Correspondence

Piezo-Junctions: Elements of a New Class of Semiconductor Devices*

Recently large effects of anisotropic stress on the resistance of *p-n* junctions were observed.[1] The magnitude of these effects is far larger than that previously reported.[2,3] Moreover, it was found that stress can cause considerable, and hitherto unreported, changes in the breakdown characteristics. The basic phenomena are illustrated in Fig. 1 where (a) the schematic stressing arrangement and (b) a family of isobaric *I-V* characteristics are shown. The radius R of the stressing element ordinarily used was of the order of 20 μ. The contact diameters d were of the order of 2 to 3 μ and local pressures obtained under reversible conditions were calculated[4] to be as high as 8×10^4 kg/cm^2. Due to the anisotropies of the stress field and the electrical field various mechanisms, which are presently being investigated, contribute to the effects described. Within the elastic range the effects are completely reversible and provide a basic element, for brevity referred to as the piezo-junction, of a new class of devices.

p-n junctions constitute basic building blocks of many semiconductor devices from diodes to tetrodes. In any such device the control of one junction will affect the performance of the whole device. For example, by stressing the emitter junction of a transistor the output across the collector junction can be controlled. Such an application is illustrated in Fig. 2 which shows a set of collector characteristics of a *p-n-p* Ge mesa transistor. Except for the indicated variations in stress on the emitter, all other experimental conditions remained identical and the changes of h_{fe} from about 40 to 5 were completely reversible. By proper selection of junction depth and the value of R of such a piezo-triode responses by more than an order of magnitude larger than those shown have been obtained. An interesting aspect of the drastic changes in transistor characteristics under controlled stress is the clear demonstration of the previously suspected[5] effects which uncontrolled built-in stresses may have on semiconductor devices.

Ge and Si piezo-diodes have been incorporated in small experimental microphone and hydrophone structures. Fig. 3 shows, for example, such a piezo-diode microphone

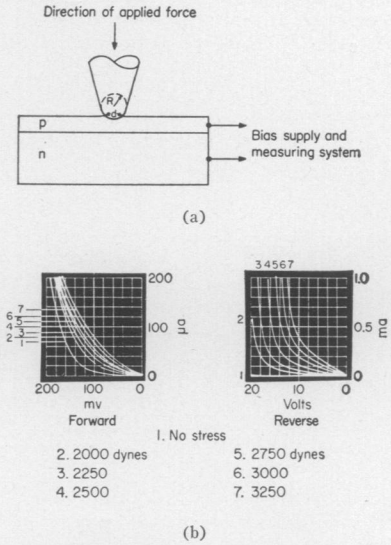


Fig. 1—(a) Schematic stressing arrangement. (b) Forward and reverse bias characteristics of a piezo-diode.

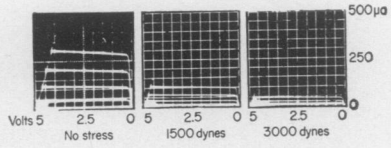


Fig. 2—h_{fe} of a piezo-triode with $I_e=2$ μa/step.

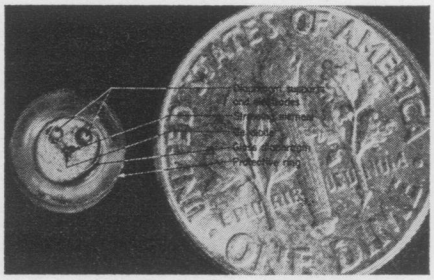


Fig. 3—An experimental piezo-diode microphone.

with the pertinent structural details. Outputs of up to about 2 v were obtained on some units in sound fields of about 60 db above 200 μ dynes. Preliminary measurements indicate that high sensitivities are possible with piezo-triodes since in the latter the inherent gain of the transistor is exploited. The frequency response has been found to extend from dc to over 100 kc, with various peaks which can be attributed to mechanical resonances.

The devices already realized in practice constitute only a fraction of the potential piezo-junction accelerometers, displacement transducers and other piezo-junction devices. In this connection some features of piezo-junctions are of particular interest. For example, the volume of the active semiconductor region extends over only a few cubic microns. Thus small dimensions and small masses and correspondingly high mechanical resonance frequencies are possible. The stress-dependent breakdown characteristics provide potential sensitivities limited mainly by noise and stability considerations.

The high sensitivity displayed suggests the possibility of coupling piezo-junctions with electromechanical transducers to provide novel four-terminal devices. This can, for example, be achieved by applying the signal input to electrostatic, magnetostrictive or piezoelectric elements in suitable mechanical contact with piezo-junctions which provide the output terminals.

ACKNOWLEDGMENT

The authors gratefully acknowledge the efficient assistance received from R. Flagg and J. Gould in the experimental work.

W. RINDNER
R. NELSON
Raytheon Res. Div.
Waltham, Mass.

* Received July 30, 1962.

[1] W. Rindner, "Anisotropic strain effects in *p-n* junctions," *Bull. Am. Phys. Soc.*, vol. 7, pp. 65–66; 1962.

[2] H. Hall, J. Bardeen, and G. Pearson, "The effects of pressure and temperature on the resistance of *p-n* junctions in Ge," *Phys. Rev.*, vol. 84, pp. 129–132; 1951.

[3] S. Miller, M. Nathan, and A. Smith, "Pressure dependence of the current-voltage characteristics of Esaki diodes," *Phys. Rev. Lett.*, vol. 4, pp. 60–62; 1960.

[4] W. Rindner and I. Braun, "On the Effects of Elastic and Plastic Deformation of *p-n* Junctions in Semiconductors," *Proc. Internat'l Conf. on the Physics of Semiconductors*, Exeter, England; July, 1962, to be published.

[5] W. G. Pfann, "Improvement of semiconductor devices by elastic strain," *Solid State Electronics*, vol. 3, pp. 261–267; November, 1961.

WWV and WWVH Standard Frequency and Time Transmissions*

The frequencies of the National Bureau of Standards radio stations WWV and WWVH are kept in agreement with respect to each other and have been maintained as constant as possible since December 1, 1957, with respect to an improved United States Frequency Standard (USFS).[1] The corrections reported here were arrived at by means of improved measurement methods based on transmissions from the NBS stations WWVB (60 kc) and WWVL (20 kc). The values given in the table are 5-day running averages of the daily 24-hour values for the period beginning at 1800 UT of each day listed.

The time signals of WWV and WWVH are also kept in agreement with each other. Since these signals are locked to the frequency of the transmissions, a continuous departure from UT2 may occur. Corrections are determined and published by the U. S. Naval Observatory. The time signals are maintained in close agreement with UT2 by properly offsetting the broadcast frequency from the USFS at the beginning of each year when necessary. This new system was commenced on January 1, 1960.

Subsequent changes were as follows:

FREQUENCY OFFSET, WITH REFERENCE TO THE USFS

January 1, 1960,	-150 parts in 10^{10}
January 1, 1962,	-130 parts in 10^{10}

* Received August 27, 1962.

[1] Refer to "National Standards of Time and Frequency in the United States," PROC. IRE, vol. 48, pp. 105–106; January, 1960.

TIME ADJUSTMENTS, WITH REFERENCE TO THE TIME SCALE UT2

December 16, 1959,	retardation,	20 milliseconds
January 1, 1961,	retardation,	5 milliseconds
August 1, 1961,	advancement,	50 milliseconds

Adjustments were made at 0000 UT on the foregoing dates; an advancement means that the signals were adjusted to occur at an earlier time than before.

WWV FREQUENCY
WITH RESPECT TO U. S. FREQUENCY STANDARD

1962		Parts in 10^{10}
July	1	−129.2
	2	−129.2
	3	−129.0†
	4	−130.0
	5	−129.9
	6	−129.9
	7	−129.9
	8	−129.9
	9	−129.9
	10	−129.8
	11	−129.7
	12	−129.6
	13	−129.6
	14	−129.5
	15	−129.5
	16	−129.4
	17	−129.4
	18	−129.4
	19	−129.3
	20	−129.3
	21	−129.3†
	22	−130.7
	23	−130.6
	24	−130.5
	25	−130.3
	26	−130.1
	27	−130.0
	28	−129.9
	29	−129.9
	30	−130.0
	31	−130.0

† A minus sign indicates that the broadcast frequency was below nominal. The uncertainty associated with these values is $\pm 5 \times 10^{-11}$.
WWV frequency adjusted as follows:
July 3, −1.1 parts $\times 10^{-10}$ at 1900 UT.
July 21, −1.4 parts $\times 10^{-10}$ at 1900 UT.

NATIONAL BUREAU OF STANDARDS
Boulder, Colo.

Miniature Electrically Tunable YIG Band-Pass Filter*

Early designs of electrically tunable band-pass filters using yttrium-iron-garnet (YIG) spheres have concentrated on single-resonator filters.[1,2] However, the performance of single-resonator filters is limited by spurious responses, (about 15 db below the main response), caused by magnetostatic modes.[3] Two-resonator filters have reduced this spurious response level to about 35 db below the main response; however, such filters are ordinarily large. Those two-resonator filters already developed for use in S band[4] and *X* band[1] are typical examples. In many applications, it is highly desirable to miniaturize these YIG filter structures because a reduction in filter size also reduces the magnet size and tuning-power requirements.

We have developed a two-resonator YIG filter (Fig. 1) that tunes from 2.4 to 4.8 Gc, which has a residual response level of 60 to 80 db, and has spot-frequency spurious resonances below 44 db. It has a volume, excluding the magnet, of less than 0.25 cubic inch. The filter is a miniaturized version of a basic structure used by Carter.[4] It consists of two garnets coupled to two strip transmission lines with a slot in the common ground plane. The slot is the coupling element between the resonators. The small size is achieved primarily by using strip-transmission-line ground-plane separations of 0.162 inch. Previously fabricated structures have had large ground-plane separations (0.6 inch).[4] Power is coupled from one line to the other when the YIG spheres are at their ferrimagnetic resonant frequencies. The filter has a two-pole response with a nominal 3-db bandwidth of 35 Mc. The filter is tuned by varying the field of an electromagnetic about a permanent-magnet biasing field. The biasing field was set for 1140 oersteds, corresponding to a frequency of 3.2 Gc.

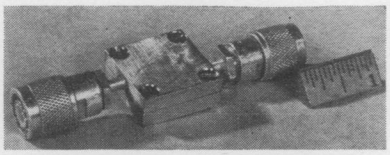


Fig. 1—Miniature two-resonator YIG band-pass filter.

The filter characteristics have been evaluated between 2.4 to 4.8 Gc. Fig. 2 shows a typical frequency response curve, and Fig. 3 shows the characteristics of the filter over the complete tuning range. The curve in Fig. 2 is centered at 3.8 Gc with a 34-Mc 3-db bandwidth and a 1.1-db insertion loss at the peak of the response. It has a frequency response equivalent to a two-pole Chebyshev filter having a 0.8-db pass-band ripple up to about 40 db of insertion loss. It then continues to a residual loss level of 83 db except for the two narrow-band responses that are 50 and 62 db down. The strongest spurious level is therefore 50 db down when the filter is tuned to 3.8 Gc. Fig. 3 is a plot of the strongest spurious-response level, the midband insertion loss, and the 3-db bandwidth vs frequency. The spurious-response level is always more than 44 db down, the 3-db bandwidth averages about 35 Mc, and the midband insertion loss is under 2 db over most of the range. The insertion loss starts to increase slightly at the low-frequency end because the unloaded *Q*'s of the YIG spheres are degraded. The filter is capable of tuning over the full 2.4 to 4.8 Gc range with 3 watts of power.

These data were measured after the resonant frequencies of the two garnets were very carefully aligned. The alignment, essential to low insertion loss, was achieved by rotating the axes of the garnets until they both had the same resonant frequencies.

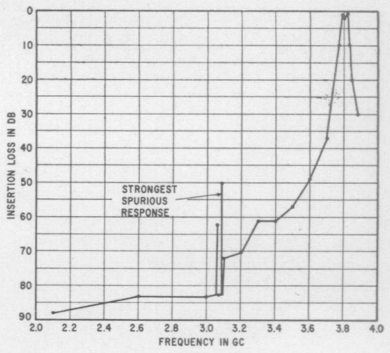


Fig. 2—Typical response curve.

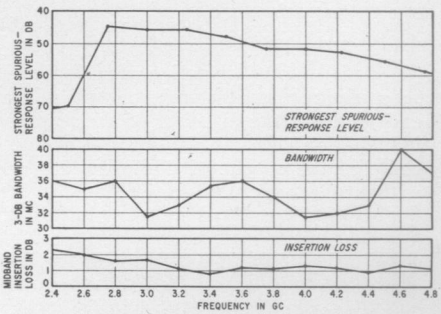


Fig. 3—Characteristics of miniature two-resonator YIG band-pass filter.

This procedure is effective because the insertion loss is only slightly higher (0.1 to 0.3 db) over the entire tuning range than when the filter was peaked to any specific frequency by individually tuning each sphere. Therefore, the filter can be accurately tuned by only one adjustment—the variation of the magnetic field.

In many receiver applications, the presence of the few, narrow-band spurious-response levels appearing at any given filter center frequency do not necessarily determine the strongest spurious response of the entire receiver. By careful choice of the IF amplifier center frequency, it is often possible to reject the few strong filter spurious levels present. Where this is feasible, the YIG filter described would have an effective spurious response level equal to its residual level of 60 to 80 db.

Work is continuing on techniques to reduce the strongest spurious response levels to more than 60 db below the main response level and to increase the tuning range. The present filter tunes from 2.1 to 5.6 Gc with slightly degraded performance between 2.1 to 2.4 Gc and 4.8 to 5.6 Gc. The results to date indicate that miniature YIG filters with two resonators can be constructed and tuned over more than one octave with a single voltage control. The rapid and remote tuning capabilities of these filters can be used extensively in microwave receiving systems.

The authors wish to thank A. H. Harvey of AIL for assisting them in the construction and evaluation of the filter.

M. L. WRIGHT
J. J. TAUB
Airborne Instruments Lab.
Division of Cutler-Hammer, Inc.
Deer Park, N. Y.

* Received June 25, 1962; revised manuscript received, July 2, 1962.

[1] P. S. Carter, Jr., "Design of magnetically-tunable microwave filters using single-crystal yttrium iron-garnet resonators," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 252–260; May, 1961.
[2] K. L. Kotzebue, "Broadband electronically-tuned microwave filters," 1960 IRE WESCON CONVENTION RECORD, pt. 1, pp. 21–27.
[3] L. R. Walker, "Magnetostatic modes in ferromagnetic resonance," *Phys. Rev.*, vol. 105, pp. 390–399; January 15, 1957.
[4] P. S. Carter, Jr., *et al.*, "Microwave Filters and Coupling Structures," Stanford Res. Inst., Menlo Park, Calif., Rept. No. 3, Contract DA-36-039-sc-87398, SRI Proj 3527; October, 1961.

Use of Electro-Optical Shutters to Stabilize Ruby Laser Operation*

Emission from a ruby laser is normally characterized by extreme fluctuations, apparently arising from an instability of stimulated emission oscillations in a resonant cavity, and further complicated by the existence of many degrees of freedom (modes) which can be excited.

The use of a feedback control to stabilize oscillation level has apparently not been previously attempted. However, the theory and practice of feedback controls is well established.

We have used a Kerr electro-optical shutter as the control element in a feedback loop to stabilize the operation of a ruby laser. The stabilizing effect is dramatic; output is changed from irregular "spiking" to essentially steady emission which generally follows the intensity of the pumping illumination. The two oscillograms of Fig. 1 compare the output of the laser with and without feedback stabilization.

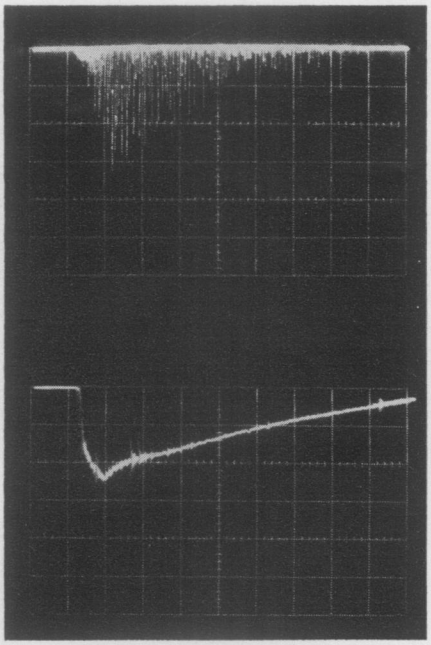


Fig. 1—Upper trace oscillogram of 1: Ruby laser output without stabilization. Lower trace, oscillogram of ruby laser with feedback stabilization. Sweep speed: 200 microseconds per major division in both traces.

The experimental arrangement is shown in Fig. 2. Both the Kerr cell and a polarizer[1] are included, together with the ruby, in the optical resonator. The mirrors are adjustable for parallelism, and the ruby is pumped with a helical xenon flash tube.

Feedback is accomplished by directing a portion of the laser output onto a high current photodiode. The photodiode drives the Kerr cell directly, the load impedance being chosen to provide suitable gain and phase characteristics. The Kerr cell is specifically designed for laser modulation. A bias voltage is applied in order to bring the Kerr cell to a more sensitive portion of its characteristic.

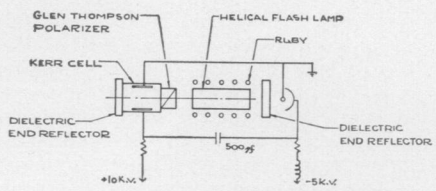


Fig. 2—Schematic of Kerr electro-optical stabilized ruby laser.

We believe that the value of this development lies in two broad areas. First, it is now feasible to control a ruby laser, and also to introduce amplitude modulation for many obvious applications. Second, this development makes the ruby laser operation much more amenable to theoretical study and analysis. Improvement in line width, coherence, and beam angle is anticipated to result from the more stable operation.

F. R. Marshall
D. L. Roberts
Quantatron, Inc.
Santa Monica, Calif.

Elimination of Re-Entry Radio Blackout*

Mr. H. Hodara has suggested[1] that the propagation of an EM wave through a collisionless plasma can be greatly improved by the use of a longitudinal magnetic field; in particular he mentioned that the propagation characteristics of a right-handed (RH) circularly polarized wave are much better than those of a left-handed (LH) circularly polarized wave. Although this is true for an infinite medium, it would appear that in the more realistic case of transmission through finite slabs this advantage is vitiated.

To see this consider the propagation of a VHF signal ($\omega=1.6\times10^9$) through a 1-cm thick homogeneous plasma when a 570-gauss ($\omega_b=10^{10}$) longitudinal magnetic field is employed. Fig. 1 shows the transmission coefficient as the plasma frequency is varied. For small values of the plasma frequency ($\omega_p\sim10^{10}$), there is substantially complete transmission for both right- and left-hand signals. As ω_p increases, however, the transmission coefficient for the LH signal decreases monitonically at a rate which becomes proportional to

$$e^{-[2x/c\sqrt{\omega_b/\omega)+1}]}\,\omega_p$$

(x is the slab thickness and c is the velocity of light).

The RH signal does not have this "decaying exponential" behavior since for large plasma frequencies its index of refraction

$$\left(n=\left[1+\frac{\omega_p^2}{\omega(\omega_b-\omega)}\right]^{1/2}\right)$$

is real. Thus, the problem is simply one of transmission through a homogeneous slab of real index of refraction as either the thickness or index of refraction is increased.

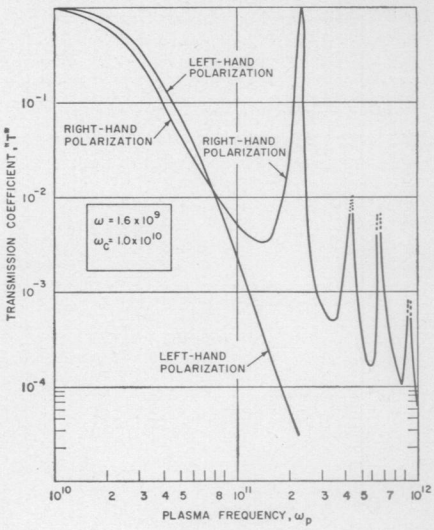


Fig. 1.

As a practical matter, however, Fig. 1 shows that the RH polarization is no improvement over the LH wave. The transmission coefficient is simply

$$T=\left[1-\frac{n^2-1}{4n^2}\sin^2 kxn\right]^{-1},$$

and for large index of refraction ($\omega_p^2\gg\omega\omega_b$) there will be, on the average, a large reflection. In fact in this case the RH signal reaches a 10-db attenuation point slightly before the LH signal.[2] The fact that aperiodically the RH signal has complete transmission is of little help since the plasma conditions are continuously changing and one would have sporadic transmission at best. For an inhomogeneous plasma the peaks are lowered and the valleys raised so that in a more realistic case one would not even get these sharp bursts of transmission.

Even though an RH signal has no advantage over an LH signal, there is still the question of whether a magnetic field has any advantage at all. The answer to this question can be seen by examining Fig. 2 where transmission has been computed for a 5700-gauss ($\omega_b=10^{11}$) longitudinal magnetic field. By increasing ω_b by a factor of 10 one is able to stand a three-fold increase in ω_p before reaching a 10-db attenuation. Thus one can get through a plasma by using a strong magnetic field, but as a practical matter the necessary field strengths are often prohibitively large.

* Received August 1, 1962.

[1] The polarizer is not essential, since the dichroism of the ruby will generally suffice.

* Received March 7, 1962.

[1] H. Hodara, "The use of magnetic fields in the elimination of the re-entry radio blackout," Proc. IRE, vol. 49, pp. 1825–1830; December, 1961.

[2] It might be objected that one might be willing to live with 20 or 30 db attenuation by going to high power systems in which case the RH system is better than the LH system. This has a disadvantage both in that high powers tend to induce antenna breakdown and also that there will be strong reflection and thus mismatch, particularly for the low-frequency systems intimated by the use of magnetic fields. (A magnetic field is of no advantage unless the signal frequency is less than the cyclotron frequency.)

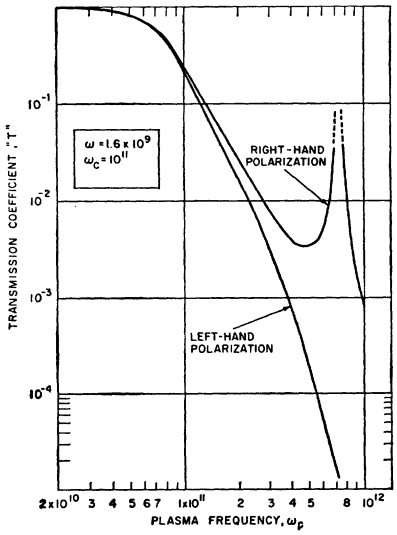


Fig. 2.

The conclusion is thus two-fold:

1) There is no advantage of one polarization over another.
2) The only way to use a magnetic field to get through a *dense* plasma is to use a *strong* magnetic field—there is no royal road to propagation.

C. R. MULLIN
Research and Advanced Development Div.
AVCO
Wilmington, Mass.

Author's Comments[3]

I would like to thank Mr. Mullin for the interest shown in my work.

Mr. Mullin's comments deserve to be discussed and I believe I can answer the questions he raises, if not to his satisfaction, at least to mine. There are essentially two statements to his thesis, first that there is no advantage in using an RH wave over an LH wave, second that in order to propagate through a dense plasma, strong static magnetic fields are required. Both of his conclusions are stated with much generality and their value can only be determined when expressed quantitatively. It is the purpose of this reply to determine this value.

Let us discuss the first statement. According to Mr. Mullin it is implied in my paper[1] that an RH wave is "much better" than an LH wave. Nowhere in the paper is there such a statement as it does not convey any quantitative information. It is stated, however, on page 1829 of the paper, that for a signal frequency below cyclotron resonance, 3 db of additional power penetrates the plasma medium if the incident wave is RH circularly polarized instead of linearly polarized. Whether 3 db is "much better" than nothing is another matter. On the other hand for the collisionless semi-infinite plasma under discussion at signal frequencies below cyclotron resonance, an RH wave is partially transmitted while an LH wave is totally reflected in which case the ratio of transmitted RH to LH wave is infinite; for this case indeed an RH wave is "much better" than an LH wave. In reality, as Mr. Mullin correctly points out, the plasma sheath surrounding a re-entering space vehicle has a finite width and is a far cry from the semi-infinite medium approximation.

A propagation analysis based on a homogeneous finite width plasma slab may be just as far a cry from reality since it is well known that the electron density exhibits a sharp discontinuity at the skin of the vehicle and decays approximately exponentially towards the shock layer boundary (the statement "sharp discontinuity" may be as vague to Mr. Mullin as the statement "much better" mentioned above). It would seem that the actual model of an inhomogeneous plasma sheath lies somewhere in between the two approximations: the homogeneous finite width slab and the homogeneous semi-infinite medium. As pointed out by Mr. Mullin, in the case of an inhomogeneous slab, the valleys in his curve of Fig. 1 are raised and the advantages of an RH wave over an LH wave are improved, provided of course it is possible to transmit with the attenuation levels corresponding to the valleys of the curve Fig. 1 rather than to the peaks. This brings up Mr. Mullin's point of contention, namely that it is impractical or at least disadvantageous to transmit with such high attenuation levels because of the deleterious effects brought about by the additional required power as described in the footnote[2] to his comments. Since this statement applies not only the the pros and cons of an RH wave over an LH wave but also to the relative advantages or disadvantages of propagation with and without magnetic field, let us discuss then Mr. Mullin's second conclusion before answering the statement in question.

Let us briefly analyze how the power transmission coefficient varies for a homogeneous plasma slab of finite width L for the cases of no magnetic field, transverse magnetic field, and longitudinal magnetic field. The transmission coefficients for these cases are respectively:

$$T_0 = \left| \frac{1}{c_0 + \frac{j}{2} s_0 \left(b_0 + \frac{1}{b_0} \right)} \right|^2 \quad \text{Isotropic case } (B_0 = 0), \tag{1}$$

$$T_T = \left| \frac{1}{c_T + \frac{j}{2} s_T \left(b_T + \frac{1}{b_T} \right)} \right|^2 \quad \text{Transverse static magnetic field case} \tag{2}$$

$$T_L = \left| \frac{1/2}{c_1 + \frac{j}{2} s_1 \left(b_1 + \frac{1}{b_1} \right)} + \frac{1/2}{c_2 + \frac{j}{2} s_2 \left(b_2 + \frac{1}{b_2} \right)} \right|^2 + \left| \frac{1/2}{c_1 + \frac{j}{2} s_1 \left(b_1 + \frac{1}{b_1} \right)} - \frac{1/2}{c_2 + \frac{j}{2} s_2 \left(b_2 + \frac{1}{b_2} \right)} \right|^2 \tag{3a}$$

Longitudinal magnetic field case; incident wave linearly polarized,

$$T_{1,2} = \left| \frac{1}{c_{1,2} + \frac{j}{2} s_{1,2} \left(b_{1,2} + \frac{1}{b_{1,2}} \right)} \right|^2 \quad \text{Longitudinal magnetic field case with incident wave circularly polarized to the right } (T_1) \text{ or to the left } (T_2). \tag{3b}$$

The following notation is used in the above equations: the subscripts 0, T, L refer, respectively, to isotropic, transverse, and longitudinal static magnetic field case. The subscripts 1 and 2 refer to an RH mode and an LH mode and the various symbols are defined below:

$$c_n = \cos b_n l \qquad (n = 0, T, 1, 2)$$
$$s_n = \sin b_n l$$
$$l = \frac{L}{\lambda\!\!\!^{-}_v}$$
$$\lambda\!\!\!^{-}_v = \frac{\lambda_v}{2\pi} \quad (\lambda = \text{wavelength of signal in free space})$$
$$b_0 = \sqrt{\epsilon_1}$$
$$b_T = \sqrt{\epsilon_1 + \frac{\epsilon_2^2}{\epsilon_1}}$$
$$b_{1,2} = \sqrt{\epsilon_1 \mp j\epsilon_2} \quad (- \text{ sign corresponds to subscript 1, } + \text{ sign to subscript 2})$$
$$\left.\begin{aligned} \epsilon_1 &= 1 + \frac{\omega_p^2}{\omega_b^2 - \omega^2} \\ \epsilon_2 &= j \frac{\omega_b}{\omega} \frac{\omega_p^2}{\omega_b^2 - \omega^2} \end{aligned}\right\} \text{(collisionless case: } \nu = 0).$$

The case of practical interest in the blackout problem is

$$\omega^2 \ll \omega_b^2 \ll \omega_p^2 \tag{4}$$

and (1), (2), and (3b) reduce to

$$T_0 \sim 16 \frac{\omega^2}{\omega_p^2} e^{-2(\omega_p/\omega)(L/\lambda\!\!\!^{-}_v)} \tag{5}$$

$$T_T \sim 16 \frac{\omega^2}{\omega_p^2} e^{-2(\omega_p/\omega)(L/\lambda\!\!\!^{-}_v)} \tag{6}$$

$$\tilde{T}_1 \approx 4 \frac{\omega_b \omega}{\omega_p^2} \tag{7}$$

$$T_2 \approx 4 \frac{\omega_b \omega}{\omega_p^2} e^{-2(\omega_p/\sqrt{\omega_b \omega})(L/\lambda\!\!\!^{-}_v)}. \tag{8}$$

Eq. (7) describes the locus of the minima of the transmission coefficient for the RH wave; it describes the worst possible condition for this case.

Although some of the parameters used by Mullin in this discussion do not quite fulfill

[3] Received July 26, 1962.

the inequality (4), the discrepancy between the results using (5)–(8) and Mr. Mullin's results is not higher than 20 per cent. This error is tolerable for the present discussion since it simplifies the analysis and does not limit the validity of the conclusions to be reached. No generality is lost and the exact formulas (1), (2), and (3) are available if desired.

Under condition (4), comparison of (5) and (6) shows no reduction in the transmission coefficient with a transverse magnetic field. On the other hand with an RH wave and a longitudinal magnetic field, although both $\breve{T}_1$ and T_0 decrease inversely with the square of the plasma frequency, T_0 decreases exponentially with the slab width while T_1 is not affected. In particular, the ratio

$$\frac{\breve{T}_1}{T_0} = \frac{\breve{T}_1}{T_T} = \frac{1}{4}\,\frac{\omega_b}{\omega}\,e^{2(\omega_p/\omega)(L/\lambda_v)} \qquad (9)$$

shows clearly the relative improvement to be obtained with a longitudinal static magnetic field and an RH wave. As ω_p and L increase, the relative decibel improvement over no magnetic field increases rapidly as seen from (9). For example if $\omega_p = 1.6 \times 10^9$, $\omega_b = 1.0 \times 10^{10}$, $L = 1$ cm, then $10 \log_{10} \breve{T}_1/T_0 = 48$ db. Thus if a "strong" longitudinal magnetic field is required to propagate through the plasma, for the same attenuation loss, a "much stronger" (48 db stronger) transverse magnetic field is necessary. If L is doubled, the relative db improvement is then $10 \log_{10} \breve{T}_1/T_0 = 94$ db. Such improvements occur at attenuation levels, of the order of 25 db; this brings up the unanswered question and the very same point of contention brought out by Mr. Mullin's footnote,[2] namely, is it advantageous or practical to transmit with as much attenuation? If one wants to assume, as Mr. Mullin does, that it is disadvantageous to raise the transmitter power to overcome the 25-db attenuation (in spite of the fact that with no magnetic field the attenuation would be 94 db for a 2-cm thick plasma sheath), the answer lies in the fact that it may be possible not to raise the transmitting power if the ground receiving system has sufficient sensitivity. In other words a criterion for transmission is not only the plasma attenuation but also the minimum detectable SNR at the receiver. As an example,[4] SNR are plotted in Fig. 3 for the re-entry path shown in Fig. 3, under two conditions: no static magnetic field, $B_0 = 0$ and $B_0 = 1000$ gauss. The curves are self-explanatory; the results show that 1000 gauss seem to be adequate to insure a minimum SNR of 25 db throughout re-entry. In making these calculations the variations of plasma sheath thickness, plasma frequency and collision frequency as a function of altitude have been taken into account. These results by themselves do not guarantee the feasibility of the magnetic field approach. Much more experimental data and flight testing are necessary before any conclusion can be drawn as to the validity or invalidity of this approach. Further theoretical studies are necessary to take into account the characteristics of

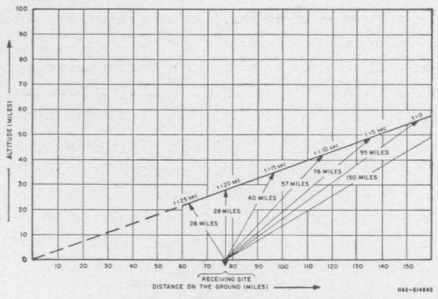


Fig. 3—Range vs altitude at a selected receiving site for typical R/V.

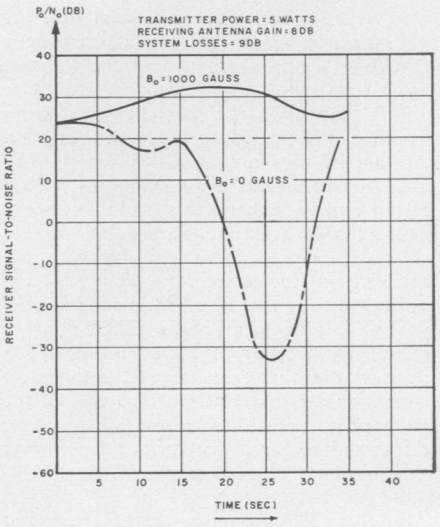


Fig. 4—SNR at selected receiving site with transmitter located aft of the R/V.

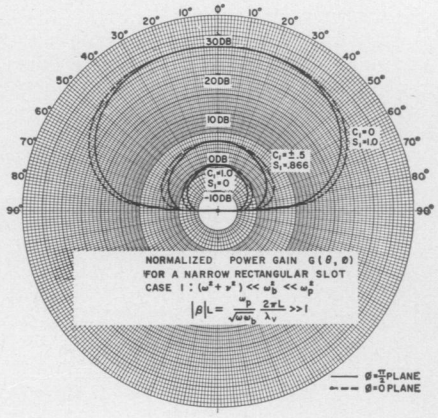


Fig. 5.

the radiating structure. Recent results[5] shown in Fig. 5 illustrate the pattern to be expected from a narrow rectangular slot covered with a homogeneous gyroplasma sheath. Each curve in Fig. 5 corresponds to a different thickness and the variations in amplitudes of the patterns match qualitatively the oscillating behavior of the transmission coefficient in Mr. Mullin's Fig. 1.

The impedance of a radiation structure covered with gyroplasma is also being investigated at this laboratory. It is believed that the results of this impedance analysis coupled with recent experimental data will shed additional light on the feasibility of the static magnetic field approach.

H. Hodara
The Hallicrafters Co.
Chicago, Ill.

Comments on "Broad-Band Parametric Amplifiers by Simple Experimental Techniques"*

Supplementary technical information is currently available which did not appear in the original paper.[1] This is included in the following miscellaneous comments:

1) All results were achieved using RG-52/U waveguide diode mounts (propagating to both pump and idle frequencies in the TE_{10} mode) and coaxial circuitry for the signal. A photograph of a typical breadboard model of the paramp is shown in Fig. 1. This paramp configuration is a modified version of structures previously developed at RCA.[2,3] The signal circuit contains a coaxial low-pass filter and a coaxial double quarter-wave transformer. A Melabs Model X-344 coaxial ferrite circulator was employed. This circulator had a useful bandwidth from 2.0 kMc to 3.0 kMc. The triple screw tuner shown in the photograph was

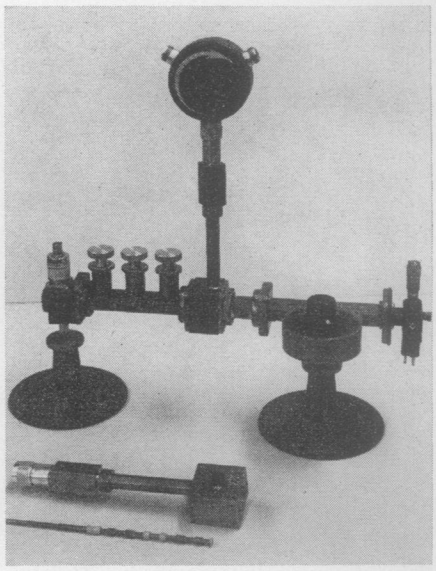


Fig. 1.

* Received March 23, 1962.

[1] B. B. Bossard and R. Pettai, "Broad-band parametric amplifiers by simple experimental techniques," PROC. IRE (*Correspondence*), vol. 50, pp. 328–329; March, 1962.

[2] R. Pettai, B. Bossard, and S. Weisbaum, "Single-diode parametric upconverter with large gain-bandwidth product," PROC. IRE (*Correspondence*), vol. 48, pp. 1323–1324; July, 1960.

[3] R. M. Kurzrok, "Parametric amplifier device," U. S. Patent No. 2,970,275; January 31, 1961.

[4] H. Hodara, "Calculations Pertinent to the Re-Entry Blackout," The Hallicrafters Co., Chicago, Ill., Tech. Rept., January, 1961.

[5] H. Hodara, "Radiations from a gyroplasma sheathed aperture," to be published in IRE TRANS. ON ANTENNAS AND PROPAGATION, January, 1963.

used to trim the idle circuit; however, the gain response shown in Fig. 1 of the original paper was achieved using three commercially available slide screw tuners in the idle circuit. More recent results at S band have entailed a gain of 11.2 db, a 1-db bandwidth of 580 Mc and a 3-db bandwidth of 700 Mc.

2) Best paramp performance was obtained using a partial height capacitive post for impedance. Similar techniques have also been used in broad-band waveguide conventional mixers.

3) External idler resistive loading was employed. Although paramp noise figures for a given varactor diode are no longer optimum, the noise figures obtained are only about 0.3 db worse for the case of no external idler loading. This should be satisfactory for most applications. Other significant broad-band microwave parametric amplifiers reported in the literature[4-7] *do not* use external idler loading.

4) A silicon cartridge-type diode with avalanche breakdown at −5 volts was used in achieving the best results. Typical amplifier gain saturation for this diode is as follows:

a) 1-db deviation from small signal gain at signal input level of −13 dbm
b) 3-db deviation from small signal gain at signal input level of −7 dbm.

The large dynamic range of the paramp yields excellent intermodulation performance. For example, a small signal of −90 dbm in the presence of a large interfering signal of −10 dbm caused no discernible intermodulation products.

5) The diode used apparently did not exhibit an abnormally high value of $C_1/2C_0$ due to anamolous minority carrier-charge storage effect; however, this has not been conclusively proven. For broad-band parametric amplifier operation at microwave frequencies, diodes must have carefully controlled values of junction capacitance, case capacitance, and lead inductance. Availability of suitable diodes is not a major problem in the advanced development of S-band paramps with bandwidths of 400 Mc; however, selected diodes are required for S-band paramps having bandwidths of 700 Mc to 900 Mc.

