Correspondence

Theory of Shot Noise in Junction Diodes and Junction Transistors*

In the above paper,[1] the shot noise in junction transistors was calculated for a one-dimensional model, both for low and for high frequencies. The low-frequency calculation was exact, the high-frequency calculation was carried out under the assumption that the equilibrium hole concentration p_n in the base region and the excess hole concentration p_{c0}' at the collector junction were negligible.

The noise has now been calculated without these assumptions. It was found that the following equations, derived in my article,[1] are of general validity:

$$\langle i_{p1}^2\rangle_{av} = 4kTG_e df - 2eI_e df \quad (1)$$

$$\langle i_{p2}^2\rangle_{av} = 2eI_c df \quad (2)$$

$$\langle i_{p1}{}^* i_{p2}\rangle_{av} = -2kT\alpha Y_e df. \quad (3)$$

Here α is the current amplification factor, Y_e the emitter admittance, G_e the emitter conductance, i_{p1} the emitter current generator and i_{p2} the collector current generator. If $V_c<0$, and $|eV_c/kT|\gg 1$, the emitter and collector current may be written

$$I_e = I_{ee}(e^{eV/kT}-1) + \alpha_0 I_{cc} \quad (4)$$

$$I_c = \alpha_0 I_{ee}(e^{eV/kT}-1) + I_{cc}, \quad (5)$$

where α_0 is the low-frequency current amplification factor and the current $I_{ee}=I_{cc}$ has been defined.[1] Substituting (4) into (5) yields

$$I_c = \alpha_0 I_e + (I_c)_{sat}. \quad (6)$$

The quantity

$$(I_c)_{sat} = I_{cc}(1-\alpha_0^2) \quad (6a)$$

is known as the *collector saturated current*, that is the collector current for zero emitter current.

The noise current generator

$$i_p = i_{p2} + \alpha i_{p1}, \quad (7)$$

corresponding to the output noise current generator for open input, has also been introduced.[1] Substituting (1), (2), (3) and (6a), we obtain

$$\langle i_p^2\rangle_{av} = 2eI_c df - |\alpha|^2 2eI_e df$$
$$= 2e(\alpha_0 - |\alpha|^2)I_e df + 2e(I_c)_{sat} df. \quad (8)$$

For low frequencies this reduces to

$$\langle i_p^2\rangle_{av} = 2e\alpha_0(1-\alpha_0)I_e df + 2e(I_c)_{sat} df. \quad (8a)$$

The reader will recognize the first term as the partition noise term introduced by van der Ziel and the second term as the shot noise term due to the collector saturated current introduced by Montgomery and Clark.[2]

The collector saturated current thus gives full shot noise at all frequencies. This is in contrast with the statement,[1] according to which the collector saturated current should *not* give full shot noise. This discrepancy is caused by a wrong definition of $(I_c)_{sat}$; instead of (6a) the definition $(I_c)_{sat}=I_{cc}$ was used. The first term in (8) gives the high-frequency equivalent of the low-frequency partition noise term.

Eqs. (1), (2), and (3) do not give any reference to the one-dimensional model for which they were derived. It thus seems likely that they hold for all geometries.

The author is indebted to Dr. K. M. van Vliet, University of Minnesota, for stimulating discussions on the subject, and to Dr. R. L. Pritchard, General Electric Company, for pointing out the correct expression for $(I_c)_{sat}$.

ALDERT VAN DER ZIEL
Elec. Eng. Dept.
University of Minnesota
Minneapolis, Minn.

* Received by the IRE, November 6, 1956. Work supported by the U. S. Signal Corps.
[1] A. van der Ziel, PROC. IRE, vol. 43, pp. 1639–1646; November, 1955.
[2] H. C. Montgomery and M. A. Clark, "Shot noise in junction transistors," *J. Appl. Phys.*, vol. 24, pp. 1397–1398; November, 1953.
A. van der Ziel, "Note on shot and partition noise in junction transistors," *J. Appl. Phys.*, vol. 25, pp. 815–816; June, 1954.

Molecular Amplification and Generation of Microwaves*

My above paper[1] was not intended to give a comprehensive survey of the literature on molecular amplification, but was meant to give a brief description of the present state of the field and an idea of its scope and promise.

However, I should like to mention here, as I did not do in my article, that the possibility of obtaining microwave amplification by means of stimulated emission was pointed out by Weber at the Tenth Annual Conference on Electron Tubes at Ottawa, Can., in June, 1952, and in the IRE TRANSACTIONS ON ELECTRON DEVICES.[2]

JAMES P. WITTKE
RCA Laboratories
Princeton, N. J.

* Received by the IRE, March 22, 1957.
[1] Proc. IRE, vol. 45, pp. 291–316; March, 1957.
[2] J. Weber, "Amplification of microwave radiation by substances not in thermal equilibrium," IRE TRANS., PGED-3, pp. 1–4; June, 1953.

Alternatives to Cathode Bias for Vacuum Tubes*

Although the use of a cathode resistor, by-passed if desired, is by far the commonest way of providing bias for tubes when some self-regulation of the operating point is desired, there are alternatives which may be attractive under certain circumstances. If a fairly well regulated negative voltage is available, as often happens in such applications as instrumentation, the scheme shown in Fig. 1 may be used. The voltage divider formed by resistors R_1, R_2, and R_3, working between the plate voltage V_p and the negative supply voltage $-V_c$, holds the grid at the proper operating point, the cathode being grounded directly. The capacitor C_1 by-passes all signal frequencies preventing degeneration of the signals.

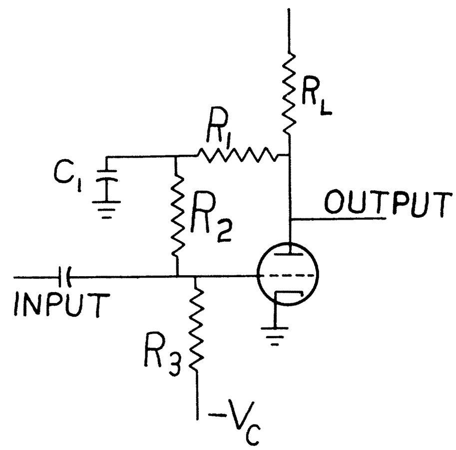

Fig. 1—The voltage divider bias system for a vacuum tube. By means of the divider chain, the plate voltage is compared with the fixed voltage $-V_c$, and the feedback from this comparison stabilizes the operating point.

It can be seen, from ordinary feedback theory, that any tendency of the operating point to change, due to a change in the properties of the tube, is attenuated approximately by a factor $\{1+(g_mR_LR_3)/(R_1+R_2+R_3)\}^{-1}$, g_m being the transconductance of the tube. If $R_1\sim R_2\sim R_3$, and $g_mR_L\gg 1$, this factor becomes approximately $3/(g_mR_L)$. When the bias is provided by a cathode resistor R_k, on the other hand, the factor is approximately $(1+g_mR_k)^{-1}$. Typically $g_mR_k \sim 1$, so the factor is around one half.

There are certain other possible advantages of the system shown in Fig. 1. Since R_1 will be as much as one thousand times as great as R_k would be for a similar situation, C_1 need be only about one thousandth the size of the capacitor required to by-pass R_k. This could mean savings in size, weight, and cost. Having the cathode grounded directly may help in reducing hum, and, in addition, it might be desirable if, for some reason, tubes with filamentary cathodes were to be used.

The same system may be used with a pentode. Fig. 2 shows another alternative which may be attractive here, especially when the screen has to be by-passed anyway. Here the screen provides the regulation, just as the plate did in the former circuit. Since the grid to screen and grid to plate

* Received by the IRE, April 1, 1957.

transconductances are in about the same ratio as the screen current and plate current, and the plate and screen run at about the same voltage, the attenuating factor for changes in the properties of the tube will be about the same when the regulation is taken from the screen as when it is taken from the plate. The screen method saves one resistor and one capacitor (if the screen has to be by-passed anyway), and is also useful if the plate load is something like a transformer, with a very low dc impedance.

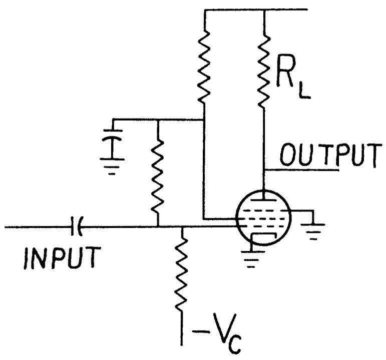

Fig. 2—The voltage divider bias system applied to a pentode by comparing the screen voltage with the fixed voltage. This method requires no additional capacitors, in case the screen has to be by-passed anyway.

HAROLD L. ARMSTRONG
Pacific Semiconductors, Inc.
Culver City, Calif.,
Formerly with Nat. Res. Council
Ottawa, Ont., Can.

A Linear Cathode-Ray Tube*

There have been numerous attempts to correct, for the nonlinear light output vs grid drive characteristic of cathode-ray tubes, (crt). In the rooter circuit developed by Oliver,[1] for instance, the crt is preceded by a circuit with a nonlinearity which is the inverse of that of the crt so that the over-all circuit is linear. It is not too difficult to match these nonlinearities with the precision required for most television applications. A further improvement in the precision of linearity, however, becomes increasingly difficult. In addition, a perfect match, if once obtained, would be difficult to maintain, since the two tubes involved would not age at precisely the proper relative rates.

The above objections are largely eliminated in a tube of the construction shown in Fig. 1. The first three electrodes are similar to the cathode, grid, and plate of a conventional lighthouse tube such as the 2C40, *i.e.*, a large area cathode, a parallel wire mesh control grid, and a parallel plate (screen grid in our tube), the plate, however, having a small hole through its center. Because of the parallel geometry, the electrons approach the plate perpendicularly, and a small, definite fraction of them penetrates the hole to form the beam which goes on to excite the phosphorescent screen. A large current (milliamperes) flows to the plate. However, this is at low (300) voltage. The power dissipated is less than that required in Oliver's rooter circuit. It is easily possible to force this plate current to vary linearly with the input voltage by using a conventional video feedback circuit. The beam current is a small, definite fraction of the plate current, and hence, its variation is accurately linear too. Note that aging of the tube no longer is a problem; it is corrected for by the feedback.

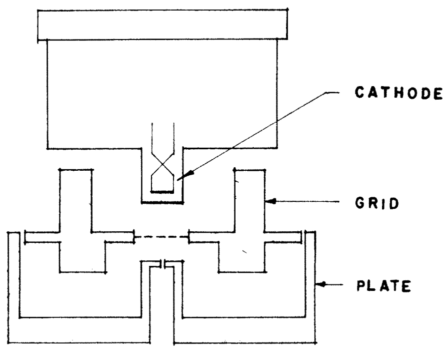

Fig. 1.

Conventional crt's can be corrected in this manner. However, the cathode current which must be sampled by the feedback circuit is so small that it becomes impossible to obtain the wide-band pass required in radar and television applications. The "g_m" on conventional tubes is of the order of 10 micromhos.

There remains the question of the linearity of the phosphor itself. As long as saturation is avoided, the phosphor in itself appears to be linear. However, its response falls off with temperature, and localized heating by the electron beam is sufficient to reduce the light out: beam current characteristic from a one to an approximate 0.8 power law within a 10 to 200 microampere current range. (Our tests have shown that P11, P15, and P16 phosphors behave this way.) It is possible, however, by properly under-correcting the gun nonlinearity, to make the over-all response sufficiently linear. The simplest feedback circuit which will do this is merely a selected cathode resistor.

The success of this device hinges upon obtaining accurate proportionality between the plate and beam currents. This is possible if several precautions are observed.

1) The grid-plate spacing should be large enough so that thermal motions of the electrons will be sufficient to fill in the "shadows" cast by the individual grid wires before the electrons reach the plate. If the hole in the plate is partially shadowed, the beam current may be too small at low currents.
2) The grid supporting structure should partially enclose the plate, as shown, to confine the field from the plate. Otherwise, as the grid bias decreases, some electrons hit the side of the plate and its supporting structure so that the fraction of the current going through the hole again is too small.
3) The space within the gun must be carefully shielded from the accelerating field to prevent electrons being drawn around rather than through the plate.

It is a pleasure to acknowledge the assistance of the staff of the University of Illinois, Electron Tube Laboratory, and especially that of Murray Babcock.

D. BITZER
AND R. D. RAWCLIFFE
Control Systems Lab.
University of Illinois
Urbana, Ill.

* Received by the IRE, March 23, 1957.
[1] B. M. Oliver, "Tone rendition in television," PROC. IRE, vol. 38, pp. 1288–1300; November, 1950.

Quantum Derivation of Energy Relations Analogous to Those for Nonlinear Reactances*

In a paper by Manley and Rowe,[1] equations are given relating the average powers at the different frequencies in nonlinear reactances. These general energy relations are very interesting and give useful information regarding the gain and stability of nonlinear reactor modulators and demodulators. However, the derivation of the above equations involves a rather complicated Fourier analysis.

It is interesting to note that in a quantum mechanical system, the Manley-Rowe relations are almost selfevident. For example, (26) and (27) or (30) and (31) of Manley and Rowe,[1] applicable to a single-sideband modulator or demodulator, can be readily obtained by using the equivalent three-level quantum mechanical system shown in Fig. 1. The principle that one must utilize is that in the steady state, the total number of quanta per second leaving any level must equal the number arriving at that level. Therefore, one can write

$$\begin{aligned} N_{12} + N_{13} &= 0 \\ N_{12} &= N_{23} \end{aligned} \qquad (1)$$

where

N_{12} = no. of quanta per second going from level 1 to level 2
N_{13} = no. of quanta per second going from level 1 to level 3
N_{23} = no. of quanta per second going from level 2 to level 3
$N_{21} = -N_{12}$ = no. of quanta per second going from level 2 to level 1.

