_i + p_r = P_i' + p_r' \qquad (1)$$

where

$$P_i = \frac{P_d}{(1-|\Gamma_L|^2)} \quad \text{and} \quad p_r = \frac{p_d |\Gamma_2|^2}{(1-|\Gamma_2|^2)}$$

Similar expressions hold for P_i' and P_r', in which all the quantities except Γ_L become primed Figure 11 may be used to relate P_d and p_d to the measured quantities of interest.

To find P_d, it is necessary to find P_s, the noise power delivered from a source at temperature T to a passive load (i.e. a load at 10°K); P_n, the power contributed by the network; and η, the efficiency of the network.

Combining the expression due to Nyquist for the average voltage squared,

$$\overline{e^2} = 4kTBRe(Z)$$

with an expression for the power delivered to a mismatch load from a mismatch generator

$$P = \frac{e^2}{4Z_o} \frac{|1-\Gamma_g|^2(1-|\Gamma_h|^2)}{|1-\Gamma_g\Gamma_h|^2}$$

k is Boltzmann's Constant, T is the temperature of the source, B is the bandwidth of the receiver, Re(Z) is the real part of the source impedance, Z_o is the characteristic impedance of the line, Γ_g is an arbitrary source reflection coefficient, and Γ_h an arbitrary load reflection coefficient. The result is

$$P = kTB \frac{(1-|\Gamma_g|^2)(1-|\Gamma_h|^2)}{|1-\Gamma_g\Gamma_h|^2} \qquad (2)$$

Using Equation (2), the power delivered by a source at temperature T, in degrees Kelvin, connected to the input plane is

$$P_s(T) = kTB \frac{(1-|\Gamma_h|^2)(1-|\Gamma_i|^2)}{|1-\Gamma_i\Gamma_h|^2} \qquad (3)$$

P_η will be determined. With the receiver section at a temperature T_a the expression for the power delivered to the left of the comparison plane is

$$P_L(T_a) = kT_aB \frac{(1-|\Gamma_2|^2)(1-|\Gamma_L|^2)}{|1-\Gamma_2\Gamma_L|^2}$$

However, when the networks on both sides of the comparison plane are at the same temperature T_a,

$$P_L(T_a) = P_s(T_a)\eta + P_\eta(T_a)$$

or solving for $P_\eta(T_a)$

$$P_\eta(T_a) = P_L(T_a) - P_s(T_a)\eta \qquad (4)$$

In general

$$P_L = P_s(T)\eta + P_\eta(T_a) \qquad (4a)$$

The expression for the efficiency η[10] of the network when applied to this system is

$$\eta = \frac{|C_{21}|^2(1-|\Gamma_L|^2)}{|1-C_{22}\Gamma_L|^2(1-|\Gamma_i|^2)} \qquad (5)$$

Where C_{22} is the appropriate element of the scattering matrix for the attenuation section, which consists of a matched isolator and an attenuator. The element C_{21} is the transmission scattering coefficient for the attenuator section and is determined by the usual method of cascading matrices. The quantities I_{ij} and A_{ij} are the elements of the scattering matrices for the isolator and the attenuator respectively

$$|C_{21}|^2 = \frac{|I_{21}A_{21}|^2}{|1-I_{22}A_{11}|^2} \qquad (6)$$

When Equations (3), (4), and (5), along with the relations

$$(1-\Gamma_i\Gamma_h)(1-C_{22}\Gamma_L) = (1-C_{11}\Gamma_h)(1-\Gamma_s\Gamma_L)$$

and

$$(1-\Gamma_i'\Gamma_s)(1-C_{22}'\Gamma_L) = (1-C_{11}'\Gamma_s)(1-\Gamma_2'\Gamma_L)$$

are substituted into an expression (Equation 4a) for the power delivered to the receiver section and when the temperature of the generator is T_n, the result is

$$P_d = kB(1-|\Gamma_L|^2)\left[\frac{(T_n-T_a)(1-|\Gamma_n|^2)\,|C_{11}|^2}{|1-C_{11}\Gamma_n|^2\,|1-\Gamma_2\Gamma_L|^2} + \frac{T_a(1-|\Gamma_2|^2)}{|1-\Gamma_2\Gamma_L|^2}\right] \qquad (7a)$$

where C_{11} is a scattering coefficient for the attenuator section. Similarly

$$P_d' = kB(1-|\Gamma_L|^2)\left[\frac{(T_s-T_a)(1-|\Gamma_s|^2)\,|C_{21}'|^2}{|1-C_{11}'\Gamma_s|^2\,|1-\Gamma_2'\Gamma_L|^2} + \frac{T_a(1-|\Gamma_2'|^2)}{|1-\Gamma_2'\Gamma_L|^2}\right]. \qquad (7b)$$

Note the common factor in the denominator of the two terms comprising the expressions for either P_d. In a similar fashion the power represented by p consisting of power from a receiver at temperature T_r and the network ξ at T_a

$$P_d' = kB\left[(T_r-T_a)F\xi + T_a(1-|\Gamma_L|^2)\right]\frac{(1-|\Gamma_2|^2)}{|1-\Gamma_2\Gamma_L|^2}$$

$$P_d' = kB\left[(T_r-T_a)F\xi + T_a(1-|\Gamma_L|^2)\right]\frac{(1-|\Gamma_2'|^2)}{|1-\Gamma_2'\Gamma_L|^2}. \qquad (8)$$

The details of the matching factor F need not be known since ξ is of the order of 10^{-12}, and hence this term may be neglected. When Equations (6), (7), and (8), are combined with (1), the resulting balance equation may be shown to be

$$\frac{(T_n-T_a)}{(T_s-T_a)}\,\frac{(1-|\Gamma_n|^2)}{(1-|\Gamma_s|^2)} = \frac{|A_{21}'|^2}{|A_{21}|^2}\,\frac{|1-C_{11}\Gamma_n|^2}{|1-C_{11}'\Gamma_s|^2} \times \frac{|1-I_{22}A_{11}|^2}{|1-I_{22}A_{11}'|^2}\,\frac{|1-\Gamma_2\Gamma_L|^2}{|1-\Gamma_2'\Gamma_L|^2} \times (1+\epsilon) \qquad (9)$$

where

$$\epsilon \approx \frac{T_a\,|1-C_{11}'\Gamma_s|^2\,|1-\Gamma_2'\Gamma_L|^2}{(T_s-T_a)\,|C_{21}'|^2(1-|\Gamma_s|^2)} \times \left[\frac{1}{|1-\Gamma_2'\Gamma_L|^2} - \frac{1}{|1-\Gamma_2\Gamma_L|^2}\right]. \qquad (10)$$

Terms of order $\xi \approx 10^{-12}$ and $|\Gamma_L|^2\,|\Gamma_2'|^2 \times (|1-\Gamma_2'\Gamma_L|^2)^{-1} \approx 10^{-6}$ have been neglected in the approximation for ϵ.

Let

$$10^{\Delta A/10} \equiv \frac{|A_{21}'|^2}{|A_{21}|^2}$$

and let T_{ne} be defined as follows:

$$(T_n-T_a)(1-|\Gamma_n|^2) \approx T_n(1-|\Gamma_n|^2) - T_a \equiv T_{ne} - T_a.$$

Using the two definitions given above it is easily shown that Equation (9) can be written as

$$10\log_{10}\frac{(T_{ne}-T_a)}{(T_s-T_a)(1-|\Gamma_s|^2)} = \Delta A + 10\log_{10}(1+\epsilon) + 20\log_{10}\frac{|1-C_{11}\Gamma_n|}{|1-C_{11}'\Gamma_s|} + 20\log_{10}\frac{|(1-I_{22}A_{11})(1-\Gamma_2\Gamma_L)|}{|(1-I_{22}A_{11}')(1-\Gamma_2'\Gamma_L)|}. \qquad (11)$$

Appendix 2

Calculation of the Attenuation of a High Temperature Waveguide

The field intensity in a rectangular waveguide of finite conductivity is known[11] to vary as $e^{j(\omega t-\gamma z)}$, where γ, the propagation constant, is defined as

$$\gamma = \gamma_0\left[1+(1-j)\frac{\mu_m\,\delta}{\mu\,\gamma_0^2}\left(\frac{\pi^2}{a^3}+\frac{k^2}{2b}\right)\right]. \qquad (1)$$

and

$$k^2 = \omega^2\mu\epsilon \quad , \quad \mu_m = \mu_0(1+\chi_m)$$

$$\gamma_0^2 = k^2 - \left(\frac{\pi}{a}\right)^2$$

$$\delta = \left(\frac{2}{\mu\omega\sigma}\right)^{1/2} = \left(\frac{2\rho}{\mu\omega}\right)^{1/2}.$$

a = large dimension of the waveguide

b = small dimension of the waveguide

ω = angular frequency in radians/sec.

σ = conductivity of the waveguide walls

μ_m = permeability of the waveguide walls

μ = permeability of the medium in the waveguide

ϵ = dielectric constant of the medium in the waveguide

ρ = resistivity of the waveguide walls

δ = skin depth in the waveguide walls

The units for the above definitions are all m.k.s. units.

The imaginary part γ_ϵ of the propagation constant causes a diminution of the field intensity and is related to the attenuation along the waveguide in db, A', by the definition

$$A' = 20(\log_{10} e)\,\gamma_i\,\Delta Z, \qquad (2)$$

or

$$A'/\Delta Z = 20(\log_{10} e)\,\gamma_i \quad \text{db/m},$$

where Δz is an incremental length along the waveguide.

$$\gamma_i = \mathrm{Im}(\gamma) = \frac{\mu_m \delta}{\mu \gamma_0}\left(\frac{\pi^2}{a^3} + \frac{k^2}{2b}\right)$$

$$= \mu_m \frac{(2\rho)^{1/2}}{(\mu^3\omega)^{1/2}} \frac{(\pi^2/a^3 + k^2/2b)}{(k^2 - \pi^2/a^2)^{1/2}}$$

Then,

$$A'/\Delta Z = 20(\log_{10} e)\,\gamma_i$$

$$= \frac{20(\log_{10} e)\,\mu_m (2\rho)^{1/2}}{(\mu^3\omega)^{1/2}} \frac{(\pi^2/a^3 + k^2/2b)}{(k^2 - \pi^2/a^2)^{1/2}},$$

which can be written as

$$A'/\Delta Z = C_1 \frac{\rho^{1/2}}{f^{1/2}} \frac{(f^2 + C_2)}{(f^2 - C_3)^{1/2}},$$

if, the constants C_1, C_2, and C_3 are adjusted so that $A'/_{\Delta z}$ reads in db/in, when the frequency is expressed in gigacycles per second and the resistivity is in micro-ohm-centimeters, their values are found to be $C_1 = 36.2 \times 10^{-5}$, $C_2 = 38.2$ and $C_3 = 43.0$ for x-band waveguide.