6) Preliminary measurements of signal and idle pass bands are accomplished using swept frequency techniques. If a particular diode looks promising, detailed point-by-point data is subsequently taken. Point-by-point measurement of the signal pass band entails use of a high-level signal input (about zero dbm) to obtain sufficient rectification. Point-by-point measurement of the idle pass band entails use of a low-level idle input (about −30 dbm). Both of these passband measurements can be made under static operating conditions (*i.e.*, without pump power).

7) The broad-band paramp can be tuned over a 1000 Mc frequency range by simply changing the pump frequency. Instantaneous single-tuned bandwidths of 60 Mc were obtained for amplifier gain levels of 16 db. Further details of this experimentation will be reported subsequently.

The experimental work of B. S. Perlman is acknowledged.

B. B. Bossard
R. M. Kurzrok
RCA
Defense Electronic Products
Surface Communications Systems Lab.
New York, N. Y.

[4] J. Kliphuis, "C-band nondegenerate parametric amplifier with 500-Mc bandwidth," Proc. IRE (*Correspondence*), vol. 49, p. 961; May, 1961.
[5] B. T. Vincent, "A C-band parametric amplifier with large bandwidth," Proc. IRE (*Correspondence*), vol. 49, p. 1682; November, 1961.
[6] K. M. Johnson, "Broad-band S-band parametric amplifier," Proc. IRE (*Correspondence*), vol. 49, p. 1943; December, 1961.
[7] K. M. Johnson, "900-Mc, non-degenerate X-band amplifier," Proc. IRE (*Correspondence*), vol. 50, p. 332; March, 1962.

New Helium-Neon Optical Maser Transitions*

This letter is intended as a summary of our optical maser investigations in RF discharges containing mixtures of helium and neon. For the purpose of making these investigations, a special Fabry-Perot interferometer was constructed. A confocal system[1] was chosen for ease in alignment and Brewster angle windows were incorporated to permit the use of external mirrors.[2] The use of external mirrors is desirable to eliminate damage to the high-reflectance dielectric coatings during bakeout of the vacuum system and to permit changing mirrors without opening the system to air. In order to increase the optical gain, a long discharge length was used; a 225-cm length (quartz) discharge tube was found to be a reasonable compromise between the various practical considerations involved. Since previous gain studies[3] in the helium-neon system and in pure neon showed that the gain varied roughly as the reciprocal of the tube diameter, a small diameter was employed. A 7-mm inside diameter was found to be a realistic compromise between gain and diffraction loss (which is primarily determined by tube-alignment errors with these dimensions). A more detailed account of the maser and of the vacuum techniques will be given elsewhere.[4]

Oscillation in the helium-neon system has been reported previously[5] on the neon transitions at 1.1177, 1.1523, 1.1614, 1.1985 and 1.2066 microns and these lines were easily observed with the present maser. Of these five transitions, the strongest oscillation is obtained at 1.1523 microns. We have obtained maser oscillation on five additional $2s$-$2p$ transitions (Paschen notation) in this system and a summary of the present results is contained in Table I. The 1.5231 micron transition is second in intensity only to the previously reported line at 1.1523 microns. The partial pressures of helium and neon were approximately in the ratio of 10:1 with the total pressure ≅1 to 2 mm Hg for these measurements.

TABLE I
New Helium-Neon Optical Maser Transitions (Paschen Notation)

Transition	λ_{air} (microns)
$2s_3-2p_7$	1.0798
$2s_2-2p_6$	1.0844
$2s_4-2p_8$	1.1143
$2s_5-2p_8$	1.1390
$2s_2-2p_2$	1.1767
$2s_2-2p_1$	1.5231

It is reasonable to expect population inversions on all of the thirty allowed $2s$-$2p$ transitions of neon (ranging from about 0.9 to 1.8 microns) in this system.[6,6,7] Needless to say, the optical gain must exceed the loss in the maser for oscillation to take place on a given transition. The gain obtainable will, of course, depend on the oscillator strength for the transition as well as on the population difference; and one expects oscillation to occur most readily on the stronger lines. Koster and Statz[8] have calculated approximate values for these relative transition probabilities. Although there is no pronounced correlation between the observed gain and the calculated transition probabilities, all but two transitions ($2s_3$-$2p_7$ at 1.0798 microns) and $2s_2$-$2p_6$ at 1.0844 microns on which oscillation has been obtained have at least been reported by Koster and Statz to have nonzero probabilities. The two discrepancies presumably arise from a failure of the *jl*-coupling approximation used in the calculation. In this connection, it should be noted that many of the "forbidden" transitions listed by Koster and Statz are conspicuously present (though generally weak) in the spontaneous emission spectrum.

We investigated the maser output in the vicinity of the strong maser line at 1.1523 microns with a high resolution (1:300,000) grating spectrometer and found no sign of oscillation on the closely neighboring $2s_4$-$2p_7$ transition at 1.1525 microns. A similar result was obtained previously[9] with a one meter (lower gain), plane parallel maser. The difference-wavelength fringes that would be provided by simultaneous oscillation on these two neighboring lines would be of considerable practical value in long-distance interferometry.

Attempts to obtain oscillation on the shortest wavelength $2s$-$2p$ transitions—in particular, the $2s_4$-$2p_{10}$, $2s_5$-$2p_{10}$, and $2s_3$-$2p_{10}$ transitions at 0.949 μ, 0.967 μ, and 0.899 μ

* Received July 31, 1962.
[1] G. D. Boyd and J. P. Gordon, "Confocal multimode resonator for millimeter through optical wavelength masers," *Bell Sys. Tech. J.*, vol. 25, pp. 489–509; March, 1961.
[2] W. W. Rigrod, H. Kogelnick, B. J. Brangaccio, and D. R. Herriott, "Gaseous optical maser with external concave mirrors," *J. Appl. Phys.*, vol. 33, pp. 743–744; February, 1962.
[3] W. R. Bennett, Jr., "Recent progress in experiments with helium-neon masers," *Bull. Am. Phys. Soc. II*, vol. 7, p. 15; January, 1962.
[4] C. K. N. Patel, "Optical power output in He-Ne and pure neon masers," *J. Appl. Phys.*, to be published.
[5] A. Javan, W. R. Bennett, Jr., and D. R. Herriott, "Population inversion and continuous optical maser oscillation in a gas discharge containing a He-Ne mixture," *Phys. Rev. Lett.*, vol. 6, pp. 106–110; February, 1961.
[6] A. Javan, "Optical maser oscillations in a gaseous discharge," in "Advances in Quantum Electronics," J. Singer, Ed., Columbia University Press, New York, N. Y., pp. 18–27; 1961.
[7] W. R. Bennett, Jr., "Radiative lifetimes and collision transfer cross-sections of excited atomic states," in "Advances in Quantum Electronics," J. Singer, Ed., Columbia University Press, New York, N. Y., pp. 28–43; 1961.
[8] G. F. Koster and H. Statz, "Probabilities for the neon laser transitions," *J. Appl. Phys.*, vol. 32, pp. 2054–2055; June, 1961.
[9] W. R. Bennett, Jr., "Hole burning effects in a He-Ne optical maser," *Phys. Rev.*, vol. 126, pp. 580–593; April, 1962.

initially proposed by Javan[10]—have been unsuccessful so far. This failure may have its origin in a larger population of the $2p_{10}$ in respect to the other $2p$ levels—for example, through collision transfer effects in the $2p$ group. However, there is no reason to believe that these transitions are not inverted at low discharge intensities and it seems probable that oscillation might be obtained by going to still longer discharge lengths and/or smaller tube diameters. Our results indicate that the discharge is at least highly transparent at 0.967 microns, and also on several of the $3d$-$2p$ transitions. The latter is particularly significant since the $3d$ levels fall too far in energy above the $He(2^3S_1)$ to be excited by this helium metastable with any appreciable degree of probability. On the other hand, many of the $3d$ levels should have large cross sections for excitation by electron impact. This general property of $3d$ levels has been confirmed by the authors[11] by obtaining maser oscillation on the $3d$-$2p$ transitions in other pure noble gases.

The above discussion has, of course, dealt entirely with the infrared transitions of neon and primarily with the $2s \rightarrow 2p$ transitions. It should be mentioned that oscillation has also been obtained [12] recently in a helium-neon maser on the $3s_2 \rightarrow 2p_4$ transition at 0.6328 microns. The processes involved in the excitation of this visible transition, of course, differ in detail from those considered here.

R. A. McFarlane,
C. K. N. Patel,
W. R. Bennett, Jr.
W. L. Faust
Bell Telephone Labs.
Murray Hill, N. J.

[10] A. Javan, "Possibility of production of negative temperature in gas discharges," *Phys. Rev. Lett.*, vol. 3, pp. 87–89; July, 1959. "Quantum Electronics," C. H. Townes, Ed., Columbia University Press, New York, N. Y., pp. 564–571; 1960.
[11] C. K. N. Patel, W. R. Bennett, Jr., W. L. Faust, and R. A. McFarlane, "Infrared spectroscopy using stimulated emission techniques," *Phys. Rev. Lett.*, vol. 9, pp. 102–109; August, 1962.
[12] A. D. White and J. D. Rigden, "Continous gas maser oscillation in the visible," Proc. IRE, (*Correspondence*), vol. 50, p. 1697; July, 1962.

The Measurement of the Velocity of Light*

I was very interested in Dr. Rapier's note.[1] Although I am not working in that field, I have put considerable thought into the problem of the measurement of the velocity of light. There seems to be some evidence for light having a constant velocity relative to its source and also for there being physical reasons for this possibility. However, it seems to me that none of the experiments to show that light has a constant velocity relative to the receiver has actually shown that it has or that it hasn't. I believe that this is true because in all these experiments an important factor has been neglected. In order to explain what this factor is, let me use an analogy in the radio-frequency part of the electromagnetic spectrum.

Suppose that we want to measure the velocity of a radio signal sent back from a rocket receding from the earth. We could measure the difference in arrival time (of different parts of some identifiable wavefront, for instance) at two spaced receivers. However, if the velocity of the signal were constant relative to the rocket and independent of the receiver, and if we interposed a receiver and reradiator in the path between the rocket and the measuring receivers, we would only measure the velocity of the wave from the reradiator and would have no information as to whether the velocity of the signal was constant relative to the rocket or not. This would be true even if this reradiator were a parasitic dipole or a conducting metal sheet, for instance. The electric vector of the wave from the rocket induces currents in the reflecting dipole or sheet. These currents then radiate energy as though this reflector had been the antenna on a transmitter and, hence, this dipole or sheet acts as a new source.

The light that is reflected from a mirror is reradiated in the same way, hence the mirror acts as a new source. If light has a velocity which is constant relative to its source, the light after reflection will have a constant velocity relative to the mirror and not to its primary source. A lens also acts in the same way. Light does not go through it without change but interacts with the molecules of the glass, and each molecule influences its neighbor till new light is radiated from the opposite side of the lens. The lens therefore acts as a new source. At microwave frequencies, lenses have been built with artificial dielectric made of pieces of metal. An analysis of their operation shows very little of the incident radiation as going through the structure. Most of what leaves this kind of lens is considered as reradiated from the components of the lens which make up the artificial dielectric.

There may be cases, such as for X and γ rays, where the radiation actually passes through matter without interaction with it, and similar cases for light for short distances through gas; but any experiment which attempts to measure the velocity of light directly from a source moving relative to the receiver when there is a possibility that this velocity is constant relative to its source or even has a component that is, cannot use mirrors or lenses or other transparent matter and must depend on the timing or phasing of mechanical shutters (toothed wheels, etc.) in vacuum.

I am afraid that the proposed experiment using a Čerenkov flash would have the same drawback, since even though the original light as generated had the velocity that is suggested, it has to pass through transparent *matter;* and therefore the velocity of the light that would be measured would be constant relative to the structure and not to the electrons of the source.

One could determine whether the velocity of light was constant relative to the receiver or source by measuring the difference in velocity of the light from the sun as it arrived at a satellite or at the moon, as the satellite or moon approached and receded from the sun.

Another way of determining whether or not light has a constant velocity relative to its source alone is that if light has a constant velocity relative to its source and if the velocity of light in glass is constant relative to the glass, independent of the velocity of the impinging light, then light from a source receding from a piece of glass will be refracted less than light from a source moving at the same velocity as the glass. Therefore, if light has a constant velocity relative to its source alone the relative velocity of the source of light could be measured by determining the focal length of a lens of which a curved surface was the first matter that intercepted the light from that source. Also, since the light reflected from a mirror would have a different velocity than that incident from a moving source, the angle of reflection would *not* be equal to the angle of incidence and therefore a similar experiment could be performed with mirrors.

A very sensitive determination of the relative velocities of light from a moving source and a stationary one can be made, if the above conditions apply, by interposing a thin piece of glass or a mirror in the path of a portion of the light from a moving source. Either the glass or the mirror will produce light coherent with the light from the moving source but now coming from a stationary source (the piece of glass or the mirror). If this portion and the unchanged portion are allowed to travel some distance and then combined after traversing equal optical path lengths, they will interfere with each other. By slight adjustment of the length of one path, cancellation of reinforcement can be obtained. If the lengths are adjusted for cancellation with a stationary source, then cancellation will not occur for a moving source. This apparatus may be considered to be a modified Rayleigh interferometer. For a distance of travel of one meter, one cycle shift in the cancellation point will be caused by difference in velocity of approximately 5×10^{-7} c for green light. This is about 300 mph or about 500 ft/sec. This is much too sensitive to be used on the moon or a satellite for measuring light from the sun, but since it should be easy to detect 1/10 of a cycle shift, a simple vibrating source (which might be the glass condensing lens of the interferometer through which light passes or a mirror from which it is reflected, as well as a direct source) driven by a loud-speaker element might be used. 1000 cps of $\frac{1}{4}$-inch amplitude gives approximately 130-ft/sec peak velocity. The output of a photocell placed in the interference area of the device would have an ac output depending on the movement of the source. As with any interferometer a visual pattern may be obtained. Then motion of the source will displace the pattern and a vibrating source will visually smear the pattern. The size of such a device is such that it could be quite reasonably placed in the vacuum of a conventional vacuum system.

An experiment of this type could add to our knowledge of the behavior of light. A positive result would indicate definitely that

* Received March 26, 1962.
[1] P. M. Rapier, "A proposed test for the existence of a Lorentz-invarient aether," Proc. IRE, vol. 50, pp. 229–230; February, 1962.

the velocity of light was constant relative to the source alone and was independent of the receiver. A negative result would be more ambiguous but would at least eliminate several possibilities.

ALFRED C. SCHROEDER
RCA Labs.
Princeton, N. J.

Intermodulation Noise in FM Troposcatter Links*

The purpose of this note is to give an approximate formula for the intermodulation noise produced by tropospheric scattering.

Measurements[1] on one path show that for small frequency deviations the signal-to-intermodulation ratio S/I as a function of peak deviation P and the highest baseband frequency F can be represented by

$$S/I = \frac{k}{(PF)^2} \tag{1}$$

where k is a constant which depends on effective distance and antenna beamwidth, and S/I is measured in the highest baseband channel. Since (1) and the shape of the measured curves for large deviations agree with the results obtained both experimentally[2] and theoretically[3] for the intermodulation produced by one delayed echo, it appears reasonable to extend these results to the troposcatter case. From Bennett, *et al.*,[3] for peak deviations much smaller or larger than the reciprocal of the echo time T,

$$S/I = \frac{k_1}{(PFT^2)^2} \qquad PT \ll 1 \tag{2a}$$

$$S/I = k_2(P/F)^3 \qquad PT \gg 1. \tag{2b}$$

Combining these gives

$$S/I = \frac{[1 + k_3T^4P^5]}{F}. \tag{3}$$

Using the measurements[1] to evaluate the constants gives

$$S/I = \frac{5700}{(PF)^2}(T_0/T)^4\,[1 + (1/F)(T/T_0)^4(P/9.4)^5]. \tag{4}$$

P and F are in Mc/sec, and $T_0 = 8\times10^{-8}$ secs. The minimum value is obtained when the second factor in square brackets in (4) is equal to 5/3. The equivalent time delay T is

$$T = (d_a/c)(d/R)_e^2\frac{(\alpha/\theta)^2}{0.92 + 4.1(\alpha/\theta)^2} \tag{5a}$$

$$= \frac{(d_a/c)\alpha^2}{0.92 + 4.1\,\alpha^2/(d/R)_e^2}. \tag{5b}$$

where

d_a = actual distance between transmitter and receiver,
c = velocity of light,
α = antenna beamwidth in radians (equal transmitting and receiving beamwidths),

$$(d/R)_e = \frac{\text{effective distance}}{\text{effective earth's radius}},$$

$\theta = (d/R)_e$ = nominal scatter angle, the angle between 1) a plane tangent to the earth and passing through the center of the transmitting antenna, and 2) a similar plane for the receiving antenna.

Eq. (5), adapted from Prosin,[4] takes into account the antenna beamwidth as well as the effective scattering volume. The above equations are for slot noise tests, pre-emphasis, no diversity, equal antenna beamwidths, and for a path which is approximately symmetrical (effective scattering volume not radically separated from the midpoint of the path). It should be noted that with pre-emphasis the lowest signal-to-intermodulation ratio is not necessarily obtained at the maximum baseband frequency.[3] In calculating T_0 for the Florida-Cuba path the effective distance was taken to be the actual distance, the effective earth's radius was taken as 2.4 times the actual radius,[5] and the antenna beamwidth was taken to be 2.55×10^{-2} radians.

Two paths in another part of the world have the following parameters:

TABLE I

Actual distance d_a	440 miles	194 miles
Effective distance d_e	393 miles	124 miles
Effective earth's radius R_e	5300 miles	5300 (assumed)
Antenna beamwidth α	1.25×10^{-2}	3.8×10^{-2}
Highest baseband frequency F	0.3 Mc	0.3 Mc

The measured and calculated values of the signal-to-intermodulation ratio are shown in Fig. 1. The discrepancy of 6 db between measurements and calculations for path A may indicate that (5) requires modification. However, more measurements on long paths are required to confirm this.

One method of reducing the intermodulation for $PT \ll 1$ is to shift the multiplex frequency band so that it occupies less than an octave in frequency.[2] (Thus, for a 0–300 kc system, shift the band to 300–600 kc.) The S/I ratio should be improved by a large factor.

For $\alpha/(d/R)_e \gg 1$, (5b) and (4) show that the signal-to-intermodulation ratio varies as the eighth power of $(d/R)_e$ so that the ratio will decrease by large factors during the winter. On the other hand, for $\alpha/(d/R)_e \ll 1$, the signal-to-intermodulation ratio does not depend on $(d/R)_e$. Experimental data are not available to verify these conclusions.

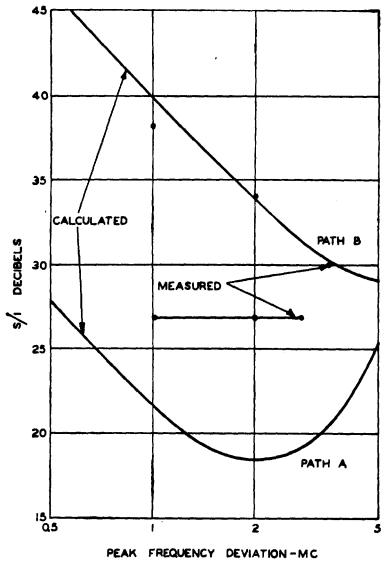


Fig. 1—Signal-to-intermodulation ratio vs frequency deviation. Parameters for the paths are given in Table I.

The theoretical formula given by Prosin[6] has the same form as (2a) for $PT \ll 1$, but the numerical coefficient is 160 times larger than the coefficient in (4). For $PT \gg 1$ the exponential factor in Prosin's formula becomes important, but only for very high deviations (over 100 Mc for the Florida-Cuba path).

W. SICHAK
R. T. ADAMS
Sichak Associates
Nutley, N. J.

Stimulated Emission from Holmium Activated Silicate Glass*

A program of studying the optical properties of very pure silicate glasses activated with rare earth ions has been in progress for over a year now at this laboratory. One of the purposes of this work is to investigate stimulated emission processes in these glasses. The preparation of the glasses, the fabrication of the etalons and the experimental measurement techniques have been reported previously.[1] The present communication describes some preliminary investigation recently made on holmium activated silicate glass.

* Received March 26, 1962. This work was sponsored in part by the Federal Electric Corp., a subsidiary of the International Telephone and Telegraph Corp., under Contract AF 19(626)-21.

[1] C. E. Clutts, R. N. Kennedy, and J. M. Trecker, "Results of bandwidth tests on the 185 mile Florida-Cuba tropospheric scatter system," IRE TRANS. ON COMMUNICATIONS SYSTEMS, vol. CS-9, pp. 434–439; December, 1961.

[2] W. J. Albershein and J. P. Schafer, "Echo distortion in the FM transmission of frequency-division multiplex," PROC. IRE, vol. 40, pp. 316–328; March, 1952.

[3] W. R. Bennett, H. E. Curtis, and S. O. Rice, "Interchannel interference in FM and PM systems under noise loading conditions," *Bell Sys. Tech. J.*, vol. 34, pp. 601–636; May, 1955.

[4] A. V. Prosin, "Calculation of cross noise power in long distance scatter communications systems," *Radio Engrg. and Electronic Phys.*, vol. 6, pp. 9–18, eq. (46); January, 1961.

[5] R. E. Gray, "Tropospheric scatter propagation and meteorological conditions in the Caribbean," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-9, pp. 492–496; September, 1961.

[6] Prosin, *op. cit.*, eq. (33).

* Received August 2, 1962.

[1] H. W. Etzel, H. W. Gandy, and R. J. Ginther, "Stimulated Emission of Infrared Radiation from Ytterbium Activated Silicate Glass," NRL Progr. Repts., pp. 27–28; February, 1962.

Stimulated emission from single crystal etalons of $CaWO_4$ activated with trivalent holmium at liquid nitrogen temperatures has been reported by Johnson, *et al.*[2] To this date, this is the only matrix in which stimulated emission from holmium has been reported and in $CaWO_4$, the emission occurs at a wavelength of 2.046 μ.

Stimulated emission of radiation has been observed in a holmium activated $LiMgAlSiO_3$ glass at liquid nitrogen temperature. This glass is of the same base composition as the glass reported previously,[1] but contains approximately one cationic mole per cent of holmium. An oscilloscopic trace of the stimulated emission from the holmium activated glass etalon as detected by a cooled InSb photovoltaic detector is shown in Fig. 1. In this experiment, a germanium filter was placed between detector and laser rod. From a knowledge of the filter's spectral transmission, it is known that the stimulated emission here occurs at wavelengths greater than 1.95 μ. It should be noted that strong well-defined bursts of stimulated emission commence about 350 μsec after the peak of the excitation pulse; this time delay is not surprising since the spontaneous lifetime, of this glass at liquid nitrogen temperature has been measured to be 700 ± 200 μsec. The shape of these spikes indicate that the glass matrix is not of high optical quality; this observation is confirmed by direct visual examination of the etalon. It is estimated that for the trace in Fig. 1, the etalon is being excited approximately 5 to 10 per cent over threshold which is about 3600 joules using a rather inefficient helical flash tube excitation method. With better optical coupling of a standard type, it is expected that the threshold would be at least an order of magnitude less.

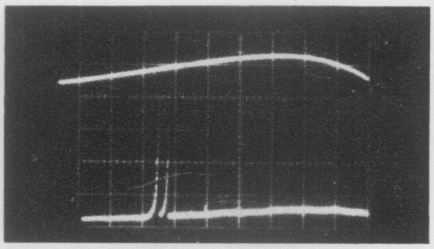


Fig. 1—Stimulated emission from holmium in silicate glass at a temperature of 78°K; excitation trace at top, laser beam trace at bottom. Writing speed is 100 μsec/cm; I_nS_b detector with germanium filters in place.

Room temperature optical absorption measurements indicate strong absorption maxima at 3615A and 4490A with absorption coefficients of 6 and 8.5 cm^{-1}, respectively; the 4490A absorption band half width is about 200A. It is believed that the principal optical pumping spectral region is from 4400 to 4600A from the similarity of the glass absorption spectrum with that of the $CaWO_4$. From the nature of the absorption spectrum and the method of preparation, it is believed that the holmium in the glass is trivalent. Following Johnson,[2] it is supposed that the stimulated emission results from the $^5I_7$ to $^5I_8$ transitions in the 4 f^{10} electronic configuration of the ion.

It should be noted that holmium is only the third ion which has been reported to exhibit laser action in a glass of any type. The first ion reported was trivalent neodymium in barium crown glass fibers[3] and the second was ytterbium in $LiMgAlSiO_3$ base glass rods of the same composition used in this work.[1] Work is now proceeding in order to improve our understanding of the stimulated emission of holmium in this glass.

H. W. Gandy
R. J. Ginther
U. S. Naval Research Lab.
Washington, D. C.

[2] L. F. Johnson, G. D. Boyd, and K. Nassau, "Optical maser characteristics of Ho^{+3} in $CaWO_4$," Proc. IRE, vol. 50, pp. 87–88; January, 1962.

[3] E. Snitzer, "Optical maser action of Nd^{+3} in a Barium crown glass," *Phys. Rev. Lett.*, vol. 7, pp. 444–446; December, 1961.

Simultaneous Laser Action of Neodymium and Ytterbium Ions in Silicate Glass*

Stimulated emission of radiation has been reported for three different ions in glasses. The first reported was neodymium in barium crown glass fibers,[1] the second was ytterbium in $LiMgAlSiO_3$ glass rods,[2] and the third was holmium in $LiMgAlSiO_3$ glass rods.[3] Laser action from neodymium had previously been reported for several matrices, the first being single crystal $CaWO_4$.[4] Trivalent neodymium represents a particularly favorable ion for laser action because the terminal energy level is essentially unpopulated at room temperature, and population inversion of the emitting energy level in this ion is relatively easy. The stimulated emission in the 4 f^{13} electronic configuration of trivalent ytterbium in silicate glass has been ascribed to transitions between the $^2F_{5/25}$ and $^2F_{7/2}$ multiplets.[2] The present work describes the investigations of a glass in which both neodymium and ytterbium are incorporated.

Stimulated emission has been observed simultaneously from two different ions contained in the same $LiMgAlSiO_3$ glass etalon while being operated at liquid nitrogen temperature. This base glass was activated with one cationic mole per cent neodymium and one cationic mole per cent ytterbium and from their optical absorption spectra both are believed to be predominantly in the trivalent state. The absorption spectra for this glass appears to be a mere superposition of the spectra of the two ions when they are separately incorporated in this base glass composition; this result suggests that there is no strong interaction between the two incorporated rare earth ion species. It is possible, however, that an interaction between the two different ions may not be detected most sensitively by optical absorption.

An oscilloscopic record of the stimulated emission from the doubly activated glass etalon is shown in Fig. 1. Stimulated emission commences about 220 μsec after initiation of the flash excitation and occurs near the peak of the excitation pulse. This result suggests that the lifetime for the species causing the initial laser action is on the same order as the rise time of the exciting pulse. The spontaneous lifetime for singly activated Nd^{+3} in this glass is about 350 μsec and this ion is believed responsible for the initial stimulated emission observed. Subsequent experiments with a monochromator placed between laser and detector and set at the Nd^{+3} stimulated emission wavelength (1.06 μ) confirm this conclusion. In Fig. 1, heavy spiking occurs for Nd and lasts for almost 1200 μsec because the etalon is being excited approximately 350 per cent over the Nd threshold in this glass.

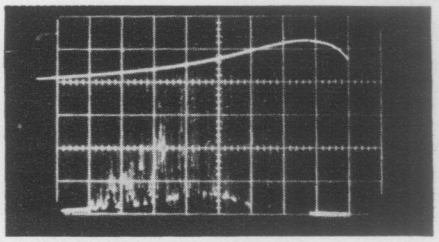


Fig. 1—Stimulated emission from neodymium and ytterbium in silicate glass at 78°K; excitation trace at top, laser trace at bottom. Writing speed is 200 μsec/cm. Sharp cut filter in place passing radiation of wavelengths greater than 1.0 μ.

In Fig. 1, an envelope commences to rise about 600 μsec after lamp flash initiation, and lasts for some 1000 μsec; the envelope peaks about 700 μsec after the maximum of the excitation pulse. This delayed emission is about what is observed in the singly activated ytterbium glass; there the lifetime for spontaneous emission is about 1450 μsec. Monochromatic investigation of this component of the emission in Fig. 1 indicates that it takes place at a wavelength of 1.015 μ and hence is due to ytterbium. The type of trace observed for the ytterbium component in this glass is the envelope as shown with spikes superimposed thereon; this emission pattern is characteristic of a fair optical quality glass etalon which is being excited just over threshold.[5] The trace in Fig. 1 is that obtained when the etalon is excited approximately 10 per cent over the ytterbium threshold. During the ytterbium (stimulated emission) period, the stimulated emission from the neodymium still persists and may be seen in Fig. 1.

Another effect observed in the present work is an apparent "fatigue effect" in the laser output from neodymium. It was found that when the etalon was excited at lamp

* Received August 2, 1962.

[1] E. Snitzer, "Optical maser action of Nd in a barium crown glass," *Phys. Rev. Lett.*, vol. 7, pp. 444–446; December, 1961.

[2] H. W. Gandy, H. W. Etzel, and R. J. Ginther, "Stimulated Emission of Infrared Radiation from Ytterbium Activated Silicate Glass," NRL Progr. Repts., pp. 27–28; February, 1962.

[3] H. W. Gandy and R. J. Ginther, this issue, pp. 2113–2114.

[4] L. F. Johnson and K. Nassau, "Infrared fluorescence and stimulated emission of Nd+3 in $CaWO_4$," Proc. IRE, vol. 49, pp. 1704–1706; November, 1961.

[5] The interpretation of this type of oscilloscope trace and its cause will be the subject of a future report.

input energies several times the neodymium threshold, the neodymium stimulated emission decreased and finally disappeared at this temperature, whereas the ytterbium emission did not. A study of this effect will be reported in the near future.

It is believed that this is the first reported work in which two different ion species have simultaneously exhibited laser action in the same host material.

H. W. Gandy
R. J. Ginther
U. S. Naval Research Lab.
Washington, D. C.

Direct Measurement of Optical Cavity Q*

The direct measurement of the bandwidth of an optical cavity resonator, in this case that of our unexcited Mark II gas maser, was undertaken as a part of a larger program involving coupled masers. The Mark II optical cavity is dimensionally identical to that of the Mark I maser furnishing the frequency-swept probe signal, both cavities being one meter long and being bounded by nominally 99 per cent reflecting plane parallel mirrors. Preliminary experiments showed the longitudinal mode frequency spacings between the two cavities to differ by no more than 50 kc out of the basic 150 Mc spacing. This gave assurance that the expected bandwidth measurements (0.3–1.2 Mc) would not be greatly perturbed by a vernier-type line broadening due to oscillation in more than one longitudinal mode by the probe maser. The same order of error arose through microphonically-induced jitter of the probe signal as well as the responding cavity.

Light waves being unguided, it was necessary in this experiment to physically align Mark I and II to extremely close angular tolerances. That is, the angular acceptance cone of Mark II as a receiving device is no doubt closely related to its beamwidth as a maser oscillator, in the order of 30 sec. Optimum transmission through Mark II, however, would probably occur only for some central portion of the cone, say 5 sec, and therefore the problem was to find and fix the optical axes of the two masers to within this angle.

An end-to-end configuration of the two masers was used for this purpose with the ultimate alignment being done by either using the mechanical hysteresis of the support table or by using the principle of a high compliance spring working against a low compliance one to obtain the fine motions required. The alignment indicator was taken to be the freedom of the detected Q-curve-like signals from off-axis modes.

Gross frequency alignment of the longitudinal modes between masers was accomplished by utilizing controllable direct current energization of a set of magnetostrictive tuning coils, one of which was wound on each of the four main Invar frame rods[1] of the Mark I maser. The tuning rate can be given roughly as 150 Mc/600 ampere-turns/coil. The oscillating maser was also swept over about a two megacycle frequency interval at 60 cps through the application of additional ac current to these same coils.

Frequency calibration of the Mark I maser was done by observing the photomixed beat-note on a spectrum analyzer between the fixed and oscillating Mark II laser and Mark I, when the latter was tuned either in a point-by-point manner with direct current or by its 60∼sweep. Fig. 1 illustrates the usual saddle-shaped power spectrum obtained with 60∼sweep, where the definition has been somewhat impaired by microphonic jitter. We found that the chosen sweep width was a good compromise between an upper limit determined by the detector response time and a lower limit set by mechanical and thermal stability considerations.

With adequate spatial and frequency alignment between the oscillating and receiving masers, the frequency modulated probe signal passed through Mark II and traced out its transmission characteristics via an Ektron detector and an oscilloscope. Fig. 2 affords an example of the best appearing Q curves. From this and pictures like it we were able to determine that the cold cavity bandwidth under optimum conditions was about 0.3 Mc. More often the bandwidth was measured to be about 0.35–0.4 Mc. The passive Q, incidentally, is very nearly one billion.

If we neglect diffraction losses, the mirror reflectance coefficient calculated from the above bandwidth value turns out to be 99.4 per cent. One has in this technique an elegant method of determining mirror reflectance coefficients in the difficult >99 per cent region.

Roy A. Paananen
Research Div.
Raytheon Co.
Waltham, Mass.

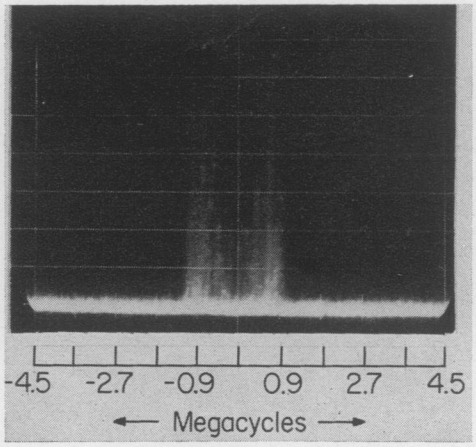


Fig. 1—Power spectrum of frequency-swept maser.

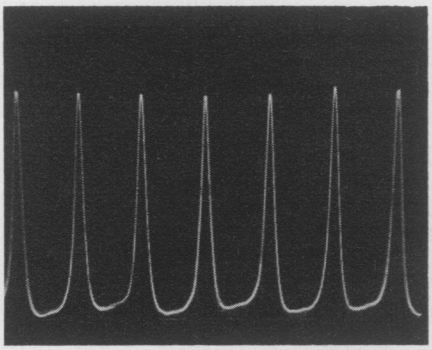


Fig. 2—Q-curves representative of the best results.

[1] C. F. Luck, R. A. Paananen, and H. Statz, "Design of a helium-neon gaseous optical maser," Proc. IRE (*Correspondence*), vol. 49, pp. 1954–1955; December, 1961.

* Received August 1, 1962.

Resonant Amplification in a Gas Maser*

The optical gain and the manner in which the gain varies with input signal level were measured on our Mark II gas maser operated at selected excitation levels up to the edge of oscillation point. In this mode of usage the Mark II amplifier could be expected to exhibit, along with gain, the special features of a resonant device, namely an extremely narrow amplifying bandwidth, and in this case, extremely narrow spatial acceptance angles.

A few general remarks are in order concerning the possible varieties of gas laser amplifiers. The central problem to be overcome in such devices is the very low specific gain γ in the equation, $P_{out} = P_{in}e^{\gamma z}$. For ordinary discharge tube diameters and pressures (of He and Ne), γ might be 0.05/m. To achieve a gain of 10, say, one would then need to traverse a gas column 46 m long. In the absence of breakthroughs towards higher γ values, all gas laser amplifiers with significant gains must, in one way or another, prove this order of column length.

The dictates of practicality suggest then that the varieties of gas laser amplifiers will be one-to-one with the ways in which one can fold an optical path.

Non-resonant gas maser amplifiers are perfectly feasible and in fact are the subject of current research. Here, however, we will only consider a feedback or resonant form, whose geometry takes advantage of the fact that there is no restriction that the same gas path cannot be traversed more than once. In return for this spatial economy, nature exacts a toll in the stringent relation between the length of the optical cavity W and frequency for any sort of transmission at all through the maser. The form of such a gas maser amplifier would be simply that of a gas maser oscillator excited below the start-oscillation level.

The on-resonance transmission through an amplifier like this is given by (1) below.

$$T_r = \frac{T^2 A_{s.p.}}{(1 - A_{s.p.}R)^2} \qquad (1)$$

* Received August 1, 1962.

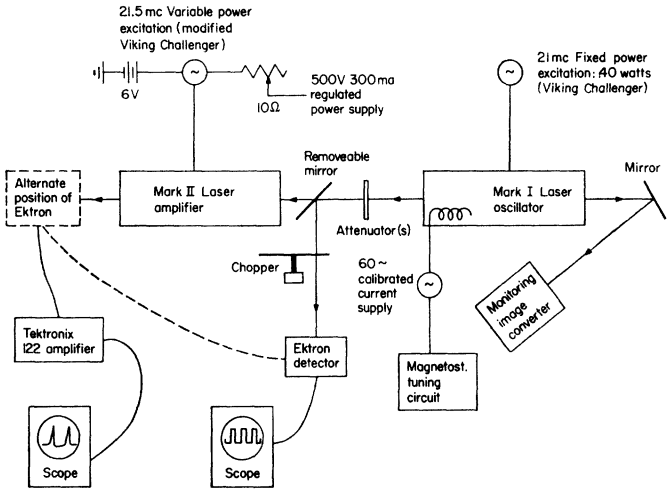


Fig. 1.

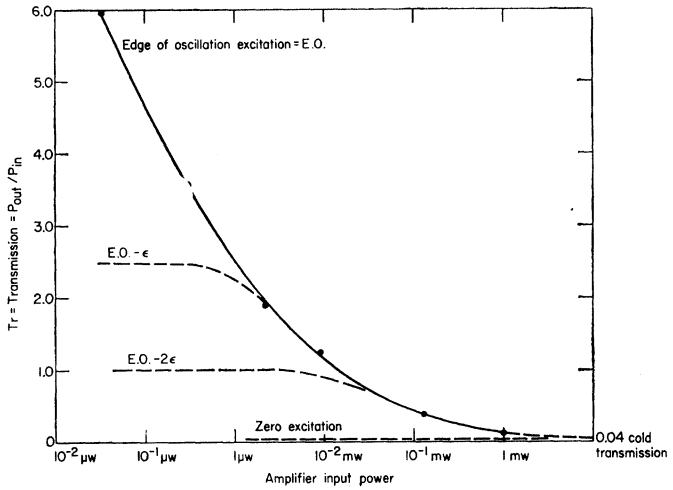


Fig. 2.

where the values in a specific case might be

$$T = 2\cdot 10^{-3}$$

(power transmission coefficient of one mirror)

$$R = 0.994$$

(power reflection coefficient)

$$\therefore\ A = 4\cdot 10^{-3}$$

$(A+R+T=1)$

$$W \cong 1 \text{ m}$$

(max. excitable length about 4/5 m)

$$A_{s.p.} = 1.00 - 1.04$$

(single pass power gain, dependent upon excitation level)

$$T_r = 10$$

(for example).

