THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XL MAY 1961 NUMBER 3

Functional Design of a Voice-Switched Speakerphone

By W. F. CLEMENCY and W. D. GOODALE, JR.

(Manuscript received January 16, 1961)

A new hands-free telephone, known as the 3A Speakerphone System, is described. It provides, by means of switched-gain techniques, almost complete freedom from distant-end talker echo and singing. The gain-switching action is virtually free of clipping or blocking — objectionable side effects that are often introduced with voice control of gain. The switching threshold is varied automatically in accordance with room noise to avoid blocking in the receive channel. Performance characteristics are shown, with particular emphasis being given to the parameters chosen to meet rather stringent performance objectives.

I. INTRODUCTION

The original 1A Speakerphone System[1] used amplification in both its receive and transmit channels to compensate for the acoustic loss that was introduced by placing telephone instruments at greater distances from users than is normal with a handset. Among the operational difficulties[2,3] with such a system are talker echo and singing, particularly under reverberant room conditions.

To avoid the limitations that are inherent in the simultaneous use of high gain in the two transmission channels, the voice-switched 3A Speakerphone System has been designed. It changes the gain in each of the two channels in accordance with the direction of the stronger speech signal. Voice-switching techniques have been used before in communica-

tion equipment, and have been found to possess limitations of their own. However, by application of certain design principles,[4] the new voice-switched speakerphone substantially eliminates talker echo and singing, and greatly reduces objectionable side effects of the type often encountered with voice control. Moreover, these advantages are realized for a wide range of conditions. This paper is intended to provide a functional description of the 3A Speakerphone System and to show its performance characteristics, with particular emphasis on the parameters chosen to meet rather stringent performance objectives.

II. TRANSMISSION DESIGN PROBLEM

From a transmission standpoint, the fundamental difference between a speakerphone and a handset telephone is the distance between the instruments and the user. The loss introduced in the receive and transmit channels with normal arrangements anticipated for the 3A Speakerphone amounts to some 20 db in each direction. The increased gain provided in each channel to compensate for this loss, and the acoustic coupling between the loudspeaker and the microphone, introduce four operational difficulties: (a) sustained feedback or singing, (b) room echoes returned to the distant talker, (c) increased levels of transmitted reverberant energy, and (d) higher transmitted room noise. While voice switching is effective in essentially eliminating the first two, it provides no relief from the latter two difficulties.[3]

The two problems to which voice switching applies as a solution are illustrated in Fig. 1, which shows acoustic and electric levels for the two types of telephones, when they are producing equivalent transmission levels.* For the marginal incoming volume level and hybrid balance selected as an illustration, it will be noted that, when the speakerphone is transmitting, Fig. 1(b), the acoustic sidetone is about 24 db above the speech pressure at the microphone. When it is receiving, Fig. 1(c), the return echo to the telephone line is about 29 db above the incoming speech level. These conditions, of course, result in singing. If the gains were reduced to a point just below singing, the return of talker echo to the connected telephone might still be objectionable on low-loss connections.

The figure suggests that, if 30 db or more of gain were interchanged in the two channels in response to signal flow, the singing problem would be eliminated and talker echo would be tolerable for the selected conditions.

* It is assumed that the appropriate sound and electric levels are measured with meters having dynamic characteristics similar to the VUmeter and used in the approved manner.

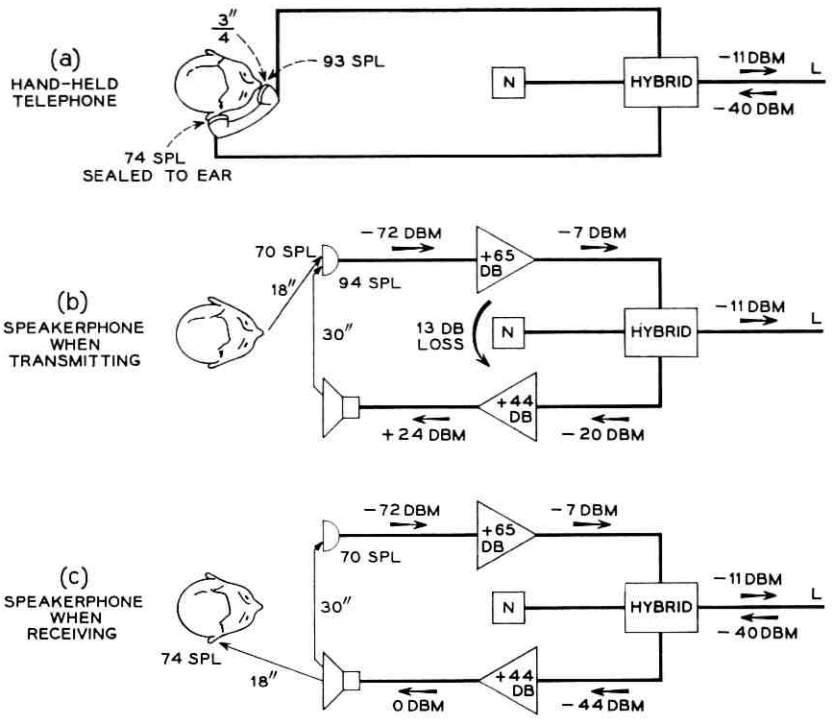


Fig. 1 — Comparison of transmission problem for handset and speakerphone

Actually, because the speakerphone's performance is more dependent on room acoustics and room noise than is that of a handset telephone, it is desirable to switch 35 to 40 db of gain between channels under certain conditions.

III. GENERAL DESCRIPTION

In the 3A Speakerphone, the interchange of gain between the receive and transmit channels is effected by control circuits operating on a linear differential basis; i.e., the channel having the stronger signal has the higher gain. To produce smooth gain changes without noticeable clipping of the speech syllables and without interference by the expected ranges of incoming line noise or room noise at the speakerphone, the control and the variable gain circuits are carefully designed on both a transient and a

steady-state basis. The time factors of speech, conversational habits, noise, and room reverberation must all be considered in determining the appropriate circuit characteristics.

In order to reduce the objectionable effects of the gain changes, the 3A Speakerphone switches only the amount of gain that is required for stability at the loudspeaker volume desired by the listener. This means that on low-loss connections, for which the volume control setting is low, a small amount of gain is switched; on higher-loss connections requiring a higher volume control setting, a greater amount of gain is switched.

Automatic variation of the switching threshold of the control circuit in accordance with room noise at the speakerphone, to avoid blocking the receive channel, eliminates the need for any adjustment by the installer or user. Also, by proper selection of the time constants and the use of the linear differential control feature, there is no need of any adjustment for almost all conditions of room reverberation.

Before describing the voice-switched gain circuit in detail, a broad over-all picture of the circuit and its operation is presented in the schematic of Fig. 2. The transmit channel consists of the microphone M, amplifiers A_M and A_T, and the transmit variolosser TVL. The receive channel consists of the loudspeaker S, amplifier A_R, and receive variolosser RVL. The two channels are coupled in the usual way with a hybrid coil to the telephone line. The balance network N of the hybrid incorporates current-sensitive variable elements which utilize the dependence of the loop direct current on the distance from the central office to compensate partially for different line impedances.

Variolossers TVL and RVL are variable-gain devices regulated in a complementary manner by the direct current flowing through them. This direct current is obtained from the manual volume control VOL, and from the combination of control amplifier A_C, rectifier R_C, and timing circuit T_C. The input to A_C is the microphone signal taken from the output of amplifier A_M and modified by the control variolosser CVL. An increase in the direct current through CVL, produced by two circuits, increases its loss. The "switchguard" circuit, consisting of amplifier A_G, rectifier R_G, and timing circuit T_G, produces a direct current which increases the loss of CVL in response to the voltage across the loudspeaker, and thus guards against false switching of the received signal due to the microphone voltage resulting from the loudspeaker output. Another direct current is produced by the NOGAD* from any nearly constant microphone voltage due to the noise at the speakerphone location.

---

* NOGAD = noise-operated gain-adjusting device.

The operation of the voice-switching circuits can be explained by considering several conditions:

First, the quiescent condition of no local speech or noise at the microphone and no incoming signal from the line. No direct current flows from the control rectifier R_C through TVL and RVL, and no direct current flows through CVL from R_N and R_G. Consequently, the gain of the receive channel is high and that of the transmit channel is low. The speakerphone is thus prepared for loudspeaker reproduction of a line signal of any level.

Second, only a line signal is present and is being reproduced by the loudspeaker. The switchguard, as a result of the loudspeaker voltage, increases the loss of CVL, and thus limits the input to A_C resulting from the acoustic coupling between the microphone and the loudspeaker, so that no current is produced through TVL and RVL. The receive channel remains in its high-gain state, allowing uninterrupted loudspeaker reception of the incoming signal.

Third, local speech at the microphone with no room noise and no incoming signal. No direct current from R_G and R_N flows through CVL; the output of A_M is coupled with minimum loss in CVL to A_C, causing direct current from R_C to flow through TVL and RVL. The gain of the transmit channel is high and that of the receive channel is low, so that speech at the microphone is readily transmitted to the line.

Fourth, the same conditions as for the third case with the addition of room noise at the microphone. The loss of CVL is increased by direct current from the NOGAD, so that the noise signal at the input of A_C produces only a small current through TVL and RVL. Local speech, however, does not cause an increase of the direct current from the NOGAD, because it is designed to give very little response to fluctuating signals. The increased loss of CVL, due to the noise at the microphone in this case, can be overcome, however, by the higher level of speech signal normally produced under noisy conditions, and sufficient speech signals can reach the input to A_C so that the resulting direct current from R_C through TVL and RVL clears the transmit channel and blocks the receive channel.

Fifth, a line signal is present and room noise exists at the microphone. Both the NOGAD and the switchguard produce direct current through CVL, and its increased loss prevents receive blocking because of noise or acoustic coupling between loudspeaker and microphone.

Sixth, an incoming speech signal is present and local speech exists at the microphone. Direct current is supplied to CVL from the switchguard, and a signal voltage exists at the output of A_M due to speech at the microphone. The relative signal strengths are evaluated in CVL: if the

line signal is great enough, the set will remain in receive; if the speech signal at the microphone is great enough, the set will transfer to the transmit state. Actually, because of the rapidly fluctuating levels characteristic of speech the set is continuously switching between the transmit and receive states, so that either party can quickly react to speech from the other party.

IV. TRANSMIT AND RECEIVE VARIOLOSSERS

The variolossers RVL and TVL in the receive and transmit channels act, in principle, like those used in the compandors of carrier systems.[5,6] In

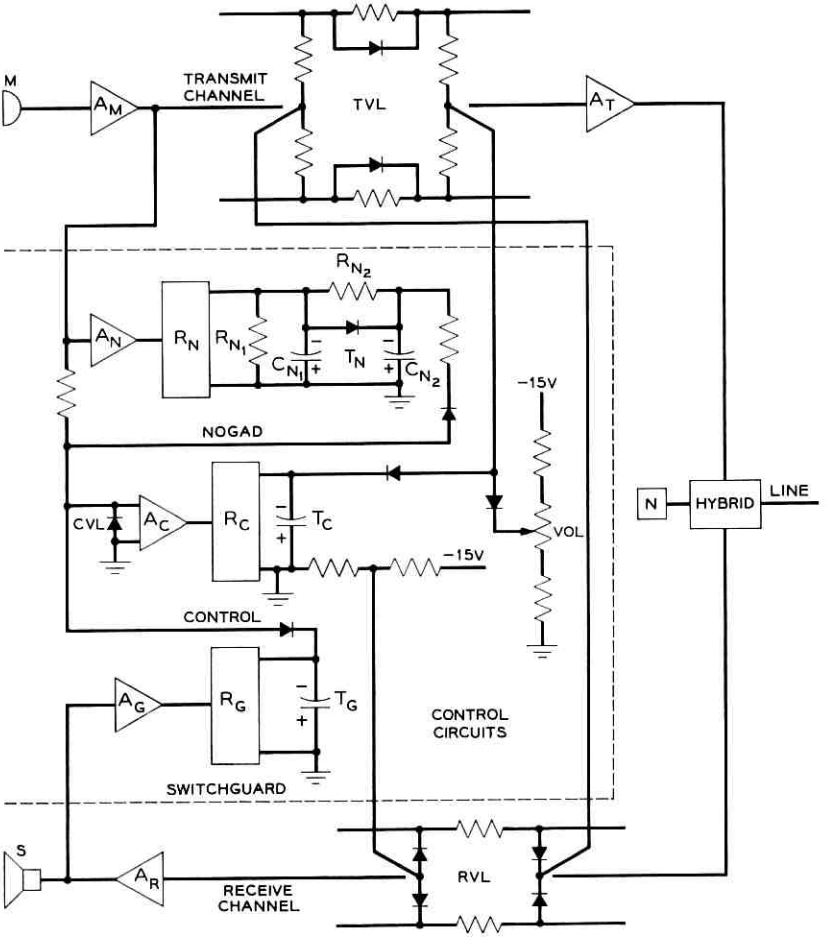


Fig. 2 — Schematic of 3A Speakerphone.

the present application, the variable impedance element is a diffused silicon varistor, the ac resistance of which varies inversely in a smooth and continuous manner as a function of the direct current through it. In TVL, as can be seen in Fig. 2, the varistors are part of the series path of the transmit channel, so that a decrease of their impedances by an increased flow of direct current increases the gain of the channel. In RVL the varistors are part of the shunt path, and a decrease of their impedances by an increased flow of direct current decreases the gain of the receive channel. The characteristics of the variolossers are shown in Fig. 3. For the range of direct current provided by the control rectifier R_C and the volume control VOL, the change of gain of TVL is 36 db and that of RVL is 42 db. Because the same control current flows through the tandem connection of the dc paths of the variolossers, their gains always change in a complementary manner.

Each variolosser is provided with a dc shunt path, not shown in Fig. 2, which serves to match the two characteristics.

As shown in Fig. 3, the sum of the transmit and receive gain changes is never greater than that for zero current. This avoids possible singing for

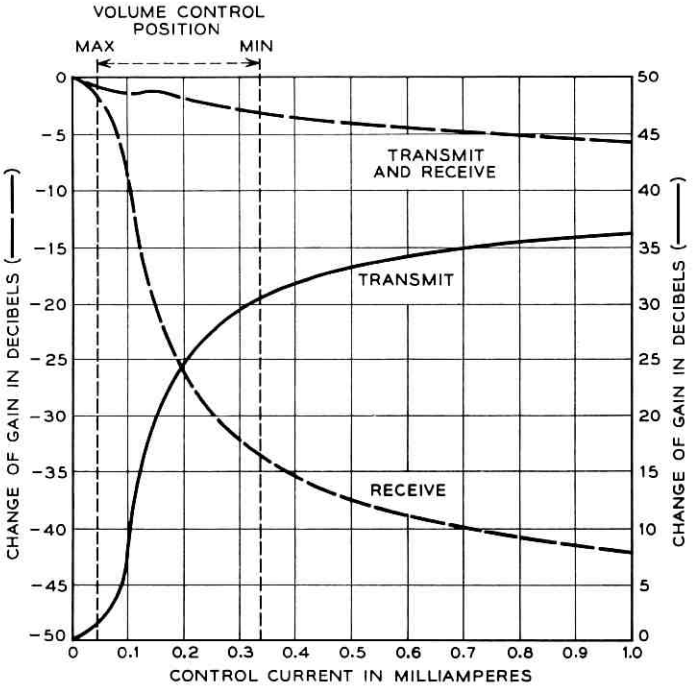


Fig. 3 — Variolosser characteristics.

any setting of the volume control and for the entire range of control current as it varies with speech level.

Because the varistors are nonlinear circuit elements, care must be taken to minimize distortion of speech signals in the variolossers. In the 3A Speakerphone, the signal levels have been adjusted so that the departure from linearity is small over the range of direct currents and signals encountered up to the overload points of the amplifiers in each channel. Furthermore, the variolossers are arranged as balanced circuits to keep control current variations from producing interference or "thump" in the speech channels.

In addition to the current from the control rectifier R_C, the loudspeaker volume control VOL supplies to the variolossers a quantity of direct current dependent on its manual setting. This, as can be seen in Fig. 3, varies the gain of the receive channel and also provides a bias current from which the control current increases in response to speech at the microphone. The resulting effect is that the amount of gain switched by the current from R_C is small for the lower volume control positions and becomes greater as the volume control is advanced, as portrayed in Fig. 4.

On the majority of calls, only a small amount of gain is switched, because low volume control settings give adequate loudspeaker output.

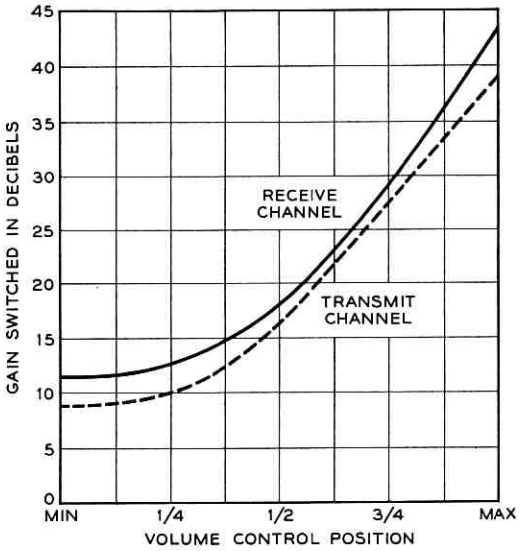


Fig. 4 — Gain switched for various volume control positions.

As a result, simultaneous speech is permitted with barely noticeable clipping, and the changes in the transmitted background noise, resulting from the gain switching, are not apparent at the connected telephone. However, sufficient loss is switched so that the talker echo retransmitted is not objectionable, even for moderately reverberant rooms. On the other hand, when the volume control is near its maximum position on high-loss connections, the resulting voice switching and noise background effects are more noticeable, but tend to be masked by noise at the distant end, because of the lower received signal there.

V. SWITCHING CONTROL CIRCUIT

Through mutual interactions, the control path, the switchguard path, and the NOGAD circuit control the state of TVL and RVL in response to signal flow in the speech channels. In the absence of both a received signal and ambient room noise, speech signals from the microphone pass through CVL with minimum loss, are further amplified by A_C, rectified by R_C, and impressed on timing circuit T_C. When the voltage developed by R_C is larger than the dc bias voltage in series with TVL and RVL, switching action occurs. Fig. 5 shows the steady-state transmit switching characteristic measured with a 1000-cycle signal from the microphone for various volume control positions, when the switchguard and NOGAD circuits are inactive. At the maximum volume setting, a gain change of 37 db occurs as the microphone voltage V_m varies from -97 to -81 dbv. At lower settings the gain change is reduced, but the point at which switching is completed remains the same.

When a loudspeaker voltage is present, the rectified output of the switchguard changes the loss of CVL in proportion to the loudspeaker voltage over a wide range. The switching action of the transmit channel with a loudspeaker voltage V_s of -20 dbv and the NOGAD inactive is shown in Fig. 6. In obtaining these data, the sidetone path is interrupted at the input of RVL and proper terminations attached. A receiving signal to produce V_s is connected at the input of RVL. Then a low microphone voltage is applied and slowly increased. At a critical value of V_m, in this case -63 dbv, the gain of the transmit channel abruptly increases. Simultaneously, the gain of the receive channel, as shown in Fig. 7, abruptly decreases.

This critical value of V_m and the value of V_s existing before the gain transfer define a receive to transmit, R TO T, transition point. In order for the set to return to the R state, V_m must be reduced considerably below the critical value, because of the removal of loss in CVL resulting from the

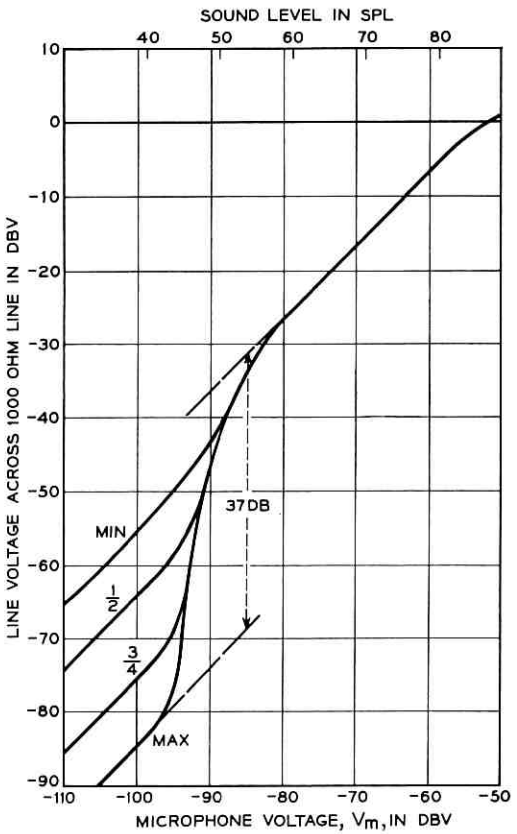


Fig. 5 — Transmit switching characteristic, with volume control position as parameter.

lower loudspeaker voltage at the input of the switchguard. At a second critical value of V_m, in this case −84 dbv, the set returns to the receive state. These values of V_m and V_s determine the T TO R transition point with $V_s = -20$ dbv. The hysteresis effect helps maintain one direction of transmission until a brief interval of low-level speech or a pause occurs in the signals passing through the channel having the higher gain, or until a deliberate attempt is made to interrupt by increasing the signal level in the other channel. The magnitude of this hysteresis effect depends upon the amount of gain switched, as determined by the volume control setting.

The R TO T transition curve of Fig. 8 shows the relationship between V_m and V_s at the transition points. Over a range of 30 db this relation-

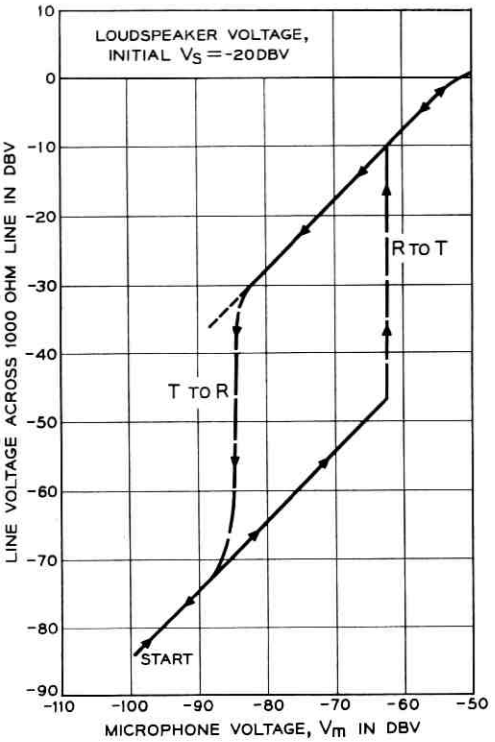


Fig. 6 — Transmit characteristic with receive signal V_s ; NOGAD inactive, maximum volume control setting, V_s applied first.

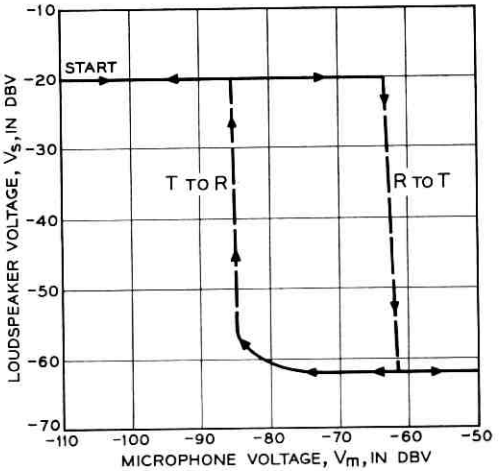


Fig. 7 — Receive characteristic with microphone signal V_m ; NOGAD inactive, maximum volume control setting, V_s applied first.

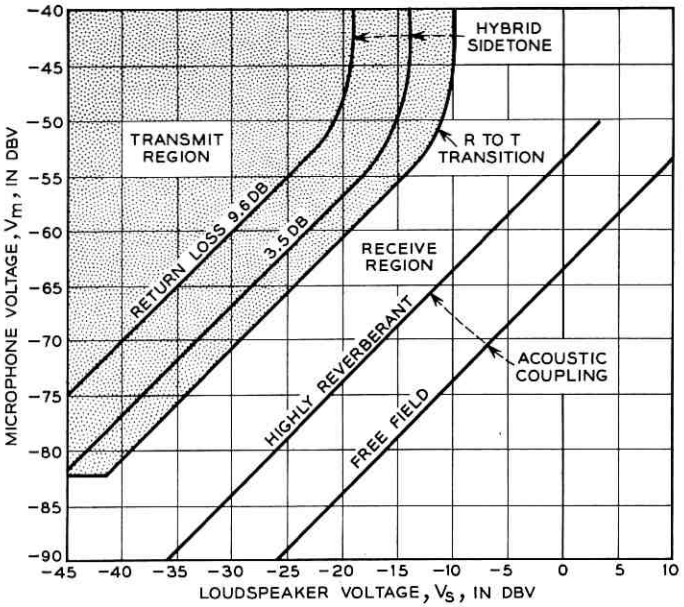


Fig. 8 — R TO T transition, with loudspeaker 3 feet from microphone on table.

ship is linear, and differential switching takes place. At its lower end the curve turns toward the horizontal, which means that V_m is constant for all values of V_s less than −42 dbv. For these low values of V_s, the switchguard does not produce enough current to increase the loss of CVL. At the upper end, overloading of amplifier A_M turns the R TO T transition curve toward the vertical. The significance of this is that, for steady loudspeaker voltages greater than −10 dbv, the set will not switch into the transmit state, regardless of the microphone input level. The electrical noise on the telephone line corresponding to this level, at maximum volume setting, is 60 dbrn,* which would only exist as a result of trouble conditions on a line. When receiving high-level speech, the loudspeaker voltage will exceed this value momentarily, but, due to the fluctuating nature of speech, high values are quickly followed by low values or pauses, during which speech at the microphone can switch the system into the transmit state.

Fig. 8 also shows the acoustic coupling and hybrid sidetone relationships. The two acoustic coupling lines depict the microphone voltage V_m produced by the loudspeaker voltage V_s for a highly reverberant (about

* dbrn = decibels above reference noise.

1 second reverberation time) and a free-field condition. The data represent the highest values observed over a 200-cps band centered at 1000 cps. Since only the microphone, the loudspeaker, and the air path are involved, this coupling is linear over the working range of voltages. The hybrid sidetone curves represent the loudspeaker voltage V_s produced by V_m as a result of hybrid coil unbalance, with the volume control at maximum and the set held in the receive state by opening the dc path of the variolossers. At high values of microphone voltage these sidetone curves are no longer linear and, because of overloading in amplifier A_M, turn towards the vertical in much the same fashion as does the transition curve. The sidetone curve for a return loss of 9.6 db represents a 2-to-1 unbalance between line and network impedance, while the return loss of 3.5 db represents an unbalance of 5 to 1.

For proper switching performance, the R TO T transition curve should lie above the acoustic coupling line but below the hybrid sidetone line.[4] If the transition curve touches or lies below the acoustic coupling line, receive blocking will occur. This means that the set will go into the transmit state instead of remaining in the receive state, because the microphone output resulting from the loudspeaker signal overpowers the switchguard. On the other hand, if the transition curve lies above the hybrid sidetone curve, transmit blocking will result; that is, the set will go into or remain in the receive state because the switchguard action of the sidetone loudspeaker voltage prevents switching into the transmit state.

In the 3A Speakerphone the margin between the transition curve and the acoustic coupling line has been made large, so that there is considerable freedom in positioning the loudspeaker and the microphone with respect to each other and with respect to walls and furniture. Less margin is provided between the transition curve and the hybrid sidetone curve, but conditions influencing this margin are better controlled. The return loss of 3.5 db, which still gives a 4 db margin, occurs only for an extreme condition, such as having three sets bridged on a loop. Aided by the increase in the total loss of TVL and RVL when transferring from the receive to the transmit state, as shown in Fig. 3, it is found that, for the transient signals of speech, hybrid sidetone will not cause transmit blocking with either the line terminals open or short-circuited.

VI. TRANSIENT SWITCHING PERFORMANCE

Both the speed with which the channel gains are switched in response to increases of speech levels and the duration of gain holdover after de-

creases of speech levels must be chosen to minimize clipping of the initial syllables, final syllables, and weaker syllables of a speech burst. In the 3A Speakerphone, a suddenly applied microphone voltage V_m of -70 dbv requires approximately 20 milliseconds to build up the line signal to within 6 db of its final value. This build-up time was selected as a compromise between slight initial clipping and too much sensitivity to impact room noise. The rate of gain change during a switching cycle is also important. By using a high rate of gain change when the gain is low and then decreasing this rate as the gain approaches its maximum, the "swishing" effect due to the rise and fall of the transmitted background noise is reduced.

The decay characteristic is shaped so that, after a microphone signal stops, the total time to revert to the receive state is approximately 300 milliseconds. In the first 115 milliseconds the transmit channel gain decreases about 5 db, and thereafter it falls at the more rapid rate of 0.17 db per millisecond. It has been found that a shorter decay time curtails the weak consonants at the ends of words and produces an undesirable expansion or "pumping" action on speech. On the other hand, a substantially longer decay time slows down the reply of the other party to an objectionable degree.

The time constant of the switchguard path, determined by T_G in Fig. 2, is similar to that of the control path. The decay time of the switchguard is of particular interest because switching performance when receiving in a reverberant environment is dependent upon it. This decay time of the switchguard, which is the recovery time of the gain of the control path, occurs at an essentially constant rate of 0.2 db per millisecond. It will completely prevent any switching from R to T in rooms having reverberation times up to 0.5 second, while the speech sounds are decaying in the room. Actually, because of other factors, satisfactory operation is obtained in rooms with reverberation times approaching 1 second. These are: (a) speech sounds do not end abruptly and their decay time adds in part to the decay time of the circuit; (b) prolonged reverberation causes partial operation of the NOGAD circuit and delays the gain increase of the control path; (c) the transient signals of incoming speech do not build up the reverberant sound to its full steady-state level.

When one party replies quickly after the other has ceased talking, the over-all transfer time, determined primarily by the release actions of the control path and the switchguard path, is dependent on the relative levels of the signals in each channel. This transfer time becomes longer when the signal from the replying party is weak with respect to the signal from the party relinquishing control. On local calls, for which low volume

control settings are adequate and less gain is switched, the receive replies can obtain control more quickly, and thus more rapid interchange is obtained, especially for the case of overlapping speech. The receive channel normally requires approximately 250 milliseconds to return to full gain under typical conditions. The maximum time is slightly over 300 milliseconds and the minimum is approximately 150 milliseconds for normal local speech levels at the microphone. In the transmit direction, the differential attack time is even more dependent upon relative signal levels, varying from 250 milliseconds when local speech levels are low with respect to the received loudspeaker speech levels to 20 milliseconds when the reverse conditions occur. A typical value can be considered to be 150 milliseconds.

VII. NOGAD CIRCUIT

To avoid noticeable clipping of the transmitted signal under quiet conditions, a sound level of approximately 45 db SPL initiates switching into the transmit state. This is less than the noise level produced by fans, air conditioners, and street and hall traffic at many locations where it is desired to use a speakerphone. Therefore, the possibility exists that the noise would block the receive channel. To prevent this, the noise-operated gain-adjusting device, NOGAD, passes a direct current, correlated with the noise, through CVL, and thus lowers the gain of the control circuit and raises the sound pressure needed to initiate switching. As people instinctively talk louder under noisy conditions, this introduces little or no switching degradation.

The NOGAD circuit recognizes the difference between the speech and the noise signals at the microphone on the basis that speech fluctuates more rapidly than most types of noise. The timing circuit T_N, shown in Fig. 2, consists of an *RC* circuit, R_{N_1}, C_{N_1}, with a time constant of about 1 millisecond, coupled to another circuit, R_{N_2}, C_{N_2}, having a build-up time of about 4 seconds and a short decay time, effected through the diode and R_{N_1} of about 100 milliseconds. Thus, when a signal consisting of both speech and noise exists at the microphone, the voltage across C_{N_2} is held closely to the nearly constant voltage component across C_{N_1} caused by the noise, and is little affected by the rapidly fluctuating voltages due to speech. The voltage across C_{N_2} then produces a direct current in CVL through a diode gate. This general method of discriminating between noise and speech is known in the echo suppressor art.[7]

With the circuit simply as described, an excessive time would be required for the set to adjust to room noise conditions after it is first turned on. To eliminate this delay, a precharging circuit, not shown in Fig. 2,

of approximately 4 db. Some of this advantage undoubtedly accrues from binaural listening with the speakerphone. Fig. 11 shows the receive frequency response. It is characterized by a small peak at 450 cps, a sharp cut-off near 300 cps, and a more gentle fall off beyond 2200 cps. The sharp cut-off near 300 cps, assures good discrimination against noise induction from 60-cycle power and is not detrimental to articulation. The high-

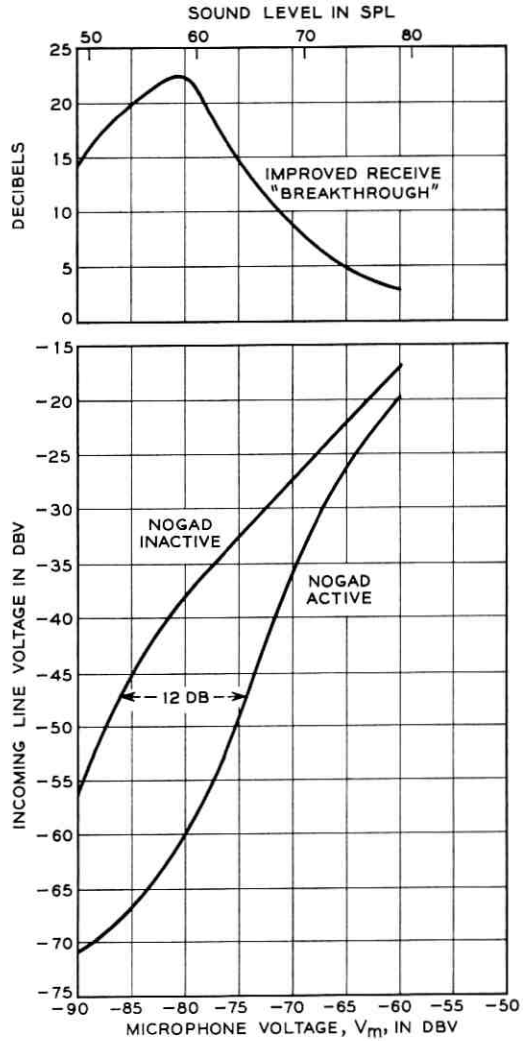


Fig. 10 — Effect of NOGAD on T TO R transition at maximum volume control setting.

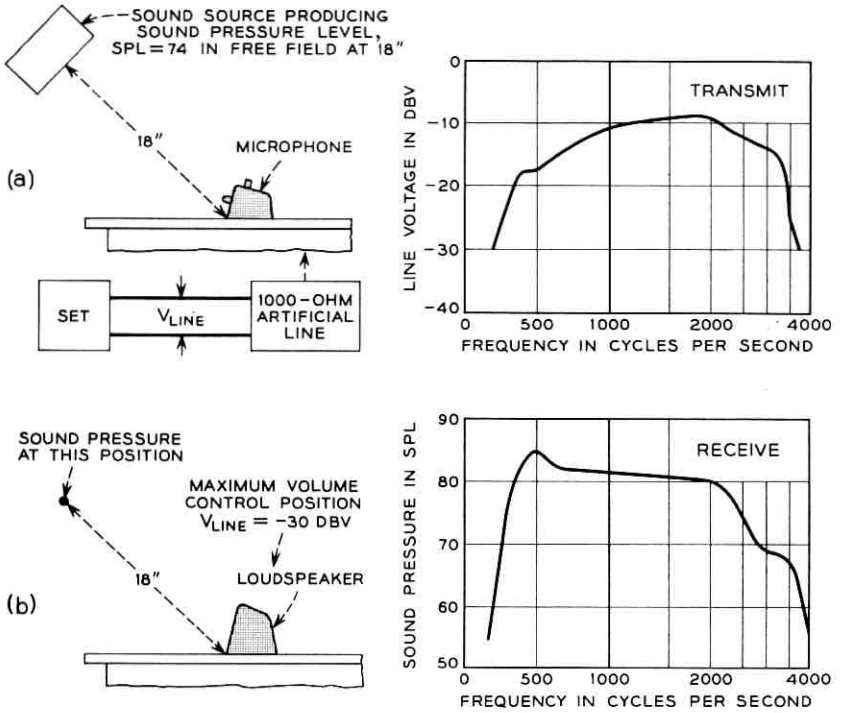


Fig. 11 — (a) Transmit and (b) receive frequency response.

frequency fall-off is caused by an acoustic interference between the direct sound wave and the wave reflected from the table top on which the loudspeaker rests. This loss is partially compensated for by sound diffraction and resonance effects occurring near the listener's ear.[8]

IX. CONCLUSIONS

The application of voice-switched gain in the 3A Speakerphone eliminates the talker echo and singing problems encountered in hands-free telephones having fixed gain, and provides very satisfactory transmission in the acoustic environment normal for homes and private offices. By careful design, the gain-switching action has been kept free of objectionable clipping or blocking, and is substantially unaffected by ordinary room noise. On most local telephone connections, the amount of gain which is switched is so small that the gain changes are barely noticeable. Finally, no adjustments are required for the noise and reverberation conditions most likely to be encountered.

X. ACKNOWLEDGMENTS

The authors wish to express their appreciation to D. Mitchell for many helpful discussions of fundamental problems, and to those other persons who participated in this development work.

REFERENCES

1. Clemency, W. F., Romanow, F. F., and Rose, A. F., The Bell System Speakerphone, A.I.E.E. Trans., Pt. I, **76,** 1957, p. 148.
2. Emling, J. W., General Aspects of Hands-Free Telephony, A.I.E.E. Trans., Pt. I, **76,** 1957, p. 201.
3. Gardner, M. B., A Study of Talking Distance and Related Parameters in Hands-Free Telephony, B.S.T.J., **39,** 1960, p. 1529.
4. Busala, A., Fundamental Considerations in the Design of a Voice-Switched Speakerphone, B.S.T.J., **39,** 1960, p. 265.
5. Stansel, F. F., N1 Carrier—Selection of Varistors for Use in Compandors, Bell Labs. Rec., **31,** 1953, p. 501.
6. Fracassi, R. D., The Compandor for N1 Carrier, Bell Labs. Rec., **31,** 1953, p. 452.
7. Wright, S. B., and Mitchell, D., U.S. Patents Nos. 1,814,017 and 1,814,018, July 14, 1931.
8. Wiener, F., and Ross, D. A., Pressure Distribution in the Auditory Canal, J. Acoust. Soc. Am., **18,** 1946, p. 401.

A Block Diagram Compiler

By JOHN L. KELLY, Jr., CAROL LOCHBAUM, and V. A. VYSSOTSKY

(Manuscript received December 7, 1960)

A computer program known as BLODI, which accepts for an input a source program written in the BLODI language, is described. The BLODI source language corresponds closely to an engineer's block diagram of a circuit and is easily learned, even by persons not familiar with computing machines. The input code consists essentially of designating the connectivity of a number of boxes drawn from an alphabet of about 30 types. These types include amplifiers, delay lines, counters, etc., which are familiar to designers of electronic circuits. The principles of the compiler are explained and applications are discussed.

I. INTRODUCTION

This paper describes a computer program known as BLODI (BLOck DIagram compiler). BLODI accepts for an input a source program written in the BLODI language, which corresponds closely to an engineer's block diagram of a circuit, and produces a machine program to simulate the circuit. BLODI has been written for both the IBM 704 and 7090 machines, and has been in use at Bell Telephone Laboratories for several months. Generally speaking, there are two situations in which it can be used profitably. One arises when a person with no knowledge of machine coding wishes to program his own problem. In this case, the BLODI language is much easier to learn than Fortran or SAP. There are, in addition, certain problems involving a rather smooth flow of data which can be most easily coded in BLODI, even by an experienced programmer.

It is rather easy to estimate the efficiency of an object program produced by BLODI. Thus a person with no knowledge of computing machines can often tell if he should code his problem in BLODI or seek the aid of an experienced programmer. This will be discussed in Section V.

BLODI was written to lighten the programming burden in problems concerning the simulation of signal-processing devices. It has the added

advantage of keeping the engineer who invents such a device in close communication with the computing machine by eliminating the middleman (expert programmer).

II. BLOCK DIAGRAM OF SAMPLED (OR PULSE) SYSTEMS

The circuits* which we wish to consider here are limited to combinations of devices which accept pulses as inputs and yield pulses as outputs. While the pulses may have arbitrary sizes within certain limits, they must all occur at multiples of a fixed clock time. In general, the output of one of the devices (or boxes) can depend on the present and all past input pulses. A box whose output is independent of the current input pulse or pulses is called a *delaying-type box*. In the current form of the compiler the only delaying-type box is a simple delay line. In addition to these boxes, the circuit may have one or more ultimate outputs and original inputs. A *circuit* then means a number of boxes, ultimate outputs, and original inputs connected in such a way that the output of any box is connected to one or more inputs to boxes and ultimate outputs, and each input to a box is connected to an output from a box or an original input. (We limit ourselves to boxes with a single output.)

A *closed loop* is a path that starts from any point, goes only through connected boxes in the direction of the pulses (i.e., input to output), and returns to its starting point. A circuit which does not contain a closed loop with no delaying-type boxes will be called an *admissible* circuit. The compiler will reject any block diagram which does not describe an admissible circuit. It is easy to see that if the pulse heights were limited to a finite number of values (as they are, of course, in the machine simulation) then an admissible circuit would be a finite-state machine. On the other hand, there is no way to interpret a block diagram corresponding to a nonadmissible circuit. To be sure, one could connect physical boxes in such a manner and something would happen. The analysis, however, would involve the precise transient behavior of the devices within the pulse width — information which is not available to the compiler.

In addition to simulating pulse circuits, the compiler may be used to simulate continuous circuits whose inputs and outputs are bandlimited time functions. One must first design a pulse circuit whose output pulses would correspond to the sample values of the desired output. Extreme

* For clarity the machine being simulated will be called the *circuit*. Its description in a certain canonical form will be called the *block diagram*. The word *machine* will always mean the IBM 704 EDPS (or 7090 EDPS).

care must be exercised here; for example, an accumulator (a device whose output is the sum of all previous input pulses) is not equivalent to an integrator. Certain continuous circuits (especially nonlinear ones) are extremely difficult to translate into pulse form. A simple circuit which is difficult to simulate on a machine (with or without the use of this compiler) is the following: Let a bandlimited input signal be connected to a full-wave rectifier, and this to a low-pass filter to return the signal to the original bandwidth.

We will see later that (with a trivial exception) the compiler can be used to simulate any admissible circuit composed of boxes drawn from a fixed list or stockpile. The boxes may have only one output (which may go to several places, however) and at most four inputs. The first constraint is no real restriction, but the second one is.* We know of no example in signal processing where a general function of more than four variables that cannot be expressed as combinations of functions of four or less variables is needed. In fact, two inputs to each box would probably be adequate but slightly awkward.

III. THE BLODI LANGUAGE

A BLODI source program is punched on standard SHARE symbolic cards in either the FAP (7090) or SAP (704) format. In general, each card corresponds to one box in the circuit; there is, however, a provision made for continuing a description of a box to the next card. The location field (columns 1 through 6) is either blank or contains the name assigned the box by the programmer. (If a box is to have any inputs, it must have a name.) The operation code field (columns 8 through 10) contains the type of box. Parameters (such as gain of an amplifier) and output connections are separated by commas and listed consecutively starting in column 12 (or column 16 for the 7090 format). The various inputs to the same box are designated by a fraction bar and numeral following the name of the box. Example:

UV AMP 5.28, XY, Z/2

Box UV is an amplifier with a gain of 5.28 which feeds box XY (first input) and the second input of box Z.

A list of all the available box types appears in Table I. Note that INP may be thought of as a box which generates signals spontaneously; actually it obtains its input from a tape designated as a parameter. Simi-

* Technically speaking, this is not true. A circuit could be designed corresponding to any finite state machine. However, we consider that it is not in the proper spirit to take advantage of the fact that the pulse heights are limited to 2^{36} values.

TABLE I — ALL THE TYPES OF BOXES WHICH ARE RECOGNIZED BY THE COMPILER

Type	Function	Inputs	Parameters
DEL	Delay	Signal	Number of units delay
AMP	Amplifier	Signal	Gain
ADR	Adder	1–4 Signals	None
SUB	Subtractor	+ Input − Input	None
MAX	Maximum circuit	1–4 Signals	None
MIN	Minimum circuit	1–4 Signals	None
CLP	Positive clipper	Signal	Clipping level
CLN	Negative clipper	Signal	Clipping level
SCL	Symmetric clipper	Signal	Clipping level
FWR	Full-wave rectifier	Signal	None
BAT	Battery or bias	(Signal)	Bias
MRP	Multiplier	2 Signals	None
DIV	Divider	Dividend Divisor	None
SQT	Square rooter	Signal	None
ACC	Accumulator	Signal	Gain
FLT	Transversal filter	Signal	Number of taps Delay per tap Gains
SLF	Symmetric filter	Signal	Number of taps Delay per tap Gains
AFL	Antisymmetric filter	Signal	Number of taps Delay per tap Gains
QNT	Quantizer	Signal	Number of levels Levels
LQT	Linear quantizer	Signal	Step size
SMP	Sampler	Signal	Period Quiescent level Initial phase
HLD	Sample and hold	Signal; control	Threshold
CNT	Counter	Signal	Countdown factor Threshold Active level Passive level Initial phase
DTS	Double-throw switch	Control; 2 signals	Threshold
FLF	Flip-flop	Signal	Low threshold High threshold Low state output High state output
PLS	Pulser	Control	Threshold Pulse length Pulse level Quiescent level
CØS	Cosine generator	None	Period Phase Amplitude
GEN	Function generator	None	Period Sample values
WNG	Noise generator	None	Standard deviation
PRT	Printer	Control; 3 signals	Threshold Record Limit

TABLE I — (*Continued*)

Type	Function	Inputs	Parameters
INP	Input	None	Tape number File maximum Samples per record Record maximum Start printing Stop printing
ØUT	Output	Signal	Tape number File maximum Samples per record Record maximum Start printing Stop printing
END	Last card of source program		

larly, OUT causes the signal appearing on its input lead to be written on the designated tape. A circuit may have several inputs and outputs. END is not a box at all but signifies the end of the source program. In addition to the types listed, it is possible for a programmer to create types of his own invention by supplying subroutines written in basic machine language. It is also quite easy to change the basic input-output programs used by BLODI to handle arbitrary tape formats.

IV. PRINCIPLE OF OPERATION

An object program produced by the compiler consists of three parts:

(a) the *prefix*, which sets up the logic for the main loop;

(b) the *main loop*, which is executed once for each sample processed;

(c) the *suffix*, which causes the main loop to be repeated the proper number of times, empties output buffers, fills input buffers, etc.

We will concern ourselves here only with the main loop. Except for some strictly local inner loops in certain boxes, this part of the object program is compiled in the same order in which it is to be executed. Simply stated, the procedure is as follows: One storage cell in the object program is assigned for each box. Each time the main loop is entered, these cells will contain values corresponding to the last *outputs* of the respective boxes. It is then the function of the main loop to compute these output values for the next (current) time slot and to fill the cells with these values.

In order to simplify the description of the algorithm used by the compiler we will at first limit ourselves to the case where all delays are unit delays. By "compiling a box" we mean writing the necessary coding to

cause the object program to fill the output cell of the box with the current pulse value. A nondelaying-type box cannot be compiled until all the boxes feeding it have been compiled. The reason is that the output of this type of box is a function of its current inputs, and this part of the object program must not be executed until the cells corresponding to input to this box have been filled with current pulse values. On the other hand, a unit delay must be compiled *before* the box which feeds it. Its output is a function of (equal to, in fact) its *last* or *old* input. At object time this value must be "moved along" before it is overwritten. To reword this second rule, no box which feeds a delay line can be compiled until that delay line has been. This second rule could be dropped if we provided an additional storage cell for each unit delay and had the object program first go through and fill each of these cells with the old input to the corresponding delay line. We will see below, however, that the only price we pay for the more efficient procedure is that the compiler will reject any block diagram containing a closed loop with nothing but delays. Such a diagram would represent an admissible circuit but would be of little value, since we could never get anything but zeros out of this loop at any point. (All delay lines are initialized at zero.)

The above two rules are effected in a fairly simple manner. A binary storage cell is assigned in the *compiler* program for each of the output cells in the *object* program. The two states of each of these cells are called "full" and "empty." Initially all cells which represent inputs to delay lines are marked "full" and all others marked "empty." A box can be compiled whenever its inputs are all marked "full" and its output "empty." When a box is compiled, its output is marked "full" and its inputs "empty." Compilation proceeds until all boxes have been compiled (successful compilation) or until no uncompiled box meets the two requirements. In the latter event the compiler prints the remark CLOSED LOOP WITH NO DELAYS OR ALL DELAYS and halts. To see that one of these conditions must prevail, note that a delay line can *always* be compiled unless it feeds another uncompiled delay line. Therefore, if any of the uncompiled boxes are delay lines there exists a closed loop with all delays. If, on the other hand, all uncompiled boxes are nondelaying, then each must have an empty input which must be the output of an uncompiled nondelaying box. Thus, working backwards, we find a closed loop with no delay.

The order of searching for uncompiled boxes which meet the tests is immaterial from a logical point of view, but this freedom can be used to optimize the use of the accumulator. By first trying to compile a box which is fed by the last compiled box, the compiler is sometimes able to save a "storage" or "fetch" order, or both. Delays are not, of course,

limited to unit delays as in the above discussion. A delay of length n sets $n - 1$ storage cells in addition to its normal output cell. For short delays the data are "stepped along" a notch each time the main loop is executed. For longer delays the same effect is produced by address modification.

The above description is merely a sketch of the general procedure used by BLODI. Actually, it has a lot more structure of purely technical character. For example, some boxes are broken down into simpler boxes by the compiler so that parts of the object program concerning a given box may not appear in consecutive locations.

V. CONCLUSIONS

BLODI has been in use at Bell Telephone Laboratories for about a year, mostly in the Department of Visual and Acoustics Research. It has been used chiefly for signal processing types of problems, and one of the authors has used the compiler to simulate a speech synthesizer of the resonant-vocoder type. It has also been used to study television coding schemes, artificial reverberation in acoustic research, and for part of the coding in a handwriting recognition problem.[1] In appraising its value one must consider two separate questions:

1. What type of problem is easily coded in the BLODI language?
2. What type of problem causes BLODI to write efficient object programs?