References

1. Dicke, R. H., "The Measurement of Thermal Radiation at Microwave Frequencies", Rev. Sci. Instr., 17, No. 7, p. 268, (1946).
2. Estin, A. J., Trembath, C. L., Wells, J.S., Daywitt, W. C., "Absolute Measurement of Temperatures of Microwave Noise Sources", Trans. IRE, I-9, No. 2, p. 209, (1960).
3. Basic Metallurgy, Edited by A. W. Grosvenor, Am. Soc. of Metals, Cleveland, Ohio, (1954).
4. Zysk, E. D., "Platinum Metal Thermocouples", Smyp. on Temp. and Meas. Control in Science and Industry, Sponsored by Am. Inst. of Phys., Instr. Soc. of Am., and NBS. Columbus, Ohio, (1961).
5. Sees, J. E., "Fundamentals in Noise Source Calibrations at Microwave Frequencies", Naval Res. Lab., Report 5051, Washington, D. C., (1958).
6. Roeser, W. F., and Wensel, H. T., "Temperature-its Measurement and Control in Science and Industry", Reinhold Publishing Corporation, New York, New York, pp. 1293-1326, (1941).
7. Northrup, E. F., "Resistivity of Pure Gold in Temperature Range 20°C to 1500°C", J. of Franklin Inst., 177, No. 3, pp. 287-292, (1914).
8. Richardson, J. M., private communication.
9. Schafer, G. E., and Bowman, R. R., "Preliminary Report on a Modulated Subcarrier Technique of Measuring Microwave Attenuation", To be published in a forthcoming issue of Journal of Institute of Electrical Engineers, London.
10. Beatty, R.W., "Mismatch Errors in the Measurement of Ultra-High-Frequency and Microwave Variable Attenuators", Journal of Research of NBS, Vol. 52, No. 1, pp. 7-9, January, 1954.
11. Kerns, D. M., and Hedberg, R. W., "Propagation Constant in Rectangular Waveguide of Finite Conductivity", Journal of Applied Physics, 25, No. 12, p. 1550, (1954).

General Reference

Bell, D. A., Electrical Noise, D. VanNostrand Company, Ltd., London, England, (1960).

Vines, R. F., and Wise, E. M., The Platinum Metals and their Alloys, The International Nickel Company, Inc., New York, New York, (1941).

TABLE I

Resistivity of Waveguide Materials as a Function of Temperature

Temp. °C	Platinum 9.83 μΩ-cm* ρ/ρo	90% Platinum 10% Rhodium 18 μΩ-cm* ρ/ρo	87% Platinum 13% Rhodium 19.0 μΩ-cm* ρ/ρo	Gold ρ
0	1.000	1.000	1.000	
100	1.392	1.166	1.156	2.97 μΩ-cm
200	1.773	1.330	1.308	3.83
300	2.142	1.490	1.456	4.72
400	2.499	1.646	1.601	
500	2.844	1.798	1.744	6.62
600	3.178	1.947	1.885	
700	3.500	2.093	2.030	
800	3.810	2.234	2.157	9.94
900	4.109	2.370	2.287	
1000	4.396	2.503	2.414	12.52
1100	4.671	2.633	2.538	
1200	4.935	2.761	2.660	
1300	5.187	2.887	2.780	
1400	5.427	3.011	2.898	
1500	5.655	3.133	3.014	

Data for Platinum alloys given by Roeser and Wensel[7]
Data for Gold given by Northrup[8]
*Resistivities at zero degree centigrade

TABLE II

Typical Calculations of Effective Temperature of Power from Gold Waveguide

$$T_{k+1} = T_k - 0.23\ A'_k\ (T_{gk} - T_k)$$

k	T_k	$.23A'_k$	$(T_k - T_{gk})$	T_{k+1}
0				963
1	963	.00169	(963 - 959)	963 - .007
2	963	.00169	(963 - 952)	963 - .026
3	963	.00167	(963 - 938)	963 - .068
4	963	.00164	(963 - 908)	963 - .158
5	963	.00160	(963 - 846)	963 - .344
6	963	.00152	(963 - 762)	963 - .650
7	962	.00140	(962 - 648)	963 - 1.090
8	962	.00129	(962 - 516)	963 - 1.666
9	961	.00114	(961 - 382)	963 - 2.326
10	961	.00101	(961 - 248)	963 - 3.046
11	960	.00087	(960 - 122)	963 - 3.775
12*	959	.00233	(959 - 44)	963 - 5.907 ≈ 957.1

Cumulative error in calculating the effective temperature

$$|\delta T_{k+1}| = |\delta T_k| + 0.23(T_k - T_{gk})|\delta A'_k| + 0.23A'_k|\delta(T_k - T_{gk})|$$

k	$\lvert\delta T_k\rvert$	$0.23(T_k - T_{gk})\ \lvert\delta A'_k\rvert$**	$0.23A'_k\ \lvert\delta(T_k - T_{gk})\rvert$	$\lvert\delta T_{k+1}\rvert$
0	3.000°			3°
1	3.000°	$0.9° \times 1.47 \times 10^{-3}$	$.00169 \times 6°$	3 + 0.011
2	3.011°	$2.5° \times 1.47 \times 10^{-3}$	$.00169 \times 6°$	3 + 0.025
3	3.025°	$5.7° \times 1.45 \times 10^{-3}$	$.00167 \times 6°$	3 + 0.043
4	3.043°	$12.7° \times 1.43 \times 10^{-3}$	$.00164 \times 6°$	3 + 0.071
5	3.071°	$26.9° \times 1.39 \times 10^{-3}$	$.00160 \times 6°$	3 + 0.118
6	3.118°	$46.2° \times 1.32 \times 10^{-3}$	$.00152 \times 6°$	3 + 0.188
7	3.188°	$72.2° \times 1.22 \times 10^{-3}$	$.00140 \times 6°$	3 + 0.285
8	3.285°	$102.6° \times 1.12 \times 10^{-3}$	$.00129 \times 6°$	3 + 0.408
9	3.408°	$133.3° \times 1.00 \times 10^{-3}$	$.00114 \times 6°$	3 + 0.548
10	3.548°	$164.0° \times 0.88 \times 10^{-3}$	$.00101 \times 6°$	3 + 0.698
11	3.698°	$192.7° \times 0.76 \times 10^{-3}$	$.00087 \times 6°$	3 + 0.860
12*	3.860°	$210.5° \times 0.68 \times 10^{-3}$	$.00233 \times 6°$	3 + 1.143 ≈ 4.1

* Includes 11 - 14 inches.

**$\delta A_k = 0.2\ A_k$

TABLE III

Computed Excess Noise Ratios from Ten Consecutive Test Runs using the Gold Waveguide with a Silicon Carbide Termination

T_m °C	T_s °C	T_a C	ΔA db	T_{ne} °K	ENR in db
958.4	952.6	25.0	10.54	10,800	15.593*
961.2	955.4	25.9	10.53	10,800	15.592
966.6	960.7	26.0	10.51	10,810	15.596
964.5	958.6	25.9	10.50	10,770	15.580
976.4	970.6	26.1	10.46	10,800	15.592
984.5	978.5	26.0	10.43	10,820	15.598
986.0	980.0	26.3	10.41	10,780	15.584
999.5	993.5	25.7	10.35	10,790	15.587
1000.0	994.0	25.8	10.35	10,790	15.589
1010.6	1004.5	25.9	10.30	10,780	15.586

Average T_{ne} = 10,790°K

ENR = 15.59 db

Computed Excess Noise Ratios from Ten Consecutive Test Runs using the Platinum-Rhodium Waveguide with a Zinc Titanate Termination

T_m °C	T_s °C	T_a °C	ΔA db	T_{ne} °K	ENR in db
947.3	934.5	22.5	10.60	10,770	15.578*
972.0	959.0	25.2	10.50	10,780	15.581
976.6	963.6	24.9	10.48	10,780	15.585
995.6	982.4	24.3	10.39	10,780	15.582
1008.0	994.7	23.3	10.34	10,800	15.593
1019.5	1006.0	24.4	10.28	10,770	15.578
1021.1	1007.6	24.5	10.28	10,780	15.585
1047.4	1033.7	23.9	10.17	10,800	15.591
1059.2	1045.4	23.3	10.11	10,780	15.584
1060.0	1046.2	23.3	10.10	10,760	15.577

Average T_{ne} = 10,780°K

ENR = 15.58* db

*The fifth place is retained for averaging only

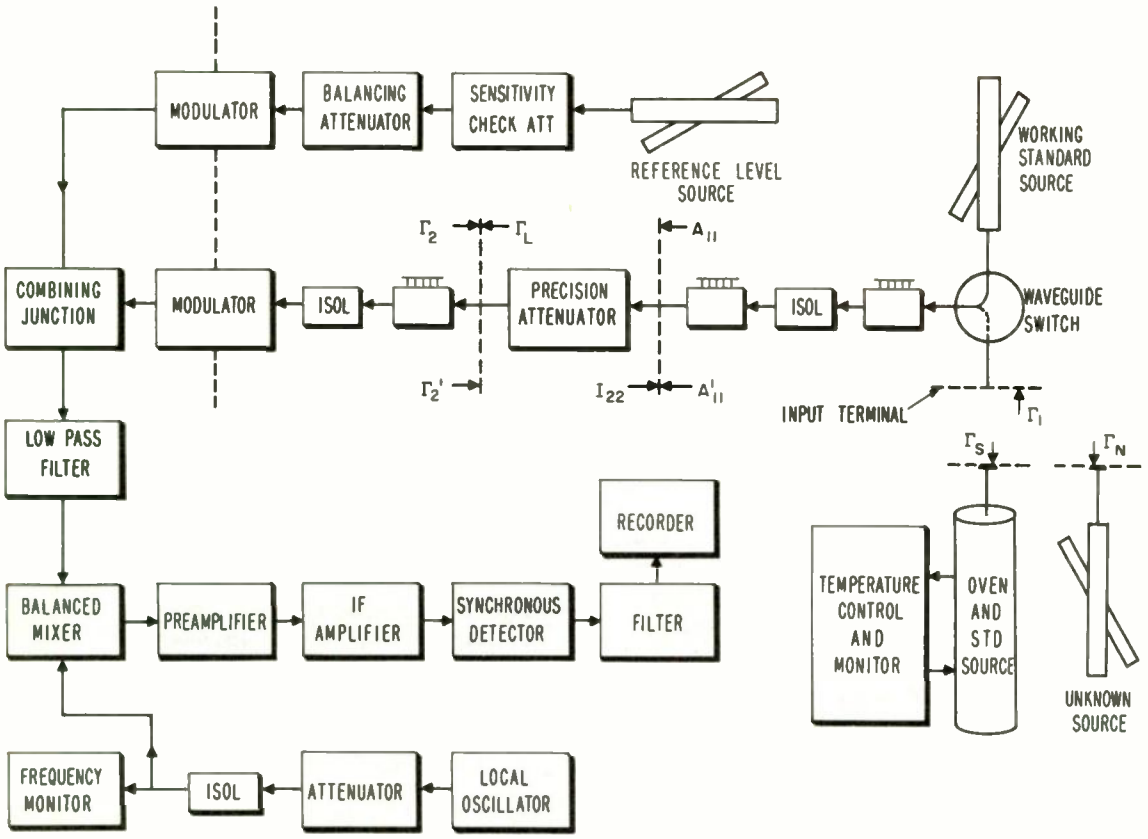


Fig. 1. Block diagram of the system.