Then,

$$(1 - A_{s.p.}R) = T\sqrt{\frac{A_{s.p.}}{T_r}} \cong 0.6\cdot 10^{-3} < 0.1 \text{ per cent.}$$

That is, the $A_{s.p.}$ value must be set and maintained to within less than 0.1 per cent of the start oscillation value to achieve a gain of 10 with the chosen mirror coefficients.

Some details of the inter-maser space and frequency alignment procedures necessary to perform coupled maser experiments are given in a companion letter (Direct Measurement of Optical Cavity Q). Fig. 1 illustrates the particular instrumentation found suitable by us for measurements of optical gain and gain saturation.

Most of the essential results of this work are reproduced as the curves of Fig. 2. The solid line represents the exact edge-of-oscillation gain data for about a $3\cdot 10^4$ power input range. Achieving gains greater than 6 was not feasible due to the instability of the regeneration calculated earlier. The dotted curves are indicative of the gain behavior for excitations not at the E.O. point, culminating, of course, in the flat cold transmission, here 4 per cent.

A complete explanation of the gain saturation behavior in Fig. 2 has not yet been derived. Eq. (2) below gives, without any theoretical justification, a reasonable fit to the E.O. curve.

$$P_{out} = 0.04P_{in} + 0.085\sqrt{P_{in}} \quad (P_{in} \text{ in mW}). \quad (2)$$

Much more satisfactory gain figures are predicted by (1) if mirrors designed for amplifier, rather than oscillator, service are utilized. Gains of several hundred seem to be possible for R values close to 0.96. Work in this direction is continuing.

Roy A. Paananen
Research Div.
Raytheon Co.
Waltham, Mass.

Junction Transistor with Reverse-Biased Input*

It was shown earlier[1] that a junction transistor can be operated as a voltage amplifier with its input (base-emitter junction) reverse biased. Although, this results in some loss in voltage gain, still, the gain remains sufficiently above unity. With the transistor type CK721 (Raytheon), it was possible to have a voltage gain of 20 to 30 db. The reverse-biased operation gives the special advantages of high input impedance and reduced feedback.

With present-day transistors the signal handling capacity of the reverse-biased input for faithful amplification is rather low (about 100 mv), compared to the forward-biased case. But with some modifications in the transistor construction[1] the situation is expected to improve.

With forward biasing, the output (collector) current (I_c) is found to increase approximately linearly with input (base) current (I_b), rather than the input voltage (V_b). [Considering a grounded emitter configuration.] This justifies the use of current gain factor $\alpha_e(=\partial I_c/\partial I_b)$ as the relevant transfer function.[2] Because, under such operation α_e, rather than transconductance $g_e=(\partial I_c/\partial V_b)$, remains constant over a wide range. But the situation is completely different with reverse-biased operation. Here, the input (base) reverse current is negligibly small (a few microamperes) and does not show a linear relation with the output current (I_c). Under such an operation the output current I_c is found to vary linearly with input voltage V_b (Fig. 4 in Chatterjee[1]), rather than input current I_b. As such, the use of a transconductance $g_e(=\partial I_c/\partial V_b)$ seems more appropriate in this case. With reverse-bias operation, the current amplification factor α_e loses its significance and rather the transconductance (g_e) is to be used for calculating the amplifier gain. With the transistor type CK721 (Raytheon) and with a collector voltage of 20 v (negative), the transconductance was found to be about 230 μmhos. This value of transconductance compares favorably well with those in voltage amplifier vacuum triodes and reasonable voltage gain can be obtained. It may be

* Received, March 2, 1962.

[1] B. Chatterjee, "A transistor voltage amplifier with reverse-biased input," *J. Inst. Telecom. Eng.* (India), vol. 7, pp. 249–254; November, 1961.

[2] "Cathode Ray," *Wireless World*, vol. 67, pp. 490; September, 1961.

mentioned here that with the input reverse biased, a transistor can be safely operated with a higher collector voltage than specified for forward-biased operation.

It may not be irrelevant to mention that a transistor with its input reverse biased is primarily useful as a voltage amplifier—especially where a high input impedance is desirable. Because of their low output current, they are not capable of delivering the maximum available power. Power amplifiers may preferably be used with a forward bias —as in high power vacuum tube amplifiers, where the grids are often driven positive to extract the maximum available power.

B. CHATTERJEE
Communication Engineering Department
Indian Institute of Technology
Kharagpur, India

Measurement of Bulk-Lifetime by Microwave Electrodeless Technique with Steady Illumination*

Jacobs, *et al.*,[1] have suggested a method for the measurement of bulk-lifetime of carriers in semiconductors in which the sample in the form of a rod is inserted in a waveguide with its axis perpendicular to the broad face of the guide. The sample is illuminated at one end by a steady source of light. Change in absorption of microwave on illumination for different values of the length d between the illuminated end-face and the point of entry within the guide is measured and the diffusion length is determined from the slope of the curve giving the logarithms of the change in absorption as a function of d. As the authors point out the method is fruitful provided 1) the distribution of excess carriers along the axial direction of the sample is exponential, 2) change in absorbed power is linearly related to the integrated excess carrier density inside the guide, and 3) $\sigma \ll \omega\epsilon$. It is further stated that the complication arising out of surface recombination may be avoided with this method. In actual practice assumption 2) may be satisfied, if, as stated by the authors, the level of illumination is low. Assumption 3) may be fulfilled through careful choice of material and ω. The validity of assumption 1) and the effect of surface recombination are, however, matters of more detailed discussion. In the present communication this aspect of the problem is discussed.

A sample that can be treated as one-dimensional is considered and the incident radiation is assumed to be monochromatic. Further, the sample is assumed to possess a thin surface layer where recombination occurs profusely. A part of the incident radiation flux is absorbed in the thin layer in a manner characteristic of the layer itself and the remainder penetrates inside the sample following a law characteristic of the bulk-absorption properties of the material. The steady-state continuity equation may, therefore, be written as

$$\frac{\partial^2 p}{\partial x^2} - \frac{p}{L^2} = -\frac{\alpha}{D}\eta G_1' e^{-\alpha x} = -\frac{\alpha}{D} G_1 e^{-\alpha x} \quad (1)$$

where p, L, D, α, η have the usual meaning. The quanta of light penetrating into the volume is G_1'.

The excess carrier density p should satisfy the boundary conditions

$$-Sp\big|_{x=0} + G_2 = -D\frac{\partial p}{\partial x}\bigg|_{x=0}; \quad p \to 0,\ x \to \alpha$$

where S=recombination velocity and G_2= carriers generated on the surface. On solving (1) the total number of carriers in the region of the sample inside the waveguide is found to be

$$p' = \frac{G_1\tau}{(\alpha^2L^2 - 1)}\left[\frac{D\alpha + S}{\frac{D}{L} + S}\alpha L(1 - e^{-b/L})e^{-d/L} - (1 - e^{-\alpha b})e^{-\alpha d}\right] + \frac{G_2 L}{\frac{D}{L} + S}(1 - e^{-b/L})e^{-d/L} \quad (2)$$

where b=smaller dimension of the waveguide.

It would be convenient to consider separately three special cases.

Case I

$$\frac{1}{L} \ll \alpha, \quad i.e., \quad \alpha L \ll 1.$$

For this case the term containing $e^{-\alpha d}$ may be neglected and the slope of the ln p' vs d curve would give L.

Case II

$$\frac{1}{L} \gg \alpha, \quad i.e., \quad \alpha L \ll 1.$$

Under this condition the first and the third term may be neglected compared to the second. The final slope of *ln* p' vs d curve, in this case, gives the absorption constant α.

From the available experimental data[2] it would appear that G_e with visible irradiation would favor Case I, whereas G_e with infrared radiation would favor Case II.

Case III

$$\frac{1}{L} \doteqdot \alpha, \quad i.e., \quad \alpha L \doteqdot 1.$$

A linear relation may be obtained in this case only for high values of d. The final slope of the *ln* p' vs d curve might give either α or $(1/L)$, whichever is smaller. However, for large values of S and for $\alpha L \doteqdot 1$, the plot of *ln* p' as a function of d would have zero slope for values of d comparable to L. Further, in some cases, the curve might have a maximum, in which its position and the maximum value of normalized p' may be used to obtain α and L using (2). Alternatively, if the maximum is not realized, α and L may be determined by comparing the values of p' for $d=L$ and $d=2L$.

It is to be noted that the curves of Jacobs, *et al.*,[1] show in one case a zero slope and in the other a maximum. They ascribe this deviation from the expected linear plot to experimental errors. The above discussion offers an alternative explanation if it is assumed that αL is of the order of unity in the experimental condition. One may then determine α and L as discussed above. Assuming that the final slope gives $1/L$, the values of α obtained in both the cases are found to be 4. On the other hand, if it is assumed that the final slope gives α, the value of lifetime comes out to be about 1 msec, which agrees reasonably well with those given by the photo-conductive decay method. In both the cases the product αL is found to be closely equal to unity. It must, however, be remembered that the experimental data were probably obtained with a composite beam of light. The α value as found from the above analysis should therefore be regarded as an effective value.

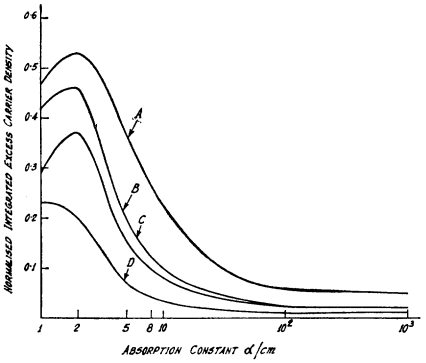


Fig. 1—Integrated excess carrier density for different values of the absorption constant. A—L =0.2 cm, $d=L$, B—L =0.2 cm, $d=2L$, C—L =0.5 cm, $d=L$, D—L =0.5 cm, $d=2L$.

The rather low values of α as found above is also corroborated by Fig. 1, which gives plots of normalized p' as a function of α for $d=L$ and $d=2L$ with $S=2000$; $b=1$ cm, $L=0.5$ and 0.2 cm. It is found that p' attains a maximum for values of α in the range 1–10. The available data on G_e show that α values in the range 1–10 are obtained with infrared radiation of wavelength 1–2 μ[2]. It may be noted that for an ordinary tungsten lamp, radiation of this range of wavelength would constitute about 30 per cent of that in the visible range for which α values are much higher.[3] If such a lamp is used as the source of irradiation then one can easily show, with the help of Fig. 1, that the analysis of experimental data would yield a low value of α and a value of αL roughly equal to unity.

In the light of the above discussions one may conclude that the method of Jacobs, *et al.*,[1] gives L correctly irrespective of the value of S if $\alpha L \gg 1$. If, however, $\alpha L \doteqdot 1$, the data may be used to obtain α and L if S is

* Received February 5, 1962.

[1] H. Jacobs, A. P. Ramsa and F. A. Brand, "Further consideration of bulk lifetime measurement with a microwave electrodeless technique," PROC. IRE, vol. 48, pp. 229–233; February, 1960.

[2] M. L. Schultz and G. A. Morton, "Photoconduction in germanium and silicon," PROC. IRE, vol. 43, pp. 1819–1828; December, 1955.

[3] V. K. Zworykin and E. G. Ramberg, "Photoelectricity and its Application," John Wiley and Sons, Inc., New York, N. Y., p. 18; 1950.

large. On the other hand, if $\alpha L \ll 1$ the method does not give L, but α.

The authors would like to express their sincere thanks to Prof. J. N. Bhar for his kind interest in the work.

S. Deb
B. R. Nag
Inst. of Radio Phys. and Electronics
University College of Tech.
Calcutta, India

Measurements of Probability Densities of Small Ensembles of Periodic Waveforms*

We wish to demonstrate with measured data some interesting properties of finite ensembles which have not to our knowledge been explicitly illustrated in the literature. With but very few exceptions[1,2] the statistics of either large ensembles (where the central limit theorem is applicable) or of single periodic waveforms are discussed.

Presented here is a set of measurements that were obtained for the summation of two and three periodic waveforms with various energy ratios.[3] These measurements have been made with advanced instrumentation techniques[4] which yield a high degree of accuracy.

Fig. 1 presents experimental measurements of the probability density for the sum of two independent sine waves as the energy ratio is varied. Curve (a) is a single sine wave and is represented by an energy ratio of 0:1. Curve (b) illustrates the large effect on the probability density of a small perturbation in the form of an additional waveform of much lower energy. In the equal energy case (d) the characteristic form is observed as reported by Slack.[1]

Fig. 2 presents density curves for the summation of three independent cosine waves where the energy of one wave is greater than or equal to the others. This illustrates an interesting concept in probability theory. The rate of convergence to a Gaussian type of distribution is influenced by the energy ratio. The convergence is most rapid when the energies are equal.[5]

Fig. 3 indicates the density function resulting when a square wave and an independent sine wave are combined (equal energy). Four modes dominate the function. These are at the maxima and minima of the resulting waveform. But yet when a third independent wave of equal energy with the first two is added, the strong modal structure disappears. Fig. 4 indicates this effect when an equal-energy triangular wave is added. Comparison with the probability density of the summation of three equal cosine waves suggests that the convergence to Gaussian may be more rapid when the summation includes nonidentical waveforms.

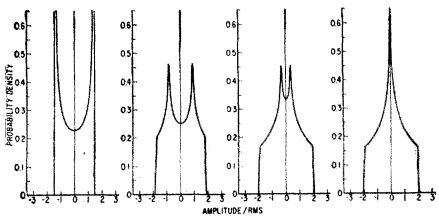


Fig. 1—Probability densities representing the sum of two independent cosine waves for various energy ratios. (a) 0:1. (b) $\frac{1}{8}$:1. (c) $\frac{1}{2}$:1. (d) 1:1.

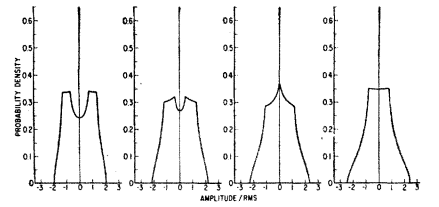


Fig. 2—Probability densities representing the sum of three independent cosine waves for various energy ratios. (a) $\frac{1}{16}:\frac{1}{16}$:1. (b) $\frac{1}{8}:\frac{1}{8}$:1. (c) $\frac{1}{4}:\frac{1}{4}$:1. (d) 1:1:1.

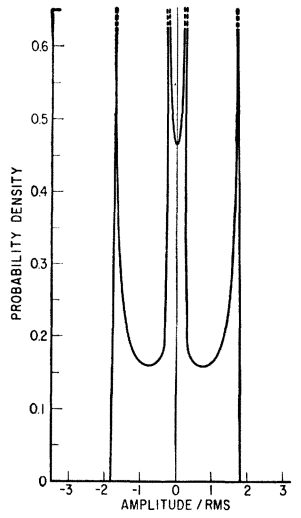


Fig. 3—Probability density of the sum of independent square and sine waves for equal energy ratios.

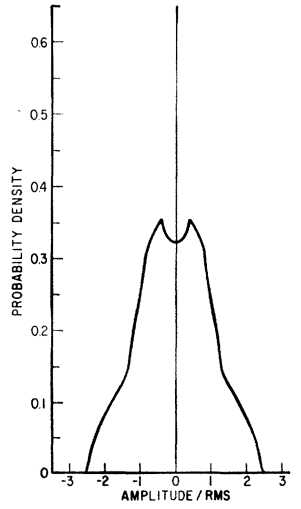


Fig. 4—Probability density of the sum of independent square, cosine, and triangular waves for equal energy.

The data described herein serves to indicate the wide variety of probability density functions that exist between the realms of single periodic waveforms and the summation of very large numbers of these waveforms or Gaussian distributions.

H. L. Fox
E. A. Starr
Bolt Beranek and Newman, Inc.
Cambridge, Mass.

Temperature Sensitivity of Twoport*

The use of a noise figure in describing the noisiness of a twoport becomes less significant in low-noise devices which are operated at temperatures much below room temperature. For this reason the "effective input noise temperature" was introduced.[1]

In a recent communication,[2] Fisher defines the "excess noise temperature" T_r of a twoport by

$$T_r = \frac{\Delta P_{eoi}}{kA_e\Delta_f}, \tag{1}$$

where A_e is the exchangeable power gain and ΔP_{eoi} is the exchangeable power output due only to the noise sources internal to the twoport. According to the definition of the "effective input noise temperature," it is the temperature of an input termination which, when it is connected to the noise-free twoport, would result in the same noise power as that of the actual twoport connected to a noise-free input termination. Hence the effective input noise temperature T_e is given by

$$T_e = \frac{\Delta P_{eoi}}{kA_e\Delta_f} \tag{2}$$

yielding the identical result as (1). Thus

$$T_r = T_e. \tag{3}$$

Similarly the "temperature sensitivity" T_s of a twoport with a particular termination having an "equivalent noise temperature" T_{eq} can be expressed, using the input noise temperature, by

$$T_{et} = T_e + T_{eq} \equiv T_s \tag{4}$$

where T_{et} is a total equivalent input noise temperature. Writing T_{eq} in terms of noise

* Received February 7, 1962; revised manuscript received, February 23, 1962.

[1] M. Slack, "The probability distribution of sinusoidal oscillations combined in random phase," *J. IEE*, vol. 93, pt. 3, pp. 76–86; March, 1945.

[2] T. Sheridan, "Time-Variable Dynamics of Human Operator Systems," AFCRC-TN-60-169; 1960.

[3] H. L. Fox and E. A. Starr, "The statistics of finite ensembles," *J. ASA*, vol. 33, p. 1679; November, 1961.

[4] The measurements were made using a B+K Instruments Probability Density Analyzer Model 160.

[5] This is the "weak" or "equal-component" central limit theorem vs Liapounov's theorem; see M. Loeve, "Probability Theory," D. Van Nostrand Co., Inc., New York, N. Y.; 1955.

* Received February 23, 1962.

[1] "IRE standards on methods of measuring noise in linear twoports, 1959," Proc. IRE, vol. 48, pp. 60–68; January, 1960.

[2] S. T. Fisher, "A completely consistent definition of temperature sensitivity including negative conductance devices," Proc. IRE (*Correspondence*), vol. 50, p. 204; February, 1962.

ratio n, which may be negative, we obtain

$$T_{eq} = nT, \tag{5}$$

where T is the room temperature in degrees K. Hence (4) has the same significance as the "effective noise figure" which includes the effect of the noisiness of the input termination with its noise ratio other than unity.

For a cascaded twoport we have

$$T_{e12} = \frac{kT_{e1}A_{e1}A_{e2}\Delta_f + kT_{e2}A_{e2}\Delta_f}{kA_{e1}A_{e2}\Delta_f}$$

$$= T_{e1} + \frac{T_{e2}}{A_{e1}}. \tag{6}$$

This is in complete agreement with Fisher's (9).

As to the heterodyne receiver, an "effective input noise temperature is defined for each pair of corresponding frequencies" according to Note 1 of the IRE Standards on Noise[1] which again is in complete agreement with the advantages pointed out by Fisher.

It seems, therefore, that the "temperature sensitivity" is a redundant term.

KATSUNORI SHIMADA
Dept. of Elec. Engrg.
University of Washington
Seattle, Wash.

Author's Comments[3]

I would like to state to Mr. Shimada and to IRE readers in general my purposes in submitting my original communication. They were:

1) To show that noise temperature and temperature sensitivity (or system noise temperature) can be uniquely defined without regard to whether source conductances (or resistances) are positive or negative;
2) To show, in particular, that a negative noise temperature is meaningful;
3) To point out the advantages of using the temperature-sensitivity concept rather than the noise-figure concept.

In many cases the symbol T_r, defined in my communication, and T_e, defined by the Standards,[1] may be used interchangeably. However, T_r, as I defined it, has some very important properties which are not expressly set forth in the Standards definition of T_e. These are as follows:

1) T_r is defined in terms of exchangeable power and exchangeable gain. The concepts of the noise temperature of the input termination and of noise-free networks are not required in the definition of T_r, although they may be helpful in understanding the physical significance of T_r.
2) T_r (also T_{eq}, the noise temperature of the input termination) is negative when, and only when, the source conductance (input termination) is negative.
3) All terms of the "cascade formula" have the same sign, which is the same as the sign of the source conductance of the *first* network in the cascade. This is independent of the signs of the output conductances of any of the networks in the cascade.
4) The excess noise temperature T_r is defined in terms of noise originating in the network itself. It does *not* include noise from the source at image frequencies in the case of a heterodyne receiver. In contrast to this, Note 3 of Section 5 of the Standards states that T_e is related to noise figure by

$$T_e = 290(F - 1).$$

If T_e is to include only noise originating in the network itself, then this relationship is valid *only* for the twoport (single-input-channel) case. It cannot be extended to the heterodyne case in which there are inputs at image frequencies.

Although T_e can be shown to have properties 1)–3) of the preceding paragraph, it being assumed that the noise temperature of the input termination has been properly defined for the case of negative conductance, it is my opinion that any future Standards article ought to state specifically the significance of the noise temperature of the source and the equivalent input noise temperature of the twoport when the source conductance is negative.

For the single-input-channel case it is true that

$$T_e = T_r = 290(F - 1),$$

where F is the noise figure. However, if the equation $T_e = 290(F-1)$ is used for a multiple-input-channel receiver, then T_e includes noise from the source at inage frequencies. This follows from the definition of F. On the other hand T_r is defined to represent noise originating only in the network itself whether the latter is a twoport or a multiple-input-channel receiver.

S. T. FISHER
Philco Scientific Lab.
Blue Bell, Pa.

[3] Received March 15, 1962.

Self-Consistent Electron Trajectories in a Magnetron*

The electron trajectories obtained numerically by Tibbs and Wright[1] are generally accepted[2] as being typical of an oscillating magnetron. However, Walker[3] has shown that these trajectories are not self-consistent.

We programed the problem of Tibbs and Wright for an IBM 704 computer. Like Tibbs and Wright, we used Hartree's method of self-consistent fields. Tibbs and Wright had performed only one cycle of the computation. We carried the computation through 36 cycles. Even so, our trajectories did not converge toward a self-consistent result.

When we dropped the requirement of space-charge-limited cathode emission and changed the values of the independent variables for a better correspondence with experimental data, we quickly obtained a self-consistent solution. A composite plot of the self-consistent trajectories is given in Fig. 1. They were computed for the CV 76 (modern designation, CV 1481). This is an S-band, 500-kw magnetron. It has 8 vanes and operates in the π mode. The calculations were made for the following conditions:

Frequency: 3000 Mc
Magnetic field: 0.23 webers/m^2
Peak anode current: 30 a
Peak anode voltage: 26.25 kv
Amplitude of traveling voltage wave on anode: 12 kv

The anode voltage and current were measured values and the RF voltage corresponded to maximum efficiency for a zero-space-charge calculation.

The computed efficiency came out to be 50 per cent, a value which agrees reasonably well with the measured value of 57 per cent. The calculated back-bombardment power was 6 per cent of the RF output power. About half the current emitted from the cathode went to the anode; the other half returned to the cathode.

Fig. 2 shows a plot of the space-charge density (averaged over the angular coor-

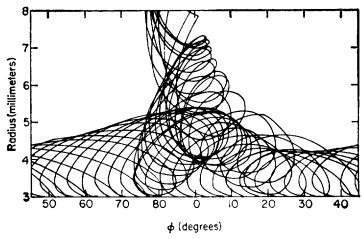


Fig. 1—Self consistent trajectories.

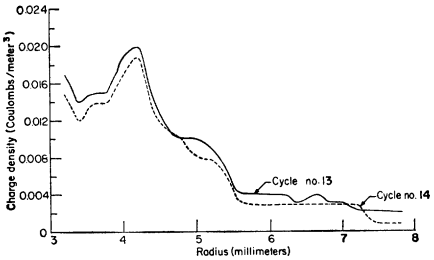


Fig. 2—Space-charge density (averaged over the angular coordinate).

* Received February 23, 1962. This work was supported by BuShips under contract no. NObsr-77592.

[1] S. R. Tibbs and F. I. Wright, "A Self-Consistent Field and the Space Charge Configuration for the CV76," CVD Report Mag. 41; March 23, 1945.

[2] See for example: G. B. Collins, Ed., "Microwave Magnetrons," McGraw-Hill Book Co., Inc., New York, N. Y., p. 270; 1948.
R. Latham, A. H. King and L. Rushforth, "The Magnetron," Chapman and Hall, London, England, p. 62; 1952.
S. I. Bychkov, "Magnetron Transmitters," Naval Publishing House, Defense Ministry of the USSR, Moscow, p. 37; 1955.

[3] G. B. Collins, Ed., "Microwave Magnetrons," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 274–282; 1948.

dinate ϕ) vs the anode radius. The charge distribution in which the trajectories of Fig. 1 were calculated is compared with the charge distribution derived from these trajectories. The figure shows the self-consistency of the charge distribution.

Figs. 1 and 2 show that the "hub" region of the space charge extends beyond the synchronous radius (3.4 mm). It goes out approximately to the Hull cutoff radius (4.4 mm). Within the hub, the space charge has more the character of a double-stream distribution (with maxima at the cathode and Hull radii) than of a Brillouin single-stream distribution.

C. G. Lehr
J. W. Lotus
Research Division
Raytheon Company
Waltham, Mass.
I. Silberman
R. C. Gunther
Missile and Space Division
Raytheon Company
Bedford, Mass.

Comments on "Analysis of AGC Loops"*

Certain discrepancies in my earlier correspondence[1] have been brought to my attention. Namely, although (9) is a solution of the differential equation (8), it is not a solution of (4), from which (8) was derived. The conclusion has been reached that (9) is not the complete solution of (8); in fact, it is suspected that (9) is a singular solution.

Efforts have been made to find a complete solution, so far without success. It should not be surprising to learn that the solution is not possible in explicit form. Communications from other interested persons would be welcome.

Despite the fact that (9) is incomplete (or singular), it is a good approximation to the error voltage, particularly if loop gain is large. This may be seen by manipulating (4) to obtain

$$E = -\frac{20}{KA}\log\frac{x}{x_0 + \dfrac{E}{G_0}}, \qquad (4a)$$

which approaches (9) for E/G_0 sufficiently small.

The material following (9) is correct, not being dependent upon the solution of the differential equation.

The author is indebted to D. Rothman of Electronic Specialty Co. for bringing the discrepancy and (4a) to his attention.

Floyd M. Gardner
Gardner Research Co.
Orange, Calif.

* Received February 27, 1962.

[1] Floyd M. Gardner, "Analysis of AGC loops," Proc. IRE (*Correspondence*), vol. 50, p. 97; January, 1962.

The Maximum Power Output of the Tunnel-Diode Oscillator*

The amplitude and power output of a tunnel-diode oscillator may be discussed on the base of approximate third order polynomial characteristics, derived by van der Pol,[1]

$$I - I_0 = a(U - U_0) + b(U - U_0)^3 \quad (1)$$

Although this formal approach in many respects truly describes the oscillation, certain properties of the oscillator, for instance the discontinuous decrease and the abrupt vanishing of the oscillation by the change of the load or the supply voltage, remain unclarified. In a previous article[2] the characteristics of a tunnel-diode were derived by quantum mechanical reasoning, as

$$I = a(U - U_m)^2 \tanh\frac{1}{2}\frac{U}{U_T} \qquad (2)$$

which approximates the real characteristics more accurately than former expressions. When treating oscillations, with regard to (2), the following approximations are appropriate:

$$I = \begin{cases} g_0 U & U < U_M \\ \dfrac{\gamma}{2}(U - U_m)^2 + I_m & U > U_M \end{cases}. \qquad (3)$$

The oscillation is determined by the conductance averaged over one cycle.[1] If the full cycle runs in the interval approximated by quadratic function (Fig. 1) then the average conductance is

$$G_{av} = \frac{1}{2\pi}\int_0^{2\pi} G(\omega t)d(\omega t) = \gamma(U_d - U_m). \qquad (4)$$

This conductance compensates for the damping conductance G_L of the oscillator circuit in a stationary state, that is

$$G_L + G_{av} = 0. \qquad (5)$$

In such a case the amplitude of the oscillation increases up to the limit of the quadratic interval ($U_1 = U_d - U_M$). It is obvious from (4) and (5) that by increasing the forward voltage at a given G_L load, the oscillation starts at the value U_M and stops at the voltage $U_d = U_M - (G_L/\gamma)$, when the amplitude of the oscillation is

$$U_1 = U_d - U_M = U_d - U_m + \Delta U = \Delta U\left(1 + \frac{G_L}{\gamma\Delta U}\right). \qquad (6)$$

Let us now consider the commencement of the oscillation when the operating point is at the start of the quadratic section (at U_M). In this case the averaged conductance of the tunnel-diode is

$$G_{av} = \frac{1}{2\pi}\int_0^{2\pi} G(\omega t)d(\omega t) = \frac{\gamma}{2}[U_1 + (U_d - U_m) + \Delta U]. \qquad (7)$$

* Received February 26, 1962; revised manuscript received March 12, 1962. The work reported in this paper was performed at the Central Measurement Research Laboratory, Budapest, Hungary.

[1] K. Tarnay, "The Tunnel-Diode," Ph.D. dissertation, Technical University, Budapest, Hungary; 1961.

[2] K. Tarnay, "Approximation of the tunnel-diode characteristics," Proc. IRE (*Correspondence*), vol. 50, pp. 202–203; February, 1962.

From this the amplitude of the oscillation at this operating point is

$$U_1 = \frac{\pi}{2}\Delta U\left(1 - \frac{g_0}{\gamma\Delta U} - 2\frac{G_L}{\gamma\Delta U}\right). \qquad (8)$$

The calculated and measured characteristics of the oscillation are plotted in Fig. 2.

The maximum power output of the tunnel-diode oscillator is obtainable near the ceasing of the oscillation, *i.e.*,

$$P_{max} = \frac{1}{2}U_1^2 G_L = \frac{1}{2}\Delta U^2\left(1 - \frac{G_L}{\gamma\Delta U}\right)^2 G_L. \qquad (9)$$

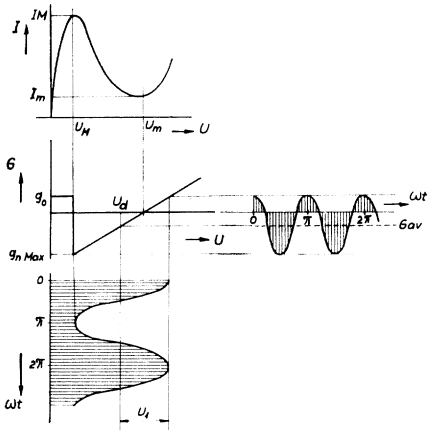
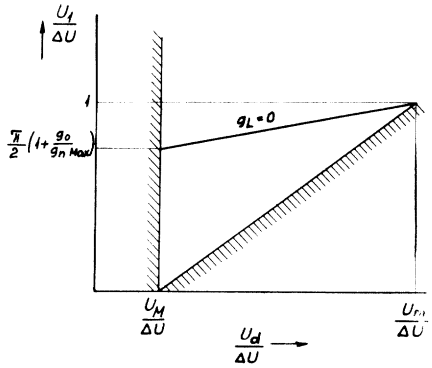
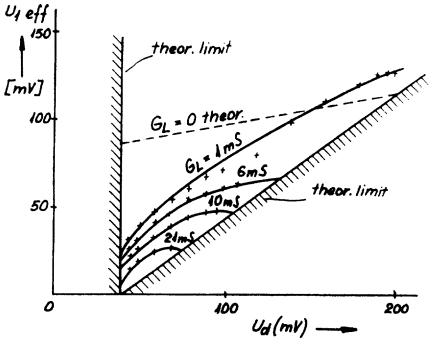


Fig. 1—The average conductance of the tunnel-diode. The full cycle runs in the quadratic interval.



(a)



(b)

Fig. 2—Oscillation characteristics of the tunnel-diode oscillator. (The amplitude of the oscillation vs operating point dc voltage.) (a) Theoretical limits. (b) Measured characteristics. (Parameter: load conductance.)

Assuming that $\gamma\Delta U = g_{n\,\max}$ and referring to (3), ΔI is

$$\Delta I = I_M - I_m = g_{n\max}\cdot\frac{U_M + U_M}{2}\cdot \quad (10)$$

A combination of these equations yields the maximum power output

$$P_{\max} = \Delta U \Delta I \frac{U_m - U_M}{U_m + U_M}\left(1 - \frac{G_L}{g_{n\max}}\right)^2 \cdot \frac{G_L}{g_{n\max}} \quad (9a)$$

The optimum power output of the oscillator may be obtained by calculating the extremum of (9a) in the function of the load conductance

$$P_{\text{opt}} = \frac{4}{27}\Delta U \Delta I \frac{U_m - U_M}{U_m + U_M}. \quad (11)$$

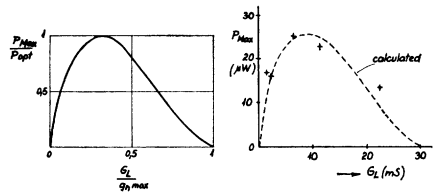
The power output is optimal at $G_L = -\frac{1}{3}g_{n\max}$ and at $U_d = U_m - \frac{1}{3}\Delta U$ voltage. Assuming $U_M = 0.2\ U_m$ as usual, with customary tunnel-diodes, the optimum power output is

$$P_{\text{opt}} = \frac{8}{81}\Delta U \Delta I. \quad (11a)$$

This is only the 52.2 per cent of the value calculated by the formal third order approximation ($3/16\Delta U\Delta I$). When the load conductance is not the optimum, the maximum output power is

$$P_{\max} = P_{\text{opt}}\left[\frac{27}{4}\left(1 - \frac{G_L}{g_{n\max}}\right)^2 \frac{G_L}{g_{n\max}}\right] \quad (9b)$$

The power output vs the load conductance is shown in Fig. 3.



(a) (b)

Fig. 3—Maximum output power of the tunnel-diode oscillator. (a) Theoretical. (b) Measured.

K. Tarnay
Senior Research Associate
Central Measurement Res. Lab.
Budapest, Hungary

Noise Measurements on Tunnel Diodes*

Since the initial studies of Yajima and Esaki[1] in 1958, relatively few papers reporting basic noise measurements in tunnel diodes have been published. Montgomery and Lee[2] reported on noise measurements at 5 Mc, intended to split the total tunnel current in its forward and reverse components. J. J. Tieman reported measurements taken by R. L. Watters of G.E. Research Laboratories.[3]

The present note will report noise spectra between 100 kc and 30 Mc on several diodes, for currents in the low-voltage positive-conductance region, the excess-current region and the high-voltage positive-conductance region. The biasing circuit involved regular RC coupled circuits at the lower frequencies and tuned LC circuits above 0.55 Mc. No step-up transformers were employed to avoid possible errors. The noise of the input 6922 cascode circuit was subtracted. The noise was calibrated with a standard noise diode. In all cases the noise spectra measured in the lower positive conductance region, taken from 100 kc to 30 Mc, were completely flat. Some data for a ZJ56A diode are listed in Table I.

If the noise corresponds to full shot noise[4] of the individual tunneling currents—the Zener current and the Esaki current—one obtains from Fermi Dirac statistics the following simple result,[5] relating the equivalent saturated diode current I_{eq} to the tunnel diode current I_d:

$$I_{\text{eq}} = I_d \coth (eV/2kT). \quad (1)$$

The calculated magnitudes according to (1) are also included in Table I. The correspondence is quite good and does indicate full shot noise, *i.e.*, the tunneling of individual carriers is independent.

In Fig. 1 we show some noise spectra in the excess current and second positive-conductance region for the same diode. It is clear that the shot noise level is only reached for frequencies over 10 Mc. The large amount of LF noise is apparently associated with the excess current and with the normal *p-n* junction current in these regions. The current dependence is complex. The shot noise levels as a function of bias for this diode are also summarized in Fig. 2.

TABLE I
Experimental Data for I_{eq} in the Lower Positive Region of ZJ56A Tunnel Diode

I_d=0.9 ma		I_d=0.8 ma		I_d=0.7 ma		I_d=0.6 ma		I_d=0.5 ma	
Freq. Mc	I_{eq} ma	Freq. Mc	I_{eq} ma	Freq. Mc	I_{eq} ma	Freq. Mc	I_{eq} ma	Freq. Mc	I_{eq} ma
0.55	1.27	0.55	1.37	0.55	1.42	0.55	1.51	0.55	1.53
1	1.31	1	1.32	1	1.44	1	1.50	1	1.53
2	1.26	2	1.33	2	1.39	2	1.49	2	1.49
4	1.31	4	1.28	4	1.40	4	1.40	4	—
8	1.30	8	1.33	8	1.41	8	1.48	8	—
15	1.26	15	1.32	15	—	15	—	15	—
30	1.26	30	—	30	—	30	—	30	—
$\overline{I_{\text{eq}}}$=1.28±0.02		$\overline{I_{\text{eq}}}$=1.33±0.015		$\overline{I_{\text{eq}}}$=1.41±0.02		$\overline{I_{\text{eq}}}$=1.51±0.02		$\overline{I_{\text{eq}}}$=1.52±0.02	
Calculated: I_{eq}=1.27		I_{eq}=1.35		I_{eq}=1.46		I_{eq}=1.55		I_{eq}=1.62	

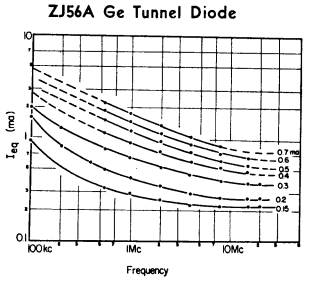


Fig. 1—Noise vs frequency for constant diode current in the upper positive region of the characteristic.

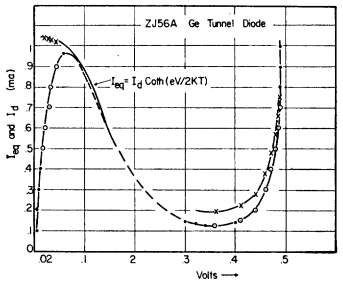


Fig. 2—Diode current and shot noise vs voltage for ZJ56A.

Somewhat different results were obtained for another type GE diode, *viz.*, IN2940. The noise for low bias is white full shot noise as before. Also for very large bias, I_{eq} approaches I_d, *i.e.*, the *p-n* junction current behaves as expected. However, the shot noise in the valley region of the characteristic exceeds full shot noise by one order of magnitude. It is not likely that this must be attributed to the occurrence of two opposite tunneling currents. Apparently, the excess current, presumably composed of electrons which tunnel via states in the forbidden gap to the valence band, is quite noisy and cannot be attributed to independent random events. Further investigation of the noise in this region as well as in the negative resistance region is in progress.

The authors would like to thank Dr. van der Ziel for his stimulating interest.

D. C. Agouridis
K. M. van Vliet
Department of Electrical Engineering
University of Minnesota
Minneapolis, Minn.

* Received February 23, 1962. This work was supported by the U. S. Army Signal Corps.

[1] T. Yajima and L. Esaki, "Excess noise in narrow *p-n* junctions," *J. Phys. Soc. Japan*, vol. 13, pp. 1281–1287; November, 1958.