The first question will be answered relative to a programmer well versed in basic and Fortran language. Whenever the problem involves a rather smooth flow of data in and out of the machine, with the output being a nearly stationary function of the input, then it can be more easily coded in BLODI than in any other existing language. When, however, the program must process input samples in a complicated order, dependent on previous results, the BLODI language becomes unbearably awkward. (This is precisely the type of circuit which is hard to design with delay lines, switches, etc.)

The second question is easily answered. Any diagram which contains many idle boxes will be inefficient, because the object program goes through the motion of calculating the state of all boxes at each clock time. For example, a program with five memory-free (delay-free) paths, only one of which is connected to the output at any one time, would result in an inefficient object program. For diagrams containing few idle boxes, however, BLODI produces object programs which are usually as efficient as those written by a competent programmer.

The version of BLODI in use at Bell Telephone Laboratories is coupled to the monitor and I-O system, BE SYS 3. Thus an installation not us-

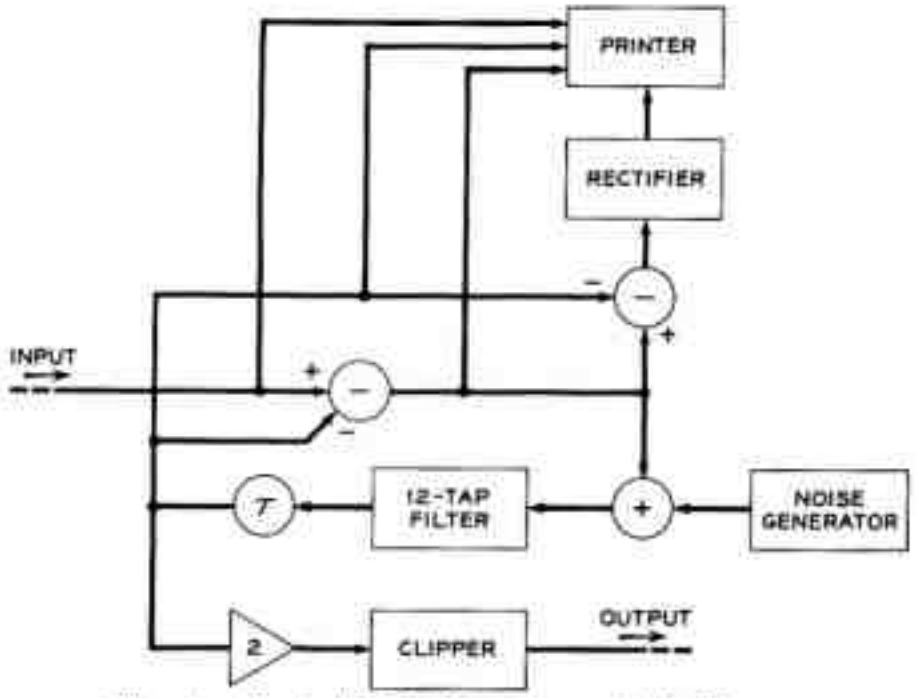


Fig. 1 — Typical BLODI program: block diagram.

```
                    SAMPLE BLODI SOURCE PROGRAM

        MNG     SUM=1,100
SUM     ADR     BUFF
BUFF    FLT     12,1,.001,.002,.004,.008,.016,.032
                .063,1/8,.230,1/4,1/5,.057,DELAY
DELAY   DEL     1,SCALE,SUB=2
                PRINT/3,SUB1/2          XXX
SCALE   AMP     2,CLIP
CLIP    CLN     T2
T2      OUT     5
                ,,,1,1                  XXX
SUB     SUB     SUM=2
                SUB1/1,PRINT/4          XXX
T1      INP     ,,,1,SUB=1
                PRINT=2,1,1             XXX

SUB1    SUB     R                       THESE CARDS AND THE CARDS MARKED
R       FWR     PRINT=1                 XXX WOULD BE OMITTED IF
PRINT   PRT     150                     PRINTING WERE NOT DESIRED

        END
```

Fig. 2 — Typical BLODI program: source program.

ing BE SYS 3 would have to modify the BLODI program. The changes, however, would only involve I-O and interaction with the FAP (or SAP) assembly program.

Figs. 1, 2, and 3 represent a sample BLODI program. Fig. 1 is a block diagram suitable for simulation using BLODI; Fig. 2 is the corresponding source program; and Fig. 3 shows the printed output that results from compiling the source program and running the simulation.

REFERENCE

1. Frishkopf, L. S., and Harmon, L. D., Machine Recognition of Cursive Script, Fourth Annual Symposium on Information Theory, 1960.

PAGE 1

```
                        02734          PRG      1,200
                        00101   HEAD   EQU 47
                        00210   WRITE  EQU      T+1
                        00401   MAN    EQU      257
                        00421   [illegible] EQU 277
                        00517   CHR    EQU      123
02734  2 00000 0 00001  01     FTB 1
02735  0146 00 0 02730  0      TDR a+1
02736  0767 01 0 00001         [illegible]
02737  0761 01 0 00000         NPR
02740  2 00401 0 02741         [illegible]
02741  0500 00 0 02741         CLS a+1
02742  0601 00 0 02760         STR a+2
02743  0534 00 1 02753         LDA L,1
02744  0600 00 1 02753         STA L,1
02745  2 00001 1 02744         TIL a-1,1,1
02746  0534 01 1 02665   a     LDA 0302,1
02747  0500 01 1 00000         CLA 0-000001,1                     T1
02750  0601 00 0 02617         STR T+1
02751  0500 00 0 02313         CLA T-2                         DELAY
02752  0601 00 0 02313         STR T+2
02753  0767 00 0 00021         ALS 17                          STALE
02754  0560 00 0 01400         LDQ 0                           ELSE
02755  0560 00 0 02757         TLQ a+2
02756  0500 00 0 07024         CLA 0
02757  0601 00 1 07777         STZ 0-07777,1                      T2
02760  0500 00 0 00027         CLA T+1                          BUFF
02761  0765 00 0 00043         LRS 43
02762  0200 00 0 00001         MPY a+1
02763  0763 00 0 00021         LLS 21    [illegible]
02764  0760 00 0 00010         RND
02765  0601 00 0 01324         STO T+18
02766  0500 00 0 02325         CLA T+12
02767  0601 00 0 02312         STO T+14
02770  0765 00 0 00043         LRS 43
02771  0200 00 0 01370         MPY C+2
02772  0763 00 0 00021         LLS 21    [illegible]
02773  0760 00 0 00010         RND
02774  0601 00 0 07324         STO T+18
02775  0500 00 0 02327         CLA T+20
02776  0601 00 0 02325         STO T+12
02777  0765 00 0 00043         LRS 43
03000  0200 00 0 02377         MPY C+3
03001  0763 00 0 00021         LLS 21    [illegible]
03002  0760 00 0 00010         RND
03003  0601 00 0 02330         STO T+21
03004  0500 00 0 02331         CLA T+30
03005  0601 00 0 02327         STO T+20
03006  0765 00 0 00043         LRS 43
03007  0200 00 0 02372         MPY C+4
03010  0763 00 0 00021         LLS 21    [illegible]
03011  0760 00 0 00010         RND
03012  0601 00 0 02332         STO T+22
03013  0500 00 0 02333         CLA T+24
03014  0601 00 0 02331         STO T+23
03015  0765 00 0 00043         LRS 43
03016  0200 00 0 02373         MPY C+5
03017  0763 00 0 00021         LLS 21    [illegible]
```

PAGE 2

```
03112  0500 00 0 02301         CLA T+43
03113  0100 00 0 02157         TZE a+42
03114  0400 00 0 00754         SUB C1
03115  0601 00 0 02701         STO T+12
03116  0771 00 1 00030         [illegible]
03117  0771 00 0 00022         ARS 22
03120  0402 00 0 02247         SUB a+27
03121  0402 00 0 02302         SLW T+24
03122  0500 00 0 02302         CLA T+7
03123  0400 00 0 02784         ADD C1
03124  0601 00 0 02302         STO T+8
03125  0074 00 4 77514         TSX (TPR),4
03126  0 77777 0 02142             [illegible]
03127  0560 00 0 02302         LDQ T+15
03130  -1 00000 0 00000        STQ
03131  0560 00 0 03302         LDQ T+14
03132  -1 00000 0 00000        STQ
03133  0560 00 0 02317         LDQ T+15
03134  -1 00000 0 00000        STQ
03135  0560 00 0 02312         LDQ T+7
03136  -1 00000 0 00000        STQ
03137  0560 00 0 02316         LDQ T+7
03140  -1 00000 0 00000        STQ
03141  0074 00 4 77024         TSX (FIL),4
03142  0020 00 0 02143         TRA a+1
03143  74632000235            BCD 3(0 PRINTBUF)
03144  [illegible]
03145  [illegible]
03146  -3 77777 0 77777            [illegible]
03147  0 00000 0 00755             [illegible]
03150  0500 00 0 02322         CLA C+5
03151  0400 00 0 02734         ADD C1
03152  0601 00 0 02563         STO C+5
03153  0074 00 4 00401         TSX MAN,4
03154  0765 00 0 00043         LRS 43
03155  0200 00 0 02304         MPY C+3
03156  0763 00 0 00021         LLS 21    [illegible]
03157  0760 00 0 00010         RND
03160  0400 00 0 02311         ADD T+3                                    BUFF
03161  0601 00 0 02347         STO T+40                                   BUFF
03162  0765 00 0 00043         LRS 43                                     BUFF
03163  0200 00 0 02305         MPY C+12
03164  0763 00 0 00021         LLS 21    [illegible]
03165  0760 00 0 00010         RND
03166  0400 00 0 02330         ADD T+40                                   BUFF
03167  0601 00 0 02311         STO T+3
03170  2 00001 0 02747         TIX a+3,1,1
03171  -0140 00 4 02201        TNX a+10
03172  0074 00 4 77000         TSX (PRINT),4
03173  0 00004 0 02170             a+2,,4
03174  0020 00 0 02207         TRA a+6
03175  0023315123[illegible]   BCD 4(OUTPUT BUFFLNORD)
03176  [illegible]
03177  [illegible]
03200  [illegible]
03201  0074 00 4 00211         TSX WRITE,4
03202  0 00065 0 02333             [illegible]
03203  0 00000 0 00000             0+5
```

PAGE 3

```
03020  0760 00 0 00010         RND
03021  0601 00 0 02334         STO T+25
03022  0500 00 0 02335         CLA T+22
03023  0601 00 0 02333         STO T+24
03024  0765 00 0 00043         LRS 43
03025  0200 00 0 02374         MPY C+6
03026  0763 00 0 00021         LLS 21    [illegible]
03027  0760 00 0 00010         RND
03030  0601 00 0 02336         STO T+26
03031  0500 00 0 02337         CLA T+30
03032  0601 00 0 02335         STO T+26
03033  0765 00 0 00043         LRS 43
03034  0200 00 0 02375         MPY C+7
03035  0763 00 0 00021         LLS 21    [illegible]
03036  0760 00 0 00010         RND
03037  0601 00 0 02340         STO T+31
03040  0500 00 0 02341         CLA T+28
03041  0601 00 0 02337         STO T+30
03042  0765 00 0 00043         LRS 43
03043  0760 00 0 00010         RND
03044  0601 00 0 02342         STO T+33
03045  0500 00 0 02343         CLA T+34
03046  0601 00 0 02341         STO T+32
03047  0765 00 0 00043         LRS 43
03050  0200 00 0 02376         MPY C+10
03051  0763 00 0 00021         LLS 21    [illegible]
03052  0760 00 0 00010         RND
03053  0601 00 0 01344         STO T+35
03054  0500 00 0 02345         CLA T+36
03055  0601 00 0 02343         STO T+34
03056  0765 00 0 00043         LRS [illegible]
03057  0760 00 0 00010         RND
03060  0601 00 0 02346         STO T+37
03061  0500 00 0 02347         CLA T+40
03062  0601 00 0 02345         STO T+36
03063  0765 00 0 00043         LRS 43
03064  0200 00 0 02377         MPY C+11
03065  0763 00 0 00021         LLS 21    [illegible]
03066  0760 00 0 00010         RND
03067  0400 00 0 02340         ADD T+37
03070  0400 00 0 02344         ADD T+35
03071  0400 00 0 02342         ADD T+33
03072  0400 00 0 02340         ADD T+31
03073  0400 00 0 02336         ADD T+27
03074  0400 00 0 02334         ADD T+25
03075  0400 00 0 02332         ADD T+23
03076  0400 00 0 02330         ADD T+21
03077  0400 00 0 02326         ADD T+17
03100  0400 00 0 02324         ADD T+15
03101  0601 00 0 02350         STO T+41
03102  0400 00 0 02311         CLA T+40                                    SUB
03103  0402 00 0 02312         SUB T+3
03104  0402 00 0 02313         SUB T+7
03105  0402 00 0 02312         SUB T+3                                    TIME
03106  0760 00 0 00003         SSP                                          Y
03107  0560 00 0 02300         LDQ C+12                                  TT(HT)
03110  [illegible]             TRA [illegible]
03111  0020 00 0 02160         TRA a+37
```

PAGE 4

```
03204  0534 00 4 02300         LXD C+0002S,4
03205  -3 00002 4 03211        TXL a+10,4,1
03206  2 00001 4 03227         TIX a+16,4,1
03207  0074 00 4 77514         TSX (SPH),4
03210  0 77777 0 02211             a+1,,7
03211  0534 00 2 02201         LXA 0120,2
03212  0560 00 2 07777         LDQ 0-07777,2
03213  -1 00000 0 00000        STQ
03214  2 00001 2 03212         TIX a-2,2,1
03215  0074 00 4 77024         TSX (FIL),4
03216  0020 00 0 02222         TRA a+4
03217  74113050000            BCD 3(OH         TITRE)
03220  [illegible]
03221  [illegible]
03222  -3 77777 0 77777            [illegible]
03223  -0634 00 4 03301        LXD C+0002S,4
03224  1 00001 4 03225        TXI a+1,4,1
03225  -0634 00 4 02300        SXD C+0002S,4
03226  2 00001 4 03227         TIX a+1,4,1
03227  0600 00 0 02252       X STZ T+00043
03230  0074 00 4 00101         TSX HEAD,4
03231  0 00065 0 00000             [illegible]
03232  0 00000 0 00000             0+5
03233  -0634 00 4 03301        LXD C+0002S,4
03234  -3 00000 4 03301        TXL a+1,4,1
03235  3 00001 4 03251         TXH a+15,4,1
03236  0074 00 4 03250         TSX [illegible]
03237  0 77777 0 02247             [illegible]
03240  0534 00 2 02267         LXA 0120,2
03241  0560 00 2 00000         LDQ 0+0120,2
03242  -1 00000 0 00000        STQ
03243  2 00001 2 01247         TIX a-2,2,1
03244  0074 00 4 77024         TSX (FIL),4
03245  0020 00 0 03252         TRA a+5
03246  [illegible]             BCD 2(OH          TITRE)
03247  [illegible]
03250  [illegible]
03251  -3 77777 0 77777            CON -1,7,-1
03252  0500 00 0 00101         CLA PAGE+1
03253  0100 00 0 02275         TZE 0
03254  0600 00 0 00100         STZ PAGE+1
03255  0770 00 0 02205         WEFB     +1
03256  0500 00 0 02774         CLA C1
03257  0534 00 4 02301         LXA T+0002S,4
03260  0601 00 0 02302         STO T+00043
03261  -3 00000 4 03230        TXL 0+1,4,5
03262  0020 00 0 02263         TRA 0+1
03263  0770 00 0 02203    EXIT WEFB     +3
03264  0020 00 0 77000    EXIT1 TRA EVERCH
03265  0 00000 0 00000    TITRE    0+TOT
                        00770  T       [illegible]
03266  +000000000000      T     OCT +000000000000
03267  +000000000000            OCT +000000000000
03270  +000000000000            OCT +000000000000
03271  +000000000000            OCT +000000000000
03272  +000000000000            OCT +000000000000
03273  +000000000000            OCT +000000000000
03274  +000000000000            OCT +000000000000
```

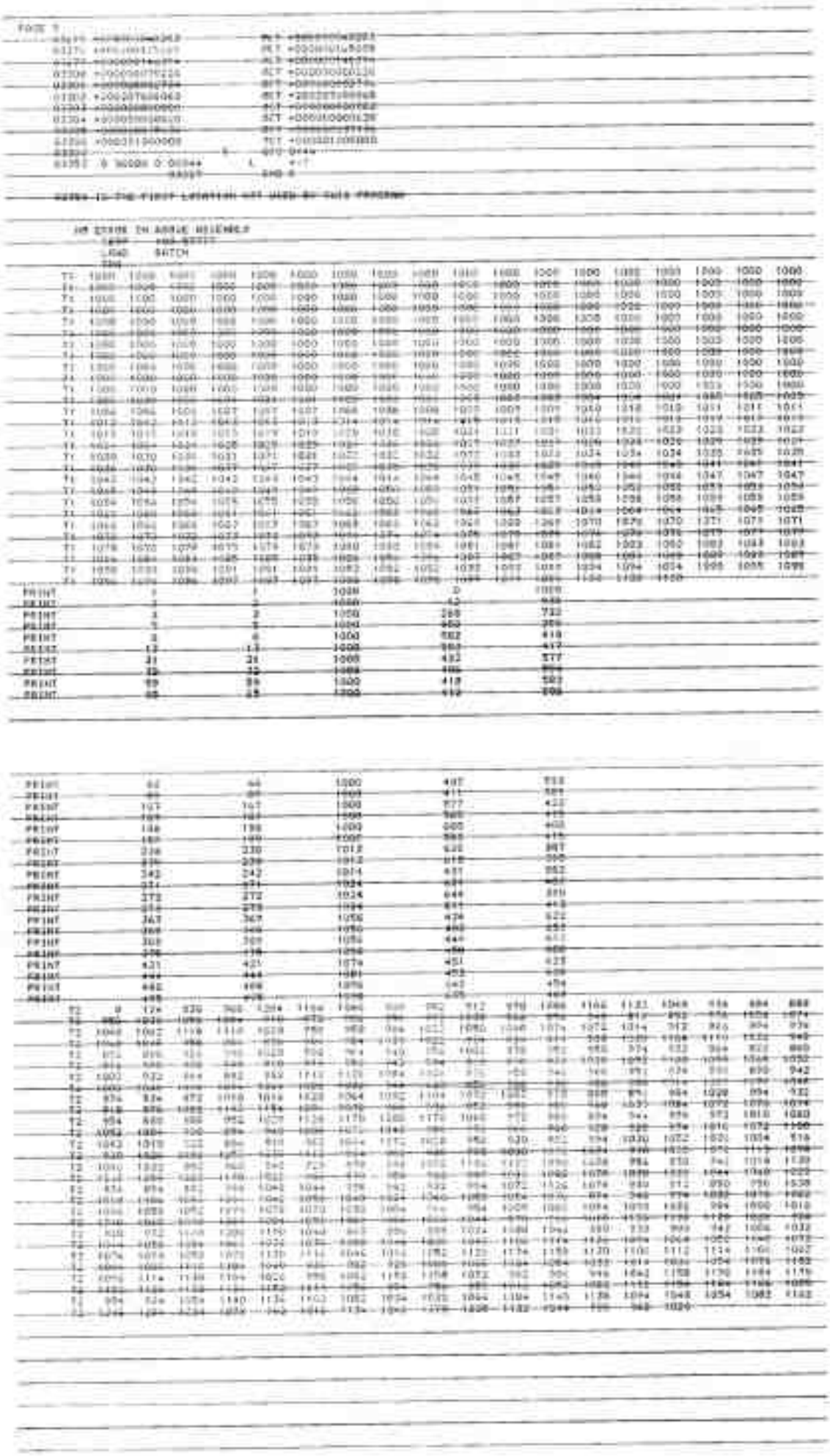


Fig. 3 — Typical BLOD1 program: printed output.

An Acoustic Compiler for Music and Psychological Stimuli

By MAX V. MATHEWS

(Manuscript received November 3, 1960)

A program for synthesizing music and psychological stimuli on a digital computer is described. The sound is produced by three operations: (a) A compiler generates the programs for a set of instruments. (b) These instruments are "played" by a sequencing program at the command of a sequence of "note" cards which contain information analogous to that given by conventional music notes. (c) The computer output, in the form of numbers on a digital magnetic tape, is converted to audible sound by a digital-to-analog converter, a desampling filter, and a loudspeaker. By virtue of the general nature of the compiling program a great variety of instruments may be produced, and the instrument programs are quite efficient in terms of computer time. The "note" cards are arranged to minimize the effort necessary to specify a composition. Preliminary compositions indicate that exceedingly interesting music and useful psychological stimuli can be generated.

I. INTRODUCTION

General translating devices for rapid conversion of numerical data into a continuous analog signal[1] make it possible for a digital computer to produce interesting and useful sounds, among them music. In this way many of the mechanical and acoustic limitations of conventional instruments and sound sources can be overcome. This paper describes the third in a series of programs written for sound production, which achieves a much greater versatility than its predecessors[2] because it includes a compiler* which writes programs for various sound generators or instruments.

Since many who are interested in the musical aspects of this subject may not be familiar with computers, technical descriptions will be minimized and programming details omitted. In addition, it may be helpful to describe briefly the digital-to-acoustic converter to which the process

* A compiler is a program which writes other programs.

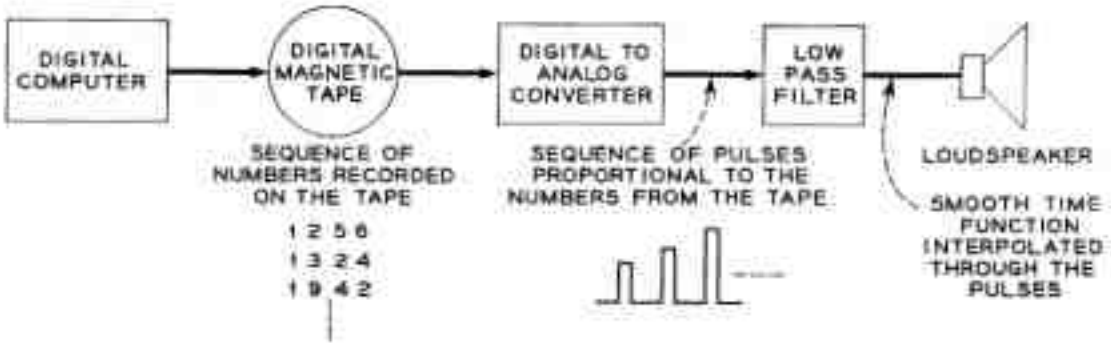


Fig. 1 — Digital-to-acoustic converter.

owes both its existence and generality. The conversion process is schematized on Fig. 1. The computer prepares a magnetic tape on which are written successive digitized samples of the acoustic output. These numbers are then converted by a digital-to-analog converter to pulses whose amplitude is proportional to the numbers. Finally, the pulses are smoothed by a low-pass filter to obtain the excitation for a loudspeaker. The maximum effective sampling rate of the present translator is 20,000 per second, permitting frequencies up to 10,000 cps to be produced.[3] Each sample is reproduced from a four decimal digit integer. Thus, the signal-to-quantizing* noise ratio is greater than 60 db. This ratio is as large as can be conveniently reproduced electronically. Within the limits of this frequency range and this signal-to-noise ratio the converter can theoretically reproduce any sound whatsoever, provided that an appropriate sequence of digital samples can be generated.

II. BASIS OF THE COMPILER

What is the basic objective of this sound generation procedure? It is not simply to produce sounds in the most general way. This generality could be achieved by having the composer list the 20,000 numbers per second which he wished converted to sound. However, such a process is impossibly tedious and, more important, does not effectively control the parameters which determine the psychological impact of the sound on the listener. The basic objective is then to find an economical way of defining and specifying these parameters while automatically supplying the numerical data through which these factors act. In addition, it is a practical necessity to select generating procedures which are economical of computer time.

In order to fulfill these objectives completely a great deal more must

* Quantizing noise is the error introduced by representing a continuous function with a number that can take on only integer values.

be learned about man as a listener. However, the following admittedly incomplete heuristics figured strongly in the design of the compiler:

1. Music can be considered a time sequence of acoustic events which might be called *notes*, although the connotations of this word are grossly inadequate for this usage. Several sequences (voices) are usually added together in all but the simplest pieces.

2. Individual notes are formed from approximately periodic functions. Their most important parameters are period, amplitude, duration, and wave shape. There are, however, notes not fitting this description which are coming into use, for example, those using random noise and those in which the pitch changes greatly over the duration of the note.

3. The ear is sensitive to a number of nuances which must be introduced to obtain interesting timbres. These effects include a wide range of attack and delay characteristics, which strongly affect *timbre*; frequency modulation or *vibrato*; and amplitude modulation or *tremolo*.

The basic form of the generating program is a scheme for producing sequences of sounds on individual instruments, whose outputs can be combined so as to effect several voices. The instruments are formed by combining a set of basic building blocks called *unit generators*, appropriate combinations of which can produce sounds of almost any desired complexity or simplicity. Such an approach has many advantages, the most obvious perhaps being that novel characteristics may be introduced by compiling new instruments. Of equal importance is that composing and computing effort be minimized for simple instruments. The cost of the compiling philosophy is some additional programming and mathematical complexity in forming the instruments and the substantial work (now completed) of writing a compiling program. But this price is small compared to the advantages gained.

The compiling program was greatly simplified by the use of macro instructions, which specify a sequence of computer instructions by means of a single statement. In this way each unit generator can be specified by a single macro statement.

In order to speed the computer operation certain basic functions which specify characteristics such as wave shape, attack, and decay are generated only once and stored in the computer memory, where they serve as references for the unit generators. The functions may be generated by Fortran subprograms.* By utilizing Fortran, a great variety of functions can easily be programmed.

* Fortran is an automatic coding procedure for the IBM 704 and 7090 computers which makes possible the simple generation of most well-known mathematical functions. A subprogram is a subsidiary program.

III. THE MECHANICS OF GENERATION

The first step in producing a musical piece is to punch a set of cards* which specify the instruments in the orchestra. Most of these cards contain a single macro instruction. These instrument cards are then fed into the computer, together with the compiling program, and the computer punches a card deck, which is the music-generating program or *orchestra*.

A sequence of note cards or *score* must now be prepared. These give the parameters such as pitch, duration, and amplitude for the notes which are to be generated. The orchestra, any Fortran subprograms required by it, and the note cards are now inserted in the computer. The numerical samples of the acoustic output are written by the computer on an output magnetic tape. This tape is then converted to a sound via the high-speed data translator[1] and a loudspeaker.

As a convenient alternative, numbers on the output tape may be copied directly from an input tape that is placed on another of the computer tape machines. This tape, for instance, might have been produced by some previous music-making attempt, and this procedure would permit modifications of a composition without regeneration of the entire piece.

Fig. 2 — Unit generator.

IV. COMPILING THE ORCHESTRA

Various kinds of unit generators are available for forming instruments. However, the one used most frequently generates quasiperiodic functions, as typified by sustained notes. This unit is diagrammed in Fig. 2 and produces samples θ_i according to the relations

* Communication between programmer and computer is carried out by punched cards. Up to 80 digits or alphabetic characters may be impressed on each card and read by either man or machine.

$$\theta_i = M_i \cdot F_j([S_i]_{\text{mod } 512}),$$

$$S_{i+1} = S_i + I_i,$$

where i is the index specifying sample sequence. The index i starts at zero at the beginning of each note and terminates at a value determined by the duration of the note. Sequencing, which controls note duration, will be discussed later. The function F_j is defined for an argument x, where $0 \leqq x < 512$ and is one of 20 functions ($j = 1, 2, \cdots, 20$) which may be stored in the computer memory. As illustrated in Fig. 3, $[S_i]_{\text{mod } 512}$ acts as a triangular scanning function which, if I_i is constant, results in θ_i having a period equal to $512/I_i$ samples and a wave shape determined

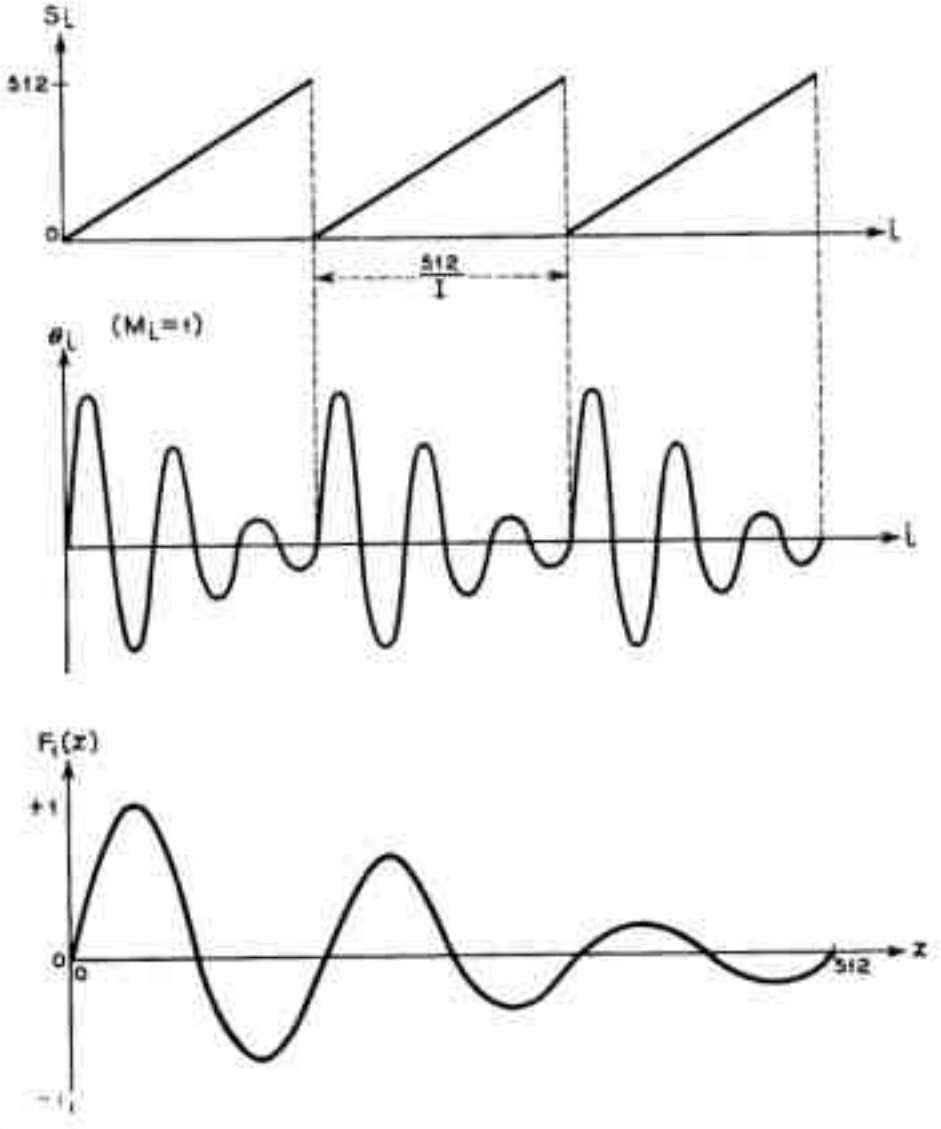


Fig. 3 — Quasiperiodic generation.

by F_j. A varying I_i produces a frequency-modulated output. The generated function is multiplied by M_i, which thus produces amplitude modulation. By this means, attack and decay characteristics can be introduced. To summarize, three functions determine θ_i, these being M_i, I_i, and F_j. In addition, one initial condition, S_8, is involved and is often set to zero at the beginning of each note.

The output θ_i may be added into the final acoustic output. Here the addition defines the process by which outputs of several instruments are combined. On the other hand, θ_i may be used as any input to another unit generator.

The simplest instrument is illustrated in Fig. 4. A periodic note with wave shape determined by F_1, amplitude determined by C1, and frequency by C2 is produced by generator 1U1. The generator 1U2 adds θ into the acoustic output. The attack and decay are instantaneous, which will result in perceptible clicks in the sound.

A more complex instrument with controllable attack and decay may be constructed as in Fig. 5. Here a new generator 2U1 and a function F_2 are added to the structure; 2U1 produces an attack and decay characteristic according to F_2 which amplitude modulates the periodic output of 2U2, while C1 and C2 again specify amplitude and frequency of the note. The new parameter C3 is set so 2U1 generates one period per note. (C3 = 512/duration of note in samples.)

An instrument with attack and decay and vibrato is shown in Fig. 6. Generators 3U1 and 3U2 have been added; 3U2 is an adder whose output is the sum of its inputs. Thus the center frequency of the tone is again specified by C2, the frequency deviation is controlled by C4 and the rate of vibrating (throb rate) by C5, and F_3 determines the wave shape of the frequency variation.

Instruments of even greater versatility can be easily developed by

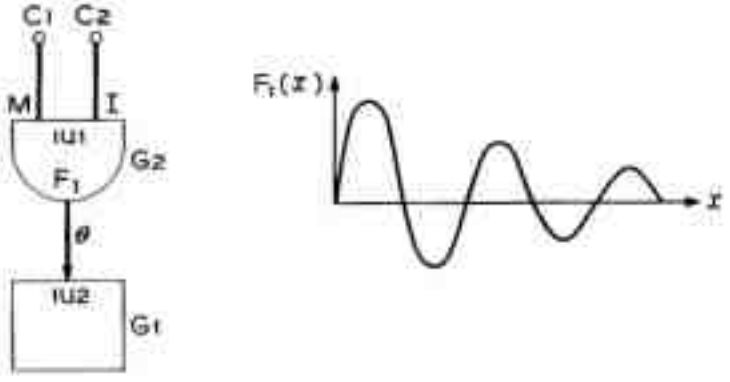


Fig. 4 — Simplest instrument.

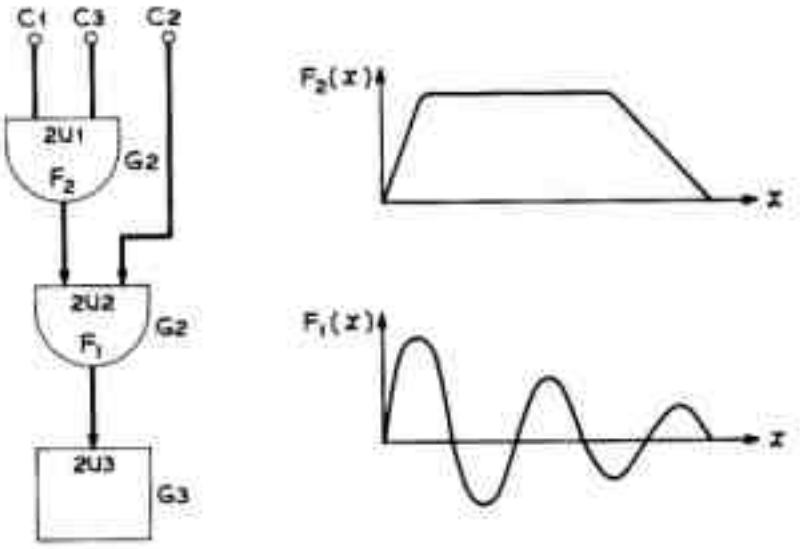


Fig. 5 — Instrument with attack and decay.

putting attacks on the vibrato generator, or adding glissando, or in many other ways. A list and brief description of some of the unit generators which may be used is included in the Appendix.

The punching of the cards from which the instruments are compiled can be illustrated as in Table I, using the cards of I3, the instrument in Fig. 6.

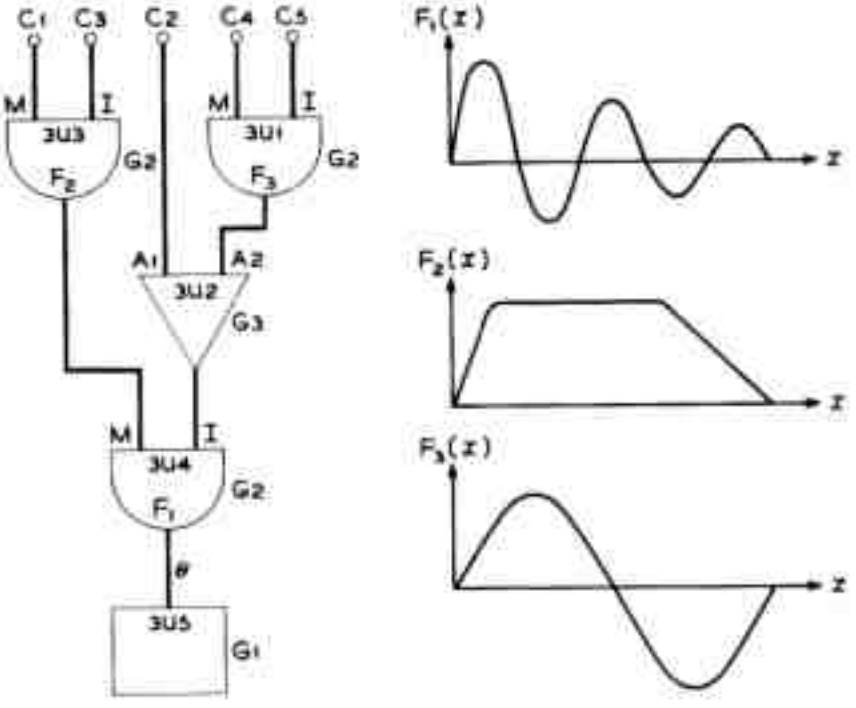


Fig. 6 — Instrument with attack, decay, and vibrato.

TABLE I — DEFINITION OF AN INSTRUMENT

Card Columns		
8 9 10	...	16 ... 72
M A C		G2, 3U1, F3, (3U2,A2), x, x
M A C		G3, 3U2, (3U4,I), x, x
M A C		G2, 3U3, F2, (3U4,M), x, x
M A C		G2, 3U4, F1, (3U5,$\emptyset$), x, x
M A C		G1, 3U5
M A C		S1, I3, ((3U3,M,C1)(3U3,I,C3) $
E T C		(3U2,A1,C2)(3U1,M,C4)(3U1,I,C5) $
E T C		(3U3,S,PO))

Comments concerning this example:

1. The MAC is a general title designating a macro instruction. Each of the first five macros specifies one unit generator.

2. The Gn in columns 16 and 17 specifies the type of generator, G1 being an output unit, G2 a semiperiodic generator, and G3 an adder.

3. The nUm (3U1 for example) designates the instrument number by n and the number of the unit generator in the instrument by m. Each instrument must have a different number, and the unit generators are numbered sequentially in each instrument.

4. The rest of the unit generator designation varies depending on the type of generator, but in general it specifies where the output of the generator is placed and provides the option of specifying inputs to the generator as constants. Thus (3U2,A2) on the first card shows that the output of 3U1 forms the A2 input of 3U2; F3 indicates that function F_3 is called on by the generator and the terminating x's allow the possibility of providing fixed inputs. For example, if the vibrato frequency were fixed at, say, 8 cps instead of being varied by C5, then a constant equal to $8 \times 512/10{,}000 = 0.4096$ (assuming a 10,000 sample-per-second rate) could be included by the statement:

MAC G2, 3U1, F3, (3U2,A2), x, 0.4096B17*

By specifying all fixed constants in the instrument definitions, the number of parameters which must be written in the score is minimized.

5. In the computer the computation of the sample proceeds from one generator to the next in the order in which they are listed on the cards. Hence, for example, if 3U2 uses as an input the output of 3U1, then 3U2 must be listed after 3U1. In addition, to execute the program, the first

* The B17, appended to the number, specifies the decimal point in the computer memory.

generator in each instrument must be number one (*n*U1) and the last generator must be of a terminating type, G1.

6. The final macro MAC S1, I3, ··· compiles instructions which set the parameters C1 through C5 at the beginning of each note. The values of these parameters are obtained from the score in a manner which will be discussed later. The macro can be interpreted in the following way: S1 is the name of the macro which is concerned with setting parameters, I3 refers to instrument 3. Each of the subparentheses sets one parameter; for example, (3U3,M,C1) means set the M input of 3U3 to the value determined by the C1 conversion function. As many subparentheses as desired may be inserted in the macro. If necessary several cards may be used by terminating each card with $ and starting the next card with ETC. The final subparenthesis (3U3,S,P0) uses a special parameter P0 which is zero and is used to set the initial value of S in 3U3 to zero. Although it is not done in this instrument, the function to which a given generator refers can also be set by the score. For example, (3U4,F,P6) would cause the function number of 3U4 to be set equal to the sixth parameter on the note cards of the score.

7. The control of the instruments may be summarized by a rule which says that each input or parameter of the unit generators must be either (a) the output of some other generator, or (b) defined as a constant, or (c) set from the score at the beginning of each note. This rule can be used as a check on the correctness of the instrument compilation.*

V. WRITING THE SCORE

After the orchestra program has been compiled it is inserted into the computer, together with any necessary subprograms and the score. The score is also punched on cards, which perform one of four general functions. These are to control the Fortran subprograms for generating the F_j functions, to set the time scale or tempo of the piece, to punctuate the piece with measures and a termination, and to specify the sequence of notes and rests.

Usually each note is specified by one card which gives the duration of the note, the instrument on which it will be played, and all parameters required by the instrument. The card has the format shown in Fig. 7. The OP code, consisting of three alphabetic letters in the first three columns, specifies the function of the card. For example, RST means

* An algorithm to check the correctness could be included in the compiler but has not yet been developed.

rest and a blank (unpunched) OP code indicates a note. The blank code was chosen to save effort, since note cards are by far the most frequently used. The remainder of the card contains space for up to 12 numbers, P1 through P12. P1 gives the instrument number and P2 the duration. The sequence of notes for each instrument is determined by the sequence of note cards. Rests may be inserted between the notes where desired. The note sequences for each instrument are treated separately, so that note cards for different instruments may be interleaved. Thus if two notes are to be sounded together it is only necessary that the sum of the durations of the preceding notes and rests for each of the two instruments be equal.

The control of sequence by note duration introduces the possibility of a duration error in one note causing all subsequent notes in that instrument to be incorrectly positioned with respect to the other instruments. To mitigate this penalty the composition is also divided into

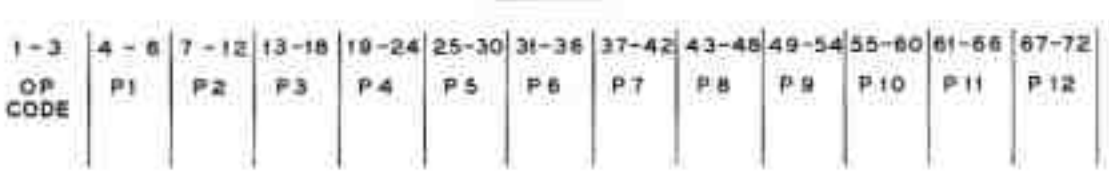


Fig. 7 — Score card.

arbitrary units called *measures*. A measure can contain any number of notes up to a maximum of 100. Durations are always computed from the beginning of the current measure, so that a mistake will affect only one measure. In general, the end of the measure marks a break in the notes of all instruments. However, special provision, by means of a slur, has been made for the rare cases where a note must be carried over from the end of one measure to the beginning of the next.

The significance of the numbers P3 through P12 on the cards depends on the particular instrument. The parameters defined by the instrument (C1 through C3 for Fig. 5 example) refer to another set of Fortran subprograms (called CVTO1 through CVTO3). Each of these subprograms can use any or all of the numbers P2 through P12 as arguments of a function to compute one parameter inserted into the instrument. These subprograms can be any of the enormous variety of functions that are specifiable by Fortran; thus exceedingly flexible conversion is possible. The additional complexity of this conversion between score-card parameters and instrument parameters is justified because it allows the composer to write in psychologically meaningful numbers. The burden of

converting from psychological to physical parameters is then carried by the computer.

The types of conversions which are usually employed, as well as the details of a score, are probably best presented by a short but liberally annotated example.

Suppose we wish to generate two measures which in conventional music notation would be written

with instrument 1 (Fig. 4) playing the upper voice and instrument 2 (Fig. 5) playing the lower voice.

Before proceeding we must decide what wave shape function F1 and what attack and decay function F2 we desire and obtain subprograms to generate these. It is unfortunately beyond the scope of this paper to discuss Fortran programming, so for the present let us consider that we have purchased two subprograms, say GEN10 and GEN11, from some competent Fortran programmer. These, when called upon by the score, will produce the damped sinusoid and the attack function illustrated on Fig. 5.

We will also need to obtain from this programmer three conversion functions CVTO1, CVTO2, and CVTO3 with which to set the parameters in our instruments, and we may well choose these functions so as to simplify our task of score writing. The function CVTO1 sets the amplitude of the note, and it is desirable to write the score in a logarithmic rather than linear scale, since the former much more closely approximates the ear's loudness scale. Hence, let us request that

$$\text{CVTO1} = 10^{P3/20}$$

and use P3 as amplitude control in decibels.

Assuming that the composition is to be played with an even-tempered frequency scale, we can easily obtain a conversion which will let us write frequencies in the form 2.0 through 2.11, where the 2 refers to the octave and the .0 through .11 to the 12 tones within the octave. For this purpose

$$\text{CVTO2} = \frac{512.0}{10{,}000.0} \times 32.70 \times 2^{(\text{I[P3]}+\text{F[P3]}/0.12)},$$

where I[P5] means the integer part of P5, F[P5] means the fractional part of P5, the sampling rate is 10,000 per second, and P5 = 0.0 refers to C three octaves below middle C having a frequency of 32.70 cps. Middle C, for example, would be 3.0 and A above middle C, 3.9.

The remaining conversion function CVTO3 causes the attack generator 2U1 to produce one cycle per note, and thus must be

$$\text{CVTO3} = \frac{512.0}{\text{P2}}.$$

Notice that this function operates on the duration P2 and requires no additional parameters on the note card.

Having obtained the necessary Fortran functions, we can now write the score as shown in Table II.

Comments concerning this example:

1. These two cards cause functions F_1 and F_2 to be generated. P1 determines the generating subroutine to be called and P2 the function to be generated. If desired, P3 through P12 may be used as parameters by the generating routine, although such was not done here.

2. This card sets the time scale so that a P2 of 1 produces 1000 samples, or one-tenth second at a 10,000 sample-per-second rate. This is the duration of an eighth note. The time scale can be reset at any point in the composition and is reduced to 750 for the second measure to accomplish the accelerando.

3. The initial rest for instrument 2 is produced by this card. How-

TABLE II — EXAMPLE OF A COMPOSITION

OP Code	P1	P2	P3	P4	P5	Comments (numbers refer to comments in text)
GEN	10	1				1
GEN	11	2				1
TME			1000			2
RST	2	1				3
	1	3	50		3.9	4
	1	2	55		3.4	4
	2	2	53		3.0	4
MES						5
TME			750			
RST	2	1				6
		4	60		3.2	6
	1	2			3.4	
	1	3			3.7	
MES						
TER						7

ever, no cards are needed for the terminal rest at the end of the first measure for instrument 2. The length of the measure is defined as the maximum sum of the durations of the notes and rests for any instrument. Automatic rests are inserted for instruments not played and between the last note of any instrument and the end of the measure.

4. These cards produce the three notes in the first measure. The instrument numbers and durations are given by P1 and P2. The amplitudes, given in decibels by P3, are arranged to effect a crescendo as requested by the score. The frequencies are specified by P5 in the 12-tone units previously defined.

5. This card terminates the measure.

6. These cards play instrument 2 in the second measure. Notice that the instrument number, P1, is not repeated on the second card. An automatic repeating feature is built into the score-reading program, so that if any parameter is left blank it is repeated from its previous value. Thus, for example, the amplitude of 60 db (P3) is not punched on the last two note cards. This feature is of great value in eliminating a quantity of redundant parameters which otherwise would have to be copied from card to card.

7. This card terminates the composition. It must be preceded by a MES card in order to generate the second measure.

This example provides at least a brief illustration of most of the functions of the score. The two most important omissions are the slur OP code which enables a note to be continued from one measure to the next and the "set" OP code which allows parameters from several cards to set one instrument. These are infrequently used functions and do not justify the space necessary to describe them adequately.

VI. SPEED OF COMPUTION

The time required to synthesize a piece of music depends directly on the number of unit generators involved and the number of samples in the piece. Thus, simple instruments can produce samples rapidly, complex instruments more slowly. For example, on the IBM 7090, the Fig. 4 instrument with two unit generators requires about 0.2 millisecond for each sample. With a rate of 20,000 samples per second, 4 seconds of computer time would be needed to generate each second of music. The Fig. 6 instrument requires 0.5 millisecond per sample, or 10 seconds computer time per second music. The computation cost to produce the music may typically come to as much as $100 per minute of music. Fortunately, computation costs are steadily decreasing.

VII. SOME PROGRAMMING DETAILS

For the benefit of programmers who may be interested, the operation of the compiling and score reading programs will be outlined very briefly. The compiling program at present consists of a symbolic assembly program for the IBM 7090 which has provisions for macro instructions. The instruments are closed subroutines assembled from these instructions. Consequently, it is quite possible to insert basic machine language instructions into the instruments simply by interspacing these with the macros. This ability to fall back on basic machine language is always desirable in a compiler.

The first instruction in any instrument bears the symbolic address nU1, where n is the instrument number. In playing the instrument, control is transferred to this point by the main program. Control is returned to the main program by the last unit generator (because of its type, G1) after one sample of acoustic output has been generated.

At the beginning of each note certain parameters in the instrument must be set. This is done by another closed subroutine, called the "setter," which is assembled along with the instrument. This subroutine delivers the parameters of the note card to the appropriate Fortran subprograms and stores the parameters computed by these programs in the instruments. The flexibility of the macro compiler is essential here, since various numbers and various types of parameters must be accommodated for the different instruments. The first instruction in the "setter" subroutine is designated In, where n is the instrument number.

The main score-reading and sound-generating program is assembled along with the instruments, so that all symbolic addresses are common to both. The main program operates in two phases, the first of which is a card-reading phase, which is terminated by a MES measure card. All the note cards in the first measure are read and their parameters stored in memory.

At the end of a measure, a sorting process must be carried out to put all the note cards into the time sequence in which the events occur in the measure, since time sequence with several instruments does not necessarily correspond to card sequence. After the sort, the setting subroutines are called to set the parameters in the instruments playing in the first time interval, and all these instrument subroutines are called N times, where N is the number of samples in the first interval. The process is repeated for the second interval, etc., until the measure is completed.

The cards for the following measure are then read, and the cycle continues until a termination card, TER, ends the composition.

The program, as written, is a compromise between efficiency, flexibility, and simplicity. The writing time was about a man-month, which is short for a compiler. This speed is attributable to the versatility of the macro assembly program. By using stored functions and setting instrument parameters only at the beginning of notes, a rather efficient program was achieved. The flexibility rests mainly on the ease with which new unit generators can be defined with new macros, and on the possibility of inserting any desired machine language instructions. So far, the program seems adequate for its objectives.

VIII. RESULTS AND CONCLUSIONS

The program has been used to generate a wide variety of sounds and sequential signals. These include musical compositions; sets of test tones to study attack, decay, and vibrato; control signals for a speaking machine; stimuli for a study of absolute pitch perception; and a set of random signals of various bandwidths and frequencies for listening tests.