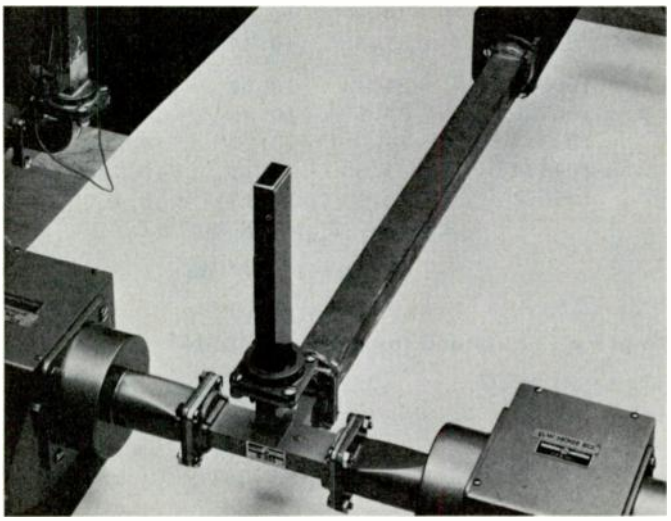


Fig. 2. Magic tee junction of the radiometer.

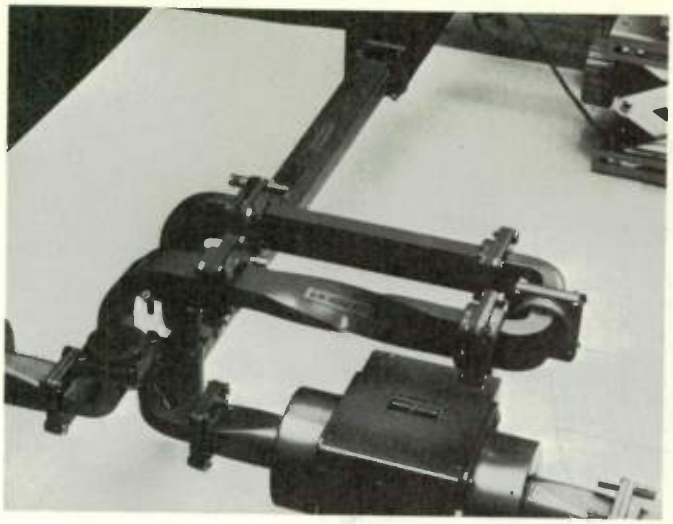


Fig. 3. Directional coupler assembly of the radiometer.

Fig. 4. The radiometer.

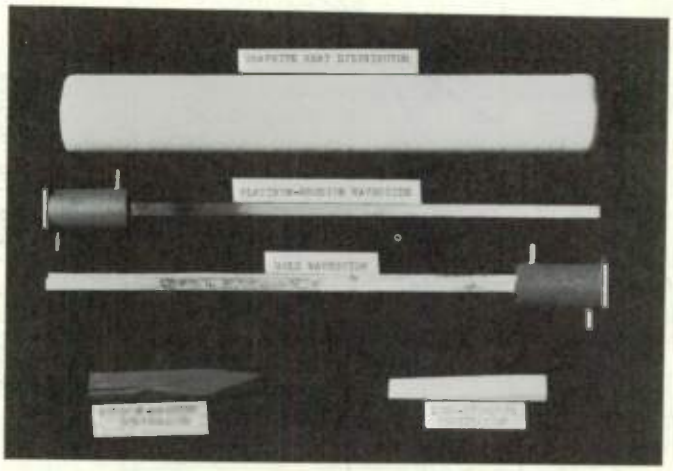


Fig. 5. Heat distributor, load, and high temperature waveguides.

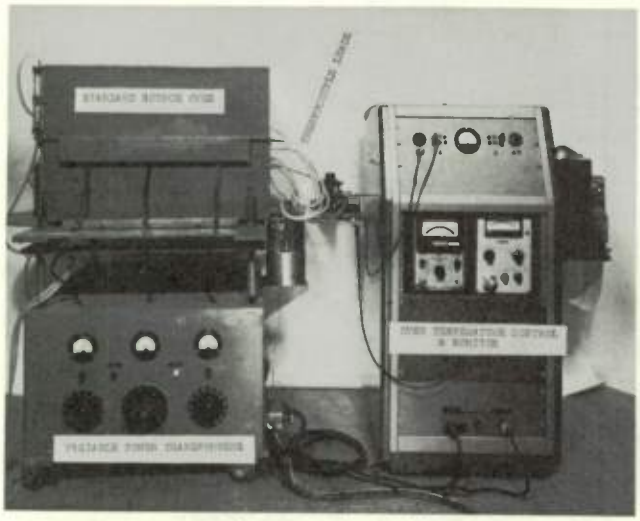


Fig. 6. Temperature controller and oven.

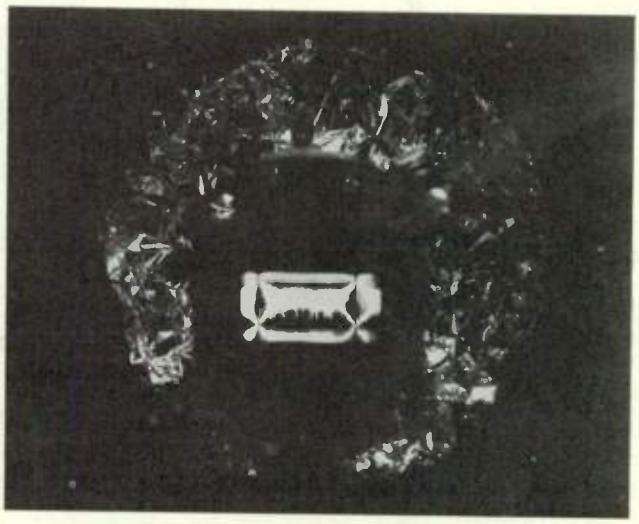


Fig. 7. Inorganic growth in an inconel waveguide.

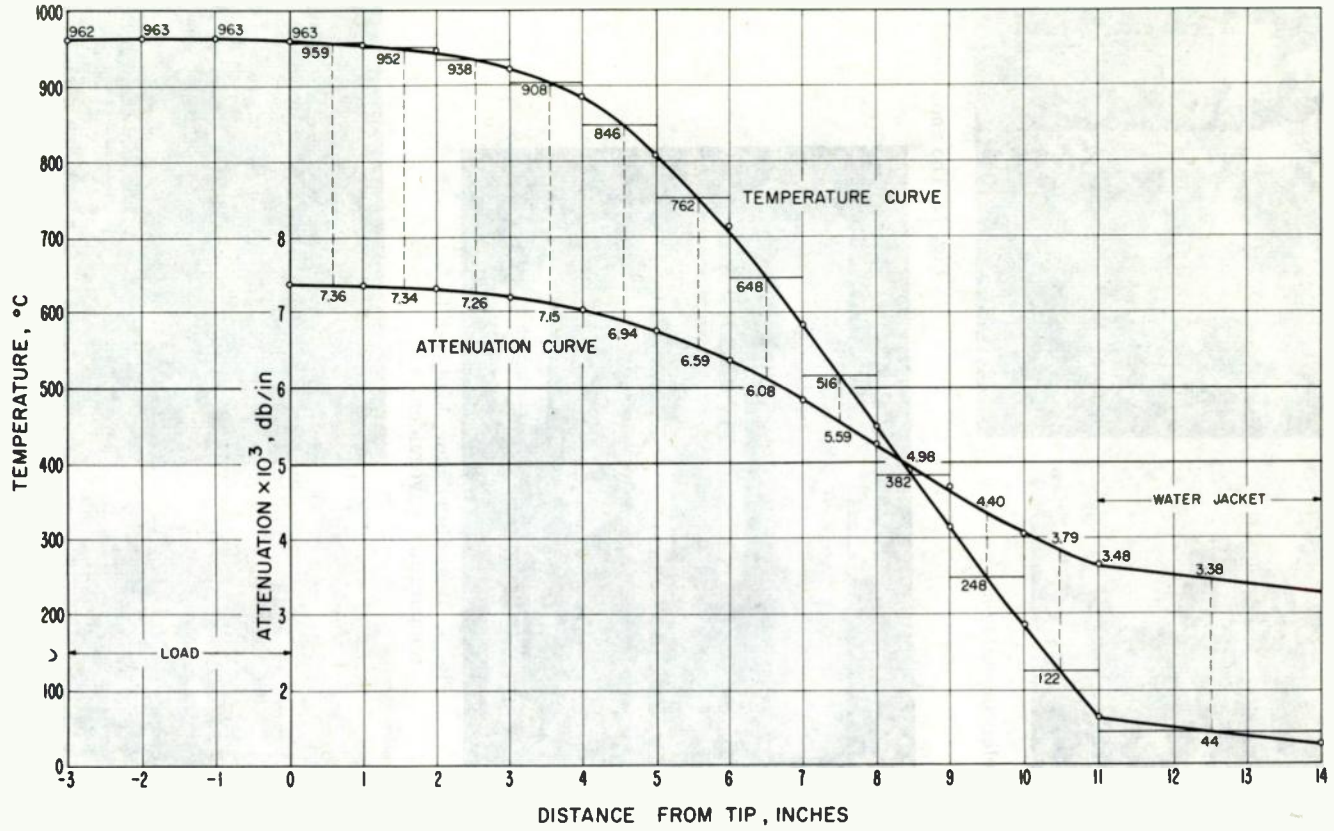


Fig. 8. Typical waveguide temperature distribution.

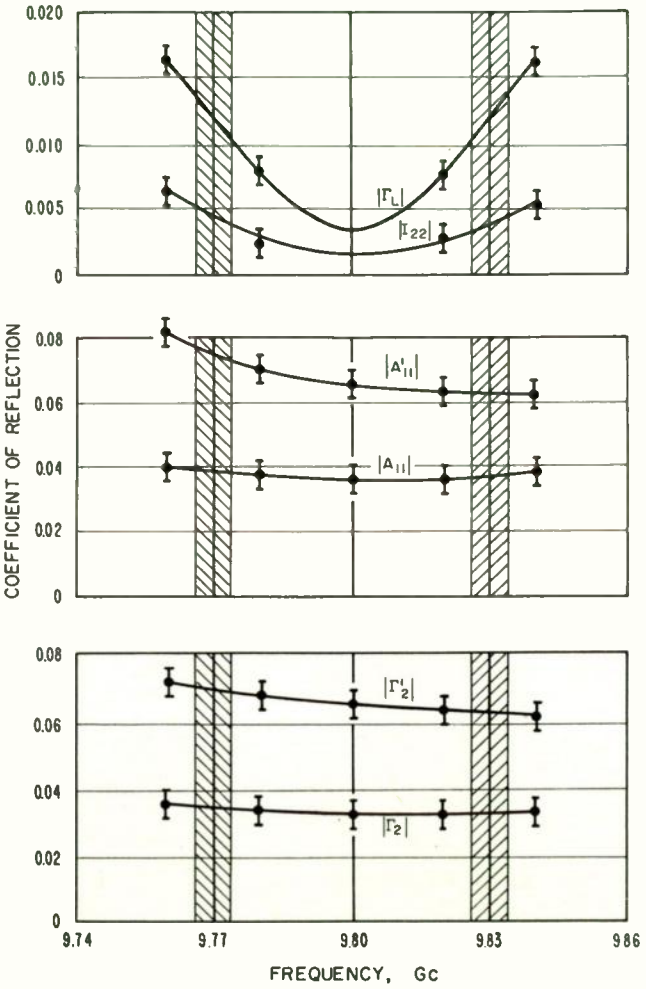


Fig. 9. Typical curves of reflection coefficient as a function of frequency.