The energy in each quantum is hf where h is Planck's constant and f is the frequency of the transition. Therefore, $N_{12}hf_1$ is equal to W_1, the power in frequency f_1, with similar relations for N_{13} and N_{23}. One can therefore write (1) as follows:

* Received by the IRE, April 2, 1957.
[1] J. M. Manley and H. E. Rowe, "Some general properties of nonlinear elements," PROC. IRE, vol. 44, pp. 904–913; July, 1956.

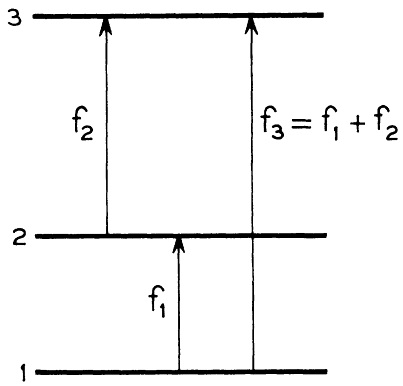

Fig. 1—Quantum level scheme equivalent of single-sideband modulator or demodulator.

$$\frac{W_1}{f_1} + \frac{W_3}{f_3} = 0$$
$$\frac{W_1}{f_1} = \frac{W_2}{f_2}. \qquad (2)$$

These equations are identical to (26) and (27) or (30) and (31) of Manley and Rowe,[1] by noting the difference in symbolism.

One can derive similar relations for the case of a nonlinear system in which both side bands carry power. Fig. 2 shows the

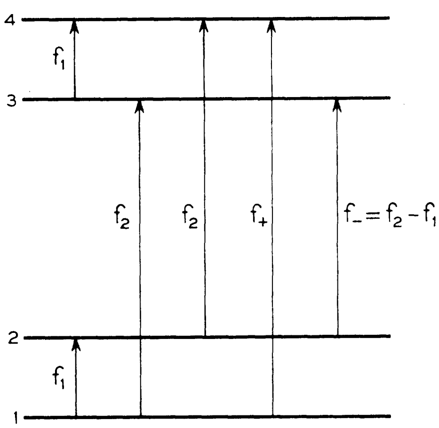

Fig. 2—Quantum level scheme equivalent of uppe and lower sideband modulator or demodulator.

equivalent quantum mechanical system which we shall analyze using the principle of conservation of quanta in the steady state.

$$N_{12} + N_{13} + N_{14} = 0$$
$$N_{23} + N_{24} - N_{12} = 0 \qquad (3)$$
$$N_{34} - N_{13} - N_{23} = 0.$$

Combining the first two of (3) we get

$$N_{13} + N_{24} + N_{14} + N_{23} = 0.$$

Similarly, combining the first and last of (3) we get

$$N_{12} + N_{34} + N_{14} - N_{23} = 0.$$

Noting that $W_2 = (N_{13}+N_{24})hf_2$ and $W_1 = (N_{12}+N_{34})hf_1$ and substituting in the above equations, we obtain

$$\frac{W_2}{f_2} + \frac{W_+}{f_+} + \frac{W_-}{f_-} = 0$$
$$\frac{W_1}{f_1} + \frac{W_+}{f_+} - \frac{W_-}{f_-} = 0 \qquad (4)$$

where W_+ and W_- are the upper and lower sideband powers respectively. These equations are again identical to those obtained from (24) and (25) of Manley and Rowe.[1]

One can alternatively derive the above relations by considering a photon interaction scheme rather than a multilevel quantum system. Thus, let us derive (4) by considering the following scheme. A material is irradiated with photons of frequency $f_+ = f_1 + f_2$ and of energy hf_+. We assume that this material is in a cavity which can absorb energy at frequencies f_1, f_2, and $f_- = f_2 - f_1$. Two different "decay" processes, shown in Fig. 3, are possible, In process A, a photon of energy hf_+ results in one photon of energy hf_1 and one photon of energy hf_2, so that energy is conserved. In process B a photon of energy hf_+ results in one photon of energy hf_- and two photons of energy hf_1, again conserving energy for this process.

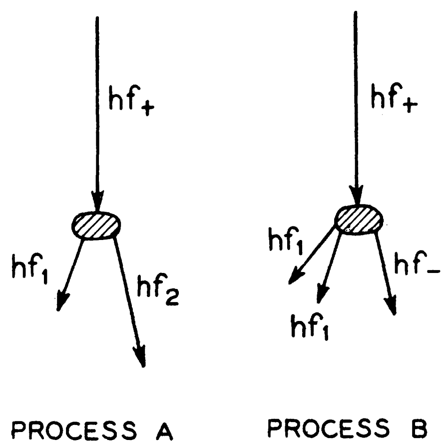

Fig. 3—Quantum interaction scheme equivalent of Fig. 2.

Let

$-N_{+A}$ = no. of hf_+ photons absorbed by process A in one second
$-N_{+B}$ = no. of hf_+ photons absorbed by process B in one second
N_{1A} = no. of hf_1 photons emitted by process A
N_{1B} = no. of hf_1 photons emitted by process B
N_2 = no. of photons of energy hf_2 emitted per second
N_- = no. of photons of energy hf_- emitted per second
$-N_+ = -(N_{+A}+N_{+B})$ = total no. of photons of energy hf_+ absorbed per second
$+N_1 = +(N_{1A}+N_{1B})$ = total no. of photons of energy hf_1 emitted per second.

From Fig. 3 one can readily see that the following equations hold:

$$\left.\begin{aligned} N_{1A} &= -N_{+A} \\ N_2 &= -N_{+A} \end{aligned}\right\}\text{process A} \qquad (5)$$

$$\left.\begin{aligned} N_{1B} &= -2N_{+B} \\ N_- &= -N_{+B} \end{aligned}\right\}\text{process B.} \qquad (6)$$

Combining the second of (5) with the second of (6) we obtain

$$N_2 + N_- = -(N_{+A} + N_{+B}) = -N_+$$

or

$$\frac{W_2}{f_2} + \frac{W_-}{f_-} + \frac{W_+}{f_+} = 0$$

which is identical to the second of (4).

Similarly, by combining the first of (5) and (6) we obtain

$$N_1 = N_{1A} + N_{1B} = -(N_{+A} + 2N_{+B})$$
$$= -N_+ - N_{+B} = -N_+ + N_-$$

or

$$N_1 + N_+ - N_- = 0$$

or

$$\frac{W_1}{f_1} + \frac{W_+}{f_+} - \frac{W_-}{f_-} = 0$$

which is the first of (4).

If one knows the relative probability of occurrence of process A to process B, one can, of course, obtain additional information regarding the powers at the different frequencies. Thus, if process A and process B were equally probable one can readily see that

$$\frac{W_1}{f_1} = -\frac{3}{2}\frac{W_+}{f_+}.$$

These simple energy relations are particularly significant in the analysis of the potentialities of the various multilevel solid-state maser[2] schemes.

MAX T. WEISS
Bell Telephone Labs., Inc.
Holmdel, N. J.

J. B. Wittke, "Molecular amplification and generation of microwaves," PROC. IRE, vol. 45, pp. 291–316; March, 1957.

Microwave Mixing and Frequency Dividing*

Data obtained from frequency mixing experiments and from related experiments in frequency division demonstrate the feasibility of achieving effective mixing and dividing in microwave tubes. In all of the experiments, the nonlinearity of an overmodulated electron beam was used to produce the desirable effects. As will be described in some detail below, mixing conversion gains as high as 30 db and strong divided-frequency signals have been obtained from traveling-wave-type tubes.

Interest in the traveling-wave mixer has been stimulated by certain characteristics which these tubes can possess. For example, such tubes may be designed to have wide IF bandwidths, to give IF outputs at microwave frequencies, to be free from burnout due to high-power inputs, to provide local-

* Received by the IRE, March 23, 1957. This paper prepared under Air Force Contract AF 19(604) 1847.

oscillator isolation from the signal input and to give high-level output power. The possibility of including the local oscillator within the tube envelope by using backward-wave structures has been investigated elsewhere by Gray.[1]

Much experimental data were obtained from an S-band traveling-wave tube having two identical helices as in Fig. 1. The tube was operated as a mixer with the local oscillator signal and the input signal applied to the input of the first helix. The IF signal was then obtained at the output of the second helix. In the experiment illustrated by Fig. 1, the measured gain through the first helix for both a high-level signal at f_1 and a low-level signal at f_2 was about 8 db. The second helix was operated at a higher voltage in the dispersive region in order to peak the gain at the difference frequency f_3. The measured gain through this helix for signals at f_3 was very nearly 0 db. The high-level signal served to drive the electron beam into a partially-saturated condition in order to produce a difference-frequency modulation on the beam. The difference-frequency modulation then was coupled to the second helix and hence to the mixer output. The conversion gain, defined as the ratio of the IF output power to the low-level signal input power, was +27 db.

The above experiment was repeated after adding an external attenuator (marked "added attenuation") to the second helix. The purpose of the attenuator was to reduce the effects of possible feedback through the second helix. Such effects may have been partially responsible for the high conversion gain because the added attenuation resulted in lowering the conversion gain to +16 db.

The operation throughout the above experimentation was very stable. For example, even with the attenuator absent, a substantial increase in the beam current would still not produce oscillations. By increasing the beam current 50 per cent the conversion gain could be increased to +30 db. It is also interesting to note that by raising the input power levels, the tube's full saturation power output of +15 dbm was obtainable at the intermediate frequency. Consequently, high-level mixing can be accomplished in this type of tube.

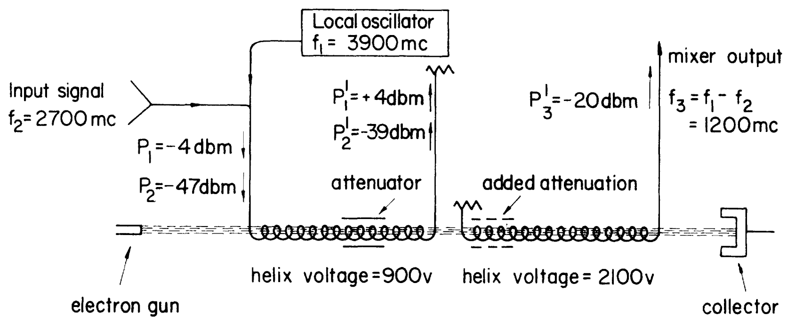

Fig. 1—Double-helix traveling-wave-tube mixer. The first helix was operated for broad-band amplification of S-band signals (including the frequencies f_1 and f_2) with a gain of +8 db. The second helix was operated in the dispersive region and was tuned for the IF frequency f_3, the gain for which was 0 db. In the experiment illustrated, the conversion gain, taken as the ratio of P_3' to P_2, was +27 db.

Another mixer experiment was performed by using a conventional, helix-type X-band traveling-wave tube. An S-band resonant cavity was fitted over the tube envelope and positioned with the cavity gap concentric with the beam and located in the region between the helix output and the beam collector. A high-level signal at 10,100 mc and a low-level signal at 7600 mc were applied to the helix input. The cavity was tuned to 2500 mc and was excited by the difference frequency current present in the electron beam passing to the collector. A conversion gain of +11 db could be obtained. At the same time the tube gain as an amplifier was +32 db. Obviously this method of obtaining a mixer output was much less effective than the previous one.

An interesting application for the double-helix mixer tube of Fig. 1 is its use as a frequency divider. This mode of operation can be understood by assuming that the frequency of the high-level signal, f_1, is exactly twice that of the low-level signal, f_2. As a consequence, f_3 equals f_2. By adding a feedback path as in Fig. 2 the tube can be made to oscillate precisely at $f_2 = f_1/2$. This, of course, constitutes dividing by two.

Because of the gain mechanism inherent in the mixing process, the oscillations at f_2 will continue so long as the signal at f_1 is present. Since the straight-through gain for f_2 is too low to sustain oscillations, the oscillations cease when f_1 is removed.

The double-helix tube was tested as a frequency divider with good results. The frequency f_1 was chosen to be 4000 mc. For the best operation it was necessary to adjust the helix voltages to obtain proper phase relationships. Under these conditions, the system could easily be excited into oscillation at 2000 mc by supplying a high-level signal at 4000 mc to saturate the beam. Thus, the tube's full saturation power output was obtainable at the divided frequency.

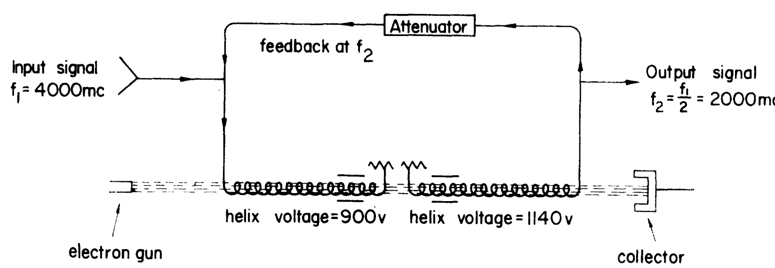

Fig. 2—Frequency divider employing a double-helix traveling-wave-tube mixer. The first helix was operated for broad-band amplification covering the 2000 to 4000-mc range. The second helix was tuned to peak the gain at 2000 mc.

Successful dividing was also obtained from a conventional (*i.e.*, a single-helix) S-band traveling-wave tube. Preliminary tests had shown that under the proper operating conditions the gain for the difference-frequency signal in this tube was substantially higher than the gain for either of the signals being mixed. This was precisely the characteristic which had permitted successful operation of the double-helix tube as a divider. Thus the conventional tube was connected with its output fed back through an attenuator to the input as in Fig. 2; except, of course, only one helix was present. The tube was operated as a frequency divider and gave successful performance over a bandwidth of about 10 mc. The operation was less stable and more critically dependent upon proper adjustment than in the case of the double-helix tube.