The musical compositions demonstrated both the facility with which scores can be written and the range of sounds which can easily be produced. The most striking effects are the continuous modulation from one instrument type into another, precisely controlled vibratos with attack and decay of the vibrato rates, the rapid sweep of frequency over many octaves in a single note, frequency as well as amplitude attacks on notes, and the representation of a melodic line by the sum or difference of the frequencies of two voices.

The compositions affirmed that a deeper understanding of how sounds are perceived is necessary before we can effectively use the new instruments that can be compiled. However, the program itself is proving an excellent tool in carrying out studies. For example, a systematic variation of attack and decay times showed the predominant influence of attack in timbre. Similar examinations were carried out for vibrato and random signals.

The program proved a convenient way of producing a random sequence of 12-tone chords in which the notes in a chord were all in octave relationship. These chords were used as psychological stimuli to test the feasibility of teaching absolute pitch. Other applications for generating psychological stimuli have been suggested.

The control signals to a speech synthesizer are fundamentally functions of time which must be flexibly specified. The music program, although conceived for producing a sequence of notes, proved ideal for this purpose. It probably would also be possible to synthesize an entire speaking machine as a complex instrument.

An important question is, "How easily can the professional composer make use of the program?" Probably the compilation of the instruments of an orchestra requires programming skill beyond that of most musicians, although a mathematically minded one would easily learn the technique. However, writing a score for an existing orchestra can be systematized to such an extent that almost anyone can create a composition. Consequently, it seems quite feasible for a musician without mathematical training to carry out his wishes with the aid of only a little programming help.

The compiler has great inherent flexibility in that new unit generators can easily be added to the group available for compiling instruments. An example is the random signal generator which was only recently programmed. In addition, the Fortran subprograms contribute to the versatility. Because of this ability to change and grow, we believe the compiler will be valuable for the production of computer music and stimuli for some time to come. We expect programs such as this, together with the cheaper, faster computers which are promised, to result in computer-generated sounds becoming of increasing utility.

APPENDIX

Unit Generators for Acoustic Compiler

A brief description of the most frequently used unit generators is given here. New generators are often added, so the list is not complete.

Output Unit — G1.

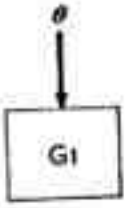
Call statement:

MAC G1, nUm.

Input designation: θ.

Function: To add the number stored in θ to the acoustic output and transfer control from the instrument to the main sequencing program. An output unit *must* form the last generator in every instrument and must not be used in any other position.

Diagram:



Periodic Function Generator — G2.

Call statement:

MAC G2, nUm, A, (pUq,B), C, D.

A = designation of stored function F_j;
(pUq, B) = location of output;
C = fixed designation of M input;
D = fixed designation of I input.

Input designation:

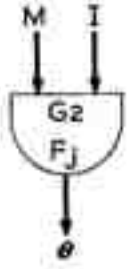
M = amplitude modulation input;
I = frequency control input;
S = initial value of S_i.

Function:

$$\theta_i = M_i F_j[(S_i)_{\text{mod } 512}];$$

$$S_{i+1} = S_i + I_i.$$

Diagram:



Adders — G3, G4, G5.

Call statement:

MAC G3, nUm, (pUq,B), C, D

MAC G4, nUm, (pUq,B), C, D, E

MAC G5, nUm, (pUq,B), C, D, E, F.

(pUq,B) = location of output;
C, D, E, F = fixed designation of inputs A1, A2, A3, A4 respectively.

Input designation: A1, A2, A3, A4

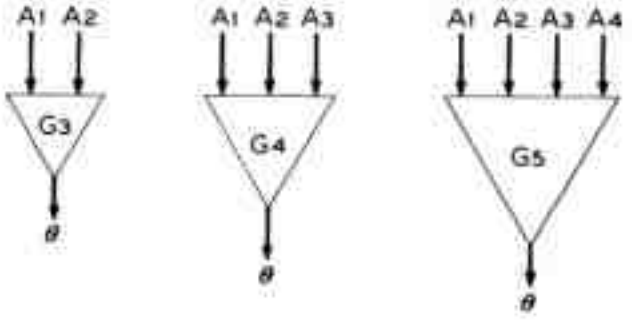
Function:

G3 θ = A1 + A2

G4 θ = A1 + A2 + A3

G5 θ = A1 + A2 + A3 + A4

Diagram:



Random Signal Generator.

Call statement:

$$\text{MAC} \quad \text{RAND}, n\text{U}m, \text{A}, \text{B}, p\text{U}q, \text{C}.$$

A = fixed designation of M input;
B = fixed designation of I input;
pUq, C = location of output.

Input Designation:

M = amplitude modulation input;
I = frequency control input.

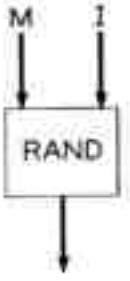
Function:

$$\theta_i = M_i R_i[I_i],$$

where $R_i[I_i]$ is a random variable whose bandwidth is controlled by I_i. The bandwidth is approximately equal to

$$\frac{\text{sampling rate}}{2} \times \frac{I_i}{512}.$$

Diagram:



REFERENCES

1. David, E. E., Jr., Mathews, M. V., and McDonald, H. S., A High-Speed Data Translator for Computer Simulation of Speech and Television Devices, Proc. Western Joint Computer Conf., March 1959.
2. Mathews, M. V., and Guttman, N., Generation of Music by a Digital Computer, Proc. Third International Congress on Acoustics, 1959, Elsever Publ. Co., Amsterdam, to be published.
3. Shannon, C. E., A Mathematical Theory of Communications, B.S.T.J., **27**, 1948, pp. 379, 623.

Minimum Noise Figure of the Variable-Capacitance Amplifier

By K. KUROKAWA and M. UENOHARA

(Manuscript received November 2, 1960)

The variable-capacitance diode is one of the most promising nonlinear elements for low-noise parametric amplifiers. In practice, however, these diodes have a small series resistance, and this limits the minimum obtainable noise figure; for the better diodes, the effect of the shunt conductance can be neglected. Taking the contribution of this series resistance into account, this paper discusses the minimum noise figure of parametric amplifiers under various conditions. It is shown that the minimum noise figures are basically determined by a dynamic quality factor of the diode, which will be defined in this paper, under the assumed model of a series resistance as the only parasitic element. Identical minimum noise figures are obtained for lower sideband amplifiers operated with optimum idler frequency, for those with the idler load at 0°K, and for the upper sideband up-converter. In terms of the over-all systems noise figure, however, the lower sideband amplifier is superior to the upper sideband up-converter, for here the gain is limited by the ratio of output to input frequency.

Experimental values are given for the figure of merit of various diodes. Universal curves are also given which demonstrate noise behavior of the various systems as a function of the network parameters and component temperatures.

I. INTRODUCTION

The variable-capacitance parametric amplifier is of interest primarily because it shows promise of very low noise amplification. However, variable capacitance diodes have a small but finite series resistance which limits the obtainable minimum noise figure; for the better diodes, the shunt conductance can be neglected.

Taking the contribution of this resistance into account, Leenov[1] has discussed the noise figure of upper sideband up-converter, and Haus and Penfield[2] and others[3,4,5] have discussed the lower sideband circulator-type amplifier. This paper discusses the lower sideband idler output

amplifier and the degenerate amplifier, in addition to the amplifiers mentioned above, and compares them to one another. Further, we shall demonstrate the unique importance of the dynamic quality factor $\tilde{Q}$, as defined here, in characterizing the diode at a given temperature for noise figure considerations.

The diode is assumed to be a series connection of a junction capacitance $C(t)$, which is a periodic function of time, and a spreading resistance R_s.

Leenov, as well as Haus and Penfield, used the open-circuit assumption for the unwanted frequencies at the diode junction. However, there are a number of published papers[3,5,6,7,8] in which the short-circuit assumption is used. We shall discuss both of these cases simultaneously, and show that the two assumptions give the same expressions for the noise figure, if the dynamic quality factor $\tilde{Q}$ is redefined for each case.

The following conclusions are obtained:†

(a) The noise figure of a diode amplifier is basically determined by the dynamic quality factor $\tilde{Q}$ of the diode; the noise figure improves with increasing value of $\tilde{Q}$.

(b) The minimum noise figure for room temperature operation of the lower sideband amplifiers is obtained when the idler load resistance approaches zero.

(c) For a given $\tilde{Q}$ at the signal frequency, there exists an optimum idler frequency which gives the smallest noise figure for the particular diode.

(d) Refrigeration of the idler load improves the noise figure only under certain conditions. As long as the idler frequency is lower than the optimum, the noise figure of the amplifier with an idler load at zero temperature absolute can be as low as that obtained using the optimum idler frequency.

(e) The minimum noise figure of the upper sideband up-converter is equal to that of the lower sideband amplifiers with the optimum idler frequency or with a zero temperature idler load.

The conclusions (b) and (c) are similar to those reached by Haus and Penfield,[2] Kotzebue,[4] and Knechtli and Weglein,[5] but are extended in this paper to include the case of lower sideband idler output amplifiers. In this case, to obtain the minimum noise figure, most of the idler power generated by the parametric action is to be dissipated in the series resistance of the diode. The output power, however, can be finite because of the unlimited gain obtainable with the negative resistance effect.

† Similar conclusions have been obtained by R. P. Rafuse of M.I.T. (private communication).

The conclusion (e) states further that, if the same diode is to be employed, a superior over-all noise figure performance is obtained with the lower sideband amplifier, inasmuch as the gain of the lower sideband amplifier can be larger than that of the upper sideband one. Nevertheless, there are other significant applications of the upper sideband upconverter which make its study important.

Two aspects of minimum noise amplification, not covered in the literature, are discussed in detail. These are

(a) the effects of load refrigeration;

(b) optimum upper sideband construction.

Universal curves are included which demonstrate noise behavior of the various systems as a function of $\tilde{Q}$.

II. EQUIVALENT CIRCUIT OF THE LOWER SIDEBAND AMPLIFIER

If the assumption is made that the only currents which flow through the diode junction are at the signal (ω_1) and idler (ω_2) frequencies, that is, making the open-circuit assumption for the unwanted frequencies, the junction is characterized by the following equation:

$$\begin{bmatrix} e_1 \\ e_2^* \end{bmatrix} = \begin{bmatrix} \dfrac{1}{j\omega_1 K_0} & -\dfrac{1}{j\omega_2(2K_1)} \\ \dfrac{1}{j\omega_1(2K_1)} & -\dfrac{1}{j\omega_2 K_0} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2^* \end{bmatrix}, \tag{1}$$

where e is the junction voltage, i is the junction current, and their subscripts 1 and 2 refer to the signal (ω_1) and idler (ω_2) frequencies respectively.

The quantities K_0 and K_1 are defined by

$$\frac{1}{C(t)} = \frac{1}{K_0} + \frac{1}{K_1}\cos \omega_p t + \cdots, \tag{2}$$

where $C(t)$ is a junction capacitance which is a periodic function of time.

If, on the other hand, the assumption is made that the only voltages which appear across the junction are at the signal and idler frequencies, that is, making the short circuit assumption for the unwanted frequencies, the junction is characterized by

$$\begin{bmatrix} i_1 \\ i_2^* \end{bmatrix} = \begin{bmatrix} j\omega_1 C_0 & -j\omega_1 \dfrac{C_1}{2} \\ j\omega_2 \dfrac{C_1}{2} & -j\omega_2 C_0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2^* \end{bmatrix}, \tag{3}$$

where C_0 and C_1 are defined by

$$C(t) = C_0 - C_1 \cos \omega_p t + \cdots . \tag{4}$$

If both sides of (3) are multiplied by the inverse of the two-by-two matrix, (3) becomes

$$\begin{bmatrix} e_1 \\ e_2^* \end{bmatrix} = \begin{bmatrix} \dfrac{C_0}{j\omega_1\left(C_0^2 - \dfrac{C_1^2}{4}\right)} & -\dfrac{\dfrac{C_1}{2}}{j\omega_2\left(C_0^2 - \dfrac{C_1^2}{4}\right)} \\ \dfrac{\dfrac{C_1}{2}}{j\omega_1\left(C_0^2 - \dfrac{C_1^2}{4}\right)} & -\dfrac{C_0}{j\omega_2\left(C_0^2 - \dfrac{C_1^2}{4}\right)} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2^* \end{bmatrix}. \tag{5}$$

The diagonal terms in the impedance matrices of (1) and (5) are just ordinary capacitances in series with the variable capacitance and, therefore, play no essential role in the parametric process. These capacitances are included in the input and output circuits, which are expressed by the impedances Z_{11} and Z_{22} respectively.

The series resistance R_s is considered as the only dissipative element of the diode,† and the dynamic quality factor of the diode is defined by

$$\tilde{Q} = \frac{1}{\omega \mid 2K_1 \mid R_s}$$

(for open-circuit assumption), (6)

$$\tilde{Q} = \frac{\dfrac{\mid C_1 \mid}{2}}{\omega \left| C_0^2 - \dfrac{C_1^2}{4} \right| R_s} = Q_0 \frac{1}{\left(\dfrac{2}{\gamma} - \dfrac{\gamma}{2}\right)}$$

(for short-circuit assumption) (7)

respectively, where‡

$$\gamma = \frac{\mid C_1 \mid}{C_0} \tag{8}$$

and Q_0 is the ordinary quality factor of the diode defined by

$$Q_0 = \frac{1}{\omega C_0 R_s} . \tag{9}$$

† For low-noise gallium arsenide diodes this assumption is good up to X-band, but for silicon mesa-type diodes a shunt conductance has to be taken into account at frequencies around this band.

‡ Our γ is twice the γ used by Kotzebue.[4]

If $C(t)$ is sinusoidal, an expression for the dynamic quality factor for the open-circuit assumption can be obtained in terms of γ and Q_0 :

$$\tilde{Q} = Q_0 \frac{1 - \sqrt{1 - \gamma^2}}{\gamma \sqrt{1 - \gamma^2}}. \tag{10}$$

When γ is small, the dynamic quality factor simplifies to

$$\tilde{Q} = \frac{\gamma}{2} Q_0 \tag{11}$$

for both assumptions, but when γ approaches unity, the open-circuit assumption gives a larger value for $\tilde{Q}$, as shown in Fig. 1.

If, on the other hand, $1/C(t)$ is sinusoidal, the same value of $\tilde{Q}$ is obtained for both short- and open-circuit assumptions.

Using these two definitions of the dynamic quality factor, the same equivalent circuit is obtained for both assumptions, as shown in Fig. 2. In this figure, E_1 is the open-circuit voltage of the input generator.

A simple calculation shows that Fig. 2 can be replaced by the more convenient form of Fig. 3, where the whole circuit is, in effect, completely separated into two parts, the ω_1 circuit and the ω_2 circuit.

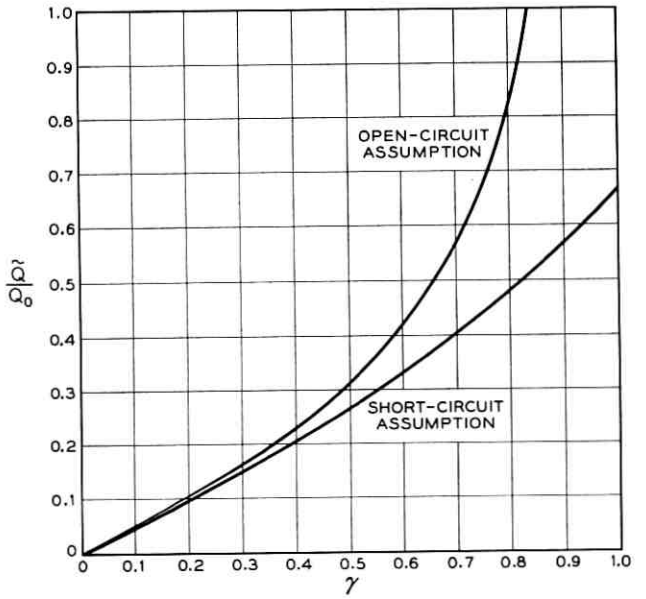


Fig. 1 — Q/Q_0 vs. γ.

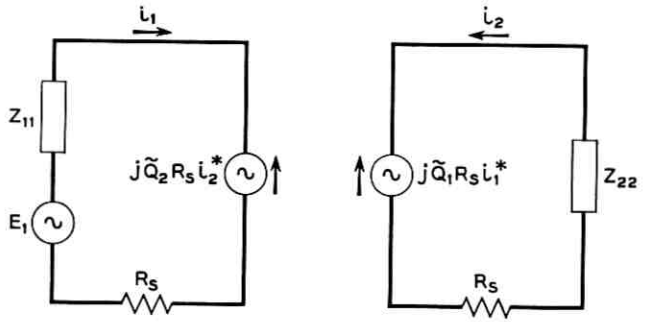


Fig. 2 — Circuit representation of lower sideband amplifier.

The real part of the input impedance of the diode at ω_1 is

$$R_s + \text{Re}\left(-\frac{\tilde{Q}_1\tilde{Q}_2 R_s^2}{R_s + Z_{22}^*}\right).$$

The largest magnitude of the negative resistance is obtained when Z_{22} is equal to zero, and the resistance is

$$R_s(1 - \tilde{Q}_1\tilde{Q}_2).$$

It is worth noting that when $\tilde{Q}_1\tilde{Q}_2$ becomes unity the diode no longer shows negative resistance and the amplifier ceases to show gain.

III. MINIMUM NOISE FIGURE FOR LARGE GAIN

From Fig. 3 the output power at the idler frequency is easily calculated. The result is

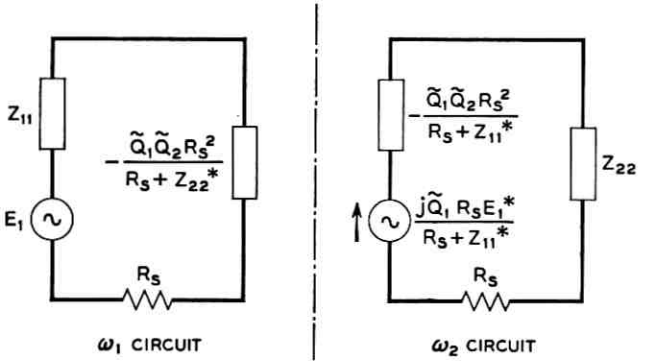


Fig. 3 — Equivalent circuit of lower sideband amplifier

$$P_{21} = \frac{\frac{R_L}{R_s^2}\tilde{Q}_1^2 \mid E_1 \mid^2}{\left|\left(1 + \frac{Z_{11}^*}{R_s}\right)\left(1 + \frac{Z_{22}}{R_s}\right) - \tilde{Q}_1\tilde{Q}_2\right|^2}, \tag{12}$$

where R_L is the load resistance and is a part of Z_{22}.

Since the available power of the source is

$$P_{av} = \frac{\mid E_1 \mid^2}{4R_g}, \tag{13}$$

where R_g is the internal resistance of the generator, the gain becomes

$$G_{21} = \frac{4\frac{R_gR_L}{R_s^2}\tilde{Q}_1^2}{\left|\left(1 + \frac{Z_{11}^*}{R_s}\right)\left(1 + \frac{Z_{22}}{R_s}\right) - \tilde{Q}_1\tilde{Q}_2\right|^2}. \tag{14}$$

Equation (14) is the gain of the amplifier whose input is at ω_1 and output at ω_2.

The noise voltage in the ω_1 circuit is given by

$$\mid e_{n_1} \mid^2 = 4kB(T_gR_g + T_1R_1 + T_sR_s), \tag{15}$$

where $R_g + R_1$ is the real part of Z_{11}; k is the Boltzmann constant; B is the bandwidth; and T_g, T_1, and T_s are the temperatures of R_g, R_1, and R_s, respectively.

Substituting (15) in place of $\mid E_1 \mid^2$ in (12), the noise output due to this noise voltage becomes

$$N_{21} = \frac{4kB(T_gR_g + T_1R_1 + T_sR_s)\frac{R_L}{R_s^2}\tilde{Q}_1^2}{\left|\left(1 + \frac{Z_{11}^*}{R_s}\right)\left(1 + \frac{Z_{22}}{R_s}\right) - \tilde{Q}_1\tilde{Q}_2\right|^2}. \tag{16}$$

Similarly, the noise output at ω_2 produced by noise sources in the ω_2 circuit is found to be†

$$N_{22} = \frac{4kB(T_LR_L + T_2R_2 + T_sR_s)\frac{R_L}{R_s^2}\left|1 + \frac{Z_{11}^*}{R_s}\right|^2}{\left|\left(1 + \frac{Z_{11}^*}{R_s}\right)\left(1 + \frac{Z_{22}}{R_s}\right) - \tilde{Q}_1\tilde{Q}_2\right|^2}, \tag{17}$$

where $R_L + R_2$ is the real part of Z_{22} and T_L and T_2 are the temperatures

† The assumption is made that the gain of the amplifier is large. If it is small, a more elaborate calculation is necessary.

of R_L and R_2 respectively. Upon combining (14), (16), and (17), the noise figure of the amplifier becomes

$$F_2 = \frac{N_{21} + N_{22}}{G_{21} k T_g B} = 1 + \frac{T_1 R_1 + T_s R_s}{T_g R_g} + \frac{T_L R_L + T_2 R_2 + T_s R_s}{T_g R_g} \frac{\left| 1 + \frac{Z_{11}^*}{R_s} \right|^2}{\tilde{Q}_1^2} . \quad (18)$$

Equation (18) is the noise figure of the lower sideband idler output amplifier. In a similar manner, one can calculate the gain and noise figure of the amplifier whose input and output are both at ω_1 but separated by means of a circulator:†

$$G_{11} \simeq \frac{4 \frac{R_g^2}{R_s^2} \left| 1 + \frac{Z_{22}^*}{R_s} \right|^2}{\left| \left(1 + \frac{Z_{11}}{R_s} \right) \left(1 + \frac{Z_{22}^*}{R_s} \right) - \tilde{Q}_1 \tilde{Q}_2 \right|^2} , \quad (19)$$

$$F_1 \simeq 1 + \frac{T_1 R_1 + T_s R_s}{T_g R_g} + \frac{T_L R_L + T_2 R_2 + T_s R_s}{T_g R_g} \frac{\tilde{Q}_2^2}{\left| 1 + \frac{Z_{22}^*}{R_s} \right|^2} . \quad (20)$$

For both cases, the large-gain condition is given by

$$\left(1 + \frac{Z_{11}}{R_s} \right) \left(1 + \frac{Z_{22}^*}{R_s} \right) = \left(1 + \frac{Z_{11}^*}{R_s} \right) \left(1 + \frac{Z_{22}}{R_s} \right) = \tilde{Q}_1 \tilde{Q}_2 , \quad (21)$$

where we have made use of the fact that $\tilde{Q}_1 \tilde{Q}_2$ is a real quantity. Substituting (21) into (18) and (20), the identical noise figure expression is obtained for both the lower sideband idler output and circulator-type amplifier, namely,

$F =$

$$1 + \frac{T_1 R_1 + T_s R_s}{T_g R_g} + \frac{T_L R_L + T_2 R_2 + T_s R_s}{T_g R_g} \frac{\omega_1}{\omega_2} \frac{R_g + R_1 + R_s}{R_L + R_2 + R_s} . \quad (22)$$

The following discussion, therefore, holds equally well for both cases.

Let us assume first that T_L is equal to T_s and that T_2 is not smaller than T_s. Under these conditions, using the high-gain condition (21) in conjunction with (22), it can be shown that R_g must be as large as possible to minimize the noise figure. Using (21), this requires that R_1, $R_L + R_2$, and the reactive components of Z_{11} and Z_{22} must all be as small as possible.

† The assumption is made that the gain of the amplifier is large. If it is small, a more elaborate calculation is necessary.

When

$$R_L = R_1 = R_2 = \text{Im } Z_{11} = \text{Im } Z_{22} = 0,\dagger \tag{23}$$

(21) becomes

$$1 + \frac{R_g}{R_s} = \tilde{Q}_1\tilde{Q}_2\,, \tag{24}$$

and the noise figure expression simplifies to give the minimum value of

$$F_m = 1 + \frac{\frac{T_s}{T_g}}{\tilde{Q}_1\tilde{Q}_2 - 1}\left(1 + \frac{\omega_1}{\omega_2}\,\tilde{Q}_1\tilde{Q}_2\right). \tag{25}$$

Equation (25) is the minimum noise figure of a lower sideband amplifier with a fixed idler frequency under the assumption of only a series-resistance parasitic element. Equation (25) is a function of ω_2, and there is an optimum idler frequency for which a minimum is obtained. Using the relation

$$\tilde{Q}_2 = \frac{\omega_1}{\omega_2}\,\tilde{Q}_1 \tag{26}$$

and (25), the smallest noise figure $F_{m,m}$ (i.e. optimized impedances and optimized idler frequency) is given by

$$F_{m,m} = 1 + 2\,\frac{T_s}{T_g}\left(\frac{1}{\tilde{Q}_1^{\,2}} + \frac{1}{\tilde{Q}_1}\sqrt{1 + \frac{1}{\tilde{Q}_1^{\,2}}}\right), \tag{27}$$

when

$$\frac{\omega_2}{\omega_1} = \sqrt{1 + \tilde{Q}_1^{\,2}} - 1. \tag{28}$$

In Fig. 4, F_m is plotted versus $\tilde{Q}_1$ for several different values of ω_2/ω_1 under the condition $T_s = T_g$. Fig. 5 gives the corresponding plot of $F_{m,m}$ versus $\tilde{Q}_1$. The optimum value for ω_2/ω_1 versus $\tilde{Q}_1$ is shown in Fig. 6.

IV. IDLER LOAD REFRIGERATION

In the noise figure expression, (25), we have assumed that $T_L = T_s$ and concluded that, to obtain a low noise figure, the idler load re-

† For the lower sideband idler output amplifier, if $R_L = 0$, the numerator of (14) becomes zero. However, by a suitable adjustment of the denominator, we may still have large gain for the limiting case of $R_L \to 0$.

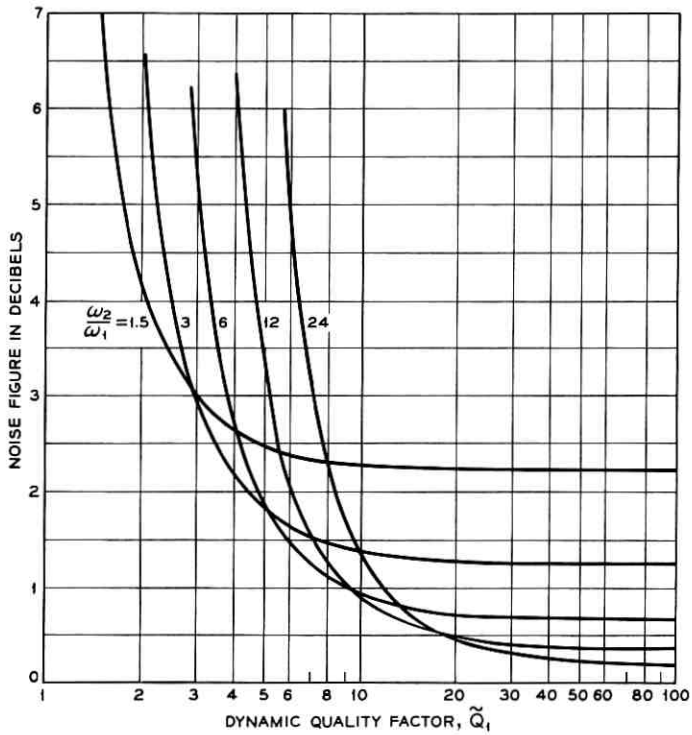


Fig. 4 — Minimum noise figure F_m of lower sideband amplifier with fixed idler frequency.

sistance R_L should be as small as possible. If $T_L < T_s$,† however, one may expect to obtain a further improvement in the noise figure for a fixed idler frequency by properly adjusting the load resistance. We shall now examine this possibility, but again under the assumption that T_2 is not smaller than T_s. From (22), under the condition of constant R_L, the value of the noise figure becomes small when R_g becomes large. Because of the large-gain condition (21), the maximum in R_g is obtained when

$$R_1 = R_2 = \text{Im } Z_{11} = \text{Im } Z_{22} = 0. \tag{29}$$

We shall assume that (29) is satisfied. The condition $R_L = 0$ is now not necessarily the case for the minimum noise figure, since the last

† There are two ways of characterizing T_L. In the treatment presented here, it is used to signify the black body emission back into the circuit (see Section III). It has a second description in terms of the noise figure of the output load which implies $T_L \gg T_s$, but this is not the present definition of T_L.

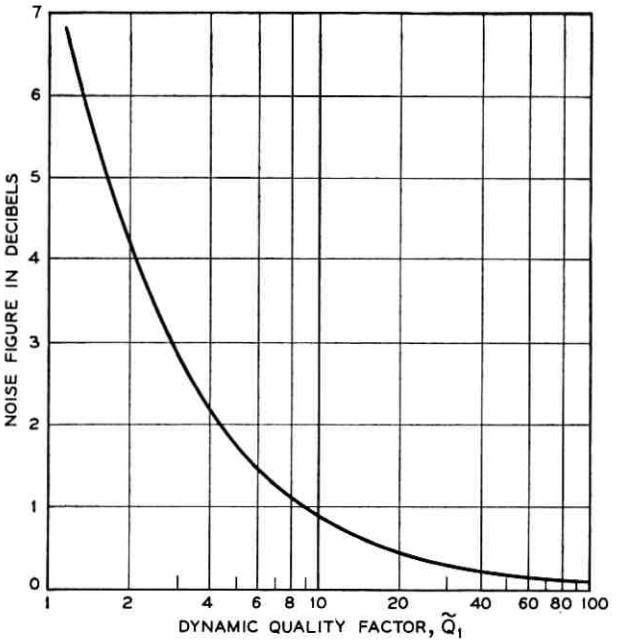


Fig. 5 — Minimum noise figure $F_{m,m}$ of lower sideband amplifier with optimized idler frequency or minimum noise figure F_m of upper sideband up-converter.

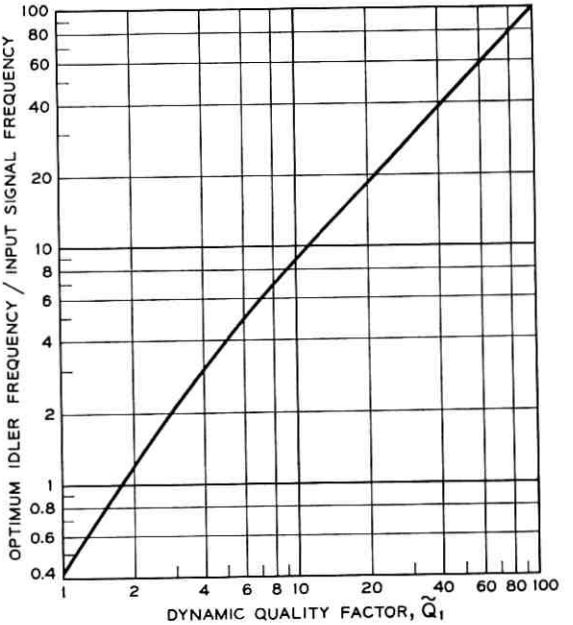


Fig. 6 — Normalized optimum idler frequency.

term of (22) can decrease when R_L increases, provided that $T_L < T_s$. If $T_L < T_s$, from (21) and (22) together with the above assumption, the conditions for the minimum noise figure are given by

$$R_0 = R_s \sqrt{1 + \left(\frac{\omega_2}{\omega_1} + \frac{T_L}{T_s}\right) \frac{\tilde{Q}_1\tilde{Q}_2}{1 - \frac{T_L}{T_s}}} \tag{30}$$

and

$$R_L = R_s \left[\frac{\tilde{Q}_1\tilde{Q}_2}{1 + \sqrt{1 + \left(\frac{\omega_2}{\omega_1} + \frac{T_L}{T_s}\right) \frac{\tilde{Q}_1\tilde{Q}_2}{1 - \frac{T_L}{T_s}}}} - 1 \right]. \tag{31}$$

The noise figure is expressed as

$$F = 1 + \frac{T_s}{T_0} \left[\frac{T_L}{T_s} \frac{\omega_1}{\omega_2} + 2 \frac{1 - \frac{T_L}{T_s}}{\tilde{Q}_1^2} + 2 \frac{1}{\tilde{Q}_1} \sqrt{\left(1 - \frac{T_L}{T_s}\right)\left(1 + \frac{T_L}{T_s} \frac{\omega_1}{\omega_2} + \frac{1 - \frac{T_L}{T_s}}{\tilde{Q}_1^2}\right)} \right]. \tag{32}$$

To obtain a real and positive value of R_L in (31), the condition

$$\frac{\frac{\omega_2}{\omega_1} + 1}{1 - \frac{T_L}{T_s}} < \tilde{Q}_1\tilde{Q}_2 - 1 \tag{33}$$

must be satisfied. If this is not the case, the minimum noise figure (25) is obtained when $R_L = 0$, and is consistent with the results of the previous theory; this will be so when $\tilde{Q}_1\tilde{Q}_2$ is close to unity, or when the load temperature is close to the diode temperature. On the other hand, if $T_L > T_s$, we see from (22) that the noise figure decreases as the idler load R_L is reduced, and the minimum noise figure is again given by (25) when $R_L = 0$. Next, we shall check whether or not the condition (33) is satisfied for the optimum idler frequency.

From (26) and (28), the condition (33) becomes

$$\frac{1}{1 - \frac{T_L}{T_s}} < 1. \tag{34}$$

Because $0 \leqq T_L \leqq T_s$, the condition (34) is not satisfied. This shows

that the effort to get further improvement in noise figure by cooling the idler load resistance proves to be useless after adjusting the idler frequency to the optimum. Finally, if we make $T_L = 0$ in (32), then we have exactly the same expression as (27). In this case, the condition (33) becomes

$$\frac{\omega_2}{\omega_1} < \sqrt{1 + \tilde{Q}_1^2} - 1. \tag{35}$$

Comparing (35) with (28), we conclude that, as long as the idler frequency is less than the optimum, the minimum noise figure obtainable with a zero-temperature idler load is equal to the best that is obtainable by optimizing the idler frequency.

As an example, taking $\omega_1/\omega_2 = 3$, the curves F versus T_L/T_s are shown in Fig. 7 for several different $\tilde{Q}$'s under the condition $T_g = T_s$.

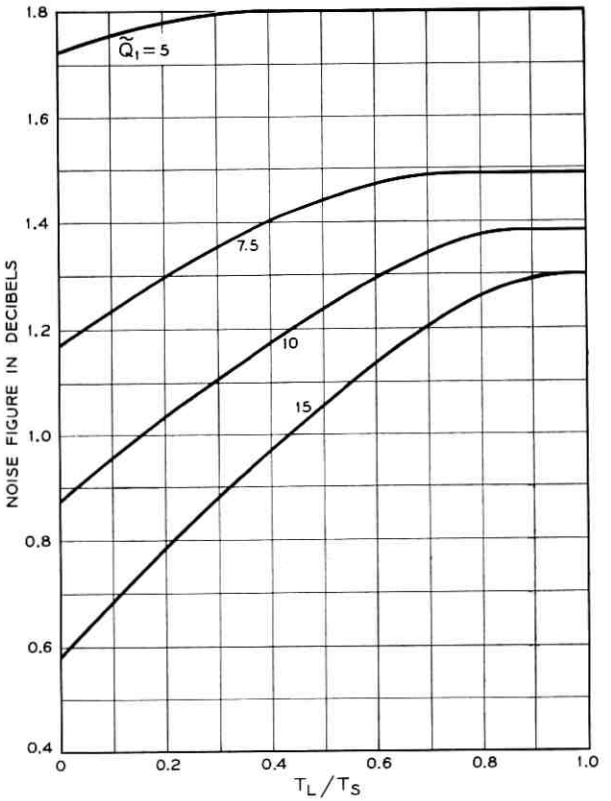


Fig. 7 — Effect of idler load refrigeration. Idler-to-signal frequency ratio of 3 is used for this calculation.

V. DOUBLE SIDEBAND OPERATION

In double sideband operation, the signal is introduced in a substantially symmetrical manner about half the pump frequency so that signal and noise are present both at ω_1 and at ω_2. Therefore, assuming $Z_{11} \simeq Z_{22}$, $\omega_1 \simeq \omega_2$, and $T_L \simeq T_g$, the noise figure expression, derived after some manipulation, is

$$F = \frac{N_{11} + N_{12}}{kTB(G_{11} + G_{12})} = 1 + \frac{T_1 R_1 + T_s R_s}{T_g R_g}. \tag{36}$$

The large-gain condition remains the same as (21). To minimize the noise figure, R_g must be as large as possible and hence

$$R_1 = R_2 = \text{Im } Z_{11} = \text{Im } Z_{22} = 0. \tag{37}$$

Then, (21) becomes

$$1 + \frac{R_g}{R_s} = \tilde{Q}. \tag{38}$$

From (36), (37), and (38), the minimum noise figure F_m is obtained. One finds

$$F_m = 1 + \frac{T_s}{T_g} \frac{1}{\tilde{Q} - 1}. \tag{39}$$

The curve F_m versus $\tilde{Q}$ for $T_s = T_g$ is shown in Fig. 8.

VI. COMPARISON WITH EXPERIMENT

A considerable number of noise figure measurements have been made for double sideband operation at 6 kmc. Some of these results (Table I) are plotted in Fig. 9 as a function of the usual quality factor as measured at zero bias voltage.†

The circles indicate the results for zero bias operation, and the triangles indicate those for biased operation with optimum adjustment. The pump power and the amplifier circuit were adjusted for optimum noise conditions. The gain of the amplifier was maintained constant at 16 db.

The theoretical noise figure curves for several different γ's are also drawn in the same figure. The values for γ given in brackets correspond to the short-circuit assumption; those without brackets correspond to the open-circuit assumption.

It is worth noting that the noise figures measured for the same kind of diode are found in the vicinity of the same γ curve. For example,

† A technique for measuring $\tilde{Q}$, and the comparison between the measured noise figures and the measured values of $\tilde{Q}$, will be published in the near future.

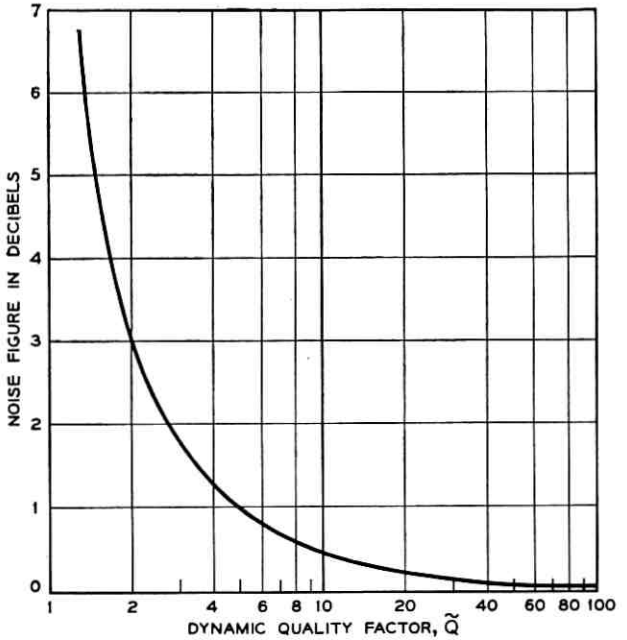


Fig. 8 — Minimum noise figure F_m of lower sideband degenerate amplifier.

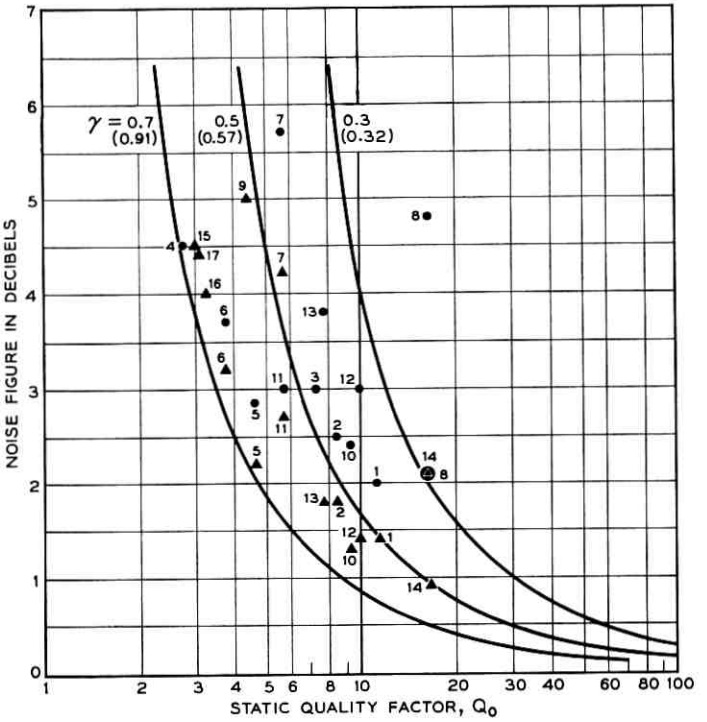


Fig. 9 — Measured results for noise figure of various diodes. Circles indicate the results for zero bias operation, triangles are for biased operation. Solid lines

TABLE I—MEASURED RESULTS OF NOISE FIGURES FOR VARIOUS DIODES

Diode Number	Material	Q	F for no bias (db)	F for bias (db)
1	silicon	11.2	2.0	1.4
2	silicon	8.5	2.5	1.8
3	silicon	7.4	3.0	—
4	silicon	2.74	4.5	—
5	silicon	4.74	2.87	2.2
6	silicon	3.88	3.7	3.2
7	germanium	5.78	5.7	4.2
8	germanium	16.65	4.3	2.1
9	silicon	4.4	—	5.0
10	silicon	9.3	2.4	1.3
11	silicon	5.85	3.0	2.7
12	gallium arsenide	10.0	3.0	1.4
13	gallium arsenide	7.7	3.8	1.8
14	gallium arsenide	16.7	2.1	0.9
15	germanium-gold	3.0	—	4.5
16	germanium-gold	3.3	—	4.0
17	germanium-gold	3.1	—	4.4

numbers 1, 2, 3, 10, and 11 are silicon p-n junction diodes, and their points are a little above the curve $\gamma = 0.5$ for zero bias operation and a little below the same curve for biased operation. Numbers 12, 13, and 14 are gallium arsenide diodes, and their noise figures are found along the $\gamma = 0.3$ curve for no bias operation and a little below the $\gamma = 0.5$ curve for biased operation. For germanium-gold bonded diodes, numbers 15, 16, and 17, the corresponding value of γ is 0.65 for biased operation.

It should be mentioned that the effective Q of the diode at the operating point is not the same as that measured at zero bias. This is because the capacitance of the diode is nonlinear and the pump voltage is swept over a wide range. Therefore, the average capacitance of the diode depends on the amplitude of the pump power and the characteristic of the capacitance. Thus, the values of γ corresponding to the measured noise figures do not give a direct indication of the γ used in the calculations. However, the difference is expected to be small. Also it should be mentioned that the Q's of the diodes were in fact measured at 1 kmc, and then calculated for the operating frequency of 6 kmc using a relation similar to (26).

VII. EQUIVALENT CIRCUIT FOR THE UP-CONVERTER

For the upper sideband up-converter, the diode junction is characterized by relationships similar to those for the lower side-band amplifier; that is,

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{j\omega_1 K_0} & \frac{1}{j\omega_2(2K_1)} \\ \frac{1}{j\omega_1(2K_1)} & \frac{1}{j\omega_2 K_0} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{40}$$

when making the open-circuit assumption for the unwanted frequencies, and

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} j\omega_1 C_0 & j\omega_1 \frac{C_1}{2} \\ j\omega_2 \frac{C_1}{2} & j\omega_2 C_0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{41}$$

when making the short-circuit assumption for the unwanted frequencies. From these equations, in a similar manner to that of Section II, the equivalent circuit of Fig. 10 is obtained for both cases.

From this equivalent circuit, the gain and noise figure are found to be

$$G = \frac{4 \frac{R_g R_L}{R_s^2} \tilde{Q}_1^2}{\left| \left(1 + \frac{Z_{11}}{R_s}\right)\left(1 + \frac{Z_{22}}{R_s}\right) + \tilde{Q}_1 \tilde{Q}_2 \right|^2}, \tag{42}$$

$$F = 1 + \frac{T_s R_s + T_1 R_1}{T_g R_g} + \frac{T_s R_s + T_2 R_2}{T_g R_g} \left| 1 + \frac{Z_{11}}{R_s} \right|^2 \frac{1.}{\tilde{Q}_1^2}. \tag{43}$$

To obtain (43), we have disregarded the noise contribution of R_L. The reasoning here follows from the conventional definition of the noise

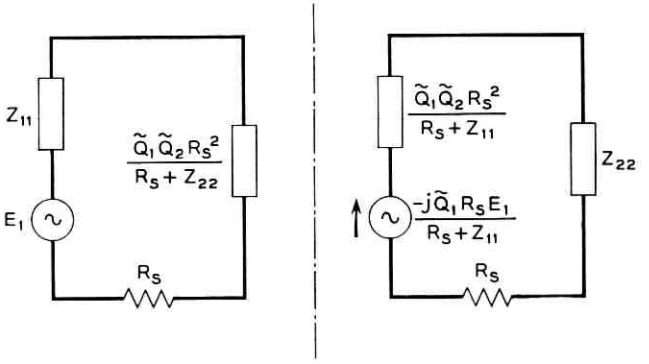


Fig. 10 — Equivalent circuit of upper sideband up-converter.

figure which is identified with a particular stage and does not include the noise contribution of the following stage. Here, R_L is taken as the input impedance of the second stage. This is quite different from the negative-resistance amplifier discussed before, where the noise power from the load is also amplified and is, therefore, taken into account as extra noise attributable to the amplification process.

VIII. MINIMUM NOISE FIGURE

Equation (43) shows that the minimum noise figure is obtained when $1 + (Z_{11}/R_s)$ is real and R_1 and R_2 are as small as possible. Therefore, by adjusting the circuit so that

$$R_1 = 0, \quad R_2 = 0, \quad 1 + \frac{Z_{11}}{R_s} = 1 + \frac{R_g}{R_s}, \tag{44}$$

(43) becomes

$$F = 1 + \frac{T_s}{T_g}\left[\frac{R_s}{R_g} + \frac{R_s}{R_g}\left(1 + \frac{R_g}{R_s}\right)^2 \frac{1}{\tilde{Q}_1^2}\right]. \tag{45}$$

Minimizing F with respect to R_g, the noise figure becomes

$$F_m = 1 + 2\frac{T_s}{T_g}\left(\frac{1}{\tilde{Q}_1^2} + \frac{1}{\tilde{Q}_1}\sqrt{1 + \frac{1}{\tilde{Q}_1^2}}\right) \tag{46}$$

when

$$R_g = R_s\sqrt{1 + \tilde{Q}_1^2} \equiv R_s L. \tag{47}$$

Comparing (46) with (27), we find that the minimum noise figure is equal to that of the lower sideband amplifier when this is used with the optimum idler frequency or with a zero temperature idler load.

The condition for the minimum noise figure is given by (47). There is a further degree of freedom in the resistance of the load, which does not appear in the noise expression (45). This degree of freedom is resolved by choosing R_L for maximum gain. The gain under the assumption of minimum noise figure is

$$G = \frac{4\dfrac{R_g R_L}{R_s^2}\tilde{Q}_1^2}{\left|\left(1 + \dfrac{R_g}{R_s}\right)\left(1 + \dfrac{R_L}{R_s}\right) + \tilde{Q}_1\tilde{Q}_2\right|^2}. \tag{48}$$

Maximizing G with respect to R_L, the gain becomes

$$G = \frac{\frac{\omega_2}{\omega_1}}{\left(1 + \frac{1}{\tilde{Q}_1\tilde{Q}_2} + \frac{1}{\tilde{Q}_1\tilde{Q}_2}\sqrt{1 + \tilde{Q}_1^2}\right)\left(1 + \frac{1}{\sqrt{1 + \tilde{Q}_1^2}}\right)} \tag{49}$$

when the output is matched, i.e.,

$$R_L = R_s\left(1 + \frac{\tilde{Q}_1\tilde{Q}_2}{1 + \frac{R_g}{R_s}}\right) \equiv R_s M. \tag{50}$$

Equation (49) is thus the maximum gain under the restriction of minimum noise figure. When $\tilde{Q}_1$ and $\tilde{Q}_2$ become large, the gain approaches ω_2/ω_1, as is to be expected. The curves F versus $\tilde{Q}_1$ and G versus $\tilde{Q}_1$ are shown in Figs. 5 and 11 respectively. Fig. 12 shows L versus $\tilde{Q}_1$ and M versus $\tilde{Q}_1$ for several different values of ω_2/ω_1.

IX. MAXIMUM GAIN

Next we shall calculate the maximum gain condition irrespective of the minimum noise figure condition imposed earlier.

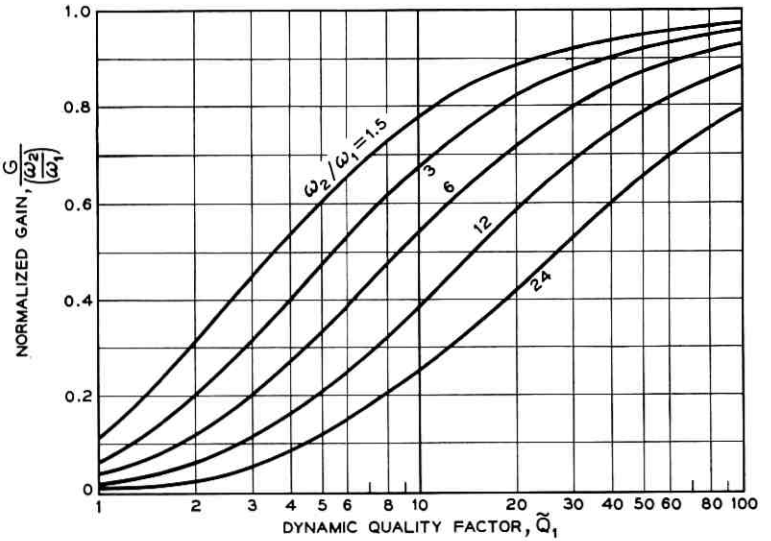


Fig. 11 — Normalized gain of upper sideband up-converter under minimum noise figure condition.