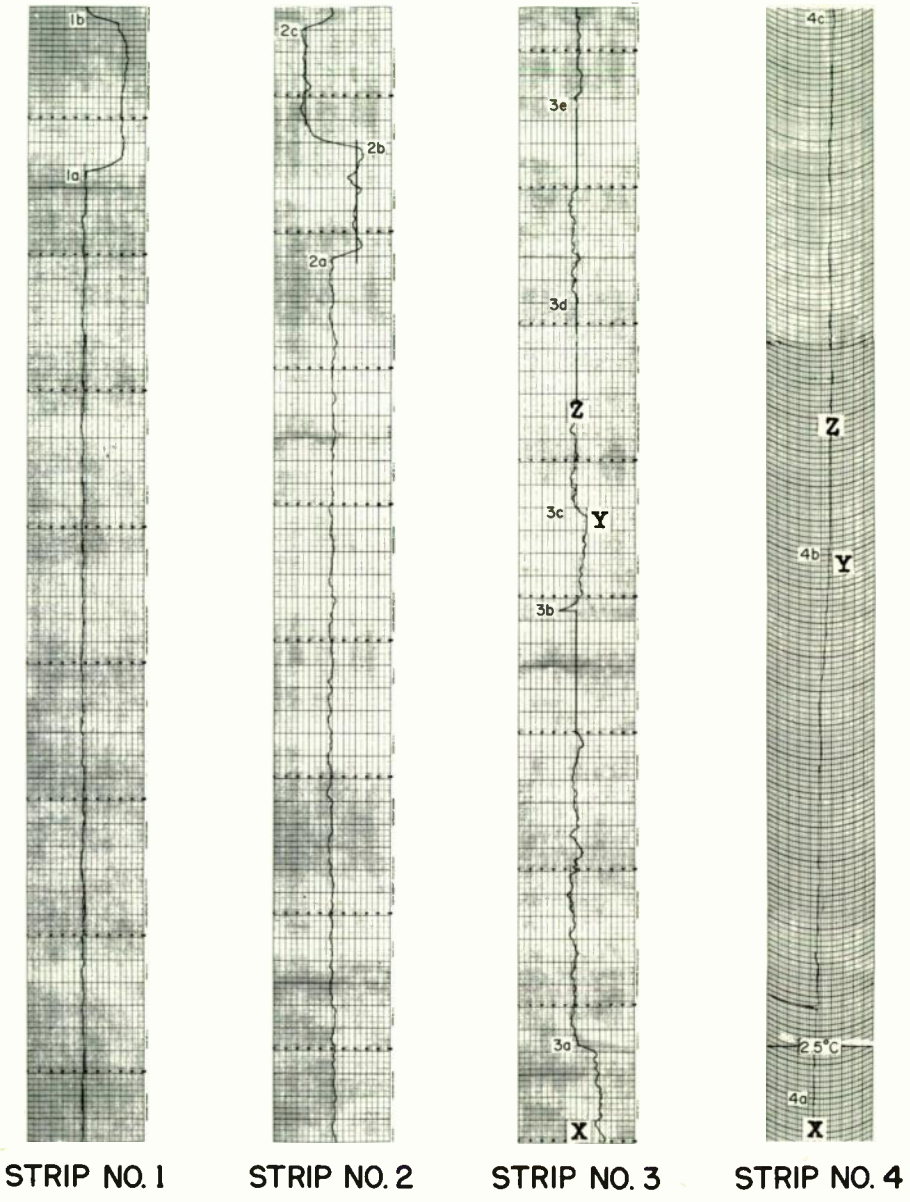


Fig. 10. Records of overall system performance.

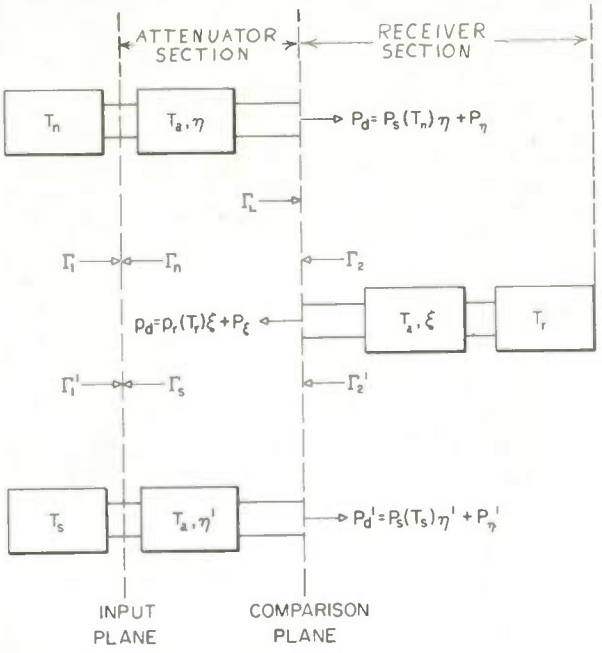


Fig. 11. Schematic showing planes for derivation of balance equation.

A 1962 REVIEW OF MILLIMETER WAVE INSTRUMENTATION

L. L. Bertan
FXR Division
Amphenol-Borg Electronics Corporation
Woodside 77, New York

Summary The increasing number of applications for millimeter and submillimeter waves (frequencies above 30 Gc) within the past few years makes desirable a review at this time of the current status of devices and components for this frequency range. Recent advances in both conventional and newer types of signal sources are reviewed. The advantages and limitations of several alternative transmission systems are discussed. Examples are shown of the development of microwave components, with emphasis placed upon devices which are particularly applicable to the shorter wavelengths. A number of areas are presented from which future developments may be expected.

INTRODUCTION

The demands for additional portions of the electromagnetic spectrum has recently occasioned a heightened interest in developments in the millimeter and submillimeter wavelength bands. Progress in these areas, although perhaps not quite as dramatic as that in the infrared and optical regions, has nevertheless resulted in a considerable increase in and refinement of techniques, devices and components within the past few years.

For the purposes of this paper, millimeter waves may be defined as wavelengths from 10 millimeters to 1 millimeter (frequencies from 30 Gc to 300 Gc). Submillimeter waves will be defined as wavelengths shorter than 1 millimeter (frequencies above 300 Gc). It is pertinent at this point to enumerate several of the more important applications for these frequencies.

Much of the earlier development work on millimeter and submillimeter wave instrumentation was performed by researchers in the field of microwave gas spectroscopy, where this instrumentation is used to measure the absorption frequencies due to molecular rotational transitions.[1] These wavelengths are also utilized for research on solid-state materials and for the analysis of energy gaps in superconductors. It is of interest to note that many of the more spectacular recent advances in microwave technology can be traced back to the painstaking efforts of the early gas and solid-state spectroscopists.

A major reason for the increased interest in millimeter waves that began about a decade ago was the requirement for additional channels of communications. At these wavelengths, bandwidths of tens of gigacycles, or many times the entire presently used microwave spectrum, could be readily made available. Accordingly, techniques and components for this application are being developed to permit its installation as future requirements dictate.[2]

Space communications, radar, and telemetry are fields particularly applicable to millimeter waves. Here the lower atmosphere absorption of these frequencies can be ignored, and advantage can be taken of the narrow antenna beams and high accuracy readily attainable with reasonable size components at the short wavelengths.

Millimeter waves have been used for microwave diagnostics of the highly ionized plasma produced in controlled fusion research. The propagation characteristics of the plasma at millimeter wave frequencies can be related to such plasma properties as temperature, electron density, and confinement time.[3] Microwave measurements have the advantage of providing the required plasma information without appreciably disturbing the reaction.

There are many additional applications that can be profitably exploited by millimeter and submillimeter waves as the necessary components and techniques become available. Among these is communication with a re-entering space vehicle, where millimeter waves may be necessary to penetrate the shock wave that accompanies the vehicle as it re-enters the upper atmosphere. Millimeter and submillimeter waves are required as high frequency pumps for parametric amplifiers and masers. Valuable scientific information can be obtained from millimeter wave radiometry and radio astronomy.

It is the intent of this paper to summarize the devices, components, and techniques currently available for utilization in the millimeter portion of the spectrum, and to point out some areas that are being explored for the required future developments in the submillimeter range.

SIGNAL SOURCES

Millimeter Wave Tubes

Traditionally one of the major difficulties confronting those working at the shorter wavelengths has been the lack of suitable self-excited coherent sources of power. Until fairly recently, tubes could be obtained only up to frequencies of approximately 75 Gc, or a wavelength of four millimeters. The small physical size of the tube structures and the resultant high current density requirements effectively set a practical upper frequency limit for the common types of microwave tubes. However, improved fabrication procedures and the application of techniques learned in the development of super-power tubes for lower frequencies have considerably extended this limit within the past few years.

Figure 1 shows graphically many of the CW and pulsed millimeter wave tubes that have been successfully built and tested.[4] Many of these sources are presently commercially available.

Among the more striking recent disclosures are the backward-wave oscillators currently being developed by CSF in France.[5] This group has fabricated a tube with a bandwidth of 10% and an output power of up to 15 milliwatts in the 0.7 millimeter wave region. They are currently working on oscillators for the 0.5 and 0.3 millimeter wave bands, and they estimate that they will shortly provide 1 watt of CW power at 1 millimeter wavelength. The slow wave structures of these tubes have dimensions of the order of small fractions of a millimeter and tolerances of the order of microns. Figure 2 is a plot of the performance of the experimental 0.7 millimeter tube, while Figure 3 is a picture of a water-cooled 2 millimeter tube in its permanent magnet.

Submillimeter Wave Devices

Despite the millimeter wave tube developments summarized above, and the spectacular advances recently announced in the field of optical masers where watts of coherent CW power and megawatts of pulsed power have been obtained, a frequency gap of several orders of magnitude still exists for sources between the millimeter and optical regions of the spectrum. Among the less conventional approaches to submillimeter wave generation are the Cerenkov Radiator proposed by Professor Coleman of the University of Illinois,[6] and the Tornadotron proposed by Dr. Weibel of the General Telephone and Electronics Laboratories.[7]

The Cerenkov Radiator When a charge particle moves thru a medium with a speed greater than the phase velocity of light in the medium, it sets up an electromagnetic field which opposes its motion. The work the charge does on this field appears as electromagnetic energy, i.e., the charge radiates Cerenkov power. Since the frequency spectrum of the radiation emitted by a single charge is not coherent, a bunched beam containing harmonics of some lower, obtainable microwave frequency must be used in order to obtain coherent Cerenkov radiation.