A number of experiments were made to investigate certain characteristics of the mixing. The results of a theoretical analysis by Putz[2] had indicated that for tubes of fixed dimensions the conversion gain should be proportional to the square of the intermediate frequency. These results were verified for low intermediate frequencies by taking the IF output signal directly from a traveling-wave-tube collector through a load resistor. As shown in Fig. 3, the variation

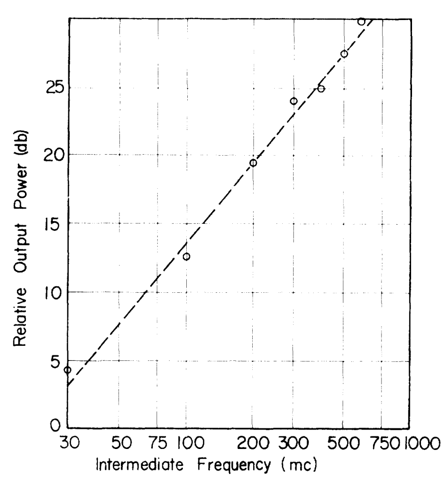

Fig. 3—Relative IF output power as a function of the intermediate frequency. For these data, the input power level of the signals which were mixed was held constant. The IF output was taken from a traveling-wave-tube collector operating into a 50-ohm load.

[1] G. A. Gray, "Investigation of Mixing in Traveling-Wave Tube Amplifiers," Ser. No. 60, Issue No. 151, Elec. Res. Lab., University of Calif., Berkeley, Calif.; November 11, 1955.

[2] J. L. Putz, "Nonlinear Phenomena in Traveling-Wave Amplifiers," Tech. Rep. No. 37 (N6onr 251), Elec. Res. Lab., Stanford University, Stanford, Calif.; October 15, 1951.

of the output power for constant input power (and thus the variation of the conversion gain) was as the square of the intermediate frequency. These data were taken with a 50-ohm collector load resistor. Since the collector capacity is fixed as in a pentode amplifier stage, the obtainable mixer conversion gain must be a function of bandwidth. This fact was verified by inserting a 30-mc tuned circuit as a collector load having a resonant impedance of about 16,000 ohms and a bandwidth of 0.6 mc. The conversion gain was thereby increased by 25 db. (This tube then had an over-all conversion gain of +7 db at 30 mc with a tube gain of 35 db.)

Additional experimental data were obtained from the above tube to determine the mixer saturation characteristics. These data are presented in Fig. 4. The relative conver-

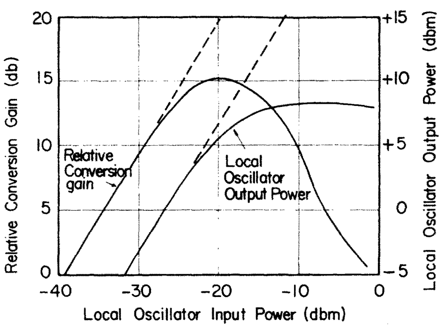

Fig. 4—Relative conversion gain and local oscillator output power as a function of local oscillator input power. The IF output was taken from the collector while the local oscillator output was taken from the helix of a traveling-wave tube.

sion gain is shown as a function of the local oscillator input power. Also shown is the curve for the local oscillator ouput power from the traveling-wave-tube helix. As is apparent, the conversion gain reaches its peak well before large signal saturation, for the tube as an amplifier, sets in. Consequently, it is felt that a first-order mixing analysis can be used successfully to predict conversion gain and mixing effects in the useful region of operation of this type of tube. Such an analysis is being made at the present time.

R. W. DeGrasse
and G. Wade
Stanford Electronics Labs.
Stanford University
Stanford, Calif.

A Name and Unit for Handling Admittances Due to Coils*

The problem of a suitable name and unit for what has been called "inverse inductance" and "inverse mutual inductance" has been discussed several times in the correspondence columns of Proceedings. The inadequacy of names such as "inverse inductance" (or "inverse" anything), and of symbols that imply inversion, is more or less obvious. The need for a better name became a practical necessity, however, when it was required to develop a simplified approach to filter theory and design, which could be readily understood and applied at all levels of engineering.

It was clear from the start that the simplest approach was via nodal analysis rather than loop analysis, for the reason that node equations translate into equivalent circuits far more readily than do the corresponding mesh equations. But, an immediate difficulty was encountered in trying to present this approach for general consumption when it was found that a very large number of engineers tend to think of admittance as the inverse of impedance. Such a notion makes nodal analysis an impossibly complicated mental contortion. It proved most essential to replace this "inverted viewpoint" with a clear understanding of admittance as a measure of the ability to admit current—measured by the current that flows per volt applied.

In terms of nodal analysis, it was found that a considerable simplification of filter theory can be made by considering all filters as band-pass configurations made up of sections coupled by "mutuals" rather than by ponderable elements. High-pass, low-pass, stop filters, and so on, then drop out as special cases of this very simple and easily manipulated basic configuration. The approach is the opposite of the usual approach which makes the low-pass filter the basic configuration. Its success depends on the simplicity with which mutually-coupled circuits can be described in terms of admittances, provided coils are described in terms of parameters related to their admittances.

The admittance of coils has therefore been expressed in terms of their *"acceptance,"* A, which is defined as a measure of the rate at which a coil accepts current. Thus, if a voltage, E, is applied to a pure acceptance, A, the current, i, that flows will be given by:

$$i = AEt.$$

The unit of acceptance is, therefore, one ampere per volt-second; hence acceptance is measured in amperes per volt-second. We can write this: i/et, where i stands for amperes, e for volts, and t for seconds; and the source of the somewhat whimsical, but completely descriptive name, "ippets" (spelled i/et), as the name for units of acceptance, is clear.

Acceptance, A, has been defined to include coefficients of magnetic coupling, therefore inductance, L, of a coil, is not simply the reciprocal of the coil's acceptance (except when the coil is not magnetically coupled to any other coils). In general, inductance, L, is related to acceptance, A, by:

$$L = \frac{1}{A(1 - \phi^2)}$$

where ϕ is the total effective coefficient of magnetic coupling. If a coil is magnetically coupled to other coils with individual coefficients of coupling, φ_{n1}, φ_{n2}, φ_{n3}, etc., then:

$$\phi^2 = \varphi_{n1}^2 + \varphi_{n2}^2 + \varphi_{n3}^2 \text{ etc.}$$

Magnetic coupling between any two coils can be replaced by a mutual acceptance, A_m. Corresponding mutual inductance, L_m, is related to mutual acceptance by:

$$L_m = \frac{\varphi^2}{A_m(1 - \varphi^2)}.$$

It will be noted that when magnetic coupling is replaced by a coil, as in Fig. 1,

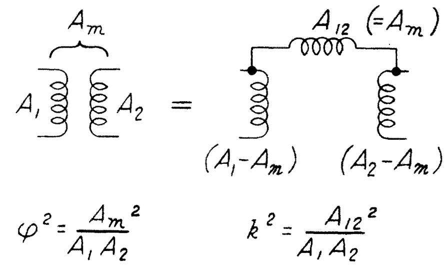

Fig. 1—Equivalent circuits and coefficients of coupling.

the coefficient of *magnetic* coupling, ϕ, becomes zero. It is replaced by the coefficient of coupling, k, which is given by:

$$k = \frac{A_{12}}{\sqrt{A_1 A_2}}$$

and the inductance of each of the coils in the equivalent π of Fig. 1 is then simply the reciprocal of the coil's acceptance. In terms of "ippets," a coil whose acceptance is 10^6 i/et (one "megippet") corresponds to a coil whose inductance is 10^{-6} henries (one microhenry) (when there are no magnetic couplings). "Megippet" is of course spelled Mi/et and "kilippet" (which corresponds to an inductance of 1 millihenry) is spelled ki/et.

The simplicity with which circuits containing mutual acceptance (or mutual inductance, which amounts to the same thing) can be handled in terms of acceptances and mutual acceptances, has indicated that it is advantageous to express a whole class of networks in terms of the mutually-coupled equivalents. Thus, the equivalent circuits of Fig. 1 are not restricted to coils. We may consider the coils to be any admittances whatsoever, coupled by a "mutual field" admittance y_m. There is then a coefficient of "field" coupling, φ, and this enters into the definition of the coupled admittances. When the circuit is eventually transformed back to a simple π, however, φ goes to zero and therefore the fact that it was ever anything other than zero can be ignored. In the same way, the presence of ϕ in the definition of inductance in terms of acceptance, can also be ignored, unless magnetically coupled circuits are to be actually constructed that way.

The filter study based on this approach is the subject of a report being prepared for the Air Force. The terms, acceptance, A, mutual acceptance, A_m, and ϕ for total effective coefficient of magnetic coupling, with φ for coefficient of magnetic coupling between any two coils, are introduced therein. The unit of acceptance, i/et, is also described, and the generalization of "field" coupling under the symbols, φ and ϕ, is used. Apart from this study, however, it was felt that the names and units for handling admittances due to coils, which have been described, would be of general interest.

F. Sutherland Macklem
Instruments for Industry, Inc.
Mineola, N. Y.

* Received by the IRE, April 8, 1957.

Two Theorems for Dissipationless Symmetrical Networks*

It is the purpose of this letter to derive two little known, but convenient theorems that enable one to quickly determine the insertion loss and insertion phase delay of a dissipationless, symmetric network from a knowledge of the input admittance of a half section of that network.

It is shown that the insertion loss $1/T^2$ of a dissipationless network of the type illustrated in Fig. 1 can be determined from

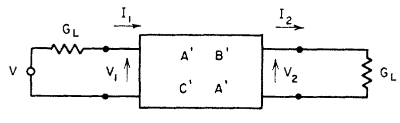

Fig. 1—Dissipationless symmetrical network.

the midplane admittance or impedance obtained when the network is terminated at each end in its load conductance $G_L=1/R_L$, by means of the relation[1]

$$\frac{1}{T^2}=1+\left[\frac{B_m}{G_m}\right]^2=1+\left[\frac{X_m}{R_m}\right]^2. \quad (1)$$

Here B_m/G_m is the ratio of midplane susceptance to conductance, while X_m/R_m is the ratio of midplane reactance to resistance. Similarly, the insertion phase of the network terminated at each end with conductance G_L can be determined from the susceptances

$$B_{oc}=-\frac{1}{X_{oc}} \quad \text{and} \quad B_{sc}=-\frac{1}{X_{sc}}$$

at the input terminals of the network, obtained when an open circuit and a short circuit are placed respectively at the midplane of the network. The relation is

$$\phi=\pi/2+\tan^{-1}\frac{B_{sc}}{G_L}+\tan^{-1}\frac{B_{oc}}{G_L}$$
$$=\frac{\pi}{2}+\tan^{-1}\frac{X_{sc}}{R_L}+\tan^{-1}\frac{X_{oc}}{R_L}\cdot \quad (2)$$

These relations may be demonstrated in the following manner.

The insertion voltage ratio $1/T$ of the network shown in Fig. 1 is defined as the voltage at the load when no network is present divided by the voltage at the load when the network is present. The input and output voltages and currents of this network are related in terms of the A', B', C', D' constants by

$$\begin{aligned} V_1 &= A'V_2+B'I_2\\ I_1 &= C'V_2+D'I_2. \end{aligned} \quad (3)$$

By applying the appropriate boundary conditions to (3) and remembering that $D'=A'$ in this symmetric network, it is easy to see that

$$\frac{1}{T}=A'+\frac{B'G_L}{2}+\frac{C'}{2G_L}\cdot \quad (4)$$

It follows from elementary matrix theory that if the network in Fig. 1 is bisected as shown in Fig. 2 and that if the $ABCD$ matrix of the left-hand section is

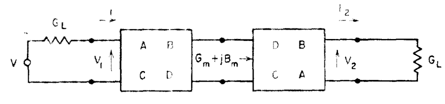

Fig. 2—Bisected dissipationless symmetrical network.

$$\left\| \begin{matrix} A & B \\ C & D \end{matrix} \right\|$$

then the $ABCD$ matrix of the right-hand section is

$$\left\| \begin{matrix} D & B \\ C & A \end{matrix} \right\|.$$

Therefore, the matrix of the whole symmetric network of Fig. 1 may be written as

$$\left\| \begin{matrix} AD+BC & 2AB \\ 2CD & AD+BC \end{matrix} \right\| \quad (5)$$

and the insertion voltage ratio for the whole network becomes

$$\frac{1}{T}=AD+BC+ABG_L+\frac{CD}{G_L}\cdot \quad (6)$$

For a dissipationless network, it is well known that A and D are real while B and C are imaginary. Hence the insertion power ratio (*i.e.*, insertion loss) $1/T^2$ is

$$\frac{1}{T^2}=(AD+BC)^2-\left(ABG_L+\frac{CD}{G_L}\right)^2. \quad (7)$$

The insertion phase delay ϕ on the other hand is given by the relation

$$\tan\phi=-j\left[\frac{1}{G_L\left(\frac{A}{C}+\frac{B}{D}\right)}+\frac{G_L}{\left(\frac{C}{A}+\frac{D}{B}\right)}\right]. \quad (8)$$

By applying the reciprocity relation $AD-BC=1$, (7) becomes

$$\frac{1}{T^2}=1-\left[ABG_L-\frac{CD}{G_L}\right]^2. \quad (9)$$

The admittance Y_m of the right-hand half of the network of Fig. 2 as measured at the midplane is

$$Y_m=G_m+jB_m=\frac{G_L+CD-ABG_L^2}{D^2-B^2G_L^2}\cdot \quad (10)$$

Substituting (10) in (9), one obtains (1), thus proving the first theorem.

Eq. (8) can be written in terms of B_{oc} and B_{sc} as

$$\tan\phi=\frac{1}{G_L\left[\frac{1}{B_{oc}}+\frac{1}{B_{sc}}\right]}-\frac{G_L}{B_{oc}+B_{sc}}$$
$$=\frac{B_{sc}B_{oc}-G_L^2}{G_L(B_{sc}+B_{oc})}\cdot \quad (11)$$

Using the trigonometric relation for the cotangent of the sum of two angles, one finds

$$\tan\phi=-\cot\left[\tan^{-1}\frac{B_{sc}}{G_L}+\tan^{-1}\frac{B_{oc}}{G_L}\right]$$
$$=\tan\left(\frac{\pi}{2}+\tan^{-1}\frac{B_{oc}}{G_L}+\tan^{-1}\frac{B_{sc}}{G_L}\right) \quad (12)$$

thus proving (2).