[2] C. A. Lee and H. C. Montgomery, "Determination of forward and reverse tunneling currents in Esaki diodes by shot noise measurement," *Bull. Am. Phys. Soc.*, vol. 5, p. 160; March, 1960, and private communication.

[3] J. J. Tieman, "Shot noise in tunnel diode amplifiers," Proc. IRE, vol. 48, pp. 1418–1423; August, 1960.

[4] We shall refer to "shot noise" as the noise observed at sufficiently high frequencies when the spectrum has leveled off to a constant value; we shall refer to full shot noise when Schottky's formula for the tunneling currents of each separate polarity is satisfied.

[5] R. A. Pucel, "The equivalent noise current of Esaki diodes," Proc. IRE, vol. 49, pp. 1080–1081; June, 1961. This expression can also be obtained, writing I_{eq} in the form of the overlap integral, which transforms into $A(Ev - E_C)^2[1 - \exp(-eV/kT)]\ Y$ where Y is a numerical integral defined by Bates. (See C. W. Bates, "Tunneling currents in Esaki diodes," *Phys. Rev.*, vol. 121, pp. 1070–1071; February, 1961.)

Ray Theory in the Ionosphere*

Ray theory remains useful as long as the ray path prescribes the path of the flow of electromagnetic energy. Assume a linearly stratified ionosphere whose electron density varies with height. Let there be no magnetic field and let collisions be neglected. At a point P in this medium let the electron density be N_e. For a frequency ν the refractive index μ at P is

$$\mu = (1 - N_e e^2/\pi m\nu^2)^{1/2}. \quad (1)$$

where e is the charge of the electron and m is its mass. It is synonymous to (1) that μ at P is isotropic so that $\mu=\mu\Psi$ at P for $0\leq\Psi\leq\pi$. The group velocity v_g of an electromagnetic wave in this medium is defined as[1]

$$v_g = c\mu \quad (2)$$

for $\mu<1$(1). The phase velocity $v_{ph}=c/\mu$ such that $v_{ph}\cdot v_g=c^2$, where c is the velocity of light attributed to free space. In this idealized ionosphere a ray path is determined by means of Snell's law,

$$\mu \sin\Psi = \sin i. \quad (3)$$

The product $\mu \sin\Psi$ is a constant along the ray path whose angle of incidence at the linear interface between free space and the ionized medium is i.

With these prerequisites an elementary approach can be used to derive: the equivalent normal-incidence frequency, the theorems for the equivalence of path, and of virtual height.[2]

For $\Psi=\pi/2$ (3) the isotropy of μ (1) yields

$$(1 - N_e e^2/\pi m\nu^2)^{1/2} = \sin i \quad (4)$$

and

$$\nu \cos i = (N_e e^2/\pi m)^{1/2}. \quad (5)$$

For a normal-incidence frequency ν' which is reflected at $P(\mu=0)$ it follows from (1) for $\nu\rightarrow\nu'$

$$\nu' = (N_e e^2/\pi m)^{1/2}, \quad (6)$$

and from (5), (6)

$$\nu \cos i = \nu' = \nu_1 \cos i_1 = \nu_2 \cos i_2 = \cdots = \nu_n \cos i_n, \quad (7)$$

where ν' is the equivalent normal-incidence frequency. Although N_e was assumed to vary with height such that $dN_e/dh\neq 0$, the isotropy of μ (1) is a valid assumption since (7) agrees with existing knowledge. Let us now derive the "equivalence of path" theorem by an equally simple reasoning.

If μ of (2) is substituted into (3) one obtains

$$v_g \sin\Psi = c \sin i = v_x, \quad (8)$$

where v_x is the velocity component of the flow of electromagnetic energy in the x direction. Since $v_x=c \sin i$ is a constant the same horizontal speed of the energy flow is obtained if the wave is unrefracted, *i.e.*, if it continues to travel along the direction of incidence with the velocity of light in free space. With the assumptions made initially (8) is true along the entire ray path. Thus, both the wave traveling with v_g along the true path and the wave traveling with c along the virtual path, reach any particular value of x within the same time interval. In regards to total travel time true and virtual path are therefore equivalent. This theorem of the equivalence of paths can therefore be understood without the use of integrals. A similar elementary approach yields the theorem for the "equivalence of virtual heights."

In order to derive the theorem for the equivalence of virtual heights the rays belonging to (ν_1, i_1), (ν_2, i_2), (ν_3, i_3), etc., which satisfy (7), are usually assumed to be returned from the same true height. Since the validity of Snell's law (3) is independent of the height variation of N_e it is possible to use any $N_e(h)$ distribution without restricting the conclusion. The simplest case is a homogeneous ionosphere $[N_e(h)=\text{constant}]$ which is linearly bounded against free space. Instead of asking for the same "reflection" height let us determine the relationship between both virtual heights y_{v_1}, y_{v_2} corresponding to the rays (ν_1, i_1) (ν_2, i_2) as they reach the same true height h. In Fig. 1 the depicted nomenclature of a ray path can be indexed and serve a double purpose. It follows by definition

$$v_{g_1}\tau_1 \cos\Psi_1 = v_{g_2}\tau_2 \cos\Psi_2 = h, \quad (9)$$

where $v_{g_1}=c\mu_1$, $v_{g_2}=c\mu_2$ and τ_1, τ_2 are the time intervals needed by (ν_1, i_1) (ν_2, i_2) to reach the true height, h. From (1)–(3), and (9) one obtains

$$\frac{\tau_1}{\tau_2} = \frac{(\cos^2 i_2 - N_e e^2/\pi m\nu_2^2)^{1/2}}{(\cos^2 i_1 - N_e e^2/\pi m\nu_1^2)^{1/2}}. \quad (10)$$

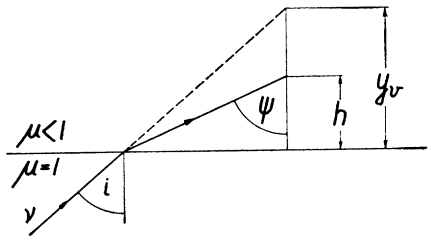


Fig. 1.

From the previous path-equivalence theorem follows:

$$y_{v_1} = c\tau_1 \cos i_1, \quad y_{v_2} = c\tau_2 \cos i_2. \quad (11)$$

From (10), (11)

$$\frac{y_{v_1}}{y_{v_2}} = \frac{\tau_1 \cos i_1}{\tau_2 \cos i_2} = \frac{(1 - N_e e^2/\pi m\nu_2^2 \cos^2 i_2)^{1/2}}{(1 - N_e e^2/\pi m\nu_1^2 \cos^2 i_1)^{1/2}}. \quad (12)$$

Because of (7) $y_{v_1}=y_{v_2}$ and the "equivalent virtual height" theorem is confirmed.

The inequality[3]

$$\frac{v_{ph}}{2\pi\mu\nu}\frac{d\mu}{dh} \ll 1 \quad (13)$$

determines whether or not a ray path does represent the path of energy flow. Neither in the above derivations, nor elsewhere[2] was (13) explicitly considered. Moreover, in the derivation of the "equivalent virtual height" theorem a discontinuity was assumed. These circumstances suggest that the theorems are a corollary of Snell's law. Their truth is contingent upon the truth of the latter. The physical meaning of the ray paths is subsequently decided by using (13).

For a circular stratification (3) is replaced by Bouguer's law

$$(\mu_1 r_1 \sin\Psi_1 = \mu_n r_n \sin\Psi_n = \text{constant})$$

and the above theorems lose their exact meaning. For frequencies for which electron collisions cannot be neglected, simple ray theory no longer applies. Allowing for the magnetic field, which makes the ionosphere an anisotropic medium, further restricts the usefulness of these theorems.

KURT TOMAN
AF Cambridge Res. Lab.
Laurence G. Hanscom Field
Bedford, Mass.

On "The Noise Figure of Negative Conductance Amplifiers"*

In a recent communication Brown[1] proved that a feedback amplifier used as a negative conductance amplifier can have an optimum noise figure much lower than the optimum noise figure of the same amplifier without feedback. It is the aim of this correspondence to show that this result is due to an error in the calculation.

The error occurs first in Brown's (8), where the factor $A^2\beta^2$ in front of E_{n2}^2 has been omitted. As a consequence Brown's (11)–(13) are in error. We shall now replace Brown's (8), (11), (12), and (13) by the corrected (8a), (11a), (12a) and (13a).

The noise figure for direct coupling between the source and the negative conductance circuit is

$$F' = 1 + \frac{A^2\beta^2 E_{n2}^2}{I^2 Z_1^2 (1 - A\beta)^2} + \frac{E_{n1}^2}{N^2 I^2 Z_1^2}. \quad (8a)$$

If the source is coupled through an ideal transformer of turns ratio N, I must be replaced by NI and Y_s by N^2Y_s, so that (10) may be written

$$N^2 Y_s \approx -\frac{1}{Z_1(1 - A\beta)}.$$

or

$$A\beta \approx \left(1 + \frac{1}{N^2 Y_s Z_1}\right).$$

Substituting into (8a), one obtains

$$F' = 1 + \frac{E_{n2}^2(N^2 Y_s + 1/Z)^2}{N^2 I^2} + \frac{E_{n1}^2}{N^2 I^2 Z_1^2}. \quad (11a)$$

* Received, March 6, 1962; revised manuscript received, March 19, 1962.

[1] H. R. Mimno, "The physics of the ionosphere," *Rev. Mod. Phys.*, vol. 9, pp. 1–43; January, 1937.

[2] S. K. Mitra, "The Upper Atmosphere," The Asiatic Society, Calcutta, India, pp. 237–240; 1952.

[3] D. E. Kerr, "Propagation of Short Radio Waves," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 13, pp. 53–58; 1951.

* Received February 5, 1962.

[1] F. W. Brown, "The noise figure of negative-conductance amplifiers," PROC. IRE (*Correspondence*), vol. 49, pp. 520–521; February, 1961.

Considered as a function of N, F' is optimized for

$$N^2 = \frac{\sqrt{E_{n2}^2 + E_{n1}^2}}{E_{n2} Y_s Z_1} \tag{12a}$$

and the optimized value is

$$F'_{\text{opt}} = 1 + \frac{2Y_s}{I^2 Z_1} \cdot [E_{n2}^2 + E_{n2}\sqrt{E_{n2}^2 + E_{n1}^2}] \tag{13a}$$

which is exactly equal to the optimum noise figure without feedback. [See Brown's (3).]

The conclusions are therefore that the two amplifiers have the *same* noise figure and that Brown's analysis cannot explain Ishii's results on negative resistance klystron amplifiers.[2]

A. VAN DER ZIEL
Dept. of Elec. Engrg.
University of Minnesota
Inst. of Tech.
Minneapolis, Minn.

[2] K. Ishii, "Noise figures of reflex klystron amplifiers," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-8, pp. 291–294; May, 1960.

Analysis and Synthesis of a Class of Nonlinear Systems*

The procedure outlined in this work is a numerical technique which can be used for the purpose of analysis and synthesis of a very broad class of nonlinear systems where the operation of the system can be represented by a linear differential equation over any arbitrarily small time interval. The work is divided into two parts:

1) Formulation of the piece-wise linear analysis,
2) Use of the results of (1) for the purpose of systematical synthesis.

The formulation of the piece-wise linear analysis and a simple example is given below. From an application point of view it compares with the Z-form technique suggested by Boxer and Thaler[1] although it is simpler to apply (there is no need for a table of transforms such as Z-form tables, and the result is more accurate for a number of problems). The shortcoming of the technique suggested here is the necessity of obtaining the roots of the characteristic equation corresponding to the linearized equation.

Let a system be described by the equation

$$f(y, \dot{y}, \ddot{y}) = 0 \tag{1}$$

and assume the solution exists and is

$$y = y(t). \tag{2}$$

The function $f(y, \dot{y}, \ddot{y})$ can readily be approximated by a linear differential equation over the period $nT \leq t \leq (n+1)T$ where T is small. For this purpose the necessary initial conditions are $y(nT)$ and $\dot{y}(nT)$, *i.e.*, the final values of the preceding period $(n-1)T < t < nT$. In terms of these initial conditions $Y_n(s)$, the Laplace transform of $y(t)$ can be obtained and is valid only during the interval $(nT < t < n+1)T$. Consequently, since

$$y(nT + t) = \frac{1}{2\pi i}\int_{c-j\infty}^{c+j\infty} Y_n(s)e^{st}ds \tag{3}$$

where $0 \leq t \leq T$,

$$y[(n+1)T] = \frac{1}{2\pi i}\int_{C-j\infty}^{C+j\infty} Y_n(s)e^{sT}ds = \sum \text{Res. of } Y_n(s)e^{sT}. \tag{4}$$

The initial conditions $y(nT)$ and $\dot{y}(nT)$ are used to determine the $Y_n(s)$. In other words $y(n+1)T$ is obtained from $f(y, \dot{y}, \ddot{y}) = 0$ and initial conditions. The steps to be followed are summarized in the following:

1) The quantity T is chosen such that $f(y, \dot{y}, \ddot{y}) = 0$ can be approximated by a linear equation over any interval with a duration T seconds.
2) The Laplace transform of $y(t)$ is obtained in terms of $y(nT)$ and $\dot{y}(nT)$ which is $Y[s, y, (nT), \dot{y}(nT)]$, valid only when $nT \leq t \leq (n+1)T$.
3) The $y(n+1)T$ is obtained which is the sum of residues of
$$Y[s, y(nT), \dot{y}(nT)]e^{st}.$$
4) The actual given initial conditions are used to obtain $y(T)$ and then systematically $y(nT)$ will be obtained up to any desired $n = N$.

Example

$$\dot{y} + y^2 = 1. \tag{5}$$

Let $y = y(nT) = \text{constant}$, over the interval $nT < t < (n+1)T$ and take $T = 0.1$ sec. Therefore

$$\dot{y} + y(nT)y = 1 \tag{6}$$

$$Y_n(s) = \frac{1 + sy(nT)}{s[s + y(nT)]}. \tag{7}$$

Hence

$$y[(n+1)T] = \sum \text{Res. of } [Y_n(s)e^{0.1s}] = \frac{1 - [1 - y^2(nT)]e^{-.1y(nt)}}{y(nT)}. \tag{8}$$

The original nonlinear equation happens to have an exact solution

$$y = \tanh(t). \tag{9}$$

The following table is a comparison between the solution obtained through the suggested technique and the exact solution.

t	$y(t)$	$y(t)$ exact
0	10	0
0.1	0.1000	0.09967
0.2	0.1941	0.1970
0.3	0.2856	0.2910
0.4	0.3816	0.3799

The second phase of the work is the synthesis of nonlinear systems by using the above technique. The procedure is given as follows:

1) Adopt a specific value for T. It is noted that the computational error involved in this case is not additive; consequently T can be made as small as required to obtain the nonlinear function as accurately as desired.
2) From the knowledge of $y(nT)$ and $y(n+1)T$ the value for the nonlinear element is obtained in a manner similar to the analysis procedure described above.

Example

Suppose the synthesis of nonlinear first-order equation of the form

$$\dot{y} + f(y)y = u(t) \tag{10}$$

is desired. The desired response to a unit step may be given by any single-valued curve. The problem is then to determine $f(y)$ as a function of y. The unknown function $f(y)$ is obtained in an approximate form in a manner that it assumes a constant value over the interval $T \leq t \leq (n+1)T$. T can be chosen arbitrarily small, based on the desired degree of approximation.

From the results of the previous section

$$y(n+1)T = \frac{1}{f(y)}\{1 - [1 - y(nT)f(y)]e^{-Tf(y)}\}. \tag{11}$$

Successively substituting the values for $y(nT)$ and $y(n+1)T$, values for $f(y)$ can be obtained as a function of y from (11).

N. E. NAHI
University of Southern California
Los Angeles, Calif.

Inductive Probability as a Criterion for Pattern Recognition*

The above communication[1] raises the question of how to apply Carnap's concept of inductive probabilities in predicting, from a finite sequence of events, a probability pattern for the next events. This problem is partly solved by splitting up the sample sequence into sections of equal length and predicting from their frequency, in terms of the wellknown Laplace formula, the probability of the future occurrence of such sections. In closing, the question is raised of how to proceed if the sample sequence is not composed of a complete period, *i.e.*, if the chosen length of the sections is not an aliquot part of the length of the sample sequence.

On first assessment, this problem appears to me to be quite readily solvable by drop-

* Received January 15, 1962.
[1] R. Boxer, and S. Thaler, "A simplified method of solving linear and nonlinear systems" PROC. IRE, vol. 44, pp. 89–101; January, 1956.

* Received March 5, 1962.
[1] B. Harris, PROC. IRE (*Correspondence*), vol. 49, pp. 1951–1952; December, 1961.

ping as many symbols at the beginning of the sequence as is necessary to obtain a complete period. This amounts, it is true, to the sacrifice of available information, but this shortcoming is inherent in the very nature of the approach described.

To make maximum possible use of the information contained in the sample sequence, the statistical analysis should, in my opinion, be extended to cover *all* pairs, triples, etc. Above all, the conditional probabilities should also be analyzed. To make this clear, I would refer to the example cited.

The sample sequence is

$$A\ B\ A\ B\ A\ B\ A\ B.$$

A statistical analysis of the individual symbols at first yields the quantities

$$n(A) = n(B) = 4;$$

hence

$$p(A) = p(B) = \frac{4+1}{8+2} = \frac{1}{2}.$$

We now proceed with calculating the conditional probability for the occurrence of the two symbols if symbol B is the condition (which amounts to the obvious assumption that the process will produce a Markov chain). The statistical analysis of the sequence yields

$$n(BA) = 3, \qquad n(BB) = 0;$$

hence

$$p(B \mid A) = \frac{3+1}{3+2} = \frac{4}{5},$$

analogous is $p(B \mid B) = \frac{1}{5}$. (In designating the conditional probability, the representation chosen here deviates from the conventional form in that the condition has been placed *ahead* of the vertical stroke and the symbol in question behind this stroke.)

The same values are obtained when investigating the triples (ABA) and (ABB). The quadruples yield

$$n(BABA) = 2,$$
$$n(BABB) = 0,$$
$$p(BAB \mid A) = \frac{2+1}{2+2} = \frac{3}{4}.$$

Thus, greatest significance in this case has the analysis of the pairs and triples, on the basis of which it can be predicted with a probability of $\frac{4}{5}$ that the next symbol will be an A.

In order to be able to predict, in addition to the next symbol, also the symbol following the next (which is different from predicting only the symbol following the next), we start with the statistical analysis of all pairs. We arrive at $m=4$ possibilities. We find that

$$n(AB) = 4,\ n(BA) = 3,\ n(AA) = n(BB) = 0$$

and

$$p(AB) = \frac{4+1}{7+4} = \frac{5}{11}$$

$$p(BA) = \frac{3+1}{7+4} = \frac{4}{11}$$

$$p(AA) = p(BB) = \frac{1}{11}.$$

Furthermore, for the conditional probabilities there results

$$n(BAB) = 3,$$
$$n(BAA) = n(BBA) = n(BBB) = 07$$

and hence

$$p(B \mid AB) = \frac{3+1}{3+4} = \frac{4}{7}$$

(the other three probabilities are each given by $\frac{1}{7}$).

When taking the last two symbols as condition, we obtain the same value; when taking more, the probability will again decline.

E. R. Berger
Central Lab.
Siemens & Halske
Munich, Germany

Author's reply[2]

This is not a criticism of Dr. Berger's comment, but rather a request for further extensions of his ideas.

The purpose of applying the concept of inductive probability to a finite sequence of events was not to determine a probability pattern for the next events (although this was the immediate result), but rather to assign measures to the various possible patterns that could be present. The patterns considered have three characteristics: basic repetition period, sequence of characters of which it is composed, and absolute location within the given sequence. The main difference between Dr. Berger's approach and mine is the attempt to maintain information as to the absolute location of the periodicity. Extending the technique to cover all pairs, triples, etc., gives a measure which is, in effect, an average over all possible absolute locations. How does one avoid this and yet not sacrifice available information?

B. Harris
College of Engrg.
New York University
New York, N. Y.

[2] Received March 28, 1962.

True RMS Measurement with Tunnel Diodes*

Neu's communication on a tunnel-diode frequency-doubling circuit[1] aroused considerable interest here, since I had previously done some work in the same area, by describing a similar circuit.[2]

I wish to stress that the same principle was also applied for measuring the true rms value of ac voltages. Namely, if the operating point of the tunnel diode is the peak point of the characteristic (where it is of quadratic character) and an ac voltage $U.\cos \omega t$ is superimposed on the dc component, Fig. 1, then the current of the tunnel diode has an ac component $\frac{1}{2}KU^2 \cdot \cos 2\omega t$, and the dc component changes with $\frac{1}{2}KU^2$. The latter is proportional to the rms value. Fig. 2 shows experimental results. The circuit measured the true rms values as was proven by a measurement using two independent ac voltages.

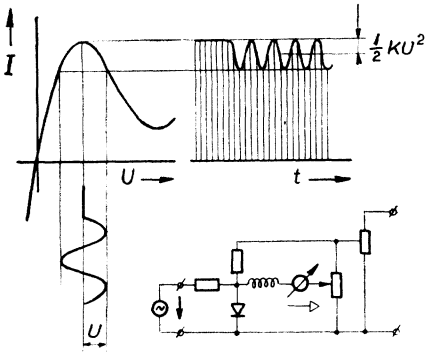


Fig. 1.

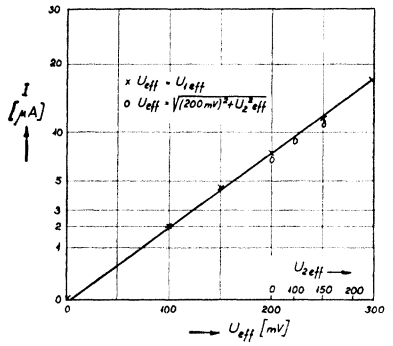


Fig. 2.

K. Tarnay
Senior Research Associate
Central Measurement Research Laboratory
Post Office Box 170
Budapest 4, Hungary

* Received March 7, 1962.

[1] F. D. Neu, "A Tunnel-diode wide-band frequency doubling circuit," PROC. IRE (*Correspondence*), vol. 49, pp. 1963–1964; December, 1961.

[2] K. Tarnay, "The Tunnel Diode," Doctoral dissertation, Technical University, Budapest, Hungary, pp. 139–141 and 151–155; May, 1961.

A Table of Solutions of Riccati's Equations*

During 1961 many workers [1–8] wrote on Riccati's nonlinear differential equations (NDE's) which have a history of well over two hundred years. Perhaps one of the many reasons why Riccati's NDE's are so important is that they replace all homogeneous

* Received March 8, 1962. A part of this note was presented at the meeting of the American Mathematical Society, Atlantic City, N. J., April 19, 1962 as "Riccati's Equation with Loosely Interrelated Coefficients."

second-order ordinary linear differential equations (LDE's) by simple changes of variables. These equations describe many physical systems, and matters of current interest such as tapered lines, $2n$-port circuits, RF waves in the ionosphere and time-varying networks. Nevertheless, the exact solution of a generalized Riccati's NDE having three independent arbitrary variable coefficients is still unknown.

Under these prevailing circumstances, a handy table of exact solutions for certain special though physically significant cases may be useful. This correspondence provides a table of solutions of Riccati's NDE's with interrelated variable coeficients, and gives a brief summary of the present state of the art.

McLachlan [9] shows in his book that an exact particular solution, $y=1/x$, satisfies

$$y'(x) + xg(x)y(x) + y^2(x) = g(x), \quad (1)$$

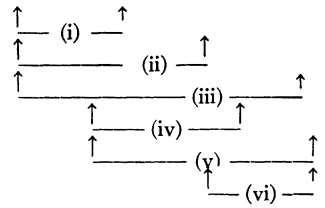
where $y(x)$ is the unknown, the prime is used for a derivative with respect to x, which is the only independent variable such as distance or time, and $g(x)$ is any known function of x. This solution is obvious if (1) is regrouped as,

$$(y' + y^2) + (xgy - g) = 0. \quad (2)$$

Suppose $f(x)$, another arbitrary known function of x, is employed instead of $g(x)$ in the right-hand side of (1), then $y=f/(xg)$ results from the second bracket group, while $y=1/x$ comes from the first bracket group. To eliminate this apparent discrepancy, the requirement of $f(x)=g(x)$ is introduced as a natural answer.

Following the same line of thought, all possible combinations of regrouping a generalized Riccati's NDE are applied. Six (as ${}_4C_2=6$) ways to bracket such an equation are indicated by linking lines in (3),

$$y'(x) + P(x)y(x) + Q(x)y^2(x) = R(x). \quad (3)$$



Unfortunately these six particular exact solutions tabulated below must accompany their corresponding necessary interrelationship among three general coefficients, $P(x)$, $Q(x)$ and $R(x)$.

Tabulated exact particular solutions of $y'(x) + P(x)y(x) + Q(x)y^2(x) = R(x)$

	A particular solution	Required interrelationship
(i)	$ae^{-\int P dx}$	$R = a^2Qe^{-2\int P dx}$
(ii)	$\dfrac{1}{\int Q dx + b}$	$R = \dfrac{P}{\int Q dx + b}$
(iii)	$\int R dx + c$	$Q = -\dfrac{P}{\int R dx + c}$
(iv)	$-\dfrac{P}{Q}$	$R = -\left(\dfrac{P}{Q}\right)'$
(v)	$\dfrac{R}{P}$	$Q = \left(\dfrac{P}{R}\right)'$
(vi)	$\sqrt{\dfrac{R}{Q}}$	$P = -\left(\ln\sqrt{\dfrac{R}{Q}}\right)'$
(vii)	$\dfrac{e^{-\int(P-2Q)dx} - \int Pe^{-\int(P-2Q)dx}dx - a}{e^{-\int(P-2Q)dx} + \int Pe^{-\int(P-2Q)dx}dx + a}$	$R = Q - P$
(viii)	$\dfrac{\int Pe^{-\int(P+2Q)dx}dx + b - e^{-\int(P+2Q)dx}}{\int Pe^{-\int(P+2Q)dx}dx + b + e^{-\int(P+2Q)dx}}$	$R = Q + P$

Note: a, b, and c are arbitrary constants.

There are actually three independent table entries in (i) through (vi) because, for example, (iii) and (iv) furnish the identical general solution of (3). Rajagopal [8] reports (i) originating from a totally different approach. Similarly (iv) and (v) were previously obtained by the writer using linearizing transforms, $y=1/(fg)$ and $y=fg$ respectively [3]. However, it appears difficult to identify the linearizing transform for a generalized Riccati's NDE for (vi). It was in the course of this searching process that a new and practical application was found.

A new transform of the dependent variable is

$$y(x) = \frac{f(x) - g(x)}{f(x) + g(x)}. \quad (4)$$

When (4) is substituted in (3), another Riccati's NDE is derived,

$$2h' - 2(Q + R)h + (P + Q - R)h^2 = (P + R - Q), \quad (5)$$

where

$$h(x) = f(x)/g(x).$$

Eq. (5) is readily solved exactly if proper interrelationships among variable coefficients are realized. When, for instance,

$$R(x) = Q(x) - P(x), \quad (6)$$

(5) is reduced to a solvable Bernoulli's NDE. Alternatively, if

$$R(x) = Q(x) + P(x), \quad (7)$$

(5) becomes a first-order LDE giving an exact solution. Since linearizing transforms of the dependent variable in a generalized Riccati's NDE exhibit dualism [3], it is easy to see that (6) and (7) can be obtained by using,

$$y(x) = \frac{f(x) + g(x)}{f(x) - g(x)}. \quad (8)$$

There are several ways to linearize Riccati's NDE [3], [10]. Two simple ways are to use transforms, $y=u'/(Qu)$ and $y=(u'-Pu)/(Qu)$, where $u(x)$ is any function of x. These transforms with (6) or (7) linearize (3) respectively,

$$u'' + \left(P - \frac{Q'}{Q}\right)u' - (Q \mp P)Qu = 0 \quad (9)$$

and

$$u'' - \left(P + \frac{Q'}{Q}\right)u' - \left[\left(\frac{P}{Q}\right)' + Q \mp P\right]Qu = 0 \quad (10)$$

where $P(x)$ and $Q(x)$ are two independent, known, arbitrary functions of x. The degree of contribution of (6) and (7) may be more or less proportional to the frequency of appearance of solved equations (9) and (10) in practical physical systems.

These "relaxed" interrelationships, (6) and (7), compared with more complicated other interrelationships such as, $R=-(P/Q)'$, $Q=(P/R)'$, and $P=-(\ln\sqrt{R/Q})'$, are new to the best knowledge of the writer

in the literature of Riccati's NDE's. These eight table entries represent a great number of specialized cases, for example, McLachlan's illustration [9], (1) is a special case of (v). Many existing text books carry exercise problems of Riccati's NDE's together with specified variable coefficients, such as $P(x)=\exp(x)$, $Q(x)=\sin x$, and $R(x)=\ln x$.

It is felt, however, that the foregoing represents a set of more useful relationships for the solutions of problems in the rapidly growing nonlinear systems.

IWAO SUGAI
ITT Federal Laboratories
Nutley, N. J.

REFERENCES

[1] K. G. Budden, "Radio Waves in the Ionosphere," Cambridge University Press, London, England, from p. 132; 1961.
[2] D. J. R. Stock and L. J. Kaplan, "Interpretations of Riccati Equations for Some Circuits," Elec. Engrg. Dept., New York University, New York, N. Y.; 2nd Sci. Rept. AF19(604)7486, May, 1961.
[3] I. Sugai, "A class of solved Riccati's equations," *Elec. Commun.*, vol. 37, pp. 56–60; May, 1961.
[4] R. Stapelfeldt and F. J. Young, "The short pulse behavior of lossy tapered transmission lines," IRE TRANS. ON MICROWAVE THEORY AND TECHNIQUES, vol. MTT-9, pp. 290–296; July, 1961.
[5] D. C. Stickler, "A note on Sugai's class of solutions to Riccati's equation," PROC. IRE (*Correspondence*), vol. 49, p. 1320; August, 1961.
[6] I. Sugai, "Riccati's and Bernoulli's equations for nonuniform transmission lines," IRE TRANS. ON CIRCUIT THEORY (*Correspondence*), vol. CT-9, pp. 359–360; September, 1961.
[7] I. T. Kolodner, "Bounds for solutions of the Riccati equation," *Am. Math. Monthly*, vol. 68, pp. 766–769; October, 1961. Note: Prob. No. 7, Part II of the 1960 W. L. Putnam Mathematical Competition (See *Am. Math. Monthly*, September, 1961) also treats a similar item.
[8] A. K. Rajagopal, "On the generalized Riccati equation," *Am. Math. Monthly*, vol. 68, pp. 777–779; October, 1961.
[9] N. W. McLachlan, "Ordinary Non-Linear Differential Equations in Engineering and Physical Sciences," Clarendon Press, Oxford, England, 2nd ed., p. 13; 1956.
[10] I. Sugai, "Riccati's nonlinear differential equation," *Am. Math. Monthly*, vol. 67, pp. 134–139; February, 1960.

On a New Security Problem*

STATEMENT OF THE PROBLEM

If we are successful in the near future in contacting an alien culture some light years distant, the following facts, which are immediately apparent, should be considered.

1) We will be participating in such communication at the earliest date that it would be technologically feasible for us to do so, *i.e.*, we would be beginners.
2) It would be positing an absurd coincidence to suggest that our interstellar correspondent would be in the same situation. In other words, arguing on probabilistic grounds, it is almost certain that the alien culture would be some centuries, millennia, or eons ahead of us in the development of physical science.
3) It follows therefore that any military contact between the communicants would find us in a position of complete helplessness.

Some weak points of this argument have already received some consideration, as follows.

4) Dr. E. M. Purcell, of Harvard University, has recently (at Green Bank, W. Va.) argued in painstaking fashion that interstellar logistics would require the expenditure of unthinkable amounts of energy over great periods of time and would therefore be impractical.
5) One may speculate that advanced cultures will have reached the point where their ability to create material wealth will have saturated their demand for it. Many social scientists have expressed the opinion that human society will lose its pugnacious characteristics when we have reached such an elevated level of technology. If this is true of human cultures, it may also be a universal law. In that case, we would have nothing to fear from contact with any culture that is advanced enough to solve the problems of interstellar logistics.

TENTATIVE CONCLUSION

While arguments 4) and 5) may find many adherents, I would like to emphasize that 3) can be affirmed with almost complete certainty, whereas 4) and 5) are highly speculative, and do not have the fail-safe characteristic. It follows, therefore, that

6) As a matter of practical security, immediate steps should be taken to ensure that no irresponsible attempts will be made to effect two-way interstellar communication. No broadcasts should be made. If informative signals are received from an intelligent stellar source, they should not be answered, but every effort should be made to decode them and to study their implications.

DISCUSSION OF POSSIBLE OBJECTIONS

I think that many scientists will be inclined to re-assert Dr. Purcell's argument with great emphasis, insisting that he has demonstrated very fundamental limits on the possibilities of interstellar travel, in respect to both the energy cost and the extreme length of the trip.

Anticipating such a conflict of opinion, I would like some consideration given to the following points.

7) The duration of the trip is objectionable to us because it represents, even under the most favorable assumptions, a substantial fraction of our life span. This is a highly anthropocentric point of view. It does not follow that an alien would take the same viewpoint. Some but not all of the reasons why he might see the matter differently are
   a) He may have a greatly longer life span.
   b) He may have cultural attitudes which do not lead to impatience under trip conditions. For example the military expansion of his culture might be undertaken not to expand his economy but by reason of instinct—a sort of generalized procreative drive. (It has even been suggested that this is true of human beings.)
   c) If his culture is able to supply normal material and social comforts during the trip then it is hard to see why any objection would arise.
8) Regarding the energy cost, Dr. Purcell has assumed total conversion of matter to energy, presumed to be the "ultimate" energy source, and found that one proposed trip would require some 400,000 tons of matter and antimatter as fuel. I would like him to give his reasons for supposing that a wealthy alien would consider this an expensive trip. How many alien bank accounts has Dr. Purcell audited?
9) The arguments that a proposed action is impossible may take two different forms. A mathematical impossibility is absolute; thus it will always be true that the rules of Euclidean geometry do not permit the exact trisection of a general angle with straight edge and compass alone. But Dr. Purcell's argument is not of this type.

   He assumes that he can describe the ultimate energy source though he cannot yet make use of it. While this notion has a definite aura of plausibility, it is far from rigorous. The history of science has shown that this general type of argument is a very hazardous one to use because the advance of technology has a way of rendering such arguments obsolete without directly proving them false. While we might indulge in a little science fiction to turn up possible reasons why Dr. Purcell's argument may eventually fail, I would prefer a more rigorous approach.

   In recent decades, mathematicians have been working on a "theory of confirmation" in order to answer questions of this type: to exactly what degree can a theory be said to be confirmed by known facts? We suggest therefore that a study of Dr. Purcell's argument be made along these lines.

FINAL CONCLUSION

If such a study should assign a low-confidence figure to the unreasonable-energy-cost argument (I, for one, think that it would), this would greatly reinforce the request for rigorous security precautions.

It seems to me merely a matter of common sense that we should be very careful about our first contact with a nonhuman culture, especially since it appears so likely that they would be much superior to us in experience, scientific knowledge, and weaponry. We may have only one chance to display our common sense in this matter.

PHILIP R. GEFFE
Axel Electronics, Inc.
Jamaica, N. Y.

* Received February 6, 1962.

The Effects of Doppler Dispersion on Matched Filter Performance*

We have noted with interest the recent correspondence of Ramp and Wingrove,[1] and Cook and Heiss,[2] commenting on the effects of Doppler dispersion on "linear FM-pulse compression." Since linear FM-pulse compression falls within the broader domain of matched filtering, the effects of Doppler dispersion can be treated with greater generality by modifying previous expressions for the "matched filter uncertainty function,"[3] in which Doppler has been approximated by a simple frequency translation.

The Doppler effect results from a continuous compression or expansion of the path length traversed by a signal during transmission. Mathematically this is exactly represented by a change in time scale and change in amplitude (to conserve energy) as follows:

$$f(t) \rightarrow \sqrt{\alpha} f(\alpha t). \tag{1}$$

It is shown first that the time-bandwidth product of a signal is not modified by Doppler. Following Gabor,[4] the normalized mean and mean-square pulse widths of a Doppler shifted signal are defined by the moments of $\sqrt{\alpha} f(\alpha t)$:

$$\overline{t_\alpha^n} = \frac{1}{E}\int_{-\infty}^{\infty} t^n \left| \sqrt{\alpha} f(\alpha t) \right|^2 dt = \overline{t^n}/\alpha^n \tag{2}$$

where E is the signal energy. The effective duration of a Doppler modified signal Δt_α in terms of the unshifted ($\alpha = 1$) signal duration Δt is found from Gabor's defining relation and our (2) to be

$$(\Delta t_\alpha)^2 = \overline{t_\alpha^2} - (\overline{t_\alpha})^2 = \frac{1}{\alpha^2}\left[\overline{t^2} - (\overline{t})^2\right]$$

$$= (\Delta t)^2/\alpha^2. \tag{3}$$

Similarly, the mean and mean-square bandwidths of $\sqrt{\alpha} f(\alpha t)$ are related by Gabor to the moments of $G(w/\alpha)/\sqrt{\alpha}$, which is defined as twice the positive spectrum of $\sqrt{\alpha} f(\alpha t)$, as follows:

$$\overline{w_\alpha^n} = \frac{1}{2E}\int_0^{\infty} w^n \left| \frac{1}{\sqrt{\alpha}} G\left(\frac{w}{\alpha}\right) \right|^2 \frac{dw}{2\pi}$$

$$= \alpha^n \overline{w^n}. \tag{4}$$

In a similar way, the effective bandwidth of a Doppler shifted signal Δw_α in terms of the unshifted ($\alpha = 1$) signal bandwidth Δw is found from Gabor's defining relation and our (4) to be

$$(\Delta w_\alpha)^2 = \alpha^2 (\Delta w)^2. \tag{5}$$

The time-bandwidth product of a Doppler shifted signal in terms of the unmodified waveform may be found from (3) and (5) to be:

$$\Delta t_\alpha \Delta w_\alpha = \left(\frac{\Delta t}{\alpha}\right)(\alpha \Delta w) = \Delta t \cdot \Delta w. \tag{6}$$

Since the time-bandwidth product of a signal is not altered by Doppler, its dimension in signal space is unchanged.

An approximation to include the effects of Doppler dispersion will now be developed. A narrow-band Doppler shifted signal may be written in polar complex notation as

$$\sqrt{\alpha} f(\alpha t) = \sqrt{\alpha}\, \mathrm{Re}\left\{ a(\alpha t) e^{j[w_0 \alpha t + \phi(\alpha t)]} \right\}, \tag{7}$$

where $a(t)$ is the envelope function, $\phi(t)$ the phase modulation, and w_0 the carrier frequency. Since α is always very close to unity, it is usually replaced by $(1+\delta)$, where δ is very closely approximated by

$$\delta \approx \pm 2v/c. \tag{8}$$

In (8), v is the rate of change of effective transmission length and c represents the velocity of propagation in the transmission medium.