A scheme for realizing a Cerenkov radiator is shown in Figure 4. The beam is passed thru a hole in medium 1 which is surrounded by a nondispersive, isotropic dielectric medium 2. The Cerenkov radiation emerging at the angle θ from medium 1 is internally reflected at A on the surface of the exterior cone, travels to point B where it strikes at the Brewster angle, and then emerges parallel to the beam axis.

Medium 1 in the Cerenkov radiator can be a scalar dielectric, a plasma, a ferrite, etc.[8] The efficiency of the system can be shown to be

$$\eta = \frac{a_n^2 R}{2 R_o}$$

where η = the efficiency of the Cerenkov system
a_n = a constant lying between 0 and 2, depending on how well the beam is bunched
R = the interaction resistance of the Cerenkov system
R_o = the DC resistance of the bunched beam, typically 2×10^6 ohms

It is thus seen that in order to couple an appreciable percentage of the power in the system, the Cerenkov interaction resistance R must be of the order of the beam resistance Ro.

The interaction resistance for scalar dielectrics varies between several hundred to several thousand ohms. However, when medium 1 is a ferrite or a plasma in an appropriate magnetic field, interaction resistances of the order of megohms can be obtained, and high efficiencies may be expected.

The Cerenkov radiator is in a sense a harmonic generator, in that it converts RF energy at a microwave frequency into RF energy at some high harmonic of this frequency. However, the high frequency microwave problems are alleviated by the use of nonresonant physical optics types of coupling structures. It is hoped that optimization of the bunched beam and the coupling

structure will result in a device producing watts of Cerenkov power in the submillimeter range.

The Tornadotron In the Tornadotron submillimeter waves are produced via the mechanism of radiation from accelerated charges. Energy conversion from microwave to submillimeter wavelengths is accomplished in a cycle of operations. First, electrons are injected and trapped inside a cylindrical chamber formed by a split anode ring as shown in Figure 5, and two end plates not shown in the diagram. The electrons are compressed in a radial direction by the magnetic field (B) parallel to the axis of the chamber, while axial confinement is achieved by electric fields. The cigar-shaped electron beam rotates around its axis as indicated by the short arrows. Outward spiraling of the beam is then accomplished by applying a microwave electric field at cyclotron frequency between the two sections of the split anode. In the following step the frequency of orbital motion, and the rotational kinetic energy associated with it, is increased by several orders of magnitude by applying a pulsed high magnetic field. Radiation of submillimeter wave power is then obtained directly from the rapidly swirling beam without the necessity of using interacting structures.

In experimental operation thus far this device has produced milliwatts of power with a magnetic field of 25,000 gauss at a wavelength at 4 millimeters. A new tube with a higher space charge is being prepared which, in conjunction with a magnetic field of 100,000 gauss, it is hoped will deliver 1 watt of power at a wavelength of 1 millimeter.

Additional Generation Schemes Other areas of generating millimeter and submillimeter wave power are being investigated. These include the moving mirror,[9] where an electromagnetic wave at an incident frequency (f_i) is reflected from an electron cloud or a plasma moving at high velocity in a direction opposite to the direction of propagation of the incident wave. The reflected wave has a frequency (f_r) that can be determined from the relation.

$$f_r = f_i \frac{(1+\beta)}{(1-\beta)}$$

where $\beta = v/c$
v = velocity of mirror
c = velocity of propagation

The practical problem involved here is the achievement of a good reflecting surface moving with sufficient velocity.

Harmonic generation in non-linear resistors, non-linear reactors, ferrites, plasmas, and gas discharges, has been used as a source of millimeter and submillimeter wave power.[10,11,12,13,14] This was, until recently, the only means of obtaining the shorter millimeter and submillimeter wavelengths. It is still a convenient and comparatively inexpensive source for those applications that can afford the resulting 10-20 db or more of conversion loss.

The use of multiple quantum transitions for harmonic generation has been suggested.[15] If a two-energy-level system is subjected to sufficiently strong RF pumping at a frequency which is close to a submultiple of the natural transition frequency, an interaction occurs between the system and the RF input. If the system can absorb power at the pump frequency and emit it back at the natural transition frequency, such an effect would constitute a procedure for high-power high-efficiency harmonic generation, and would be applicable at very short wavelengths.

Recent developments in optical masers suggest the possibility of mixing two different optical wavelengths in an appropriate semiconductor to provide a difference frequency in the submillimeter region. There is also the intriguing possibility that the tunnel diode may prove to be a practical means of converting DC to millimeter wave energy.[16]

It is certainly possible that one or more of the schemes mentioned above will prove to be a practical generator of millimeter and submillimeter energy. It is at least equally possible that some other device, perhaps as yet unthought of, will emerge as the dominant source of submillimeter wave power. It seems fairly safe at this time to assume that if one would expect to generate submillimeter wave power of the level now being generated at microwave and optical wavelengths, reasonably efficient high energy devices would be required with internal dimensions quite large as compared to the operating wavelength.

TRANSMISSION LINES

Millimeter Waveguides

The two types of waveguide transmission that have heretofore been generally used for the guiding of millimeter wave energy are the TE_{10} rectangular waveguide mode and the TE_{01} circular waveguide mode. These modes are transmitted in fully enclosed metallic waveguides.

The adaptability of fundamental mode TE_{10} rectangular waveguide to the transmission of microwave energy and to the design and fabrication of most of the common microwave components has resulted in the choice of this waveguide for the great majority of microwave applications. However, as the shorter millimeter wave frequencies are reached, attenuation in this mode becomes quite high, and its use must be restricted to those components that can be kept to fairly short lengths of transmission line.

The TE_{01} circular waveguide mode has the desirable characteristic of continuously decreasing loss as frequency increases for a given waveguide inside diameter. However, when those higher frequencies are reached where the TE_{01} mode attenuation becomes low, it will be found that many other waveguide modes can also propagate. If these undesired modes are excited, they will contribute attenuation and distortion to the microwave signal. Thus, great care must be taken to insure both the launching and maintaining of a pure TE_{01} mode in a circular waveguide system.[17]

A number of non-enclosed waveguides have been considered for the transmission of millimeter and submillimeter energy. These include the H-type waveguide[18] which like the TE_{01} circular metallic guide has the property of continuously decreasing loss with increasing frequency, the dielectric rod waveguide, the dielectric image guide which consists of half of a dielectric rod waveguide mounted on a conducting surface, and the dielectric coated wire.[19] All of these transmission lines have the capability of lower transmission losses than fundamental mode rectangular waveguide. They have the disadvantage, however, of tending to radiate at bends, supports, and other discontinuities.

Beam Waveguide

The sizes and tolerances required in any of the single mode waveguide transmission systems limits their use to the millimeter and longer submillimeter wavelengths. As the shorter submillimeter wavelengths are approached, it becomes necessary to turn to optical or quasi-optical transmission schemes. A promising transmission line for these shorter wavelengths is the recently disclosed beam waveguide.[20]

The beam waveguide differs from the usual waveguide in that the energy is transmitted in the form of a wave beam rather than a wave mode. A wave mode is characterized by a propagation constant and field pattern which do not vary down the length of the guide. The field in a beam waveguide consists of a bundle of waves characterized by a spectrum of propagation constants and a field distribution that varies along the guide, but is periodically repeated. The beam wave is guided by passing through phase transformers which reset the cross-sectional phase distribution in the beam at predetermined intervals.

Figure 6 is a schematic drawing of the beam waveguide, including the launching and guiding structures. The phase transformers consist of dielectric lenses which advance the phase of the outer portion of the beam relative to the center in accordance with the relation

$$\varphi = \frac{k\rho^2}{D}$$

where φ = phase advance at radius ρ
D = spacing of the phase transformers
$k = 2\pi/\lambda$

The maximum phase shift (at the outer radius R of the phase transformer) is related to the diffraction loss of the phase transformer. This loss, caused by the cross-sectional limitation of the beam, can be held to less than 0.01 db per phase transformer when the maximum phase shift approaches 2π. Experimental data on a beam waveguide at a frequency of 35 Gc shows an iteration loss (total of diffraction, reflection, and absorption losses) of approximately 0.03 db per transformer, or about 1.2 db per kilometer. A photograph of a beam waveguide, as set up at the Ultramicrowave Section of the University of Illinois, is shown in Figure 7.

COMPONENTS

Millimeter Wave Components

In view of the recent interest in millimeter waves, most of the standard microwave components have recently been extended to cover the rectangular waveguide sizes down to wavelengths of 1.5 to 2 millimeters. Because of the high attenuation and close tolerances inherent at these wavelengths, careful fabrication techniques, including precision milled block and electroformed construction, are generally used. Among the novel ideas that have been applied to the shorter wavelength components are those incorporated in the design of two different slotted sections, or impedance meters, that will be described.

The dielectric vane slotted section,[12] shown schematically in Figure 8, consists of a fixed

position coupling probe followed by a calibrated adjustable phase shifter. The phase shifter is a thin eccentric mounted mica disc, the insertion of which can be varied within the primary waveguide. Thus, the phase of the standing wave produced by a load under test can be moved past the fixed position probe, which couples a portion of this wave into this secondary waveguide. The standing wave pattern coupled into the secondary waveguide is similar to that provided by the moving probe of a conventional slotted section.

The variable impedance meter [21] measures the modulus and argument of the reflection coefficient by using the principle of the rotary attenuator.[22] Referring to the schematic representation of Figure 9, the variable impedance is used in conjunction with a hybrid tee to balance the reflection of a load under test. This balance is determined from a null indicator in the fourth arm of the hybrid.

The impedance meter consists of a transition from rectangular to circular waveguide and a section of circular waveguide in which are contained two thin attenuating vanes and a reflecting plunger that can be both rotated and axially displaced. The input attenuating vane is fixed perpendicular to the electric field of the incident wave. The second attenuating vane is fixed at its far end to the reflecting plunger, and thus rotates and moves axially with the plunger. The modulus of the reflection coefficient of the impedance meter can be shown to be equal to the cosine squared of the angle between the two attenuating vanes, while the argument of the reflection coefficient can be shown to be proportional to the axial movement of the plunger.

The design of ferrite devices has proceeded well into the millimeter wave region, and a Y circulator has been constructed with good performance at 140 Gc.[23] The development of permanent magnet type ferrite materials with very high anisotropic fields suggest that broadband resonance isolators requiring little or no externally applied magnetic fields can be developed for a substantial portion of the millimeter wave spectrum.[24] The ferrite phase-shifter[25] and the YIG resonator[26] have also been extended to operate at millimeter wave frequencies.