E. M. T. Jones
and S. B. Cohn
Stanford Res. Inst.
Menlo Park, Calif.

* Received by the IRE, March 5, 1957; revised manuscript received, April 15, 1957.

[1] This relation has been stated without proof by J. Reed, "Low-Q microwave filters," Proc. IRE, vol. 38, pp. 793–796; July, 1950.

Smooth Random Functions Need Not Have Smooth Correlation Functions*

This note is concerned with certain questions relating to the behavior of autocorrelation functions, defined by

$$\rho(\tau)\equiv\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t+\tau)f(t)dt \quad (1)$$

for functions $f(t)$ such that the limit (1) exists for every (real) value of τ. Sometimes this definition is modified by subtracting out the square of the mean, $\langle f\rangle^2$, where

$$\langle f\rangle\equiv\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)dt, \quad (2)$$

and dividing by the variance $\sigma^2\equiv\langle f^2\rangle-\langle f\rangle^2$, assumed positive. The two definitions coincide if $\langle f\rangle=0$ and $\sigma^2=1$, as we shall assume for the time being.

It is quite widely believed that the autocorrelation function obtained from an $f(t)$ with no "jump" discontinuities is differentiable at the origin. That is, it is thought that the graph of $\rho(\tau)$ necessarily has a horizontal tangent at $\tau=0$, provided only that $f(t)$ has no "jumps" such as those in an ideal square wave. This proposition seems to be some sort of "folk theorem;" *e.g.*, one finds:

> "Of course, the theoretical function . . . given in the figure cannot fit the data in the vicinity of the origin; this function has a finite slope [at the origin] which signifies infinitely sharp or steplike boundaries, a physical impossibility."[1]

This has even been cited as the basis, at least in part, for modifying or rejecting a proposed physical theory:

> "Finally, it should be noted that the first derivative of the correlation function (2) [exists] at the origin, as it must . . . to describe a physical process, and this eliminates one of the objections to the model proposed by Booker and Gordon which involved an exponential correlation function with a cusp at the origin."[2]

Such examples are quite common, and may be found in several branches of physics and engineering. However, there is one feature in common to every example—namely, that every time the "folk theorem" is invoked, it is always without reference to a proof. As a matter of fact, however, there is no hope of proving it; the object of this note is to point out that it is false, and to state mathematical conditions under which it is actually true that an autocorrelation function is differentiable at the origin. We shall also discuss certain practical considerations.

The fact that it is false has actually been known, though not widely so, since 1933, when Wiener[3] gave an example of a function

* Received by the IRE, March 15, 1957. This research was supported jointly by the U. S. Army, Navy, and Air Force under contract with the Mass. Inst. of Tech., Cambridge, Mass.

[1] L. Liebermann, "The effect of temperature inhomogeneities in the ocean on the propagation of sound," *J. Acoust. Soc. Amer.*, vol. 23, pp. 563–570; September, 1951.

[2] K. A. Norton, "Point-to-point radio relaying via the scatter mode of tropospheric propagation," IRE Trans., vol. CS-4, p. 42; March, 1956.

[3] N. Wiener, "The Fourier Integral and Certain of its Applications," Cambridge University Press (reprinted, Dover Press, New York, N. Y.), p. 151; 1953.

that was everywhere *analytic* (not merely continuous!) and uniformly bounded, but which had the autocorrelation function

$$\rho(\tau) = \begin{cases} 1, & \tau = 0; \\ 0, & \tau \neq 0, \end{cases} \qquad (3)$$

which is not even continuous at the origin, let alone differentiable there. Once it is appreciated that the "folk theorem" is indeed false, it is not very difficult to give other counter-examples; *e.g.*, one can construct a uniformly-bounded function with continuous, everywhere-finite derivatives of *all* orders, and which has the correlation function $\rho(\tau) = \exp - |\tau|$.

It is of interest to state conditions under which it is unequivocally true that an autocorrelation function is differentiable at the origin. In order to do this, we shall first review some definitions. A function $f(t)$ is said to satisfy a uniform Lipschitz condition of order p if there is a number M such that for $|\tau| \leq 1$,

$$|f(t+\tau) - f(t)| \leq |\tau|^p M \qquad (4)$$

for all t. A function $f(t)$ is said to be "uniformly differentiable" if it is differentiable and if, given any $\epsilon > 0$, there is a $\delta > 0$ such that for every τ in $0 < |\tau| < \delta$,

$$\left| \frac{f(t+\tau) - f(t)}{\tau} - f'(t) \right| < \epsilon \qquad (5)$$

for all t. It is then true that $\rho(\tau)$ will have a horizontal tangent at the origin whenever any one of the following three conditions is satisfied:

1) $f(t)$ satisfies a uniform Lipschitz condition of order $p > \frac{1}{2}$.
2) $f(t)$ is uniformly differentiable and has a finite mean square derivative $\langle (f')^2 \rangle$.
3) $f(t)$ is uniformly differentiable and the limit $\langle f f' \rangle$ exists and is equal to zero.

Condition 1) follows readily from the easily established relation

$$\left| \frac{1 - \rho(\tau)}{\tau} \right| = \frac{|\tau|^{2p-1}}{2\sigma^2} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left[\frac{f(t+\tau) - f(t)}{\tau^p} \right]^2 dt \qquad (\tau \neq 0) \qquad (6)$$

where we now use the normalized definition of $\rho(\tau)$ mentioned after (1). Condition 2) follows from (6) (with $p = 1$) and the Schwarz inequality. Condition 3) follows from the trivial fact that, for $\tau \neq 0$,

$$\frac{1 - \rho(\tau)}{\tau} = \frac{1}{\sigma^2} \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t) \frac{f(t) - f(t+\tau)}{\tau} dt. \qquad (7)$$

In some sense, condition 1) is the most general of the three, and cannot be sharpened very much; in particular, it may happen that 2) or 3) implies 1).

It has been conjectured by Eckart[4] that a function with a finite mean square derivative essentially satisfies a Lipschitz condition, in the sense that if $\langle (f')^2 \rangle$ is finite, then there is a function $g(t)$ satisfying a uniform Lipschitz condition and such that $\langle (f-g)^2 \rangle = 0$. If so, this would shed light on 2). This conjecture is quite attractive, though it does not appear to be a simple matter to establish the true state of affairs concerning it.

We might remark that 1) will hold (with $p = 1$) if in particular $f(t)$ has a uniformly bounded first derivative. Condition 3) will hold if in particular

$$\lim_{|t| \to \infty} [f^2(t)/t] = 0$$

(*e.g.*, if f is uniformly bounded) and if f has a uniformly continuous first derivative. If $f(t)$ has a uniformly-bounded second derivative, then $f'(t)$ is uniformly continuous, which implies that $f(t)$ is uniformly differentiable.

It is clear from the conditions that it would certainly be safe to assume that one or more of them held in a great many practical cases. On the other hand, there are a great many practical cases where it would be equally safe to ignore the conditions, and use a model correlation function that did not have a horizontal tangent at the origin. For example, consider Fig. 1, which is an

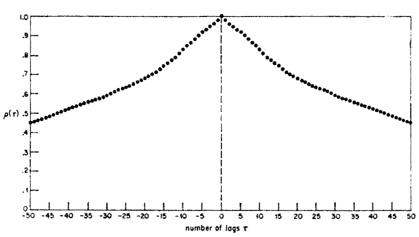

Fig. 1—Correlogram of the envelope of a fading radio wave.

empirical autocorrelogram of the envelope of a fading radio wave. One would be quite prepared to assume that one or more of the conditions held for the original envelope function. However, in spite of the fact that the correlogram has been computed with resolution a full order of magnitude better than is usually obtained $[1 - \rho(1) = 0.01$ in Fig. 1$]$, the obvious model correlation function to fit to the points of Fig. 1 would be one which did not have a derivative at the origin. (As a matter of fact, one of the form $\exp - |a\tau|$ does quite well.) The only circumstances in which this would be undesirable would be where some desired relation actually diverged or otherwise failed to exist in consequence of the nonexistent derivative,[5] in which case one would naturally use something else.

This situation is completely analogous to the many situations in physics and engineering where one uses (as a matter of convenience) the Rayleigh, normal, or some other probability distribution of a definitely unbounded sort to represent the distribution of a variable that would actually be bounded. For example, no one would actually expect to find gas molecules running about a room at speeds in excess of the velocity of light, even though the Maxwell-Boltzmann distribution would assign positive probability to this event. As always, such distributions are approximately valid only over certain finite ranges, which do not include either arbitrarily small or arbitrarily large values.

In the case of autocorrelation functions, meaningful questions will always involve values of τ bounded away from zero by some significant margin—*not* arbitrarily small values. The question of whether or not a given model correlation function actually has a derivative at the origin is thus ultimately vacuous, so far as physics is concerned, though the writer would be the last to deny the utility of classical analysis for physics and engineering. It might be added that analogous statements are true of the behavior of power spectra at high frequencies.

We might point out that (6) is useful in investigating the "reasonableness" of proposed model correlation functions at finite values of τ, and for interpreting the meaning of empirical correlograms. Setting $p = 1$, (6) yields the interpretation: The magnitude of the slope from the origin to the point $[\tau, \rho(\tau)]$ is equal to $(\tau/2)$ times the mean square slope of f (measured in units of the standard deviation σ) over intervals of fixed length τ. This is useful for, among other things, estimating limits of validity for proposed correlation functions.

The writer is indebted to several people, including Martin Balser, T. J. Carroll, Carl Eckart, C. L. Mack, and Robert Price for critical comments and stimulating discussions on various aspects of this subject.

Donald G. Brennan
Lincoln Laboratory
Mass. Inst. Tech.
Lexington, Mass.

[4] C. Eckart, Scripps Institution of Oceanography, La Jolla, Calif.; private communication.

[5] For example (3.3–11), of S. O. Rice, *Bell Sys. Tech. J.*, vol. 24, p. 54; January, 1945, could not be used with model correlation functions of this type.

Transient Response in FM*

In the above paper,[1] Gumowski derived expressions for the frequency-step and phase-step responses of networks in terms of the impulse responses of their zero-center-frequency analogs. This note is intended to present derivations which are considerably simpler and which sidestep the mathematical complications of Gumowski.[1]

Let the input be

$$e(t) = z(t) \exp(j\omega_0 t),$$

where $z(t)$ is, in general, a complex function. The output of the network may be written as

$$E(t) = Z(t) \exp(j\omega_0 t); \; Z(t) = X(t) + jY(t);$$

and its instantaneous frequency is

$$\omega(t) = \omega_0 + d/dt[\tan^{-1}(Y/X)].$$

The zero-center-frequency analog $A(j\omega)$ of the network's transfer function $A'(j\omega)$ is defined by

$$A'(j\omega) \equiv A[j(\omega - \omega_0)].$$

* Received by the IRE, April 5, 1957.

[1] I. Gumowski, Proc. IRE, vol. 42, pp. 819–822; May, 1954.

Using a well-known transform property, $Z(t)$ may be found[2] as the response of the analog network $A(j\omega)$ to the excitation $z(t)$. In particular, $Z(t)$ may be expressed in terms of the impulse response $u(t)$ of the analog network $A(j\omega)$.[3]

As an example, consider deviations from center frequency and zero phase, starting at $t=0$; *i.e.*, let

$$z(t) = 1, \quad (t < 0)$$
$$= \exp [jF(t)], \quad (t>0).$$

The output $Z(t)$ consists of the superposition of three responses: The steady-state response to a unit dc input; the response to a negative unit step occurring at $t=0$; and the response to the input $\exp [jF(t)]$, applied at $t=0$. The first of these responses is simply $A(0)$. The other two may be written in terms of $u(t)$. One obtains immediately

$$Z(t)=A(0)-\int_0^t u(x)dx+\int_0^t e^{jF(t-x)}u(x)dx,$$

which is essentially (10) of Gumowski.[1] An input frequency jump from center frequency ω_0 to (ω_0+a) is described by $F(t) = \exp (jat)$. The corresponding output is[4]

$$Z(t)=A(0)-\int_0^t u(x)dx+e^{jat}\int_0^t e^{-jax}u(x)dx.$$

Similarly, $F(t)=\exp (j\Phi)$ represents a carrier phase jump from zero to Φ radians, and the output function is[5]

$$Z(t) = A(0) - (1 - e^{j\Phi}) \int_0^t u(x)dx.$$

An extension of Gumowski's results will provide a final example: consider a frequency jump from (ω_0-b) to (ω_0+a). Here

$$z(t) = \exp (-jbt), \qquad (t < (0)$$
$$= \exp (jat), \qquad (t > 0).$$

Again considering the output as the sum of three responses, one may write

$$Z(t) = A(-jb)e^{-jbt} - \int_0^t e^{-j(b-x)}u(x)dx + \int_0^t e^{ja(t-x)}u(x)dx.$$

Inputs other than frequency and phase steps may be handled in the same manner. Further, the output $Z(t)$ can be expressed in terms of the unit step response of $A(j\omega)$ rather than its impulse response.

If the zero-frequency analog is realizable, *i.e.*, if the band-pass network is narrow and symmetrical, the above derivations may be rephrased in terms of in-phase and quadrature components. For example, an input frequency jump from ω_0 to (ω_0+a) corresponds to an input function

$$e(t) = \cos \omega_0 t, \qquad (t < 0)$$
$$= \cos (\omega_0 + a)t, \qquad (t > 0).$$

Alternatively,

$$e(t) = x(t) \cos \omega_0 t - y(t) \sin \omega_0 t,$$

where

$$x(t) = 1, \qquad (t < 0)$$
$$= \cos at, \qquad (t > 0)$$

and

$$y(t) = 0, \qquad (t < 0)$$
$$= \sin at, \qquad (t > 0).$$

The corresponding output is

$$E(t) = X(t) \cos \omega_0 t - Y(t) \sin \omega_0 t,$$

where $X(t)$ and $Y(t)$ are the responses of the low-pass analog to $x(t)$ and $y(t)$, respectively. The original network is thus represented by two identical low-pass analogs, one in the in-phase and one in the quadrature channel.[6] The instantaneous output frequency is again given by the time derivative of arc tan (Y/X).