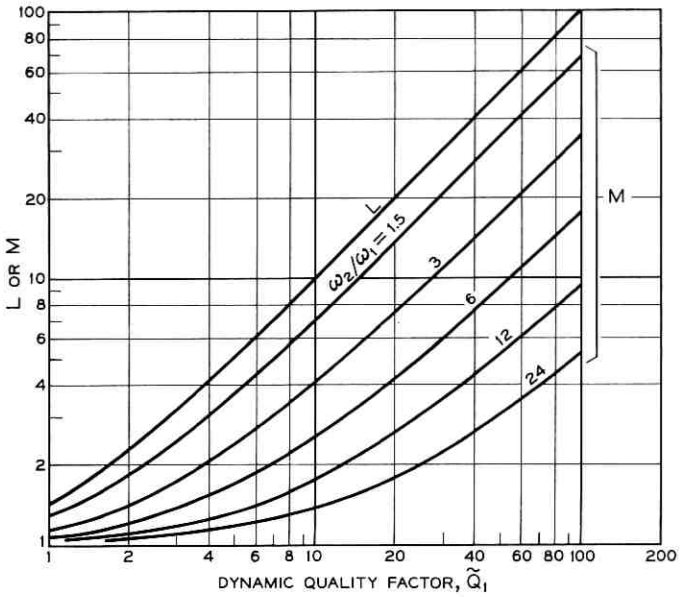


Fig. 12 — L and M vs. Q_1.

From (42), assuming R_g and R_L to be constant, the largest gain is provided by the conditions that Z_{11} and Z_{22} are real and that R_1 and R_2 vanish. We therefore only need to investigate the expression (48). Upon maximizing the right-hand side of (48) with respect to R_g and R_L, the gain becomes

$$G = \frac{\omega_2}{\omega_1}\frac{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2} - 1}{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2} + 1} = \frac{\omega_2}{\omega_1}\frac{K - 1}{K + 1} \tag{51}$$

when

$$R_g = R_L = R_s\sqrt{1 + \tilde{Q}_1\tilde{Q}_2} \equiv R_sK. \tag{52}$$

The noise figure under this condition is

$$\begin{aligned} F &= 1 + \frac{T_s}{T_g}\left(\frac{1}{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2}} + \frac{\omega_1}{\omega_2}\frac{1}{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2}}\frac{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2} + 1}{\sqrt{1 + \tilde{Q}_1\tilde{Q}_2} - 1}\right) \\ &= 1 + \frac{T_s}{T_g}\left(\frac{1}{K} + \frac{\omega_1}{\omega_2}\frac{1}{K}\frac{K + 1}{K - 1}\right). \end{aligned} \tag{53}$$

If $\tilde{Q}_1\tilde{Q}_2$ is large, K is large, G approaches ω_2/ω_1, and F tends to unity, as

is to be expected. The curves F versus $\tilde{Q}_1$, G versus $\tilde{Q}_1$ and K versus $\tilde{Q}_1$ are shown in Figs. 13, 14, and 15.

X. MINIMUM OVER-ALL NOISE FIGURE

In the previous discussions of the upper sideband up-converter, only the noise figure of the up-converter was considered and no attention was paid to the following stage. However, in a practical system, the over-all noise figure is more important than that of the preamplifier itself. This is especially so when the gain of the preamplifier is low, or the over-all noise figure is much higher than that of the preamplifier alone. We shall therefore consider in this section the over-all noise figure.

As discussed in the previous section, the condition for the minimum noise figure of the up-converter does not coincide with that for the maximum gain. Therefore, the best over-all noise performance is ob-

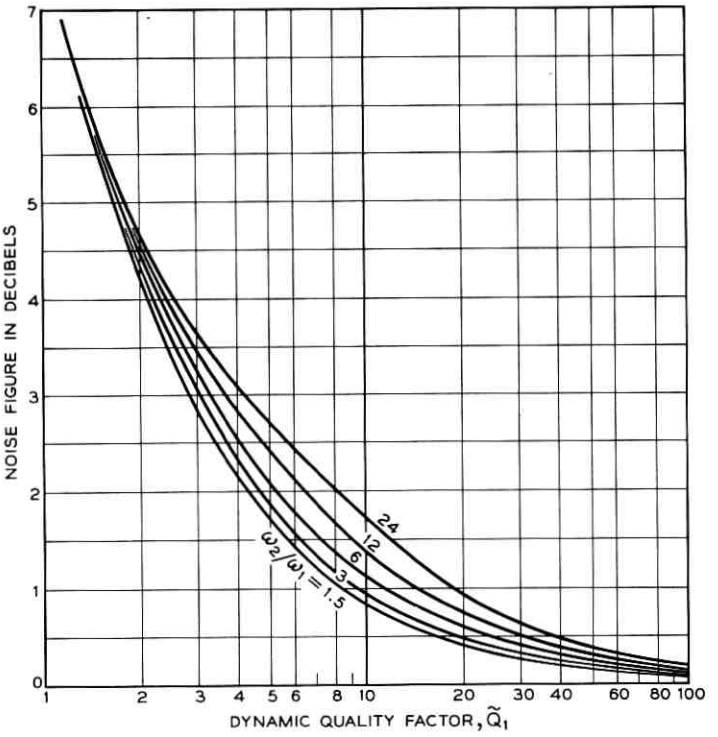


Fig. 13 — Noise figure of upper sideband up-converter under maximum gain condition.

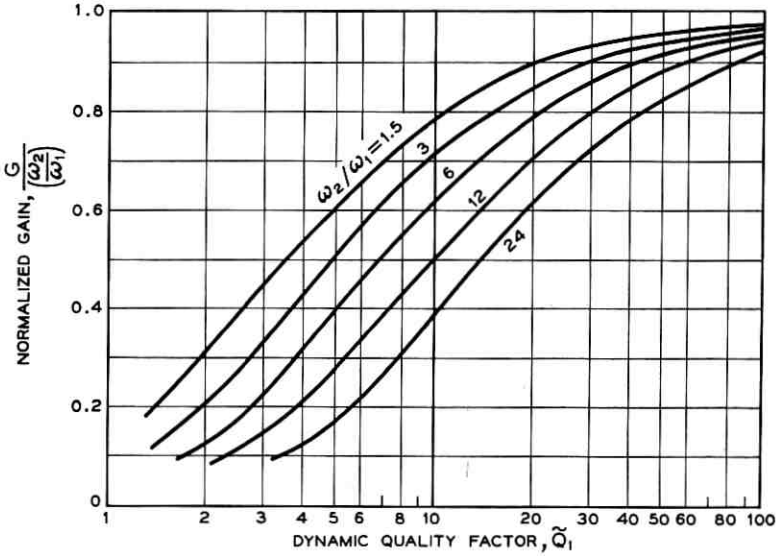


Fig. 14 — Normalized maximum gain of upper sideband up-converter.

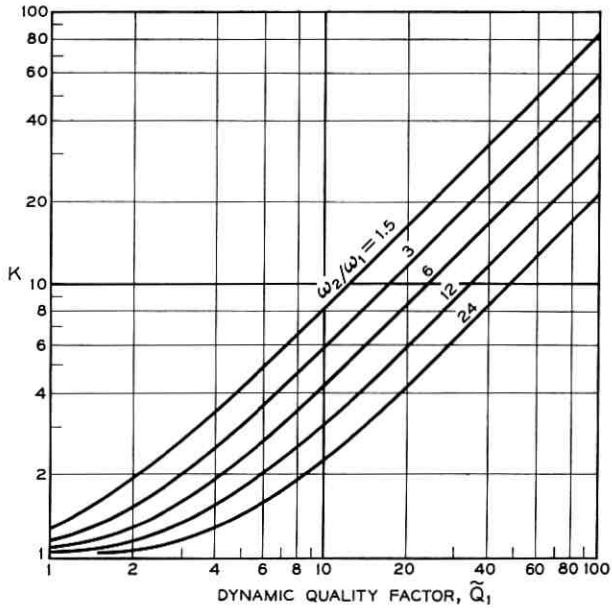


Fig. 15 — K vs. Q_1 .

tained neither at the minimum noise figure condition nor with maximum gain.

The noise figure of the second stage depends on the input impedance. In this discussion, however, we shall assume that the second-stage input impedance is kept constant by connecting an isolator in front of it. Thus, the noise figure of the second stage is defined regardless of any possible mismatch in the output impedance of the up-converter.

The over-all noise figure is given by

$$F_0 = F_1 + \frac{F_2 - 1}{G_1}, \tag{54}$$

where F_1 is the noise figure of the up-converter including the isolator, G_1 is the gain, and F_2 is the noise figure of the second stage.

Since the best over-all noise figure is obtained when F_1 is small and G_1 is large, we have only to investigate the case where Z_{11} and Z_{22} are real and R_1 and R_2 are equal to zero.

From (45), (48), and (54), the over-all noise figure becomes

$$\begin{aligned} F_0 = 1 + \frac{T_s}{T_g}\left[\frac{1}{L_0} + \frac{1}{L_0}\frac{(1+L_0)^2}{\tilde{Q}_1^2}\right] \\ + \frac{T_L}{T_g}\frac{\left[(1+L_0)(M_0-1) - \tilde{Q}_1^2\frac{\omega_1}{\omega_2}\right]^2}{4L_0M_0\tilde{Q}_1^2} \\ + (F_2 - 1)\frac{\left[(1+L_0)(1+M_0) + \tilde{Q}_1^2\frac{\omega_1}{\omega_2}\right]^2}{4L_0M_0\tilde{Q}_1^2}, \end{aligned} \tag{55}$$

where

$$R_g \equiv R_s L_0, \quad R_L \equiv R_s M_0. \tag{56}$$

Minimizing F_0 with respect to M_0, the optimum M_0 is given by

$$M_0 = 1 + \frac{\omega_1}{\omega_2}\frac{\tilde{Q}_1^2}{1+L_0}. \tag{57}$$

This is the same condition as (50). Under this condition, the optimum value L_0 is found to be

$$L_0 = \tilde{Q}_1\left[\frac{1}{\tilde{Q}_1^2} + \frac{\frac{T_s}{T_g} + \frac{\omega_1}{\omega_2}(F_2 - 1)}{\frac{T_s}{T_g} + F_2 - 1}\right]^{\frac{1}{2}}. \tag{58}$$

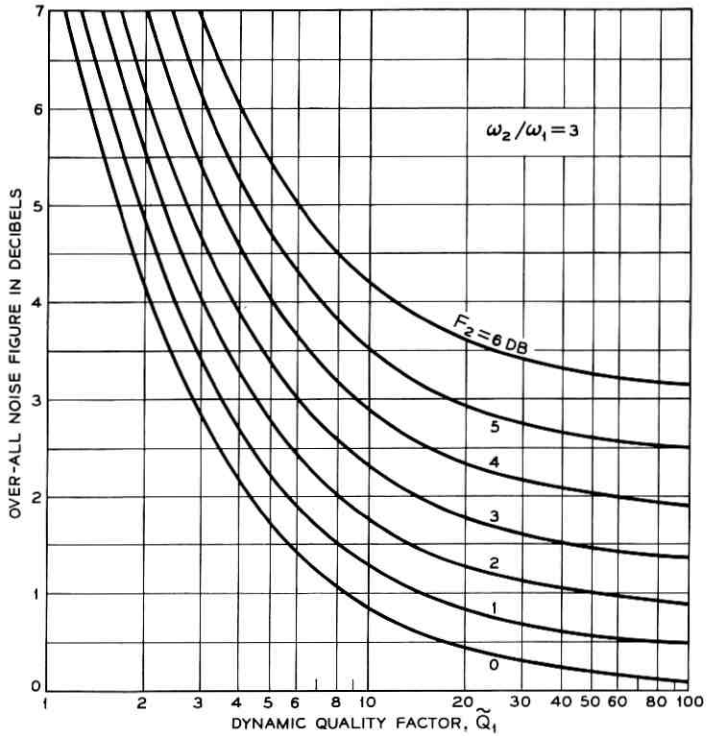


Fig. 16 — Over-all noise figure F_0. Output-to-signal frequency ratio of 3 is used for this calculation.

If we make $F_2 = 1$, (58) becomes

$$L_0 = \sqrt{1 + \tilde{Q}_1^2},$$

which is the same as (47), the condition for the minimum noise figure. If we make $F_2 \to \infty$, (57) and (58) become

$$M_0 = \sqrt{1 + \tilde{Q}_1\tilde{Q}_2},$$

$$L_0 = \sqrt{1 + \tilde{Q}_1\tilde{Q}_2}.$$

These are the maximum gain conditions which appear in (52). Therefore, as expected, when the second stage has a very poor noise figure, the up-converter should be adjusted for the maximum gain condition, while it must be at the minimum noise figure condition if the second stage noise figure is close to unity.

From (55), (57), and (58), the minimum over-all noise figure is given by

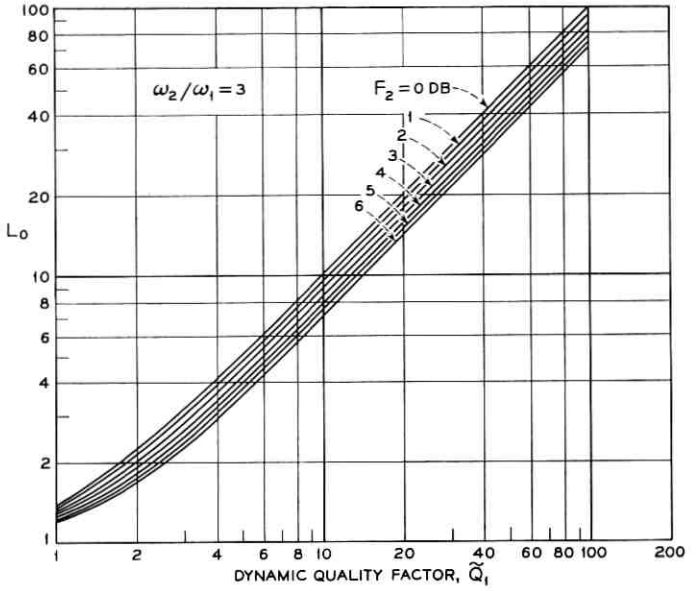


Fig. 17 — L_0 vs. Q_1 .

$$F_0 = 1 + \frac{2}{\tilde{Q}_1^2}\left(\frac{T_s}{T_g} + F_2 - 1\right) + \frac{\omega_1}{\omega_2}(F_2 - 1) + \frac{2}{\tilde{Q}_1}\sqrt{\left[\frac{T_s}{T_g} + \frac{\omega_1}{\omega_2}(F_2 - 1) + \frac{1}{\tilde{Q}_1^2}\left(\frac{T_s}{T_g} + F_2 - 1\right)\right]\left(\frac{T_s}{T_g} + F_2 - 1\right)}. \tag{59}$$

As an example, taking $T_s/T_g = 1$ and $\omega_2/\omega_1 = 3$, the curves F_0 versus $\tilde{Q}_1$, L_0 versus $\tilde{Q}_1$, and M_0 versus $\tilde{Q}_1$ are shown in Figs. 16, 17, and 18, respectively, to show the general tendency of the variations in these parameters.

XI. CONCLUSION

On the assumption that the series resistance of the diode is the only parasitic element, we have calculated the minimum noise figures for both the lower sideband and the upper sideband parametric amplifiers under various conditions. For each case, the noise figure is basically determined by the dynamic quality factor $\tilde{Q}$ of the diode. The larger $\tilde{Q}$ is, the lower is the noise figure which can be obtained. If the practically obtainable values of $\tilde{Q}$ are equal for different diodes, the same minimum noise figure is expected. Even for the static diode Q, i.e., Q_0 is very high, it is impossible to build a low-noise amplifier if the capacitance varia-

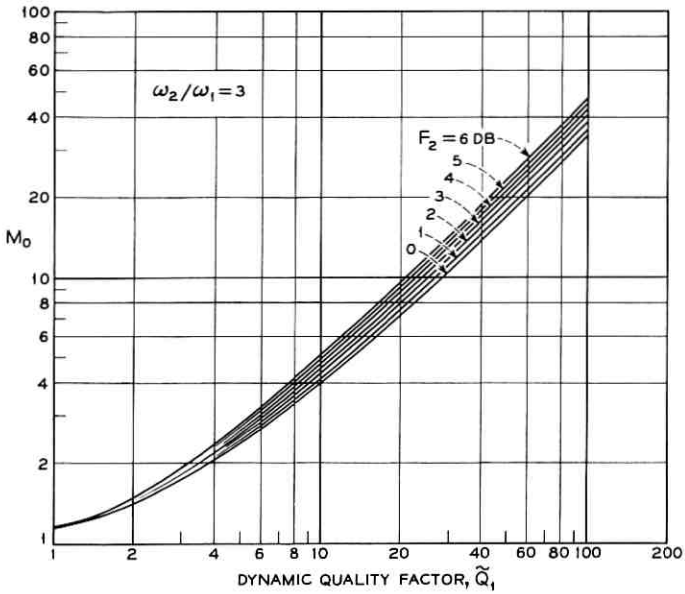


Fig. 18 — M_0 vs. $\tilde{Q}_1$.

tion is small. Therefore, we conclude that the dynamic quality factor $\tilde{Q}$ is more appropriate than Q_0 as a measure of the quality of a variable capacitance diode.

The identical noise figure expression is obtained for both the lower sideband idler output and the circulator-type amplifier, if the gain in each case is large. The minimum noise figure of the lower sideband amplifier for room-temperature operation is obtained when the idler load resistance approaches zero. For a given $\tilde{Q}_1$ there exists an optimum idler frequency at which the realizable noise figure is equal to the best that is obtainable with a zero temperature idler load.

The minimum noise figure of the upper sideband up-converter is equal to that of the lower sideband amplifier. Since the maximum gain of the upper sideband up-converter is limited by the ratio of output to input frequency, while the gain of the lower sideband amplifier is unlimited, a superior over-all noise figure is expected with the lower sideband amplifier.

XII. ACKNOWLEDGMENTS

Acknowledgments are due to H. Seidel and K. D. Bowers for stimulating discussions and helpful criticism in preparation of this manuscript.

APPENDIX

Stability Comparison for Two Different Types of Lower Sideband Amplifier

Since the lower sideband idler-output and circulator-type amplifiers give the same noise figure under the large-gain condition, the question of which is more stable arises. With large gain and minimum noise operation, the idler-output type is less stable than the circulator type. To see that this is so, consider the gain expressions (14) and (19). The major cause for instability comes from the denominator, since the two terms in the denominator almost cancel each other and a small variation of either term brings about a large variation in the gain. For instance, a small increase in the pump power can cause enough variation in $\tilde{Q}_1\tilde{Q}_2$ for the amplifier to break into oscillation. Similarly, a small change in the input impedance Z_{11} gives a large variation.

For a given gain, if the numerator is small, the cancellation of the two terms in the denominator must be more complete; in other words, the negative resistance effect must be more fully utilized, making the amplifier less stable. Thus, for the comparison of the stability of the two types, we only have to investigate the numerator of the gain expressions.

For the minimum noise figure,

$$R_g = R_s(\tilde{Q}_1\tilde{Q}_2 - 1),$$

$$R_L \to 0.$$

Hence we see at once that the idler-output type has the smaller numerator, making it the less stable amplifier.

Let us now consider the case where R_L is small but finite, i.e., where we accept some degradation in noise performance. The comparison should then be made between $R_g \simeq R_s\tilde{Q}_1\tilde{Q}_2$ and $R_L\tilde{Q}_1^2$, for $|\ 1 + (Z_{22}^*/R_s)\ |^2$ is approximately unity and $\tilde{Q}_1\tilde{Q}_2 \gg 1$. From this comparison, we find that $R_L/R_s > \omega_1/\omega_2$ is the approximate condition for the idler-output type to be the more stable. The noise figure does not change very rapidly when R_L/R_s changes a little from its optimum value of zero. In fact, the noise figure thus obtainable corresponds to a decrease in $\tilde{Q}_1\tilde{Q}_2$ by the factor of $R_s/(R_s + R_L)$. We therefore conclude that, when the ratio of the idler frequency to the signal frequency is large, the idler-output type can be made the more stable with only a small sacrifice in the noise figure.

There is still another type of operation, namely the transmission type, where the output load is directly coupled at the signal frequency. It can be shown in a way similar to the above discussion that, using this

scheme, it is never possible to achieve more stable operation than with the circulator type, and the obtainable noise figure is poorer than or at best equal to that of the circulator type.

It should be remarked that the major limitation in the bandwidth also comes from the denominator of the gain expression. The above argument therefore holds equally well for a bandwidth comparison.

The above discussion holds only for room-temperature operation. If load refrigeration is available (i.e., a cold isolator at the idler frequency), a different conclusion is reached; i.e., under certain conditions, it is possible to make the idler-output type more stable without a sacrifice in the noise figure.

REFERENCES

1. Leenov, D., Gain and Noise Figure of a Variable Capacitance Up-Converter, B.S.T.J., **37**, 1958, p. 989.
2. Haus, H. A., and Penfield, P., Jr., On the Noise Performance of Parametric Amplifiers, Internal Memorandum No. 19, Massachusetts Inst. of Technology, Cambridge, Mass.
3. Uenohara, M., Noise Consideration of the Variable Capacitance Parametric Amplifier, Proc. I.R.E., **48**, 1960, p. 169.
4. Kotzebue, K. L., Optimum Noise Performance of Parametric Amplifiers, Proc. I.R.E., **48**, 1960, p. 1324.
5. Knechtli, R. C., and Weglein, R. D., Low-Noise Parametric Amplifier, Proc. I.R.E., **48**, 1960, p. 1218.
6. Bloom, S., and Chang, K. K. N., Theory of Parametric Amplification Using Nonlinear Reactances, R.C.A. Rev., **18**, 1957, p. 578.
7. Rowe, H. E., Some General Properties of Nonlinear Elements. II — Small Signal Theory, Proc. I.R.E., **46**, 1958, p. 850.
8. Heffner, H., and Wade, G., Gain, Bandwidth and Noise Characteristics of the Variable Parameter Amplifier, J. Appl. Phys., **29**, 1958, p. 1321.

Design of a High-Resolution Electrostatic Cathode Ray Tube for the Flying Spot Store

By H. G. COOPER

(Manuscript received November 22, 1960)

A high-resolution electrostatically deflected cathode ray tube is required for the flying spot store of an experimental electronic switching system. This tube is used to obtain random access, by optical means, to 2.5 × 10^6 bits of information stored on a photographic plate. High degrees of both resolution uniformity and faceplate optical quality are required to achieve large storage capacity and error-free performance.

In this paper the design criteria for optimum gun performance and minimum deflection focusing are analytically and empirically evolved. A novel result of this work is a dual shield placed between the two pairs of deflection plates, which substantially reduces beam aberrations due to the deflection fringing fields. A precision tube is described that fulfills the flying spot store design objectives and has performed reliably in a field trial at Morris, Illinois.

1. INTRODUCTION

High-resolution cathode ray tubes (CRT's) with electrostatic deflection are normally limited to relatively small deflection angles because of beam-focusing effects introduced by the deflection plates. The deflection distortion causes a loss of resolution at the edges of the screen and severely limits the usable screen diameter of high-resolution tubes.

A large-capacity, high-speed, semipermanent memory,[1,2] denoted as the *flying spot store*, has recently been developed for an experimental electronic telephone switching system.[3] One of the critical components of this memory is an electrostatically deflected CRT. The tube is used to provide random access to binary information stored in the form of transparent or opaque dots on photographic plates, as shown in Fig. 1. The electron beam is deflected to the desired address on the screen of the CRT, the luminescent spot is then focused onto an array of n informa-

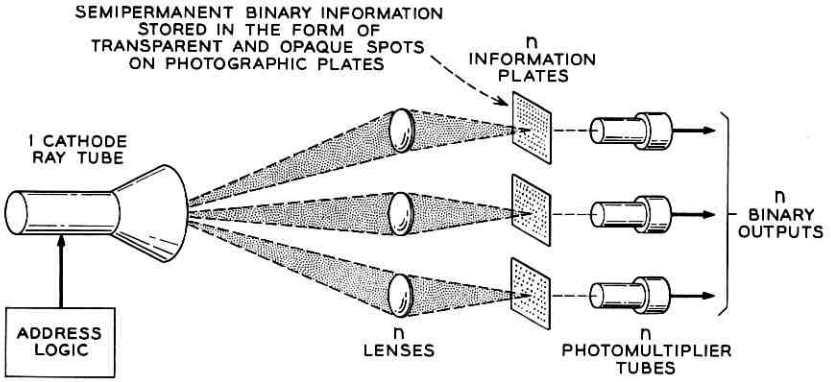


Fig. 1 — Role of the CRT in the flying spot store.

tion plates by optical lenses, and an n-digit binary output is obtained from photomultiplier tubes located behind each information plate.

System design objectives include virtually error-free performance (less than one error per 10^{10} reading operations) and a large memory capacity (2.5×10^6 bits). As a result, a precision CRT with electrostatic deflection is needed which has unusual characteristics. Among these are a small and very uniform beam size over a relatively large screen area, high deflection sensitivity, and uniform resolution over a wide beam current range without adjustment of the beam focus voltage. These points will be covered in more detail in the next section.

It was concluded that the spot size uniformity requirement at the desired resolution, screen size, etc. constituted performance appreciably better than had been previously attained. Development[4] of a CRT for the flying spot store was consequently undertaken at Bell Telephone Laboratories.

The primary objective of this article is to present the electron-optical design considerations for optimum tube performance in the system. The subjects covered, in order of presentation, are (a) tube design objectives, (b) electron gun, (c) objective lens, (d) deflection system, (e) screen, and (f) electrical characteristics of present tubes. An appreciable portion of the paper is devoted to the electrostatic deflection system, since it posed the most difficult design problem and includes a novel feature of the tube.

II. TUBE DESIGN OBJECTIVES

Development of the cathode ray tube was stimulated by system needs; thus, tube design objectives were imposed almost entirely by system

considerations. These objectives, in terms of desired tube characteristics, will be summarized in this section and serve as a basis for the three subsequent sections on tube design. Most of the values listed below have been discussed in previous articles[1,2] on system design, and others were determined in private discussions with systems personnel.

2.1 *Resolution and Spot Size Uniformity*

Cathode ray tube spot size usually refers to the diameter of the luminescent spot on the screen, which may be denoted as the *optical spot size*. This may be larger than the electron beam size (or *electrical spot size*) if there is a significant scattering of light within the screen phosphor. Experiments conducted on screen materials used in the flying spot store CRT indicated that such scattering could be made negligible if proper screen thickness and method of deposition were used. Hence it will be assumed here that electrical and optical spot sizes are the same, and the terms will be used interchangeably in the article.

A quantitative specification of spot size has long been subject to ambiguity, because of the ill-defined edge of the electron beam. Among the various methods of defining spot size are (a) shrinking raster resolution, (b) TV line resolution, and (c) the standard deviation σ of a Gaussian distribution.* The last method was chosen for this work, since experiments showed that the beam cross section in the CRT was substantially Gaussian. Another specification of spot size[1,2] is the size of a square that, when centered on the luminescent spot, will contain 90 per cent of the radiant light flux. For a Gaussian spot, the side of such a square is 4σ.

Flying spot store objectives of large storage capacity and essentially error-free operation result in CRT resolution values of $\sigma = 0.0045 \pm 0.00075$ inch. This corresponds to a spot size uniformity ratio ($\sigma_{max}/\sigma_{min}$) of $\leqq 1.4$. It should be emphasized that the tolerances on spot size must be maintained at all points on the quality screen area (6 inches square), and over the required beam-current range without changing focus potentials. The best spot size uniformity that could be obtained in commercially-developed electrostatic CRT's was a $\sigma_{max}/\sigma_{min}$ value of three, when the median σ was 0.0045 inch and the beam current was held constant at 10 microamperes.

2.2 *Accelerating Voltage*

Final CRT accelerating potential is determined by system needs for high deflection sensitivity and a high level of radiant flux from the

* See, for example, Klemperer.[5] It may be noted that σ of a Gaussian is approximately equal to one TV line or one-half a shrinking raster line.

screen. Since an improvement in one parameter is achieved only at a sacrifice in the other, a compromise must be made. On this basis, the design value for accelerating voltage was chosen as 10 kilovolts.[1,2]

2.3 *Beam Current*

Another CRT characteristic determined by the minimum permissible radiant flux from the screen is beam current. In addition, system design requires that the radiant flux must have a constant absolute value which is independent of tube life or beam location on the screen. Accordingly, variations in luminescent intensity, due to nonuniformities in screen deposition or to degradation in luminescent efficiency of the phosphor by electron bombardment, are compensated by an appropriate electronic adjustment of the beam current. Thus, an extended operating range in beam current must be provided over which the spot size limits, specified in Section 2.1, are to be maintained under conditions of constant focus voltage. It should be noted that space charge and lens aberrations produce beam size enlargements at increased beam currents.

In view of the above considerations, the required beam current operating range was selected as 4 to 20 microamperes. The value of four would be used for the most intense spot on the screen of a new tube and 20 at the dimmest location when the tube reaches end of life.

2.4 *Deflection System*

The random access feature of the flying spot store necessitates an electrostatic deflection system. In principle, magnetic coils could be used, but the power required in the deflection circuitry would be prohibitive. Objectives for the deflection system are low capacitance, less than 25 micromicrofarads per plate when driven push-pull, and an average sensitivity of 150 $\pm$ 10 volts per inch for both pairs of plates. In addition, the two sets of plates should be orthogonal to within ± 0.5 degree. Variations in deflection factor due to barrel and/or pin-cushion distortion must be less than ± 0.7 per cent over the quality screen area, which is defined later in Section 2.5.1. It is permissible to obtain the 150 $\pm$ 10 volts per inch average deflection factor by varying the final acceleration potential in the range between 9 and 11 kilovolts.

Beam focusing effects introduced by the deflection system are very deleterious to the spot size uniformity ratio. Dynamic correction for deflection focusing is frequently made by feeding part of the deflection signal back to the focus electrode(s) via an appropriate shaping circuit. This might be tolerated in the system design but is quite undesirable.

2.5 *Screen and Faceplate*

2.5.1 *Quality Screen Area*

The area of the phosphor over which tolerances on beam size, deflection factor, etc. are to be maintained is denoted the *quality screen area.* The minimum quality area is determined by many interacting system considerations, and has been specified as a square 6 inches on a side centered on the mechanical center of the faceplate.

2.5.2 *Phosphor*

The phosphor must have a short persistence ($\approx 10^{-7}$ second) to permit high-speed system operation and should possess a luminescent efficiency as high as possible. From a study of high-speed phosphors,[1,2] it was concluded that P16 was the optimum screen material. It has a decay time the order of 100 millimicroseconds and an energy efficiency of about one per cent after preaging.

Another important screen parameter is uniformity of light output from the quality area. The desired objective is less than +20 and −40 per cent variation from the median radiant flux value, in order to keep the system error rate sufficiently low.

2.5.3 *Faceplate*

The flying spot store includes a very precise optical system, which focuses the luminescent spot on the photographic information plates. Since the CRT faceplate is in the light path between the phosphor and lenses, it is a part of the optical system. One system objective is that CRT's should be interchangeable without the necessity of rewriting the information stored on their photographic plates. As a result, the faceplate is made flat (rather than curved) and must meet very rigid specifications with regard to optical flatness, thickness, plate curvature, and freedom from flaws. Briefly stated, the surface must be ground flat to within six fringes of 5890 angstroms light per inch, the thickness must be uniform (0.465 ± 0.010 inch), the radius of curvature must be greater than 1900 inches (corresponding to a bow on the axis of ±0.00375 inch measured from a plane passing through a 7.6-inch diameter circled centered thereon), and the flaw size must be maintained below 0.004-inch diameter with a minimum spacing of 0.050 inch between such flaws.

The tube design objectives discussed above are summarized in Table I.

TABLE I —SYSTEM OBJECTIVES FOR THE FLYING SPOT STORE CRT

Spot size	$\sigma = 0.0045 \pm 0.00075$ inch (for all beam currents between 4 and 20 microamperes and no change in focus potentials).
Accelerating voltage	10 kilovolts
Beam current	Variable between 4 and 20 microamperes
Deflection system	Electrostatic
Deflection factor	150 ± 10 volts per inch*
Deflection linearity	± 0.7 per cent†
Deflection plate orthogonality	90 ± 0.5 degrees
Quality area	6×6 inches square
Phosphor type	P16
Phosphor light output uniformity in the quality area	+20 and −40 per cent from the mean value
Faceplate tolerances in quality area	Flat to <6 fringes/inch of 5890 angstroms; thickness 0.465 ± 0.010 inch; minimum radius of curvature 1900 inches; maximum flaw size of 0.004 inch; flaw spacing less than 0.050 inch

* The accelerating voltage can be varied between 9 and 11 kilovolts to achieve the 150 volts per inch average deflection factor.

† Deflection factor is constant in the quality area to within ± 0.7 per cent.

III. ELECTRON GUN DESIGN

Three primary objectives in the electron gun design are:

(a) small beam size at the crossover,

(b) moderate control grid transconductance, and

(c) long cathode life.

Two important factors contributing to long life are low cathode current density (or loading) and high current efficiency (the ratio of beam to cathode currents). The latter minimizes positive ion production, which in turn reduces cathode degradation due to ion bombardment. Since the ratio of surface area within the CRT to the cathode-emitting area is exceedingly large ($>10^8$), minimization of ion bombardment warrants serious consideration in the gun design.

An immersion-lens triode gun, shown in Fig. 2(a), is used in most CRT's made today. It contains three elements which focus the electrons to a small crossover located in the control grid-anode region. The crossover in turn is imaged on the screen by the objective lens. Moss[6] has investigated the triode gun in great detail. Calculations of crossover size, cathode loading, and current efficiency were carried out, using the results of Moss, for a number of triode configurations with the anode at 10 kilovolts. This approach appeared unsatisfactory because the beam current was either very low (less than 10 microamperes) or cathode loading was high (>1 ampere per cm^2), to achieve a suitably small crossover size.

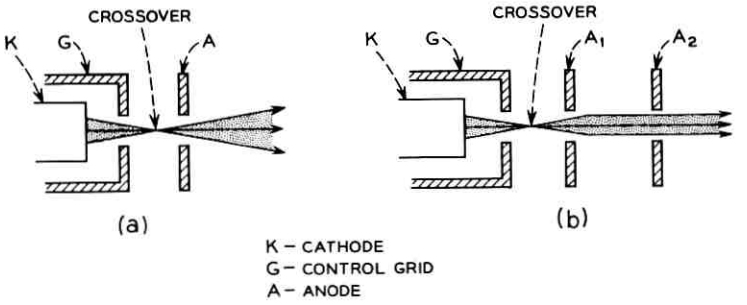


Fig. 2 — Schematics of (a) triode and (b) tetrode guns.

Much better results, which will be discussed in detail later, were obtained with the tetrode gun arrangement of Fig. 2(b), containing cathode K, control grid G, first anode A_1, and second anode A_2. Consequently, the tetrode gun was selected for the flying spot store CRT and was designed to meet the objectives outlined in Section II. The first three electrodes (K, G, and A_1) constitute an immersion-lens triode, forming a crossover in the G–A_1 space as drawn in Fig. 2(b). Therefore, the extensive knowledge available on triode guns may be applied to the tetrode front end design. A model previously developed* at Bell Telephone Laboratories for high resolution storage tubes (designed for an A_1 potential $\simeq$1 kilovolt) was selected for this purpose. It has a small crossover size and moderately high transconductance ($\approx$2 micromhos).

A very important factor affecting beam size at the screen is the electrode geometry of anodes A_1 and A_2 of the tetrode gun, where A_2 is operated at the nominal final accelerating potential of 10 kilovolts. The analysis utilized to attain a suitable design is described in the next section.

3.1 *Design of A_1 and A_2 Electrodes*

The electrode arrangement at the A_1–A_2 gap, the region in which the beam is accelerated from 1 to 10 kilovolts, interacts critically with the over-all crossover magnification M (defined as the ratio of spot size at the screen divided by the crossover size). Since the A_1–A_2 lens is converging, it should be made as weak as possible in order to minimize M. Because of its convergent action, the A_1–A_2 lens geometry also strongly affects the cathode loading and current efficiency (the fraction of cathode current that passes through the limiting aperture of the objective

* This work was done by R. W. Sears and W. E. Kirkpatrick.

lens). The latter point is illustrated by beam trajectories in Fig. 2(b), where the beam is confined closer to the axis (after it passes through the anode) for the case of the tetrode than for that of the triode. The current efficiency increases as the A_1–A_2 lens becomes more convergent (which reduces cathode loading) and also when the A_1–A_2 gap is placed closer to the crossover.

Three different A_1–A_2 configurations, shown in Fig. 3, are analyzed for crossover magnification in Appendix A. The lens constants, namely focal lengths and locations of principal planes, used in this analysis were obtained from the work of Spangenberg and Field.[7,8,9] The D_1, D_2, and S values listed in the figure were selected (from those tabulated in Ref. 7) to provide the weakest lens. The results are summarized in Fig. 4, where the magnification, M, due to all lenses between the crossover and screen is plotted as a function of the distance X between the crossover and midplane of the A_1–A_2 gap. The midplane for each geometry is indicated in Fig. 3, and the electron-optical diagram of the CRT is shown in Fig. 20(b). It may be noted from Fig. 4 that M is smallest for the A_1–A_2 configuration in (b) and that M increases monotonically with X. Thus X should be small, which, fortunately, is also the condition for maximum current efficiency. The final tetrode design incorporating a "(b)" A_1–A_2 arrangement (with a slightly reduced D_2/D_1 ratio) is included as Fig. 5. Because of electrostatic field and mechanical considerations, the distance X from the crossover to A_1–A_2 midplane could not be reduced completely to zero. One reason for this is that the crossover moves closer to the cathode as the A_1–A_2 spacing approaches zero. A 0.8-inch X value was selected, therefore, as a good compromise. Also the D_2/D_1 ratio was made small, 0.33, in order to make the A_1–A_2 lens as weak as possible. This ratio was not plotted in Fig. 4, since the lowest D_2/D_1 ratio studied by Spangenberg and Field[8] for the "(b)" geometry was 0.67.

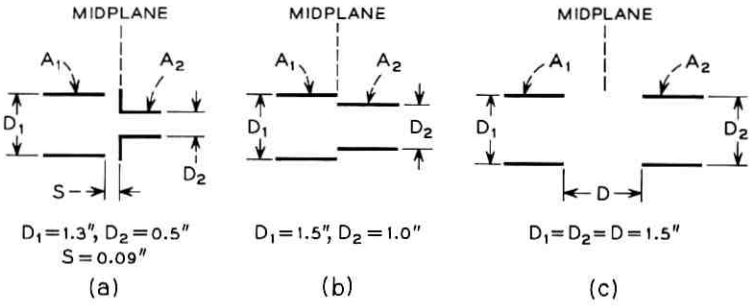


Fig. 3 — Three proposed A_1–A_2 configurations; D and S values have been selected to provide a weak A_1–A_2 lens.

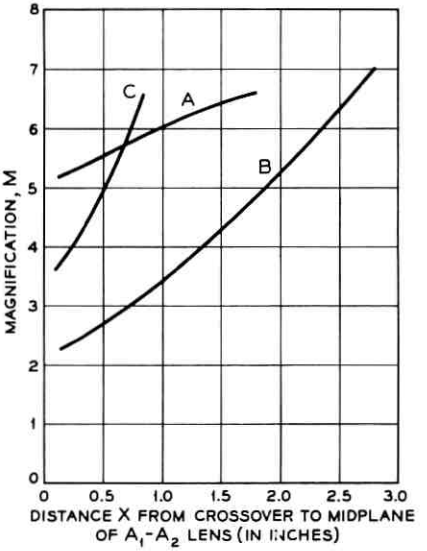


Fig. 4 — Magnification M vs. A_1–A_2 midplane position X (M = ratio of beam size at screen to crossover size). Note: M is negative for configuration (c) of Fig. 3 and positive for (a) and (b).

An important advantage of this gun design is that the average cathode loading is $\leqq 150$ milliamperes per cm^2. Also, the current efficiency is very high; for example, only about 15 microamperes of cathode current I_K are required to provide 10 microamperes of beam current I_B (67 per cent current efficiency).

Another gun feature, utilized in the flying spot store, is the capability of varying the average beam size, σ_{avg}, over a wide range by controlling

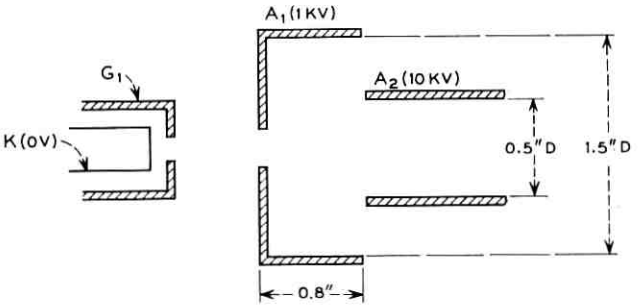


Fig. 5 — Tetrode gun design for CRT.

the voltage of A_1. This and the electrical characteristics of the tetrode gun will be presented in detail in Section VII.

IV. OBJECTIVE LENS

The crossover formed by the electron gun is imaged on the screen by the combined convergent focusing action of the A_1–A_2 lens and the objective lens [see Fig. 20(a)]. The objective lens, located close to the beam entrance side of the deflection system, provides most of the beam focusing. It also contains the limiting aperture that defines the maximum beam diameter as it passes through the objective lens and deflection system.

An electrostatic, rather than magnetic, lens was chosen for the objective for reasons of compactness and reduction of pattern distortion on the screen due to magnetic fringe fields. Fig. 6 shows the lens used, denoted *crossed-elliptical*, which has low spherical aberrations and provides orthogonal focus control. It is a four-element modified einzel (or unipotential) design, in which the first and last electrodes are both at A_2 (10 kilovolt) potential. This lens will be described in detail elsewhere[10] and only the general features will be pointed out here. (The CRT lens is a 1.5:1 scale-up of the lens described in Ref. 10.) There are two focus electrodes, A_3 and A_4, having mutually perpendicular elliptical apertures (from which the name is derived). The unique feature of the lens is substantially independent focus control in the horizontal and vertical directions by means of A_3 and A_4 respectively. It is achieved in the crossed-elliptical lens with a minimum number of electrodes. Also, the axial

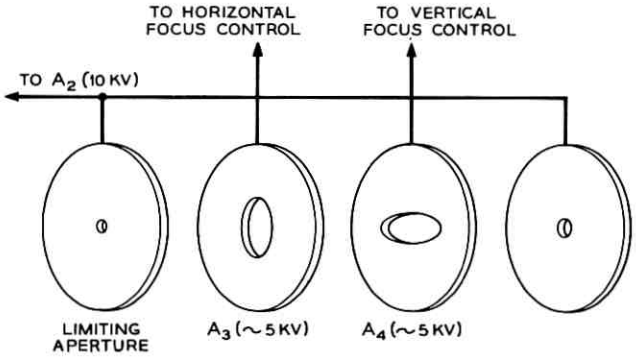


Fig. 6 — Crossed-elliptical lens.

distance occupied by the lens is very small. Both A_3 and A_4 operate at a potential near 5 kilovolts in the CRT design.

4.1 *Advantages of the Crossed-Elliptical Lens*

There are several means by which the independent focus properties of the lens are utilized in the flying spot store CRT:

i. It permits astigmatism correction at the objective lens, rather than at the deflection system. This is advantageous in the design of the deflection amplifier for the flying spot store.

ii. The conditions for optimum focus (best spot size uniformity ratio) can be attained more conveniently. The beam-positioning servo-loop of the flying spot store contains separate vertical and horizontal bar patterns located in the plane of the photographic plates. From them, it can be ascertained when best vertical and horizontal focus is reached, and the adjustment can be made quickly with virtually no interaction in the two directions.

iii. The crossed-elliptical lens can provide independent dynamic deflection focus correction; at present this is not used, but it is available if needed. Circuitry for dynamic correction with this lens has been developed in connection with the barrier-grid storage tube.[10,11]

4.2 *Lens Location and Length of CRT*

Magnification of the objective lens (M_2 in Appendix A) is determined by the spot size required at the screen, crossover size, and magnification M_1 of the A_1–A_2 lens. With the tetrode gun design described in Section 3.1, it was found that the image-to-object distance ratio q_2/C of the objective lens (see Fig. 20) should be about 5 to obtain the required average beam size, σ_{avg}, of 0.0045 inch. The image distance q_2, from the center of the objective lens to the screen, is determined almost entirely by the minimum permissible distance from the beginning of the deflection system to the screen. This in turn is defined by the amount of deflection focusing that can be tolerated. In the Section V it is concluded that q_2 should be 25 inches. Hence the object distance C from crossover to the objective lens, is equal to 5 inches. The lens is placed as close to the entrance of the deflection system as fringing field aberrations will permit. In the CRT, this axial spacing (1.0 inch) was made twice the entrance separation (0.5 inch) of the gun's set of deflection plates.

The over-all tube length is equal to the cathode-to-screen distance ($C + q_2$) plus the axial length needed at the cathode end for leads,

supports, stem, and base (about 5 inches for this CRT). Hence the total tube length becomes $(C + q_2 + 5)$, which is approximately 35 inches.

4.3 *Limiting Aperture*

The electron beam diameter, as it passes through the objective lens and deflection system, is an important tube-design parameter. It is one of the dominant factors that determine lens and deflection aberrations. Thus it influences both the average spot size σ_{avg} and the uniformity ratio $\sigma_{max}/\sigma_{min}$.

A very suitable position for the limiting aperture was found to be at the first electrode of the crossed-elliptical lens (see Fig. 6). Effects of secondary emission from that location were negligible, and it was also convenient from mechanical considerations. Two aperture diameters were studied, 0.113 and 0.075 inch. The spot size uniformity ratio was slightly better (5 to 10 per cent) for the smaller aperture, but cathode loading was increased considerably. Since spot size objectives could be met with the 0.113-inch aperture, it was used in the final design.

V. DEFLECTION SYSTEM

The most formidable CRT design problem, and the area where the greatest improvement over previously existing tubes was needed, was that of the electrostatic deflection system. The primary problem was reduction of deflection focusing and aberration effects to the point where sufficient uniformity in beam size would be achieved over the relatively large quality screen area.

The general approach was as follows. First, the deflection plates were contoured to maximize sensitivity. The distance from the screen to the point where the beam enters the deflection system was then selected to maintain deflection focusing effects appropriately small. Finally, the separation, length, and termination for the two pairs of plates were designed for the lowest $\sigma_{max}/\sigma_{min}$ ratio. Clearly, the optimum spot uniformity is achieved when the maximum spot enlargement produced by the vertical plates is equal to that of the horizontal set. The problem then is to evaluate quantitatively, the spot size enlargement for a given electrostatic deflection system, which will be done below.

5.1 *Plate Contour*

It has long been known[12] that the sensitivity of electrostatic deflections can be optimized by shaping them to the beam contour at the maximum

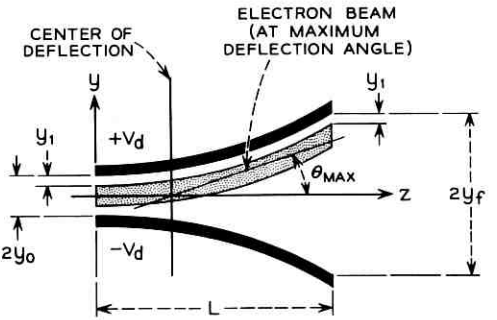


Fig. 7 — Deflection plate contour for maximum sensitivity: V_d is measured relative to average deflection plate potential, V_0 ; distance y_1 from beam edge to deflection plate is constant for all values of z.

deflection angle. Under these conditions, as may be seen from Fig. 7, the beam edge is always a constant distance y_1 (perpendicular to the axis) from the positive plate. The equation[12] for the optimum shape, where the axial dimension z is expressed as a function of the transverse distance y, may be written as

$$z - z_0 = 2y_0 \left(\frac{V_0}{V_d}\right)^{\frac{1}{2}} \int_0^{\sqrt{\ln (y/y_0)}} e^{u^2}\, du. \tag{1}$$

Parameters in (1), illustrated in Fig. 7, are the initial plate separation $2y_0$, the accelerating (or average deflection plate) potential V_0 , and the maximum push-pull deflection voltage V_d (measured relative to V_0). It may be noted that the so-called "peak-to-peak" deflection voltage is $4V_d$, since each plate has a maximum variation of $\pm V_d$. A boundary condition is that $z = z_0$ when $y = y_0$. Equation (1) is an integral for which no closed-form solution was found. Fortunately, it is the same as that for space charge spreading in a cylindrical beam, and solutions have been tabulated in generalized graphical form (Ref. 13, p. 149). A specific plot of interest in deflection plate design is

$$\frac{y}{y_0} \quad \text{vs.} \quad \frac{z - z_0}{2y_0 \left(\frac{V_0}{V_d}\right)^{\frac{1}{2}}},$$

which is included as Fig. 8. The final plate separation, $2y_f$ in Fig. 7, is given by the value of y when z is equal to the deflection plate length L. All intermediate points can be obtained from Fig. 8 once V_0 , V_d , and y_0 are determined.

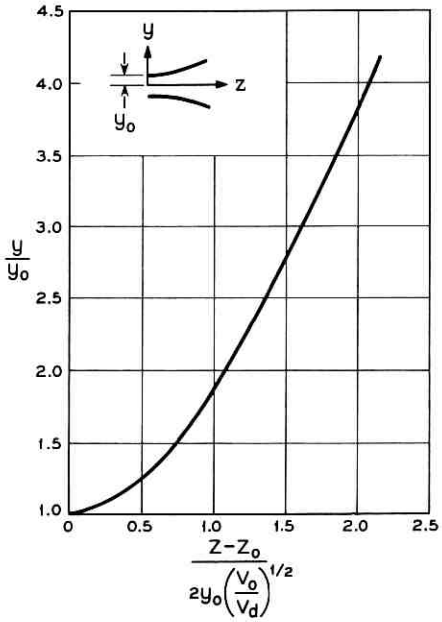


Fig. 8 — Fundamental curve for designing deflection plates contoured for maximum sensitivity.