Quasi-Optical Components The applicability of optical techniques at microwave frequencies has long been realized by the antenna and propagation engineer. As the requirements arise for components in the submillimeter wave region, the size limitations of conventional microwave devices will certainly require their replacement, in the majority of cases, by quasi-optical components. Many optical components can be more readily realized at millimeter than at optical frequencies. The Fabry-Perot interferometer shows promise of being useful not only as a high Q wavemeter, but as a resonator in millimeter wave generation schemes.[27]

An illustration of the combination of microwave and optical techniques at millimeter wave frequencies can be seen in Figure 10. This shows a Michelson interferometer[28] as set up for the measurement of relative dielectric constant at the General Telephone and Electronics Laboratories. At the lower left corner can be seen a BWO signal source, voltage-tunable over the frequency range 50-80 Gc. The output of the BWO feeds thru a horn antenna into a polyethylene lens, which directs the output into a Teflon split prism. The spacing of the two halves of the prism is adjusted to give a 3 db power split between the two arms in the upper left and upper right corners of the illustration. The outside surfaces of the Teflon prism are coated with a $\lambda/4$ thick material of relative dielectric constant equal to the square-root of that of Teflon, thus providing a match at the dielectric-air interface. The energy in the upper left arm passes thru a frame holder for the dielectric under test, and is then reflected back into the prism by the sliding mirror. A reflecting mirror can also be seen in the upper right corner. The equipment behind the upper right mirror is a Golay cell which is used for power detection when the mirror is removed. The interference measurement is made in the detecting element following the focusing lens and horn antenna in the lower right corner.

CONCLUSIONS

Standard microwave devices and components, or slight modifications thereof, have recently been extended to cover the millimeter wavelength frequency range up to 300 Gc. It is possible that these techniques can be extended somewhat into the longer submillimeter wavelengths. However, at some point it becomes necessary to concede that these microwave techniques cannot be indefinitely scaled down in size until the optical wavelengths are reached.

Considerable progress has been made in the application of optical and quasi-optical techniques to millimeter and submillimeter wave transmission systems and components. However, the major obstacle to the exploitation of the submillimeter wave band has been the inability up to now of both classical and quantum

electronics techniques to produce an adequate coherent source of submillimeter wave energy. It is fervently hoped that this situation will shortly be corrected.

ACKNOWLEDGMENT

The assistance of the many individuals and organizations that provided the information contained in this paper is gratefully acknowledged. In particular, thanks are due to Professor P. D. Coleman of the University of Illinois, Mr. M de Thomasson of the American Radio Company (CSF), Dr. G. Wade of the Raytheon Company, and Dr. G. Weibel of the General Telephone and Electronics Laboratories.

REFERENCES

1. W. Gordy, W. F. Smith, and R. F. Trambarulo, Microwave Spectroscopy, John Wiley and Sons, New York, 1953.

2. S. E. Miller, "Millimeter Waves in Communication," MRI Symposia Series, vol. 9, pp. 25-44; April 1959.

3. M. S. Heald, "Microwave Measurements in Controlled Fusion Research," IRE Conv. Rec., part 9, pp. 14-18; March 1958.

4. G. Wade, "A Survey of Research in Millimeter Wave Generation," Paper presented at AIEE Winter General Meeting, January 1962.

5. G. Convert, "Limitations to the Generation of Submillimeter Waves by Electron Devices," Paper presented at AIEE Winter General Meeting, January 1962.

6. P. D. Coleman, "Cerenkov Radiation Approach to the Submillimeter Wave Problem," Advances in Quantum Electronics edited by J. Singer, Columbia University Press, New York, pp. 581-594; 1961.

7. G. E. Weibel, "High Magnetic Field Submillimeter Wave Generators with Parametric Excitation," MRI Symposia Series, vol. 8, pp. 389-406; April 1958.

8. P. D. Coleman, Private Communication.

9. P. D. Coleman, "Electron Devices for the Millimeter-Infrared Gap," IRE Conv. Rec., part 3, pp. 54-63; March 1961.

10. I. Kaufman, "The Band Between Microwave and Infrared Regions," Proc. IRE, vol. 47, pp. 381-396; March 1959.

11. P. D. Coleman and R. C. Becker, "Present State of the Millimeter Wave Generation and Technique Art," IRE Trans., vol. MTT-7, pp. 42-61; January 1959.

12. L. L. Bertan, "Considerations in the Design of Two Millimeter Wave Components," Proc. Fourth Annual Joint Military - Industrial Electronic Test Equipment Symposium, pp. 113-130; September 1960.

13. C. B. Swan, "Generation of Microwave Harmonics in an Electrodeless Discharge at Low Pressure," Proc. IRE, vol. 49, pp. 1941-1942; December 1961.

14. J. R. Baird and P. D. Coleman, "Frequency Conversion in a Microwave Discharge," Proc. IRE, vol. 49, pp. 1890-1900; December 1961.

15. J. Fontana, R. Pantell, and R. Smith, "Harmonic Generation by Means of Multiple Quantum Transitions," Advances in Quantum Electronics edited by J. Singer, Columbia University Press, New York, pp. 612-618; 1961.

16. C. A. Burrus, "Millimeter Wave Esaki Diode Oscillators," Proc. IRE, vol. 48, pg. 2024; December 1960.

17. A. P. King, "The Observed 50-90 KMC Attenuation of Two Inch Improved Waveguide," IRE Wescon Conv. Rec., part 1, pp. 28-33; August 1960.

18. F. J. Tischer, "Properties of the H-Guide at Microwaves and Millimeter Waves," IRE Wescon Conv. Rec., part 1, pp. 4-12; August 1958.

19. F. Sobel, F. L. Wentworth, and J. C. Wiltse, "Quasi-Optical Surface Waveguide and Other Components for the 100 to 300 Gc Region," IRE Trans., vol. MTT-9, pp. 512-518; November 1961.

20. J. R. Christian and G. Goubau, "Experimental Studies on a Beam Waveguide for Millimeter Waves," IRE Trans., vol. AP-9, pp. 256-263; May 1961.

21. C. W. van Es, M. Gevers, and F. C. deRonde, "Waveguide Equipment for 2 mm Microwaves," Philips Technical Review, vol. 22, pp. 113-125; February 1961.

22. G. C. Southworth, Principles and Applications of Waveguide Transmission, D. Van Nostrand Company, Princeton, New Jersey, pg. 374; 1950

23. J. B. Thaxter and G. S. Heller, "Circulators at 70 and 140 KMC," Proc. IRE, vol. 48, pp. 110-111; January 1960

24. D. J. DeBitetto, F. K. du Pré, and F. G. Brockman, "Highly Anisotropic Magnetic Materials for Millimeter Wave Applications," MRI Symposia Series, vol. 9, pp. 95-108; April 1959.

25. R. S. McCarter and E. F. Landry, "Ka-Band Ferrite Phase Shifter," IRE Trans., vol. MTT-9 pg. 271; May 1961

26. D. Douthett and I. Kaufman, "The Unloaded Q of a YIG Resonator from X-Band to 4 Millimeters," IRE Trans., vol. MTT-9, pp. 261-262; May 1961.

27. W. Culshaw, "Resonators for Millimeter and Submillimeter Wavelengths," IRE Trans., vol. MTT-9, pp. 135-144; March 1961.

28. L. Page, Introduction to Theoretical Physics, Third Edition, D. Van Nostrand Company, Princeton, New Jersey, pg. 640; 1952.

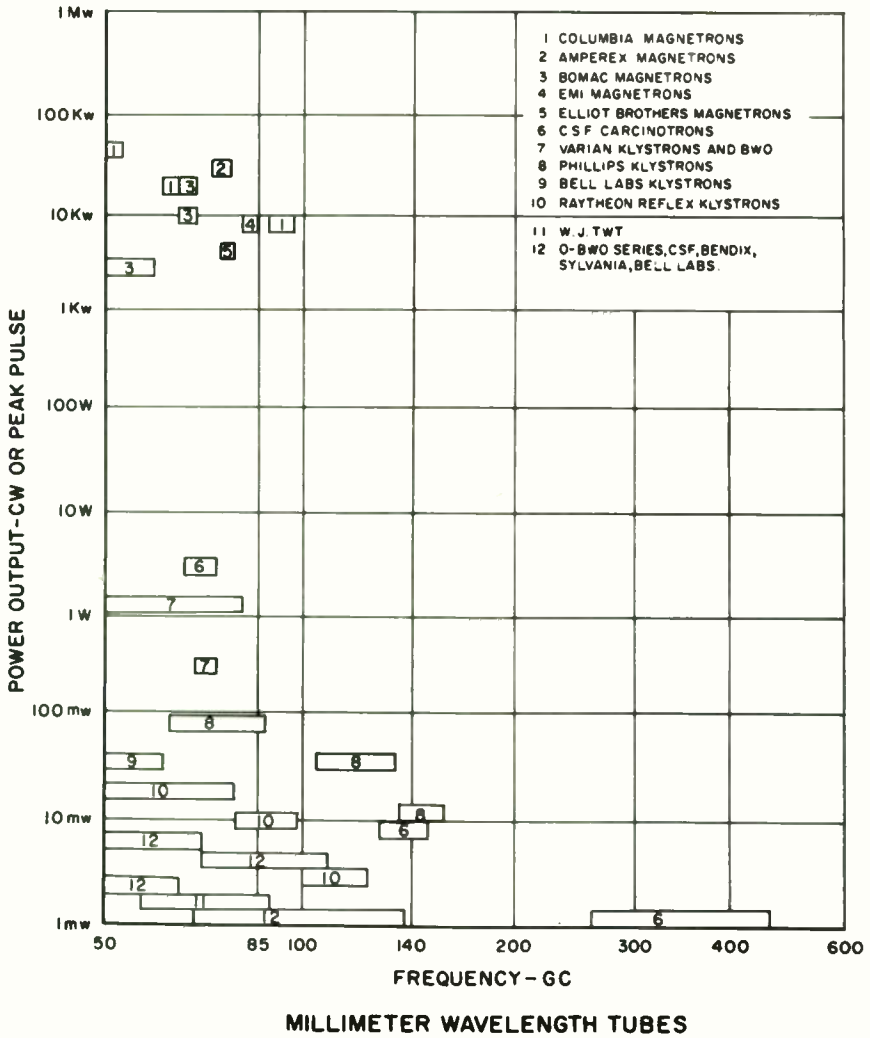
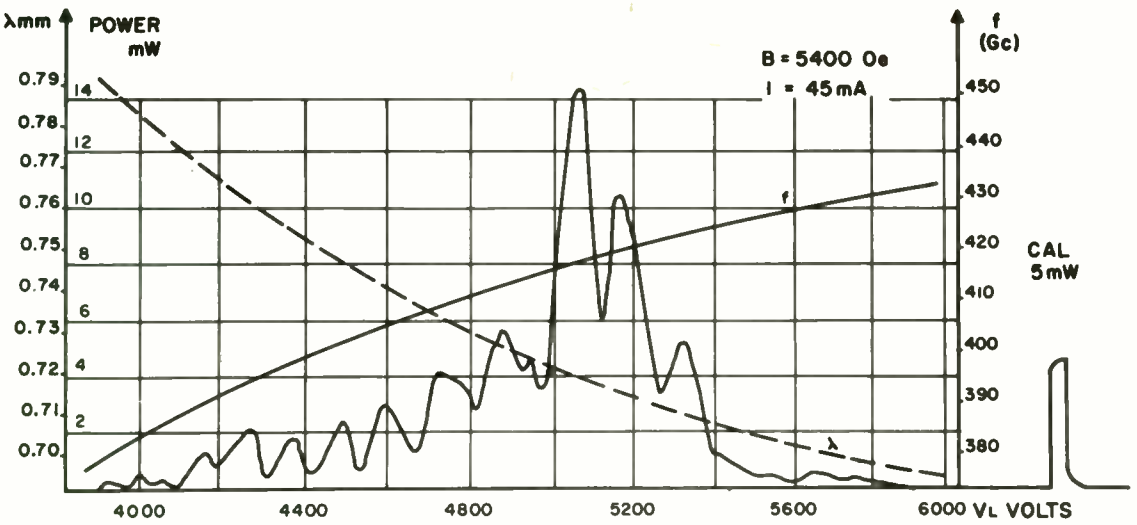


Fig. 1. Millimeter wavelength CW and pulsed tubes.