DONALD A. LINDEN
Philco Corp.
Philadelphia, Pa.

[2] For example, E. A. Guillemin, "Communication Networks," John Wiley and Sons, New York, N. Y., vol. II, p. 470; 1935.

[3] The functions $z(t)$ and $Z(t)$ are "vector envelopes" in the sense defined by H. A. Wheeler, "The solution of unsymmetrical-sideband problems with the aid of zero-frequency carrier," PROC. IRE, vol. 29, pp. 446–458; August, 1941.

[4] Gumowski, *op. cit.*, (5a).

[5] *Ibid.*, (7a) represents the output in response to a carrier phase jump from Φ to zero radians, and is therefore not identical with this result.

[6] This concept has been used extensively by M. J. E. Golay in the analysis of fm phenomena.

Phase Error of a Two-Phase Resolver*

In the application of two-phase resolvers, it is sometimes useful to know the error resulting from an inequality between the two-phase voltages. This error may be expressed as the difference between the space angle of the rotor, θ_m, and the electrical angle of the rotor voltage, θ_e.

If the amplitude of the reference phase voltage is set equal to unity, and the amplitude of the quadrature voltage is A, then the phasor diagram (Fig. 1) can be drawn

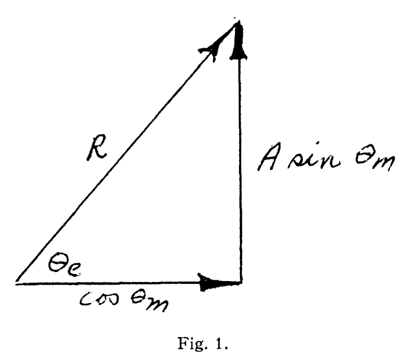

Fig. 1.

for the induced voltages in the rotor. This leads to the relationship:

$$\theta_e = \tan^{-1} (A \tan \theta_m).$$

Subtracting the angle θ_m from each side:

$$\theta_e - \theta_m = \tan^{-1} (A \tan \theta_m) - \theta_m.$$

This expression gives the angular error for any rotor position. The maximum angular error may be obtained by differentiating the above expression and setting the result equal to zero.

$$\frac{d(\theta_e - \theta_m)}{d\theta_m} = \frac{A \sec^2 \theta_m}{1 + A^2 \tan^2 \theta_m} - 1 = 0$$

$$A \sec^2 \theta_m = 1 + A^2 \tan^2 \theta_m$$

$$A(1 + \tan^2\theta_m) = 1 + A^2 \tan^2 \theta_m$$

$$\tan \theta_m = \sqrt{\frac{1}{A}}.$$

Substituting this result in the expression for the angular error and taking the tangent of both sides:

$$\tan (\theta_e - \theta_m) = \frac{\sqrt{A} - \frac{1}{\sqrt{A}}}{2}.$$

It can be seen from this result that an unbalance of only one per cent in the phase voltages will result in a maximum angular error of more than 17 minutes which is considerably in excess of the inherent error of many of these devices.

JACOB SCHACHTER
Pitometer Log Corp.
New York, N. Y.

* Received by the IRE, March 29, 1957.

A Simplified Procedure for Finding Fourier Coefficients*

Application of Gibbons' methods in the above article[1] to a different representation of the Fourier series may be of interest.

Consider the function $f(x)$ with period X and Fourier series:

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos \omega nx + b_n \sin \omega nx). \quad (1)$$

An alternate representation of $f(x)$ is

$$f(x) = \sum_{n=-\infty}^{\infty} \alpha_n e^{jn\omega x} \quad (2)$$

where

$$\omega = \frac{2\pi}{X}; \quad \alpha_n = \frac{1}{X}\int_0^X f(x)e^{-jn\omega x}dx \quad (3)$$

$$\alpha_n = \begin{cases} \frac{1}{2}(a_n + jb_n) \text{ if } n \geq 0 \\ \alpha_{-n}^* \text{ if } n < 0 \end{cases} \quad (4)$$

$$\int f(x)dx = \int \sum_{n=-\infty}^{\infty} \alpha_n e^{jn\omega x}dx$$
$$= {\sum_{n=-\infty}^{\infty}}' \frac{\alpha_n}{jn\omega} e^{jn\omega x} + \alpha_0 x + K. \quad (5)$$

Σ' indicates omission of the $n=0$ term.

The α_n's for a train of delta functions

$$\sum_{i=0}^{s} \sum_{j=0}^{\infty} \delta(x - x_i - j X);$$
$$0 \leq x_0 < x_1 < \cdots < x_s < X \quad (6)$$

* Received by the IRE, April 5, 1957.

[1] J. F. Gibbons, PROC. IRE, vol. 45, p. 243; February, 1957.

are

$$\sum_{i=0}^{s} \frac{1}{X} e^{-jn\omega x_i}. \tag{7}$$

Consider a function $g(x)$ which has $h(x)$ as its derivative. Since $g(x)$ is periodic, $h(x)$ is also periodic, and has zero average value.

$$g(x) = \sum_{n=-\infty}^{\infty}{}' \frac{\alpha_n}{jn\omega} + K. \tag{8}$$

Now, assume that by differentiating some function $g(x)$ m times, we arrive at a function $k(x)$ which is a sum of delta functions as given by (6). The α_n's are given by (7). That is,

$$\alpha_n = \frac{1}{X} \sum_{i=0}^{s} e^{-jn\omega x_i}.$$

Integrating m times gives

$$g(x) = \sum_{n=-\infty}^{\infty}{}' \frac{\alpha_n}{(jn\omega)^m} e^{jn\omega x} + K \tag{9}$$

$$= \sum_{n=-\infty}^{\infty} \alpha_{n_g} e^{jn\omega x}. \tag{10}$$

In conclusion, to find the α_n's of a function $g(x)$, differentiate it until the mth derivative appears as a sum of delta functions. Find the α_n's of

$$\frac{d^m[g(x)]}{dx^m}.$$

Divide by $(jn\omega)^m$ to get the α_n's of $g(x)$. Add the average value of $g(x)$.

MARVIN I. GANG
1681 49th St.
Brooklyn 4, N. Y.

High Performance Silicon Tetrode Transistors*

Since the grown-diffused technique was announced one year ago,[1] this technique has been further exploited. The result is a developmental silicon tetrode transistor which seriously rivals germanium transistors in high-frequency performance.

Fig. 1 illustrates the frequency response possible for the common emitter and common base short-circuit current gains. The alpha cutoff frequency is approximately 350 megacycles.

One prevalent misconception concerning tetrode transistors is that tetrode operation necessarily precludes a high alpha. Referring to Fig. 1, it is seen that this unit has a low-frequency common emitter current gain of 37 db. Since the common emitter cutoff frequency is about 5 megacycles, this unit has a very large gain bandwidth product.

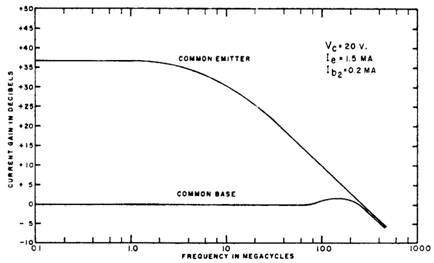

Fig. 1—Short-circuit current gain as a function of frequency in the common emitter and common base connections for a developmental grown-diffused silicon tetrode transistor.

The collector capacity for this type of unit is very low, approximately 0.6 mmf. The emitter and collector series bulk resistances are often thought to prohibit very high-frequency performance in a grown structure. However, with improved fabrication techniques these extrinsic elements may be reduced to quite low values. For this particular unit, the emitter series bulk resistance was about 0.5 ohm and the collector bulk resistance was 60 ohms. The high-frequency base resistance is 150 ohms at 150 megacycles.

Fig. 2 shows the power gain capabilities of this device. The curve was calculated assuming a common emitter conjugately matched input and ouptut stage with admittance neutralization. With moderate care, it is possible to realize within 2 or 3 db of this calculated power gain.

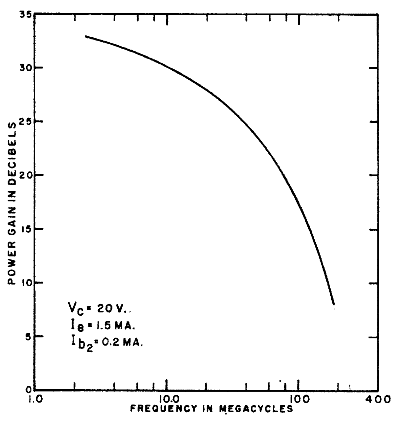

Fig. 2—Calculated common emitter power gain, assuming conjugate matched input and output and admittance neutralization.

The maximum collector to emitter voltage is 30 volts. The frequency response is relatively constant for collector voltages from 1.5 volts to 30 volts. The I_{co} has been a small fraction of a microampere as is usual in siilcon transistors.

The tetrode bias current is not critical. A value of 0.2 ma was used in this example. It should be pointed out that in many applications, such as common emitter circuits, adequate tetrode bias may be obtained by merely connecting the second base to the emitter.

Fig. 3 illustrates not only the high tetrode alpha obtained but also the constancy of alpha with emitter current. Alpha remains essentially constant from 0.2 ma to greater than 15-ma emitter current.

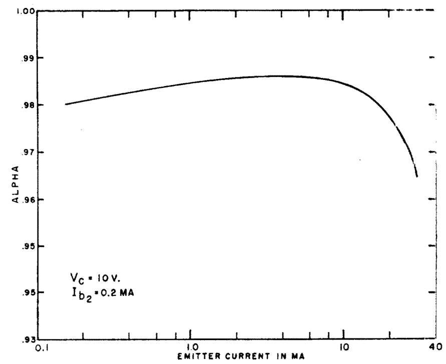

Fig. 3—Variation of tetrode alpha with emitter current.

In this effort, many individuals have made significant contributions. Members of the project group who contributed to the development and evaluation of these devices in addition to the author were W. C. Brower and Drs. R. E. Anderson, M. E. Jones, and W. R. Runyan.

RICHARD F. STEWART
Texas Instruments Inc.
Dallas, Texas

* Received by the IRE, April 1, 1957. This work was supported by the Signal Corps under Industrial Preparedness Study Contract No. DA-36-039-SC-72703.

[1] R. F. Stewart, B. Cornelison, and W. A. Adcock, "High-frequency tetrodes," 1956 IRE NATIONAL CONVENTION RECORD, Pt. 3, pp. 166–171.

Relation Between Ratio of Diffusion Lengths of Minority Carriers and Ratio of Conductivities*

There is a simple relationship between the diffusion length L_p of the holes in the n region, L_n of the electrons in the p region, and the conductivities σ_n and σ_p of the n and p regions respectively. Namely,

$$\left(\frac{L_p}{L_n}\right)^2 = \frac{\sigma_p}{\sigma_n}. \tag{1}$$

Eq. (1) is derived on the basis of two commonly-made assumptions of junction transistor theory.

1) The conductivities in both regions can be closely approximated by that of the majority carriers alone.
2) The presence of injected minority carriers does not increase the number of majority carriers substantially enough to have detectable effect on the diffusion lengths.

Eq. (1) can be readily derived from the following well-established relations:

$$L_p^2 = D_p \tau_p = \frac{kT}{q} \mu_p \tau_p \tag{2}$$

$$L_n^2 = D_n \tau_n = \frac{kT}{q} \mu_n \tau_n \tag{3}$$

$$\tau_p = \frac{1}{rn_n} \tag{4}$$

$$\tau_n = \frac{1}{rp_p} \tag{5}$$

$$\sigma_p \approx q\mu_p p_p \tag{6}$$

$$\sigma_n \approx q\mu_n n_n. \tag{7}$$

* Received by the IRE, April 3, 1957.

In the above equations, τ_p and τ_n are lifetimes of holes and electrons as minority carriers; D_p, D_n and μ_p, μ_n are diffusion constants and mobilities of the holes and electrons, respectively. N_n and P_p are densities of majority carriers, q is the electronic charge and r is a constant of recombination. From (2) and (3),

$$\frac{L_p^2}{L_n^2} = \frac{\mu_p \tau_p}{\mu_n \tau_n}. \quad (8)$$

From (4) and (5),

$$\frac{\tau_p}{\tau_n} = \frac{p_p}{n_n}. \quad (9)$$

From (6) and (7),

$$\frac{\sigma_p}{\sigma_n} = \frac{\mu_p p_p}{\mu_n n_n}. \quad (10)$$

Eq. (1) follows from (8), (9), and (10).

As there are usually more than a single p region and a single n region, it is desirable to put (1) into an alternative form,

$$\sigma_j L_j^2 = \frac{kT}{r} \mu_p \mu_n \quad (11)$$

where σ_j is the conductivity of the majority carrier in the j-th region, and L_j is the diffusion length of the minority carrier in the j-th region. The right-hand side of (11) is a constant throughout the entire crystal, independent of the local densities of donors and acceptors. Eq. (11) can be easily derived for an n region from (2), (4), and (7), and for a p region from (3), (5), and (6).

SHELDON S. L. CHANG
Dept. of Elec. Eng.
New York University
New York, N. Y.

The Measurement and Specification of Nonlinear Amplitude Response Characteristics in Television*

It is gratifying to find Doba, in the above paper,[1] placing emphasis on linearity as such rather than on its effect—distortion. The widespread practice of specifying linearity in terms of the distortion which it produces has persisted too long.

I would, however, like to suggest an alternate procedure for plotting the linearity of an amplifier which is more direct and also simpler in application than the method of superposition of two signals. The term linearity refers specifically to an input-output characteristic and the definition of differential gain corresponds, in essence, to the slope of this curve at a particular point referred to some other point on the curve as a reference. If a ramp function is used for the input voltage waveform, the output waveform as plotted against a linear time base is the input-output characteristic. By means of a simple RC differentiating network, it is possible to differentiate the output waveform with respect to the input (since $e_{in} = kt$) and thus obtain a plot of the slope as a function of signal amplitude. The definition "differential gain" corresponds very closely to the "figure of demerit" which I proposed[2] and is defined in Fig. 1.