If $a(t)$ and $\phi(t)$ are well-behaved functions, $a(t+\delta t)$ and $\phi(t+\delta t)$ may be expanded in a Taylor series about $(\delta t = 0)$:

$$a(t + \delta t) = a(t) + \dot{a}(t)\delta t + O[(\delta t)^2] \tag{9}$$

$$\phi(t + \delta t) = \phi(t) + \dot{\phi}(t)\delta t + O[(\delta t)^2]. \tag{10}$$

Substituting (9) and (10) into (7) and ignoring terms in $(\delta t)^2$ or higher, results in

$$\sqrt{\alpha} f(\alpha t) = \sqrt{\alpha}\, \mathrm{Re}\left\{ \left[a(t) + \dot{a}(t)\delta t \right] \cdot e^{j[w_0 t + w_0 \delta t + \phi(t) + \dot{\phi}(t)\delta t]} \right\}. \tag{11}$$

The term $\exp\{jw_0 \delta t\}$ represents the earlier Doppler approximation, for which $\delta w_0 = w_d$. Assuming $\dot{\phi}(t)\delta t$ is small, the exponential containing this term may be replaced by the first two terms of its series expansion. Neglecting terms in $(\delta t)^2$, (11) become

$$\sqrt{\alpha} f(\alpha t) = \sqrt{\alpha}\, \mathrm{Re}\left\{ \left[a(t) e^{j(w_0 t + \phi(t) + w_d t)} \right] \cdot \left[1 + \frac{\dot{a}(t)}{a(t)}\delta t + j\dot{\phi}(t)\delta t \right] \right\}. \tag{12}$$

Note that the first bracketed term in (12) approximates Doppler by a linear frequency translation, while the remainder contains first-order correction terms reflecting the effects of frequency dispersion.

Applying the modified Doppler approximation to Siebert's[3] "uncertainty function" yields

$$\psi_m(\tau, w_d) = \frac{\left| \sqrt{\alpha} \int_{-\infty}^{\infty} u(t) u^*(t+\tau) e^{jw_d t} \left\{ 1 + \frac{\dot{a}(t)}{a(t)}\delta t + j\dot{\phi}(t)\delta t \right\} dt \right|}{\int_{-\infty}^{\infty} |u(t)|^2 dt}, \tag{13}$$

where

$$\int_{-\infty}^{\infty} |u(t)|^2 dt = 2E, \tag{14}$$

and

$$u(t) = a(t) e^{j\phi(t)}. \tag{15}$$

The absolute magnitude of the first term in the integral is the (conventional) "uncertainty function" $\psi(\tau, w_d)$ in which the effect of Doppler has been approximated by a linear translation.

An upper bound on the difference between (13) and $\psi(\tau, w_d)$ can be found by neglecting $\sqrt{\alpha}$, which is approximately unity, and rewriting (13) as an inequality:

$$\psi_m(\tau, w_d) \leq \psi(\tau, w_d) + \frac{|\delta| \int_{-\infty}^{\infty} [|t|\, a(t+\tau)][\dot{a}^2(t) + a^2(t)\dot{\phi}^2(t)]^{1/2} dt}{2E}. \tag{16}$$

Hölder's Inequality states that

$$\int |f \cdot g| \leq \left\{ \int |f|^p \right\}^{1/p} \cdot \left\{ \int |g|^{p'} \right\}^{1/p'}, \tag{17}$$

where f and g are members of class L^p and $L^{p'}$ respectively. Since the two bracketed expressions in the second term of (16) will usually be members of Class L^2 for pulsed signals, (16) becomes

$$\psi_m(\tau, w_d) \leq \psi(\tau, w_d) + |\delta| \left\{ \frac{\int_{-\infty}^{\infty} [|t|\, a(t+\tau)]^2 dt}{2E} \right\}^{1/2} \cdot \left\{ \frac{\int_{-\infty}^{\infty} [\dot{a}^2(t) + a^2(t)\dot{\phi}^2(t)] dt}{2E} \right\}^{1/2}. \tag{18}$$

The second bracketed integral in (18) has been interpreted in the literature[5] as the mean-square frequency about w_0, which is equal to the effective bandwidth Δw when $a(t)$ is symmetrical and $\dot{\phi}(t)$ is zero or has odd symmetry. This relation follows from a substitution of the Fourier transform of (15) into (4). Defining $a(t)$ from $-T/2$ to $T/2$, an upper bound on the integral in the first bracket is found to be $T^2/4$. Hence (18) may be rewritten as

$$\psi_m(\tau, w_d) \leq \psi(\tau, w_d) + |\delta|\, \tau \Delta w/2 < \psi(\tau, w_d) + |\delta| \frac{k}{2}(\Delta t \cdot \Delta w), \tag{19}$$

* Received March 9, 1962.

[1] H. O. Ramp and E. R. Wingrove, Jr., "Performance degradation of linear FM-pulse-compression systems due to the Doppler effect," PROC. IRE (*Correspondence*), vol. 49, pp. 1693–1694; November, 1961.

[2] C. E. Cook and W. H. Heiss, "Linear FM-pulse compression Doppler distortion effects," to be published in PROC. IRE (*Correspondence*).

[3] W. McC. Siebert, "A radar detection philosophy," IRE TRANS. ON INFORMATION THEORY, vol. IT-2, pp. 204–221; September, 1956.

[4] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pt. III, pp. 429–457; November, 1946.

[5] P. Bello, "Joint estimation of delay, Doppler and Doppler rate," IRE TRANS. ON INFORMATION THEORY, vol. IT-6, pp. 330–341; June, 1960.

where k relates the effective pulse duration Δt to the actual pulse duration T. Note that the difference term in (19) is proportional to the time-bandwidth product, envelope shape factor k, and δ which is related to velocity.

For a pulse signal with a rectangular envelope, k is approximately 3.5. Assuming relative velocities in the order of 17,000 miles per hour, $|\delta|$ is approximately 0.5×10^{-4}. For this case, the conventional and modified uncertainty expressions will differ by less than 10 per cent for all values of (τ, w_d), if

$$\Delta t \cdot \Delta w < 1100. \tag{20}$$

For a specific value of w_d, the "uncertainty function" represents the output waveform from a matched filter. Hence the above result provides a bound on the reduction in matched-filter peak signal output when the conventional Doppler approximation is used. It is clear that for very large time-bandwidth products and large Doppler velocities, a better Doppler approximation, such as (13), should be employed.

W. L. Rubin
J. V. DiFranco
Advanced Studies Dept.
Surface Armament Div.
Sperry Gyroscope Co.
Great Neck, N. Y.

Mechanical Quadratures to Synthesize Nonuniformly-Spaced Antenna Arrays*

A new method of synthesizing nonuniformly-spaced antenna arrays has been developed and several examples using different field patterns have been used to verify this synthesis technique. In the examples chosen, broadside directive radiation patterns with constant phase variation have been used to synthesize nonuniformly-spaced symmetric arrays.

For a finite aperture of over-all width $2a$, the radiation pattern can be written as

$$F(u) = \int_{-a}^{a} g(x) \cos uxdx, \tag{1}$$

where $u=\pi \sin\theta$, θ is the angle measured from the aperture normal and x is the distance from the origin to any point along the aperture measured in half wavelengths. The constant involved has been suppressed, as the field pattern is to be normalized in its final form.

The integral involved in (1) can be evaluated by mechanical quadratures in a form which can be identified with the formulation of an array of discrete elements located within the interval of integration. The specific form of the mechanical quadrature[1] used is

$$\int_a^b f(x)dx = \sum_{j=1}^{n} H_j f(a_j), \tag{2}$$

where the values of a_j and H_j can be evaluated uniquely. As indicated in (2), a suitably weighted sum of n values of $f(a_j)$ can evaluate the integral exactly if $f(x)$ can be written as a polynomial of degree less than or equal to $(2n-1)$. Kopal[1] has shown that the n values of a_j are real, distinct and within the interval (a, b). In addition all values of H_j are real and positive.

Identifying $f(x)$ in (2) with the integrand in (1), $f(x)=g(x)\cos(ux)$, and remembering that u is a constant for this integral, then $F(u)$ can be written as

$$F(u) = \sum_{j=1}^{n} H_j g(a_j) \cos(ua_j), \tag{3}$$

which can be readily identified as the formulation for an array of $2n$ discrete elements with spacings governed by the n values of a_j.

A more general approach to the array synthesis procedure involves the use of weight functions. In this form the mechanical quadrature[1] becomes

$$\int_a^b w(x)f(x)dx = \sum_{j=1}^{n} H_j f(a_j). \tag{4}$$

The n values of a_j are the roots of a polynomial orthogonal with respect to the weight function $w(x)$ on the interval (a, b). The n values of H_j can be determined by solving the set of linear equations generated by letting $f(x)=x^k$ for $k=0, 1, 2, \cdots, 2n-1$. In this form it is more efficient to identify the weight function as the aperture excitation $g(x)$. The only requirement on the weight function is that it does not vanish within the interval of integration and that it is integrable. Using this notation the Fourier integral in (1) becomes

$$F(u) = \int_{-1}^{1} g(y) \cos uydy = \sum_{j=1}^{n} H_j \cos(ua_j), \tag{5}$$

where the transformation $ay=x$ has normalized the limits of integration. In this form the element positions are given by the n values of a_j and the current excitation coefficients are given by the n values of H_j.

One of the radiation patterns suggested by Unz[2] is the function

$$F(u) = e^{-(au/\pi)^2} \cos\left(\frac{2b}{\pi} u\right). \tag{6}$$

An advantage of this pattern is the independent control of beamwidth and sidelobe level. The constants a and b are determined from the design parameters for the beamwidth and sidelobe level and the relationships may be found in the original paper.

From (6) and the Fourier transform, $g(x)$ can be determined[3] and is given by

$$g(x) = K_1 e^{-K_2 x^2} \cosh(K_3 x), \tag{7}$$

where K_1, K_2, and K_3 are constants which have been evaluated. Using (1) and (7), $F(u)$ in (6) can be written as

$$F(u) = \frac{K_1}{\sqrt{K_2}} \int_{-\infty}^{\infty} e^{-y^2} \cosh\left(\frac{K_3}{\sqrt{K_2}} y\right) \cdot \cos\left(\frac{u}{\sqrt{K_2}} y\right) dy \tag{8}$$

by a suitable change of variable. Identifying e^{-y^2} as the weight function $w(y)$ in (4), the field pattern $F(u)$ can be expressed by

$$F(u) = K \sum_{j=1}^{n} H_j \cosh\left(\frac{K_3}{\sqrt{K_2}} a_j\right) \cdot \cos\left(\frac{u}{\sqrt{K_2}} a_j\right). \tag{9}$$

Kopal[1] has shown that the n values of a_j in this case, $w(y)=e^{-y^2}$, are the roots of the classic Hermite polynomial of degree n. The values of a_j and H_j are tabulated in Kopal.[1] Several examples using different beamwidths and sidelobe levels have been calculated and they give very good results, which will be published at a later date.

Another pattern used with excellent results is the approximation to the Dolph-Tchebycheff distribution given by Taylor.[4] In this case the orthogonal polynomials were generated by approximating the aperture excitation by a polynomial $w(x)$ and using this entire function as the weight function in the integral in (4). An alternative approach which has been used for the generation of the aperture distribution $g(x)$ employs the envelope of the standard tabulated[5] Dolph-Tchebycheff current distributions for uniformly-spaced arrays.

Programs have been written and used for several exponential and Dolph-Tchebycheff radiation patterns on an IBM 650 digital computer at the University of Kansas. In each case the results were within the prescribed design limits. Detailed analysis of the above method and more complete numerical results will be published at a later date.

In reference to an abstract of a technical paper[6] which was presented at the 1962 IRE International Convention, it appears that some work along similar lines is being done independently by Dr. Y. T. Lo at the University of Illinois.

J. D. Bruce
Research and Development
Apparatus Division
Texas Instruments, Incorporated
Dallas, Texas
H. Unz
Consultant to Texas Instruments, Inc.
Electrical Engineering Department
University of Kansas
Lawrence, Kan.

* Received March 15, 1962; revised manuscript received, March 22, 1962.

[1] Z. Kopal, "Numerical Analysis," Chapman and Hall, Ltd., London, England; 1955.

[2] H. Unz, "Array Current Distribution Which Will Produce Near Optimum Broadside and Scanning Narrow Radiation Beam," Res. and Dev. Dept., Apparatus Div., Texas Instruments Incorporated, Dallas, Tex., Interim Sci. Rept. No. 3, for the period May 1, 1961–July 31, 1961; to be published.

[3] H. B. Dwight, "Tables of Integrals and Other Mathematical Data," The Macmillan Co., New York, N. Y., see especially 863.3, p. 201; 1947.

[4] T. T. Taylor, "Desigh of line-source antennas for narrow beamwidth and low side lobes," IRE Trans. on Antennas and Propagation, vol. AP-3, pp. 16–28; January, 1955.

[5] L. B. Brown and G. A. Scharp, "Tschebyscheff Antenna Distribution, Beamwidth, and Gain Tables," Naval Ordnance Lab., Corona, Calif., NAVORD Rept. 4629; February 28, 1958.

[6] Y. T. Lo, "A Spacing Weighted Antenna Array," presented at the 1962 IRE International Convention.

Analysis of Nonlinear Sampled-Data Systems by Frequency Response Method*

Sampled-data control systems containing nonlinear element have been analyzed by both Tou with z-series method[1] and Kinnen with multirate sampling method.[2] Now in this correspondence another method by frequency response is introduced. Fig. 1 shows an error-sampled nonlinear system in which the nonlinear element is in series with $G(s)$. If the nonlinear element is temporarily left out, we have for the linear system

$$C^*(s) = \frac{G^*(s)}{1 + GH^*(s)} R^*(s). \quad (1)$$

Then for a complex input $r(t) = e^{j\omega t}$, the output of the system in terms of z transforms becomes

$$C(z) = z \frac{G(z)}{(z - e^{j\omega T})[1 + GH(z)]} \quad (2)$$

where T is the sampling period. Now if $G(z)/(z-e^{j\omega T})[1+GH(z)]$ can be expressed as

$$= c_0/(z - z_0) + c_1/(z - z_1) + c_2/(z - z_2) + \cdots + c_n/(z - z_n) \quad (3)$$

then with the aid of L'Hospital's Rule we can prove for stable systems where z_r's <1 that at sampling instants

$$c(t)_{t\to\infty} = \frac{G^*(j\omega)}{1 + GH^*(j\omega)} e^{j\omega t}. \quad (4)$$

Now put the nonlinear block back to the system and let its output be represented by

$$y^*(e) = N_l e^*(t) + f^*(e) \quad (5)$$

where $e^*(t)$ is the input and N_l is a linear transfer function being either a constant or a complex quantity. Thus the nonlinear system can be replaced by an equivalent linear system (Fig. 2) in which its total output $c(t)$ is divided into two parts: $c_1(t)$, the system output when $f(e)=0$ and $\delta(t)$, the correction term due to $f(e)$. Thus according to (4)

$$c_1(t)_{t\to\infty} = \frac{N_l G^*(j)}{1 + N_l G^*(j)} \cos(t + \gamma_0) \quad (6)$$

when $r(t)=\cos t$ and $H(s)=1$. Suppose $e_1(t)$ represents the error when $f(e)=0$, then we have for $r(t) = \cos t$ (using subscript c)

$$e_1(t) = r(t) - c_1(t) = a_0 \cdot \cos(t + \theta_0) \quad (7)$$

and

$$e(t) = e_1(t) - \Delta_c(t). \quad (8)$$

Thus for a particular nonlinear element whose $f(e)=-e^3$, we have an approximate expression through (8)

$$f(e) = e_1^3 - 3e_1^2\Delta_c(t). \quad (9)$$

Now considering only the fundamental and third-harmonic terms of $\Delta_c(t)$ we assume

$$\Delta_c(t) = b_1 \cos(t + \theta_1) + b_3 \cos(3t + \theta_3). \quad (10)$$

* Received March 12, 1962; revised manuscript received March 26, 1962.

[1] J. Tou, "Analysis of sampled-data systems containing nonlinear element," PROC. IRE (*Correspondence*), vol. 46, p. 915; May, 1958.

[2] E. Kinnen, "An Analysis of Nonlinear Sampled-Data Control Systems," Ph.D. dissertation, Purdue University, Lafayette, Ind.; 1958.

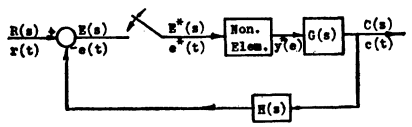


Fig. 1—A basic nonlinear sampled-data control system.

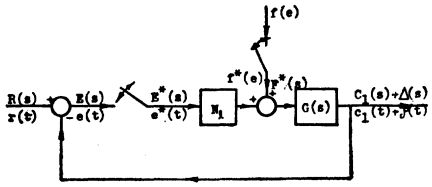


Fig. 2—The equivalent linear sampled-data control system.

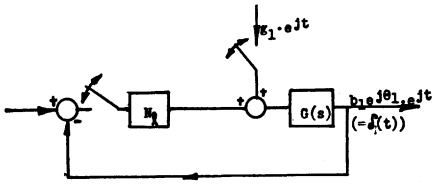


Fig. 3—The system considering $g_1 \cdot e^{tj}$ only.

Then substituting (7) and (10) into (9) gives

$$f_c(e) = -a_0^3 \cdot \cos^3(t + \theta_0) + 3a_0^2 \cdot \cos^2(t + \theta_0)\,\Delta_c(t) \quad (11)$$

$$= f_{1c}(t) + f_{3c}(t), \quad (12)$$

where $f_{1c}(t)$ and $f_{3c}(t)$ contain terms associated with $\cos t$ and $\cos 3t$ respectively. Now consider an input $r(t)=\sin t$ and let

$$\Delta_s(t) = b_1 \sin(t + \theta_1) + b_3 \sin(t + \theta_3), \quad (13)$$

similarly we obtain

$$f_s(e) = f_{1s}(t) + f_{3s}(t). \quad (14)$$

Then for a complex input $r(t) = e^{jt}$ we have

$$f(e) = f_c(e) + jf_s(e) \quad (15)$$

$$= g_1(b_1 e^{j\theta_1}, b_3 e^{j\theta_3}) e^{jt} + g_3(b_1 e^{j\theta_1}, b_3 e^{j\theta_3}) e^{j3t}. \quad (15a)$$

So according to (4) we finally obtain (Fig. 3)

$$b_1 e^{j\theta_1} \cdot e^{jt} = \frac{G^*(j)}{1 + N_l G^*(j)} (g_1 \cdot e^{jt}). \quad (16)$$

Now write

$$b_1 e^{j\theta_1} = R_1 + jI_1$$
$$b_3 e^{j\theta_3} = R_3 + jI_3. \quad (17)$$

Then by equating respectively the real and imaginary terms on both sides of (16) we have

$$k_1R_1 + k_2R_3 + k_3I_1 + k_4I_3 = k_5$$
$$n_1R_1 + n_2R_3 + n_3I_1 + n_4I_3 = n_5. \quad (18)$$

where all k_n's and n_n's are constants. Similarly, when g_3 is considered, another set of equations can be drawn for the third-harmonic components. Thus b_1, b_3, θ_1 and θ_3 can be solved and both $\delta(t)$ and $c(t)$ can be found. The $c(t)$ thus obtained is valid only at sampling instants.

C. Y. LEE
Aeronautical Research Lab.
P.O. Box 63, Taichung
Taiwan, China

The Microwave Diagnosis of a Column of Ionized Gas*

We shall present a method of determining the electron distribution in a dilute column of ionized gas. The term "dilute" means here that the electron density is small enough that each electron behaves as an independent scatterer and that the frequency ν of collisions between electrons and neutral particles is much smaller than the frequency ω of the incident wave. These conditions ensure that the gas column may, to a good approximation, be considered to be uniformly irradiated by the incident wave, which we take to be plane, transverse-magnetic and incident normal to the column's axis.

The signal scattered from the column is assumed to be received by a receiver situated at a distance R from the axis and at an angle α from the direction of the incident wave vector k. The electron density N is independent of the coordinate z along the axis and of the azimuth about the axis. Thus, $N=N(r)$, where r is the radial distance from the axis. The column has a radius $a \ll R$ and is assumed to be sufficiently long that the scattered signal at the receiver is virtually that which is scattered by an infinitely long column.

Under the above conditions, Eshleman[1] has shown that the back-scattered electric field ($\alpha=\pi$) is approximately

$$E_z^s|_\pi = A\,|E_z^i|\,\exp\left[i\left(kR - \frac{\pi}{4} - \omega t\right)\right] \cdot \left(\frac{2}{\pi kR}\right)^{1/2} \int_0^a N(r) J_0(2kr)\,rdr, \quad (1)$$

where $A = -i\mu_0 e^2\pi/2m_e$, $|E_z^i|$ is the absolute value of the incident field and J_0 is the zeroth-order Bessel function; μ_0, e, and m_e are the permeability of space and the charge and mass of an electron, respectively. We define the scattering coefficient ρ_α by the relation

$$E_z^s|_\alpha = |E_z^i| \left(\frac{2}{\pi kR}\right)^{1/2} \rho_\alpha; \quad (2)$$

therefore, the back-scattering coefficient is

$$\rho_\pi(k) = A \int_0^a N(r) J_0(2kr)\,rdr. \quad (3)$$

It is easy to extend (1) to the forward scattering case,[2] $\alpha=0$. The result is

$$E_z^s|_0 = A\,|E_z^i|\,\exp\left[i\left(kR - \frac{\pi}{4} - \omega t\right)\right] \cdot \left(\frac{2}{\pi kR}\right)^{1/2} q/2\pi, \quad (4)$$

where q is the number of electrons per unit length,

$$2\pi \int_0^a Nrdr.$$

Thus

$$\rho_0 = Aq/2\pi. \quad (5)$$

* Received March 19, 1962. This work was supported by the Army Rocket and Guided Missile Agency, Contract DA-04-495-ORD-3112.

[1] V. R. Eshleman, "The Mechanism of Radio Reflections from Meteoric Ionization," Electronics Res. Lab., Stanford University, Calif., Tech. Rept. No. 49; July, 1952.

[2] J. R. Barthel, "The Microwave Diagnosis of a Column of Ionized Gas," General Dynamics/Convair, San Diego, Calif., Rept. No. ZPh-096; May, 1961.

In the narrow column (or long wave) limit, $ka \ll 1$, we find $\rho_0 = \rho_\pi$.

If we determine $\rho_\pi(k)$ experimentally, we may find $N(r)$ from (3) by applying the Fourier-Bessel integral,[3] which states

$$f(z) = \int_0^\infty J_m(kz)kdk \cdot \int_0^\infty f(\xi)J_m(k\xi)\xi d\xi \quad (m > -\tfrac{1}{2}). \quad (6)$$

By using (3) in (6), we find

$$N(r) = (4/A)\int_0^\infty J_0(2kr)\rho_\pi(k)kdk. \quad (7)$$

Eq. (7) holds only in a dilute column. The values of k at which this model is inapplicable must be sufficiently limited that the $\rho(k)$ at those values make a negligible contribution to the integral in (7).

At any point in the column, the dielectric constant

$$K = 1 - \omega_p^2/\omega^2 = 1 - Ne^2/m_e\epsilon_0\omega^2$$

becomes negative when $\omega < \omega_p(r)$. In a region where $K < 0$, the incident wave attenuates, contrary to our assumption. Let δ be the distance over which the wave attenuates by $1/e$ (skin depth). If $\omega_p(\max)/c = k'$, δ is real for $k > k'$. As $k \to 0$,

$$\delta \to \delta_{\min} = c/\omega_p(\max) = 1/k'. \quad (8)$$

Now suppose $\delta_{\min} \gg a$, or $ka' \ll 1$; then attenuation can be neglected even at very low frequencies. This condition implies that

$$m_e\epsilon_0c^2/e^2 \gg N_{\max}a^2 \quad (> q/\pi). \quad (9)$$

Since $m_e\epsilon_0c^2/e^2 = 10^{14}/\pi$ electrons per meter, the condition $q \ll 10^{14}/m$ is a necessary (but not sufficient) condition for $\delta_{\min} \gg a$. $\delta_{\min} \gg a$ is sufficient for the validity of (7).

We have used k to mean vacuum wave number ω/c. In the plasma the wave number is

$$\bar{k} = \frac{\omega}{c}\sqrt{1 - \frac{\omega_p^2}{\omega^2}}.$$

An additional requirement for the validity of (7) is, therefore, that the contribution to the integral is negligible for values of k for which $(k - \bar{k})a$ is of the order unity or greater. If $\omega \geq \omega_p$ (max),

$$(k - \bar{k})a = k'a\left(\frac{\omega}{\omega_p} - \sqrt{\frac{\omega^2}{\omega_p^2} - 1} \leq k'2.\right.$$

Therefore if $k'a \ll 1$, significant phase errors are confined to the $\omega < \omega_p$, which does not contribute significantly to (7) if condition (9) is met. If $k'a$ is of the order unity or greater, we must in general take account of the variation of $\bar{k}$ in the gas column, although for certain electron distributions the contribution to (7) of the k region is which there are phase effects may be small.

It is readily seen that collisions have no effect if $\nu \ll k'c$.

As an independent check on the applicability of the dilute column model in a given case, we may use the forward scattered signal (if available) in (4) to determine q. If our model is valid, (4) should give identical values of q for each $k > k'$.

JAMES R. BARTHEL
Space Science Section
General Dynamics/Astronautics
San Diego, Calif.

[3] P. Morse and H. Feshback, "Methods of Theoretical Physics," McGraw-Hill Book Co. Inc., New York, N. Y., p. 766; 1953.

Causality*

Causality as invoked by Deutsch is axiomatic in nature.[1] Then there are several rather fundamental limitations that may be imposed on this monobloc computer:

1) There is no known logical system, whose consistency is proved, that can be used as a basis for the computer.[2]
2) The mandatory uncertainties in the computer's knowledge of the monobloc's structure must be such that no future event is thereby affected. Heisenberg's Principle of Quantum Indeterminancy may be invoked during the time that the computer is acquiring data about the formation of the monobloc. The simultaneous invocation of a monobloc and a supercomputer would be pressing a *deus ex machina* too far.

One should more properly infer that, given our local solar system, one might predict the ineluctable evolution of life. In any event, predictability does not necessarily alter the operative value of such concepts as free will and human creativity. If there is no physically realizable means of executing such predictability, these actions and events remain unpredictable, even to ourselves.

This conversely means that the boundary level of realizable predictability is worthy of investigation and is capable of some quantification.

Although some men seem to be intelligent, it seems too early to ascribe this quality to man. It appears that social activity has a generic propensity to sink to the lowest common denominator. On the other hand it appears incredible that it is impossible to raise the level of this lowest common denominator to at least a survival level.

WALTER B. MORTON, JR.
1245 Siesta St.,
Anaheim, Calif.

* Received March 14, 1962.
[1] S. Deutch, "Causality," PROC. IRE (*Correspondence*), vol. 50, p. 222; February, 1962.
[2] E. Nagel and J. R. Newman, "The World of Mathematics," Simon and Schuster, Inc., New York, N. Y., vol. 3, pt. XI-4; 1956. This is a discussion of K. Goedel's proof from "On Formally Undecidable Propositions of Principia Mathematica and Related Systems."

Comments on "Symmetrical RC Distributed Networks"*

Hellstrom[1] described a class of RC distributed lines where the specific series resistance $r(x)$, was varied along the line, and the ratio $r(x)/c(x)$ was kept constant. It is interesting to note that this class is essentially identical with the uniform line case if only the measuring scale dx is chosen correctly. This becomes evident if one performs the following scale transformation

$$du = \frac{r(x)}{r_0}dx = f(x)dx \quad (1)$$

in the differential equation of the symmetrical network

$$\frac{d^2V}{dx^2} - \frac{f'}{f}\frac{dV}{dx} - j\omega r_0c_0f^2V = 0. \quad (2)$$

The result is recognized as the differential equation governing the uniform line:

$$\frac{d^2V}{du^2} - j\omega r_0c_0V = 0. \quad (3)$$

Consequently the chain parameters of the symmetrical line can be directly obtained from those of the uniform line by the same transformation (1).

On the other hand, the case where the product $r(x)c(x)$ remains constant leads to possible realization of network functions appreciably different from those obtainable with uniform lines. This class of RC distributed lines is governed by

$$\frac{d^2V}{dx^2} - \frac{f'}{f}\frac{dV}{dx} - j\omega r_0c_0V = 0. \quad (4)$$

Unlike (2) only special cases of $f(x)$ will permit the analytical solution of (4). The exponential line is among the most interesting cases of this class. Considering a distributed RC ladder where the series resistance varies exponentially with the distance

$$r(x) = r_0e^{hx} \quad (5)$$

and the shunt capacitance correspondingly,

$$c(x) = c_0e^{-hx}. \quad (6)$$

Eq. (4) then becomes

$$\frac{d^2V}{dx^2} - h\frac{dV}{dx} - j\omega r_0c_0V = 0. \quad (7)$$

Assuming a solution in the form of $\exp(\gamma x)$, the possible values for γ are found,

$$\gamma_{1,2} = \frac{h}{2}\left(1 \pm \sqrt{1 + 4j\frac{\omega c_0r_0}{h^2}}\right) \quad (8)$$

and after defining

$$\gamma_1 = -\alpha_1 - j\beta_1; \qquad \gamma_2 = \alpha_2 + j\beta_2;$$
$$\theta = \tan^{-1}\frac{4\omega r_0c_0}{h^2},$$

one finds that

$$\beta_1 = \beta_2 = \frac{h}{2}\left(\frac{1}{\sqrt{1 - \tan^2\theta/2}} - 1\right) = \beta$$
$$\alpha_1 = \frac{h}{2}\left(\frac{1}{\sqrt{1 - \tan^2\theta/2}} - 1\right)$$
$$\alpha_2 = \frac{h}{2}\left(\frac{1}{\sqrt{1 - \tan^2\theta/2}} + 1\right). \quad (9)$$

γ_1 and γ_2 are the voltage propagation constants in the positive and negative directions of x. It is seen from (9) that the attenuation constant in the positive direction (α_1) is many times smaller than the attenuation constant in the negative direction (α_2), while the phase constants β_1 and β_2 are equal.

The elimination of θ from (9) results in the following relations between α, β and h:

$$\beta = \sqrt{\alpha_1(\alpha_1 + h)} = \sqrt{\alpha_2(\alpha_2 - h)}. \quad (10)$$

* Received March 19, 1962.
[1] M. J. Hellstrom, PROC. IRE (*Correspondence*), vol. 50, pp. 97–98; January, 1962.

Eq. (10) is geometrically represented by the circle of Fig. 1 which is very useful for the design of exponential RC lines. It gives a clear picture of the phase shift that one obtains within the practical limitations of line loss and physical size. For prescribed values of α_1 and h one can find the values of β and α_2 as indicated on the circle. From the uniform line ($h=0$) one always gets $\alpha=\beta$, while from the exponential line a larger phase shift can be obtained for the same value of attenuaton allowed. This effect is greatest for high values of h/α_1.

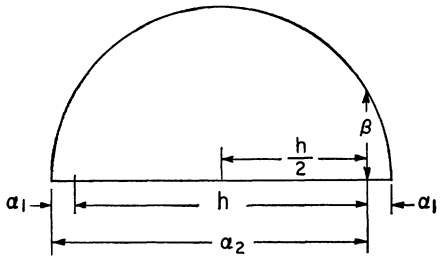


Fig. 1—Geometrical representation of the relations between α_1, α_2, β and h.

If the following approximation is made:

$$e^{-\alpha_2 x} \ll e^{+\alpha_1 x},$$

then the complete solution can be put in a form of a simple chain matrix,

$$\begin{bmatrix} V(0) \\ I(0) \end{bmatrix} = \frac{e^{-\gamma_1 x}}{\gamma_2 - \gamma_1} \begin{bmatrix} \gamma_2 & r_0 \\ j\omega c_0 e^{hx} & -\gamma_1 e^{hx} \end{bmatrix} \begin{bmatrix} V(x) \\ I(x) \end{bmatrix}$$

$$= \frac{\sqrt{\cos\theta}\cdot e^{\alpha_1 x + j(\beta x - \theta/2)}}{h} \cdot \begin{bmatrix} \gamma_2 & r_0 \\ \pi\omega c(x) \dfrac{r(x)}{r_0} (\alpha_1 + j\beta) & \end{bmatrix} \cdot \begin{bmatrix} V(x) \\ I(x) \end{bmatrix}.$$

It is also important to mention that lines of constant RC product can be easily built in the form of resistive thin films of variable width.

J. Pearl
RCA Labs.
Princeton, N. J.

Author's Reply[2]

The above comments afford the opportunity to add a point to my note[1] which presents the requirements on the resistance and capacitance density functions such that the network will be symmetrical *at all lengths*. It is possible to construct networks, not in this class, which are symmetrical at a specific network length, but not at all lengths.

Note that although the performance of the symmetrical network is identical to that of the uniform network (and it is in this sense that the proportional network defines a group of equivalent networks) the network itself is *not identical* since its r and c functions may have unlimited forms, subject to the proportionality requirement.

The transformation suggested by Pearl,

or its equivalent integral form

$$u = \int_0^x f(x)dx,$$

is a special case of one given by Liouville,[3] *i.e.*,

$$z = \int_0^x \sqrt{rc}\, dx.$$

Its usefulness in changing a general nonuniform transmission-line problem (or distributed RC network) into one with constant propagation factor (or RC product) has been noted recently.[4] This same transformation can be used to assist in defining groups of equivalent networks, *i.e.*, networks with the same terminal performance.[5] The proportional network constitutes such a group with the uniform network being a single member.

The case, RC = constant, is not a "Class" of networks in the sense that it does not prescribe anything about network performance. Mr. Pearl's differential equation, (4) is no closer to solution due to the choice RC = constant than the original network equation because the differential equation of any network can always be put into form (4).

The exact exponential-network open-circuit impedance parameters and a discussion of the propagation factor, incidentally, have both been published.[6,7]

M. J. Hellstrom
TV-Radio Div.
Westinghouse Elec. Corp.
Metuchen, N. J.

[2] Received, March 26, 1962.

[3] E. Kamke, "Differentialgleichungen, Lösungsmethoden und Lösungen," Edwards Bros., Inc., Ann Arbor, Mich., p. 261; 1945.

[4] I. Jacobs, "A generalization of the exponential transmission line," Proc. IRE, vol. 47, pp. 97–98; January, 1959.

[5] M. G. Hellstrom, "Equivalent distributed RC networks or transmission lines," IRE Trans. on Circuit Theory Vol.-CT-9, pp. 247–251; September, 1962.

[6] B. L. H. Wilson and R. B. Wilson, "Shaping of distributed RC networks," Proc. IRE, vol. 49, pp. 1330–1331; August, 1961.

[7] W. A. Edson, "Tapered distributed RC lines for phase shift oscillators," Proc. IRE, vol. 49, pp. 1021–1024; June, 1961.

Capacitance Calculations for Several Simple Two-Dimensional Geometries*

A number of situations arise in solid-state device research in which capacitance estimates are required for two-dimensional configurations where edge effects are not negligible, and for which confinement of the displacement to the dielectric medium because of its high permittivity prohibits the use of handbook formulas.[1] Typical of these are the complementary configurations of Fig. 1. In the figure, the constant potential segments of the boundaries are indicated by solid lines, while the dotted lines designate segments on which the normal derivative of the potential vanishes.

Fig. 1(a) represents the right half of an infinite sheet of dielectric of thickness b, having a completely electroded lower surface ($\phi=0$), and bearing an infinitely long strip electrode ($\phi=V$) of width $2a$ on the upper surface. The boundary conditions $\phi_n=0$ result from considerations of symmetry at the left, and of confinement of the displacement to the dielectric sheet at the upper right.

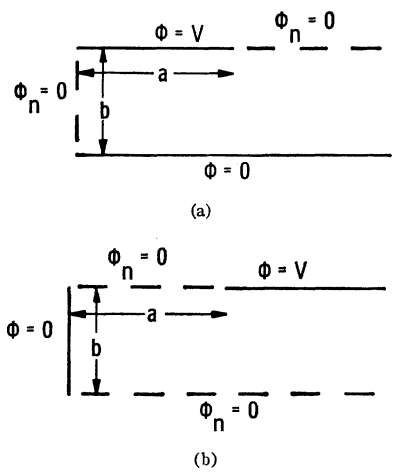


Fig. 1—Boundary conditions.

By placing a z-plane origin at the lower left corner of the figure, one can employ the Schwarz-Christoffel transformation

$$\tanh\frac{\pi z}{2b} = \mathrm{sn}\,(w, k)$$

to map the dielectric region into a K by K' rectangular region in the w plane, where K and K' are complete elliptic integrals of the first kind for moduli $k=\tanh \pi a/2b$ and $k'=(1-k^2)^{1/2}$, respectively.[2] The sides of the w-plane rectangle correspond to the four boundary segments of the z-plane figure, from which it follows that the geometric (dimensionless) capacitance c_λ for Fig. 1(a) is K/K'. Because of the complementary nature of the geometries of Fig. 1(a) and (b), the geometric capacitance of the latter is K'/K. These, of course, must be multiplied by the permittivity of the dielectric in question in order to obtain the desired capacitance per unit length in usable units:

$$C_\lambda = \epsilon c_\lambda = \mathrm{K}\epsilon_0 c_\lambda,$$

where K is the dielectric constant.

Fig. 2 summarizes the above results, as well as those for other configurations, which are solvable by a similar, although somewhat more involved transformation.

Fortunately, in the application of these formulas some rather good approximations are available. These are based on the fact that, for $k\to1$, $K'\to\pi/2$ and $K\to\ln(4/k')$. Thus, in the case of $c_{\lambda 1}$ for $a\gg b$,

$$c_{\lambda 1} \approx (a + 0.44b)/b$$

* Received March 19, 1962.

[1] See, for example, "American Institute of Physics Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., sect. 5b; 1957.

[2] A helpful monograph on this subject is F. Bowman, "Introduction to Elliptic Functions with Applications," John Wiley and Sons, Inc., New York, N. Y.; 1953.