CO. 0.7mm EXPERIMENTAL BWO

Fig. 2. Performance of the CSF 0.7 millimeter experimental BWO.

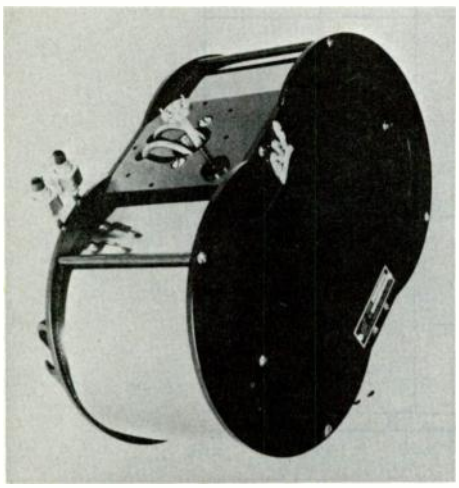


Fig. 3. A 2 millimeter BWO in its permanent magnet.

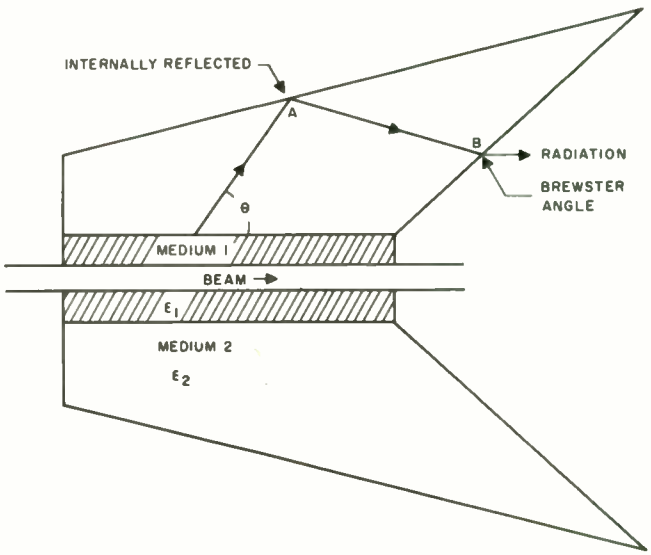


Fig. 4. Schematic representation of the Cerenkov Radiator.

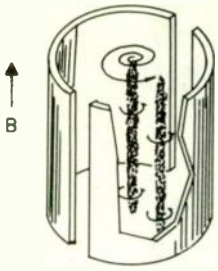


Fig. 5. Motion of electron cloud in the Tornadotron.

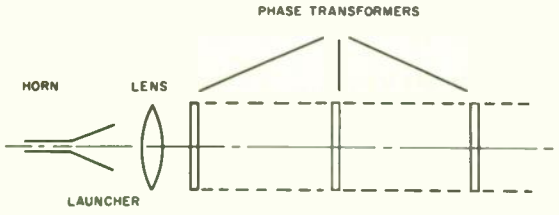


Fig. 6. Schematic representation of the beam waveguide.

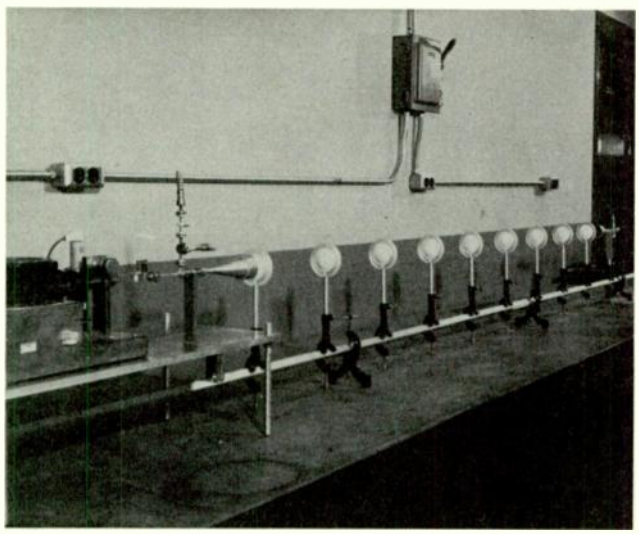


Fig. 7. A beam waveguide transmission line.

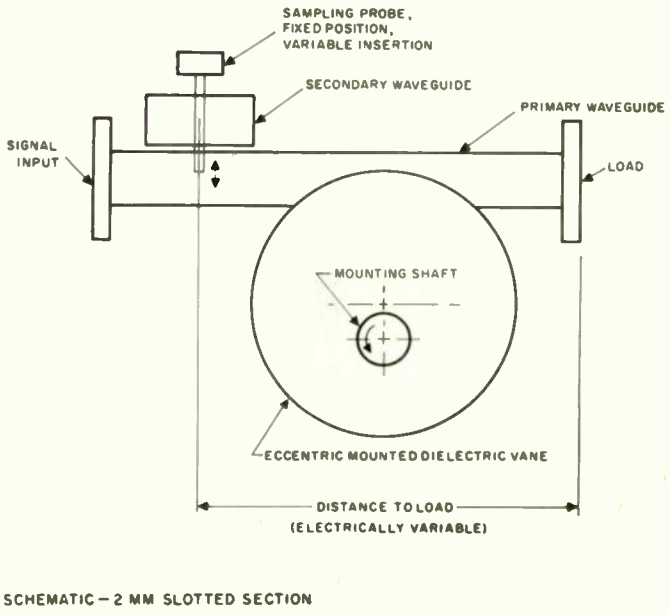


Fig. 8. Schematic representation of the dielectric vane slotted section.

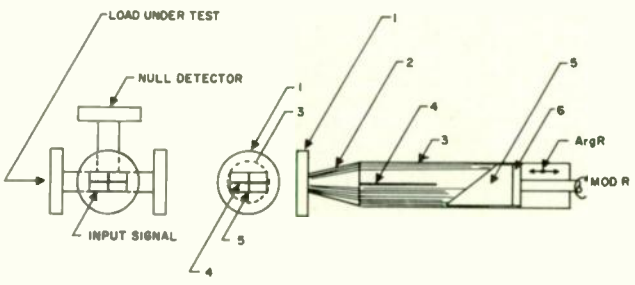


Fig. 9. The variable impedance meter, consisting of a hybrid T and a variable impedance shown as a section and end view. 1) Flange with rectangular opening. 2) Transition from rectangular to circular cross-section. 3) Circular waveguide. 4) Fixed vane. 5) Vane capable of being rotated and axially displaced, fixed to plunger. 6) When 5 is rotated, the modulus of the complex reflection coefficient is changed; when 5 is displaced axially, the argument is changed.

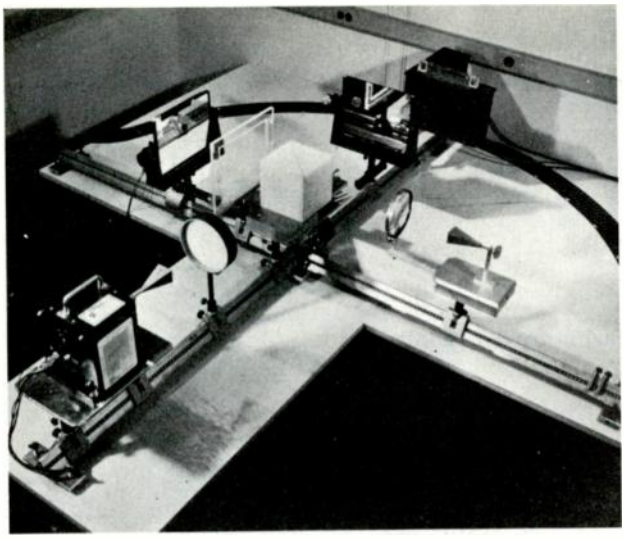


Fig. 10. A millimeter wave Michelson interferometer.

LOW LOSS X-BAND TRAVELING RESONATOR

E. Wantuch and A. H. Reeves
Airtron Division of Litton Industries
Morris Plains, New Jersey

Summary

The use of the resonant ring to simulate higher power levels than available from the power source is well known. At lower microwave frequencies substantial power multiplication may be obtained with this technique. At X-band the use of a reasonably sized ring becomes a limiting factor due to the loss of the waveguide making up the ring.

This limitation has been overcome by the use of square waveguide operating in the TE_{01} mode. This scheme reduces ring losses by almost one-half. The use of this configuration requires the elimination of the cross-polarized TE, as well as the lowest TM modes. Power multiplication factors up to 40 to 1 have been measured.

The problem of temperature variation can cause serious detuning of this ring due to dimensional changes. A second version of this equipment was therefore designed using silver laminated invar tubing.

Several design features such as 180° bends, mode launchers, and other pertinent details are described.

* * * * * *

The use of the traveling wave resonator to simulate high microwave power levels is well known. At the lower microwave frequencies the multiplication factor obtainable is generally limited by the insertion loss of the device to be tested and losses of the ring can be neglected. At the higher microwave frequencies such as X-band, this assumption is no longer justified. A rather complete review of this situation has recently been made by S. Miller.[1]

The traveling wave resonance is shown schematically in Figure 1. It consists of a variable coupler which has been designed utilizing two directional couplers with a variable phase-shifter inserted in one line. In this manner the overall coupling may be varied from approximately 6 db tighter than one of the fixed couplers to any arbitrarily low value limited only by component imperfections. A similar coupler may be designed using only 3 db hybrids. The latter arrangement has the disadvantage that phase-shifter losses affect ring losses to a greater extent than the use of looser couplers would entail. Moreover, this device is only useful, if it does provide appreciable multiplication, and therefore extremely tight coupling to the ring is not necessary.