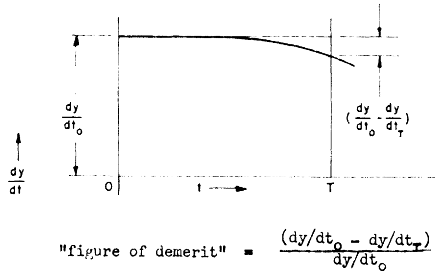

Fig. 1.

This procedure for plotting linearity requires fewer components, less bandwidth, and by use of a repetitive waveform of low-duty cycle permits measurements at power levels beyond the continuous duty ratings of the components involved.

STANLEY I. KRAMER
Guided Missiles Div.
Fairchild Engine and Airplane Corp.
Wyandanch, N. Y.

* Received by the IRE, March 29, 1957.
[1] S. Doba, Jr., PROC. IRE, vol. 45, pp. 161–165; February, 1957.
[2] S. I. Kramer, "A sensitive method for the measurement of amplitude linearity," PROC. IRE, vol. 44, pp. 1059–1060; August, 1956.

Author's Comment[3]

Although I am not in disagreement with Mr. Kramer's method of measuring and defining the "figure of demerit," I would like to point out that this is not quite the same as "differential gain."

The methods proposed by Mr. Kramer are correct for devices which have no time-storage effects. By this, I mean that they are valid when applied to a dc amplifier if the ramp function used as the input waveform changes at a "suitably" slow rate. As is obvious, in an ac amplifier there may be difficulty in making the ramp function too slow. With a finite bandwidth, there also may be difficulties in making it too fast. Hence, the determination of the meaning of the word "suitable" above may be quite a stumbling block, as was indicated by Mr. Kramer.

A more serious difficulty arises when we consider systems in general (such as feedback amplifiers), where the nonlinearity is a function of frequency. Here the distortion is affected not because of bandwidth limitations, but because a rapidly varying wave may be distorted differently from a slowly varying wave.

It is important then that the definitions and methods of measurement be stated clearly and without ambiguity. I believe our terms "differential gain" and "differential phase" are successful in meeting this objective. For the simple dc amplifier "differential gain" may be closely correlated with "figure of demerit." For more complex systems where it may be necessary to specify the "differential gain" as a function of frequency, this can be done without violence either to the definition or to the method of measurement.

STEPHEN DOBA, JR.
Bell Telephone Labs.
Murray Hill, N. J.

[3] Received by the IRE, April 11, 1957.

Automatic Dictionaries for Machine Translation*

Mechanical translation is one of the newer objectives of the communications engineer. Until now his concern has been chiefly with single-language messages. As communication has become more widespread it is more and more necessary to perform rapid multilingual transformations. An increasing number of communications engineers are concerned with the problems with which this note is concerned, and for which a partial solution is outlined.

In the field of machine translation, automatic dictionaries are generally recognized as the first requirement even by those who hope to go beyond word-for-word translation to sentence-for-sentence translation: "If translations better than word-for-word are wanted, work on an automatic dictionary should still be undertaken because any machine that translates will need a dictionary."[1]

At the present time, no translation machine, not even word-for-word, has been constructed which has a usefully large vocabulary.[2] The main problems for automatic translation machines are: 1) overcoming the many-one and one-many word relations which arise in translating words with several meanings; 2) reducing memory-search time; and 3) reducing input and output or print-out time. This note is addressed mainly to 2), although the word-for-word translation is evidently also a first approximation to a solution to 1).

As a specific kind of storage and retrieval device (a "pure memory" look-up system, rather than a computing system), an automatic dictionary is subject to the same logical considerations as storage and retrieval systems in general. Therefore an automatic dictionary involves two main processes: input of data, and search of stored data. Discussion of input is omitted here except for the observation that the information which can be gotten out of any storage and retrieval system can be no better than the information and organization put into it.

Now an automatic dictionary must search its word-memory either sequentially or by parallel (simultaneous) search. Obviously, parallel search is to be preferred for reduction of search time. Unfortunately, up to the present, the principal large-scale mem-

* Received by the IRE, April 11, 1957.
[1] V. H. Yngve, "The technical feasibility of translating languages by machine," *Elec. Eng.*, vol. 75, pp. 994–999 (p. 994); November, 1956.
[2] R. E. O'Dette, "Russian translation," *Science*, vol. 125, pp. 579–585 (p. 584); March 29 1957.

ories are tapes or their analogs, and cards, all of which require sequential search.

We have found that rapid, accurate parallel scanning of a large word-collection can be achieved by scanning, not the collection directly, but a class or classes of which the word sought is a member. We have constructed several systems for isolating class members, in the storage and retrieval of large collections of data.[3,4] They have the merit that access to an item can be had through a number of starting points, so that the devices are almost certain to deliver a word if it is in the system.

The operating principle is the construction of data storage classes from which it is possible to form logical products. Such products have continually decreasing membership as the number of products is increased: if a, b, c are the members of class X, and b, c, d are the members of class Y, then only b, c are the members of the class (X and Y). Further, if c, d, e are the members of class Z, then only c is a member of class [(X and Y) and Z]. If a device is set up which forms the logical product or "intersects" (in the form of an "and"-gate) classes X, Y, and Z, the intersection is simultaneous. The search is parallel search. In the case of a dictionary the final membership of the logical product is comprised of a word or words. The word-for-word dictionary restricts the output membership to a single word, but in the more general cases which may in time become feasible the dictionary could print out the entire set of translated words required to satisfy the one-many relationship.

Now there is a particularly easy way to locate words as members of classes, which is based upon the concept that the letters of words determine classes. If a letter of the alphabet is the name of a class, then a word is a member of the class formed by the logical intersection of its letter-classes. This is true for any language which has been alphabetized. In word-for-word translation, the input word is the intersection of the letter-classes which form it; and the output word is the intersection of other letter classes. However, it is not necessary to use this device for both input and output. Only the letters of the input word are needed, and of course only these are known initially. The output word can be considered as a single element. That is, in one-way translation the letters of the input terms are interpreted as the names of classes; the output terms as the members of these classes.

Suppose a French-to-English dictionary is to be used. The desired outputs will be English words, the input will be letters of the French words which are to be translated. Each letter of the French input will define a class which has as members a set of English words. The class distinguishes between the position of letters. Thus, the letter C in the French word, *chateau*, will be the class C_1, whereas the C's in *accomplir* will be C_2 and C_3; and the H in *chateau* will be H_2, etc. Consider now the normal procedure of looking up a word in a printed French-English dictionary, *e.g.*, the word *chateau*.

1) First look for the letter C (that is, for the French words beginning with C). This isolates from the total number of English words in the dictionary a partial set consisting of such members as: basket (cabas); coffee (café); hood (capuchon); ticket (carte); warm (chaud); cat (chat); hat (chapeau); castle (chateau); . . . swan (cygne); etc.
2) Next look for CH. This is equivalent to the class $(C_1 \cap H_2)$ and has as typical members: warm (chaud); hat (chapeau); and castle (chateau).
3) If the search is continued until it reaches the class $[((((((C_1 \cap H_2) \cap A_3) \cap T_4) \cap E_5) \cap A_6) \cap U_7]$, the output will be "castle," the single member of this class.

The above is a strictly accurate description only for a single word output (one-one) type dictionary such as would be used in an elementary word-for-word device. In general, in order to convert a printed dictionary to an automatic one, it is only necessary to construct a system of product circuits in which the selection of a set of switches (input letters) will select a single path through the system (the output word). In this case the system can be arranged as many-one or as one-one, depending on the ratio of input to output words. If the system is arranged as one-many, the reader of the translation has to do part of the work of final selection, since he has the different meanings of the input word presented to him. For example, the French noun *revers* might be printed out: (back, wrong side, back-stroke, counterpart, reverse, facing). As mentioned above, several devices using the class intersection principle have already been constructed for general storage and retrieval problems, and additional machines are under design.[3,4] These machines ordinarily have multiple outputs and multiple inputs as well. Thus modification to automatic dictionaries is a simple matter of substituting different meanings for the variables, namely, the input and output signals, by redefining what are regarded as classes and as members.

After basic design of the product-making automatic dictionary, certain interesting problems of efficiency remain. For instance, the sequential-scanning automatic dictionary may be operated on the assumption that the root of a word can be recognized by stripping away letters one by one, starting at the end of a word (that is, from right to left, in English). For example, the memory can be greatly reduced in size by storing only the root-stock *judg*, instead of *judge*, *judging*, *judgment*, etc.; and storing the endings separately. A product-making dictionary will proceed from the beginning of the word (*i.e.*, from left to right, in English). In many cases the last letters need not even be employed. Thus, the classes defined by C_1, H_2, A_3, T_4, E_5, A_6, U_7 and by C_1, H_2, A_3, T_4, E_5, A_6 have the same member. U_7 is redundant and can be disregarded. An efficient machine will not utilize products longer than are required to provide unique outputs. In our experience seven-letter intersections cover most cases, and need for more than ten is rare. And, of course, word-endings do not require separate storage.

The suggested solution for the search problem of automatic dictionaries has been directed mainly to word-for-word translation, on the assumption that this will come first. The principle should apply also in the case of dictionaries with many-one and one-many input-output relations. If such complexity can indeed be handled by a reasonably-sized memory, then the letter-class "and"-gate method appears to be the most efficient method of simultaneous search. The operating principle is that of the actual use of a dictionary: mere typing out of the input word forms the logical product. Not only will this save search time and some extra storage space, but the input typewriter itself can be a high-speed printer. The time for translation then will depend almost entirely on the output "readout" device.

M. Taube
and L. B. Heilprin
Documentation, Inc.
Washington, D. C

[3] E. Miller, "Final Report to the National Science Foundation on the MATREX Indexing Machine," Documentation, Inc., Washington, D. C.; January, 1957. See also, "The Prototype of the Mechanical Alpha-MATREX Indexing Machine," in "Studies in Coordinate Indexing," Documentation, Inc., vol. 4; June, 1957.

[4] M. Taube, *et al.*, "The Logic and Mechanics of Storage and Retrieval Systems," Tech. Rep. No. 14, prepared under Contract Nonr-1305(00) for the Office of Naval Res.; February, 1956. See also "Studies in Coordinate Indexing," Documentation, Inc., Washington, D. C., vol. 3, pp. 58–100; 1956.

On a Property of Wiener Filters*

The theory of optimum filters as originated by Wiener[1] and extended by many others[2] is invariably based upon a minimization of mean-squared error. A property of Wiener filters acting on stationary Gaussian inputs, which seems to have escaped general notice, is that they also minimize any function of the error of the form

$$f(|\epsilon|) = \sum_n \overline{|\epsilon(t)|^n},$$

where n is positive (but not necessarily an integer). This may be proved as follows.

If the input $\{x(t)\}$ to a linear time-invariant system $h(t)$ is Gaussian, the output $\{y(t)\}$ and hence the error

$$\{\epsilon(t)\} = \{y(t)\} - \{x(t+\alpha)\}$$

will also be Gaussian. Then

$$\overline{|\epsilon(t)|^n} = \frac{2}{\sqrt{2\pi\overline{\epsilon^2}}} \int_0^\infty \epsilon^n e^{-\epsilon^2/2\overline{\epsilon^2}} d\epsilon$$
$$= \frac{2^{n/2}}{\sqrt{\pi}} \Gamma\left(\frac{n+1}{2}\right) (\overline{\epsilon^2})^{n/2},$$

where

$$\overline{\epsilon^2} = \phi_{xx}(0) - 2\int_0^\infty h(\tau)\phi_{xx}(\tau+\alpha)d\tau$$
$$+ \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2)\phi_{xx}(\tau_1 - \tau_2)d\tau_2 d\tau_1$$

* Received by the IRE, April 22, 1957.

[1] N. Wiener, "Extrapolation, Interpolation and Smoothing of Stationary Time Series," John Wiley and Sons, New York, N. Y.; 1949.

[2] J. Bendat, "A general theory of linear prediction and filtering," *J. Soc. Ind. Appl. Math.*, vol. 4, pp. 131–151; September, 1956.

and $\phi_{xx}(\tau)$ is the autocorrelation of the input $\{x(t)\}$. The minimization of $f(|\epsilon|)$ by the usual calculus of variations technique of varying $h(\tau)$ to $h(\tau)+ak(\tau)$, differentiating with respect to a, letting $a=0$, and subsequently setting this derivative equal to 0 gives

$$\left[\sum_n \frac{2^{n/2}n}{\sqrt{\pi}}\Gamma\left(\frac{n+1}{2}\right)(\bar{\epsilon^2})^{n/2-1}\right]$$
$$\cdot\left[-\int_0^\infty k(\tau)\phi_{xx}(\tau+\alpha)d\tau + \int_0^\infty k(\tau_1)\int_0^\infty h(\tau_2)\phi_{xx}(\tau_1-\tau_2)d\tau_2 d\tau_1\right]=0$$

where $h(\tau)$ is the optimum filter impulse response. The quantity within the first brackets cannot vanish since $\bar{\epsilon^2}>0$; hence a necessary condition for a minimum is that the second factor must vanish. This yields the requirement

$$\phi_{xx}(\tau+\alpha)=\int_0^\infty h(\sigma)\phi_{xx}(\tau-\sigma)d\sigma$$

which may be recognized as the Wiener-Hopf equation arising in the classical analysis based on the usual mean-square error criterion. The sufficiency of this condition may be proved in the usual manner.