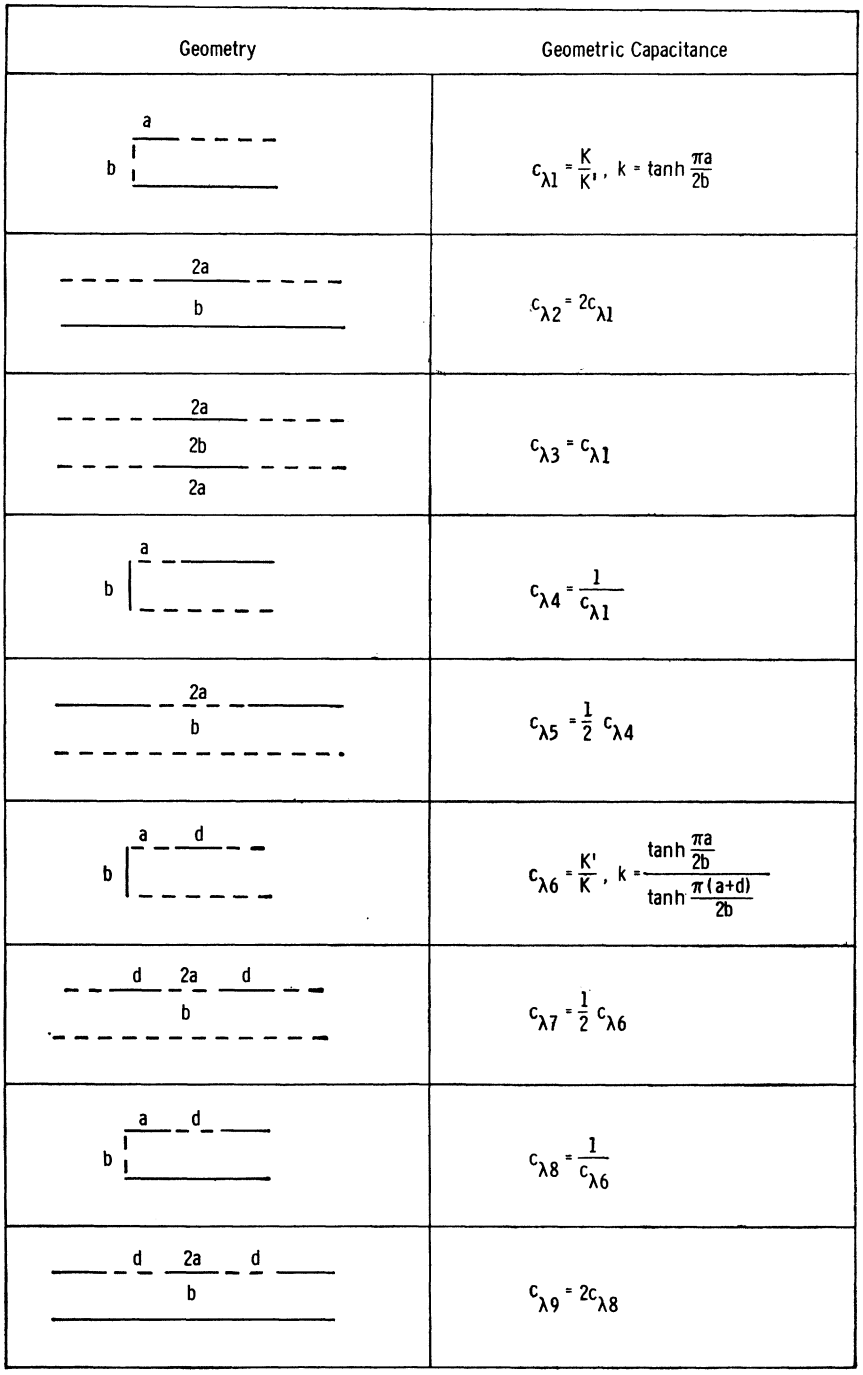


Fig. 2—Capacitances of various geometries.

where 0.44 is $(2 \ln 2)/\pi$. This result is applicable to all of the cases, $c_{\lambda 2}, \cdots, c_{\lambda 5}$, dependent on $c_{\lambda 1}$. Furthermore, although the approximation is derived for $a \gg b$, it is accurate within less than one per cent for $a = b$.

Similarly, $c_{\lambda 6}$ for $a \gg b$ becomes

$$c_{\lambda 6} \approx b/(a + 0.44 + \Delta)$$

$$\Delta = \frac{1}{\pi} \ln \left(\frac{1 + D}{2D} \right)$$

$$D = \tanh \frac{\pi d}{2b}.$$

This approximation was tested for $a = b$, $d = 0.1b$ and found to be accurate to within less than 1 per cent. The term Δ has the value 0.44 for $d = 0.091b$, and a value 10 times smaller for $d = 0.65b$. For this reason, it is practical in most cases for which $d \geq b$ to use the simpler expression for $c_{\lambda 4}$ $(d = \infty)$, rather than $c_{\lambda 6}$.

The derivation of the results herein assumes homogeneity and isotropy throughout a dielectric sheet of infinite extent. In most practical applications, uniformity is satisfied and neglect of the requirement on infinite extent has been found to be of little consequence, probably because significant edge-effect fields extend on the order of only one sheet thickness past the electrode edges (as indicated by the approximation formulas). Discrepancies between calculated and measured values of capacitance have been noted, however, for very high dielectric-constant ferroelectric ceramics with narrow electrodes. Such materials are known to be anisotropic and to possess surface layers having permittivities quite different from those of the bulk. Even in these cases, calculations of capacitance made by the methods described and utilizing the bulk dielectric constant have been found useful for estimating purposes.

P. N. Wolfe
Westinghouse Central Labs.
Pittsburgh, Pa.

Synchronization Error*

That there will be some error when a receiver synchronizes with an incoming noisy signal is intuitively obvious. The magnitude of the error in the case of ideal sync signal and receiver processing is indicated by a simple application of information theory. A great deal of insight into the nature of the problem can be obtained from the analysis.

Let the time period over which the receiver expects to receive the synchronizing signal (whose waveform is exactly known) be L, and let the total energy in the signal be U. If the bandwidth of the signal is B, and the noise is white with density N_0, the noise power received with the signal is BN_0. Shannon's theorem on the capacity of a channel states that the maximum amount of information which can be received is

$$I = LB \ln (1 + U/LBN_0) \text{ natural bits.}$$

It is apparent that I increases with B. If B is very large, the argument of the logarithm is nearly unity, and a good approximation for fixed U/N_0 is

$$\ln (1 + U/LBN_0) \doteq U/LBN_0, \quad B \gg U/LN_0.$$

In the limit

$$I = U/N_0 \text{ natural bits.}$$

This is the maximum amount of information which can be received in time L when the signal has energy U and the noise density is N_0. Furthermore, according to Shannon's results, if the proper signal and reception technique is used, this information is perfectly reliable, *i.e.*, there is zero probability of error.

Now let the information contained in the signal be an indication of the time at which the signal is received. Exactly $n = \exp(U/N_0)$ states can be specified by the information. If the interval L is divided into n equal subintervals,[1] each message state can be assigned to correspond to reception of the signal within one of the subintervals. Thus synchronization occurs at some time within the subinterval corresponding to the signal. If the mid-point of the subinterval is used, the maximum synchronization error e for the

* Received March 12, 1962.

[1] The problem is only academic if n is not an integer.

ideal system is then one-half the width of the subinterval.

$$e = (L/2) \exp(-U/N_0).$$

This result is remarkable in the same way that Shannon's result is remarkable: There is no probability distribution associated with the error; an absolutely firm bound on the ideal error has been achieved. It is also inadequate in the same sense that Shannon's result is. It does not indicate the practical manner by which this error rate can be achieved.

As a corollary to the result note that the same final error occurs whether two synchronizing signals, each with energy U, are sent sequentially with the second, reducing the error left by the first, or whether one signal with energy $2U$ is sent.

C. M. Hackett, Jr.
Light Military Electronics Dept.
Advanced Electronics Ctr.
at Cornell University
GE Company
Ithaca, N. Y.

On Visual Perception and Retinal Motions*

The experimental evidence concerning the basis of visual perception presents a confusing turmoil of data to the designer of pattern recognition systems. The critical question seems to be: with what capacities should the system be "born" and how much of its capability is either required or, better, left to be "learned"? There seems some demonstration of predetermined or innate ability to perceive form, texture, and depth—at least to some nontrivial extent—in human, cat, and goat infants.[1,2] Other evidence suggests an initial period of nervous system "plasticitiy" subsequent to which a "hardening" sets in and precludes development of important properties if there has been sensory deprivation in this initial time of growth; if this is true for man, as it seems to be true for monkey and cat, it provides some small reinforcement of the notion that there is a good deal more specification in the sensory system (visual) than is suggested by such statistical principles as are embodied in the Perceptron. At least it reveals that later perception and/or learning still arises in neural networks with largely fixed, nonstatistical properties and organization. While this would not rule out the importance or desirability of initially "freely-connected" (Rosenblatt) circuitry as a possible context for processes of self-organization, it does raise the question of the theoretical need of statistical procedures in perception or learning if determinate networks have the same abilities.

The importance of retinal motions[1-4] in form recognition has been underestimated. Some of these motions are oscillatory, do not yield to voluntary control, and are not consciously perceived save when the *effects* of their having been stabilized with respect to stimulus objects in experimental situations reveal their presence to subject or experimenter.

Let us assume for a moment an array of receptors containing the representation of an object-pattern as a distribution of receptor excitation intensities, and stipulate that there is no relative motion of any sort between this retinal system and the external stimulus. It is a simple task to project only the contours in the retinal image onto another array A_1 of neurons. A well-known difficulty is involved here in having the system recognize figures which have not been somehow position "normalized," that is, brought into some standardizing region of A_1. So long as the discrimination process is motion independent and depends only on the excitation of special, more-or-less unique sets of neurons, this normalization procedure, in one form or another, is mandatory; it is implicit even in the procedure of repeatedly presenting the object in different positions in the visual field together with some suitable type of response coercion which we may speak of as "learning." The particulars here are inconsequential; rather it is important that a) the system must rely on the classification of the whole pattern presented as a unit entity and not, for example, on the "straightness" or "curviness," etc., of the various regions in the pattern of contours, and b), if a) is true, the spatial structure (geometry) of the stimulus is lost as it must be recoded into a new geometry and energy distribution for internal system representation. By considering only these last two items, a) and b), it may be seen that systems embodying such characteristics can claim little similarity either to structural or functional processes in the visual sensory system of man.

We can begin to demonstrate our assertion most simply and convincingly by contemplating the consequences of item b) above; all we need do is require such a system to "draw," for example, its interpretation of the circle it has been able to perceive (discriminate from all other geometrical entities). Unless it can perform a quite difficult inverse transform, the system's output will be its internal representation of a circle which, to us (man), will be a meaningless assortment of spatial patterns and energy amplitudes. Since man can reproduce his perceptions with rather good accuracy, it seems a safe assumption that no such recoding as has been described occurs in man's sensory system, unless one can postulate with any confidence this unwieldy inverse transform. As for item a), the forms customarily perceived by man are not the same perceived when the retinal oscillations are stabilized[3,4] relative to the stimulus, *i.e.*, there is no perception as we "know" it without these retinal motions under control of the autonomic nervous system.

One can speculate that there is some provision in the visual system for mapping the image on A_1 onto another array A_2 in such a way that the image geometry is preserved to a degree depending on the complexity of the pattern; and that, in general, in a world largely without "simple" patterns, this mapping fails to preserve certain portions of the image on A_1. Therefore, without some motion of the retinal system, there would be some regions of the image that seem to "fade" out in passage from A_1 to A_2. The phenomenon of "fading" and partitioning has been observed experimentally[3,4] with the feeling that those parts of the image that are preserved after retinal system-object image stabilization, are "meaningful" units of perception. Be this as it may, it would appear that the only principle compatible with the preservation of the images of the external world after passage through certain neural network processes and normalization to eliminate variations with size and position, must depend upon the discrimination of such local properties as straightness and curviness. Cognition of pattern then becomes a matter of the conditioning of the association between certain nervous system activity connected with the control of retinal motions (which may seem currently to be a substratum of "noise" in neurophysiological studies) and the geometry-preserved patterns of activity representing stimulus objects.

Robert Bernhard
Research Dept.
Grumman Aircraft Engnrg. Corp.
Bethpage, N.Y.

Demodulation Effect of an Envelope Detector at Low Signal-to-Noise Ratios*

An envelope detector suffers an effective loss of percentage modulation when the signal-to-noise ratio (SNR) approaches unity. The mechanism responsible is discussed, and the degradation of the percentage modulation is computed for different SNR. A method for partially correcting this condition is suggested.

If we think of an envelope detector as a synchronous detector (which in fact it is), then it is readily seen that noise can interfere with the detection process at low SNR. The synchronous feature comes about from the fact that the radio-frequency carrier switches the diode off during negative half cycles and on during positive half cycles. Noise amplitude peaks comparable to the signal amplitude can interfere in two ways. The diode may be switched on during a negative-signal half cycle, or the noise peaks may switch the diode off during a positive-

* Received by the IRE, March 6, 1962.

[1] R. L. Frantz, "The origin of form perception," *Sci. Am.*, vol. 204, p. 66; May, 1961.

[2] E. J. Gibson, "The visual cliff," *Sci. Am.*, vol. 202, p. 64; April, 1960.

[3] R. Ditchburn and D. Fender, "The stabilized retinal image," *Optica Acta*, vol. 2, p. 128.

[4] R. Pritchard, W. Heron, and D. Hebb, "Visual perception approached by the method of stabilized images," *Can. J. Psychol.*, vol. 14, p. 6.

* Received March 13, 1962; revised manuscript received April 3, 1962.

signal half cycle. At large SNR there is little danger of this occurring, but at low ratios (of the order of magnitude of one to one) it is apparent that the normal detection process will be interfered with. The over-all effect is an apparent reduction in the percentage of modulation of the amplitude-modulated radio-frequency signal.

The reduction in the percentage of modulation may be calculated for different SNR where the noise is Gaussian in a narrow band. We are primarily interested in SNR in the neighborhood of one.

Bennett[1] has shown that Gaussian noise in a narrow band has a Rayleigh distribution, and for a noise power of σ^2 watts in the band combined with a sinusoidal carrier signal of amplitude A, the amplitude distribution is given as

$$p(v) = (v/\sigma^2) \cdot \{\exp - [(v^2 + A^2)/2\sigma^2]\} I_0(Av/\sigma^2) \quad \text{for } v \geq 0$$
$$p(v) = 0 \quad \text{for } v < 0. \quad (1)$$

The signal or carrier is unmodulated, and we will determine the output of the detector for this case first. We will then consider modulation as simply varying the amplitude of the carrier, which is actually the case. By considering the amplitude as positive only (or envelope), we essentially introduce the rectification action of a linear detector. We can account for the negative swings by considering a full-wave linear detector. The average value out of the detector is then seen to be

$$\bar{v} = \int_0^\infty v p(v) dv \quad (2)$$

$$\bar{v} = (1/\sigma^2)[\exp - (A^2/2\sigma^2)] \cdot \int_0^\infty v^2 [\exp - (v^2/2\sigma^2)] I_0(Av/\sigma^2) dv. \quad (3)$$

The integral may be found in Foster and Campbell,[2] and after some manipulation is seen to yield

$$\bar{v} = (\sqrt{2\pi}\,\sigma/2)[\exp - (S/N)] \cdot {}_1F_1(3/2; 1; S/N) \quad (4)$$

where

$$S/N = A^2/2\sigma^2$$

the unmodulated SNR. The function ${}_1F_1$ is Kummer's confluent hypergeometric series, and is plotted for these indices in Jahnke and Emde[3] for the range of argument of interest.

This plot is duplicated in Fig. 1 together with the exponential term and the product of both the exponential term and hypergeometric term. Over most of the range of interest, the product of these two is a straight line and this approximation will be made. Eq. (4) then may be written as

$$\bar{v} = \sqrt{2\pi}\,(\sigma/2)[0.4(S/N) + 1]. \quad (5)$$

If we consider the modulation as a slow variation of the carrier, then we need only compute the output as given by (5) for the maximum and minimum of the carrier (assuming sinusoidal modulation). We may then write

$$A^2(1+m)^2/2\sigma^2 = S_{\max}/N = S(1+m)^2/N \quad (6a)$$

and

$$A^2(1-m)^2/2\sigma^2 = S_{\min}/N = S(1-m)^2/N. \quad (6b)$$

The dc output for the extremes of carrier output is then

$$\bar{v}_{\max} = \sqrt{2\pi}\,(\sigma/2)[0.4(S/N)(1+m)^2 + 1] \quad (7a)$$
$$\bar{v}_{\min} = \sqrt{2\pi}\,(\sigma/2)[0.4(S/N)(1-m)^2 + 1]. \quad (7b)$$

If we now define

$$m_{\text{out}} = (\bar{v}_{\max} - \bar{v}_{\min})/(\bar{v}_{\max} + \bar{v}_{\min}) \quad (8)$$

consistent with the concept of modulation index we find upon substituting (7) into (8)

$$m_{\text{out}} = 2m(S/N)/[(S/N)(1+m^2) + 2.5]. \quad (9)$$

A plot of this final result is shown in Fig. 2 for SNR up to 3. Because of the straight-line approximation made, a slight error (less than 10 per cent) may be encountered at large modulation indices near 1. Using a different approach Ragazzini[4] has demonstrated the demodulation effect.

If we present the output of the envelope detector on an oscilloscope and slowly increase the noise level into the detector, we observe that the demodulated output begins indeed to shrink in size as the noise becomes comparable to the signal. The output SNR, therefore, deteriorates at low input SNR for two reasons: the output noise increases, and the output signal decreases. Over 3-db improvement could be realized if carrier phase information would be available to cut off the diode during the half cycle when it is not conducting. In addition to eliminating the noise for half the carrier cycle, the demodulation effect of the above is reduced. Whether or not the addition of a phase-locked loop is justified to supply the carrier phase information will probably be contingent upon other factors.

ACKNOWLEDGMENT

Thanks are due Dr. E. Pflumm for his valuable assistance in reducing the mathematics.

ALEX GRUMET
Republic Aviation Corp.
Farmingdale, L. I., N. Y.

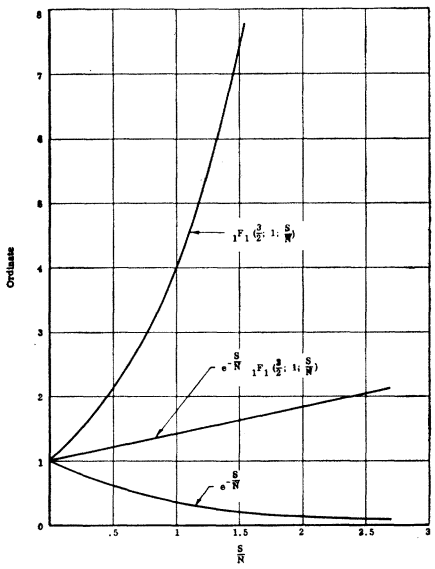


Fig. 1.

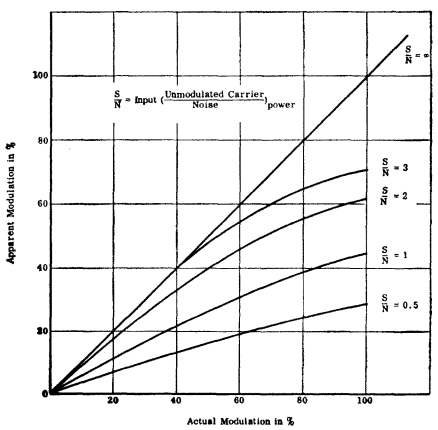


Fig. 2.

[1] W. R. Bennett, "Methods of solving noise problems," PROC. IRE, vol. 44, pp. 609–638; May, 1956.
[2] R. M. Foster and G. A. Campbell, "Fourier Integrals for Practical Applications," D. Van Nostrand Co., Inc., New York, N. Y.; 1948.
[3] E. Jahnke and F. Emde, "Tables of Functions," Dover Publications, New York, N. Y., p. 278; 1945.
[4] J. R. Ragazzini, "The effects of fluctuation voltage on the linear detector," PROC. IRE, vol. 30, pp. 277–288; June, 1942.

Cerenkov Radiation as Leaky Surface Waves*

An argument based on physical considerations [1] points out that Cerenkov radiation can be considered as the equivalent of reflection of an electromagnetic wave approaching a dielectric vacuum interface from infinity in a semi-infinite dielectric. The reflection of this incident wave results in leaky surface waves on the dielectric side of the boundary, and satisfies the Cerenkov condition, *i.e.*, the angle θ between the normal to the wave front and the boundary is $\cos\theta = (\beta\sqrt{\mu_r\epsilon_r})^{-1}$. In the absence of the incident wave, a beam of charged particles moving along the interface on the vacuum side supplies the equivalent fields which, under proper conditions, correspond to the leaky surface waves of the structure. That these waves are indeed the natural modes of the system can be shown as follows:

Consider a semi-infinite lossless dielectric μ, ϵ, with a cylindrical hole of radius r_0. In the hole $\mu = \mu_0$, $\epsilon = \epsilon_0$. Assuming propagation of the form $\exp[i(\omega t - kz)]$, simultaneous solution of the two Maxwell curl equations can be reduced to the Helmholtz scalar equation

$$(\nabla_t^2 + h^2)\begin{bmatrix} E_z \\ H_z \end{bmatrix} = 0 \quad (1)$$

* Received March 12, 1962. This work was supported by Rome Air Development Center under contract number AF30(602)2151.

where h^2 is the eigenvalue of ∇_t^2—the Laplacian operating on transverse coordinates.

This equation has to be satisfied in the hole (subscript 0) and in the dielectric (subscript 1). It is well known that all the field components can be obtained from the electric- and magnetic-field components in the direction of propagation [2].

On the vacuum side of the boundary

$$h_0^2 = k_0^2 - k^2 \tag{2}$$

where k = propagation constant, and $k_0^2 = \omega^2\mu_0\epsilon_0$.

In the dielectric

$$h_1^2 = k_1^2 - k^2 \tag{3}$$

with $k_1^2 = \omega^2\mu_1\epsilon_1$.

The simultaneous solution of (2) and (3) yields

$$\frac{\omega^2}{c^2} = \frac{h_1^2 - h_0^2}{\mu_r\epsilon_r - 1} \tag{4}$$

$$k^2 = \frac{h_1^2 - \mu_r\epsilon_r h_0^2}{\mu_r\epsilon_r - 1} \tag{5}$$

for the frequency and propagation constant dependence on the eigenvalues. Here $c^2\mu_0\epsilon_0 = 1$, $\mu_r\mu_0 = \mu_1$, and $\epsilon_r\epsilon_0 = \epsilon_1$. Thus far, the equations are quite general, and the only restriction is that a coordinate system be used in which the Helmholtz equation for the z-component is separable [3]. We can easily satisfy this restriction by using a circular coordinate system, which is most convenient for the geometry under consideration.

Several points are clear: 1) since the selected structure is lossless, the propagation constant cannot be complex [4]; 2) the radian frequency (ω) and the phase velocity must be real for propagation to take place, implying that $k^2 > 0$; 3) if $h_1^2 > 0$, the fields in the dielectric medium will be described by a Hankel function due to the fact that there is only one finite boundary condition at $r = r_0$, and the fact that Hankel functions have no finite zeros; 4) if $h_1^2 < 0$, the field will be described by a modified Bessel function of the second kind and will represent a trapped surface wave propagating along the interface [5], [8]. This also follows from well known solutions of (1) in a circular cylindrical coordinate system.

If $h_1^2 > 0$, a leaky surface wave exists in the dielectric medium with propagation of the form $\exp[i(\omega t - h_1 r - kz)]$ when the asymptotic representation for Hankel function is used. The phase velocity along the normal to the wave front is easily calculated [2]:

$$v_{ph} = \frac{\omega}{\sqrt{h_1^2 + k^2}} \tag{6}$$

at an angle to the z-axis

$$\theta = \cos^{-1}\left(\frac{k}{\sqrt{h_1^2 + k^2}}\right). \tag{7}$$

The substitution of (3) into (7) yields

$$\cos\theta = \frac{k}{k_1} = \frac{1}{\left(\frac{\omega}{k}\right)\frac{\sqrt{\mu_r\epsilon_r}}{c}} = \frac{1}{\beta\sqrt{\mu_r\epsilon_r}} \tag{8}$$

where $c\beta = (\omega/k) = v_z$. But (8) exactly represents the Cerenkov condition [6], requiring the wave front velocity along the interface (in the z direction) to be equal to the velocity of the particle moving along the interface; and requiring that this velocity, v_z, be greater than the phase velocity of the medium. Comparison of (6) and (8) shows this to be the case when $h_1^2 > 0$.

The limiting case occurs when $h_1^2 \to 0$, where transition from leaky to trapped surface waves takes place [5, 8]. This condition corresponds to incidence at the critical angle ϕ_c [7] at the interface in the limit as $\beta \to 1$. Since the angle of incidence ϕ is related to the Cerenkov angle by $\frac{1}{2}\pi$, *i.e.*, $\theta = \frac{1}{2}\pi - \phi$, then $\cos\theta = \sin\phi$, and in the limit as $\beta \to 1$, $\sin\phi \to \sin\phi_c = (\mu_r\epsilon_r)^{-1/2}$.

Thus, making use of the formalism usually invoked in solution of problems of propagation of electromagnetic surface waves, the results obtained corroborate the viewpoint presented by Brumbaugh [1]; namely, that Cerenkov radiation can be viewed as a result of the continuous reflection of incident waves from a vacuum-dielectric boundary, with the necessary excitation fields supplied by the motion of charged particles on the vacuum side of the boundary.

L. W. Zelby
Defense Electronics Products
Radio Corporation of America
Camden, N. J.

References

[1] J. M. Brumbaugh, "A simplified approach to Cerenkov radiation," to be published.
[2] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., 1st ed.; 1941.
[3] D. E. Spencer, "Separation of variables in electromagnetic theory," *J. Appl. Phys.*, vol. 22, pp. 386–389; April, 1951.
[4] R. B. Adler, "Propagation on inhomogeneous cylindrical structures," Proc. IRE, vol. 40, pp. 339–348; March, 1952.
[5] L. W. Zelby, "Propagation modes on dielectric coated wires," to be published.
[6] See, for instance, J. V. Jelly, "Cerenkov Radiation and Its Applications," Pergamon Press, Inc., New York, N. Y.; 1958.
[7] L. Brekhovskikh, "Waves in Layered Media," McGraw-Hill Book Co., Inc., New York, N. Y.; 1961.
[8] F. J. Zucker, "The guiding and radiation of surface waves," *Proc. of the Symp. on Modern Advances in Microwave Techniques*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.; November, 1954.

Noise Figure for Negative Source Resistance*

When parametric amplifiers and Esaki diode amplifiers are used as the first stage in a cascade of twoport amplifiers, it sometimes happens that the stage following such an amplifier "looks" into a source impedance (the output impedance of the preceding stage) with a negative real part. For brevity, we shall call a twoport that exhibits an output impedance with a negative real part "negative resistance twoport."

The measurement or computation of the over-all noise figure of the cascade does not encounter conceptual difficulties because the source (usually antenna) impedance connected to the input of a receiving system has a positive real part. However, often one is interested in determining experimentally the noise figure of the first stage from measurements on the cascade, because all pertinent noise parameters of the first stage may be determined from the value of noise figure for various source impedances.[1] In the evaluation of such measurements one cannot have recourse to the cascading formula using the noise figure definition of the IRE. Indeed, the source impedance connected to the second stage has a negative real part and the IRE definition of the noise figure does not apply to this case. Yet it is convenient to use a cascading formula because it leads to the determination of the noise figure of the first stage in a systematic way. This is possible if the noise figure definition based on the exchangeable power concept is used that is consistent with the cascading formula.[2] One has for the over-all noise figure of two stages in cascade

$$F_e = F_{e1} + \frac{F_{e2} - 1}{G_{e1}}. \tag{1}$$

Here the F_e's are the noise figures of the first and second stages defined on the basis of exchangeable power,[2] and G_{e1} is the exchangeable gain of the first stage. In order to find F_{e1} from a measurement of F_e one must know G_{e1} and $(F_{e2} - 1)$ for the value of source impedance that is driving the second stage in the cascade. If this impedance has a negative real part, one must overcome the practical problem of determining $(F_{e2} - 1)$ experimentally for a negative source resistance, a rather difficult and inconvenient measurement to perform directly.

The purpose of this communication is to show how one may determine the second term in (1) from signal measurements on the first stage, and from noise measurements on the second stage by using only source admittances with a positive real part. For this purpose, we rewrite the second term in a form containing terms that are more conveniently measured.

$$\frac{F_{e2} - 1}{G_{e1}} = \frac{(F_{e2} - 1)G_{1\,\text{out}}}{G_{e1}G_{1\,\text{out}}}. \tag{2}$$

$G_{1\,\text{out}}$ is the output conductance of the first stage which acts as the source conductance of the second stage. It is well-known[1] that the conventional IRE definition of the noise figure of a twoport can be expressed as a function of the source admittance $Y_s = G_s + jB_s$ for $G_s > 0$ as follows:

$$F = F_0 + \frac{R_n}{G_s}[(G_s - G_0)^2 + (B_s - B_0)^2]. \tag{3}$$

The quantities F_0, R_n, G_0 and B_0 are the four noise parameters that describe the noise of the twoport (now the second stage). The extended definition[2] of the noise figure F_e preserves the analytic form of F in terms of G_s

* Received March 29, 1962; revised manuscript received, April 9, 1962. This work was supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research.

[1] A. G. Th. Becking, H. Groendijk, and K. S. Knol, "The noise factor of 4 terminal networks," *Philips Res. Rept.*, vol. 10, pp. 349–357· Octobe-, 1955.
IRE Subcommittee 7.9 on Noise, "Representation of noise in linear twoports," Proc. IRE, vol. 48, pp. 69–74; January, 1960.
[2] H. A. Haus and R. B. Adler, "An extension of the noise figure definition," Proc. IRE (*Correspondence*), vol. 45, p. 690; May, 1957.

and B_s for negative values of G_s. If one plots

$$H = (F_e - 1)G_s = (F_0 - 1)G_s + R_n[(G_s - G_0)^2 + (B_s - B_0)^2] \quad (4)$$

as a function of G_s, one finds it to be a parabola with a minimum

$$H_{\min} = G_0(F_0 - 1) - \frac{(F_0 - 1)^2}{4R_n} + R_n(B_s - B_0)^2 \quad (5)$$

at

$$G_s - G_0 = -\frac{F_0 - 1}{2R_n}. \quad (6)$$

It can be shown that $H_{\min} > 0$ in all cases. A typical plot of H vs G_s has the appearance of Fig. 1. (It should be noted, however, that $H_{\min}$ may occur for negative values of G_s.) A measurement of H for positive G_s may be extrapolated to negative values of G_s. One simple way of doing this is to plot

$$(H - H_{\min}) \text{ vs } x = \left(G_s - G_0 + \frac{F_0 - 1}{R_n}\right)^2. \quad (7)$$

This plot is a straight line, a fact that can serve for a check of the experimental accuracy of the measurements with positive values of G_s, as well as a simple means of extrapolating H to negative values of G_s. The measurement of $G_{1\,\text{out}}$ is straightforward. The experimentally obtained value of $G_{1\,\text{out}}$ is used as the source conductance on the previously obtained plot to read off the value of $H=(F_e-1)G_s$ for the second stage.

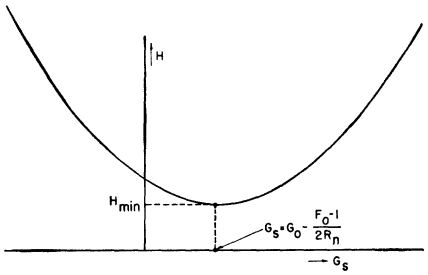


Fig. 1—Plot of H vs G_s.

The value of $G_{e1}G_{1\,\text{out}}$ may be obtained from a transducer gain G_T measurement on the first stage. If we use Norton's equivalent signal current source i_s and admittance Y_s of the source, and the corresponding signal quantities i_0, $Y_{1\,\text{out}}$ at the output port of the first stage connected to the source, we have for the transducer gain of the first stage

$$G_{T1} = \frac{\overline{i_0^2}}{i_s^2} \frac{4G_L G_s}{|Y_{1\,\text{out}} + Y_L|^2} \quad (8)$$

with $G_L=\text{Re}\,(Y_L)$, where Y_L is the admittance of the load connected to the output of the first stage. The exchangeable gain in turn is

$$G_{e1} = \frac{\overline{i_0^2}}{i_s^2} \frac{G_s}{G_{1\,\text{out}}}. \quad (9)$$

Therefore

$$G_{e1}G_{1\,\text{out}} = G_{T1} \frac{|Y_{1\,\text{out}} + Y_L|^2}{4G_L}. \quad (10)$$

Thus one may obtain $G_{e1}G_{1\,\text{out}}$ from measurements of the transducer gain of the first stage, of the load admittance Y_L and of the output admittance of the first stage. With these measurements one may determine $(F_{e2}-1)/G_{e1}$ as it appears in the cascading formula. $(F_{e1}-1)$ may thus be evaluated.

H. A. Haus
Dept. of Elec. Engrg. and
Res. Lab. of Electronics
Mass. Inst. Tech.
Cambridge, Mass.

Optimization of Pulse Transmission*

Gerst and Diamond[1] have considered the elimination of intersymbol interference in time-invariant linear systems by properly specifying the input waveform. In the transmission of digital data it is also of interest to design the signaling waveform to insure maximum energy at the output per unit of input energy. This communication is concerned with the design of signaling waveforms for linear systems which will not only insure a time-limited output signal but will also insure maximum energy transfer. If the system transfer function in Fig. 1 is

$$F(s) = \frac{\sum_{i=0}^{m} k_i s^i}{\sum_{i=0}^{n} h_i s^i}, \quad \text{with } m \le n \text{ and } h_n = 1 \quad (1)$$

then we let

$$e_0(t) = k_m \frac{d^m}{dt^m} f(t) + k_{m-1} \frac{d^{m-1}}{dt^{m-1}} f(t) + \cdots + k_0 f(t) \quad (2)$$

and

$$e_{\text{in}}(t) = \frac{d^n}{dt^n} f(t) + h_{n-1} \frac{d^{n-1}}{dt^{n-1}} f(t) + \cdots + h_0 f(t). \quad (3)$$

Fig. 1.

Without being too restrictive we consider the class of $n+1$ times differentiable output signal shapes to which the calculus of variations is easily applied. We then minimize

$$\int_0^T [e_0^2(t) - e_{\text{in}}^2(t)]dt = \int_0^T g(t)dt \quad (4)$$

subject to the constraint

$$\int_0^T e_{\text{in}}^2(t)dt = E_{\text{in}} \quad (5)$$

and require $f(0)=f(T)=0$. The solution may be obtained by solving the Euler-Lagrange equation.

For the simple case of a low-pass RC network the above steps are easily applied. Here

$$e_0(t) = \frac{1}{RC} f(t) \quad (6)$$

$$e_{\text{in}}(t) = \frac{d}{dt} f(t) + \frac{1}{RC} f(t) \quad (7)$$

and the Euler-Lagrange equation becomes

$$\frac{\partial g}{\partial f} - \frac{d}{dt}\left(\frac{\partial g}{\partial f'}\right) = 0 \quad (8)$$

from which we find

$$f(t) = \sqrt{\frac{2E/T}{1 + \frac{\pi^2 R^2 C^2}{T^2}}} \sin \frac{\pi t}{T}. \quad (9)$$

Hence,

$$e_{\text{in}}(t) = \frac{1}{RC}\sqrt{\frac{2E}{T}} \sin\left(\frac{\pi}{T}t + \tan^{-1}\frac{1}{RC}\frac{T}{\pi}\right) \quad 0 \le t \le T \quad (10)$$

and

$$e_0(t) = \frac{1}{RC}\sqrt{\frac{2E/T}{1 + \frac{\pi^2 R^2 C^2}{T^2}}} \sin \frac{\pi t}{T} \quad 0 \le t \le T. \quad (11)$$

Note that the signaling waveform is a gated sinusoid whose phase depends upon the baud length and the RC time constant.

It is interesting to examine the cost of eliminating intersymbol interference. Illustrated in Fig. 2 is the energy transfer for a rectangular signaling pulse of duration T, and the pulse given by (10). For example, when $T=RC$, the additional cost in energy is over 6 db; however, the time required to receive 90 per cent of the rectangular transmitted pulse is 3 times the duration of the optimum received pulse given by (11).

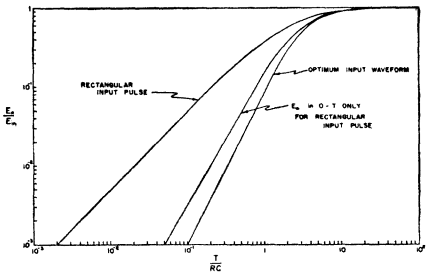


Fig. 2—Energy transfer ratio for the RC filter.

J. C. Hancock
H. Schwarzlander
R. E. Totty
Communication Sciences Lab.
School of Elec. Engrg.
Purdue University
Lafayette, Ind.

* Received March 26, 1962.

[1] I. Gerst and J. Diamond, "The elimination of intersymbol interference by input signal shaping," Proc. IRE, vol. 49, pp. 1195–1203; July, 1961.

Frequency Folding with Up-Converters*

The purpose of this communication is to describe the results of an experiment in frequency folding in a noninverting parametric up-converter.

Various papers[1,2] have described the sensitivity of nonlinear resistance mixers driven by distributed local oscillators where the RF bandwidth is considerably greater than the IF bandwidth. Such devices are effective in folding a wide RF band into a narrow IF band; however, the conversion efficiency of a mixer with distributed local oscillators decreases approximately in the same ratio as the number of local oscillators added and the sensitivity of such systems is very poor. The degradation in conversion efficiency can be attributed to a decrease in the Fourier series coefficients of the dynamic conductance of the mixer crystal,[1] and to the generation of parasitic frequencies which cannot be reactively terminated and thus absorb power from the signal.[2]

Detrimental effects of this nature using nonlinear resistive mixers lead to speculation as to whether a distributively pumped nonlinear capacitance (*i.e.*, varactor) would yield similar results. Analysis of a double-pumped varactor[3] has shown that the Fourier series coefficients of the dynamic capacitance may be actually increased through double pumping. When a double-pumped varactor is used in a noninverting up-converter configuration where the input bandwidth is greater than the output bandwidth, parasitic frequencies are also generated; however, in this type of device the parasitics which cannot be reactively terminated absorb power from the pumps rather than the signal.

The small signal analysis of a double-pumped noninverting up-converter having frequency relations as shown in Fig. 1 has been performed[3] resulting in the Manley-Rowe energy relation as given in (1).

$$P_{f_5} = -\left(P_{f_1}\frac{f_5}{f_1} + P_{f_2}\frac{f_5}{f_2}\right). \qquad (1)$$

Eq. (1) shows that the double-pumped up-converter is essentially a device which converts energy from a bandwidth 2β into a bandwidth β with a conversion gain f_5/f_1 for signals at f_1 and f_5/f_2 for signals at f_2. The same generalization would hold true if the device were operated as a double-pumped down-converter except that conversion loss would result rather than conversion gain.

In order to verify the conclusions of the small signal analysis, an S- to X-band up-converter was implemented using a Microwave Associates MA 4298 varactor. The up-converter frequency relations selected are shown in Fig. 2. The up-converter had a calculated gain of 2.8 db based on an effective $f_{co}=70$ kMc and a filling ratio of 1/4. The measured gain was 2.5 db. The frequency relations are slightly different from the case analyzed and shown in Fig. 1 in that two outputs are obtained (f_5 and f_5') rather than folding the two input signals (f_1 and f_2) into one single output (f_5). The output spacing was deliberate such that the single- and double-pumped outputs might readily be displayed and compared on a spectrum analyzer.