5.2 *Spot Size Enlargement Produced by Deflection Focusing*

Fig. 9 depicts the phenomenon of deflection focusing. The trajectories of edge electrons, as the beam passes from the deflection system to the screen, are shown for zero and maximum deflection in Figs. 9(a) and 9(b) respectively. With no deflection, the electrons are converged to a point on the screen by the objective lens. Upon application of deflection voltages, the electrons are given additional convergence (in the direction of deflection only) by the deflecting field such that the edge electrons cross over before reaching the screen. This results in an enlarged spot in the direction of deflection and is denoted as *deflection focusing*. A qualitative explanation of the effect is that the electrons nearest to the positive deflection plate are at a higher average potential than other electrons in the beam. Consequently, they pass through the deflecting field faster. Since the electrostatic deflection field, E, is essentially constant in the y direction, the faster electrons are deflected less than are those moving more slowly. Hence the beam is converged by the deflection system and comes to a focus (in the y direction) in front of the screen. The electrostatic plates may then be regarded as a cylindrical

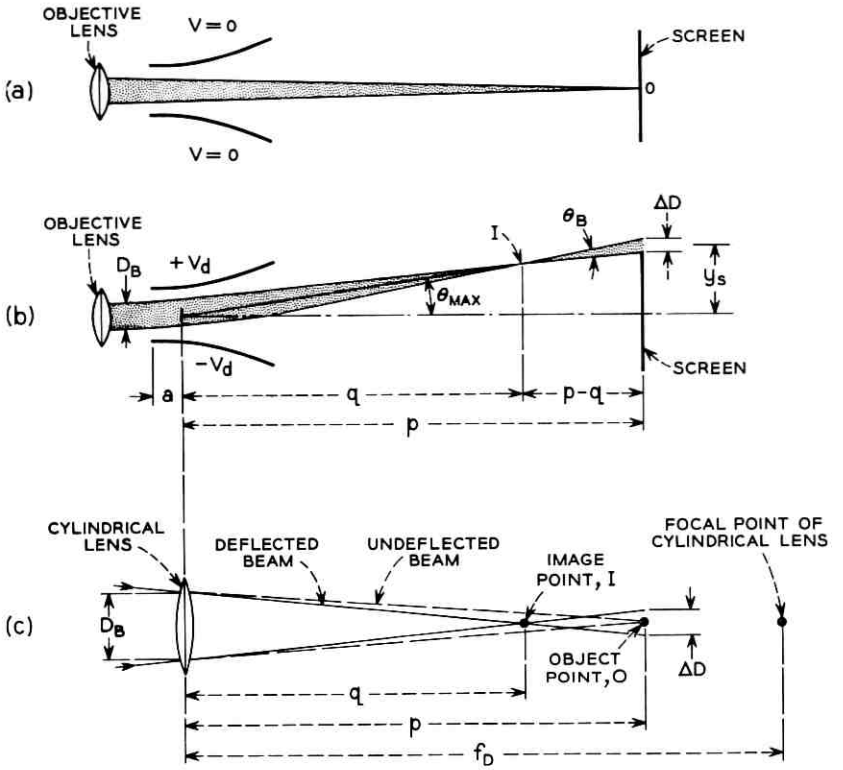


Fig. 9 — Illustration of deflection focusing for (a) undeflected beam; (b) beam at maximum deflection angle; (c) lens equivalent of deflection plates.

converging lens with an associated focal length f_D, as sketched in Fig. 9(c).

5.2.1 *Deflection Plate Focal Length, f_D, for Contoured Deflection Plates*

Pierce (Ref. 13, pp. 41–46) has derived the generalized equation for deflection plate focal length f_D, which is

$$\frac{1}{f_D} = 2\int_0^s \frac{1}{\cos^2\varphi}\left(\frac{d\theta}{ds}\right)^2 ds. \tag{2}$$

In (2), s is the distance along the electron path, φ is the angle between the deflection field E and the normal to the electron path, and θ is the angle between the electron path and the tube axis.

Focal length, f_D, can now be evaluated for the contoured plates which

were discussed in Section 5.1. It will be assumed that the deflection angle, θ, is sufficiently small that (a) $\cos \varphi \simeq 1$ (the deflecting field is normal to the path); (b) $s \simeq z$, where z is the axial distance; and (c) $\tan \theta \simeq \theta$. In practice, θ is usually less than 10 degrees, so that all three of the above should be good approximations (neglecting fringing fields). Equation (2) can then be written as

$$\frac{1}{f_D} = 2 \int_0^L \left(\frac{d \tan \theta}{dz} \right)^2 dz, \tag{3}$$

where L is the total axial length of the deflection plates. From the derivation[12] of (1) for the contoured plates, it can be shown that

$$\frac{d \tan \theta}{dz} = \frac{V_d}{2 V_0 y} \quad \text{and} \quad dz = \frac{dy}{\left(\frac{V_d}{V_0} \ln \frac{y}{y_0} \right)^{\frac{1}{2}}}, \tag{4}$$

where V_d, V_0, and y are defined in Section 5.1 and Fig. 7. From (4) and the substitution $y/y_0 = e^{u^2}$, (3) becomes

$$\frac{1}{f_D} = \frac{\sqrt{\pi}}{2 y_0} \left(\frac{V_d}{V_0} \right)^{\frac{1}{2}} \operatorname{erf} \left(\ln \frac{y_f}{y_0} \right)^{\frac{1}{2}}, \tag{5}$$

where erf is the error function* for which

$$\operatorname{erf} X \equiv \int_0^X \frac{2}{\sqrt{\pi}} e^{-u^2} du.$$

In Appendix B, f_D is related to the "zero spot size enlargement," ΔD, of a beam with zero undeflected beam size [see Fig. 9(c)]. From (17), ΔD is given by the equation

$$\Delta D = \frac{p}{f_D} D_B, \tag{6}$$

where p is the distance from the center of deflection to screen and D_B is the beam diameter at the deflection plates. The enlargement $\Delta\sigma$ of a Gaussian beam, unlike ΔD, is a function of beam size σ for any given deflection plate design. Hence the zero spot size enlargement, ΔD, is a fundamental parameter of a deflection system, where $\Delta\sigma$ may be obtained from the curve of Fig. 22 for any specific σ and ΔD.

5.3 *Deflection System to Screen Distance, Plate Separation, and Plate Length*

For a fixed distance, $z_{D\text{-}S}$, from the deflection system to screen, two critical parameters in the design of a precision electrostatic deflection

* See, for example, Peirce.[14]

system are the plate separation $2y_0$ and the axial plate length L. A small separation increases deflection sensitivity, but at the same time it enhances beam aberrations produced by fringing fields. Similarly increasing the plate length diminishes the deflection focusing effects (Ref. 13, pp. 41–46), but increases tube length and plate capacitance. Hence not all of the tube characteristics can be optimized simultaneously, and compromises must be made in accordance with design objectives.

The deflection plate design procedure used for the CRT can be summarized in two basic steps:

1. The plate length, L, was computed for both pairs such that the maximum spot enlargement, $\Delta\sigma$, due to deflection focusing was just equal to the maximum permissible value. This allows the deflection plate-to-screen distance $z_{D\text{-}S}$ to be minimized.

2. After L was obtained, the spacing $2y_0$ was made as large as deflection sensitivity requirements would permit. This minimized fringing field aberrations.

As discussed previously, the spot size objective was $\sigma = 0.0045 \pm 0.00075$ inch. The ΔD and $\Delta\sigma$ values computed in Appendix B are based on best focus at the screen center. Although this restriction permits a formulation of deflection design theory, it does not yield the best spot uniformity. That is, the CRT focus control should be adjusted for minimum beam diameter at a screen location intermediate between the center and corner to achieve best performance. Experience indicates that a midway point is about optimum and that, under this condition of best focus, the $\Delta\sigma = \sigma_{\max} - \sigma_{\min}$ value is about one half of that computed when it is assumed that the best focus is at the screen center. Hence the $\Delta\sigma$ value used to obtain ΔD from Fig. 23 was $2(\sigma_{\max} - \sigma_{\min})$. This is 0.0015 inch $\times$ 2 = 0.003 inch, which corresponds to $\Delta D = 0.021$ inch. Theoretically this ΔD figure should be used for both sets of plates, but experience indicated that the calculated ΔD should be slightly less for the target set than for the gun plates, in order that the experimentally observed $\Delta\sigma$'s would be the same for both. Thus ΔD values of 0.025 and 0.018 inch were selected for gun and target plates respectively. The result was that this overshot the goal slightly (they should have been closer together) but not enough to warrant deflection plate redesign.

It should be stressed that the primary objective of the deflection design procedure outlined below is to equalize ΔD (and hence $\Delta\sigma$) for the two pairs of plates. The particular ΔD value selected depends on the CRT to be designed but, once parameters such as deflection-to-screen distance $z_{D\text{-}S}$, accelerating voltage V_0, screen size y_s, etc. have been specified, the best $\sigma_{\max}/\sigma_{\min}$ ratio will be achieved *when the observed* ΔD *(or* $\Delta\sigma$*) is made the same for both horizontal and vertical deflection plates.*

The design values for spacings $2y_0$ and $2y_f$, deflection-to-screen distance $z_{D\text{-}S}$, plate length L, and zero beam size enlargement ΔD were obtained by the following iterative procedure:

1. A value is assumed for distance $z_{D\text{-}S}$ from the beginning of the deflection system to the screen and also for the initial plate separation $2y_0$. The location of center of deflection is estimated and the distance p from it to the screen is computed.

2. From the estimated values of p and the required screen deflection distance y_s, the approximate deflection angle θ (taken relative to the axis) is computed, where $\tan \theta = y_s/p$.

3. The deflection plate flare ratio y_f/y_0 is computed from the equation

$$\frac{y_f}{y_0} = e^{V_0/V_d \tan^2 \theta}$$

[obtained by integrating (4)], where $V_0 = 10$ kilovolts and $V_d = 250$ volts are fixed by tube design objectives. From this ratio and the assumed $2y_0$ value, the final plate separation $2y_f$ is obtained.

4. From the computed y_f/y_0 ratio, the quantity

$$\frac{z_f - z_0}{2y_0 \sqrt{\frac{V_0}{V_d}}}$$

is obtained from Fig. 8, and the axial plate length L, which is equal to $z_f - z_0$, is determined (y_0, V_0, and V_d are known).

5. The exact location of center of deflection (quantity a in Fig. 9) is calculated from the equation

$$a = L - \left(\frac{y_f - y_0}{\tan \theta}\right).$$

If it differs from the value assumed in step 1, steps 2 to 5 are repeated with the new a value. The zero spot size enlargement ΔD is determined using (5) and (6), where $\Delta D = pD_B/f_D$.

If the ΔD value obtained in step 5 differs appreciably from that desired, as it almost certainly will after the first attempt, then the values assumed for $z_{D\text{-}S}$ and/or $2y_0$ are altered until the proper ΔD is attained. This iterative procedure is followed first for the pair of plates nearest the electron gun (gun set) and then for the target set until the desired ΔD values are reached for the two pairs of plates. After some experience, the final design can be reached after about six series of computations. Table II includes the final sequence of calculations for the CRT design. Important design values are: (a) distance from the gun plates to the

TABLE II—FINAL VALUES FOR DEFLECTION PLATE DESIGN

Fixed parameters.....	V_0 = 10 kilovolts; V_d = 250 volts;* y_s = 3 inches; $D_{\text{lim.ap.}}$ = 0.113-inch diameter	
Assumed parameters....	$2y_0$ = 0.500 inch, $z_{D\text{-}S}$ = 24.5 inches; axial distance between the two pairs of deflection plates = 0.75 inch	
	Gun Set	Target Set
$\tan\theta \left(= \frac{y_s}{p}\right)$	$\frac{3 \text{ inches}}{23.0 \text{ inches}} = 0.130$	$\frac{3 \text{ inches}}{18.35 \text{ inches}} = 0.165$
θ (relative to axis)	7.5 deg	9.4 degrees
$\frac{y_f}{y_0} = e^{V_0/V_d \tan^2\theta}$	1.98	2.98
$\frac{L}{2y_0}\left(\frac{V_0}{V_d}\right)^{\frac{1}{4}}$ (from Fig. 8)	1.07	1.60
Center of deflection, $\left(a = L - \frac{y_f - y_0}{\tan\theta}\right)$	1.5 inches	2.0 inches
Center of deflection to screen, p	23.0 inches	18.35 inches
$D_B = \frac{D_{\text{lim.ap.}}p}{z_{D-S} + 1 \text{ inch}}$†	$\frac{0.113 \text{ inch} \times 23.0 \text{ inches}}{25.5 \text{ inches}} = 0.102 \text{ inch}$	$\frac{0.113 \text{ inch} \times 18.35 \text{ inches}}{25.5 \text{ inches}} = 0.081 \text{ inch}$
$\frac{1}{f_D} = \frac{\sqrt{\pi}}{2y_0}\left(\frac{V_d}{V_0}\right)^{\frac{3}{4}} \cdot \text{erf}\left(\ln\sqrt{\frac{y_f}{y_0}}\right)$	$\frac{1}{95 \text{ inches}}$	$\frac{1}{83 \text{ inches}}$
$\Delta D = \frac{D_B p}{f_D}$	0.025 inch	0.018 inch

* Corresponding to a deflection sensitivity of approximately 150 volts per inch.
† Distance from objective lens to entrance of the gun plates is 1 inch.

screen $z_{D\text{-}S}$ = 24.5 inches, (b) initial plate separation $2y_0$ = 0.500 inch for both sets of plates, and (c) axial lengths of 3.4 inches and 5.1 inches for gun and target plates respectively. It may be noted that, with a limiting aperture of 0.113-inch diameter at the objective lens, the electron beam occupies only about 20 to 25 per cent of the initial plate spacing $2y_0$. Considerably higher deflection sensitivity (lower V_d values) could be achieved by decreasing $2y_0$, but this was not done since the sensitivity was adequate for system design. Also, fringing field and deflection focusing aberrations would then be increased. The plate contour (y versus z) can be obtained from Fig. 8, letting $z = L$ when $y = y_f$.

5.4 *Effects of Fringing Fields*

In all of the deflection-system considerations thus far, fringing fields have been neglected. During the course of the CRT development, it was found that very appreciable aberrations were introduced by fields between the two sets of conventionally designed electrostatic plates. For purposes of discussion, the gun and target sets of plates will be denoted as the vertical and horizontal pairs respectively. Similarly, the respective enlargements along the vertical and horizontal axes are designated $\Delta\sigma_V$ and $\Delta\sigma_H$. The effects of deflection aberrations, as observed on the screen, are shown in the distortion patterns of Fig. 10, where enlargements are exaggerated for purposes of illustration. The left side of the figure depicts the spot distortion pattern expected from theory, where the deflection plates are assumed to be converging cylindrical lenses that distort the beam in the direction of deflection only. It may be noted, for example, that when only vertical deflection voltages are applied the beam is expected to enlarge only along the vertical axis (the enlargement being denoted as $\Delta\sigma_{V_0}$). When measurements were made on experimental tubes, however, the distortion pattern observed was as shown at the right of Fig. 10. It differs considerably from the expected, or "normal," pattern and consequently is denoted as "abnormal." The normal spot size enlargements are denoted as $\Delta\sigma_{V_0}$ and $\Delta\sigma_{H_0}$, and the abnormal ones as $\Delta\sigma_{V_1}$ and $\Delta\sigma_{H_1}$. As will be described below, the abnormal focusing effects are due to the fringing fields between the two pairs of plates, and they can be restored to a normal pattern by appropriate shaping of these fields.

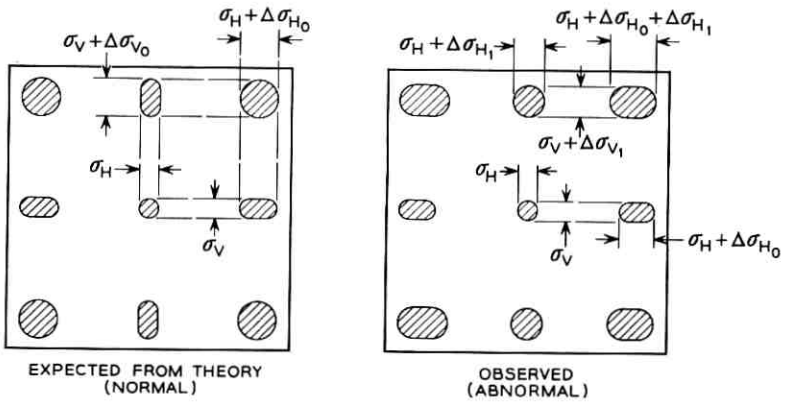


Fig. 10 — Deflection distortion patterns as observed on the CRT screen.

5.4.1 *Explanation of Distortion Pattern*

A feature of the abnormal pattern worth noting is that the beam behaves essentially normally for horizontal deflection only (where $\Delta\sigma_V \simeq 0$ and $\Delta\sigma_H \simeq \Delta\sigma_{H_0}$). When the beam was only deflected vertically, however, $\Delta\sigma_V$ was less than expected, but the beam enlarged nearly as much along the horizontal as in the vertical direction ($\Delta\sigma_{H_1} \simeq \Delta\sigma_{V_1}$). Thus, when the beam is deflected to the corners of the raster, the horizontal enlargements $\Delta\sigma_{H_0}$ and $\Delta\sigma_{H_1}$ are substantially additive. This results in a highly egg-shaped beam at the corners and the $\sigma_{\max}/\sigma_{\min}$ ratio reaches exceedingly large values. It should be noted that it is primarily the vertical deflection that produces the tremendously large horizontal enlargement at the corners. Also, the observed enlargement $\Delta\sigma_{V_1}$ in the vertical direction is less than $\Delta\sigma_{V_0}$ predicted by theory. This is not very consoling, however, since the damage to spot size uniformity already has been done.

The abnormal distortion pattern can be explained qualitatively by the focusing action of the fringing field between the two sets of deflection plates. Section A-A of Fig. 11 is a view of this interplate fringing field looking down the axis toward the screen with only vertical deflection voltages applied. There are skew-line electric forces acting on the beam, which are schematically shown in the figure. The beam position in the fringing field is above the axis, near the positive vertical plate. Looking qualitatively at the net focusing action, it may be seen (from the diagram of forces acting on peripheral electrons in Fig. 11) that the fringing field is *diverging* along the vertical and *convergent* in the horizontal direction.

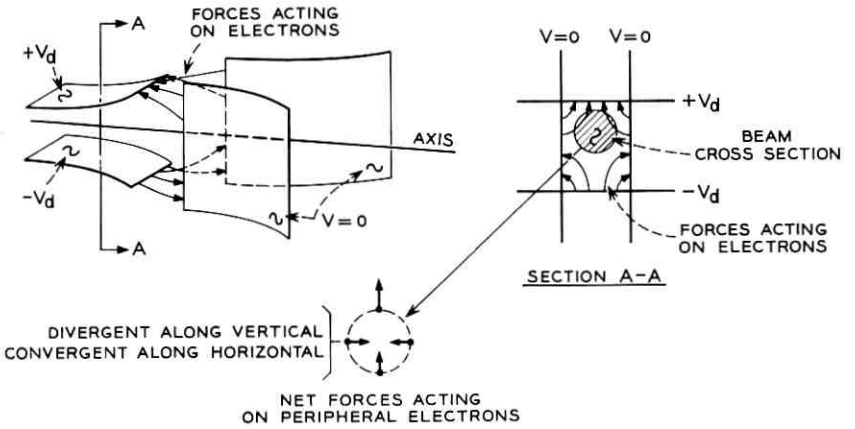


Fig. 11 — Focusing action of interplate fringing field without shields.

Hence there is a net horizontal convergence even though the beam is deflected only vertically, which explains the $\Delta\sigma_{H_1}$ of Fig. 10. Also, the net divergent action of the interplate fringing field in the vertical direction cancels out some of the predominant convergence of the vertical deflection field. Consequently, as previously noted, the amount of spot size enlargement $\Delta\sigma_{V_1}$ is less than the $\Delta\sigma_{V_0}$ calculated.

5.5 *Dual Deflection Shield*

It is clear that the abnormal deflection focusing can be produced by the skew fringing fields in the region between the two sets of plates. These skew lines are in turn caused by overlap or interpenetration of the fringing fields where the beam exits the vertical plates and enters the horizontal set. Hence some type of isolation (or shielding) of the two fields is needed. A single shield between plates, as indicated in Fig. 12(a), is frequently used, presumably to perform fringing-field isolation and provide electrostatic shielding between horizontal and vertical deflection signals. When the single shield was evaluated in experimental testers, no improvement in resolution uniformity was found. As may be seen from the data presented in Fig. 13, $\Delta\sigma_{H_1}$ increased while $\Delta\sigma_{H_0}$ decreased when just a single shield was used. This represents an enhanced abnormal effect, and the $\sigma_{\max}/\sigma_{\min}$ ratio remained very high but about constant for all the no-shield and single-shield testers that were studied.

It was concluded that more complete shielding was needed between the two pairs of plates. One approach to the problem is to increase the axial separation between the two plate pairs, but this has disadvantages of increasing tube length and/or the deflection angle of the target set of plates. An alternative method is to provide better electrostatic shielding between the two interplate fringing fields. The dual shield of Fig. 12(b)

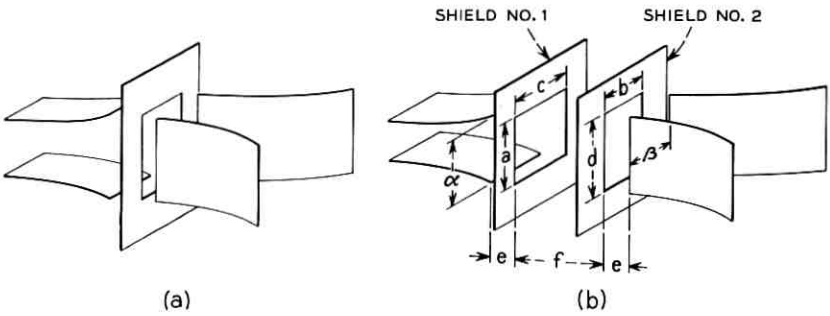


Fig. 12 — Deflection shields: (a) single; (b) dual.

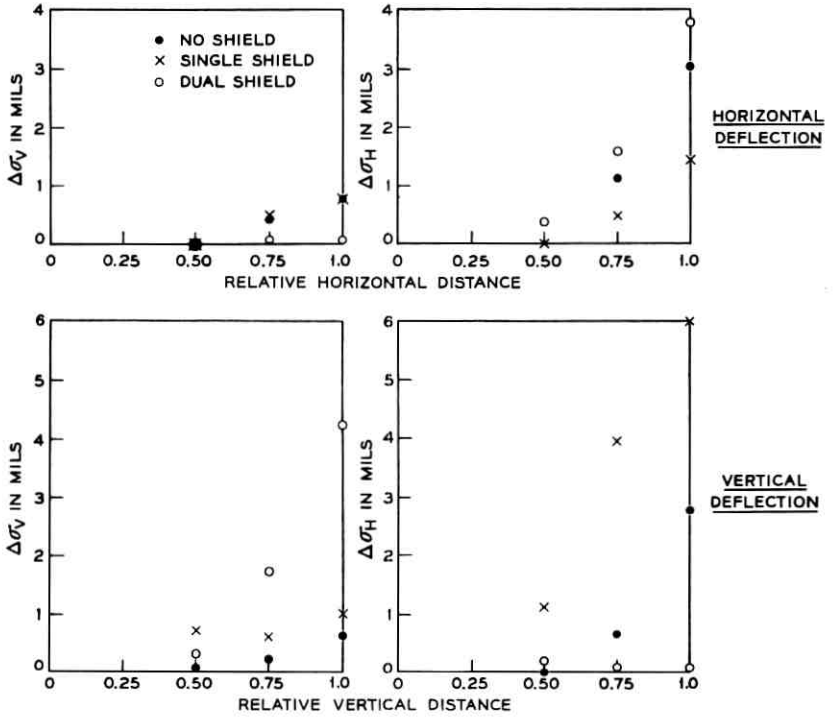


Fig. 13 — Spot distortion as a function of distance deflected on screen.

was devised for this purpose. It is compact and, as may be seen from Fig. 13, it virtually eliminates the abnormal deflection focusing effect (both $\Delta\sigma_{H_1}$ and $\Delta\sigma_{V_1}$ are very nearly normal for the dual shield). The normal spot size enlargements, $\Delta\sigma_{V_0}$ and $\Delta\sigma_{H_0}$, are increased with the dual shield. In spite of this, the over-all result is a marked improvement in resolution uniformity since the sum of $\Delta\sigma_{H_0} + \Delta\sigma_{H_1}$ is less (see Section 5.4.1).

The two shields have mutually perpendicular rectangular apertures, in which the short dimensions, a and b in Fig. 12(b), are approximately equal to the separation, α and β respectively, of the nearest set of plates. The long dimensions, c and d, are made as large as is mechanically feasible. Plate-to-shield axial spacing e is as low as possible, and the shield separation f (0.5 inch for the CRT) is just large enough to provide adequate fringing field isolation. Basically the dual shield concentrates the exit field of the vertical set of plates and the entrance field of the horizontal pair such that the field overlap is effectively zero. The shields are

operated at the average deflection plate potential (final accelerating voltage of the CRT) and therefore no external leads to the shields are necessary.

5.6 *Alignment of Beam Axis with Midplane of Target Deflection Plates*

With the high degree of precision required from the flying spot store CRT, there are many tube dimensions that must be held to unusually small tolerances. One of the most important considerations in this regard is alignment of the electron beam with the center (or midplane) of the horizontal (target) deflection plates. The degree of accuracy required was studied by applying a horizontally deflecting magnetic field between the objective lens and target set of plates. It was found that the spot size uniformity ratio was degraded noticeably when the beam was deviated only slightly from horizontal deflection plate midplane. The maximum amount of misalignment which could be tolerated in the CRT was estimated as ± 0.025 inch. This necessitates very precise alignment of the tetrode gun and objective lens axes with that of the target deflection plates. From a similar investigation at the gun plates, it was concluded that the degree of alignment with the midplane of the vertical set is not nearly so critical.

VI. FACEPLATE AND SCREEN

The screen consists of a P16 phosphor settled on a flat faceplate and then aluminized. The phosphor decay time (or persistence) is approximately 100 millimicroseconds after preaging. A component of radiated light with longer decay time is aged out by a treatment of 0.04 coulomb per cm^2 of cumulative charge at 10 kilovolts (this takes about three days with $I_B = 50$ microamperes). The aging treatment also yields a more uniform light output over the screen area during system use, which prolongs the screen life. A screen weight of 1.5 milligrams per cm^2 yielded the optimum light output for the batch of P16 material that was used. Measurements of the electron beam size and optical observations of the luminous spot size made on the same CRT showed that the increase in σ due to light scatter in the screen was less than one per cent, when σ was 0.0045 inch.

Another screen requirement was uniformity of light output (radiant flux) from the quality area. The necessary uniformity was achieved using screen fabrication techniques developed by G. Helmke and A. Pfahnl. Maximum variations in radiant flux are typically maintained between $+20$ and -30 per cent from the median flux.

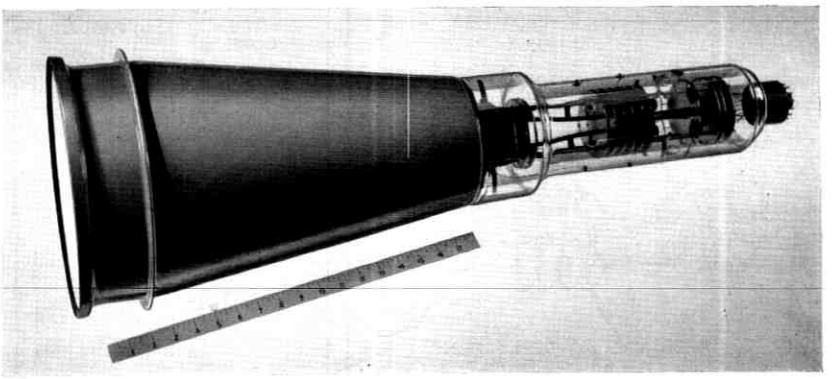


Fig. 14 — The CRT for the flying spot store.

As discussed in Section II, the glass faceplate must be of optical quality, very free from flaws and must meet rigid flatness and dimensional specifications. The objectives on glass blemishes were 0.004 inch maximum flaw diameter with a minimum spacing of 0.050 inch between flaws. This was achieved* by custom melting the glass, rolling to the proper thickness, and selecting only those areas that met specified tolerances. The plates were then cut, ground to an "F" optical quality surface, and sealed into a metal rim. The radius of faceplate curvature was maintained above 1900 inches (after tube evacuation) by means of special sealing techniques. The faceplate diameter is 10 inches, which permits an $8\frac{1}{2}$-inch diameter quality area and results in a maximum tube diameter of $10\frac{1}{2}$ inches.

VII. CHARACTERISTICS AND PERFORMANCE

A completed tube is shown in Fig. 14, with a schematic cutaway view in Fig. 15. Typical operating voltages and other tube characteristics are tabulated in Table III. Of the various characteristics listed in the table, it may be noted that the degree of spot size uniformity is somewhat better than the original system objectives. In the average tube, the spot size σ is maintained within the range 0.0045 $\pm$ 0.0006 inch for constant focus and over a beam current range between 4 and 20 microamperes. Beam current, I_B, is plotted as a function of both cathode current, I_K, and control grid bias, V_g, in Fig. 16 for $V_{A_1} = 625$ and 1000 volts and

* Faceplate development, as well as all mechanical design, was done by a group under the direction of C. Maggs and J. W. West.

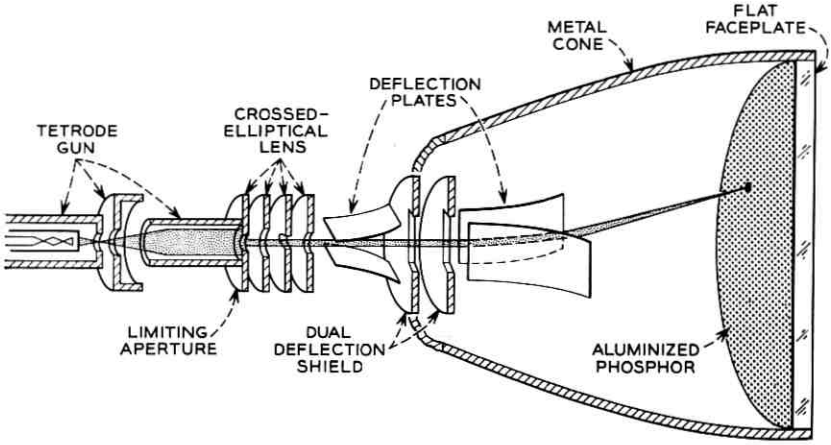


Fig. 15 — Cutaway schematic view of the flying spot store CRT.

TABLE III—FLYING SPOT STORE CRT CHARACTERISTICS

Typical Operating Potentials*	
First anode (A_1)	0.5 to 1.0 kilovolt
Second anode (A_2)	10 kilovolts
Vertical focus anode (A_4)	5 kilovolts
Horizontal focus anode (A_3)	5 kilovolts
Control grid	
(a) cutoff	−100 to −70 volts
(b) operating	−90 to −50 volts
Heater	6.8 volts
Tube Characteristics	
Deflection sensitivity	150 ± 10 volts per inch
Beam current, I_B	4 to 20 microamperes
Maximum over-all length	36 inches
Maximum over-all diameter	10½ inches
Minimum faceplate radius of curvature	1900 inches
Quality area	6 × 6 inches
Screen	
(a) phosphor type	P16 (aluminized)
(b) uniformity of radiant flux	+20 per cent, −30 per cent from average flux
Spot size σ (for I_B between 4 and 20 microamperes at constant focus)	0.0045 ± .0006 inch
Spot size uniformity ratio, $\sigma_{max}/\sigma_{min}$	<1.3
Deflection plate capacitance (per plate for push-pull deflection)	16 μμf (gun set) 24 μμf (target set)

* All expressed relative to cathode potential.

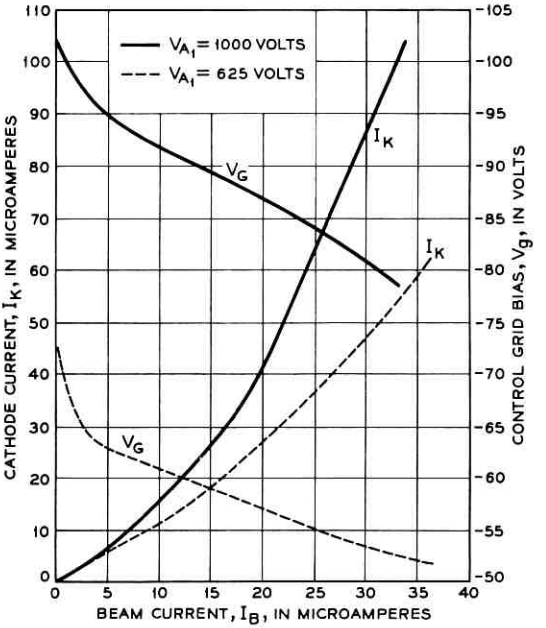


Fig. 16 — Tetrode gun characteristics.

V_{A_2} = 10 kilovolts. It is worth noting that the current efficiency I_B/I_K increases markedly as V_{A_1} is reduced from 1000 to 625 volts. This is due to the convergent lens action of the A_1–A_2 electrodes, as discussed in Section III.

7.1 σ_{avg} *as a Function of* V_{A_1} *and* V_{A_2}

One of the system objectives is to maintain the mean spot size, σ_{avg}, at a constant value of 0.0045 inch for all tubes. Because of the extreme sensitivity to gun electrode spacings, it is quite difficult to control spot size to that degree of precision from tube to tube when the crossover is formed by an immersion-lens gun. Therefore some electrical control of σ_{avg} is desirable, which can be done very conveniently in this CRT design by means of the first anode, A_1, potential. Fig. 17 is a plot of σ_{avg} vs V_{A_1} for a typical tube. It may be seen that a relatively wide σ_{avg} range, from 0.0040 to 0.0048 inch, is achieved as V_{A_1} is decreased from 1000 to 500 volts. The 0.0045 inch system objective is attained in a typical tube with V_{A_1} = 625 volts. Also shown in Fig. 17 is the change in optimum V_{A_3} and V_{A_4} focus voltages with V_{A_1}. A very desirable feature of the

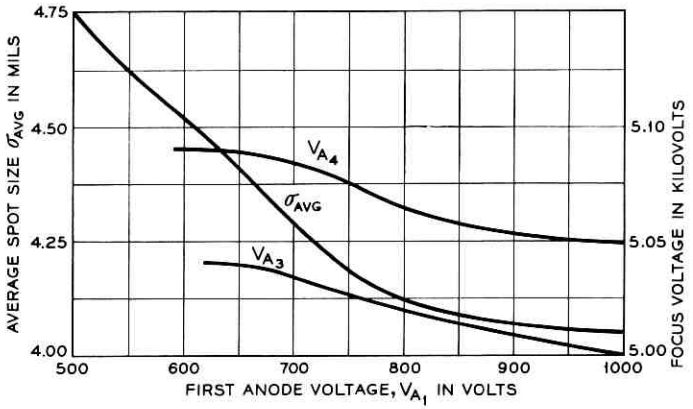


Fig. 17 — Focus voltages and σ_{avg} as a function of V_{A1}.

gun and lens design is that this variation is small, less than a one per cent increase (50 volts) in both V_{A_3} and V_{A_4}, as V_{A_1} is reduced from 1000 to 600 volts.

Another interesting capability of the CRT is that of decreasing σ_{avg} by changing the A_2 potential of the tetrode gun (the last aperture of the crossed-elliptical lens, and all subsequent electrodes are maintained at 10 kilovolts). The reason for this is that the magnification M is proportional to $\sqrt{V_{A_2}/V_{acc}}$. Fig. 18 is a plot of σ_{avg} as a function of V_{A_2} for $V_{A_1} = 1$ kilovolt, and the final accelerating potential, $V_{acc} = 10$ kilovolt.

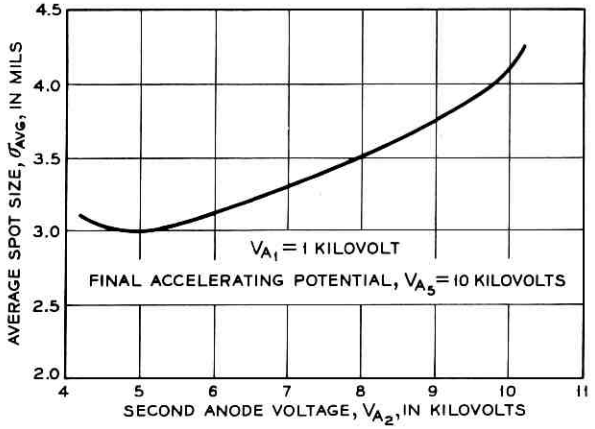


Fig. 18 — Average spot size as a function of V_{A_2}.

The average σ value may be reduced to as low as 0.003 inch by this technique, but it should be noted that current efficiency is decreased. Also, as may be seen from Fig. 23, the spot size uniformity ratio is inherently degraded with decreasing σ. This is because $\Delta D/\sigma$ increases at lower σ values, which results in an enlarged $\Delta\sigma/\sigma$ value. Since the system was designed for σ_{avg} at 0.0045 inch, V_{A_1}, rather than V_{A_2}, is varied to control the value of σ.

7.2 *Uniformity of Radiant Flux from Screen*

A typical distribution curve of radiant light flux from the quality screen area is given in Fig. 19. The brightest and dimmest spots are 20 per cent above and 30 per cent below the mean light output respectively. However, the radiant flux of over 95 per cent of the quality area is within a range of ± 10 per cent of the median output level. It should be pointed out that precision screen settling techniques are required to achieve the relatively narrow distribution of Fig. 19. Important considerations are rigid dust control in the settling room, removal of oversize phosphor particles, uniform screen weight, and aluminizing techniques which yield a smooth continuous aluminum layer. Also, it was found that the aging operation, in addition to removing the long decay time component of P16, produced a more uniform radiant flux distribution.

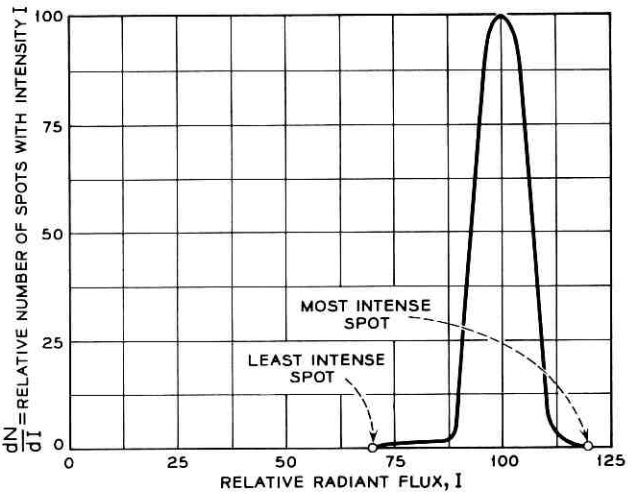


Fig. 19 — Typical distribution curve of light output from quality area.

VIII. SUMMARY AND CONCLUSIONS

Generalized step-by-step procedures have been presented for designing large-area electrostatically deflected cathode ray tubes for high and exceedingly uniform resolution. Both electron gun and electrostatic deflection plate design have been theoretically analyzed. Equations were evolved which yield an optimum combination of resolution, spot uniformity, and tube length.

There are three basic electron-optical design problems, namely those of the deflection system, electron gun, and lens. They may be summarized as follows:

1. Deflection plate length and deflection to screen distance are designed such that the maximum spot size enlargement, $\Delta\sigma_{max}$, produced by deflection distortion is the same for both pairs of plates and is exactly equal to the maximum permissible value. Equations for computing $\Delta\sigma_{max}$ as a function of beam size σ were derived for the case of deflection plates contoured for maximum sensitivity. Plate contour and minimum separation were then selected to provide the desired sensitivity, the separation between the plates in a given set being made as large as possible to reduce fringing field and deflection focusing aberrations. A novel feature that resulted from this work was a dual shield between the two pairs of plates which provides a substantial reduction of deflection aberrations.

2. A tetrode gun design is described which has properties of small crossover size, low cathode current density, high current efficiency, and low magnification M of the crossover at the screen. Equations for M have been developed, from which the minimum value of M can be found when the focal lengths and principal planes of the lens formed by the first and second anodes are known. Once the M value is fixed, the overall tube length is then determined from the deflection-to-screen distance obtained in step 1.

3. The electron lens and limiting aperture size are selected such that aberrations of both the lens and deflection fringing field are sufficiently small.

Using the above design principles, a cathode ray tube has been developed for the flying spot store which has exceptional spot size uniformity and meets system objectives. Typical performance characteristics of this tube are a Gaussian beam size, σ, of 0.0045 ± 0.0006 inch which can be maintained over a 6- by 6-inch square screen area for all beam currents from 4 to 20 microamperes without changing focus potentials. The 0.0045-inch σ value corresponds to a resolution of 2000 TV lines per useful target diameter and is electrically variable by controlling the first anode potential.

The faceplate is very flat (greater than 1900-inch radius), of moderate optical quality, and very free of flaws (less than 0.004-inch diameter with a minimum spacing of 0.050 inch). Precision preparation methods of the settled P16 screen yield a very uniform radiant (or light) flux output (typically within +20 and −30 per cent of the median value) at all points in the quality area.

IX. ACKNOWLEDGMENTS

Many people have made development of the flying spot store CRT possible. The project was under the general direction of J. A. McCarthy and R. W. Sears, whose suggestions and guidance have been most valuable. C. Maggs and J. W. West directed the faceplate development and mechanical design aspects of the work. Other important contributors are: G. E. Helmke and A. Pfahnl, development of the screen settling techniques; M. Poulsen, glass-to-metal faceplate seal; A. M. Johnson and J. J. Ulozas, tube evaluation and processing; and H. W. Ericsson and G. J. Kossyk, mechanical design.

APPENDIX A

The electron optical focusing action on the electron beam as it passes from the crossover to screen is depicted in Fig. 20(a). There are two lenses, the A_1–A_2 gap and the objective lens, which act to image the crossover onto the target. The purpose of this appendix is to compute the magnification M of the complete lens system. Fig. 20(b) shows the principal planes, object, image, and focal points of the two lenses. Conventional electron lens terminology[7] is used, where the quantities are positive as shown in Fig. 20(b). A beam acceleration from 1 to 10 kilovolts occurs at the A_1–A_2 gap, and it should be regarded as a thick rather than a thin lens.[7] Hence, it has two principal planes, P_1 for the object space and P_1' for the image space. Focal lengths are f_1 and f_1' respectively for object and image. Likewise, the distances from object and image points to their respective focal points are X_1 and X_1'. Similar notation holds for the objective lens, except that this lens is an einzel type and can to a good approximation be represented as a thin lens (Ref. 13, pp. 98–101). Thus the principal planes are coincident ($P_2 = P_2'$) and also $f_2 = f_2'$. It should be noted that Fig. 20(b) is drawn to indicate positive lens parameters and that actually the image point of the A_1–A_2 lens will lie beyond the target outside the tube, with the result that X_2 is negative. Also, the object focal point of the A_1–A_2 lens can be to the left of the crossover, and X_1 then becomes negative.

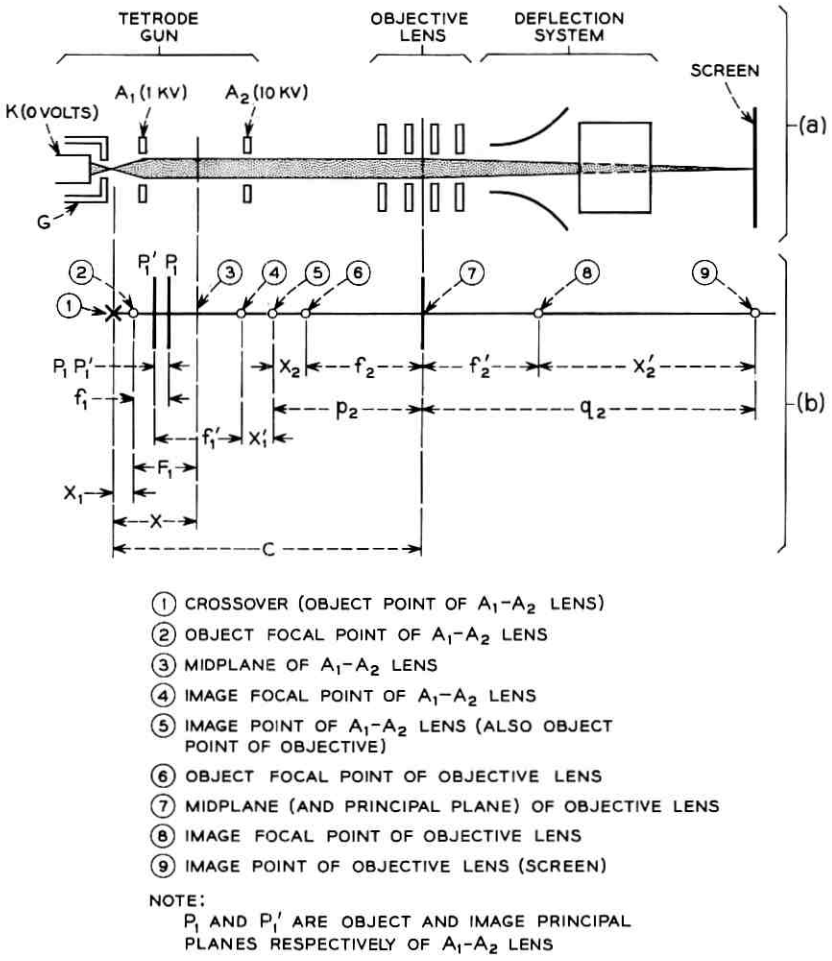


Fig. 20 — Notation used for calculation of crossover magnification M by A_1–A_2 and objective lenses.

First, the magnification M_1 of the A_1–A_2 lens and M_2 of the objective lens will be computed independently and then the combined magnification M of the two lenses in series will be derived. The thick-lens formula must be used for the A_1–A_2 lens, and the magnification M_1 is given by

$$M_1 = \frac{f_1}{X_1} = \frac{X_1'}{f_1'}, \tag{7}$$

where the quantities are as shown in Fig. 20(b). For the case of the thin lens (objective lens) $f_2 = f_2'$ and the expression for M_2 is therefore[7]

$$M_2 = \frac{X_2' + f_2}{X_2 + f_2} = \frac{q_2}{p_2}. \tag{8}$$

The total magnification M of both lenses[7] is

$$M = M_1 M_2 = \frac{f_1}{X_1} \frac{q_2}{p_2}. \tag{9}$$

Now the distance C in Fig. 20(b) between the crossover and objective lens midplane, a basic design value of the tube, is equal to

$$C = p_2 + X_1' + f_1' + X_1 + f_1 - \overline{P_1P_1'}. \tag{10}$$

Substituting (10) into (9), the expression for M becomes

$$M = \frac{f_1}{X_1} \frac{q_2}{(C + \overline{P_1P_1'} - X_1 - f_1 - X_1' - f_1')}. \tag{11}$$

The image distance X_1' and image focal length f_1' of the A_1–A_2 lens can be eliminated from (11) by the fundamental electron lens equations,[7]

$$X_1' = \frac{f_1 f_1'}{X_1}, \qquad f_1' = \sqrt{\frac{V_2}{V_1}} f_1. \tag{12}$$

Equation (11) then becomes

$$M = \frac{f_1 q_2}{X_1 \left(C + \overline{P_1P_1'} - X_1 - f_1 - \sqrt{\frac{V_2}{V_1}} \frac{f_1^2}{X_1} - \sqrt{\frac{V_2}{V_1}} f_1 \right)}. \tag{13}$$

Note that signs for all quantities in (13) must be carefully designated, where positive values are shown in Fig. 20. The distance X from the crossover to the midplane of the A_1–A_2 lens is given by relation

$$X = X_1 + F_1, \tag{14}$$

where F_1 is the focal length measured from the midplane.[7]

The parameters C and q_2 are basic design values of the tube and, for reasons of beam size, beam size uniformity, and deflection factor, were selected as $C = 5.0$ inches and $q_2 = 25.5$ inches. As mentioned previously $V_1 = 1$ kilovolt and $V_2 = 10$ kilovolts. Values of f_1, F_1, and $\overline{P_1P_1'}$ for electrode arrangements (a), (b), and (c) of Fig. 3 have been evaluated by Spangenberg and Field[8] and are tabulated in Table IV. Substituting these values in (13) and (14), the magnification M can then be

TABLE IV

A_1–A_2 Geometry (see Fig. 3)	f_1 (inches)	F_1 (inches)	$\overline{P_1P_1'}$ (inches)	$\frac{V_2}{V_1}$
(a) cylinder — aperture	0.8	1.3	0.15	10
(b) two cylinders of unequal diameter	1.2	3.2	−0.2	10
(c) two equidiameter cylinders	1.5	2.55	−0.9	10

computed as a function of the distance X from crossover to the A_1–A_2 midplane. The results are plotted in Fig. 4.

APPENDIX B

Spot size enlargement, $\Delta\sigma$, due to deflection focusing can be calculated from knowledge of the equivalent deflection plate focal length, f_D, and distance, p, from the center of deflection to the screen. It will be done in two steps. First the enlargement, ΔD, will be computed for a beam whose spot size is equal to zero at the undeflected screen position. Next the spot size increase, $\Delta\sigma$, for a Gaussian beam will be evaluated from ΔD.

It will be assumed that f_D can be represented by a thin convergent cylindrical lens with the principal plane at the center of deflection (the point at which the projected center of the deflected beam intersects the axis). The equivalent electron-optical diagram is shown in Fig. 9(c). If an ideal beam is focused by the objective lens, to a point o on the screen (σ = zero) at the undeflected position, and then deflection signals are applied, the distance q between the center of deflection and new focal point I (see Fig. 9) is

$$\frac{1}{-p} + \frac{1}{q} = \frac{1}{f_D} \qquad \text{or} \qquad q = \frac{pf_D}{p + f_D}. \tag{15}$$

The parameter p is the distance from center of deflection to screen and is the object point for the lens. It should be noted that, for convenience, p is defined positive for the object point to the right of the principal plane. Hence $-p$ must be used in the thin lens equation (15). The angle θ_B at which the beam converges to the image point I in Fig. 9(b) is

$$\theta_B = \frac{D_B}{q}. \tag{16}$$

where D_B is the beam diameter at the center of deflection. The enlargement ΔD is then, using (15) and (16),

$$\Delta D = (p - q)\theta_B = \frac{p}{f_D} D_B. \tag{17}$$

Spot size enlargement, $\Delta\sigma$, for a Gaussian beam can now be obtained from ΔD. It will be assumed for purposes of discussion, that the beam is deflected in the y direction, and attention will be confined only to the current density distribution $J(y)$ with its associated standard deviation σ_y. At the undeflected screen position, $J(y)$ is a Gaussian beam distribution denoted as $J_1(y)$ and is given by

$$J_1(y) = J_0 e^{-y^2/2\sigma_y^2}, \tag{18}$$

where J_0 is the current density at the beam center ($y = 0$). The integral of $J_1(y)\,dy$ from $-\infty$ to $+\infty$ is the beam current I_B, from which it may be shown that $J_0 = I_B/\sqrt{2\pi}\sigma_y$. Upon deflecting the beam to the maximum angle, $\theta_{\max}$, the Gaussian $J_1(y)$ appears in front of the screen at point 1 of Fig. 21. The value of σ_y probably will be reduced if the distance from 1 to the screen becomes large, but for high resolution CRT's this effect should be of second order and therefore will be neglected here. As illustrated in Fig. 21, a point y_1 on the Gaussian at position 1 becomes enlarged to an area ΔD at the screen. Thus there is a new (and enlarged) distribution $J_2(y)$ at the screen given by

$$J_2(y) = \int_{y-(\Delta D/2)}^{y+(\Delta D/2)} \frac{J_1(y_1)\,dy_1}{\Delta D}. \tag{19}$$

The distribution $J_2(y)$ is no longer a Gaussian but, if ΔD is sufficiently small, it will not deviate greatly. Hence, in order to retain the previous spot size definition of σ, it will be assumed that J_2 is also a Gaussian and can be written as

$$J_2(y) = J_0' e^{-y^2/2(\sigma_y+\Delta\sigma_y)^2},$$

where $J_0' = I_B/\sqrt{2\pi}(\sigma_y + \Delta\sigma_y)$ and $\Delta\sigma_y$ is the spot size enlargement.