T. R. Benedict
and M. M. Sondhi
Dept. of Elec. Eng.
University of Wisconsin
Madison, Wis.

A Simplified Procedure for Finding Fourier Coefficients

*Letter from Mr. Brenner**

In the above article, Gibbons[1] pointed out that the Fourier series of periodic waveforms can be deduced with ease if successive differentiation of the waveform results in a simpler waveform, impulses being particularly desirable. The same technique can be applied to the problem of formulating the Fourier transform of pulse-type waveforms. If a pulse-type waveform is defined as a signal which lasts for a finite time and is Fourier transformable, *i.e.*,

$$f(t)\equiv 0, |t| \geq a, a \text{ positive}, \quad (1)$$

then the (complex) Fourier transformation has the form:

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}F(j\omega)\exp(j\omega t)d\omega \quad (2a)$$

where

$$F(j\omega)=\mathfrak{F}f(t)=\int_{-a}^{+a}f(t)\exp(-j\omega t)dt. \quad (2b)$$

From (2b) it follows that

$$i\omega F(j\omega)+f(a)\exp(-j\omega a) + f(-a)\exp(j\omega a)=\mathfrak{F}df/dt.$$

* Received by the IRE, March 22, 1957.
[1] J. F. Gibbons, Proc. IRE, vol. 45, p. 243; February, 1957.

Since a has been chosen so that $f(a)=f(-a)=0$, one may write

$$\mathfrak{F}f(t)=1/(j\omega)\mathfrak{F}df/dt. \quad (3)$$

In addition to (3), the relationship for the Fourier transform of the unit impulse

$$\mathfrak{F}\delta(t-b)=\exp(-j\omega b) \quad (4)$$

is sufficient to find the Fourier transform of many waveforms. (The same scheme is applied to the Laplace transform if the waveform is shifted so that $f(t)=0$ for $t\leqq 0$.)

Examples

1) Find the Fourier transform of the rectangular pulse shown in Fig. 1(a). Solution: Since

$$f(t)=U(t)-U(t-T),$$
$$df/dt=\delta(t)-\delta(t-T)$$

as shown in Fig. 1(b). Hence

$$i\omega F(j\omega)=1-\exp(-j\omega T)$$

and

$$F(j\omega)=[1-\exp(-j\omega T)]/j\omega.$$

2) Find the Fourier transform of the pulse

$$f(t)=(V/T)tU(t)-(V/T)(t-T)U(t-T) - VU(t-T)$$

shown in Fig. 2(a). The derivative df/dt, shown in Fig. 2(b) has the form

$$df/dt=(V/T)U(t)-(V/T)U(t-T) - V\delta(t-T);$$

hence

$$\frac{d}{dt}[df/dt-V(t-T)] = (V/T)\delta(t)-\delta(t-T)$$

which corresponds to Fig. 1(b) (with a scale change). Hence

$$j\omega\mathfrak{F}[df/dt-V\delta(t-T)] = (V/T)[1-\exp(-j\omega T)]$$

and

$$\mathfrak{F}df/dt=(V/j\omega T)[1-\exp(-j\omega T)]] - V\exp(-j\omega T)$$

so that

$$\mathfrak{F}f(t)=F(j\omega)=\frac{V}{T}\cdot\frac{1-\exp(-j\omega T)-j\omega T\exp(-j\omega T)}{(j\omega)^2}.$$

3) Find the Fourier transform of the sine pulse shown in Fig. 3(a):

$$f(t)=\sin(\omega_0 t)U(t) + \sin\omega_0(t-\pi/\omega_0)U(t-\pi/\omega_0).$$

For this pulse, the derivative [see Fig. 3(b)] is

$$df/dt=\omega_0\cos(\omega_0 t)U(t) + \omega_0\cos\omega_0(t-\pi/\omega_0)U(t-\pi/\omega_0)$$

and the second derivative [see Fig. 3(c)] is

$$d^2f/dt^2=-\omega_0^2\sin(\omega_0 t)U(t) - \omega_0^2\sin[\omega_0(t-\pi/\omega_0)U(t-\pi/\omega_0)] + \omega_0[\delta(t)+\delta(t-\pi/\omega_0)].$$

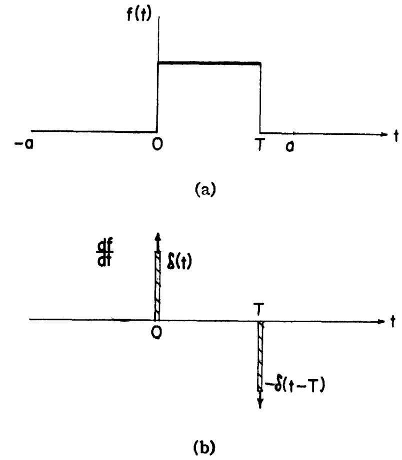

Fig. 1.

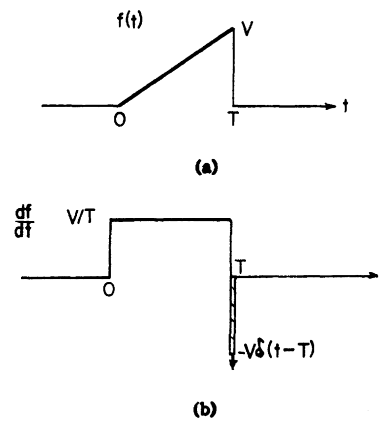

Fig. 2.

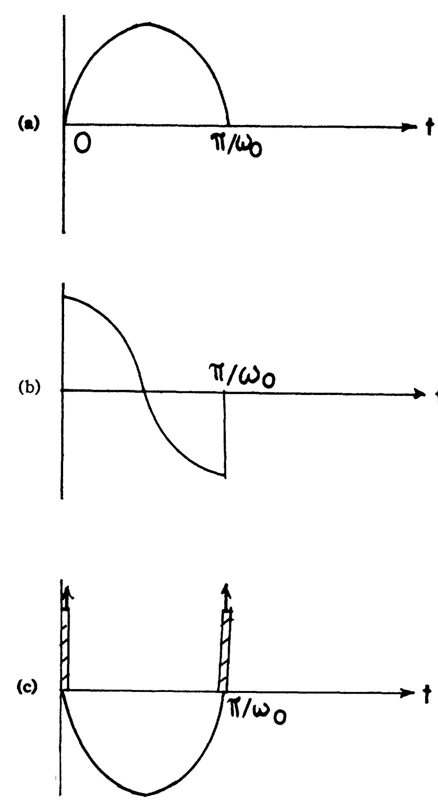

Fig. 3.

Hence

$$\mathfrak{F}[\omega_0^2 f(t) + d^2f/dt^2] = \omega_0[1 + \exp(-j\omega\pi/\omega_0)],$$

so that

$$F(j\omega) = \frac{\omega_0[1 + \exp(-j\omega\pi/\omega_0)]}{\omega_0^2 + (j\omega)^2}.$$

EGON BRENNER
The City College of New York
New York 31, N. Y.

Letter from Dr. Fatehchand[2]

The simplified procedure for finding Fourier coefficients given by Gibbons[1] appears a most useful one. It may easily be extended to the transient case, when it is required to determine the frequency spectrum $g(\omega)$ of a time function $f(t)$. Here the Fourier integral relationships are necessary:

$$\left.\begin{aligned} g(\omega) &= \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt \\ f(t) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} g(\omega)e^{i\omega t}d\omega \end{aligned}\right\}. \quad (1)$$

If $f^n(t)$ is the nth time derivative of $f(t)$ differentiating within the integral sign gives

$$\left.\begin{aligned} f^n(t) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} g(\omega)(i\omega)^n e^{i\omega t}d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} g_n(\omega)e^{i\omega t}d\omega \end{aligned}\right\}. \quad (2)$$

Therefore, if $g_n(\omega)$ is the frequency spectrum of $f^n(t)$, the frequency spectrum of $f(t)$ is given by

$$g_0(\omega) = g(\omega) = \frac{g_n(\omega)}{(i\omega)^n}. \quad (3)$$

We also have the result that if $\delta(t-T)$ is a δ function centered on $t=T$, and $h(t)$ is a function continuous in the neighborhood of $t=T$,

$$\int_{-\infty}^{+\infty} h(t)\delta(t-T)dt = h(T) \quad (4)$$

Hence the frequency spectrum of $\delta(t-T)$ is given by

$$\int_{-\infty}^{+\infty} e^{-i\omega t}\delta(t-T)dt = e^{-i\omega t}. \quad (5)$$

Thus, for example, if

$$f^n(t) = \delta(T-t), \quad g_n(\omega) = e^{-i\omega t}$$

and

$$g(\omega) = \frac{e^{-i\omega T}}{(i\omega)^n}.$$

As an illustration of the simplified method for deriving the frequency spectrum, consider a triangular pulse centered at $t=0$ (Fig. 1). The time function is differentiated until δ functions appear. In this case two differentiations are necessary (Figs. 2, 3). Then by (5):

$$g_2(\omega) = \frac{2}{T}e^{i\omega T/2} - \frac{4}{T} + \frac{2}{T}e^{-i\omega T/2}$$

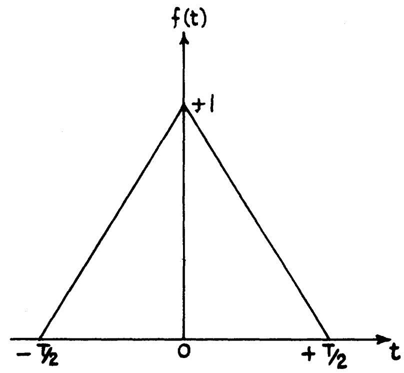

Fig. 1.

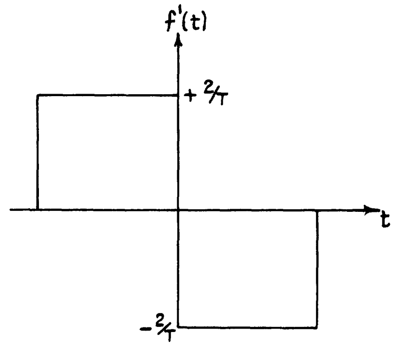

Fig. 2.

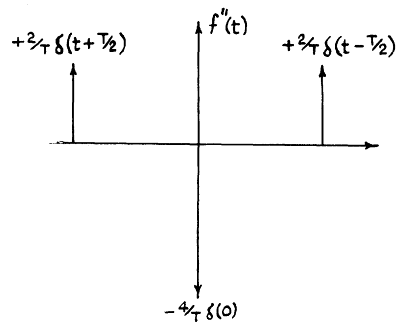

Fig. 3.

and

$$g(\omega) = \frac{4}{T}\cdot\frac{\cos\omega T/2 - 1}{(i\omega)^2}.$$

Here, the δ functions all appear at the same stage, but as Gibbons shows, this is not necessary for the solution. The method may obviously also be applied to the inverse problem, *i.e.*, to obtain the time function which corresponds to a given frequency spectrum.

RICHARD FATEHCHAND
British Broadcasting Corp.
London, England

Letter from Mr. Klotter[3]

Mr. Gibbons[1] has drawn attention to a very valuable tool for finding the Fourier coefficients of functions which exhibit discontinuities in the function itself or in any of its derivatives. Because his note does not quote any literature on the subject, it may be worth while to point out that the method has both a history and considerable ramifications, and to give some pertinent references.

After some early publications by Lalesco[4] and Abason[5] which failed to receive much attention, the first (and still best) presentation in English of the method for obtaining the Fourier coefficients from the discontinuities of a function's derivatives was given by Eagle.[6] In spite of its usefulness and the very adequate presentation by Eagle, the method seems to have gone generally unnoticed in the English-speaking countries. There seems to exist only a single reference to it and application of it by Jolley.[7]

On the European continent, however, the method caught on, was used extensively, and was developed further. Propagandists for and developers of the method were, in addition to the Roumanian school,[4,5,8] mainly Walther and his fellow workers in Germany. The first paper[9] of that group essentially reports on Eagle's work. Some items in that paper led to a discussion with Feinberg.[10,11] Later Zech[12] investigated some refinements and seemingly paradoxical results. A paper by Popesco[8] treats essentially the same items as Zech's paper.

The idea of finding the coefficients of expansions from the discontinuities of the function is not restricted to expansions into Fourier series but can be extended to other expansions; this was shown in a paper by Walther and Brinkmann.[13] Here the idea is extended to series progressing according to any orthogonal system with particular attention to expansions into spherical harmonics. The paper also gives an extensive list of references dealing with the development and applications of the method.

A fairly recent (English language) presentation of the essential features of the method has been given by Janssen.[14] In this paper, the method, furthermore, is extended and linked to the Fourier integrals, and it is shown how the frequency even of continuous functions may be found by suitably applying

[2] Received by the IRE, April 4, 1957.

[3] Received by the IRE, April 23, 1957.

[4] T. Lalesco, "Sur les fonctions polygonales périodiques," *Rev. Gén. Électr.*, vol. 5, pp. 43–45; January, 1919.

[5] E. Abason, "Asupra determinării pe cale grafică a armonicelor unei functiuni periodice (On finding the harmonics of a periodic function by graphical means)," *Gaz. mat.*, vol. 26, pp. 81–85, 105–108; 1920.

[6] A. Eagle, "Fourier's Theorem and Harmonic Analysis," Longman, Green and Co., London, England; 1925.

[7] L. B. W. Jolley, "Alternating Current Rectification and Allied Problems," Chapman and Hall, London, England, 3rd ed.; 1928.

[8] A. T. Popesco, "Sur l'application de la méthode des discontinuités a l'analyse harmonique des fonctions sinusoidales," *Bull. Math. Phys.*, (Ecole Polytech, Bucharest), vol. 9, p. 83; 1939.

[9] G. Koehler and A. Walther, "Fouriersche Analyse von Funktionen mit Sprüngen Ecken und ähnlichen Besonderheiten," *Arch. Elektrotech.*, vol. 25, pp. 747–758; October, 1931.

[10] R. Feinberg, "Bemerkung zu der Arbeit von G. Koehler und A. Walther über die Fouriersche Analyse von Funktionen mit Sprüngen, Ecken und ähnlichen Besonderheiten," *Arch. Elektrotech.*, vol. 27, pp. 15–19; January, 1933.

[11] A. Walther, "Stellungnahme zu der Bemerkung von Herrn Feinberg und geschichtliche Ergänzung zur Fourierschen Analyse von Funktionen mit Sprüngen, Ecken und ähnlichen Besonderheiten," *Arch. Elektrotech.*, vol. 27, pp. 19–20; January, 1933.