Fig. 3 illustrates the results obtained in the single- and double-pumped cases. In Fig. 3(a), the device was operated as a standard single-pumped up-converter with the input $f_2=3060$ Mc and the output $f_5'=10,200$ Mc. In Fig. 3(b), the device was again single-pumped with the input $f_1=2840$ Mc and the output $f_5=10,195$ Mc. Fig. 3(c) shows the results of applying both signals and both pumps simultaneously and some slight degradation in output at $f_5=10,195$ Mc can be observed. The slight decrease was found to be due to improper termination of the parasitic frequency at $f_1+f_3=9980$ Mc where an output existed which was approximately 13 db below f_5.

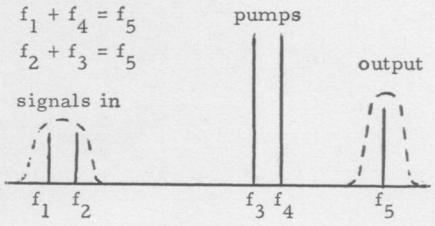


Fig. 1—Noninverting up-converter frequency relations.

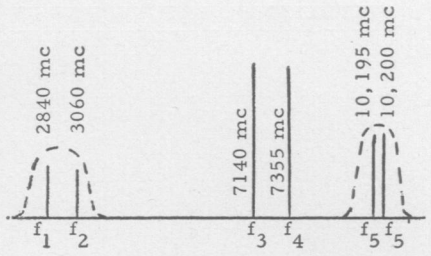


Fig. 2—Experimental up-converter frequency relations.

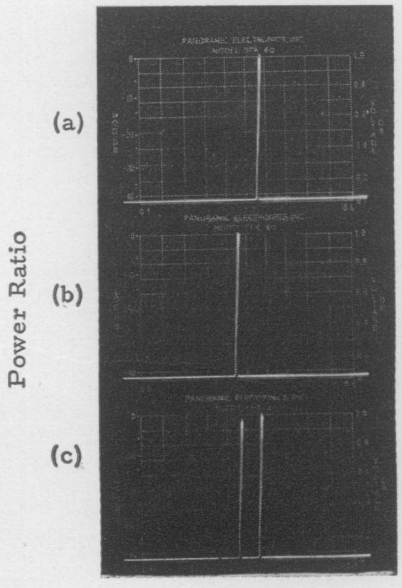


Fig. 3—Up-converter output. (a) and (b) Single-pumped. (c) Double-pumped.

Another parasitic output frequency was found to exist at $f_2+f_4=10,415$ Mc which was about 18 db below f_5' and apparently not large enough to produce any visible degradation in that output.

The results of the experiment verify the conclusions of the small signal analysis and (1). It would thus appear that distributive pumping of noninverting up-converters can, with proper design, provide a means of folding a wide signal bandwidth into a narrow output bandwidth without suffering any degradation in conversion efficiency.

R. F. Brewster
C. G. Spacek
Defense Res. Labs.
General Motors Corp.
Santa Barbara, Calif.

* Received March 21, 1962; revised manuscript received, March 23, 1962.

[1] J. L. Grigsby, "An Investigation of the Sensitivity of Some Receivers Having Large R-F Bandwidth," Stanford Electronics Labs., Stanford University, Calif., ASTIA Rept. No. AD 206201; 1958.

[2] C. G. Spacek and R. F. Brewster, "Low Noise Superheterodyne Radiometer Investigation," Research Source, AC Spark Plug Div., General Motors Corp., Milwaukee, Wis., ASTIA Rept. No. AD 247344; 1960.

[3] C. G. Spacek and R. F. Brewster, "Investigation and Development of Multiple-Pumped Parametric Amplifiers," Defense Res. Labs., General Motors Corp., Santa Barbara, Calif., Internal Rept. ER-61-110; 1961.

Antenna Gain-Loss Estimates*

The "gain-loss" of a large antenna is the quotient of two ratios: the ratio of the power received by a large aperture antenna to that received by a reference antenna of small aperture on the same transhorizon path, divided by the ratio of the corresponding powers in free space, with a single plane wave incident normally on the apertures.

Yeh recently published a collection of experimental measurements of gain-loss [1], for guidance of tropospheric system designers. Unfortunately, his curve does not include two important experimental data, and two important data seem to be plotted incorrectly. When the latter are corrected and additional available data at 9000 Mc at 46 miles and 400 Mc at 600 miles are included, no curve such as Yeh's can claim to approximate the experimental facts.

The incorrectly plotted points are those taken from Crawford [2] (4110 Mc, 171 miles) and from Geiger [3] (9640 Mc, 173 miles). Crawford [4] reports the ratio of the yearly median levels of received power to be 5.7 db on the 60-foot and 8-foot antennas. The reported ratio of plane-wave gains for these two antennas is 16.5 db, so the apparent gain-loss for one 60-foot antenna, compared to an 8-foot antenna, becomes $16.5-5.7=10.8$ db, and not 5.7 db as plotted by Yeh [1]. Geiger's measurements [3] show the gain-loss for *one* 8-foot antenna, compared to a horn, to be 4.4 db rather than 4.5 db for *two* antennas as reported by Yeh. Geiger [6] estimates the total gain-loss to be 8 db when two 8-foot paraboloids replace horns.

Since the gain-loss is measured with respect to a reference antenna with varying

* Received March 26, 1962.

TABLE I
EXPERIMENTAL ANTENNA GAIN-LOSS DATA

Reference Author	Frequency (Mc)	Distance (Miles)	Antenna Diameter (Feet)	θ/α	Gain-Loss	
					One Antenna	Two Antennas
Trolese [7]	9375	46.3	4	0.288	5 ± 0.5	10 ± 1
M.I.T. Lincoln Lab.	412.8	618	60	2.5	0	0
M.I.T. Lincoln Lab.	399	639	120	3.76	0.5	1.0
Bullington [9]	4090	173	28	2.4	3.5	7.0
Chisholm *et al.* [10]	3670	188	28	2.7	3.5	7.0
Chisholm, *et al.*	3670	188	28	2.7	7.9	15.8

directivity from one experiment to another, measured losses should be referred to comparable reference antennas before being compared. In plotting the measured data of Chisholm [8], 5-foot dishes were assumed herein for reference antennas because their directivity is roughly comparable to that of the other data plotted.

For present purposes, where gain-loss measurements for large aperture antennas at both ends of the path are not available, the loss measured for one narrow-beam antenna will be doubled to obtain the loss for two such antennas. This is by no means always true of course, but it at least approximates the truth in Trolese's experiments.

Additional available data on gain-loss is listed in Table I [7]–[10]. All data are plotted in Fig. 1 as a function of θ/α (scatter angle/antenna beamwidth). It is evident that a well-fitting empirical curve cannot be drawn through these data. The lack of fit is even more serious for the many diverse theoretical curves, especially when *all* of the data are presented. The disagreement with Trolese's 9375-Mc, 46-mile data was noted earlier both by Waterman [11] and by Armand and Vvedenskii [12]. The lack of significant gain-loss on the newer Lincoln Lab., 400-Mc, 600-mile observations serves to emphasize this disagreement. These considerations lead one to suspect that the observed antenna gain-loss on long tropospheric paths beyond the horizon is not a simple function of θ/α.

Armand and Vvedenskii [12] have recently suggested that gain-loss may be proportional to the ratio of the angular spread θ_0 occupied by the ensemble of received waves caused by the mechanism. This essentially revives the suggestion originally offered in 1951 by Schott [13] in the first paper on this phenomenon. Armand and Vvedenskii further hypothesize that this angular spread θ_0 may be inversely proportional to the path length d, and thus that the gain-loss measurements may be expected to be a function of fD/d since the numerator is proportional to the beam angle (frequency f in Gc, antenna diameter D in feet, and distance d in hundreds of miles). Fig. 2 shows the measurements plotted against fD/d. A single line seems to go well through the points in contrast to Fig. 1. We do not extend the line to the lowest point on the right because this yearly median value certainly includes losses due to local refraction variations greater than the 0.3° beamwidth used. These variations are not properly classified as gain-loss. Also for this point, the rough assumption of doubling the measured db loss for one antenna is especially an exaggeration of what would be measured with large antennas at both ends of the path. The lower of the two points at abscissa 54 is partly a sea reflection effect at a low-sited terminal. Despite the well-known time variability of the gain-loss effect, the measured median losses appear to be a monotonic function of fD/d.

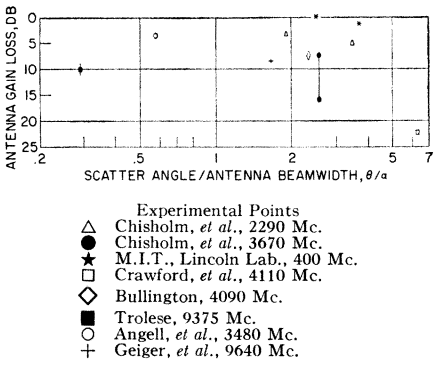


Fig. 1—Antenna gain-loss for two antennas as a function of θ/α.

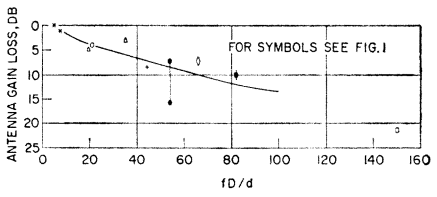


Fig. 2—Empirical curve for estimating antenna gain-loss for two antennas.

One implication of this plot is particularly interesting. It suggests that large aperture antennas may more nearly realize their plane-wave gains over long distances, thus making the use of SHF more feasible for long distance tropospheric circuits than heretofore assumed. There is some limited experimental justification for this optimism: Trolese [7] at 9375 Mc, over 46 miles, measured a gain degradation of 10 ± 1 db for two 4-foot dishes, while Geiger [3], at 9640 Mc, over 173 miles, measured 4.4 db for one 8-foot antenna, or 8.8 db for two 8-foot antennas, showing a definite decrease in gain-loss with distance. Also, Potter [14] has reported some NEL measurements at 9375 Mc, which in two cases show a decrease of gain-loss with distance.

Until more is known, both experimentally and theoretically, about the phenomenon, Fig. 2 seems to offer a safer empirical estimate for system designers than the usual curves plotted as a function of θ/α.

J. L. LEVATICH
Radio Div.
The Bendix Corporation
Towson, Md.

REFERENCES

[1] L. P. Yeh, "Experimental aperture medium coupling loss," PROC. IRE (*Correspondence*), vol. 50, p. 205; February, 1962.
[2] A. B. Crawford, D. C. Hogg, and W. H. Kummer, "Studies in tropospheric propagation beyond the horizon," *Bell Sys. Tech. J.*, vol. 38, pp. 1067–1178; September, 1959.
[3] G. V. Geiger, N. D. LaFrenais, and W. J. Lucas, "Propagation measurements at 3480 and 9640 Mc/s beyond the radio horizon," *Proc. IEE*, pt. B, vol. 107, pp. 531–546; November, 1960.
[4] Crawford, *op. cit.*, see p. 1080.
[5] JTAC Report, "Radio transmission by ionospheric scatter," part II: "Long-range tropospheric transmission," PROC. IRE, vol. 48, pp. 30–44; January, 1960.
[6] Geiger, *op. cit.*, p. 546.
[7] L. G. Trolese, "Characteristics of tropospheric scattered fields," PROC. IRE, vol. 43, pp. 1300–1305; October, 1955.
[8] J. H. Chisholm, W. E. Morrow, B. E. Nichols, J. F. Roche, and A. E. Teachman, "Properties of 400 MCPS Long Distance Tropospheric Circuits," (To be published); and B. E. Nichols, "The AN/FRC-47 (XD-1) mock-up system performance results," M.I.T. Lincoln Lab. Tech. Rept. No. 203, AD 216 370; April 23, 1959.
[9] K. Bullington, "Characteristics of beyond-the-horizon radio transmission," PROC. IRE, vol. 43, pp. 1175–1180; October, 1955; and personal communication.
[10] J. H. Chisholm, P. A. Portmann, J. T. de Bettencourt, and J. F. Roche, "Investigations of angular scattering and multipath properties of tropospheric propagation of short radio waves beyond the horizon," PROC. IRE, vol. 43, pp. 1317–1335; October, 1955.
[11] A. T. Waterman, Jr., "Some generalized scattering relationships in transhorizon propagation," PROC. IRE, vol. 46, pp. 1842–1848; November, 1958.
[12] N. A. Armand and B. A. Vvedenskii, "On the loss of gain mechanism with narrow-beam antennas in long distance ultra-short-wave tropospheric propagation," *Radiotekh. Elektron.*, vol. 4, no. 10, pp. 1594–1601; 1959. (English translation, pp. 33–47.)
[13] F. W. Schott, "On the response of a directive antenna to incoherent radiation," PROC. IRE, vol. 39, pp. 677–680; June, 1951.
[14] C. A. Potter, "Tropospheric scattering of microwaves," *Proc. Decennial Symp. of ONR*, pp. 469–479; March, 1957.

The Assistor, a Component with Bipolar Negative Resistance*

I had lately resumed—after 36 years—a study of solid-state negative resistors. I now have evolved a new and promising basic component. In its simplest form it consists of a few grains of powdered semiconductor pressed between metal electrodes.

The voltage vs current characteristic of the new device exhibits a smoothly curved region of negative resistance for either polarity, and for rising and falling test current. Fig. 1 shows a typical plot as traced with the conventional oscilloscope test circuit of Fig. 2. For very low currents the resistance is linear and large, of the order of +50 kohm; the curve turns over, *e.g.*, at ±3 v and 0.1 ma, to drop with a slope of, *e.g.*, initially −5 kohm and more to level out at, *e.g.*, ±2 volt. The positive and negative branches are, as a rule, strictly symmetrical and identical for forward and return trace. The lowest leveling voltage observed is ±2 v with peaks then at ±3 v; with higher final levels the peaks are usually much more than proportionately higher. The characteristic does not vary from dc to—in some cases—the megacycles. One can plot the familiar

* Received March 19, 1962.

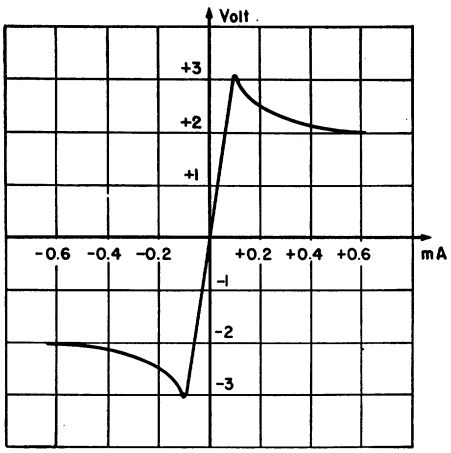


Fig. 1—Typical voltage vs current characteristic.

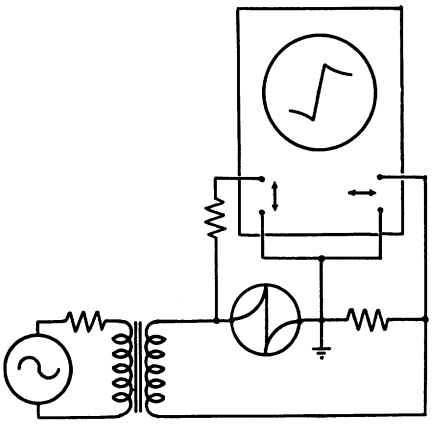


Fig. 2—Circuit for curve tracing with oscilloscope.

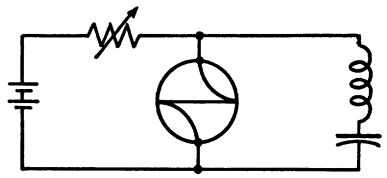


Fig. 3—Oscillator circuit.

square loop by varying a dc source voltage applied via a fixed series resistor.

Given a chance—by shunting with a series-tuned circuit—the device will readily oscillate as in the circuit of Fig. 3 from, *e.g.*, 65 cps to over a megacycle. In some (dubious) conditions the device contributes apparent inductance so that the frequency of oscillation is lower than that expected from the tuned circuit. The device differs from any now known. It has a voltage rather than the current peak of the tunnel diode and it is bipolar, but unlike the carbon arc has negative resistance for decreasing as well as for rising current.

The metal electrodes may be slightly domed flat, or simply two round wires crossing at a right angle, such as steel wires of 0.25 mm diameter, with a few grains of powder sprinkled between them; the required pressure is then of the order of a few grams which in view of the small area amounts to at least a thousand pounds per square inch. As semiconductors I have tested a few metal oxides, none failed, powdered to about 0.1 mm grain size. I use—mainly from habit—zinc oxide, ZnO. Even table salt, NaCl, finely powdered, will work at least in the kilocycle region. In its present shape the device is easily disturbed; and it will seldom dissipate more than about 2–5 mw without early harm.

I feel that, rather than the double negation "negative resistor," the device merits a new name, such as "assistor" and a new circuit symbol, such as that used in Figs. 2 and 3.

H. E. Kallmann
417 Riverside Drive
New York, N. Y.

Microwave Phonon Instabilities Associated with Ferromagnetic Resonance*

Parametric coupling between the uniform precession and the low-frequency resonant vibrational modes of a ferrimagnetic insulator has been reported by Spencer and LeCraw[1] and more recently by Comstock, Auld, and Wade.[2] In this note conditions are given under which microwave phonons, propagating in the cubic lattice of a ferrimagnetic insulator, can be parametrically excited by the uniform precession undergoing ferromagnetic resonance at frequency ω_0.

Let $\varrho = i\xi + j\eta + k\zeta$ be the displacement vector in a deformable cubic crystal where the (100) directions correspond to $x_i(x, y, z$ for $i = 1, 2, 3)$. The strain tensor components are $\epsilon_{ij} = \partial\rho_i/\partial x_j$ and the elastic energy U (including magnetoelastic energy) is of the form

$$\begin{aligned} U = {} & \tfrac{1}{2}\gamma(\epsilon_{11}^2 + \epsilon_{22}^2 + \epsilon_{33}^2) \\ & + \lambda(\epsilon_{11}\epsilon_{22} + \epsilon_{11}\epsilon_{33} + \epsilon_{22}\epsilon_{33}) \\ & + \tfrac{1}{2}\mu[(\epsilon_{12} + \epsilon_{21})^2 + (\epsilon_{13} + \epsilon_{31})^2 \\ & + (\epsilon_{23} + \epsilon_{32})^2] + a(M_1^2\epsilon_{11} + M_2^2\epsilon_{22} \\ & + M_3^2\epsilon_{33}) + b[M_1M_2(\epsilon_{12} + \epsilon_{21}) \\ & + M_1M_3(\epsilon_{13} + \epsilon_{31}) + M_2M_3(\epsilon_{23} + \epsilon_{32})] \\ & + \tfrac{1}{2}c(M_1^2\epsilon_{11}^2 + M_2^2\epsilon_{22}^2 + M_3^2\epsilon_{33}^2) \\ & + d(M_1^2\epsilon_{22}\epsilon_{33} + M_2^2\epsilon_{11}\epsilon_{33} + M_3^2\epsilon_{11}\epsilon_{22}) \\ & + \tfrac{1}{2}e[M_1^2(\epsilon_{23} + \epsilon_{32})^2 + M_2^2(\epsilon_{13} + \epsilon_{31})^2 \\ & + M_3^2(\epsilon_{12} + \epsilon_{21})^2] + \cdots . \end{aligned}$$

The differential equations of motion governing ϱ are given by

$$D\ddot{\rho}_i + \alpha\dot{\rho}_i - \frac{\partial}{\partial x_j}\left(\frac{\partial U}{\partial \epsilon_{ij}}\right) = 0 \quad (i = 1, 2, 3) \quad (1)$$

where a repeated index (in this case j) indicates summation over that index. D is the mass density of the crystal and α is a damping parameter for the elastic waves.

Assume that the crystal is magnetized along some z' axis and is undergoing ferromagnetic resonance. Then

$$\boldsymbol{M} = M\sin\theta(i'\cos\omega_0 t + j'\sin\omega_0 t) + k'M\cos\theta$$

where $\bar{\imath}$, $\bar{\jmath}$ and $\bar{k}$ are related to $\bar{\imath}'$, $\bar{\jmath}'$ and $\bar{k}'$ by a pure rotation transformation so that $M_i = l_{ij}M_j'$.

Longitudinal Waves

Consider a longitudinal wave propagating in the x direction. In this case $\eta = \zeta = 0$ and $\xi = \xi_0(t)e^{-ikx}$. Eq. (1) becomes

$$D\ddot{\xi}_0 + \alpha\dot{\xi}_0 + k^2(\gamma + cM_1^2)\xi_0 = 0. \quad (2)$$

If $l_{ij} = \delta_{ij}$ (the Kronecker delta),

$$M_1^2 = M^2\sin^2\theta\tfrac{1}{2}(1 + \cos 2\omega_0 t).$$

Since (2) is a Mathieu equation there will be a second-order instability, provided that

$$k \simeq \frac{\omega_0}{\sqrt{\gamma/D}} \quad \text{and} \quad \sin\theta \geq 2\sqrt{\frac{(\alpha/D)\gamma}{\omega_0 cM^2}}.$$

The threshold is naturally the same for longitudinal waves propagating along y; there is no coupling to longitudinal waves propagating along z.

If

$$\begin{aligned} & l_{ij} \neq \delta_{ij}, \\ & M_1^2 \simeq l_{13}^2M^2 + 2l_{13}\sqrt{1 - l_{13}^2}\,M^2 \\ & \cdot\sin\theta\cos(\omega_0 t + \phi_0) \end{aligned}$$

(neglecting terms in $\sin^2\theta$ etc.). A first-order instability can occur if

$$k = \frac{\omega_0}{2}\Big/\sqrt{\frac{\gamma + l_{13}^2cM^2}{D}}$$

and

$$\sin\theta \geq \frac{2(\alpha/D)}{\omega_0}\,\frac{\gamma + l_{13}^2cM^2}{cM^2l_{13}\sqrt{1 - l_{13}^2}}.$$

If $cM^2 \ll \gamma$, θ will be minimized when $l_{13} = 1/\sqrt{2}$.

Transverse Waves

Consider a transverse wave linearly polarized along x and propagating in the z direction. In this case $\eta = \zeta = 0$ and

$$\xi = \xi_0(t)e^{-ikz}.$$

Eq. (1) becomes

$$D\ddot{\xi}_0 + \alpha\dot{\xi}_0 + k^2(\mu + eM_2^2)\xi_0 = 0. \quad (3)$$

If $l_{ij} = \delta_{ij}$, the instability occurs when

$$k \simeq \frac{\omega_0}{\sqrt{\mu/D}} \quad \text{and} \quad \sin\theta \geq 2\sqrt{\frac{(\alpha/D)\mu}{\omega_0 eM^2}}.$$

The threshold is the same for transverse waves polarized along y propagating in the z direction; however, there is a 90° phase shift between the $\xi(z)$ and $\eta(z)$ components (circularly polarized if amplitudes are equal). Transverse waves $\xi(y)$ or $\eta(x)$ are not coupled to the uniform precession.

* Received March 27, 1962; revised manuscript received April 10, 1962. This research was sponsored by the Ford Foundation.

[1] E. G. Spencer and R. C. LeCraw, "Magnetoacoustic resonance in yttrium iron garnet," *J. Appl. Phys.*, vol. 30, pp. 149S–150S; April, 1959.

[2] R. L. Comstock, B. A. Auld, and G. Wade, "Parametric effects in magnetoacoustic resonance," *J. Appl. Phys.*, vol. 32, pp. 225S–226S; March, 1961.

If $l_{ij} \neq \delta_{ij}$, a first order instability occurs when

$$k = \frac{\omega_0}{2} \Big/ \sqrt{\frac{\mu + l_{23}^2 eM^2}{D}}$$

and

$$\sin\theta \geq \frac{2(\alpha/D)}{\omega_0} \frac{\mu + l_{23}^2 eM^2}{eM^2 l_{23}\sqrt{1 - l_{23}^2}}.$$

If $eM^2 \ll \mu$, θ will be minimized when $l_{23} = 1/\sqrt{2}$.

Conclusion

The instability thresholds given are for phonons propagating along principal axes of the crystal and may be considered representative of thresholds for other propagation directions. The ratio of $\mu c/\gamma e$ together with the magnetizing geometry will determine which type of phonon (longitudinal or transverse) will have the lowest threshold.[3] If $\omega_0/2$ spinwaves do not exist (this can always be arranged by making ω_0 sufficiently large) degenerate spinwave instabilities occur if

$$\theta \geq \sqrt{\frac{2\Delta H_k}{M}}$$ [4]

where $2\Delta H_k$ is the pertinent spinwave linewidth). Assuming that the first-order phonon coupling is dominant one must compare

$$\frac{4(\alpha/D)}{\omega_0}\frac{\mu}{eM^2} \text{ or } \frac{4(\alpha/D)}{\omega_0}\frac{\gamma}{cM^2} \text{ with } \sqrt{\frac{2\Delta H_k}{M}}$$

where the intrinsic Q of the phonon is

$$\left(\frac{\omega_0}{2}\right) \Big/ \left(\frac{\alpha}{D}\right).$$

It would appear that if α is independent of ω_0 the phonon coupling dominates above some characteristic frequency.

For yttrium iron garnet (YIG):

$$\sqrt{\frac{2\Delta H_k}{M}} \simeq 1/40.$$

This means that for a supposed Q of 10^5 one would expect the phonon coupling to dominate provided that

$$\frac{eM^2}{\mu} \text{ or } \frac{cM^2}{\gamma} \geq 10^{-3}.$$

Even if this condition is not satisfied (or the Q's are lower) it might be possible to increase ΔH_k through rare earth additions to the point where the inequality is satisfied. This might work if α were relatively insensitive to the rare earth impurity.

If the phonon coupling is desired and YIG turns out to be unsuitable, one might profitably investigate one of the high anisotropy materials such as barium ferrite which undoubtedly has greater "eM^2/μ" ratios. A study of phonon Q's and ΔH_k's in these materials is essential. Of course, the results derived here are for cubic symmetry and would have to be modified accordingly.

The author[5] has suggested that the thresholds of spinwaves degenerate with the uniform precession can be raised by modulating the spinwave frequencies without at the same time modulating the uniform precession. Such a technique will *not* raise the phonon thresholds and so apart from other limitations of the method there is a new one. If large-amplitude resonance is the desired goal, it appears that one wants to magnetize along a (100) direction of the crystal so as to forbid first-order phonon coupling and, if possible, to use a material with low eM^2/μ, cM^2/γ ratios. Another suggestion made by the author[6] concerns the possibility of dropping the uniform precession *below* the spinwave manifold so as to weaken spinwave coupling and allow large-amplitude resonance. In this regard it should be noted that one *always* has phonons which are degenerate with the uniform precession and they may well be the dominant limitation. This point was not considered in the previous reference.[6]

Frederic R. Morgenthaler
Dept. of Elec. Engrg.
Mass. Inst. Tech.
Cambridge, Mass.

[3] It is possible that hybrid modes may have comparable or even lower thresholds.

[4] H. Suhl, "The nonlinear behavior of ferrites at high microwave signal levels," Proc. IRE, vol. 45, pp. 1270–1284; October, 1956. See especially p. 1271.

[5] F. R. Morgenthaler, G. E. Bennett, and F. A. Olson, "Suppression of spinwave instabilities associated with ferromagnetic resonance," *Proc. 1961 Internat'l. Conf. on Magnetism and Crystallography*, Kyoto, Japan, September 25–30, to be published, *J. Phys. Soc. Japan*, vol. 17, Suppl., 1962.

[6] F. R. Morgenthaler, "On the possibility of obtaining large amplitude resonance in very thin ferrimagnetic disks," *J. Appl. Phys.*, vol. 33, Suppl., pp. 1297–1299; March, 1962.

Two Inequalities Useful in Transient Analysis*

In a recent communication[1] the following inequality, which is useful in the transient analysis of physical systems, was announced:

Let $w(t)$ be a real-valued function of the real variable t, and let it be zero for $t<0$. Furthermore let its Laplace transform $W(s)$ be a rational function denoted by

$$W(s) = \frac{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_0}{s^m + b_{m-1}s^{m-1} + \cdots + b_0}, \quad (1)$$

where $m \geq 2n$ and all the roots of the denominator have nonpositive real parts. Then for $t \geq 0$

$$|w(t)| \leq \frac{|a_n| t^{m-n-1}}{(m-n-1)!} + \frac{|a_{n-1}| t^{m-n}}{(m-n)!} + \cdots + \frac{|a_0| t^{m-1}}{(m-1)!}. \quad (2)$$

* Received March 23, 1962.

[1] A. H. Zemanian, "Further properties of certain classes of transfer functions: II," *Quart. Appl. Math.*, vol. 19, pp. 158–159; July 1961.

This result was obtained as a special case of a more general conclusion whose proof consisted of some comparatively involved arguments. The first purpose here is to present a direct and considerably simpler proof of this special case. In the following discussion an upper case letter will denote a rational function of s and the corresponding lower case letter will designate that function of t which is zero for $t<0$ and is the inverse Laplace transform of the function of s.

First let us note that every rational function of the form

$$F(s) = \frac{s}{s^2 + c_1 s + c_0}, \quad (3)$$

where c_1 and c_0 are real and the roots of the denominator have nonpositive real parts, is a positive-real function. This means that the function $f(t)$ will satisfy the following inequality for $t \geq 0$:

$$|f(t)| \leq 1. \quad (4)$$

Similarly for

$$G(s) = \frac{1}{s+d}; \quad (d \geqq 0), \quad (5)$$

the function $g(t)$ satisfies

$$|g(t)| \leq 1. \quad (6)$$

Also consider the expression

$$H(s) = \frac{1}{(s+\gamma)(s+\bar{\gamma})}; \quad (\mathrm{Re}\,\gamma \geqq 0) \quad (7)$$

where $\bar{\gamma}$ denotes the complex conjugate of γ. The corresponding function of t satisfies

$$|h(t)| \leq t. \quad (8)$$

Secondly, consider the function

$$M(s) = \frac{s^q}{s^m + b_{m-1}s^{m-1} + \cdots + b_0} \quad (9)$$

where $q \leq m/2$. It can be factored into a product of functions such that q of them have the form of (3) and if $q<m/2$, the remaining part is the reciprocal of a polynomial. This remaining part can be factored into a product of terms, each of which has the form of either (5) or (7). The function of t, corresponding to the first product of q factors of the form (3), is the convolution of q functions, each of which is zero for $t<0$ and bounded by one for $t \geq 0$. It therefore is bounded in magnitude by

$$\frac{t^{q-1}}{(q-1)!}u(t) \quad (10)$$

where $u(t)$ is the unit step function. By a similar argument we see that when $q<m/2$, the function of t, corresponding to the product of factors such as (5) and (7), is bounded in magnitude by

$$\frac{t^{m-2q-1}}{(m-2q-1)!}u(t). \quad (11)$$

Again, convolving (10) and (11), we obtain

$$|m(t)| \leq u(t)\int_0^t \frac{x^{q-1}(t-x)^{m-2q-1}}{(q-1)!(m-2q-1)!}dx = \frac{t^{m-q-1}}{(m-q-1)!}u(t). \quad (12)$$

Since $W(s)$ is a linear combination of terms of the form (9), it follows that (10) and (12) imply (2). Q.E.D.

It can be shown through examples that the conclusion (2) cannot be strengthened if nothing more is assumed known about $W(s)$.

On the other hand if we weaken our assumptions on the denominator of $W(s)$ and the degrees of the numerator and denominator of $W(s)$, a similar inequality can be obtained when we add the assumption that $w(t)$ is non-negative. More specifically, *if $w(t)$ is a real-valued non-negative function of the real variable t and is zero for $t<0$, and if its Laplace transform is given by* (1) *where $m>n$ and the $b_i(i=1, 2, \cdots, m)$ are non-negative, then for $t\geq 0$*

$$w(t) \leq \frac{a_n t^{m-n-1}}{(m-n-1)!} + \frac{a_{n-1} t^{m-n}}{(m-n)!} + \cdots + \frac{a_0 t^{m-1}}{(m-1)!} \tag{13}$$

and thus $W(s)$ has no poles in the half plane Re $s>0$.

The usefulness of this result, whose proof can be found elsewhere,[2] is enhanced by the fact that a number of criteria exist by which one can ascertain whether $w(t)$ is non-negative merely by inspecting the pole and zero locations of $W(s)$.[3,4]

A. H. Zemanian
New York University
New York, N. Y.

[2] A. H. Zemanian, "An upper bound on nonnegative transient responses," *Quart. Appl. Math.* (to be published).
[3] A. H. Zemanian, "The properties of pole and zero locations for nondecreasing step responses," *Trans. AIEE (Communications and Electronics*, no. 50), pp. 421–426; September, 1960.
[4] A. H. Zemanian, "On the pole and zero locations of rational Laplace transformations of nonnegative functions," *Proc. Am. Math. Soc.*, vol. 10, pp. 868–872; December, 1959.

Some Comments on "A New Precision Low-Level Bolometer Bridge"*

The following comments on the above article[1] are unfortunately necessary:

1) Reisener and Birx misrepresent when they claim, "we have been able to achieve an over-all detection limit of 3×10^{-10} w." While an article in the PROCEEDINGS is not a patent application, generally authorship implicitly, and here explicitly, testifies to technical, not merely literary, achievement. I had reported[2] the theory of the over-all system of the instrument about a year before Mr. Reisener became familiar with it. I had the burden for all aspects of the bridge development including the construction and testing of both breadboard and prototype models. Mr. Reisener designed many of the circuits and was invaluable in the trouble-shooting, but he quite properly refused to be cited as coinventor for the patent application. Dr. Birx's contribution was very slight. The bridge described is exclusively my invention.

2) What the article refers to as "bolometer instability" is really ambient temperature fluctuation. Ambient temperature fluctuation has greatly limited the sensitivity of all bolometer bridges. The method indicated by Fig. 1 dynamically compensates for temperature fluctuations sufficiently so that in principle substitution bolometric measurements can be made to the limits of sensitivity inherent in the bolometer used. To establish the correspondence of the principle of the bridge to its performance requires the determination of the sensitivity of the bolometer.

3) The statement, "tests showed that the predominant sources of noise were the Johnson noise of the fixed resistors and bolometers and the temperature fluctuation noise of the bolometers," cannot be sustained. To identify the nature of the noise it would be necessary to obtain the noise spectrum over the frequency band 0 cps to about $1/20t$ (t = thermal response time of the bolometer) and to measure the absolute value of the noise. To the best of my knowledge this has not yet been done.

4) There are several errors in the analytical evaluation of the bolometer sensitivity. a) The effective temperature of the bolometer is not 90.8°C as given, but 189°C. The bolometer is a fine platinum wire. Its effective temperature at nominal resistance can be calculated by dividing the difference between the nominal resistance, 100 ohms, and the "cold" resistance, about 60 ohms, by the temperature coefficient of resistance of platinum, 0.00352 according to the International Critical Tables, and the cold resistance. The formula used in the article is not applicable to large temperature changes since the heat loss characteristic β is not a constant but a function of temperature. b) Since both β and T, the temperature, vary along the length of the fine wire constituting the bolometer, and since T occurs as a squared term, the effective temperature cannot be used and the formula for total noise is not valid. c) As Van der Ziel[3] shows, the sensitivity of a bolometer and a bolometer bridge are two different things. No such difference is acknowledged in this evaluation. d) The expression for Johnson noise is too large by a factor of 4. The "maximum available power" is the most that can be drawn from a signal source.

5) If the analysis referred to above were accurate and the tests had been made, the latter would validate the former. While a ratio of the inherent sensitivity of the bolometer to the measured sensitivity of the bridge, as attempted in the article, would be a valid measure of bridge performance, it says nothing as to the correspondence of the principle of the bridge to its performance. For this it would be necessary to calculate the degradation caused by design limitations, *i.e.*, noise figure of the amplifier, band-pass of amplifier, equal arm basic bridge, etc. Time constants greater than the response time of the bolometer are, of course, not relevant to either evaluation.

6) The precise calculation of the bolometer sensitivity is very difficult, but a fair approximation could be made analytically. Furthermore there is sufficient test data to check this approximation with an accuracy comparable to that of the noise figure of the ac amplifier " . . . estimated to be 1.8 db." The law of the instrument allows the calculation of the sensitivity in accordance with the nature and the measured performance of the components so that the correspondence of principle to performance can be estimated also.

MAX HIRSCH
Universal Associates
Philadelphia, Pa.

Authors' Comments[4]

The authors are pleased to acknowledge that the conception and initial development of the low-level bridge were the contributions of Mr. Hirsch.

The Franklin Institute has made application for a patent in his name.

To reduce complexity and in the interest of saving space, only a simplified analysis was presented in the paper. Mr. Hirsch is quite correct in pointing out possible errors resulting from the use of mean values for the bolometer temperature rise and the heat loss coefficient. The Johnson noise mentioned is expressed in the paper as an equivalent input power fluctuation.

The performance data presented are reliable and the analysis, while not rigorous, does convey a proper conception of the nearness with which the performance approaches that theoretically possible. It was our purpose, not to present an exhaustive treatise, but to bring to the attention of the engineering community the capabilities of this bridge which we feel contributes significantly to the art of low-level RF measurement.

W. C. REISENER, JR.
D. L. BIRX
Labs. for Research and Development
The Franklin Institute
Philadelphia, Pa.

[4] Received April 18, 1962.

Causality*

It is most surprising and disappointing to find Deutsch's reflections on Causality[1] in a scientific journal of the year 1962. One hundred years ago, his remarks would have been timely and appropriate. At that time,

* Received March 26, 1962.
[1] W. C. Reisener and D. L. Birx, "A new precision low-level bolometer bridge," PROC. IRE, vol. 50, pp. 39–42; January, 1962.
[2] M. M. Hirsch and D. L. Birx, "Investigation of Low-Level RF Voltage and Power Measurement," Franklin Institute, Philadelphia, Pa., 7th Quart. Engrg. Rept., Signal Corps Contract DA-36-039-SC-72827; March, 1958.
[3] A. Van der Ziel, "Noise," Prentice-Hall, Inc., Englewood Cliffs, N. J., pp. 409–412; 1954.

* Received March 30, 1962.
[1] S. Deutsch, "Causality," PROC. IRE (*Correspondence*), vol. 50, p. 222; February, 1962.

the work of the great mathematicians Laplace and Lagrange had led to the belief that a super-mathematician would be able to predict the entire future of the universe if the present conditions could fully be made known to him. This situation was entirely changed by the advent of quantum mechanics. It is now established that the laws of nature can be formulated only in a statistical manner. The classical physical laws are averages that are accurate with a very high degree of probability, but they are not fundamentally correct. For example, in a radioactive substance it is *in principle* impossible to predict which atom is going to disintegrate next, and no super-computer will be able to provide this knowledge. The laws of quantum mechanics have been verified by such an overwhelming mass of experimental evidence that today there is not the slightest experimental indication that the probabilistic interpretation of these laws is incorrect.

Turning now to the human brain and the human mind, Deutsch correctly states that the nerve cells are complicated electrochemical engines. Their behavior is fairly well understood and, of course, obeys the laws of physics and chemistry. It is further known that the number of molecules involved in a single occurrence at the synapsis is so small that there is, in accordance with Heisenberg's indeterminacy relation, a significant degree of uncertainty as to their location over the period of time of the process. Therefore, our thoughts are *in principle* unpredictable, and certainly a billion years ago neither Golay's paper nor Deutsch's paper (nor any other paper) could have been predicted.