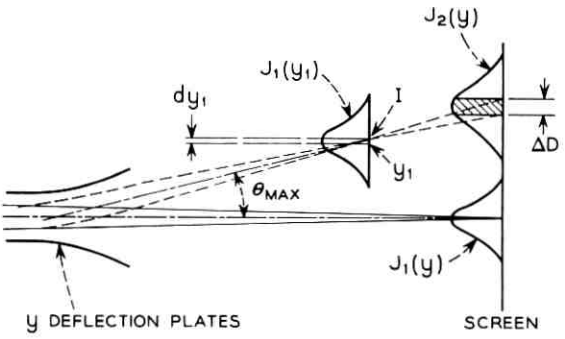


Fig. 21 — Relation of ΔD to Gaussian beam.

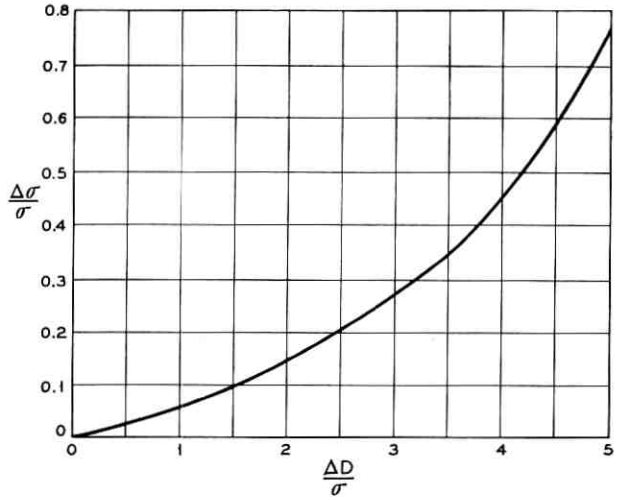


Fig. 22— Generalized relation of $\Delta\sigma$ to ΔD.

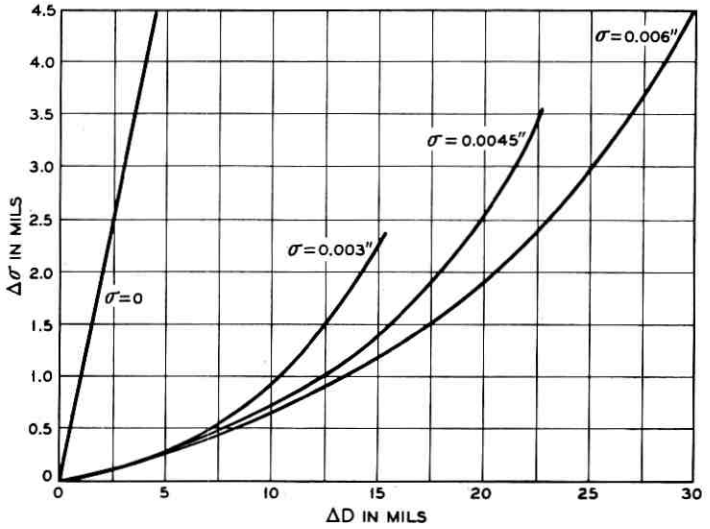


Fig. 23 — Enlargement $\Delta\sigma$ of a Gaussian beam as a function of zero beam size enlargement ΔD.

When $y = 0$, $J_2(y) = J_0'$ and (19) becomes

$$\frac{I_B}{\sqrt{2\pi}(\sigma_y + \Delta\sigma_y)} = \int_{-\Delta D/2}^{+\Delta D/2} J_0 \frac{e^{-y_1^2/2\sigma_y^2}}{\Delta D}\, dy_1. \tag{20}$$

By defining $u = \Delta D/2\sqrt{2}\sigma_y$, (20) may be solved for $\Delta\sigma_y/\sigma_y$, with the result

$$\frac{\Delta\sigma_y}{\sigma_y} = \frac{2u}{\sqrt{\pi}\ \text{erf}\ u} - 1, \tag{21}$$

where erf u is as defined in (5) of Section V. Equation (21) can be solved graphically and a plot of $\Delta\sigma/\sigma$ versus $\Delta D/\sigma$ is presented in Fig. 22 [the y subscripts in (21) are dropped in order to generalize]. In Fig. 23, $\Delta\sigma$ is shown as a function of ΔD for specific σ values of 0.003, 0.0045, and 0.006 inch. It may be noted that $\Delta\sigma$ increases superlinearly with ΔD. Also it decreases markedly with increasing σ at a fixed ΔD value. Thus it becomes clear that, when one is designing a CRT for a high degree of spot size uniformity (low $\Delta\sigma$), improvements are obtained at an exponential rate as ΔD is reduced and/or σ_{avg} is increased.

REFERENCES

1. Hoover, C. W., Jr., Staehler, R. E., and Ketchledge, R. W., Fundamental Concepts in the Design of the Flying Spot Store, B.S.T.J., **37,** 1958, p. 1161.
2. Hoover, C. W., Jr., Haugk, G., and Herriott, D. R., System Design of the Flying Spot Store, B.S.T.J., **38,** 1959, p. 365.
3. Joel, A. E., Jr., An Experimental Switching System Using New Electronic Techniques, B.S.T.J., **37,** 1958, p. 1091.
4. Cooper, H. G., McCarthy, J. A., and Sears, R. W., A High-Resolution CRT Employing Electrostatic Deflection, I.R.E. Prof. Group on Electron Devices Meeting, Washington, D. C., October 30, 1959.
5. Klemperer, O., *Electron Optics*, 2nd ed., Cambridge Univ. Press, Cambridge, 1953, p. 254.
6. Moss, H., The Electron Gun of the Cathode Ray Tube, J. Brit. I.R.E., **5,** 1945, p. 10; **6,** 1946, p. 99.
7. Spangenberg, K. R., *Vacuum Tubes*, McGraw-Hill, New York, 1948, Ch. 13.
8. Spangenberg, K. R., and Field, L. M., The Measured Characteristics of Some Electrostatic Electron Lenses, Elect. Comm., **21,** 1943, p. 194.
9. Spangenberg, K. R., and Field, L. M., Some Simplified Methods of Determining the Optical Characteristics of Electron Lenses, Proc. I.R.E., **30,** 1942, p. 138.
10. Kirkpatrick, W. E., et al., to be published.
11. Greenwood, T. S., and Staehler, R. E., A High-Speed Barrier Grid Store, B.S.T. J., **37,** 1958, p. 1195.
12. Maloff, I. G., and Epstein, D. W., *Electron Optics*, McGraw-Hill, New York, 1938, p. 200.
13. Pierce, J. R., *Theory and Design of Electron Beams*, 2nd ed., D. Van Nostrand, New York, 1954.
14. Peirce, B. O., *A Short Table of Integrals*, 3rd ed., Ginn & Co., Boston, p. 116.

Synthesis of Transformerless Active N-Port Networks

By I. W. SANDBERG

(Manuscript received August 16, 1960)

The following theorem is proved:

Theorem: An arbitrary symmetric $N \times N$ matrix of real rational functions in the complex-frequency variable (a) can be realized as the immittance matrix of an N-port network containing only resistors, capacitors, and N negative-RC impedances, and (b) cannot, in general, be realized as the immittance matrix of an N-port network containing resistors, capacitors, inductors, ideal transformers, and M negative-RC impedances if $M < N$.

The necessary and sufficient conditions for the immittance-matrix realization of transformerless networks of capacitors, self-inductors, resistors, and negative resistors follow as a special case of the theorem. In addition, an earlier result is extended by presenting a procedure for the realization of an arbitrary $N \times N$ short-circuit admittance matrix as an unbalanced transformerless active RC network requiring no more than N controlled sources. The passive RC structure has the interesting property that it can always be realized as a $(3N + 1)$-terminal network of two-terminal impedances with common reference node and no internal nodes. The active subnetwork can always be realized with N negative-impedance converters.

I. INTRODUCTION

The development of the transistor has provided the network synthesist with an efficient low-cost active element and has stimulated considerable interest in the theory of active RC networks during the last decade.

Several techniques have been proposed for the transformerless active RC realization of transfer and driving-point functions.[1-18] It has, in fact, been established that any real rational fraction (in the complex frequency variable) can be realized as the transfer or driving-point function of a transformerless active RC network containing one active element. In particular, Linvill's technique[3] has been the basis for much of the later work.

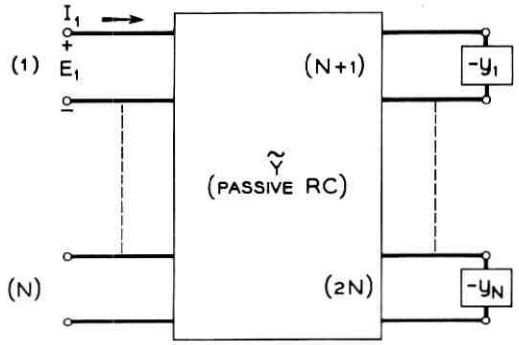


Fig. 1 — Realization of an arbitrary $N \times N$ symmetric immittance matrix.

It has recently been shown[19] that an arbitrary $N \times N$ matrix of real rational functions can be realized as the short-circuit admittance matrix of a transformerless N-port active RC network containing N controlled sources, and that in general all N controlled sources are required. These results have suggested the possibility of establishing the theorem stated in the abstract to this paper. The proof, presented in the next section, is based on a technique developed in an earlier paper for factoring a class of matrix-coefficient polynomials in a scalar variable. For the special case $N = 1$, our result reduces to that of Sipress.[18] *

We also present in Section II a procedure for the realization of an arbitrary $N \times N$ short-circuit admittance matrix as an unbalanced active RC network requiring no more than N controlled sources. The required passive RC network has the interesting property that it can always be realized as a $(3N + 1)$-terminal network of two-terminal impedances with common reference node and no internal nodes. This result not only displaces the balanced network assumption implicit in the proof given in Ref. 19, but is of considerable interest in its own right.

II. REALIZATION OF A SYMMETRIC IMMITTANCE MATRIX AS AN ACTIVE *RC* NETWORK CONTAINING NEGATIVE-*RC* IMPEDANCES

Consider a $2N$-port network of resistors and capacitors characterized by the short-circuit admittance matrix $\tilde{\mathbf{Y}}$ and suppose that a negative-RC admittance $-y_k$ is connected to port $N + k$ ($k = 1,2,\cdots,N$), as shown in Fig. 1. It is convenient to partition $\tilde{\mathbf{Y}}$ as follows:

* This case was first considered in detail by Kinariwala,[13] who showed that a broad class of driving-point functions could be realized.

$$\tilde{\mathbf{Y}} = \begin{matrix} \\ N \\ N \end{matrix} \begin{bmatrix} N & N \\ \mathbf{Y}_{11} & \mathbf{Y}_{12} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} \end{bmatrix} \begin{matrix} \\ N \\ N \end{matrix}. \tag{1}$$

The short-circuit admittance matrix $\mathbf{Y}$ relating the voltages and currents at ports $k(k = 1,2,\cdots,N)$ can readily be shown to be

$$\mathbf{Y} = \mathbf{Y}_{11} - \mathbf{Y}_{12}[\mathbf{Y}_{22} - \text{diag } (y_1,y_2,\cdots,y_N)]^{-1}\mathbf{Y}_{12}{}^t, \tag{2}$$

where the superscript t indicates matrix transposition.

We assume that $\mathbf{Y} = (1/D)[N_{ij}]$ is an arbitrary prescribed symmetric $N \times N$ matrix of real rational functions, where $[N_{ij}]$ is a matrix of polynomials and D is a common denominator polynomial. The synthesis technique requires that the three submatrices in (2) be determined so that $\tilde{\mathbf{Y}}$ is realizable as a transformerless RC network and that the elements in diag $(y_1,y_2,\cdots,y_N)$ be RC driving-point admittances.

The matrix $\tilde{\mathbf{Y}}$ can be expressed as

$$\tilde{\mathbf{Y}} = s\mathbf{K}_\infty + \sum_{m=0}^{M} \mathbf{K}_m \frac{s}{s + \gamma_m}, \tag{3}$$

where $\mathbf{K}_\infty$ and $\mathbf{K}_m$ are real symmetric coefficient matrices and the γ_m are real and satisfy

$$0 = \gamma_0 < \gamma_1 < \gamma_2 \cdots < \gamma_M. \tag{4}$$

It is well known that, if the coefficient matrices in (3) are "dominant-diagonal" matrices,* $\tilde{\mathbf{Y}}$ can be realized as a transformerless balanced RC network.[20] Our objective is to determine the submatrices in (1) so that $\tilde{\mathbf{Y}}$ satisfies the dominant-diagonal condition. To simplify the discussion it is assumed that $\tilde{\mathbf{Y}}$ is to be regular at infinity.

2.1 *The Synthesis Technique*

Consider the class of matrices $\mathbf{Y}_{11}$, $\mathbf{Y}_{12}$, $\mathbf{Y}_{22}$, and diag $(y_1,y_2,\cdots,y_N)$ satisfying (2) such that $\mathbf{Y}_{12}$ and $[\mathbf{Y} - \mathbf{Y}_{11}]$ possess inverses. As a first step in obtaining insight into the realization problem we rewrite (2) in the following form:

$$-\mathbf{Y}_{12}{}^t[\mathbf{Y} - \mathbf{Y}_{11}]^{-1}\mathbf{Y}_{12} = \mathbf{Y}_{22} - \text{diag } (y_1,y_2,\cdots,y_N). \tag{5}$$

* A dominant-diagonal matrix M has elements m_{jk} which satisfy

$$m_{jj} \geqq \sum_{k \neq j} |\, m_{jk} \,|.$$

It is convenient to employ the following notation:

$$\begin{aligned} \mathbf{Y}_{11} &= \frac{1}{q}[x_{ij}] = \frac{1}{q}\mathbf{X}_{11}, \\ \mathbf{P} &= [qN_{ij} - Dx_{ij}], \\ \mathbf{Y}_{12} &= \frac{1}{q}\mathbf{X}_{12}, \end{aligned} \tag{6}$$

where $\mathbf{X}_{11}$, $\mathbf{P}$, and $\mathbf{X}_{12}$ are $N \times N$ matrices of polynomials and q is a common denominator polynomial.

From (5) and (6),

$$-\frac{D}{q}\mathbf{X}_{12}{}^{t}\mathbf{P}^{-1}\mathbf{X}_{12} = \mathbf{Y}_{22} - \text{diag}\,(y_1,y_2,\cdots,y_N). \tag{7}$$

The left-hand side of (7) can be written before cancellation of common factors as a matrix of real rational functions with common denominator polynomial q det $\mathbf{P}$. Since the poles of the right-hand side of (7) are required to be distinct and on the negative-real axis, $\mathbf{X}_{12}$ must be chosen so that the least common denominator polynomial of the matrix of rational functions has only zeros that are distinct and on the negative-real axis. To satisfy this condition, we employ a matric polynomial factorization technique developed in an earlier paper.[19] Specifically, it is shown in Appendix A that, given $\mathbf{Y}$, a realizable submatrix $\mathbf{Y}_{11} = (1/q)[x_{ij}]$ can be chosen so that:

(a) deg x_{ii} = deg q = $NL_0 (i = 1,2,\cdots,N)$, where* L_0 = max [max deg N_{ij}, deg D];

(b) the off-diagonal numerator polynomials $x_{ij} (i \neq j)$ are any set of real polynomials consistent with $x_{ij} = x_{ji}$ and deg $x_{ij} \leqq$ deg q;

(c) $\mathbf{Y}_{11}$ has only coefficient matrices that satisfy the dominant-diagonal condition with the inequality sign;

(d) the matric polynomial $\mathbf{P}$ [defined in (6)], of degree* deg $q + L_0$ can be written as the product $\mathbf{P}_1\mathbf{P}_2$ of two matric polynomials $\mathbf{P}_1$ and $\mathbf{P}_2$ (with $N \times N$ matrix coefficients) of degrees respectively deg q and L_0;

(e) det $\mathbf{P}$ does not vanish identically; and

(f) the matric polynomial $\mathbf{P}_2$ has the property that det $\mathbf{P}_2$, a polynomial of degree NL_0, has only distinct negative-real zeros that are different from those of q.

In that which follows, we shall assume that conditions (a) through (f) are satisfied.

* The degree requirement is merely a sufficient condition.

In accordance with (d) and (f), note that the left-hand side of (7) can have only distinct negative-real poles if $\mathbf{X}_{12}$ is chosen to be $(1/\alpha)\mathbf{P}_1$, where α is any nonzero real constant, for then (7) reduces to*

$$\frac{-D}{\alpha^2 q \det \mathbf{P}_2} \mathbf{P}_1{}^t \text{ adj } \mathbf{P}_2 = \mathbf{Y}_{22} - \text{diag } (y_1, y_2, \cdots, y_N). \tag{8}$$

In addition, with this choice of $\mathbf{X}_{12}$, $\mathbf{Y}_{12}$ is regular at infinity [see (6) and (d)]. Therefore, by choosing the magnitude of α sufficiently large it is always possible [see (c)] to satisfy the dominant-diagonal condition for the first N rows of $\tilde{\mathbf{Y}}$. Hence let

$$\mathbf{Y}_{12} = \frac{1}{\alpha q} \mathbf{P}_1. \tag{9}$$

It remains to identify $\mathbf{Y}_{22}$ and the y_i such that the dominant-diagonal condition can be satisfied in the last N rows of $\tilde{\mathbf{Y}}$.

The left-hand side of (8) also is regular at infinity since the required condition:

$$\deg D + \deg \mathbf{P}_1 + \deg \text{adj } \mathbf{P}_2 \leqq \deg q + NL_0 \tag{10}$$

reduces to

$$\deg D \leqq \max [\max \deg N_{ij}, \deg D]. \tag{11}$$

From (f),

$$q \det \mathbf{P}_2 = \lambda \prod_{m=1}^{M} (s + \gamma_m), \tag{12}$$

where λ is a nonzero real constant, $M = \deg q + NL_0$, and

$$0 < \gamma_1 < \gamma_2 \cdots < \gamma_M .$$

In view of (10) and (12), (8) can be rewritten as

$$\mathbf{Y}_{22} - \text{diag } (y_1, y_2, \cdots, y_N) = \sum_{m=0}^{M} \mathbf{A}_m \frac{s}{s + \gamma_m}, \tag{13}$$

where

$$0 = \gamma_0 < \gamma_1 < \gamma_2 \cdots < \gamma_M$$

and the $\mathbf{A}_m$ are real symmetric coefficient matrices. It is clear from (13) that each off-diagonal term in $\mathbf{Y}_{22}$ is equal to the corresponding sum on

* In (8), adj $\mathbf{P}_2$ refers to the adjoint of $\mathbf{P}_2$ which is defined by $\mathbf{P}_2$ adj $\mathbf{P}_2 = \mathbf{U} \det \mathbf{P}_2$, where $\mathbf{U}$ is the identity matrix.

the right-hand side and that

$$\text{diag } (\tilde{y}_{N+1,N+1},\tilde{y}_{N+2,N+2},\cdots,\tilde{y}_{2N,2N}) - \text{diag } (y_1,y_2,\cdots,y_N) = \sum_{m=0}^{M} \frac{s}{s+\gamma_m} \text{diag } (a_{11m},a_{22m},\cdots,a_{NNm}). \tag{14}$$

Let

$$\text{diag } (a_{11m},a_{22m},\cdots,a_{NNm}) = \text{diag } (b_{11m},b_{22m},\cdots,b_{NNm}) - \text{diag } (c_{11m},c_{22m},\cdots,c_{NNm}),$$

where

$$b_{iim},c_{iim} \geqq 0 (i = 1,2,\cdots,N).$$

The $\tilde{y}_{N+i,N+i}$ and y_i can be identified as follows:

$$\text{diag } (\tilde{y}_{N+1,N+1},\tilde{y}_{N+2,N+2},\cdots,\tilde{y}_{2N,2N}) = \sum_{m=0}^{M} \frac{s}{s+\gamma_m} \text{diag } (b_{11m} + d_{11m}, b_{22m} + d_{22m}, \cdots, b_{NNm} + d_{NNm}), \tag{15}$$

$$\text{diag } (y_1,y_2,\cdots,y_N) = \sum_{m=0}^{M} \frac{s}{s+\gamma_m} \text{diag } (c_{11m} + d_{11m}, c_{22m} + d_{22m}, \cdots, c_{NNm} + d_{NNm}), \tag{16}$$

where the matrices diag $(d_{11m},d_{22m},\cdots,d_{NNm})$ are chosen to satisfy the dominant-diagonal condition in the last N rows of $\tilde{\mathbf{Y}}$. Hence the matrix $\tilde{\mathbf{Y}}$ is realizable as a transformerless balanced $2N$-port RC network for all symmetric $N \times N$ matrices $\mathbf{Y}$ of real rational functions.

The realization of an arbitrary symmetric open-circuit impedance matrix $\mathbf{Z}$ can be treated as follows. The elements of a matrix $\mathbf{R} =$ diag $(r_1,r_2,\cdots,r_N)$ can be chosen nonnegative and sufficiently large so that $\mathbf{Y}' = [\mathbf{Z} - \mathbf{R}]^{-1}$ exists. Therefore, $\mathbf{Z}$ can be realized by inserting a (nonnegative) resistor r_k in series with each port $k (k = 1,2,\cdots,N)$ of a network characterized by $\mathbf{Y}'$.*

The proof relating to the necessity of N negative-RC admittances follows directly from a more general result developed previously.[19] †

* Similarly, the theorem proved in Ref. 19 remains valid if the words "short-circuit admittance" are replaced with "open-circuit impedance."

† In connection with the analysis in Ref. 19, it is worthwhile to point out that any controlled voltage (current) source can be replaced with an arbitrarily chosen finite impedance (admittance) in series (parallel) with a new controlled voltage (current) source whose output differs from that of the original source by a term which nullifies the effect of the impedance (admittance). With this understanding, it is not necessary to consider further the degenerate cases which can arise if zero and/or infinite impedance paths appear when the controlled sources are set equal to zero.

The techniques presented in this section bear heavily on the problem of realizing unbalanced transformerless N-port active RC networks. These considerations are treated in detail in the following section.

III. UNBALANCED ACTIVE RC REALIZATION OF AN ARBITRARY SHORT-CIRCUIT ADMITTANCE MATRIX

We consider a $(3N + 1)$-terminal RC network to which is connected at terminals $N + k$ $(k = 1,2,\cdots,2N)$ and the common reference node a $(2N + 1)$-terminal active network as shown in Fig. 2. Denote by $\mathbf{E}_a$, $\mathbf{E}_b$, $\mathbf{E}_c$, $\mathbf{I}_a$, $\mathbf{I}_b$, and $\mathbf{I}_c$ the following column matrices of voltages and currents:

$$\mathbf{E}_a = \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_N \end{bmatrix}, \quad \mathbf{E}_b = \begin{bmatrix} E_{N+1} \\ E_{N+2} \\ \vdots \\ E_{2N} \end{bmatrix}, \quad \mathbf{E}_c = \begin{bmatrix} E_{2N+1} \\ E_{2N+2} \\ \vdots \\ E_{3N} \end{bmatrix},$$

$$\mathbf{I}_a = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_N \end{bmatrix}, \quad \mathbf{I}_b = \begin{bmatrix} I_{N+1} \\ I_{N+2} \\ \vdots \\ I_{2N} \end{bmatrix}, \quad \mathbf{I}_c = \begin{bmatrix} I_{2N+1} \\ I_{2N+2} \\ \vdots \\ I_{3N} \end{bmatrix}. \tag{17}$$

It is convenient to partition $\hat{\mathbf{Y}}$, the short-circuit admittance matrix of

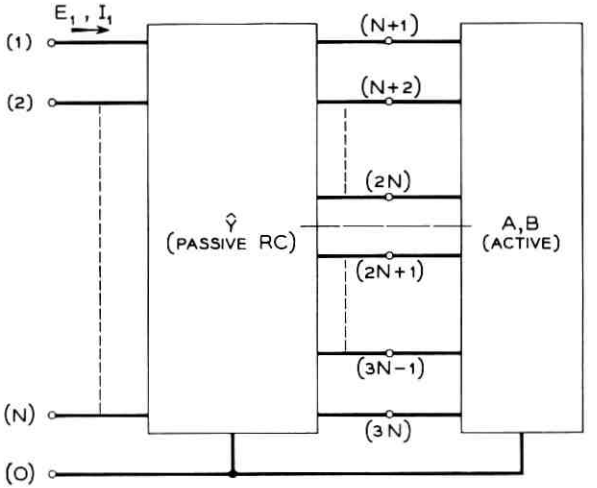


Fig. 2 — Unbalanced realization of an arbitrary $N \times N$ short-circuit admittance matrix.

the $(3N + 1)$-terminal network, as follows:

$$\hat{\mathbf{Y}} = \begin{matrix} \\ N \\ N \\ N \end{matrix} \begin{matrix} \begin{matrix} N & N & N \end{matrix} \\ \begin{bmatrix} \mathbf{Y}_{11} & \mathbf{Y}_{12} & \mathbf{Y}_{13} \\ \mathbf{Y}_{21} & \mathbf{Y}_{22} & \mathbf{Y}_{23} \\ \mathbf{Y}_{31} & \mathbf{Y}_{32} & \mathbf{Y}_{33} \end{bmatrix} \end{matrix} . \tag{18}$$

The active network is assumed to impose the constraints*

$$\begin{aligned} \mathbf{I}_b &= -\mathbf{A}\mathbf{I}_c , \\ \mathbf{E}_c &= -\mathbf{B}\mathbf{E}_b , \end{aligned} \tag{19}$$

where $\mathbf{A}$ and $\mathbf{B}$ are $N \times N$ coefficient matrices. It is not difficult to derive the following expression for the short-circuit admittance matrix $\mathbf{Y}$ relating $\mathbf{E}_a$ and $\mathbf{I}_a$, the voltages and currents at the N accessible ports in Fig. 2:

$$\begin{aligned} \mathbf{Y} = \mathbf{Y}_{11} + (\mathbf{Y}_{12} - \mathbf{Y}_{13}\mathbf{B}) \\ \cdot[\mathbf{A}\mathbf{Y}_{33}\mathbf{B} - \mathbf{Y}_{22} - \mathbf{A}\mathbf{Y}_{32} + \mathbf{Y}_{32}{}^t\mathbf{B}]^{-1}(\mathbf{Y}_{12}{}^t + \mathbf{A}\mathbf{Y}_{13}{}^t). \end{aligned} \tag{20}$$

We shall simplify the discussion by assuming that the matrices $\mathbf{A}$ and $\mathbf{B}$ are given by

$$\mathbf{A} = a\mathbf{U}, \qquad \mathbf{B} = b\mathbf{U}, \tag{21}$$

where $\mathbf{U}$ is the identity matrix of order N and a and b are real constants such that $ab > 0$. The synthesis technique does not further restrict the choice of a and b so that the $(2N + 1)$-terminal active network can always be realized with N voltage-inversion or N current-inversion negative-impedance converters by choosing respectively $a = 1$, $b > 0$ or $b = -1$, $a < 0$. Note therefore that the realization can always be accomplished with N controlled sources.

We shall consider explicitly the case in which a, $b > 0$ and indicate the modifications necessary to treat the remaining case.

Our objective is to prove for all prescribed matrices $\mathbf{Y}$ that $\hat{\mathbf{Y}}$ can be realized as a $(3N + 1)$-terminal network of two-terminal impedances with common reference node and no internal nodes. It is well known[20] that the necessary and sufficient conditions for achieving this type of realization are that the coefficient matrices in

$$\hat{\mathbf{Y}} = s\mathbf{K}_\infty + \sum_{m=0}^{M} \mathbf{K}_m \frac{s}{s + \gamma_m} \tag{22}$$

* The matrices $\mathbf{A}$ and $\mathbf{B}$ should not be confused with the diagonal matrices $\mathbf{A}_m$ and $\mathbf{B}_m$ introduced in Section 2.1.

be real symmetric dominant-diagonal matrices with nonpositive off-diagonal terms, and

$$0 = \gamma_0 < \gamma_1 < \gamma_2 \cdots < \gamma_M, \qquad \gamma_m \text{ real.} \tag{23}$$

It is clear that all off-diagonal terms in $\hat{\mathbf{Y}}$ are required to be negative-RC driving-point admittance functions. For simplicity we assume that $\hat{\mathbf{Y}}$ is not to have a pole at infinity ($\mathbf{K}_\infty = 0$).

3.1 *The Realization Technique*

Our notation is identical to that used in the preceding Section 2.1:

$$\begin{aligned} \mathbf{Y}_{11} &= \frac{1}{q}[x_{ij}] = \frac{1}{q}\mathbf{X}_{11}, & \mathbf{P} &= [qN_{ij} - Dx_{ij}], \\ \mathbf{Y}_{12} &= \frac{1}{q}\mathbf{X}_{12}, & \mathbf{Y}_{13} &= \frac{1}{q}\mathbf{X}_{13}. \end{aligned} \tag{24}$$

By paralleling the development in Section 2.1* and using (20), (21), and (24), we obtain†

$$\frac{D}{q}(\mathbf{X}_{12}{}^t + a\mathbf{X}_{13}{}^t)\mathbf{P}^{-1}(\mathbf{X}_{12} - b\mathbf{X}_{13}) = ab\mathbf{Y}_{33} - \mathbf{Y}_{22} - a\mathbf{Y}_{32} + b\mathbf{Y}_{32}{}^t. \tag{25}$$

We again assume that $\mathbf{Y}_{11}$ is chosen so that (a) through (f) (Section 2.1) are satisfied. It is assumed in addition that the off-diagonal terms in $\mathbf{Y}_{11}$ are chosen to be negative-RC driving-point admittance functions [see (b)].

Next let

$$\begin{aligned} \mathbf{X}_{12} - b\mathbf{X}_{13} &= \frac{1}{\beta_1}\mathbf{P}_1, \\ \mathbf{X}_{12}{}^t + a\mathbf{X}_{13}{}^t &= \frac{1}{\beta_2}\mathbf{P}_3, \end{aligned} \tag{26}$$

where β_1 and β_2 are nonzero real parameters to be chosen in accordance with the discussion below and $\mathbf{P}_3$ is a nonsingular matrix of N^2 polynomials chosen so that each entry in $(1/q)\mathbf{P}_3$ is a negative-RC driving-point admittance function that is nonzero at the origin and finite at infinity. It is clear that deg $\mathbf{P}_3$ = deg q.

We consider the matrices $\mathbf{Y}_{12}$ and $\mathbf{Y}_{13}$. From (24) and (26) we find

* It is assumed that $[\mathbf{Y}_{12} - b\mathbf{Y}_{13}]$, $[\mathbf{Y} - \mathbf{Y}_{11}]$, and $[\mathbf{Y}_{12}{}^t + a\mathbf{Y}_{13}{}^t]$ possess inverses.

† The writer is indebted to J. M. Sipress for suggesting a study of (25) by exploiting the essential similarities between it and (7).

$$\mathbf{Y}_{12} = \frac{1}{q}\frac{b}{\beta_2(a+b)}\left[\mathbf{P}_3{}^t + \frac{a\beta_2}{b\beta_1}\mathbf{P}_1\right],$$
$$\mathbf{Y}_{13} = \frac{1}{q}\frac{1}{\beta_2(a+b)}\left[\mathbf{P}_3{}^t - \frac{\beta_2}{\beta_1}\mathbf{P}_1\right]. \tag{27}$$

Suppose that* $a,b,\beta_2 > 0$. Note that, since $\deg \mathbf{P}_1 = \deg \mathbf{P}_3 = \deg q$, it is possible to choose $|\beta_2/\beta_1|$ sufficiently small such that each element in $\mathbf{Y}_{12}$ and $\mathbf{Y}_{13}$ is a negative-RC driving-point admittance function. It is clear that this ratio can be held invariant while β_2 is chosen sufficiently large to satisfy the dominant-diagonal condition in the first N rows of $\hat{\mathbf{Y}}$.

At this point the synthesis problem reduces to the determination of the submatrices $\mathbf{Y}_{23}$, $\mathbf{Y}_{33}$, and $\mathbf{Y}_{22}$.

3.2 *Determination of* $\mathbf{Y}_{23}$, $\mathbf{Y}_{33}$, *and* $\mathbf{Y}_{22}$

Substituting (26) into (25) gives

$$\frac{1}{\beta_1\beta_2}\frac{D}{q}\mathbf{P}_3\mathbf{P}_2^{-1} = ab\mathbf{Y}_{33} - \mathbf{Y}_{22} - a\mathbf{Y}_{32} + b\mathbf{Y}_{32}{}^t, \tag{28}$$

where

$$q \det \mathbf{P}_2 = \lambda \prod_{m=1}^{M} (s + \gamma_m).$$

It can easily be shown that the left-hand side of (28) is regular at infinity. Hence it can be written as

$$\sum_{m=0}^{M} \mathbf{F}_m \frac{s}{s+\gamma_m} = \sum_{m=0}^{M} \mathbf{G}_m \frac{s}{s+\gamma_m} - \sum_{m=0}^{M} \mathbf{H}_m \frac{s}{s+\gamma_m}, \tag{29}$$

where the $\mathbf{F}_m$ are real (in general nonsymmetric) coefficient matrices,

$$0 = \gamma_0 < \gamma_1 < \gamma_2 \cdots < \gamma_M,$$

and the elements in $\mathbf{G}_m$ and $\mathbf{H}_m$ are nonnegative.

It is clear from (28) that the asymmetry in the $\mathbf{F}_m$ must be absorbed by the terms $-a\mathbf{Y}_{32} + b\mathbf{Y}_{32}{}^t$. By equating the antisymmetric part of (29) to the antisymmetric part of (28), we obtain

$$\frac{b+a}{2}[\mathbf{Y}_{32}{}^t - \mathbf{Y}_{32}] = \frac{1}{2}\sum_m \frac{s}{s+\gamma_m}[\mathbf{G}_m - \mathbf{G}_m{}^t - \mathbf{H}_m + \mathbf{H}_m{}^t]. \tag{30}$$

* The case in which $a, b < 0$ can be treated by an entirely analogous method, which involves interchanging the properties assigned to the matric polynomials $\mathbf{P}_1$ and $\mathbf{P}_2$ in (25). The required factorization can be obtained by factoring $\mathbf{P}^t$ and taking the transpose of the resulting product.

Equation (30) is satisfied* with

$$\mathbf{Y}_{32} = -\frac{1}{a+b}\sum_m \frac{s}{s+\gamma_m}[\mathbf{G}_m + \mathbf{H}_m{}^t]. \tag{31}$$

The equation corresponding to (30) for the symmetric parts, with $\mathbf{Y}_{32}$ given by (31), is

$$ab\mathbf{Y}_{33} - \mathbf{Y}_{22} = \frac{b}{a+b}\sum_m \frac{s}{s+\gamma_m}[\mathbf{G}_m + \mathbf{G}_m{}^t] - \frac{a}{a+b}\sum_m \frac{s}{s+\gamma_m}[\mathbf{H}_m + \mathbf{H}_m{}^t]. \tag{32}$$

The identification of $\mathbf{Y}_{33}$ and $\mathbf{Y}_{22}$ can be made as follows:

$$\begin{aligned}
\mathbf{Y}_{33} &\overset{od}{=} -\frac{1}{b(a+b)}\sum_m \frac{s}{s+\gamma_m}[\mathbf{H}_m + \mathbf{H}_m{}^t],\\
\mathbf{Y}_{22} &\overset{od}{=} -\frac{b}{(a+b)}\sum_m \frac{s}{s+\gamma_m}[\mathbf{G}_m + \mathbf{G}_m{}^t],\\
\mathbf{Y}_{33} &\overset{d}{=} \frac{1}{a(a+b)}\sum_m \frac{s}{s+\gamma_m}[\mathbf{G}_m + \mathbf{G}_m{}^t] + \sum_m \frac{s}{s+\gamma_m}\mathbf{J}_m,\\
\mathbf{Y}_{22} &\overset{d}{=} \frac{a}{a+b}\sum_m \frac{s}{s+\gamma_m}[\mathbf{H}_m + \mathbf{H}_m{}^t] + ab\sum_m \frac{s}{s+\gamma_m}\mathbf{J}_m,
\end{aligned} \tag{33}$$

where "od" or "d" over an equal sign signifies that equality holds respectively only for the off-diagonal and on-diagonal elements. The diagonal matrices $\mathbf{J}_m$ in (33) are chosen to satisfy the dominant-diagonal condition for the last $2N$ rows of $\hat{\mathbf{Y}}$.

For the special case when all the $\mathbf{F}_m$ are symmetric matrices the structure can be simplified by setting $\mathbf{Y}_{32} = 0$. This leads to the identification:

$$\begin{aligned}
\mathbf{Y}_{33} &\overset{od}{=} -\frac{1}{ab}\sum_m \mathbf{H}_m \frac{s}{s+\gamma_m},\\
\mathbf{Y}_{22} &\overset{od}{=} -\sum_m \mathbf{G}_m \frac{s}{s+\gamma_m},\\
\mathbf{Y}_{33} &\overset{d}{=} \frac{1}{ab}\sum_m \mathbf{G}_m \frac{s}{s+\gamma_m} + \sum_m \mathbf{J}_m \frac{s}{s+\gamma_m},\\
\mathbf{Y}_{22} &\overset{d}{=} \sum_m \mathbf{H}_m \frac{s}{s+\gamma_m} + ab\sum_m \mathbf{J}_m \frac{s}{s+\gamma_m}.
\end{aligned} \tag{34}$$

* There are, of course, other solutions of (30).

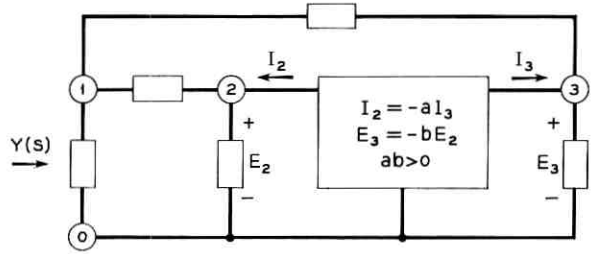


Fig. 3 — Realization of a general driving-point function.

Note that the elements in $\mathbf{Y}_{32}$ and the off-diagonal elements in $\mathbf{Y}_{22}$ and $\mathbf{Y}_{33}$ given by (33) and (34) are, as required, negative-*RC* driving-point admittance functions.

Hence, an arbitrary $N \times N$ matrix of real rational functions can be realized as the short-circuit admittance matrix of the structure shown in Fig. 2 in which the $(3N + 1)$-terminal network requires no internal nodes and contains only resistors and capacitors.* A numerical example is considered in Appendix B. The freedom implicit in the synthesis procedure can be exploited further to yield certain simplifications and other types of structures. Some of these possibilities may already have occurred to the sufficiently interested reader.

IV. DISCUSSION

In Section II it is shown that N is the sufficient and, in general, minimum number of negative-*RC* driving-point immittances that must be embedded in an N-port network of resistors and capacitors to realize as its immittance matrix an arbitrary symmetric $N \times N$ matrix of real rational functions in the complex-frequency variable.

Since any negative-*RC* driving-point admittance function which is regular at infinity can be written as the sum of a negative constant and an *RL* driving-point admittance function, it follows [recall from (16) that the y_i need not have a pole at infinity] that

Theorem:† *An arbitrary symmetric $N \times N$ matrix of real rational functions can be realized as the immittance matrix of an N-port transformerless RLC network containing N negative resistors. A canonical form is a 2N-port network of resistors and capacitors terminated at each of N ports with an RL driving-point impedance in parallel with a negative resistor.*

* The complete structure for the special case $N = 1$ (and $\mathbf{Y}_{32} = 0$) is shown in Fig. 3.

† Carlin has established[21] some interesting related results for networks containing resistors, capacitors, inductors, gyrators, ideal transformers, and negative resistors.

The unbalanced realization of an N-port active RC network described in Section III leads to a particularly simple structural form for the required passive subnetwork. Possibilities of determining other structures are implicit in the method. An intriguing class of unsolved problems relate to the determination of structures which optimize some measure of performance such as the sensitivity function.

V. ACKNOWLEDGMENT

The writer is grateful to S. Darlington for his constructive criticism and advice.

APPENDIX A

Selection of $\mathbf{Y}_{11}$ *and Decomposition of* $\mathbf{P}$

The submatrix $\mathbf{Y}_{11}$ can be made to have dominant-diagonal coefficient matrices by choosing any realizable $N \times N$ RC admittance matrix, with elements of suitable degree as determined subsequently, and multiplying each diagonal entry by a sufficiently large positive real constant ρ. Denote the matrix determined in this way by

$$\mathbf{Y}_{11} = \frac{1}{q}\begin{bmatrix} \rho x_{11}' & x_{12} & \cdots & x_{1N} \\ \vdots & & \rho x_{22}' & \vdots \\ x_{N1} & & \cdots & \rho x_{NN}' \end{bmatrix}. \tag{35}$$

The polynomial $\det \mathbf{P}$ can be written as

$$\det \mathbf{P} = \det [qN_{ij} - Dx_{ij}] = (-\rho)^N \left\{ D^N \prod_{i=1}^{N} x_{ii}' + \frac{R(s)}{\rho^N} \right\}, \tag{36}$$

where $R(s)/\rho^N$ is a polynomial with all coefficients that approach zero as ρ approaches infinity. We shall assume that $\deg x_{ii} = \deg q$ $(i = 1,2,\cdots,N)$, and that the x_{ii} are nonzero at the origin. Note that, as ρ approaches infinity, $N\deg q$ zeros of $\det \mathbf{P}$ approach the zeros of

$$\prod_{i=1}^{N} x_{ii}'.$$

The zeros of this product can be chosen to be distinct and different from those of D. Hence, for a sufficiently large value of ρ, condition (c) of Section 2.1 is satisfied, and $\det \mathbf{P}$ has at least N $\deg q$ distinct negative-real zeros that are different from those of q.

We next consider a sufficient condition for the removal of a linear factor of $\mathbf{P}$.

A.1 *Factorization of the Matric Polynomial* **P***

Let L be the degree of the highest degree polynomial in $\mathbf{P}$ and suppose that the zeros of

$$\det \mathbf{P} = \sum_{j=0}^{J} a_j s^j$$

include K distinct zeros.

Consider the result of determining a nonsingular matrix $\mathbf{Q}$ with constant elements such that every element in the ith column of $\mathbf{PQ}$ has a zero at $s = s_i\,(i = 1,2,\cdots,N)$, where s_i is a zero of $\det \mathbf{P}$. If indeed this can be done, $\mathbf{P}$ can be written as

$$\mathbf{P} = (\mathbf{PQ})\mathbf{Q}^{-1} = \mathbf{P}'(\mathbf{DQ}^{-1}), \tag{37}$$

where $\mathbf{D}$ is the diagonal matrix diag $[s - s_1, s - s_2, \cdots, s - s_N]$, and the degree of the highest degree polynomial in $\mathbf{P}'$ is $L - 1$. This is equivalent to removing a linear factor of the matric polynomial $\mathbf{P}$:

$$\begin{aligned}\mathbf{P} = \sum_{j=1}^{L} s^j \mathbf{A}_j &= \left[\sum_{j=1}^{L-1} s^j \mathbf{A}_j'\right] \mathbf{DQ}^{-1} \\ &= \left[\sum_{j=1}^{L-1} s^j \mathbf{A}_j' \mathbf{Q}^{-1}\right] \mathbf{QDQ}^{-1} \\ &= \left[\sum_{j=1}^{L-1} s^j \mathbf{A}_j''\right] (s\mathbf{U} - \mathbf{B}),\end{aligned} \tag{38}$$

where $\mathbf{U}$ is the identity matrix of order N and

$$\mathbf{B} = \mathbf{Q}\ \text{diag}\ [s_1, s_2, \cdots, s_N]\,\mathbf{Q}^{-1}.$$

We first develop a sufficient condition for the existence of a nonsingular matrix of constants $\mathbf{Q}_k$ such that every element in the kth column of $\mathbf{PQ}_k$ has a zero at $s = s_k$. It is then shown that $\mathbf{Q}$ can be constructed as the product of N matrices of this type.

At any zero of $\det \mathbf{P}$, say at $s = s_l$, the column rank of $\mathbf{P}$ is necessarily less than N, and hence there exists a relationship of the form

$$0 = \sum_{j=1}^{N} q_{jl} \mathbf{P}_j(s_l), \tag{39}$$

where $\mathbf{P}_j(s_l)$ is the jth column vector of $\mathbf{P}$ evaluated at $s = s_l$ and the

* The discussion is more general than is required for the purposes of this paper.

constants q_{jl} are not all zero. If, in addition, for some value k of the index l there exists a relationship of the form (39) with $q_{kk} \neq 0$, a matrix $\mathbf{Q}_k$ having the desired properties exists and in fact is given by

$$\mathbf{Q}_k = \begin{bmatrix} 1 & & q_{1k} & & \\ & \ddots & q_{2k} & & \\ & & 1 & & \\ & & \vdots & \ddots & \\ & & q_{kk} & & \\ & & & \ddots & \\ & & \vdots & & 1 \\ & & & & \ddots \\ & & q_{Nk} & & & 1 \end{bmatrix}.$$

Consider

$$\det \mathbf{P} = \sum_{i=1}^{N} p_{ik}\Delta_{ik},$$

where the Δ_{ik} are the appropriate cofactors constructed from columns $1,2,\cdots,k-1, k+1,\cdots,N$ of $\det \mathbf{P}$. Denote by $C_k(s)$ the polynomial which is the greatest common factor of all the Δ_{ik}. It follows that

$$\det \mathbf{P} = C_k(s) \sum_{i=1}^{N} p_{ik}\Delta_{ik}', \tag{40}$$

in which there are no factors common to all the Δ_{ik}'. It is evident that all $(N-1)$-rowed minors of $\det \mathbf{P}$ constructed from columns $1,2,\cdots,k-1, k+1,\cdots,N$ cannot vanish at $s = s_k$, if s_k is a zero of

$$\sum_{i=1}^{N} p_{ik}\Delta_{ik}'$$

that is different from those of $C_k(s)$. In such cases the following set of equations yields only the trivial solution for the q_{jk}:

$$0 = \sum_{j\neq k}^{N} q_{jk}\mathbf{P}_j(s_k) \tag{41}$$

and hence

$$0 = \sum_{j=1}^{N} q_{jk}\mathbf{P}_j(s_k), \tag{42}$$

where $q_{kk} \neq 0$.

In other words, if det $\mathbf{P}$ has at least one zero which is different* from those of $C_k(s)$, a nonsingular matrix of constants, $\mathbf{Q}_k$, can be determined such that each element in the kth column of $\mathbf{PQ}_k$ has a zero at $s = s_k$.

Since the number of zeros of the polynomial $C_k(s)$ cannot exceed $(N-1)L$, it is obviously sufficient that K, the number of distinct zeros of det $\mathbf{P}$, exceed $(N-1)L$. Note that the degree of the highest degree polynomial in $\mathbf{P}$ and the zeros of det $\mathbf{P}$ are identical to the corresponding quantities in $\mathbf{PQ}_k$. Note also that the elements in all columns of $\mathbf{PQ}_k$ except the kth remain unchanged. Hence, if $K > (N-1)L$, the matrix $\mathbf{Q}$ can be constructed as a product of N matrices $\mathbf{Q}_k$ chosen so that every element in the ith column of

$$\mathbf{P} \prod_{k=1}^{m} \mathbf{Q}_k, \qquad (i = 1,2,\cdots,m)$$

has a zero at $s = s_i$.

To summarize, if $(N-1)L < K$, N zeros of Det. $\mathbf{P}$ can be removed as a linear factor of the matric polynomial $\mathbf{P}$. The remaining polynomial is of degree $L - 1$.†

The removal of a linear factor can be ensured under a weaker condition if $\mathbf{A}_L$, the leading coefficient of the matric polynomial, is singular. This matter is discussed in the following paragraph.

Let $\mathbf{R}$ be a nonsingular matrix of real constants chosen so that $\mathbf{A}_L\mathbf{R}$ has $N - r$ vanishing columns, where r is the rank of $\mathbf{A}_L$. Assume for the purposes of discussion that the last $N - r$ columns of $\mathbf{A}_L\mathbf{R}$ vanish. It follows that the elements in the last $N - r$ columns of $\mathbf{PR}$ have degrees not exceeding $L - 1$. In accordance with the discussion presented above, it is possible to determine a nonsingular matrix of constants $\mathbf{Q}_k$ such that each element in column k of $\mathbf{PRQ}_k$ has a zero at $s = s_k$ if det $\mathbf{P}$ has at least one zero that is different from those of $C_k'(s)$ [the greatest common factor of the $(N-1)$-rowed minors of $\mathbf{PR}$ analogous to those of $\mathbf{P}$ above]. Note that if $1 \leqq k \leqq r$ the degree of $C_k'(s)$ cannot exceed

* A suitable $\mathbf{Q}_k$ corresponding to a multiple root of det $\mathbf{P}$ at $s = s_k$ can be determined if the nullity of $\mathbf{P}$ at $s = s_k$ exceeds the number of linearly independent nontrivial solutions for the q_{jk} in (41).

† This implies that the matric polynomial $\mathbf{P}$ can be written as

$$\mathbf{P} = \mathbf{C} \prod_{i=1}^{L} (s\mathbf{U} - \mathbf{B}_i),$$

when det $\mathbf{P}$ has NL distinct zeros. When these zeros are all real the coefficient matrices $\mathbf{C}$ and $\mathbf{B}_i$ are also real.

$(N-1)L-(N-r)$. Therefore, if $K > (N-1)L-(N=r)$, a nonsingular matrix

$$\mathbf{Q}' = \prod_{k=1}^{r} \mathbf{Q}_k$$

can certainly be determined such that each element in the kth column of $\mathbf{PRQ}'$ has a zero at $s = s_k$ $(k = 1,2,\cdots r)$, while each element in the last $N-r$ columns of $\mathbf{PRQ}'$ is of degree not exceeding $L-1$. Hence, P can be written as follows:

$$\begin{aligned}\mathbf{P} &= (\mathbf{PRQ}')(\mathbf{RQ}')^{-1} \qquad (43)\\ &= \mathbf{P}'' \operatorname{diag}[s-s_1, s-s_2, \cdots, s-s_r, \underbrace{1,1,\cdots,1}_{N-r}](\mathbf{RQ}')^{-1},\end{aligned}$$

$$\mathbf{P} = \mathbf{P}''[s\mathbf{F}+\mathbf{G}], \qquad (44)$$

where $\mathbf{P}''$ is of degree $L-1$ and $\mathbf{F}$ and $\mathbf{G}$ are constant $N \times N$ matrices. In particular, $\mathbf{F}$ is of rank r.