[12] T. Zech, "Über das Sprungstellenverfahren zur harmonischen Analyse," *Arch. Elektrotech.*, vol. 36, pp. 322–328; May, 1942.

[13] A. Walther and K. Brinkmann, "Zum sprungstellen-verfahren, insbesondere für die entwicklung nach kugelfunktionen," *Ing.-Arch.*, vol. 13, pp. 1–8; January, 1942.

[14] J. M. L. Janssen, "The method of discontinuities in Fourier analysis,' *Phillips Res. Rep.*, vol. 5, pp. 435–460; December, 1950.

the ideas of the "Method of Discontinuities."

The writer wishes to close with an appeal to the textbook authors in this country to take notice of this extremely useful method; especially because the ideas underlying the method are intimately related to intrinsic properties of the expansions which are treated in the textbooks anyway.

KARL KLOTTER
Stanford University
Stanford, Calif.

Unit-Distance Number-Representation Systems, a Generalization of the Gray Code*

A digit-distance function is any distance function, ρ, satisfying the metric space axioms, and defined on a finite set of digits. For binary digits, $\rho=0$ when the digits are the same and 1 when they are different. If d_1 and d_2 denote the numerical value of arbitrary digits, then for any base, β, one distance function is

$$\rho_1(d_1, d_2) = |d_1 - d_2|. \qquad (1)$$

Another one is

$$\rho_2(d_1, d_2) = \min(|d_1 - d_2|, |d_1 - d_2 + \beta|, |d_2 - d_1 + \beta|). \qquad (2)$$

If $\beta=2$, (1) and (2) reduce to the same function. This is false for $\beta>2$, since the maximum value of ρ in (1) is $\beta-1$, whereas in (2) it is the largest integer not greater than $\beta/2$. Digit-distance function (1) corresponds to placing the digits, in natural order, opposite the teeth of a rack, and measuring the distance between digits along the pitch line of the rack. Function (2) corresponds to placing the digits in natural order, opposite the teeth of a gear, and measuring the distance along the pitch circle of the gear.

Given a digit-distance function, we define an expression-distance function (syntactical distance) between two digit strings, e_1 and e_2, of length n, as the sum of the n digit-distance functions for corresponding digit pairs, d_{1i}, d_{2i}.

$$\rho_j^{(n)}(e_1, e_2) = \sum_{i=n}^{n} \rho_j(d_{1i}, d_{2i}) \qquad (3)$$

where $j=1$ or $j=2$, or j may represent some other digit-distance function. This corresponds to metrizing the Cartesian product of the n individual metric spaces. It is known that a function of a finite number of factor metrics which vanishes if and only if all the factor metrics vanish, and which is subadditive and suitably isotone, is a product metric. The above is perhaps the simplest function having these properties.

We now assign some (or all) of these β^n-distinct n-digit strings as names of integers. At most, there will be β^n integers, and if all strings are used, there will be exactly that many. For example, let $n=2$, $\beta=3$, and the integers be 0 to 8 inclusive:

integer, i	string of n digits $e=\widehat{d_1d_0}$
0	'00'
1	'01'
2	'02'
3	'12'
4	'10'
5	'11'
6	'21'
7	'22'
8	'20'.

The integers have a natural distance function:

$$p(i_1, i_2) = |i_1 - i_2|.$$

This integer distance function also induces a (semantical) expression-distance function

$$P(e_1, e_2) = p(i_1, i_2) = |i_1 - i_2|, \qquad (4)$$

where i_j is the integer whose name (string of n digits) is e_j, $j=1, 2$.

A unit-distance code is one for which $P(e_1, e_2)=1 \Longrightarrow \rho(e_1, e_2)=1$. The arrow is unilateral except in the degenerate case $\beta=2$.

Thus the names of two consecutive integers are at unit (syntactical) distance from each other, which for digit-distance functions (1) and (2) means that they differ only in a single digit and that those single digits are at unit digit distance from each other.

The Gray code ($\beta=2$) is the most familiar example of a unit-distance code. The writer, in a paper delivered orally at a meeting of the Association for Computing Machinery in 1951, gave examples of codes for $\beta>2$ based on digit-distance functions (1) and (2), calling the first type "reflecting" and the second type "progressive" and "retrogressive."

This terminology is based on the fact that in reflecting codes the digit cycle runs through the digits from smallest to largest, and then reflects and runs back down again, yielding a digit cycle of length twice the base. In the progressive code, the digit cycle is β times the base, and the digits increase to maximum, then dwell on the largest digit while another column changes, then they start with 0 again and progress (increase) until the next-to-the-largest digit, where another dwell occurs. The dwell precesses down to zero, at which point the digit cycle is complete. This system might be more natural (than the reflecting system) to mechanize on physical code wheels, since the wheels would always rotate in the same direction. A brief example is given above for $n=2$, $\beta=3$. One complete cycle of length 9 is shown for the least significant digit (d_0). The next digit (d_1) has a period of 27.

The retrogressive system is similar, except the digit cycle runs in the opposite direction and the dwell precesses through successively larger digits toward zero. It should be emphasized that all these codes reduce to the Gray code for $\beta=2$.

Since the generalized reflected code has been already discussed,[1] it does not seem necessary here to give any conversion rules for this code. The following conversion rules are for the progressive and retrogressive codes. Let d_i, Δ_i and δ_i be the ith digits of the name of a number, expressed in the progressive, retrogressive and ordinary number system, respectively, all to the base β. The 0th place or column contains the least significant, or units, digit.

Then

$$d_i = \delta_i \ominus_\beta \delta_{i+1},$$
$$\Delta_i = \delta_{i+1} \ominus_\beta \delta_i,$$

where $\ominus_\beta$ represents subtraction modulo β. In order to convert into normal code, we use the following equations:

$$\delta_i = \sum_{l=i}^{k}{}_{(\beta)}\, d_l,$$

$$\delta_i = \sum_{l=i}^{k}{}_{(\beta)}\, (-\Delta_l),$$

where k is any digit position such that all digits further to the left would be zero (if n were larger), and the addition is modulo β. The subscript (β) is thus a homomorphism operator mapping the sum onto the residue-class ring representatives 0 to $\beta-1$.

GEORGE W. PATTERSON
The Moore School of Elec. Eng.
University of Pennsylvania
Philadelphia, Pa.

* Received by the IRE, April 11, 1957. This communication was prepared under the sponsorship of the Dept. of the Army, U. S. Army Signal Corps. Eng. Labs., Ft. Monmouth, N. J., Contract No. DA36-039 SC-72344.

[1] For example, I. Flores, "Reflected number systems," IRE TRANS., vol. EC-5, pp. 79–82; June, 1956.

On the Order of the Differential Equation Describing an Electrical Network*

The question of determining the order of the differential equation describing an electrical network without writing down and expanding the network determinant is quite basic and must have been answered correctly by many, even in the early years of network analysis. Yet, a search through literature failed to provide the writer with a full answer to the problem. Hence this note, in which the results of an investigation are presented, rather than the complete discussion.

The order of the differential equation describing an electrical network is equal to the number of energy-storing elements, *i.e.*, inductances and capacitances, less the number of certain independent algebraic equations which can be written for the network. One type of the equations relates currents through inductances, a second relates currents through capacitances, a third relates voltages across inductances, and a fourth relates voltages across capacitances. Such algebraic equations, when dealing with the currents, will be referred to as current-interdependence relations, and when dealing with the voltages, as voltage-interdependence relations.

Current-interdependence relations stem from Kirchhoff's first law, the law of currents, and arise when all the currents flowing between two parts of a network pass through

* Received by the IRE, March 20, 1957.

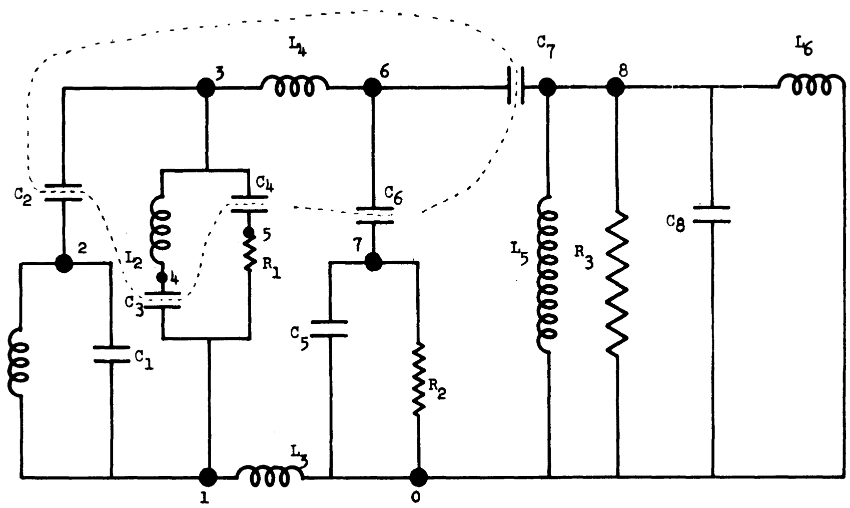

Fig. 1—The positive direction of currents and voltage-drops is chosen upwards and from left to right.

inductances (capacitances). In other words, a current-interdependence relation arises whenever a number of inductances (a number of capacitances) form a cut-set.[1] As a simple case of such an interdependence, consider two inductances (two capacitances) in series. For other examples, consider the network of Fig. 1. Examination of this network will show that the capacitances C_2, C_3, C_4, C_6, and C_7 form a cut-set, since they isolate nodes nos. 3, 4, and 6, from the rest of the network. This is indicated in Fig. 1 by a dotted line drawn through these capacitances. Thus, by Kirchhoff's law of currents, as applied to an isolated part of the network rather than to a single node, we have

$$I_{C_2} + I_{C_3} + I_{C_4} + I_{C_6} - I_{C_7} = 0. \qquad (1)$$

Further examination of the network will show that the inductances L_3 and L_4 form a cut-set, and therefore by Kirchhoff's law of currents

$$I_{L_3} + I_L = 0. \qquad (2)$$

We thus have two current-interdependence relations in the network.

Voltage-interdependence relations stem from Kirchhoff's second law, the law of voltages, and arise when a closed loop exists which consists of inductances only (capacitances only). As a simple case of such an interdependence relation, consider two inductances (two capacitances) in parallel. This is the case with the inductances L_5 and L_6 in Fig. 1. Thus, we have by Kirchhoff's law of voltages

$$V_{L_5} + V_{L_6} = 0. \qquad (3)$$

A capacitance loop is created in this network by capacitances C_5, C_6, C_7, and C_8. By Kirchhoff's law of voltages, we have

$$V_{C_5} + V_{C_6} + V_{C_7} - V_{C_8} = 0. \qquad (4)$$

Thus, there are two voltage-interdependence relations in the network.

The order of the differential equation describing the network of Fig. 1 will be 14 (the number of energy-storing elements), minus 4 (number of independent interdependence relations), or 10.

In most cases, as in the above example, the order of the differential equation can be obtained by rather simple inspection. For large and complicated networks, the question of counting all the independent interdependence relations is not a simple one. Care should be taken that dependent interdependence relations, *i.e.*, interdependence relations which are a linear combination of interdependence relations already counted, are not included. The procedure outlined herewith enables one to count quite easily the number of independent interdependence relations, even for large and complicated networks. The procedure applies to planar and nonplanar networks, including networks which cannot be mapped on a sphere.

Redraw the network, short-circuiting all the voltage sources and removing all the current sources. Disregard any mutual inductive coupling. Eliminate the ideal transformers by connecting the output from the secondary windings in parallel to the input of the primary. A network redrawn in this way will be referred to as the simplified network. Then follow these steps:

1) Count the number of the energy-storing elements.
2) Obtain the number of current-interdependence relations for capacitances by redrawing the simplified network, replacing any element other than a capacitance by a short-circuit. Count the total number of nodes.[2] If n_t is the total number of nodes, then $n = n_t - 1$ will be the number of independent node-pairs,[3] which is equal to the number of current-interdependence relations for capacitances in the original network.
3) Obtain the number of independent current-interdependence relations for inductances by following step 2) and reading inductance, inductances, for capacitance, capacitances.
4) Obtain the number of independent voltage-interdependence relations for capacitances by removing from the simplified network all elements other than capacitances.[4] Let b be the number of elements, n the number of nodes, and s the number of separate parts in this final network. The number of independent loops of this final network is given by $l = b - n + s$, the number of independent voltage-interdependence relations for capacitances of the original network is equal to l.
5) Obtain the number of independent voltage-interdependence relations for inductances by following step 4) and reading inductances for capacitances.
6) Subtract the numbers obtained in steps 2)–5) from the number obtained in step 1).

The order of the differential equation determines the number of the independent voltage and current initial conditions that we are free to specify, and which determine all the currents and voltages in the network for $t > 0$. Moreover, from a mathematical standpoint we can specify a number of "petrified" conditions which are superimposed on the solution and which remain constant unless forced to change by appropriate generators. These are dc currents circulating in loops of inductances, infinite frequency currents circulating in loops of capacitances, infinite frequency voltage-drops between parts of network separated by inductance cut-sets, and dc voltage-drops between parts of a network separated by capacitance cut-sets. Of these four cases, it seems that certain physical significance can be attached only to the last case.

JOSEPH OTTERMAN
Eng. Res. Inst.
University of Michigan
Ann Arbor, Mich.

[1] For a definition of a cut-set, see E. A. Guillemin, "Introductory Circuit Theory," John Wiley and Sons, New York, N. Y., p. 21; 1953.

[2] We define a node as a junction of two or more elements, or a junction of two ends of the same element, or an isolated end of a single element.

[3] The discussion here is limited to the case when the simplified network consists of a single separate part. When this is not the case, determine the order of the differential equations separately for each separate part and add the results.

[4] The removal from the resulting network of all the capacitances that do not form part of any loop will simplify the count of elements and nodes but does not affect the answer.