If Deutsch feels that free will depends on the absence of causality, then the existence of free will has abundantly been proved.

A. P. Speiser
IBM Res. Lab.
Adliswil-Zurich, Switzerland

Author's Comment[2]

In 1922, the living cell was seen as a spherical blob having few structural details; in 1962, the electron microscope reveals an unbelievably complex electrochemical engine, with over 1000 different kinds of molecules. A short time ago, a protein molecule was little more than a tiny dot with an impressive atomic weight and some vague chemical bonds; today, the three-dimensional structure is an awe-inspiring miracle.[3] A few years ago, matter was composed of electrons and protons; today there is a bewildering array of "fundamental" particles, with electrons that somehow jump from one quantum orbit to another, with nuclei that remain stable for a million years and then disintegrate for no apparent reason. A multitude of questions remain unanswered, and will never be answered if we accept the naïve viewpoint that present concepts, including quantum theory, give more than a dim outline of reality. Matter may originate in particles that are much smaller than an electron; the units that we regard as fundamental may be the "microorganisms" that have survived in an evolutionary sense. The uncertainty principle, if anything, begs for this interpretation.[4]

Statistical methods are necessary in describing the activities of large populations. The conclusion that individuals do not obey causality, however, is most unjustified. We may not know when a particular atom will disintegrate, but we can be certain that a definite combination of circumstances *causes* this *effect* each time, without exception. A corollary of all this is that the human brain can be duplicated, on a small scale, by means of pattern-recognition units and memory devices.[5]

Although every facet of man's future is written in the present state of the universe, the record is too complicated for us to read. We may yet destroy ourselves because of the hostility that evolution has bestowed upon our genes.

Sid Deutsch
Polytechnic Inst. of Brooklyn
Brooklyn, N. Y.

[2] Received April 19, 1962.
[3] J. C. Kendrew, "The three-dimensional structure of a protein molecule," *Sci. American*, vol. 205, pp. 96–110; December, 1961.
[4] D. Bohm, "Causality and Chance in Modern Physics," D. Van Nostrand Co., Inc., Princeton, N. J.; 1957.
[5] S. Deutsch, "Causality, consciousness, and creativity," *Cybernetica*, vol. 4, pp. 154–170; 1961.

Capacitance Between Thin Film Conductors Deposited on a High-Dielectric-Constant Substrate*

Introduction

In microsystems electronics functional assemblies, it is sometimes desirable to use as a substrate a material of very high dielectric constant in order to achieve capacity between the opposite sides of the substrate. This may be required for lumped capacitors wherein two film conductors are deposited on either side of the substrate, or for distributed parameter *RC* networks, in which case a film resistor replaces one or both of the conductors.

In either case, the presence of the substrate may have deleterious effects on the over-all circuit performance by introducing stray capacitance between various conductors, where no capacitance is acceptable.

It is necessary to have a means for estimating these stray capacitances in order to account for their effects in the completed function morphology.

The idealized case of two parallel conductors of equal width on the same side of a plane dielectric of finite thickness and infinite extent may be analyzed analytically. This case is a very close approximation to the practical case of finite substrate dimensions, provided the substrate extends a few thicknesses beyond the conductors.

* Received April 6, 1962.

Analysis

The conductor-substrate morphology to be considered is illustrated in Fig. 1(a). The conductors and dielectric are of infinite length (perpendicular to the Z plane) and the capacitance between these conductors will be calculated per unit length. Since the dielectric constant of the substrate is assumed very large compared to that of the surrounding medium, all of the flux is considered confined to the dielectric.

A Schwarz-Christoffel transformation is used to map the Z plane of Fig. 1(a) onto the ω plane of Fig. 1(b). An inverse Schwarz-Christoffel transformation is then used to map the ω plane of Fig. 1(b) onto the ζ plane of Fig. 1(c). The dielectric region shown shaded in the Z plane maps onto the shaded regions in the ω and ζ planes. Thus, the capacitance calculation is reduced to the elementary calculation of the capacitance of parallel plates without fringing fields.

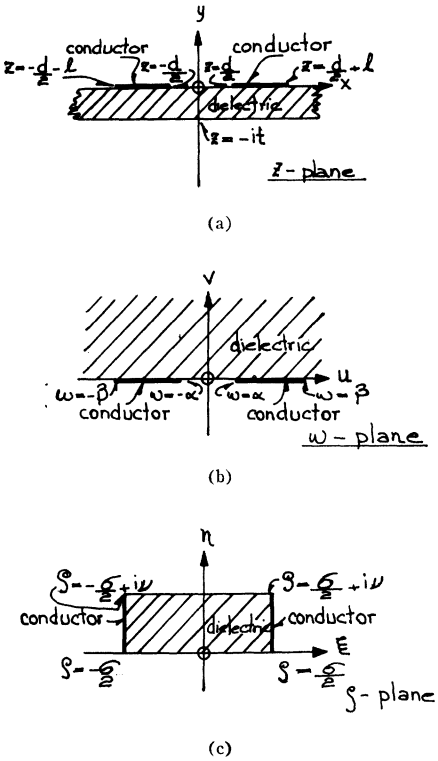


Fig. 1.

The transformation relating the Z and ω planes is readily obtained by the method of Schwarz-Christoffel as

$$Z = \int_0^{\omega} \frac{d\omega}{\omega^2 - \gamma^2} \tag{1}$$

or

$$Z = \frac{1}{2\gamma}\left[\log\frac{\omega - \gamma}{\omega + \gamma} - i\pi\right]. \tag{2}$$

The quantity γ is evaluated as $\pi/2t$ by noting that the point $Z = -it$ goes into the point $\omega = \infty$.

From (2) we obtain

$$\omega = -\frac{\pi}{2t}\tanh\frac{\pi Z}{2t} \tag{3}$$

so that

$$\alpha = \frac{\pi}{2t}\tanh\frac{\pi d}{4t} \tag{4}$$

and

$$\beta = \frac{\pi}{2t}\tanh\left(\frac{\pi l}{2t}+\frac{\pi d}{4t}\right) \tag{5}$$

in Fig. 1(b).

Similarly, the transformation relating the ζ and ω planes is

$$\zeta = \int_0^{\omega} \frac{d\omega}{\sqrt{(\alpha^2-\omega^2)(\beta^2-\omega^2)}} \tag{6}$$

which yields

$$\nu = \int_{\alpha}^{\beta} \frac{d\omega}{\sqrt{(\omega^2-\alpha^2)(\beta^2-\omega^2)}} \tag{7}$$

and

$$\sigma = \int_0^{\alpha} \frac{2d\omega}{\sqrt{(\alpha^2-\omega^2)(\beta^2-\omega^2)}} \tag{8}$$

in Fig. 1(c). The elliptic integral (7) can be put into the Legendre normal integral form of the first kind by means of the transformation

$$\omega = \beta\sqrt{1-x^2\left(1-\frac{\alpha^2}{\beta^2}\right)}$$

to yield

$$\nu = \frac{1}{\beta}F\left(\sqrt{1-\frac{\alpha^2}{\beta^2}},\frac{\pi}{2}\right) \tag{9}$$

in standard elliptic integral notation.

Similarly, the elliptic integral (8) is transformed by putting $\omega=\alpha x$ to yield

$$\sigma = \frac{2}{\beta}F\left(\frac{\alpha}{\beta},\frac{\pi}{2}\right). \tag{10}$$

The capacitance in the ζ plane [Fig. 1(c)] is simply

$$C = \frac{\epsilon_0 K\nu}{\sigma}$$

and from (9) and (10)

$$C = \frac{\epsilon_0 K}{2}\frac{F\left(\sqrt{1-\frac{\alpha^2}{\beta^2}},\frac{\pi}{2}\right)}{F\left(\frac{\alpha}{\beta},\frac{\pi}{2}\right)}\ \text{farad/meter.} \tag{11}$$

From (4) and (5)

$$\frac{\alpha}{\beta} = \frac{\tanh\frac{\pi d}{4t}}{\tanh\left(\frac{\pi l}{2t}+\frac{\pi d}{4t}\right)}$$

or

$$\frac{\alpha}{\beta} = \frac{1+\tanh\frac{\pi l}{2t}\tanh\frac{\pi d}{4t}}{1+\tanh\frac{\pi l}{2t}\Big/\tanh\frac{\pi d}{4t}}. \tag{12}$$

To calculate C for given l, d, and t, the quantity α/β is first calculated from (12) and this value used to evaluate (11) by means of tables of elliptic integrals.

Numerical Results

The capacitance in $\mu\mu$f/inch is given by

$$C(\mu\mu f/\text{inch}) = 0.1125K\frac{F\left(\sqrt{1-\frac{\alpha^2}{\beta^2}},\frac{\pi}{2}\right)}{F\left(\frac{\alpha}{\beta},\frac{\pi}{2}\right)}$$

and is plotted in Fig. 2 for unit dielectric constant K vs d/t for various values of l/t.

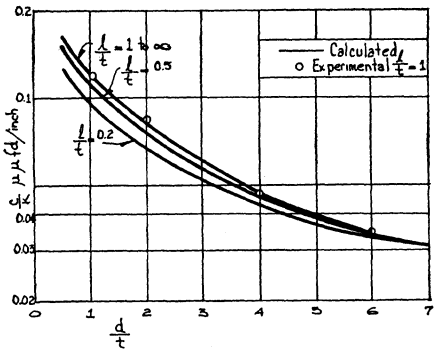


Fig. 2.

Measurements performed on a barium titanate substrate of 0.020 inch thickness and dielectric constant 1773 are also shown in Fig. 2 as circled points.

Acknowledgment

The preparation of substrates and the measurements are due to A. J. Nichols III.

H. R. Kaiser
P. S. Castro
Lockheed Missiles and Space Co.
Microsystems Electronics Dept.
Sunnyvale, Calif.

Hilbert Transforms of Band-Pass Functions*

The Hilbert transform $\hat{x}(t)$ of a (generally complex-valued) function $x(t)$ of a real variable t is defined by

$$\hat{x}(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(v)dv}{t-v} \tag{1}$$

where the integral is a Cauchy principal value. The form of (1) shows that $\hat{x}(t)$ is the convolution of $x(t)$ and $1/(\pi t)$. Therefore, the Fourier transform of $\hat{x}(t)$ is the product of the Fourier transform of $x(t)$ and that of $1/(\pi t)$. That is, if $X(\omega)$, $\hat{X}(\omega)$ are the Fourier transforms of $x(t)$ and $\hat{x}(t)$, respectively, then

$$\begin{aligned}\hat{X}(\omega) &= -jX(\omega), && \omega > 0\\ &= jX(\omega), && \omega < 0.\end{aligned} \tag{2}$$

Now, let

$$x(t) = a(t)\exp(j\omega_c t), \qquad \omega_c > 0$$

where, in general, $a(t)$ may be complex-valued. If $A(\omega)$ is the Fourier transform of $a(t)$, the Fourier transform of $x(t)$ will be

$$X(\omega) = A(\omega - w_c).$$

Then

$$\begin{aligned}\hat{X}(\omega) &= -jA(\omega-\omega_c), && \omega > 0\\ &= jA(\omega-\omega c), && \omega < 0.\end{aligned}$$

If $x(t)$ is a band-pass function, $a(t)$ will be slowly varying compared to ω_c. In particular, let it be required that the total, nonzero extent of $A(\omega)$, including negative ω, be such that $A(\omega-\omega_c)=0$ for $\omega<0$. This will ordinarily require that the "video" bandwidth of $a(t)$ be less than ω_c. With this requirement,

$$\hat{X}(\omega) = -jA(\omega-\omega_c)$$

and

$$\hat{x}(t) = -ja(t)\exp(j\omega_c t).$$

Generally, for any real ω_c,

$$\hat{x}(t) = -j(\operatorname{sgn}\omega_c)a(t)\exp(j\omega_c t) \tag{3}$$

where "sgn" stands for "signum" and is $+1$ if $\omega_c>0$ and -1 if $\omega_c<0$.

Eq. (3) is the basic formula from which the entries in Table I may be derived. In every case, $a(t)$ is a function with transform $A(\omega)$, such that $A(\omega-\omega_c)=0$, for $\omega<0$, $\omega_c>0$ and for $\omega>0$, $\omega_c<0$.

TABLE I

Function, $x(t)$	Hilbert Transform, $\hat{x}(t)$
$a(t)\exp(j\omega_c t+j\theta)$	$-j(\operatorname{sgn}\omega_c)a(t)\exp(j\omega_c t+j\theta)$
$a(t)\cos(\omega_c t+\theta)$	$(\operatorname{sgn}\omega_c)a(t)\sin(\omega_c t+\theta)$
$a(t)\sin(\omega_c t+\theta)$	$-(\operatorname{sgn}\omega_c)a(t)\cos(\omega_c+\theta)$
	θ = constant

Harry Urkowitz
Philco Scientific Lab.
Blue Bell, Pa.

* Received April 5, 1962.

A Large Signal Behavior of a Low-Noise Traveling-Wave Tube and Its Application to Pulse Radar*

A novel pulse radar employing a circulator and low-noise traveling-wave tube as a switching device is suggested [1] and experimental results in X-band 50-kw and S-band 5-Mw radars are reported. In this duplexing system, by taking advantage of limiter action of the traveling-wave tube, the mixer crystals are completely protected from the leaking power of the transmitter. Differing from usual radar, our radars use no TR tubes and naturally there is no spike and therefore no fear about deterioration of the mixer crystals. It is another merit that

* Received April 3, 1962.

some extension of the radar range can be obtained in a low-noise traveling-wave tube. In S-band experiments, a ferrite switch is needed in order to protect the traveling-wave tube. High-power behavior of low-noise traveling-wave tubes is shown in Figs. 1 and 2. Gain of the X-band tube which was made on a trial basis at our laboratory bench is more than 8 db for the input power less than −20 dbm, for more power than −20 dbm the gain turns to decrease and gets into the attenuation band. In the S-band tube

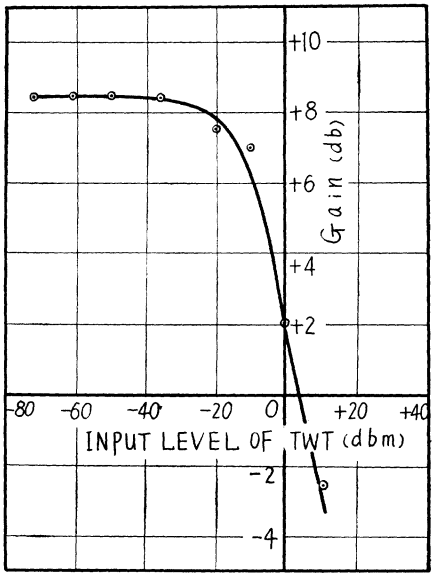


Fig. 1—High-power behavior of the X-band traveling-wave tube (f=9375 Mc).

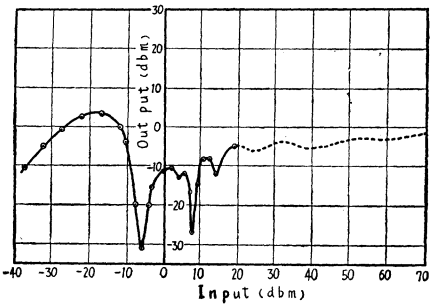


Fig. 2—High-power behavior of the traveling-wave tube 6861.

RCA 6861, however, the attenuation characteristics as a function of input power show a complicated figure. The output power of the tube drops sharply to −30 dbm for the input power of −5 dbm and +8 dbm. This complicated behavior seems to stem from the close coupling between the helix and electron beam.

A schematic diagram of the duplexing device in S-band 5-Mw radar is shown in Fig. 3. Though decoupling between ports 1 and 3 of the circulator is fairly large—say 25 db—the reflected power from the antenna, which is about 20 db down from the radiation power, adds to the former. Therefore, some kind of protection to the TWT was needed. For this purpose a Faraday-type ferrite switch was employed where insertion loss was 0.7 db and attenuation 30 db with an exciting coil current of 0.9 a. These were measured at the ambient temperature of 40°C.

Over-all noise figure was measured as 8.7 db in which 6.9 db was due to the TWT only. In an ordinary 5-Mw radar where GE 6621 was employed for the duplexing device, the NF of this system was 9 db in normal operation which sounded almost the same as our radar. However, the striking merit of our radar was its small recovery time. The recovery time of GE 6621 was more than 160 μsec in 5-Mw operation, but in our radar, there was no spike and no recovery time. However, some kind of pulse delay due to magnetizing coil in the ferrite switch had to be considered which measured less than 10 μsec.

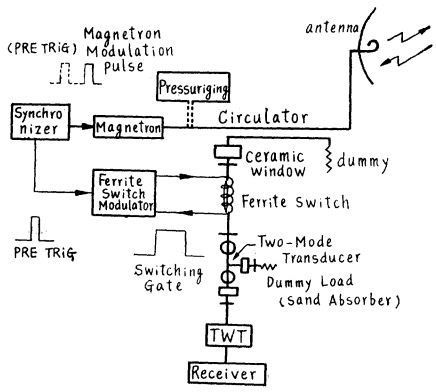


Fig. 3—A schematic diagram of a circulator-TWT-radar in 5-Mw operation.

In the experiment of the S-band 500-kw radar, the arrangement was almost the same as in the 5-Mw radar except for the circulator. As the 5-Mw circulator was too bulky, a complete tee-type circulator was employed where the applied magnetic field was above resonance of the ferrite.

Field tests of the two radars have been successfully done through the courtesy of the Japan Self-Defense Force. Sensitivity of the receivers were compared by echoes from a target plane. It was confirmed in 500-kw radar that the radar range was extended by 27 per cent in our TWT radar whose over-all NF was measured as 7.3 db. This was due to 4-db improvement of the NF compared to TR radars.

The authors are indebted to K. Kurihara who developed the ferrite and S. Takeuchi and T. Nagai for making the necessary measurements.

S. Mita
T. Takeya
S. Hayasi
K. Kakizaki
T. Yoshida
Tokyo Shibaura Electric Co., Ltd.
Kawaski, Japan

References

[1] S. Mita, S. Hayasi, Japanese Patent No. 219,555, February 2, 1956; Brit. Patent No. 812,082, April 26, 1960; U. S. Patent No. 2,934,638, April 26, 1960.
[2] S. Mita, S. Hayasi, and K. Kakizaki, "A new pulse radar employing a low-noise TWT and a ferrite duplexer," *L'onde electrique*, vol. 38, Annec S109, 1958.

Modified Radar Range Equation Relating Pulse Energy to Antenna Aperture or Gain in a Radar with a Specified Angular Accuracy*

The specifications for new radars usually include not only the requirement of maximum range but also angular accuracy both of which are related to signal-to-noise ratio. The relationship between transmitted power and antenna gain or area for a given maximum range differs from that where both maximum range and angular accuracy must be considered. This difference is discussed briefly in the following.

The usual radar range equation when the same antenna is used both on transmission and reception is

$$R^4 \propto \frac{P_p \tau G^2 \lambda^2 \sigma}{KTL(S/N)} \propto \frac{P_p \tau A^2 \sigma}{KTL(S/N)\lambda^2}$$

where

P_p = peak power transmitted
τ = pulse length
G = antenna gain
λ = wavelength
σ = target area
S/N = signal-to-noise power ratio
K = Boltzmann's constant
T = system noise temperature
L = system losses
A = antenna area.

These expressions imply that, for given values of R, required S/N, T, σ, and L, the energy

$$P_p \tau \propto \frac{\lambda^2}{A^2} \propto \frac{1}{G^2 \lambda^2}.$$

Thus, if the antenna size is changed so that the gain is reduced by 3 db, the required power must be increased by 6 db in order to maintain the range capability. This commonly used relationship applies if maximum range is the prime consideration.

If the basic radar equation is modified slightly, a similar expression is available relating energy to gain or antenna area with both range and rms angular error fixed. The rms angular error in the x plane of an antenna is[1]

$$\sigma_x \propto \frac{\lambda}{D_x} (S/N)^{1/2}$$

where D_x is the x dimension of the antenna. Similarly,

$$\sigma_y \propto \frac{\lambda}{D_y} (S/N)^{-1/2}$$

and

$$\sigma_x \sigma_y \propto \frac{\lambda^2}{D_x D_y} (S/N)^{-1} \propto \frac{\lambda^2}{A} (S/N)^{-1}.$$

Substituting for S/N in the basic radar equation,

$$R^4 \propto \frac{P_p \tau G^3 \lambda^2 \sigma \sigma_x \sigma_y}{KTL} \propto \frac{P_p \tau \sigma A^3 \sigma_x \sigma_y}{KTL\lambda^4}.$$

* Received April 2, 1962.
[1] R. Manasse, "Maximum angular accuracy of tracking a radio star by lobe comparison," IRE Trans. on Antennas and Propagation, vol. AP-8, pp. 50–56; January, 1960.

Now, with $\sigma_x\sigma_y$ fixed, along with R, T, L, and σ,

$$P_{pT} \propto \frac{\lambda^4}{A^3} \propto \frac{1}{G^3\lambda^2}.$$

Returning to the example used previously, if, in this case, the antenna size is changed so that the gain is reduced by 3 db, the power must be increased by 9 db in order to maintain both range and angle measuring capabilities instead of the 6 db required to measure range only.

R. A. ENSTROM
Radar Equipment Engrg.
Westinghouse Electric Corp.
Baltimore, Md.

A Relation Between α and Q*

In 1944, Davidson and Simmonds[1] derived a relation between the Q of a cavity composed of a uniform transmission line with short-circuiting ends and the attenuation constant α of such a transmission line. Later, in 1950, Barlow and Cullin[2] rederived this relation. These authors showed that this relation is quite general and is applicable to arbitrary cross-section uniform metal tube waveguides. Since then one of the standard techniques for the measurements of the attenuation constant α has become the use of the cavity method.[3] This method offers an excellent way of measuring the attenuation constant of a waveguide when the loss is quite small. Later on this method was generalized and applied to open waveguides,[4] such as the single-wire transmission line and the dielectric waveguide, by various authors.

However, it is noted that the formula by Davidson, etc., was derived under the assumption that there exists a single equivalent transmission line for the mode under consideration. This assumption is true for a pure TE, TM, or TEM mode, but it is not clear that such a single equivalent transmission line exists for a hybrid wave. This suspicion originates from the fact that 1) the TE and TM waves are intimately coupled to each other, and 2) the characteristic impedance defined by Schelkunoff[5] is not constant with respect to the transverse coordinates. It is, therefore, very difficult to conceive the possibility that there exists a single equivalent transmission line for this hybrid wave; at best the hybrid wave may be represented by a set of transmission lines coupled tightly with one another. Hence the formula by Davidson, etc., may not be applicable to a hybrid wave.

A more general relation[6] between Q and α can be obtained without using the transmission line equivalent circuit, provided that α is very small compared with the phase constant β and that the loss contributed by the short-circuiting end plates is negligible compared with the total loss of the waveguide section under consideration. The propagation constant of a guided wave with a small attenuation constant at ω_0 is

$$\Gamma(\omega_0) = \alpha(\omega_0) + i\beta(\omega_0). \tag{1}$$

At resonance, the following relation is true:

$$\Gamma(\omega_0) + \frac{\partial\Gamma}{\partial\omega}\Delta\omega \approx i\beta(\omega_0). \tag{2}$$

Combining (1) and (2) gives

$$\alpha(\omega_0) = -i\frac{\partial\beta}{\partial\omega}\Delta\omega. \tag{3}$$

Since the group velocity v_g and Q are given by the relations

$$v_g = \frac{\partial\omega}{\partial\beta},$$

and

$$Q = \frac{\omega_0}{2(\Delta\omega/i)},$$

one arrives at the relation

$$\alpha = \frac{\omega_0}{2Qv_g} = \frac{v_p}{v_g}\frac{\beta}{2Q} \tag{4}$$

where v_p is the phase velocity of the wave. This is the general relation that is sought.

Substituting the values of v_p/v_g for a TE, TM, or TEM wave into (4), one gets the relations derived by Davidson, etc. For a TM or TE wave in a metal waveguide,

$$\frac{v_p}{v_g} = \frac{1}{1-\left(\frac{\lambda}{\lambda_c}\right)^2}, \tag{5}$$

where λ_c is the cutoff wavelength of the wave under consideration. Thus

$$\alpha_{TE,TM} = \frac{1}{1-\left(\frac{\lambda}{\lambda_c}\right)^2}\frac{\beta}{2Q}. \tag{6}$$

For a TEM wave,

$$\frac{v_p}{v_g} = 1. \tag{7}$$

Hence

$$\alpha_{TEM} = \frac{\beta}{2Q}. \tag{8}$$

For a hybrid wave, v_g and v_p are not simply related. They may be obtained graphically from the ω-β diagram. However, for a dominant hybrid wave on a dielectric rod at very low frequencies or at very high frequencies, the relation (8) is a good approximation since at these frequencies $v_p \approx v_g$.[7]

C. YEH
Elec. Engrg. Dept.
Univ. So. Calif.
Los Angeles, Calif.
Formerly with
Calif. Inst. Tech.
Pasadena, Calif.

* Received April 9, 1962.

[1] C. F. Davidson and J. C. Simmonds, "Cylindrical cavity resonators," *Wireless Engr.*, vol. 21, pp. 420–424; September, 1944.

[2] H. M. Barlow and A. L. Cullen, "Microwave Measurements," Constable and Company, Ltd., London, England; 1950.

[3] E. L. Ginzton, "Microwave Measurements," McGraw-Hill Book Co., Inc., New York, N. Y.; 1957.

[4] C. H. Chandler, "An investigation of dielectric rod as waveguides," *J. Appl. Phys.*, vol. 20, pp. 1188–1192; December, 1949.
E. H. Sheibe, B. G. King, and D. L. Van Zeeland, "Loss measurements of surface wave transmission lines," *J. Appl. Phys.*, vol. 25, pp. 790–797; June, 1954.

[5] S. A. Schelkunoff, "The impedance concept and its application to problems of reflection, refraction, shielding and power absorption," *Bell Sys. Tech. J.*, vol. 17, pp. 17–48; January, 1938.

[6] Private communication with Prof. R. W. Gould, Calif. Inst. Tech., Pasadena, Calif.

[7] C. Yeh, "Electromagnetic surface wave propagation along a dielectric cylinder of elliptical cross section," Ph.D. thesis, Calif. Inst. Tech., Pasadena, 1962.

Man-Machine Communication

*M. J. Pedelty**

Mr. Radcliffe appears to have overlooked most of the important points concerning the man-computer communication problem.[1] He seems to propose a new class of specialist trained to respond to the characteristics of the machine (whatever that may mean). This is precisely the situation we are trying to avoid in the construction of new computer "languages," for example, for reasons which hardly bear repeating. In any case a central point is that the characteristics of computer hardware and software are changing in a time scale incomparable to that of human characteristics. What long-term good, then, would "two semesters of computer pseudo-speech" do a twelfth grader?

Even more strenuous objection might be raised due to what Mr. Radcliffe does not say. First, in referring to man's need "to give up his native language . . . in order to earn his living." one may remark that this is hardly the "last straw" but rather the first since machine language has been constantly giving way to "people-oriented" language. *A fortiori*, to refer in this context to "a machine of lightning speed and powerful memory" is totally inept in the light of the clearly demonstrated properties of self-organizing systems and "intelligent" machines. It is analogous to designing the Niagara Falls power station as though it were to be driven by the steam engine of James Watt. Ultimately the only mismatch which will exist between man and machine will be due to his relatively low upper bound on communication (and *conscious* computation) rate, his fallibility and his need to operate with comparatively informal structures. Every effort must be made to take advantage of whatever sensory motor discrimination a man has,[2] but it is too much to hope that the computer user of the future will adapt very much to the machine.

M. J. PEDELTY
Adaptronics, Inc.
Annandale, Va.

* Received April 23, 1962.

[1] A. J. Radcliffe, Jr., "Man-machine communication," PROC. IRE, vol. 50, p. 482; April, 1962.

[2] J. C. Bliss, "Kinesthetic-tactile communications," IRE TRANS. ON INFORMATION THEORY, vol. IT-8, pp. 92–98; February, 1962.

P. I. Hershberg[3]

In a recent paper,[1] Radcliffe suggests the interconnection of the human mind to the computing machine for purposes of obtaining advantages absent in either device. In one case, he advises " coupling a living organism in a feedback loop with a computer in a manner devised to condition the organism to communication with the machine." Is not this condition present in the many teaching machines utilizing digital computers, with the reward and punishment signals appearing in the form of wrong and right answers?

Problems involved in man-machine communication are similar to those consequent to machine-machine communication. While both machines may have capabilities realized through megacycle operating speeds, the communication facility between computers may be capable of only kilocycle data transmission rates. The result is that if we wish to have extensive communication between computers, the performance of each must be degraded to that of the communication facility.

This situation is analogous to man-machine communication. If we are to couple the machine to some peripheral nerve of the body, the maximum data rate of the nerve is given by

$$V_n = Kmd^2$$

where V_n is the maximum data rate in bits per second, K is a constant, m is the myelin factor of the nerve fiber and d is the nerve fiber diameter. Since most peripheral nerves have relatively low myelin factors and diameters, the maximum data rate for these nerve fibers is of the order of fifty bits per second.

If the optic nerve is utilized, it must be recognized that a basically slow chemical reaction converts light to nervous impulses (rhodopsin-iodopsin reaction). This may be verified by walking from a bright room to a dark one and noting that it requires a few seconds for the eyes to once again pick up input information.

In contrast with these lower data rates, the human brain has an extremely fast clock rate. For example, it takes no more than a few seconds for the cortex to scan several hundreds of thousands of bits of data "stored" in the human memory. This clock rate is of the order of megacycles and is comparable with the digital computer.

Of course if one is to couple data into the cortex, it is necessary to understand the function of the cortex in terms of its component circuitry. Yet Eccles has stated that there has been little correlation between the enormous amounts of data gathered on the brain and the "activity of cortical neurones."[4] In other words we don't really know anything about how the brain functions in terms of its basic building blocks.

Clearly, a better understanding of the human data processor is necessary before we attempt to couple this device to another machine.

PHILIP HERSHBERG
AF Cambridge Research Labs.
Bedford, Mass.

[3] Received April 26, 1962.

[4] J. C. Eccles, "The Neurophysiological Basis of the Mind," Oxford University Press, New York, N. Y., p. 258; 1953.

Author's Comment[5]

In response to the letter of Mr. Pedelty, I wish to state two of the basic premises of my previous letter:

1) Verbal or written language, man's primary means of conveying concepts which he might wish to communicate to a computer, has evolved over only a relatively short period of time and has been shaped toward a maximum basic survival capability in a crudely hostile environment over most of that period.
2) The speech process and the sonal structure of our various languages are inherently limited by evolutionary processes, extending over a million years, which shaped the structure of our vocal and hearing organs on the basis of an environment encompassing factors quite foreign to the man-machine relationship.

With reference to 1) man has difficulty in conveying concepts clearly to another human due to the uniquely implicit values tied to words by each individual as a consequence of his own living experience. If semantic problems present a barrier to human-human communication, then communication with a machine may similarly be plagued. While humans might communicate with each other purely on the basis of the "dictionary meaning" of words, this is not the actual case. Human communication is assisted materially by interpretations on the part of the listener which are influenced by past experience with the particular speaker. Much meaning is added to speech by accompanying gestures or facial expressions.

My second premise leads to a questioning of the efficiency of speech for man-machine communication. The speech process involves concessions on the part of our vocal organs to limitations of the ear; and very probably the listening process has limitations related to limitations of the vocal organs. These mutual limitations are not inherent in man-machine communications.

It must also be recognized that the human brain has adaptive and learning capability which can be brought to use at rather low cost to overcome limitations in everyday language and speech.

Why then employ expensive input-output machine hardware to circumvent weaknesses in the structure of our language and inefficiencies in the speech process? It appears to me that we will not continue indefinitely to base computer input and output languages (which need not even be the same for man-machine communication) on the language and speech that evolution and circumstance have left for us. When machines of magnitude sufficient to generate genuine semantic problems in man-machine communication have been built, it may prove more practical for man to adapt to the machine, as its potentialities are explored, rather than continually modify a massive piece of hardware.

One of the best ways to train man, as Mr. Hershberg points out, is to employ a "teaching machine" which involves programmed responses by the machine that are conditional upon input signals from the learner. This is feedback of a sort, but not what I had in mind for improving man-machine communication.

Two strangers, motivated to communicate but lacking a common language, may be expected to evolve means for communicating, given sufficient time. Recognizing in advance that a problem of this sort was to be encountered would lead them to preparations that would improve the efficiency of this mutual learning process. It appears that man and machine might be coupled in a similar closed learning loop with hopeful prospects for improved communication as a consequence. The time constants of this closed loop would be a great deal shorter than the time constants of the communication loop, concerned with this same problem, of which this letter is part.

Both Mr. Hershberg and Mr. Pedelty refer to the discrepancy in speed between computation rate in man or machine and in the information rate available in visual, aural or tactile communication channels to or from a human. While we cannot change the propagation velocity of nerve impulses, we can perhaps speed communication by greater emphasis of the parallel mode of operation as opposed to series. There is a great deal more information in a symbol from the Chinese written language than in a letter of the Latin alphabet. Granted that one is more amenable to typesetting than the other, still one presents more information than the other. Is there not a lesson to be learned here in respect to improved communication between man and machine? Similarly, it takes no longer to say crag, steed or gloom than rock, horse or dark, but the former certainly seem to convey more meaning with the same input-output effort.

The associative and adaptive capabilities present in the human brain are not likely to be exceeded by a machine for some time to come. In the meantime it appears to be desirable to determine how man can modify his customary means of communication to improve man-machine communication.

A. J. RADCLIFFE, JR.
2407 S. Summerlin
Orlando, Fla.

[5] Received May 8, 1962.

The Discrete Binomial Time Interval Distribution*

The time interval distribution as derived from the Poisson distribution is a useful relation when dealing with random pulses. In some pulse systems, the time interval is restricted to multiples of a finite resolving time. Hence, it would be desirable to have a discrete time interval distribution describing this condition. A discrete distribution is derived from the binomial distribution that can be numerically evaluated from standard tables.

* Received April 11, 1962.

For Poisson-distributed pulses the probability of x pulses in time t is

$$P_x(t) = \frac{(at)^x}{x!} e^{-at} \tag{1}$$

where a is the average pulse rate. The probability of no pulses ($x=0$) in time t is

$$P_0(t) = e^{-at}. \tag{2}$$

If the period t is to be a time interval between successive pulses, the interval must terminate with the next pulse. The probability of a pulse in dt at t is (adt). The combined probability of no pulse in time t followed by a pulse in dt at t is the product of these two probabilities.

$$dP(t) = ae^{-at}dt. \tag{3}$$

The probability of a time interval less than t is obtained by integrating (3)

$$P(t) = \int_0^t dP(t) = (1 - e^{-at}). \tag{4}$$

Note that (3) and (4) are continuous functions.

A discrete relation analog to (3) and (4) can be derived from the binomial distribution. The probability of x pulses in r time increments of length τ is

$$B_x(r) = \binom{r}{x} (a\tau)^x (1 - a\tau)^{r-x} \tag{5}$$

where $\binom{r}{x}$ is the binomial coefficient and $r=t/\tau$ is an integer. The probability of no pulses ($x=0$) in r increments is

$$B_0(r) = (1 - a\tau)^r. \tag{6}$$

The probability of a pulse in the $(r+1)$ increment is $a\tau$. The combined probability of no pulses in r increments followed by a pulse in the $(r+1)$ increment, *i.e.*, the probability of a time interval r increments long is

$$B(r) = a\tau(1 - a\tau)^r. \tag{7}$$

The probability of a time interval less than n increments long is obtained by summing (7)

$$B(n) = \sum_{r=0}^{n} B(r) = 1 - (1 - a\tau)^{n+1}. \tag{8}$$

Eqs. (7) and (8), the desired discrete relations, are analogs to (3) and (4), respectively.

It is readily shown that (4) is the limiting expression for (8), just as the Poisson distribution can be considered as the limiting distribution for the binomial distribution. Writing (8) in terms of t and τ

$$B(n) = 1 - (1 - a\tau)^{(t/\tau)+1}. \tag{9}$$

If $a\tau \ll 1$, $(1-a\tau)$ is approximately $e^{-a\tau}$ and

$$B(n) \sim 1 - e^{-a(t+\tau)}. \tag{10}$$

Letting the increment τ approach zero and n approach infinity such that $t=n\tau$ is finite yields

$$\begin{aligned} B(n) &\sim 1 - e^{-at} \\ B(n) &\sim P(t). \end{aligned} \tag{11}$$

Numerical values for $B(n)$ are given in the Table of the Binomial Probability Distribution.[1] For table use $B(n)$ is rewritten in terms of the binomial expansion as

$$B(n) = \sum_{r=1}^{n+1} \binom{n+1}{r} (a\tau)^r (1 - a\tau)^{n+1-r}. \tag{12}$$

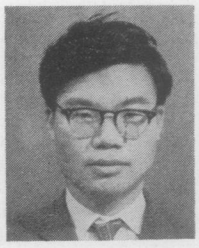
This form is tabulated in the tables.

The preceding relations have proven useful in evaluating sampling rates for digital data systems handling random Poisson-distributed pulses.

Ned Wilde
Atomic Energy Division
Phillips Petroleum Co.
Idaho Falls, Idaho

[1] "Table of the Binomial Probability Distribution," Natl. Bureau of Standards, Applied Mathematics Series No. 6, 1950.

Contributors



Suguru Arimoto was born in Hiroshima, Japan, on August 3, 1936. He received the B.S. degree in mathematics from the University of Kyoto, Kyoto, Japan, in 1959.

Following his graduation he joined the Digital Computer Section of the OKI Electric Company, Japan. He is presently at the University of Tokyo as a Research Assistant in applied physics.

Mr. Arimoto is a member of the Institute of Electrical Communication Engineers of Japan.

Lee F. Barash (A'50–M'55) was born in New York, N. Y. on December 28, 1927. She received the B.A. degree in physics from New York University, in 1947, and the M.S. degree in physics from Stanford University, Calif., in 1948.

From 1948 to 1954 she worked in the Electron Tube Department of the Sperry Gyroscope Company, Great Neck, N. Y. She was engaged in research and development of klystrons and traveling-wave tubes. In 1955 she worked at Jasik Laboratories, Westbury, N. Y., on antenna design. During the years 1959 to 1961 she was at Columbia University, New York, N. Y., for graduate study and held a Graduate Assistantship at Columbia Radiation Laboratories. At present she is retired for the third time.

Mrs. Barash is a member of the American Physical Society and Sigma Xi.

❖

Edward Bedrosian (S'51–A'53–M'54–SM'56) was born in Chicago, Ill., on May 22, 1922. He received the B.S.Ae.E. degree from the Aeronautical University, Chicago, in 1942, and the B.S.E. and M.S. degrees, from Northwestern University, Evanston,