It should be clear that the factorization (44) is not dependent upon which $N-r$ columns of $\mathbf{A}_L\mathbf{R}$ vanish.

For our purposes it is sufficient to consider only the negative-real zeros of det $\mathbf{P}$. A moment's reflection will show that if $N\deg q$, the minimum number of distinct negative-real zeros of det $\mathbf{P}$, satisfies $N\deg q > (N-1)L$, N distinct negative-real zeros of det $\mathbf{P}$ can be removed as a linear factor of $\mathbf{P}$. The remaining polynomial is of degree $L-1$ and the matrix of constants $\mathbf{B}$ [in (38)] is real. It follows that Nk distinct negative-real zeros of det $\mathbf{P}$ can be removed as k linear factors if

$$(N-1)[L-(k-1)] < N\deg q - N(k-1). \qquad (45)$$

The degree of $\mathbf{P}$ is $L = \deg q + L_0$, where $L_0 = \max[\max \deg N_{ij}, \deg D]$. To ensure that $k = L_0$ linear factors of $\mathbf{P}$ can be removed, we have, from (45),

$$NL_0 - 1 < \deg q. \qquad (46)$$

APPENDIX B

Synthesis of a Two-Port Network — A Numerical Example

To illustrate the main points in the synthesis technique presented in Section 3.1, we consider in detail the synthesis of a two-port network. Since the factorization of $\mathbf{P}$ is described elsewhere,[19] we select an example

for which it is possible to choose $\mathbf{Y}_{11}$ so that the required factoring is trivial. It is assumed that $a = b = 1$ [see (19) and (21)]:

Let the prescribed 2×2 matrix be

$$\mathbf{Y} = \frac{1}{D}[N_{ij}] = \frac{1}{s+3}\begin{bmatrix} 1 & s+3 \\ s-3 & 2 \end{bmatrix}. \tag{47}$$

The following matrix $\mathbf{Y}_{11}$ obviously satisfies the dominance condition with inequality:

$$\mathbf{Y}_{11} = \frac{1}{q}[x_{ij}] = \frac{1}{s+3}\begin{bmatrix} \rho(s+1) & 0 \\ 0 & \rho(s+2) \end{bmatrix}, \qquad \rho > 0. \tag{48}$$

Since $q = D$, the factorization of $\mathbf{P}$ is trivial. Specifically, we have

$$\mathbf{P} = (s+3)\begin{bmatrix} (1-\rho-\rho s) & s+3 \\ s-3 & (2-2\rho-\rho s) \end{bmatrix} = \mathbf{P}_1\mathbf{P}_2, \tag{49}$$

where

$$\mathbf{P}_1 = (s+3)\mathbf{U}, \qquad \mathbf{P}_2 = \begin{bmatrix} (1-\rho-\rho s) & s+3 \\ s-3 & (2-2\rho-\rho s) \end{bmatrix}, \tag{50}$$

and $\mathbf{U}$ is the identity matrix of order two. It is clear from (50) that ρ can be chosen so that $\det \mathbf{P}_2$ has two distinct negative-real zeros. We choose $\rho = 10$, which yields

$$\begin{aligned} \mathbf{P}_2 &= \begin{bmatrix} -(10s+9) & s+3 \\ s-3 & -(10s+18) \end{bmatrix}, \\ \det \mathbf{P}_2 &= 99s^2 + 270s + 171 \\ &= 99(s+1.0000)(s+1.7273). \end{aligned} \tag{51}$$

Hence $\hat{\mathbf{Y}}$ will be of the form

$$\sum_{m=0}^{3} \mathbf{K}_m \frac{s}{s+\gamma_m}, \tag{52}$$

where $\gamma_0 = 0$, $\gamma_1 = 1.0000$, $\gamma_2 = 1.7273$, and $\gamma_3 = 3.0000$.

Since $\mathbf{P}_1$ is a diagonal matrix, $\mathbf{P}_3$ can be chosen to be a diagonal matrix. Let

$$\mathbf{P}_3 = -(s+2)\mathbf{U}. \tag{53}$$

Note that $(1/q)\mathbf{P}_3$ is a matrix of negative-RC driving-point admittances. Using (27), we can determine values of β_2/β_1 for which $\mathbf{Y}_{12}$ and $\mathbf{Y}_{13}$ are matrices of negative-RC driving-point admittances. Accordingly, with $\beta_2/\beta_1 = 0.5$ we obtain:

$$\mathbf{Y}_{12} = \frac{1}{\beta_2}\begin{bmatrix} -0.0833 & 0 \\ 0 & -0.0833 \end{bmatrix} + \frac{s}{(s+3)\beta_2}\begin{bmatrix} -0.1666 & 0 \\ 0 & -0.1666 \end{bmatrix},$$
$$\mathbf{Y}_{13} = \frac{1}{\beta_2}\begin{bmatrix} -0.5833 & 0 \\ 0 & -0.5833 \end{bmatrix} + \frac{s}{(s+3)\beta_2}\begin{bmatrix} -0.1666 & 0 \\ 0 & -0.1666 \end{bmatrix}. \tag{54}$$

From (48) with $\rho = 10$,

$$\mathbf{Y}_{11} = \begin{bmatrix} 3.3333 & 0 \\ 0 & 6.6666 \end{bmatrix} + \frac{s}{s+3}\begin{bmatrix} 6.6666 & 0 \\ 0 & 3.3333 \end{bmatrix}. \tag{55}$$

The choice $\beta_2 = 0.2$ satisfies the dominant-diagonal condition for the first two rows of $\mathbf{K}_0$ and $\mathbf{K}_3$. This condition is satisfied with the equality sign in the first row of $\mathbf{K}_0$, and for this reason reduces by one the number of resistors necessary to realize $\mathbf{K}_0$.

Using $(99\beta_1\beta_2)^{-1} = 0.1263$, we obtain from (28), (29), (51), and (53),

$$\frac{0.1263(s+2)}{(s+1.0000)(s+1.7273)}\begin{bmatrix} 10s+18 & s+3 \\ s-3 & 10s+9 \end{bmatrix} = \sum_{m=0}^{2} \mathbf{F}_m \frac{s}{s+\gamma_m}. \tag{56}$$

Equation (56) can be expressed as

$$\sum_{m=0}^{2} \mathbf{F}_m \frac{s}{s+\gamma_m} = \begin{bmatrix} 2.6316 & 0.4386 \\ -0.4386 & 1.3159 \end{bmatrix} + \frac{s}{s+1.0000}\begin{bmatrix} -1.3889 & -0.3472 \\ 0.6944 & 0.17361 \end{bmatrix} + \frac{s}{s+1.7273}\begin{bmatrix} 0.0199 & 0.0348 \\ -0.1296 & -0.2267 \end{bmatrix}. \tag{57}$$

The coefficient matrices $\mathbf{K}_m$ can readily be constructed with the aid of (31), (33), (54), (55), and (57). Consider for example $\mathbf{K}_0$. From (57),

$$\mathbf{G}_0 = \begin{bmatrix} 2.6315 & 0.4386 \\ 0 & 1.3159 \end{bmatrix}, \qquad \mathbf{H}_0 = \begin{bmatrix} 0 & 0 \\ 0.4386 & 0 \end{bmatrix}. \tag{58}$$

Using (31), (33), and (58),

$$\mathbf{Y}_{32_0} = \begin{bmatrix} -1.3157 & -0.4386 \\ 0 & -0.6579 \end{bmatrix},$$

$$\mathbf{Y}_{33_0} = \begin{bmatrix} 2.6315 + j_{10} & -0.2193 \\ -0.2193 & 1.3159 + j_{20} \end{bmatrix}, \tag{59}$$

$$\mathbf{Y}_{22_0} = \begin{bmatrix} j_{10} & -0.2193 \\ -0.2193 & j_{20} \end{bmatrix},$$

where j_{10} and j_{20} are the diagonal elements in $\mathbf{J}_0$ [see (33)].

From (54) with $\beta_2 = 0.2$, (55) and (59)

$$K_0 = \left|\begin{array}{cc:cc:cc} 3.3333 & 0 & -0.4166 & 0 & -2.9166 & 0 \\ 0 & 6.6666 & 0 & -0.4166 & 0 & -2.9166 \\ \hdashline -0.4166 & 0 & j_{10} & -0.2193 & -1.3157 & 0 \\ 0 & -0.4166 & -0.2193 & j_{20} & -0.4386 & -0.6579 \\ \hdashline -2.9166 & 0 & -1.3157 & -0.4386 & 2.6315 + j_{10} & -0.2139 \\ 0 & -2.9166 & 0 & -0.6579 & -0.2139 & 1.3159 + j_{20} \end{array}\right|. \tag{60}$$

It is easy to verify that the choice $j_{01} = 2.2534$, $j_{02} = 2.4725$ satisfies the dominant-diagonal condition for the last four rows of $\mathbf{K}_0$ and in particular satisfies the condition with equality in rows five and six.

The remaining coefficient matrices $\mathbf{K}_1$, $\mathbf{K}_2$, and $\mathbf{K}_3$ can be constructed in a similar manner. The realization of the matrix $\hat{\mathbf{Y}}$ is straightforward.

B.1 *An Alternative Synthesis*

Some reflection will show that a large number of elements are required to realize $\hat{\mathbf{Y}}$. This number can be reduced by choosing the elements in $\mathbf{P}_3$ differently. For this reason it is worth while to consider the following alternative synthesis technique.

If $\mathbf{P}$ could be written as $\mathbf{P}_1'\mathbf{P}_2'$, where $\mathbf{P}_2'$ has the properties previously associated with $\mathbf{P}_3$ (here denoted by $\mathbf{P}_3'$), the sum in (29), with $\mathbf{P}_2' = \mathbf{P}_3'$, would contain simply one term [see (28) and recall that $D = q$] while the properties assigned to $\mathbf{Y}_{12}$ and $\mathbf{Y}_{13}$ are permitted to remain invariant.

Consider the matrix $\mathbf{P}_2$ given in (51) and repeated below for convenience:

$$\mathbf{P}_2 = \begin{bmatrix} -(10s + 9) & s + 3 \\ s - 3 & -(10s + 18) \end{bmatrix}. \tag{61}$$

By adding the first row of $\mathbf{P}_2$ to the second, and then adding the new second row to the first, we obtain

$$\mathbf{P}_2' = \begin{bmatrix} -(19s + 21) & -(8s + 12) \\ -(9s + 12) & -(9s + 15) \end{bmatrix}. \tag{62}$$

Note that each element in $(1/q)\mathbf{P}_2'$ is a negative-RC driving-point admittance function. Since $\mathbf{P}_2'$ can be obtained from $\mathbf{P}_2$ by successive elementary operations on rows, the relation between $\mathbf{P}_2$ and $\mathbf{P}_2'$ can be expressed by

$$\mathbf{P}_2' = \mathbf{T}\mathbf{P}_2, \tag{63}$$

where $\mathbf{T}$ is a 2×2 nonsingular matrix of real constants. Specifically,

$$\mathbf{T} = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}. \tag{64}$$

The matrix $\mathbf{P}$ can be written as

$$\mathbf{P} = \mathbf{P}_1\mathbf{T}^{-1}\mathbf{T}\mathbf{P}_2 = \mathbf{P}_1'\mathbf{P}_2', \tag{65}$$

where $\mathbf{P}_2' = \mathbf{T}\mathbf{P}_2$ and $\mathbf{P}_1' = \mathbf{P}_1\mathbf{T}^{-1}$. Using (50), (64), and (65),

$$\mathbf{P}_1' = \begin{bmatrix} (s+3) & -(s+3) \\ -(s+3) & 2(s+3) \end{bmatrix}, \tag{66}$$

At this point we let $\mathbf{P}_3' = \mathbf{P}_2'$ and return to the procedure demonstrated earlier.

From (27) with $\beta_2/\beta_1 = 1$, (62), (66), and (55),

$$\begin{aligned}
\mathbf{Y}_{12} &= \frac{1}{\beta_2}\begin{bmatrix} -3.0 & -2.5 \\ -2.5 & -1.5 \end{bmatrix} + \frac{s}{\beta_2(s+3)}\begin{bmatrix} -6.0 & -2.5 \\ -2.0 & -2.0 \end{bmatrix}, \\
\mathbf{Y}_{13} &= \frac{1}{\beta_2}\begin{bmatrix} -4.0 & -1.5 \\ -1.5 & -3.5 \end{bmatrix} + \frac{s}{\beta_2(s+3)}\begin{bmatrix} -6.0 & -2.5 \\ -2.0 & -2.0 \end{bmatrix}, \\
\mathbf{Y}_{11} &= \begin{bmatrix} 3.3333 & 0 \\ 0 & 6.6666 \end{bmatrix} + \frac{s}{s+3}\begin{bmatrix} 6.6666 & 0 \\ 0 & 3.3333 \end{bmatrix}.
\end{aligned} \tag{67}$$

The dominance condition is satisfied in the first and second rows of $\mathbf{K}_0'$ and $\mathbf{K}_1'$ with $\beta_2 = 3.3000$. The condition is satisfied with equality in the first row of $\mathbf{K}_0'$.

The left-hand side of (28) is

$$\frac{1}{\beta_1\beta_2}\frac{D}{q}\mathbf{P}_3'\mathbf{P}_2'^{-1} = 0.0918\,\mathbf{U}, \tag{68}$$

where $\mathbf{U}$ is the identity matrix of order two.

Equations (34) and (68) lead to

$$\begin{aligned}
\mathbf{Y}_{33} &= \begin{bmatrix} 0.0918 + j_{10}' & 0 \\ 0 & 0.0918 + j_{20}' \end{bmatrix} + \frac{s}{s+3}\begin{bmatrix} j_{11}' & 0 \\ 0 & j_{21}' \end{bmatrix}, \\
\mathbf{Y}_{22} &= \begin{bmatrix} j_{10}' & 0 \\ 0 & j_{20}' \end{bmatrix} + \frac{s}{s+3}\begin{bmatrix} j_{11}' & 0 \\ 0 & j_{21}' \end{bmatrix}, \\
\mathbf{Y}_{32} &= 0.
\end{aligned} \tag{69}$$

The coefficient matrix $\mathbf{K}_0'$ is

$$\mathbf{K}_0' = \left[\begin{array}{cc|cc|cc} 3.3333 & 0 & -0.9091 & -0.7576 & -1.2121 & -0.4545 \\ 0 & 6.6666 & -0.7576 & -0.4545 & -0.4545 & -1.0606 \\ \hline -0.9091 & -0.7576 & j_{10}' & 0 & 0 & 0 \\ -0.7576 & -0.4545 & 0 & j_{20}' & 0 & 0 \\ \hline -1.2121 & -0.4545 & 0 & 0 & 0.0918 + j_{10}' & 0 \\ -0.4545 & -1.0606 & 0 & 0 & 0 & 0.0918 + j_{20}' \end{array}\right]$$

The dominance condition is satisfied in the last four rows of $\mathbf{K}_0'$ (satisfied with equality in the third and sixth rows) with $j_{10}' - 1.6667$, $j_{20}' = 1.4233$.

The remaining coefficient matrix $\mathbf{K}_1'$ is given by

$$\mathbf{K}_1' = \left[\begin{array}{cc|cc|cc} 6.6666 & 0 & -1.8182 & -0.7576 & -1.8182 & -0.7576 \\ 0 & 3.3333 & -0.6061 & -0.6061 & -0.6061 & -0.6061 \\ \hline -1.8182 & -0.6061 & j_{11}' & 0 & 0 & 0 \\ -0.7576 & -0.6061 & 0 & j_{21}' & 0 & 0 \\ \hline -1.8182 & -0.6061 & 0 & 0 & j_{11}' & 0 \\ -0.7576 & -0.6061 & 0 & 0 & 0 & j_{21}' \end{array}\right]$$

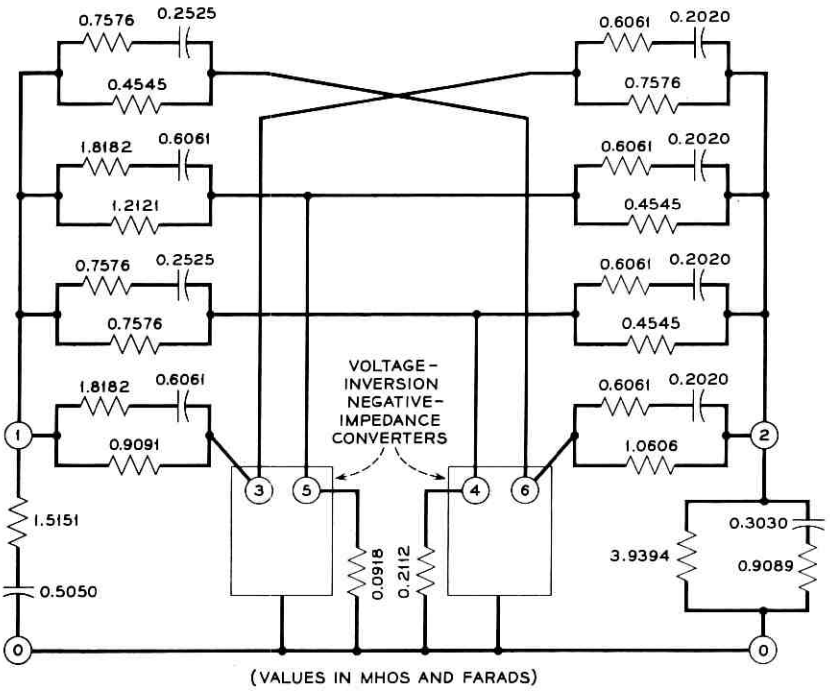


Fig. 4 — Realization of two-port network example.

For this matrix the dominance condition is satisfied with $j_{11}' = 2.4243$, $j_{21}' = 1.3637$.

The final network is shown in Fig. 4.

REFERENCES

1. Dietzold, R. L., Frequency Discriminative Electric Transducer, U. S. Patent No. 2,549,965, April 17, 1951.
2. Bangert, J. T., The Transistor as a Network Element, B.S.T.J., **33,** 1954, p. 329.
3. Linvill, J. G., RC Active Filters, Proc. I.R.E., **42,** 1954, p. 555.
4. Armstrong, D. B., and Reza, F. M., Synthesis of Transfer Functions by Active RC Networks, I.R.E. Trans., **CT-1,** 1954, p. 8.
5. Sallen, R. P., and Key, E. L., A Practical Method of Designing RC Active Filters, I.R.E. Trans., **CT-2,** 1955, p. 74.
6. Horowitz, I. M., RC-Transistor Network Synthesis, Proc. Nat. Elec. Conf., October 1956, p. 818.
7. Horowitz, I. M., Synthesis of Active RC Transfer Functions, Research Report R-507-56, PIB-437, Microwave Research Inst., Polytechnic Inst. of Brooklyn, November 1956.
8. Sipress, J. M., Active RC Partitioning Synthesis of High-Q Band-pass Filters, M.E.E. Thesis, Polytechnic Inst. of Brooklyn, 1957.
9. Bongiorno, J. J., Synthesis of Active RC Single-Tuned Bandpass Filters, I.R.E. Nat. Conv. Rec., March 1958, Pt. 2, p. 30.
10. Sandberg, I. W., Active RC Networks, Research Report R-662-58, PIB-590, Microwave Research Inst., Polytechnic Inst. of Brooklyn, May 1958.
11. DeClaris, N., Synthesis of Active Networks—Driving-Point Functions, I.R.E. Nat. Conv. Rec., March 1959, Pt. 2, p. 23.
12. Myers, B. R., Transistor-RC Network Synthesis, I.R.E. Wescon Conv. Rec., August 1959, Pt. 2, p. 65.
13. Kinariwala, B. K., Synthesis of Active RC Networks, B.S.T.J., **38,** 1959, p. 1269.
14. Horowitz, I. M., Optimization of Negative Impedance Converter Synthesis Techniques, I.R.E. Trans., **CT-6,** 1959, p. 296.
15. Blecher, F. H., Application of Synthesis Techniques to Electronic Circuit Design, I.R.E. Nat. Conv. Rec., March 1960, Pt. 2, p. 210.
16. Kuh, E. S., Transfer Function Synthesis of Active RC Networks, I.R.E. Nat. Conv. Rec., March 1960, Pt. 2, p. 134.
17. Sandberg, I. W., Synthesis of Driving-Point Impedances with Active RC Networks, B.S.T.J., **39,** 1960, p. 947.
18. Sipress, J. M., Synthesis of Active RC Networks, to be published.
19. Sandberg, I. W., Synthesis of N-Port Active RC Networks, B.S.T.J., **40,** 1961, p. 329.
20. Slepian, P., and Weinberg, L., Synthesis Applications of Paramount and Dominant Matrices, Proc. Nat. Elec. Conf., October 1958, p. 611.
21. Carlin, H. J., General N-Port Synthesis with Negative Resistors, Proc. I.R.E., **48,** 1960, p. 1174.

Delays for Last-Come First-Served Service and the Busy Period

By JOHN RIORDAN

(Manuscript received November 10, 1960)

For Poisson input to a single server, it is shown that the stationary delay distribution function for last-come first-served service is equal to the distribution function for the busy period (the interval of time during which the server is continuously busy) only for exponential distribution of service time. A similar argument shows that the identity persists for a group of fully accessible servers, each with exponential distribution of service times — a result which has been a curiosity for some time. Finally, the delay distribution for last-come first-served service and a single server with constant service time is derived.

I. INTRODUCTION

Delays for last-come first-served service were first considered by Vaulot[1] for a system with Poisson input to a group of fully accessible servers, each with exponential service time. For this order of service, an arrival which is not served immediately, following the biblical edict, goes to the head of the waiting line. Its consideration has a natural theoretical interest, because, as the opposite of first-come first-served, it seems to be a bound for the gamut of possible service assignments, or at least of those with simple structure. Indeed Vaulot[1] has used it to find the envelope of delay functions for all service assignments.

Very briefly, Vaulot's formulation is as follows. Let $v_n(t)$ be the probability that a waiting demand for service, which at a given moment has just become $n + 1$ in line, waits at least t [$v_n(t)$ is the complement of a distribution function; $v_0(t)$ is the complement of the conditional delay distribution function]. Then, if a is the arrival rate, and b the service rate for each of the c servers, the set of differential recurrence relations for the $v_n(t)$ is

$$v_n'(t) = bcv_{n-1}(t) - (a + bc)v_n(t) + av_{n+1}(t), \qquad n = 0, 1, \cdots .$$

Vaulot's solution of these equations will be given later. Here it is sufficient to notice that the same relations had already appeared in the formulation of the busy period, for the same traffic system, by Palm.[2] The correspondent to $v_n(t)$ is $f_n(t)$, the probability that a busy period (all servers busy), which at a given moment has n waiting customers, continues as a busy period at least t. Since the boundary conditions also agree, $v_n(t) = f_n(t)$, and in particular $v_0(t) = f_0(t)$; that is, the conditional delay distribution function for last-come first-served service is equal to the distribution function of the busy period, for the given system. This is the curious and puzzling result mentioned in the abstract.

The first result of this paper is the proof that, for Poisson input to a single server, the two distributions are alike only for exponential service. For more than one server, it is plausible that the same thing is true, though no easy line of proof seems open. For many servers, each with exponential service time, proof of identity may be given in a way parallel to the single server case. A second result is the determination of the delay distribution for Poisson input to a single server with constant service time; this result belongs in the book with that of Burke[3] on random service for the same system.

II. LAST-COME FIRST-SERVED DELAY FOR POISSON INPUT TO A SINGLE SERVER

For last-come first-served service, an arrival finding the server busy is delayed until all subsequent arrivals finding the server busy or just completing service have been served. The waiting demands at the arrival epoch have no effect on this delay because the new arrival goes to the head of the line.

Following Takács,[4] the delay distribution may be derived as follows. Take the arrival rate as a, the service rate (reciprocal of the average service time) as b, the service time distribution function as $B(t)$, and the delay distribution function as $G^*(t)$ — the star indicating that $G^*(t)$ differs from $G(t)$, the distribution function for the busy period, in the starting epoch of the corresponding intervals.

Note first that the (stationary) distribution function between an arbitrary time epoch in a service interval and the epoch of next service completion is given by†

$$C(t) = b \int_0^t [1 - B(u)]\, du, \tag{1}$$

† The earliest proof of this result known to me is in Palm[5]; it may have been proved much earlier.

so that

$$\gamma(s) = \int_0^\infty e^{-st}\, dC(t) \qquad (2)$$
$$= (b/s)[1 - \beta(s)],$$

with

$$\beta(s) = \int_0^\infty e^{-st}\, dB(t).$$

Note that $\gamma(s) = \beta(s)$ if and only if $\beta(s) = b(b + s)^{-1}$, or what is the same thing, $B(t) = 1 - e^{-bt}$.

Now consider the interval between the arrival epoch of a delayed demand and the epoch of the next service completion. The probability that this interval has length $(y, y + dy)$ is $dC(y)$. The probability of n arrivals in this interval, when its length is y, is the Poisson term $e^{-ay}(ay)^n/n!$ If $n = 0$, the delay of the given arrival for last-come first-served service is simply $C(t)$; if $n = 1$, the delay consists of the interval to the first service completion plus the busy interval occasioned by this arrival and *all subsequent arrivals during subsequent service periods*, the distribution function for which is $G(t)$. Hence for $n = 1$ the delay is the convolution of $C(t)$ and $G(t)$. In the same way, for $n = 2$ the delay is the convolution of $C(t)$,$G(t)$ and $G(t)$, and so on.

If $G_n(t)$ is the distribution function for the sum of n variables, each with distribution function $G(t)$, and $G_0(t) = 1$, $G_1(t) = G(t)$, then the conditional delay distribution function is given by

$$G^*(t) = \int_0^t \sum_{n=0} e^{-ay} \frac{(ay)^n}{n!} G_n(t - y)\, dC(y). \qquad (3)$$

If

$$\Gamma^*(s) = \int_0^\infty e^{-st}\, dG^*(t)$$

and $\Gamma(s)$ has a similar significance, then the transform of (3) is†

$$\Gamma^*(s) = \gamma[s + a - a\Gamma(s)], \qquad (4)$$

with $\gamma(s)$ defined by (2). Note that $\Gamma(s)$ is determined by Takács' equation

$$\Gamma(s) = \beta[s + a - a\Gamma(s)]. \qquad (5)$$

† A similar and equivalent result appears in Wishart[6]; note that Wishart considers the unconditional delay distribution function, which is less directly comparable (than the conditional) with the distribution function of the busy period.

Hence $\Gamma^*(s) = \Gamma(s)$ when and only when $\gamma(s) = \beta(s)$. As already noted, this implies $B(t) = 1 - e^{-bt}$, the exponential service distribution.

The common distribution function for exponential service time has been given by Takács,[4] and is obtained as follows. First, since $\beta(s) = b(b + s)^{-1}$, the solution of (5) is

$$\Gamma(s) = \frac{a + b + s - \sqrt{(a + b + s)^2 - 4ab}}{2s}.$$

Next, the inverse of this† is

$$G(t) = \int_0^t \frac{dx}{x\sqrt{\rho}} e^{-(a+b)x} I_1(2x\sqrt{ab}), \qquad \rho = a/b, \tag{6}$$

with $I_1(x)$ the Bessel function of the first kind and imaginary argument. An equivalent expression due to Vaulot,[1] which seems better adapted to numerical work, is

$$1 - G(t) = \frac{2}{\pi} \int_0^\pi \frac{dx}{A} e^{-Abt} \sin^2 x, \qquad A = 1 + \rho - 2\sqrt{\rho} \cos x. \tag{6a}$$

III. DELAYS FOR POISSON INPUT TO MANY EXPONENTIAL SERVERS

The modifications of the argument above for c fully accessible servers, each with exponential distribution of service time (exponential servers, for brevity), are relatively minor. First, in the derivation of the busy-period distribution, a service period in the single-server case is replaced by the interval between the epoch at which the last idle server becomes busy and the epoch at which the first of c busy servers becomes idle. The distribution function of this interval is that of the least of c random variables, each with the same exponential distribution, the service time distribution. If this function is $H(t)$ and the service time distribution is $B(t) = 1 - e^{-bt}$, then

$$1 - H(t) = [1 - B(t)]^c = e^{-bct}. \tag{7}$$

Hence the distribution function $G_c(t)$ is obtained from its single server (exponential service time) correspondent $G(t)$ simply by replacing b by bc.

The last-come first-served delay distribution is proved identical simply by the remark that $H(t)$ is also the distribution of the interval between an arbitrary point in the interval for which $H(t)$ is the distribution function and its termination, because $H(t)$ is of exponential form.

† Ref. 7, pair 556.1, p. 59.

Fig. 1 shows a comparison of conditional delay curves for various orders of service at an occupancy $\rho = a/bc$ of 0.9. The orders are order of arrival, random, and inverse order of arrival (last-come first-served). The abscissae are values of an auxiliary time variable $u = bct$. The ordinates are values of $F(u)$, the probability that the delay of a delayed arrival is at least u. The curve marked "envelope" is an upper bound for all possible orders of service; actually it is a plot of $1 - v(u)$ with

$$v(u) = (1 - \rho) \sum \rho^n v_n(u)$$

and $v_n(u)$ as defined in the introduction.

A detailed study of delays for last-come first-served service and the busy period, with and without defections from waiting, is planned for a later paper.

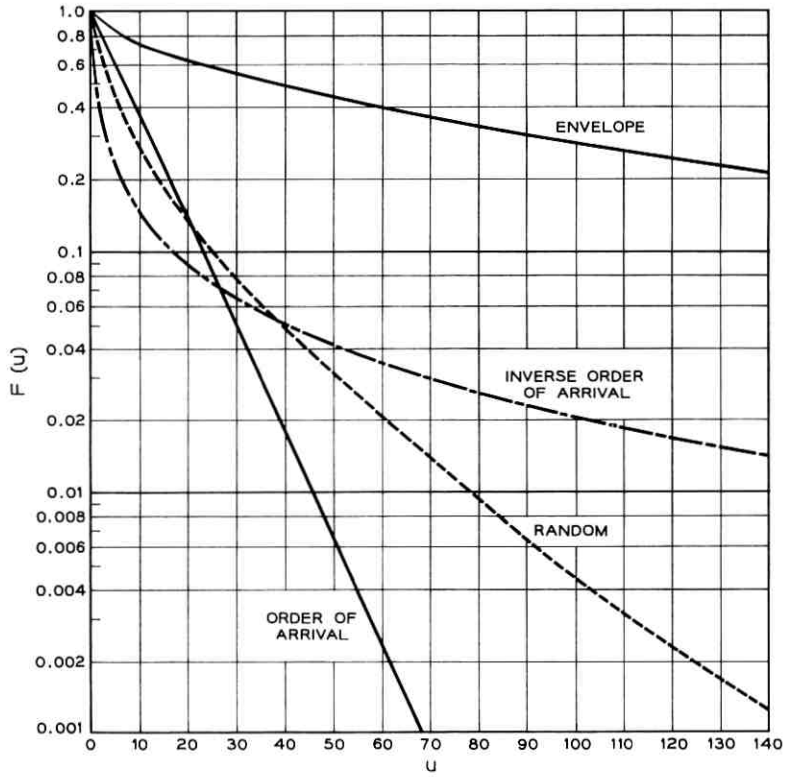


Fig. 1 — Comparison of delay curves for various orders of service occupancy 0.9.

IV. DELAY FOR SINGLE SERVER WITH CONSTANT SERVICE TIME

Although this is a limit case of (4), both delay and busy-period distribution functions are obtained more easily, following Takács for the latter, from the classification of the busy period by number served.

Note first that the generating function for number of arrivals in an interval with distribution function $C(t)$ is given by

$$\begin{aligned} p^*(x) &= \sum p_n^* x^n \\ &= \sum_{n=0} a^n x^n \int_0^\infty e^{-ay}(y^n/n!)\, dC(y) \\ &= \gamma(a - ax). \end{aligned} \tag{8}$$

The corresponding generating function for number of arrivals in a service interval is given by

$$p(x) = \beta(a - ax) \tag{9}$$

and, by (2),

$$p^*(x) = \frac{1 - p(x)}{\rho(1 - x)}, \qquad \rho = a/b. \tag{10}$$

Note that

$$p^*(1) = \frac{p'(1)}{\rho} = \frac{a\beta'(0)}{\rho} = 1,$$

$$\rho p_n^* = 1 - p_0 - p_1 - \cdots - p_n.$$

Now write f_n^* for the probability of n services in a delay period for last-come first-served service, and

$$f^*(x) = \sum_{n=1} f_n^* x^n$$

for its generating function; f_n and $f(x)$ are the corresponding entities for a busy period. Then, following Takács,[4]

$$\begin{aligned} f_1^* &= p_0^*, \\ f_2^* &= p_1^* f_1, \\ f_3^* &= p_1^* f_2 + p_2^* f_1^2, \end{aligned}$$

and

$$f_n^* = p_1^* f_{n-1} + p_2^* \sum f_j f_{n-1-j} + p_3^* \sum f_j f_k f_{n-1-j-k} + \cdots, \tag{11}$$

$$f^*(x) = xp^*[f(x)] = x\gamma[a - af(x)]. \tag{12}$$

Note that the corresponding result for $f(x)$ obtained by Takács is

$$f(x) = x\beta[a - af(x)]. \tag{13}$$

Hence, by (2), (12), and (13),

$$f^*(x) = \frac{x - f(x)}{\rho[1 - f(x)]}, \qquad \rho = a/b. \tag{14}$$

Note that

$$f'(x) = \frac{f(x)}{x} - axf'(x)\beta'[a - af(x)].$$

Note also that (14) may be rewritten as

$$\begin{aligned} \rho f^*(x) &= 1 + (x - 1)[1 - f(x)]^{-1} \\ &= 1 + (x - 1)g(x), \end{aligned} \tag{14a}$$

where

$$\begin{aligned} g(x)[1 - f(x)] &= 1, \\ g(0) &= g_0 = 1. \end{aligned} \tag{15}$$

Hence

$$\rho f_n^* = g_{n-1} - g_n . \tag{16}$$

For constant service time with service rate b,

$$\begin{aligned} B(t) &= 0, \qquad t < b^{-1}, \\ &= 1, \qquad t > b^{-1} \end{aligned}$$

and $\beta(s) = e^{-s/b}$. Hence

$$f(x) = xe^{-\rho+\rho f(x)},$$

whose solution is given by

$$f(x) = \sum_{n=1} \frac{(\rho n)^{n-1}e^{-n\rho}}{n!} x^n,$$

so that

$$f_n = \frac{(\rho n)^{n-1}e^{-n\rho}}{n!}. \tag{17}$$

Then, using (15), the coefficients g_n of the generating function $g(x)$, are given by

$$g_0 = 1,$$
$$g_1 = f_1 = e^{-\rho},$$
$$g_2 = f_2 + f_1^2 = (1 + \rho)e^{-2\rho},$$
$$g_3 = f_3 + 2f_2f_1 + f_1^3 = (1 + 2\rho + 3\rho^2/2!)e^{-3\rho}.$$

These results suggest writing

$$g_n = g_n(\rho) = \sum_{k=0}^{n-1} \frac{e^{-n\rho} g_{nk} \rho^k}{k!}. \tag{18}$$

From (15) it follows that

$$g_n = \sum_{j=1}^{n} f_j g_{n-j}. \tag{18a}$$

From this and (17) and (18), it is found that

$$g_{nk} = \sum_{j=0}^{k} \binom{k}{j} (j + 1)^{j-1} g_{n-j-1,k-j}. \tag{19}$$

From (19) it is found in succession that $g_{n0} = 1$, $g_{n1} = n - 1$, $g_{n2} = n(n - 2)$, all of which are contained in the single formula $g_{nk} = n^{k-1}(n - k)$. If this is correct, substitution in (19) shows that the following must be an identity:

$$n^{k-1}(n - k) = \sum_{j=0}^{k} \binom{k}{j} (j - 1)^{j-1}(n - j - 1)^{k-j-1}(n - k - 1). \tag{20}$$

Equation (20) is in fact one of the forms associated with Abel's generalization of the binomial formula; it is a special case of Equation (1b) of Salie,[8] which is due to Jensen[9] (of whom it is proper to remember here that he was chief engineer of the Copenhagen Telephone Company as well as a mathematician of note).

Thus, finally,†

$$g_n = e^{-n\rho} \sum_{k=0}^{n} (n - k)n^{k-1}\rho^k/k! \tag{21}$$

which, with (16), determines $f_n^*(x)$.

Note also that the polynomials $g_n \equiv g_n(\rho)$ appear also in the dual

† An equivalent result has been found independently by Frank A. Haight, of the University of California.

problem of the busy period of a single server with exponential service time and regular arrivals considered by Connolly.[10]

The first service has the distribution function

$$\begin{aligned} S_1(t) &= bt, \qquad t \leqq b^{-1}, \\ &= 1, \qquad t \geqq b^{-1} \end{aligned} \tag{22}$$

and the distribution function for n services is

$$\begin{aligned} S_n(t) &= 0, && t \leqq (n-1)b^{-1}, \\ &= bt, && (n-1)b^{-1} < t < nb^{-1}, \\ &= 1, && nb^{-1} < t. \end{aligned} \tag{23}$$

So

$$G^*(t) = \sum_{n=1} f_n{}^* S_n(t) \tag{24}$$

is the (conditional) delay distribution function for last-come first-served service with Poisson input to a single server with constant service time. The distribution function for the busy period (Takács[4]) is

$$G(t) = \sum_{n=0}^{[bt]} e^{-\rho n}(\rho n)^{n-1}/n!, \tag{25}$$

with $[bt]$ the integral part of bt.

REFERENCES

1. Vaulot, E., Delais d'attente des appels téléphoniques dans l'ordre inverse de leur arrivée, C. R. Acad. Sci. (Paris), **238,** 1954, pp. 1188–1189.
2. Palm, C., Specialnummer för teletrafikteknik, Tekniska Meddelanden från Kungl. Telegrafstyrelsen, 1946; English version: Research on Telephone Traffic Carried by Full Availability Groups, Tele, No. 1, 1957.
3. Burke, P. J., Equilibrium Delay Distribution for One Channel with Constant Holding Time, Poisson Input and Random Service, B.S.T.J., **38,** 1959, pp. 1021–1031.
4. Takács, L., Investigation of Waiting Time Problems by Reduction to Markov Processes, Acta Math. Acad. Sci. Hung., **6,** 1955, pp. 101–129.
5. Palm, C., Intensitätsschwankungen im Fernsprechverkehr, Ericsson Tech., No. 44, 1943.
6. Wishart, D. M. G., Queuing Systems in Which the Discipline is "Last-Come, First-Served", Oper. Res., **8,** 1960, pp. 591–599.
7. Campbell, G. A., and Foster, R. M., *Fourier Integrals for Practical Applications*, D. Van Nostrand Co., New York, 1948.
8. Salie, H., Über Abel's Verallgemeinerung der binomischen Formel, Ber. Verh. Sächs. Akad. Wiss. Leipzig Math.-Nat. Kl., **98,** 1951, pp. 19–22.
9. Jensen, J. L. W. V., Sur une identite d'Abel et sur d'autres formules analogures, Acta Math. **26,** 1902, pp. 307–318.
10. Connolly, B. W., The Busy Period in Relation to the Queuing Process GI/M/1, Biometrika, **46,** 1959, pp. 246–251.

Stochastic Processes with Balking in the Theory of Telephone Traffic*

By LAJOS TAKÁCS

(Manuscript received January 16, 1961)

It is supposed that at a telephone exchange calls are arriving according to a recurrent process. If an incoming call finds exactly j lines busy then it either realizes a connection with probability p_j or balks with probability q_j $(p_j + q_j = 1)$. The holding times are mutually independent random variables with common exponential distribution. In this paper the stochastic behavior of the fluctuation of the number of the busy lines is studied.

I. INTRODUCTION

Many results in telephone traffic theory (and elsewhere) may be unified by the introduction of *balking*. A call is said to balk if for any reason it refuses service on arrival. A mathematical model for balking is constructed by assigning a probability to balking dependent only on the state of the system; if an incoming call finds exactly j lines busy, then it realizes a connection with probability p_j and balks with probability q_j $(p_j + q_j = 1)$. Thus if $p_j = 1$ $(j = 0, 1, \cdots)$ the system is one with an infinite number of lines and with no loss and no delay, the ideal for for any service, while if $p_j = 1$ $(j = 0, 1, \cdots, m - 1)$ and $p_j = 0$ $(j = m, m + 1, \cdots)$ the system is a loss system with m lines.

This balking model is examined here for recurrent input and exponential distribution of holding times. More specifically, the call arrival times are taken as the instants $\tau_1, \tau_2, \cdots, \tau_n, \cdots$, where the interarrival times $\theta_n = \tau_{n+1} - \tau_n$ $(n = 0, 1, \cdots; \tau_0 = 0)$ are identically distributed, mutually independent, positive random variables with distribution function

$$\mathbf{P}\{\theta_n \leqq x\} = F(x). \tag{1}$$

* Dedicated to the memory of my professor Charles Jordan (December 16, 1871–December 24, 1959)

The holding times are identically distributed, mutually independent random variables with distribution function

$$H(x) = \begin{cases} 1 - e^{-\mu x} & \text{if } x \geqq 0, \\ 0 & \text{if } x < 0. \end{cases} \tag{2}$$

The holding times are independent of the $\{\tau_n\}$ as well.

Let us denote by $\xi(t)$ the number of busy lines at the instant t. Define $\xi_n = \xi(\tau_n - 0)$; that is, ξ_n is the number of busy lines immediately before the arrival of the nth call. The system is said to be in state E_k at the instant t if $\xi(t) = k$. Let us denote by m the smallest nonnegative integer such that $p_m = 0$. If $p_j > 0$ $(j = 0, 1, 2, \cdots)$ then $m = \infty$.

In the present paper we shall give a method to determine the distribution of ξ_n for every n, the distribution of $\xi(t)$ for finite t values, and the limiting distributions of ξ_n and $\xi(t)$ as $n \to \infty$ and $t \to \infty$ respectively. Further, we shall determine the stochastic law of the transitions $E_k \to E_{k+1}$ $(k = 0, 1, 2, \cdots)$.

II. NOTATION

The Laplace-Stieltjes transform of the distribution function of the interarrival times will be denoted by

$$\varphi(s) = \mathbf{E}\{e^{-s\theta_n}\} = \int_0^\infty e^{-sx}\, dF(x),$$

which is convergent if $\Re(s) \geqq 0$. The expectation of the interarrival times will be denoted by

$$\alpha = \mathbf{E}\{\theta_n\} = \int_0^\infty x\, dF(x).$$

Let $\mathbf{P}\{\xi_n = k\} = P_k^{(n)}$ and $\mathbf{P}\{\xi(t) = k\} = P_k(t)$. Define

$$\Pi_k(s) = \int_0^\infty e^{-st} P_k(t)\, dt,$$

which is convergent if $\Re(s) > 0$. Let

$$\lim_{n\to\infty} P_k^{(n)} = P_k \qquad \text{and} \qquad \lim_{t\to\infty} P_k(t) = P_k^*,$$

provided that the limits exist.

Define

$$C_r = \prod_{i=1}^{r} \left(\frac{\varphi(i\mu)}{1 - \varphi(i\mu)} \right) \qquad (r = 0, 1, 2, \cdots), \tag{3}$$

where the empty product means 1; that is, $C_0 = 1$. We shall also use the abbreviation

$$\varphi_r = \varphi(r\mu) = \int_0^\infty e^{-r\mu x}\, dF(x) \qquad (r = 0, 1, 2, \cdots). \tag{4}$$

Denote by $M_k(t)$ the expected number of calls occurring in the time interval $(0,t]$ which find exactly k lines busy. The expected number of transitions $E_k \to E_{k+1}$ occurring in the time interval $(0,t]$ is clearly $p_k M_k(t)$. Denote by $N_k(t)$ the expected number of transitions $E_{k+1} \to E_k$ occurring in the time interval $(0,t]$.

Let $G_k(x)(k = 0, 1, 2, \cdots)$ be the distribution function of the time differences between successive transitions $E_{k-1} \to E_k$ and $E_k \to E_{k+1}$, while $R_k(x)(k = 0, 1, 2, \cdots)$ is the distribution function of the time differences between consecutive transitions $E_k \to E_{k+1}$. If $\xi(0) = 0$ then we say that a transition $E_{-1} \to E_0$ takes place at time $t = -0$. Write

$$\gamma_k(s) = \int_0^\infty e^{-sx}\, dG_k(x)$$

and

$$\rho_k(s) = \int_0^\infty e^{-sx}\, dR_k(x)$$

which are convergent if $\Re(s) \geqq 0$.

III. PREVIOUS RESULTS

3.1 *A. K. Erlang*

Erlang[1] has proved that, if $\{\tau_n\}$ forms a Poisson process of intensity λ—that is, $F(x) = 1 - e^{-\lambda x}$ for $x \geqq 0$—and further, $p_j = 1$ when $j < m$, $p_j = 0$ when $j \geqq m$, then

$$P_k^* = \frac{\dfrac{(\lambda/\mu)^k}{k!}}{\displaystyle\sum_{j=0}^{m} \frac{(\lambda/\mu)^j}{j!}} \qquad (k = 0, 1, \cdots, m). \tag{5}$$

In this case $P_k = P_k^*$ $(k = 0, 1, \cdots, m)$ also holds. This is the simplest loss system.

3.2 *Conny Palm*

Palm[2] has generalized the above result of Erlang for the case when $\{\tau_n\}$ forms a recurrent process and otherwise every assumption remains unchanged. Palm has proved that

$$P_m = \frac{1}{\sum_{r=0}^{m} \binom{m}{r} \frac{1}{C_r}}, \tag{6}$$

where C_r is defined by (3). In this case the complete limiting distributions $\{P_k\}$ and $\{P_k^*\}$ have been determined by Pollaczek,[3] Cohen,[4] and the author.[5,6] The transient behavior of the sequence $\{\xi_n\}$ was determined by Pollaczek[3] and Beneš,[7] and the transient behavior of the process $\{\xi(t)\}$ by Beneš[8] and by the author.[9]

3.3 *The Infinite Line Case*

The case when $p_j = 1$ $(j = 0, 1, 2, \cdots)$ has been investigated by the author,[10,11] who has proved that

$$P_k = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r \qquad (k = 0, 1, 2, \cdots) \tag{7}$$

and, if $F(x)$ is not a lattice distribution and if $\alpha < \infty$, then the limiting distribution $\{P_k^*\}$ exists and

$$\begin{aligned} P_k^* &= \frac{P_{k-1}}{k\alpha\mu} \qquad (k = 1, 2, \cdots), \\ P_0^* &= 1 - \frac{1}{\alpha\mu} \sum_{k=1}^{\infty} \frac{P_{k-1}}{k}. \end{aligned} \tag{8}$$

The transient behavior of the process $\{\xi(t)\}$ is also treated in Refs. 10 and 11.

3.4 *The Case* $p_0 = 1$, $p_j = p$ $(j = 1, 2, \cdots)$, $q_j = q$ $(j = 1, 2, \cdots)$

This case, where $p + q = 1$, plays an important role in the theory of particle counters and has been investigated by the author,[12] who has found that

$$P_0 = \frac{p \sum_{r=0}^{\infty} (-p)^r C_r}{1 - q \sum_{r=0}^{\infty} (-p)^r C_r} \tag{9}$$

and

$$P_k = \frac{\sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} C_r}{1 - q \sum_{r=0}^{\infty} (-p)^r C_r} \qquad (k = 1, 2, \cdots). \tag{10}$$

If $F(x)$ is not a lattice distribution and if $\alpha < \infty$, then the limiting distribution $\{P_k^*\}$ exists and

$$\begin{aligned} P_{k+1}^* &= \frac{p_k P_k}{(k+1)\alpha\mu} \qquad (k = 0, 1, 2, \cdots), \\ P_0^* &= 1 - \frac{1}{\alpha\mu} \sum_{k=0}^{\infty} \frac{p_k P_k}{(k+1)}. \end{aligned} \tag{11}$$

The transient behavior of the process $\{\xi(t)\}$ is also treated in Ref. 12.

3.5 *The Distribution Function $G_k(x)$*

This function plays an important role in the investigation of overflow traffic. In the infinite line case, i.e., when $p_j = 1$ $(j = 0, 1, 2, \cdots)$, Palm[2] has proved that $\gamma_k(s)(k = 0, 1, 2, \cdots)$ satisfies the following recurrence formula:

$$\gamma_k(s) = \frac{\gamma_{k-1}(s+\mu)}{1 - \gamma_{k-1}(s) + \gamma_{k-1}(s+\mu)} \qquad (k = 1, 2, \cdots), \tag{12}$$

where $\gamma_0(s) = \varphi(s)$. Palm has obtained $\gamma_k(s)$ explicitly when $\{\tau_n\}$ is a Poisson process; that is, $\varphi(s) = \lambda/(\lambda + s)$. Then

$$\gamma_k(s) = \frac{\sum_{j=0}^{k} \binom{k}{j} \frac{s(s+\mu)\cdots[s+(j-1)\mu]}{\lambda^j}}{\sum_{j=0}^{k+1} \binom{k+1}{j} \frac{s(s+\mu)\cdots[s+(j-1)\mu]}{\lambda^j}}. \tag{13}$$

The general solution of the recurrence formula (12) is

$$\gamma_k(s) = \frac{\sum_{r=0}^{k} \binom{k}{r} \prod_{i=0}^{r-1} \left[\frac{1 - \varphi(s+i\mu)}{\varphi(s+i\mu)}\right]}{\sum_{r=0}^{k+1} \binom{k+1}{r} \prod_{i=0}^{r-1} \left[\frac{1 - \varphi(s+i\mu)}{\varphi(s+i\mu)}\right]} \qquad (k = 0, 1, 2, \cdots), \tag{14}$$

where the empty product means 1. The formula (14) is proved in Refs. 10 and 11.

In the particular case $p_0 = 1$, $p_j = p$ $(j = 1, 2, \cdots)$, $q_j =$

$q(j = 1, 2, \cdots)$, where $p + q = 1$, the Laplace-Stieltjes transform $\gamma_k(s)$ has been given explicitly in Ref. 12. We have

$$\gamma_k(s) = \frac{D_k(s)}{D_{k+1}(s)} \qquad (k = 0, 1, 2, \cdots), \tag{15}$$

where $D_0(s) \equiv 1$ and

$$D_k(s) = \left\{ p \sum_{r=0}^{k} \binom{k}{r} \prod_{i=0}^{r-1} \left[\frac{1 - \varphi(s + i\mu)}{p\varphi(s + i\mu)} \right] - \frac{q[1 - \varphi(s)]}{p\varphi(s)} \sum_{r=0}^{k} \binom{k}{r} \sum_{j=1}^{r-1} (-1)^j \prod_{i=j+1}^{r-1} \left[\frac{1 - \varphi(s + i\mu)}{p\varphi(s + i\mu)} \right] \right\} \tag{16}$$

if $k = 1, 2, \cdots$.