In the present design the fixed coupler had coupling values of 13 db. These couplers also served as mode transducers from rectangular to square waveguide. The coupling apertures consist of eight slots of full rectangular waveguide height. The wall left between adjacent slots was adjusted until the desired coupling is obtained. This arrangement is indicated in Figure 2.

Figure 3 summarizes typical attenuation values for different transmission lines operating at 9.375 KMc/s. These figures are theoretical figures with actual losses generally running twenty-five per cent higher than these figures.

Figure 4 shows the theoretical curve for power multiplication as a function of ring loss. In order to obtain a power multiplication of 40 (16 db) the total ring insertion loss must be held to 0.1 db. If the ring is to be kept to reasonable physical size such as six feet, the use of rectangular waveguide is clearly ruled out. Circular waveguide operating in the TE_{01}^{o} mode should provide an optimum configuration.

Considerable work has been done on microwave components utilizing the circular electric mode. Literature on the subject generally dis-

cusses the extremely low attenuation figures in runs of several hundred feet, but neglects to mention the power loss at the mode coupler due to unwanted mode conversion. This energy is generally dissipated in a selective mode absorber.

Our initial experiments were therefore attempted to realize a low-loss mode exciter consisting of a four-way power divider with each of the four waveguides fed to a two-way power divider to obtain a directional injection scheme utilizing eight coupling apertures into the circular waveguide. Despite this elaborate approach, no satisfactory low-loss solution to the coupling problem could be obtained. This approach was therefore abandoned in favor of the next most attractive scheme which is the use of square waveguide.

Again, referring to the previous figure, a substantially lower attenuation may be obtained in square waveguide as compared to rectangular waveguide of the same width. The use of square waveguide would require the elimination of two undesired modes, namely, the cross-polarized TE mode as well as the first TM mode. Fortunately, it is found that by preserving strict symmetry in the ring couplers, these two modes were not set up. Such symmetry would have to be violated in bends necessary to close the ring. The problem of making waveguide bends in square waveguide was avoided by the use of 3 db sidewall hybrids as 180° bends. These hybrids could be scaled from corresponding designs in rectangular waveguide with only a readjustment of the capacitive loading in the coupling region.

Particular attention must be paid to the flange design to avoid appreciable losses. A configuration shown in Figure 5 was evolved and proved quite satisfactory. It consists of a raised portion immediately surrounding the waveguide opening. Next, a groove is milled which contains a suitable pressure gasket. The remaining flange face, except for a raised portion at the outer periphery of the flange, is machined to be 0.005 to 0.008" below the inner and outer raised portions. The screws joining these flanges are located in this relieved region. Tightening of this connection will result in a distortion of the flange but will insure good mechanical contact at the inner and outer portions of the contact flange. As a further precaution, these two surfaces were lapped to a surface finish of better than 16μin.

Since a traveling wave resonator may be considered as a high Q cavity, the effect of changing ambient temperature must be considered as a factor in ring tuning. The final version of the traveling wave resonator was therefore constructed of silver laminated invar tubing. The waveguide had a wall thickness of 0.040" with a silver layer 0.005" lining the inside of the waveguide.

The traveling wave resonator showed a power gain of 16 db measured at 9.375 Mc/s. An input power up to 400 KW peak could be handled without pressurization with an r.f. pulse length of 1μsec.

Figure 6 shows the variable coupler assembly with the addition of the monitoring coupler on the return branch of the ring. Figure 7 shows the side-wall hybrid in square waveguide with its associated movable shorting plunger assembly.

Figure 8 shows the complete resonant ring.

Reference

1. Stanley J. Miller, "The Traveling Wave Resonator and High Power Microwave Testing," The Microwave Journal, Sept. 1960.

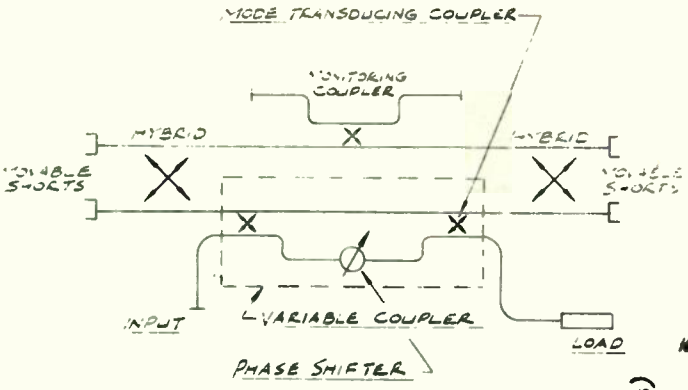


Fig. 1. Resonant ring schematic.

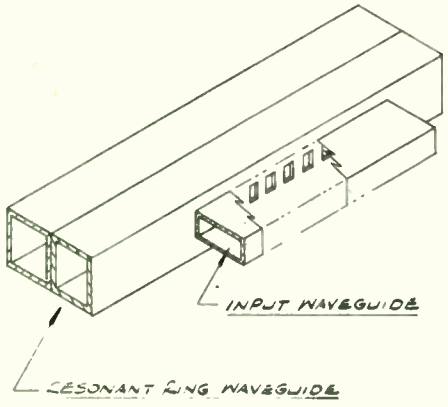


Fig. 2. Mode transducing coupler.

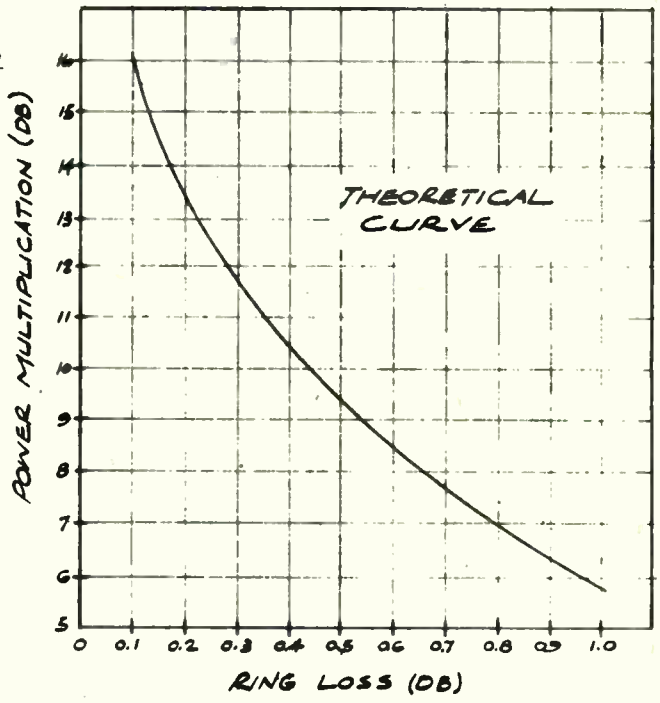


Fig. 4. Power multiplication as a function of ring loss.

WAVEGUIDE CROSS-SECTION	OPERATING MODE	LOSS PER FT. (Silver)
0.400 X 0.900"	$TE_{10}^{\square}$	0.031 db
0.497" X 1.122"	$TE_{10}^{\square}$	0.021 db
0.938" I.D.	TE_{11}^{O}	0.024 db
1.164" Square	$TE_{10}^{\square}$	0.011 db
2.50" I.D.	TE_{01}^{O}	0.0023 db

Fig. 3. Transmission line comparison at 9.375 kMc.

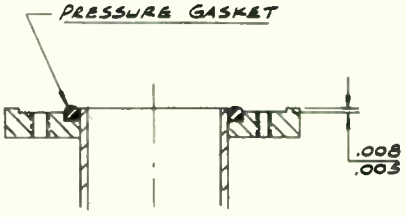


Fig. 5. Low loss flange design.

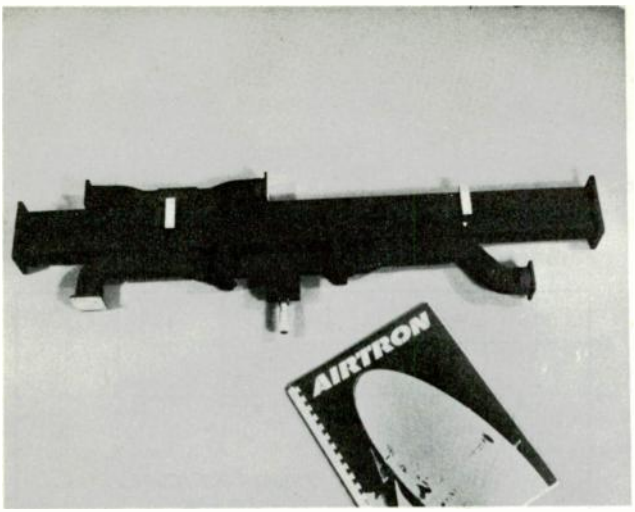


Fig. 6. Variable coupler assembly.

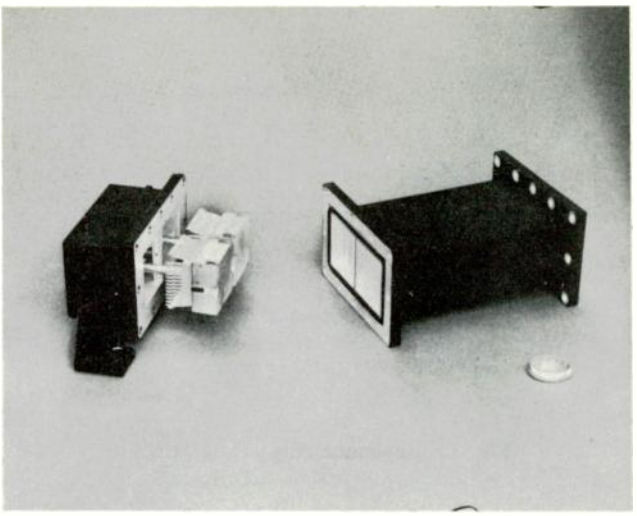


Fig. 7. Side wall hybrid.

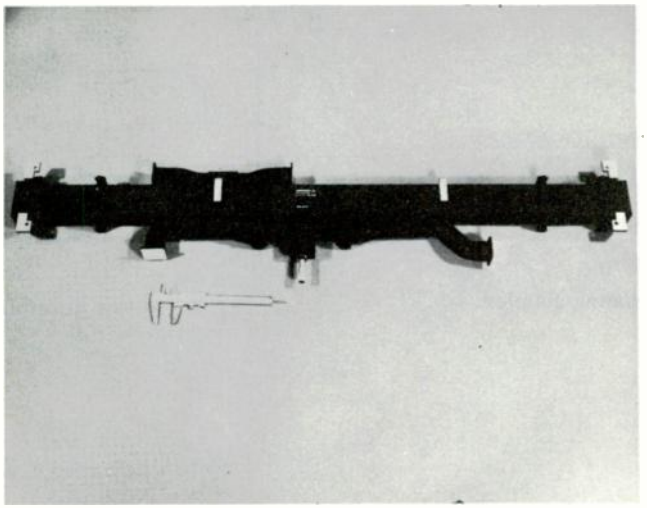


Fig. 8. Complete resonant ring.