IV. THE TRANSIENT BEHAVIOR OF $\{\xi_n\}$

It is easy to see that the sequence of random variables $\{\xi_n\}$ forms a homogeneous Markov chain with transition probabilities

$$p_{jk} = \mathbf{P}\{\xi_{n+1} = k \mid \xi_n = j\} = \int_0^\infty \pi_{jk}(x)\, dF(x), \tag{17}$$

where

$$\pi_{jk}(x) = p_j \binom{j+1}{k} e^{-k\mu x}(1 - e^{-\mu x})^{j+1-k} + q_j \binom{j}{k} e^{-k\mu x}(1 - e^{-\mu x})^{j-k} \tag{18}$$

is the conditional transition probability given that the interarrival time $\theta_n = x$ (constant). For, if $\xi_n = j$ and $\theta_n = x$, then ξ_{n+1} has a Bernoulli distribution, either with parameters $j + 1$ and $e^{-\mu x}$ when the nth call realizes a connection, or with parameters j and $e^{-\mu x}$ when the nth call does not. The system is said to be in state E_k at the nth step if $\xi_n = k$.

Starting from the initial distribution $\{P_k^{(1)}\}$ the distributions $\{P_k^{(n)}\}$ can be determined successively by the following formulas:

$$P_k^{(n+1)} = \sum_{j=k-1}^{\infty} p_{jk} P_j^{(n)} \qquad (n = 1, 2, \cdots). \tag{19}$$

However, it turns out that in many cases it is more convenient to determine the binomial moments of $\{P_k^{(n)}\}$ first. By definition,

$$U_r^{(n)} = \mathbf{E}\left\{\binom{\xi_n}{r}\right\} = \sum_{k=r}^{\infty} \binom{k}{r} P_k^{(n)} \qquad (r = 0, 1, 2, \cdots) \tag{20}$$

is the rth binomial moment of $\{P_k^{(n)}\}$. If we suppose that $U_r^{(1)} < C_1^r/r!$ where C_1 is a constant, then it can be proved that every $U_r^{(n)}$ exists and $U_r^{(n)} < C^r/r!$ where C is a constant. Thus the distribution $\{P_k^{(n)}\}$ is uniquely determined by $\{U_r^{(n)}\}$. We obtain from (20) that

$$P_k^{(n)} = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} U_r^{(n)} \qquad (k = 0, 1, 2, \cdots). \tag{21}$$

This is the inversion formula of Jordan.[13]

It is convenient to use the related quantities

$$V_r^{(n)} = \mathbf{E}\left\{\binom{\xi_n}{r} p_{\xi_n}\right\} = \sum_{k=r}^{\infty} \binom{k}{r} p_k P_k^{(n)} \qquad (r = 0, 1, 2, \cdots), \tag{22}$$

whence by inversion

$$p_k P_k^{(n)} = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} V_r^{(n)}. \tag{23}$$

Now we shall prove

Theorem 1. We have $U_0^{(n)} = 1$ $(n = 1, 2, \cdots)$ *and*

$$U_r^{(n+1)} = \varphi_r(U_r^{(n)} + V_{r-1}^{(n)}) \qquad (n = 1, 2, \cdots; \quad r = 1, 2, \cdots), \tag{24}$$

where $\varphi_r = \varphi(r\mu)$. *Further*

$$V_r^{(n)} = \sum_{j=r}^{\infty} \binom{j}{r} (\Delta^{j-r} p_r) U_j^{(n)} \qquad (r = 0, 1, 2, \cdots), \tag{25}$$

where

$$\Delta^{j-r} p_r = \sum_{\nu=0}^{j-r} (-1)^{\nu} \binom{j-r}{\nu} p_{j-\nu}. \tag{26}$$

Proof. First of all we note that the rth binomial moment of the Bernoulli distribution $\{Q_k\}$ with parameters n and p, that is, that of

$$Q_k = \binom{n}{k} p^k (1 - p)^{n-k} \qquad (k = 0, 1, \cdots, n),$$

is given by

$$B_r = \sum_{k=r}^{n} \binom{k}{r} Q_k = \binom{n}{r} p^r \qquad (r = 0, 1, \cdots, n). \tag{27}$$

Using (27), we get by (18) that

$$\mathbf{E}\left\{\binom{\xi_{n+1}}{r} \mid \xi_n = j, \theta_n = x\right\} = p_j \binom{j+1}{r} e^{-r\mu x} + q_j \binom{j}{r} e^{-r\mu x},$$

whence

$$\begin{aligned}\mathbf{E}\left\{\binom{\xi_{n+1}}{r} \mid \xi_n = j\right\} &= \varphi_r\left[p_j\binom{j+1}{r} + q_j\binom{j}{r}\right] \\ &= \varphi_r\left[\binom{j}{r} + p_j\binom{j}{r-1}\right].\end{aligned} \tag{28}$$

If we multiply both sides of (28) by $P_j^{(n)}$ and add them for every j, then we get (24). We obtain (25) if we put (21) into (22). This completes the proof of the theorem.

Starting from $U_r^{(1)}(r = 1, 2, \cdots)$ the binomial moments $U_r^{(n)}(n = 2, 3, \cdots)$ can be obtained recursively by (24) and (25). If, specifically, $\xi(0) = i$ and $\tau_1 = x$ then ξ_1 has a Bernoulli distribution with parameters i and $e^{-\mu x}$ and thus, for $\xi(0) = i$,

$$U_r^{(1)} = \mathbf{E}\left\{\binom{\xi_1}{r}\right\} = \binom{i}{r}\varphi_r \qquad (r = 0, 1, 2, \cdots). \tag{29}$$

Remark 1. If we introduce the generating functions

$$U_r(w) = \sum_{n=1}^{\infty} U_r^{(n)} w^n \tag{30}$$

and

$$V_r(w) = \sum_{n=1}^{\infty} V_r^{(n)} w^n \tag{31}$$

and suppose that $\xi(0) = i$, then by (24) and (29) we get that

$$U_r(w) = \frac{w\varphi_r}{1 - w\varphi_r}\left[\binom{i}{r} + V_{r-1}(w)\right] \qquad (r = 1, 2, \cdots), \tag{32}$$

and evidently

$$U_0(w) = \frac{w}{1 - w}. \tag{33}$$

Note also that (21) implies that

$$\sum_{n=1}^{\infty} P_k^{(n)} w^n = \sum_{r=k}^{\infty} (-1)^{r-k}\binom{r}{k} U_r(w). \tag{34}$$

Example 1. In the infinite line case, i.e., when $p_j = 1(j = 0, 1, 2, \cdots)$, $V_r^{(n)} = U_r^{(n)}$ and $V_r(w) = U_r(w)$ for $r = 0, 1, 2, \cdots$. If we suppose that $\xi(0) = i$, then by (32) we get

$$U_r(w) = \frac{w\varphi_r}{1 - w\varphi_r}\left[\binom{i}{r} + U_{r-1}(w)\right] \qquad (r = 1, 2, \cdots) \tag{35}$$

and $U_0(w) = w/(1 - w)$. The solution of these equations is given by

$$U_r(w) = \left\{\prod_{j=0}^{r}\left(\frac{w\varphi_j}{1 - w\varphi_j}\right)\right\}\left\{\sum_{j=0}^{r}\binom{i}{j}\prod_{\nu=0}^{j-1}\left(\frac{1 - w\varphi_\nu}{w\varphi_\nu}\right)\right\} \qquad (r = 0, 1, 2, \cdots),$$

where the empty product means 1. The distribution $\{P_k^{(n)}\}$ is determined by (34).

Example 2. For a loss system with m lines, i.e., when $p_j = 1$ $(j < m)$ and $p_j = 0$ $(j \geqq m)$, in the case $\xi(0) = i \leqq m$ we have

$$V_r^{(n)} = U_r^{(n)} - \binom{m}{r} U_m^{(n)} \qquad (r = 0, 1, 2, \cdots, m - 1)$$

and

$$V_r^{(n)} = U_r^{(n)} = 0 \qquad (r = m, m + 1, \cdots).$$

Thus,

$$V_r(w) = U_r(w) - \binom{m}{r} U_m(w) \qquad (r = 0, 1, 2, \cdots, m - 1)$$

and

$$V_r(w) = U_r(w) = 0 \qquad (r = m, m + 1, \cdots).$$

By (32) we get

$$U_r(w) = \frac{w\varphi_r}{1 - w\varphi_r}\left[\binom{i}{r} + U_{r-1}(w) - \binom{m}{r - 1} U_m(w)\right] \qquad (r = 1, 2, \cdots, m) \tag{36}$$

and $U_0(w) = w/(1 - w)$. The solution of these equations for $r = 0, 1, 2, \cdots, m$ is given by

$$U_r(w) = \frac{\Gamma_r(w)}{\sum_{j=0}^{m}\binom{m}{j}\frac{1}{\Gamma_j(w)}}\left\{\left[\sum_{j=r}^{m}\binom{m}{j}\frac{1}{\Gamma_j(w)}\right]\left[\sum_{j=0}^{r}\binom{i}{j}\frac{1}{\Gamma_{j-1}(w)}\right] - \left[\sum_{j=0}^{r-1}\binom{m}{j}\frac{1}{\Gamma_j(w)}\right]\left[\sum_{j=r+1}^{m}\binom{i}{j}\frac{1}{\Gamma_{j-1}(w)}\right]\right\} \tag{37}$$

where

$$\Gamma_r(w) = \prod_{i=0}^{r}\left(\frac{w\varphi_i}{1 - w\varphi_i}\right), \qquad (r = 0, 1, 2, \cdots)$$

and $\Gamma_{-1}(w) \equiv 1$. Finally, $\{P_k^{(n)}\}$ can be obtained by (34).

V. THE LIMITING DISTRIBUTION $\{P_k\}$

In the Markov chain $\{\xi_n\}$ the states E_0, E_1, $\cdots$, E_m form an irreducible closed set, while E_m, E_{m+1}, $\cdots$ are transient states. If either $m = \infty$ or $m < \infty$, but we restrict ourselves to the states E_0, E_1, $\cdots$, E_m, then the Markov chain $\{\xi_n\}$ is irreducible. The Markov chain $\{\xi_n\}$ is always aperiodic. Accordingly

$$\lim_{n\to\infty} P_k^{(n)} = P_k \qquad (k = 0, 1, 2, \cdots)$$

always exists and is independent of the initial distribution. There are two possibilities: either every $P_k = 0$ $(k = 0, 1, 2, \cdots)$ or $\{P_k\}$ is a probability distribution. (In the second case $P_k > 0$ if $k \leqq m$ and $P_k = 0$ if $k > m$.) In the second case $\{P_k\}$ is the unique stationary distribution of the Markov chain $\{\xi_n\}$ and conversely if there exists a stationary distribution then it is unique and agrees with the limiting distribution $\{P_k\}$.

In the particular case $p_j = 1$ $(j = 0, 1, 2, \cdots)$ the limiting distribution always exists, as has been proved in Ref. 10. In this special case

$$P_0 = \sum_{r=0}^{\infty} (-1)^r C_r > 0.$$

If we consider an arbitrary sequence $\{p_j\}$ then evidently

$$P_0 \geqq \sum_{r=0}^{\infty} (-1)^r C_r > 0,$$

whence it follows that $\{\xi_n\}$ belongs to the second class; that is, $\{P_k\}$ is a probability distribution.

The stationary distribution $\{P_k\}$ is uniquely determined by the following system of linear equations:

$$P_k = \sum_{j=k-1}^{\infty} p_{jk} P_j \tag{38}$$

and

$$\sum_{k=1}^{\infty} P_k = 1. \tag{39}$$

Since in this case $P_k^{(n)} = P_k$ for every n, we get (38) by (19). Now let us introduce the binomial moments

$$U_r = \sum_{k=r}^{\infty} \binom{k}{r} P_k \qquad (r = 0, 1, 2, \cdots) \tag{40}$$

and define

$$V_r = \sum_{k=r}^{\infty} \binom{k}{r} p_k P_k . \tag{41}$$

By inversion we get, from (40),

$$P_k = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} U_r \qquad (k = 0, 1, 2, \cdots) \tag{42}$$

and similarly, from (41),

$$p_k P_k = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} V_r . \tag{43}$$

The binomial moments $U_r (r = 0, 1, 2, \cdots)$ can be obtained by the following

Theorem 2. We have $U_0 = 1$ *and*

$$U_r = \frac{\varphi_r}{1 - \varphi_r} V_{r-1} \qquad (r = 1, 2, \cdots), \tag{44}$$

where $\varphi_r = \varphi(r\mu)$. *Further,*

$$V_r = \sum_{j=r}^{\infty} \binom{j}{r} (\Delta^{j-r} p_r) U_j \qquad (r = 0, 1, 2, \cdots), \tag{45}$$

where

$$\Delta^{j-r} p_r = \sum_{\nu=0}^{j-r} (-1)^{\nu} \binom{j-r}{\nu} p_{j-\nu} . \tag{46}$$

Proof. This theorem immediately follows from Theorem 1 if we put $U_r^{(n)} = U_r$, $V_r^{(n)} = V_r$ in (24) and (25).

Remark 2. In many cases there is a simple relation between the generating functions

$$U(z) = \sum_{k=0}^{\infty} P_k z^k \tag{47}$$

and

$$V(z) = \sum_{k=0}^{\infty} p_k P_k z^k \tag{48}$$

when U_r $(r = 0, 1, \cdots)$ can easily be obtained by (44). For,

$$U_r = \frac{1}{r!} \left[\frac{d^r U(z)}{dz^r} \right]_{z=1} \qquad (r = 0, 1, 2, \cdots) \tag{49}$$

and

$$V_r = \frac{1}{r!}\left[\frac{d^r V(z)}{dz^r}\right]_{z=1} \qquad (r = 0, 1, 2, \cdots). \tag{50}$$

Theorem 3. The binomial moments U_r $(r = 0, 1, 2, \cdots)$ *satisfy the following system of linear equations*:

$$\sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \left(p_k U_r - \frac{1 - \varphi_{r+1}}{\varphi_{r+1}} U_{r+1}\right) = 0 \qquad (r = 0, 1, 2, \cdots) \tag{51}$$

and

$$U_{r+1} = \frac{\varphi_{r+1}}{1 - \varphi_{r+1}} \sum_{j=r}^{\infty} \binom{j}{r} (\Delta^{j-r} p_r) U_j \qquad (r = 0, 1, 2, \cdots), \tag{52}$$

where $\Delta^{j-r} p_r$ *is defined by* (46).

Proof. If we put (42) into (43) and use the relation (44) then we get (51). If we eliminate V_r from (44) and (45) then we get (52).

Remark 3. If $p_m = 0$ then $U_r = 0$ for $r > m$, and in this case, starting from U_m, the unknowns U_{m-1}, U_{m-2}, $\cdots$, U_0 can be obtained successively either by (51) or by (52) and finally $U_0 = 1$ determines U_m. If the higher differences of p_r vanish, then (52) can be used successfully for the determination of the binomial moments U_r.

Example 3. If $p_j = 1$ $(j = 0, 1, 2, \cdots)$ then $V_r = U_r$ $(r = 0, 1, 2, \cdots)$ and, by (44),

$$U_r = \frac{\varphi_r}{1 - \varphi_r} U_{r-1} \qquad (r = 1, 2, \cdots),$$

whence

$$U_r = \prod_{j=1}^{r} \left(\frac{\varphi_j}{1 - \varphi_j}\right) \qquad (r = 1, 2, \cdots) \tag{53}$$

and $U_0 = 1$. The distribution $\{P_k\}$ is given by (42).

Example 4. Let $p_j = 1$ if $j < m$ and $p_j = 0$ if $j \geqq m$. Then

$$V_r = U_r - \binom{m}{r} U_m \qquad (r = 0, 1, \cdots, m)$$

and

$$V_r = U_r = 0 \qquad (r = m + 1, m + 2, \cdots).$$

By (44)

$$U_r = \frac{\varphi_r}{1 - \varphi_r}\left[U_{r-1} - \binom{m}{r-1} U_m\right] \qquad (r = 1, 2, \cdots, m),$$

and the solution of this equation is

$$U_r = C_r \frac{\sum_{j=r}^{m} \binom{m}{j} \frac{1}{C_j}}{\sum_{j=0}^{m} \binom{m}{j} \frac{1}{C_j}} \qquad (r = 0, 1, \cdots, m), \tag{54}$$

where C_r is defined by (3). $U_r = 0$ if $r > m$. Finally, $\{P_k\}$ is given by (42).

Example 5. Let $p_0 = 1$ and $p_j = p$ $(j = 1, 2, \cdots)$, $q_j = q$ $(j = 1, 2, \cdots)$, where $p + q = 1$. Then

$$V_r = pU_r \qquad (r = 1, 2, \cdots),$$

$$V_0 = pU_0 + qP_0 = 1 - q\,(U_1 - U_2 + U_3 - \cdots).$$

Putting V_r into (44) we get

$$U_r = \frac{p\varphi_r}{1 - \varphi_r} U_{r-1} \qquad (r = 1, 2, \cdots)$$

and

$$U_r = \frac{\varphi_1}{1 - \varphi_1}[1 - q(U_1 - U_2 + U_3 - \cdots)].$$

The solution of this system of linear equations is

$$U_r = \frac{p^r C_r}{1 - q\sum_{j=0}^{\infty} (-p)^j C_j} \qquad (r = 0, 1, 2, \cdots), \tag{55}$$

where C_r is defined by (3). Finally, $\{P_k\}$ is given by (42).

Example 6. Let $p_0 = 1$, $p_j = p$, and $q_j = q$ if $j = 1, 2, \cdots, m - 1$, where $p + q = 1$, and $p_j = 0$ if $j > m$. Then

$$V_0 = p + qP_0 - pP_m$$

$$= p + q[U_0 - U_1 + U_2 - \cdots + (-1)^m U_m] - pU_m,$$

$$V_r = pU_r - p\binom{m}{r} U_m \qquad (r = 1, 2, \cdots, m),$$

$$V_r = U_r = 0 \qquad (r = m + 1, m + 2, \cdots).$$

Now $U_0 = 1$ and, by (44),

$$U_r = \frac{p^r C_r \sum_{j=r}^{m} \binom{m}{j} \frac{1}{C_j p^j}}{\sum_{j=0}^{m} \binom{m}{j} \frac{1}{C_j p^j} - q \sum_{j=0}^{m} (-1)^j C_j p^j \sum_{i=j}^{m} \binom{m}{i} \frac{1}{C_i p^i}} \qquad (r = 1, 2, \cdots, m). \tag{56}$$

The distribution $\{P_k\}$ is given by (42).

Example 7. If, in particular, $F(x) = 1 - e^{-\lambda x}$ for $x \geqq 0$, then $\varphi(s) = \lambda/(\lambda + s)$ and $\varphi_r = \lambda/(\lambda + r\mu)(r = 0, 1, 2, \cdots)$. In this case by (24) we have

$$r\mu U_r = \lambda V_{r-1} \qquad (r = 1, 2, \cdots),$$

whence

$$\mu U'(z) = \lambda V(z).$$

Forming the coefficient of z^{k-1} we obtain that

$$\mu k P_k = \lambda p_{k-1} P_{k-1} \qquad (k = 1, 2, \cdots), \tag{57}$$

whence

$$P_k = P_0 \frac{\left(\frac{\lambda}{\mu}\right)^k}{k!} p_0 p_1 \cdots p_k \qquad (k = 0, 1, 2, \cdots),$$

and P_0 is determined by the requirement that

$$\sum_{k=0}^{\infty} P_k = 1.$$

VI. THE TRANSIENT BEHAVIOR OF $\{\xi(t)\}$

In this section we suppose that $\xi(0) = i$ always. Denote by $M_j(t)$ the expectation of the number of calls occurring in the time interval $(0,t]$ which find exactly j lines busy. Let

$$\mu_j(s) = \int_0^{\infty} e^{-st}\, dM_j(t), \tag{58}$$

which is convergent if $\Re(s) > 0$. Now we shall prove the following

Lemma 1. Define

$$\Phi_r(s) = \sum_{j=r}^{\infty} \binom{j}{r} \mu_j(s) \qquad (r = 0, 1, 2, \cdots) \tag{59}$$

and

$$\Psi_r(s) = \sum_{j=r}^{\infty} \binom{j}{r} p_j \mu_j(s) \qquad (r = 0, 1, 2, \cdots), \tag{60}$$

which are convergent if $\Re(s) > 0$. *Then*

$$\Phi_0(s) = \frac{\varphi(s)}{1 - \varphi(s)} \tag{61}$$

and if $\xi(0) = i$ *then*

$$\Phi_r(s) = \frac{\varphi(s + r\mu)}{1 - \varphi(s + r\mu)} \left[\binom{i}{r} + \Psi_{r-1}(s) \right]. \tag{62}$$

Proof. Since evidently

$$M_j(t) = \sum_{n=1}^{\infty} \mathbf{P}\{\tau_n \leqq t, \quad \xi_n = j\}, \tag{63}$$

we have

$$\Phi_r(s) = \sum_{j=r}^{\infty} \binom{j}{r} \mu_j(s) = \sum_{n=1}^{\infty} \mathbf{E}\left\{ e^{-s\tau_n} \binom{\xi_n}{r} \right\} \tag{64}$$

and similarly

$$\Psi_r(s) = \sum_{j=r}^{\infty} \binom{j}{r} p_j \mu_j(s) = \sum_{n=1}^{\infty} \mathbf{E}\left\{ e^{-s\tau_n} \binom{\xi_n}{r} p_{\xi_n} \right\}. \tag{65}$$

Now we shall prove that

$$\mathbf{E}\left\{ e^{-s\tau_{n+1}} \binom{\xi_{n+1}}{r} \mid \xi_n = j, \quad \theta_n = x, \quad \tau_n = y \right\}$$
$$= \left[p_j \binom{j+1}{r} + q_j \binom{j}{r} \right] e^{-r\mu x} e^{-s(x+y)}.$$

This follows from the fact that under the given condition ξ_{n+1} has a Bernoulli distribution either with parameters $j + 1$ and $e^{-\mu x}$ when the nth call gives rise to a connection, or with parameters j and $e^{-\mu x}$ when the nth call does not. Unconditionally we get

$$\mathbf{E}\left\{ e^{-s\tau_{n+1}} \binom{\xi_{n+1}}{r} \right\}$$
$$= \varphi(s + r\mu) \left[\mathbf{E}\left\{ e^{-s\tau_n} \binom{\xi_n}{r} \right\} + \mathbf{E}\left\{ e^{-s\tau_n} \binom{\xi_n}{r-1} p_{\xi_n} \right\} \right]. \tag{66}$$

If $\xi(0) = i$ then

$$\mathbf{E}\left\{e^{-s\tau_1}\binom{\xi_1}{r}\right\} = \binom{i}{r}\varphi(s + r\mu). \tag{67}$$

If we add (66) for $n = 1, 2, \cdots$ and (67) then we get

$$\Phi_r(s) = \varphi(s + r\mu)\left[\binom{i}{r} + \Phi_r(s) + \Psi_{r-1}(s)\right] \qquad (r = 0, 1, 2, \cdots), \tag{68}$$

where $\Psi_{-1}(s) \equiv 0$. Thus we get (61) and (62). In many cases use of Lemma 1 determines $\Phi_r(s)$ $(r = 0, 1, 2, \cdots)$ explicitly.

Remark 4. From (59) we obtain by inversion

$$\mu_k(s) = \sum_{r=k}^{\infty} (-1)^{r-k}\binom{r}{k}\Phi_r(s). \tag{69}$$

The functions $\mu_k(s)(k = 0, 1, 2, \cdots)$ can be determined also by the following system of linear equations:

$$\sum_{k=r}^{\infty}\binom{k}{r}\mu_k(s) = \frac{\varphi(s + r\mu)}{1 - \varphi(s + r\mu)}\left[\binom{i}{r} + \sum_{k=r-1}^{\infty}\binom{k}{r-1}p_k\mu_k(s)\right], \tag{70}$$

which we get if we put (59) and (60) into (62).

If we know $\Phi_r(s)$ $(r = 0, 1, 2, \cdots)$ then $P_k(t)$ can be determined by the following

Theorem 4. The Laplace transform $\Pi_k(s)$ is given by

$$\Pi_k(s) = \sum_{r=k}^{\infty} (-1)^{r-k}\binom{r}{k}\beta_r(s), \tag{71}$$

where

$$\beta_r(s) = \frac{[1 - \varphi(s + r\mu)]\Phi_r(s)}{\varphi(s + r\mu)(s + r\mu)} \qquad (r = 0, 1, 2, \cdots). \tag{72}$$

Proof. Let the rth binomial moment of $\{P_k(t)\}$ be defined by

$$B_r(t) = \mathbf{E}\left\{\binom{\xi(t)}{r}\right\} = \sum_{k=r}^{\infty}\binom{k}{r}P_k(t) \qquad (r = 0, 1, 2, \cdots). \tag{73}$$

By using the results of Ref. 10 we can see that $B_r(t) \leqq C^r/r!$ for every $t \geqq 0$, where C is a constant. Thus the probability distribution $\{P_k(t)\}$ is uniquely determined by its binomial moments. From (73) we get by inversion

$$P_k(t) = \sum_{r=k}^{\infty} (-1)^{r-k}\binom{r}{k}B_r(t). \tag{74}$$

If

$$\beta_r(s) = \int_0^\infty e^{-st} B_r(t)\, dt$$

and we form the Laplace transform of (74), we get (71). Now let us determine $\beta_r(s)$ $(r = 0, 1, 2, \cdots)$.

If $\xi(0) = i$, then

$$B_r(t) = \binom{i}{r} e^{-r\mu t}[1 - F(t)] + \sum_{j=0}^{\infty}\left[p_j\binom{j+1}{r} + q_j\binom{j}{r}\right]\int_0^t e^{-r\mu(t-u)}[1 - F(t-u)]\, dM_j(u), \tag{75}$$

where $M_j(t)$ is defined by (63). For, if there is no call in the time interval $(0,t]$ then $\xi(t)$ has a Bernoulli distribution with parameters i and $e^{-\mu t}$. If the last call in the time interval $(0,t]$ occurs at the instant u and in that instant the number of busy lines is j, then $\xi(t)$ has a Bernoulli distribution, either with parameters $j + 1$ and $e^{-\mu(t-\mu)}$ when this call gives rise to a connection or with parameters j and $e^{-\mu(t-\mu)}$ when this call does not. If we also take into consideration that the last call occurring in the time interval $(0,t]$ may be the 1st, 2nd, $\cdots$, nth, $\cdots$ one, then we get (75). Forming the Laplace transform of (74) we get

$$\beta_r(s) = \frac{1 - \varphi(s + r\mu)}{s + r\mu}\left\{\binom{i}{r} + \sum_{j=0}^{\infty}\left[p_j\binom{j+1}{r} + q_j\binom{j}{r}\right]\mu_j(s)\right\} \tag{76}$$

where $\mu_j(s)$ is defined by (58). By using the notations (59) and (60) we can write also that

$$\beta_r(s) = \frac{1 - \varphi(s + r\mu)}{s + r\mu}\left\{\binom{i}{r} + \Phi_r(s) + \Psi_{r-1}(s)\right\}. \tag{77}$$

Taking into consideration the relation (68) we obtain finally

$$\beta_r(s) = \frac{1 - \varphi(s + r\mu)}{(s + r\mu)}\,\frac{\Phi_r(s)}{\varphi(s + r\mu)}, \tag{78}$$

which was to be proved.

Example 8. Define

$$C_r(s) = \prod_{i=0}^{r}\left(\frac{\varphi(s + i\mu)}{1 - \varphi(s + i\mu)}\right) \qquad (r = 0, 1, 2, \cdots) \tag{79}$$

and

$$C_{-1}(s) \equiv 1.$$

If $p_j = 1$ $(j = 0, 1, 2, \cdots)$ and $\xi(0) = i$, then $\Psi_r(s) = \Phi_r(s)$ $(r = 0, 1, 2, \cdots)$ and, by (62), we get

$$\Phi_r(s) = \frac{\varphi(s + r\mu)}{[1 - \varphi(s + r\mu)]}\left[\binom{i}{r} + \Phi_{r-1}(s)\right] \qquad (r = 0, 1, \cdots), \quad (80)$$

where $\Phi_{-1}(s) = 0$. The solution of this recurrence formula is

$$\Phi_r(s) = C_r(s) \sum_{j=0}^{r} \binom{i}{j} \frac{1}{C_{j-1}(s)}, \quad (81)$$

where $C_r(s)$ is defined by (79).

Example 9. If $p_j = 1$ when $j < m$ and $p_j = 0$ when $j \geqq m$ and $\xi(0) = i \leqq m$, then

$$\Psi_r(s) = \Phi_r(s) - \binom{m}{r}\Phi_m(s) \qquad (r = 0, 1, \cdots, m)$$

and

$$\Psi_r(s) = \Phi_r(s) = 0 \qquad (r = m + 1, m + 2, \cdots).$$

By (62)

$$\Phi_r(s) = \frac{\varphi(s + r\mu)}{[1 - \varphi(s + r\mu)]}\left[\binom{i}{r} + \Phi_{r-1}(s) - \binom{m}{r-1}\Phi_m(s)\right] \quad (82)$$

for $r = 1, 2, \cdots, m$. The solution of this equation is

$$\Phi_r(s) = \frac{C_r(s)}{\sum_{j=0}^{m}\binom{m}{j}\frac{1}{C_j(s)}}\left\{\left[\sum_{j=r}^{m}\binom{m}{j}\frac{1}{C_j(s)}\right]\left[\sum_{j=0}^{r}\binom{i}{j}\frac{1}{C_{j-1}(s)}\right] - \left[\sum_{j=0}^{r-1}\binom{m}{j}\frac{1}{C_j(s)}\right]\left[\sum_{j=r+1}^{m}\binom{i}{j}\frac{1}{C_{j-1}(s)}\right]\right\} \quad (83)$$

where $C_j(s)$ is defined by (79).

VII. THE LIMITING DISTRIBUTION $\{P_k^*\}$

Now we shall prove

Theorem 5. If $F(x)$ is not a lattice distribution and its mean α is finite, then the limiting distribution

$$\lim_{t\to\infty} P_k(t) = P_k^* \qquad (k = 0, 1, \cdots)$$

exists and is independent of the initial distribution. We have

$$P_{k+1}^* = \frac{p_k P_k}{(k+1)\alpha\mu} \qquad (k = 0, 1, 2, \cdots) \tag{84}$$

and

$$P_0^* = 1 - \frac{1}{\alpha\mu} \sum_{k=0}^{\infty} \frac{p_k P_k}{k+1}, \tag{85}$$

where $\{P_k\}$ *is defined by* (38).

Proof. By the theory of Markov chains we can conclude that

$$\lim_{t\to\infty} \frac{M_k(t)}{t} = \frac{P_k}{\alpha}. \tag{86}$$

Furthermore, it is clear that the difference of the number of transitions $E_k \to E_{k+1}$ and $E_{k+1} \to E_k$ occurring in the time interval $(0,t]$ is at most 1. Accordingly, if we denote by $N_k(t)$ the expectation of the number of transitions $E_{k+1} \to E_k$ occurring in the time interval $(0,t]$, then

$$| p_k M_k(t) - N_k(t) | \leqq 1 \tag{87}$$

for all $t \geqq 0$. Further,

$$N_k(t) = (k+1)\mu \int_0^t P_{k+1}(u)\, du, \tag{88}$$

for, if we consider the process $\{\xi(t)\}$ only at those instants when there is state E_{k+1}, then the transitions $E_{k+1} \to E_k$ form a Poisson process of density $(k+1)\mu$. Thus, by (86), (87), and (88),

$$\lim_{t\to\infty} \frac{(k+1)\mu}{t} \int_0^t P_{k+1}(u)\, du = \lim_{t\to\infty} \frac{N_k(t)}{t} = \lim_{t\to\infty} \frac{p_k M_k(t)}{t} = \frac{p_k P_k}{\alpha};$$

that is,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t P_{k+1}(u)\, du = \frac{p_k P_k}{(k+1)\alpha\mu} \qquad (k = 0, 1, 2, \cdots). \tag{89}$$

If we prove that the limiting distribution

$$\lim_{t\to\infty} P_k(t) = P_k^* \qquad (k = 0, 1, 2, \cdots)$$

exists, then it follows by (89) that

$$P_{k+1}^* = \frac{p_k P_k}{(k+1)\alpha\mu} \qquad (k = 0, 1, 2, \cdots), \tag{90}$$

and so

$$P_0{}^* = 1 - \sum_{k=0}^{\infty} P_{k+1}{}^* = 1 - \frac{1}{\alpha\mu} \sum_{k=0}^{\infty} \frac{p_k P_k}{(k+1)}. \tag{91}$$

To prove the existence of the limiting distribution we need the following auxiliary theorem: If $F(x)$ is not a lattice distribution, then

$$\lim_{t\to\infty} \frac{p_k M_k(t+h) - p_k M_k(t)}{h} \tag{92}$$

exists for every $h > 0$ and is independent of h and the initial state. This is a consequence of a theorem of Blackwell.[14] For the time differences between successive transitions $E_k \to E_{k+1}$ are identically distributed, independent, positive random variables, and, if $F(x)$ is not a lattice distribution, then these random variables have no lattice distribution either. If (92) exists, then it follows that

$$\lim_{t\to\infty} \frac{M_k(t+h) - M_k(t)}{h} = \lim_{t\to\infty} \frac{M_k(t)}{t} = \frac{P_k}{\alpha} \qquad (k = 0, 1, 2, \cdots). \tag{93}$$

Now, by the theorem of total probability, we can write

$$\begin{aligned} P_k(t) = \binom{i}{k} e^{-k\mu x} (1 - e^{-\mu x})^{i-k} [1 - F(t)] \\ + \sum_{j=k-1}^{\infty} \int_0^t \pi_{jk}(t-u)[1 - F(t-u)]\, dM_j(u), \end{aligned} \tag{94}$$

where $\pi_{jk}(t)$ is defined by (18) and it is supposed that $\xi(0) = i$. The event $\xi(t) = k$ may occur in several mutually exclusive ways: there is no call in the time interval $(0,t]$ and, with the exception of k, all the i connections terminate by t; or the last call in the time interval $(0,t]$ is the nth $(n = 1, 2, \cdots)$ one and it finds state E_j $(j = k-1, k, \cdots)$. If $\tau_n = u$ $(0 < u \leqq t)$, then during the time interval $(u,t]$ no new call arrives [the probability of which is $1 - F(t-u)$] and with the exception of k connections every connection terminates by t [the probability of which is $\pi_{jk}(t-u)$].

Applying Blackwell's theorem to (94) and using $\alpha < \infty$, it follows that

$$\lim_{t\to\infty} P_k(t) = P_k{}^* \qquad (k = 0, 1, \cdots)$$

exists and

$$P_k{}^* = \sum_{j=k-1}^{\infty} p_{jk}{}^* P_j, \tag{95}$$

where

$$p_{jk}^* = \frac{1}{\alpha}\int_0^\infty \pi_{jk}(x)[1 - F(x)]\,dx. \tag{96}$$

It is easy to see from (95) that $\{P_k^*\}$ is a probability distribution.

VIII. THE DETERMINATION OF $\gamma_k(s)$

Define

$$\gamma_k(s) = \int_0^\infty e^{-sx}\,dG_k(x) = \frac{D_k(s)}{D_{k+1}(s)}, \tag{97}$$

where $D_0(s) = 1$. We are going to determine $D_r(s)(r = 1, 2, \cdots)$.

Write $D_r(s)$ in the following form:

$$D_r(s) = \sum_{j=0}^{r}\binom{r}{j}\Delta^j D_0(s), \tag{98}$$

where $\Delta^j D_0(s)$ is the jth difference of $D_r(s)$ at $r = 0$; that is,

$$\Delta^j D_0(s) = \sum_{i=0}^{j}(-1)^{j-i}\binom{j}{i}D_i(s). \tag{99}$$

Then $D_r(s)$ is uniquely determined by its differences.

Now we shall prove

Theorem 6. Starting from $D_0(s) = \Delta^0 D_0(s) = 1$, *the functions* $D_r(s)$ $(r = 0, 1, 2, \cdots)$ *and the differences* $\Delta^j D_0(s)$ $(j = 0, 1, 2, \cdots)$ *can be obtained successively by the recurrence formulas*

$$\begin{aligned}\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}D_j(s)\\ = \varphi(s + j\mu)\sum_{j=0}^{r}(-1)^{r-j}\binom{r}{j}[p_j D_{j+1}(s) + q_j D_j(s)]\end{aligned} \tag{100}$$

and

$$\Delta^j D_0(s) = \frac{\varphi(s + j\mu)}{1 - \varphi(s + j\mu)}\sum_{i=0}^{j}\binom{j}{i}(\Delta^{j-i}p_i)\Delta^{i+1}D_0(s) \tag{101}$$

respectively. Here

$$\Delta^{j-i}p_i = \sum_{\nu=0}^{j-i}(-1)^\nu\binom{j-i}{\nu}p_{j-\nu}. \tag{102}$$

Proof. By the theorem of total probability we can write for $r = 0, 1, 2, \cdots$ that

$$G_r(x) = \int_0^x \sum_{j=0}^{r} \binom{r}{j} e^{-j\mu y} (1 - e^{-\mu y})^{r-j} \cdot [p_j G_{j+1}(x - y) * \cdots * G_r(x - y) + q_j G_j(x - y) * \cdots * G_r(x - y)] \, dF(y), \tag{103}$$

where the empty convolution product is equal to 1. Let us consider the instant of a transition $E_{r-1} \to E_r$ and measure time from this instant. Then $G_r(x)$ is the probability that the next transition $E_r \to E_{r+1}$ occurs in the time interval $(0,x]$. This event may occur in the following mutually exclusive ways: the first call in the time interval $(0,x]$ arrives at the instant $y (0 < y \leqq x)$, it finds state E_j $(j = 0, 1, \cdots, r)$, the probability of which is

$$\binom{r}{j} e^{-j\mu y} (1 - e^{-\mu y})^{r-j},$$

and, in the time interval $(y,x]$, a transition $E_r \to E_{r+1}$ occurs, the probability of which is

$$p_j G_{j+1}(x - y) * \cdots * G_r(x - y) + q_j G_j(x - y) * \cdots * G_r(x - y).$$

Introduce the notation

$$q_{r,j}(s) = \binom{r}{j} \int_0^\infty e^{-sx} e^{-j\mu x} (1 - e^{-\mu x})^{r-j} \, dF(x) \tag{104}$$

and form the Laplace-Stieltjes transform of (103); then

$$\gamma_r(s) = \sum_{j=0}^{r} q_{rj}(s) \left[p_j \prod_{i=j+1}^{r} \gamma_i(s) + q_j \prod_{i=j}^{r} \gamma_i(s) \right] \qquad (r = 0, 1, 2, \cdots),$$

where the empty product is 1. Now using (97) we find

$$D_r(s) = \sum_{j=0}^{r} q_{rj}(s)[p_j D_{j+1}(s) + q_j D_j(s)] \qquad (r = 0, 1, 2, \cdots). \tag{105}$$

This is already a recurrence formula for the determination of $D_r(s)$ $(r = 0, 1, 2, \cdots)$, but the coefficients can be simplified further.

If we form

$$\Delta^j D_0(s) = \sum_{l=0}^{j} (-1)^{j-l} \binom{j}{l} D_l(s),$$

where $D_l(s)$ is replaced by (105), and take into consideration that

$$\sum_{l=i}^{j} (-1)^{j-l} \binom{j}{l} q_{l,i}(s) = (-1)^{j-i} \binom{j}{i} \varphi(s + j\mu), \tag{106}$$

then we obtain

$$\Delta^j D_0(s) = \varphi(s + j\mu) \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} [p_i D_{i+1}(s) + q_i D_i(s)]. \tag{107}$$

Now, comparing (99) and (107), we obtain (100).

On the other hand, by (107) it follows that

$$\Delta^j D_0(s) = \varphi(s + j\mu)\Delta^j D_0(s) + \varphi(s + j\mu) \sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} p_i \Delta D_i(s),$$

whence

$$\Delta^j D_0(s) = \frac{\varphi(s + j\mu)}{1 - \varphi(s + j\mu)} \Delta^j [p_0 \Delta D_0(s)] \tag{108}$$

and here

$$\Delta^j [p_0 \Delta D_0(s)] = \sum_{i=0}^{j} \binom{j}{i} (\Delta^{j-i} p_i) \Delta^{i+1} D_0(s), \tag{109}$$

where

$$\Delta^{j-i} p_i = \sum_{\nu=0}^{j-i} (-1)^{\nu} \binom{j-i}{\nu} p_{j-\nu}. \tag{110}$$

This proves (101).

Example 10. In the infinite line case, i.e., when $p_j = 1$ ($j = 0, 1, 2, \cdots$), (101) has the following simple form:

$$\Delta^{j+1} D_0(s) = \frac{1 - \varphi(s + j\mu)}{\varphi(s + j\mu)} \Delta^j D_0(s) \qquad (j = 0, 1, 2, \cdots), \tag{111}$$

whence

$$\Delta^j D_0(s) = \prod_{i=0}^{j-1} \left[\frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)} \right] \tag{112}$$

and

$$D_r(s) = \sum_{j=0}^{r} \binom{r}{j} \prod_{i=0}^{j-1} \left(\frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)} \right). \tag{113}$$

Example 11. If $p_0 = 1$ and $p_j = p$ $(j = 1, 2, \cdots)$, then (101) reduces to the following difference equation:

$$\Delta^{j+1}D_0(s) - \frac{1 - \varphi(s + j\mu)}{\varphi(s + j\mu)}\Delta^j D_0(s) + (-1)^j \frac{q[1 - \varphi(s)]}{p\varphi(s)} = 0 \qquad (j = 0, 1. 2, \cdots). \tag{114}$$

A simple calculation shows that the solution of (114) is

$$\Delta^j D_0(s) = \left\{ p \prod_{i=0}^{j-1} \left[\frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)} \right] - \frac{q[1 - \varphi(s)]}{p\varphi(s)} \sum_{\nu=1}^{j-1} (-1)^\nu \prod_{i=\nu+1}^{j-1} \left[\frac{1 - \varphi(s + i\mu)}{p\varphi(s + i\mu)} \right] \right\}, \tag{115}$$

and finally,

$$D_r(s) = \sum_{j=0}^{r} \binom{r}{j} \Delta^j D_0(s). \tag{116}$$

Theorem 7. Suppose that $\xi(0) = 0$ *and under this condition denote by* $M_k(t)$ *the expectation of the number of calls arriving in the time interval* $(0,t]$ *which find exactly* k *lines busy. Let*

$$\mu_k(s) = \int_0^\infty e^{-st}\, dM_k(t). \tag{117}$$

Then

$$\rho_k(s) = 1 - \frac{1}{p_k D_{k+1}(s)\mu_k(s)}, \tag{118}$$

where $D_{k+1}(s)$ *is given by Theorem 6 and* $\mu_k(s)$ *is given by*

$$\mu_k(s) = \sum_{r=k}^{\infty} (-1)^{r-k} \binom{r}{k} \Phi_r(s), \tag{119}$$

where $\Phi_r(s)$ *can be obtained by Lemma 1.*

Proof. The expected number of transitions $E_k \to E_{k+1}$ occurring in the time interval $(0,t]$ is evidently $p_k M_k(t)$. The time differences between consecutive transitions $E_k \to E_{k+1}$ are identically distributed, independent random variables with distribution function $R_k(x)$. By using renewal theory we can write that

$$p_k M_k(t) = G_0(t) * G_1(t) * \cdots * G_k(t) * [I(t) + R_k(t) + R_k(t) * R_k(t) + \cdots], \tag{120}$$

where $I(t) = 1$ if $t \geqq 0$ and $I(t) = 0$ if $t < 0$. Forming the Laplace-

Stieltjes transform of (121), we obtain

$$p_k\mu_k(s) = \frac{\gamma_0(s)\gamma_1(s)\cdots\gamma_k(s)}{1 - \rho_k(s)} = \frac{1}{D_{k+1}(s)[1 - \rho_k(s)]}, \qquad (121)$$

whence (118) follows.

Since we know the distribution functions $G_k(x)$ and $R_k(x)$ ($k = 0, 1, 2, \cdots$), the distribution of the number of transitions $E_k \to E_{k+1}$ occurring in the time interval $(0,t]$ can be obtained easily.

IX. THE OVERFLOW TRAFFIC

Suppose that $p_j = 1$ ($j = 0, 1, 2, \cdots$) and that the telephone lines are numbered by 1, 2, 3, $\cdots$. Further suppose that an incoming call realizes a connection through the idle line that has the lowest serial number. Consider the group $(1, 2, \cdots, m)$. Denote by $\pi_m^{(n)}$ the probability that the nth call finds every line busy in the group $(1, 2, \cdots, m)$. The distances between successive calls which find every line busy in the group $(1, 2, \cdots, m)$ are evidently identically distributed, independent random variables with distribution function, say, $G_m(x)$.

Palm[2] proved that

$$\pi_m = \lim_{n\to\infty} \pi_m^{(n)} = \frac{1}{\sum_{r=0}^{m} \binom{m}{r} \frac{1}{C_r}}, \qquad (122)$$

where C_r is defined by (3). This is in agreement with (6). In this case it is easy to see that $\pi_m^{(n)} = P_m^{(n)}$, where the distribution $\{P_k^{(n)}\}$ is defined in Example 2 of Section IV.

In Refs. 10 and 11 it is shown that

$$\int_0^\infty e^{-sx}\, dG_m(x) = \frac{\sum_{r=0}^{m} \binom{m}{r} \prod_{i=0}^{r-1} \left[\frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)}\right]}{\sum_{r=0}^{m+1} \binom{m+1}{r} \prod_{i=0}^{r-1} \left[\frac{1 - \varphi(s + i\mu)}{\varphi(s + i\mu)}\right]}, \qquad (123)$$

where the empty product means 1. It is easy to see that $G_m(x)$ agrees with the corresponding $G_m(x)$ defined in Section VIII when $p_j = 1$ ($j = 0, 1, 2, \cdots$). Thus (123) can be obtained from (97) and (113).

Remark 5. Denote by Γ_m the expectation of the random variable which is the difference of call numbers of successive calls, both of which find all lines busy in the group $(1, 2, \cdots, m)$. Knowing Γ_m, we can write that

$$\pi_m = \lim_{n\to\infty} \pi_m^{(n)} = \frac{1}{\Gamma_m} \qquad (124)$$

and

$$\int_0^\infty x\, dG_m(x) = \alpha\Gamma_m. \tag{125}$$

In Ref. 6 it is shown that Γ_r $(r = 1, 2, \cdots)$ satisfies the following recurrence formula:

$$\begin{aligned}\Gamma_r = q_{r,0}(\Gamma_1 + \Gamma_2 + \cdots + \Gamma_r) + q_{r,1}(\Gamma_2 + \Gamma_3 + \cdots + \Gamma_r) \\ + \cdots + q_{r,r-2}(\Gamma_{r-1} + \Gamma_r) + q_{r,r-1}\Gamma_r + 1,\end{aligned} \tag{126}$$

where

$$q_{r,j} = \binom{r}{j}\int_0^\infty e^{-j\mu x}(1 - e^{-\mu x})^{r-j}\, dF(x) \qquad (j = 0, 1, \cdots, r). \tag{127}$$

The solution of (126) is given by

$$\Gamma_r = \sum_{j=0}^{r}\binom{r}{j}\prod_{i=1}^{j}\left(\frac{1 - \varphi_i}{\varphi_i}\right) \qquad (r = 1, 2, \cdots). \tag{128}$$

REFERENCES

1. Erlang, A. K., Solution of Some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges, P. O. Elect. Eng. J., **10,** 1918, p. 189.
2. Palm, C., Intensitätsschwankungen im Fernsprechverkehr, Eric. Tech., No. 44, 1943.
3. Pollaczek, F., Généralisation de la théorie probabiliste des systèmes téléphoniques sans dispositif d'attente, C. R. Acad. Sci. (Paris), **236,** 1953, p. 1469.
4. Cohen, J. W., The Full Availability Group of Trunks with an Arbitrary Distribution of the Interarrival Times and a Negative Exponential Holding-Time Distribution, Simon Stevin Wis-en Natuurkundig Tijdschrift, **31,** 1957, p. 169.
5. Takács, L., On the Generalization of Erlang's Formula, Acta Math. Acad. Sci. Hung., **7,** 1956, p. 419.
6. Takács, L., On a Probability Problem Concerning Telephone Traffic, Acta Math. Acad. Sci. Hung., **8,** 1957, p. 319.
7. Beneš, V. E., On Trunks with Negative Exponential Holding Times Serving a Renewal Process, B.S.T.J., **38,** 1959, p. 211.
8. Beneš, V. E., Transition Probabilities for Telephone Traffic, B.S.T.J., **39,** 1960, p. 1297.
9. Takács, L., The Time Dependence of Palm's Loss Formula, to be published.
10. Takács, L., On a Coincidence Problem Concerning Telephone Traffic, Acta Math. Acad. Sci. Hung., **9,** 1958, p. 45.
11. Takács, L., On the Limiting Distribution of the Number of Coincidences Concerning Telephone Traffic, Ann. Math. Stat., **30,** 1959, p. 134.
12. Takács, L., On a Coincidence Problem Concerning Particle Counters, to be published.
13. Jordan, C., *Chapters on the Classical Probability Theory* (in Hungarian), Akadémiai Kiadó, Budapest, 1956.
14. Blackwell, D., A Renewal Theorem, Duke Math. J., **15,** 1948, p. 145.

A New Technique for Increasing the Flexibility of Recursive Least Squares Data Smoothing

By N. LEVINE

(Manuscript received October 19, 1960)

A method of performing recursive least squares data smoothing is described in which optimum (or arbitrary) weights can be assigned to the observations. The usual restriction of a constant data interval can be removed without affecting the optimum weighting or recursive features. The method also provides an instantaneous (i.e. real time) estimate of the statistical accuracy in the smoothed coordinates for a set of arbitrary data intervals. Optimum gate sizes for arbitrary predictions can be determined. These features greatly increase the flexibility of recursive least squares data smoothing, and several applications are discussed.

I. INTRODUCTION

During the past few years, a need has arisen for data smoothing techniques which can be applied, in real time, to radar observations of bodies traveling along highly predictable trajectories. The observations are usually processed in digital computers, so that much effort has been expended in devising techniques suited to the advantages and limitations of computers. This paper is concerned with one such technique, recursive least squares smoothing, whose theoretical foundation was established several years ago.[1] It is our purpose to show how this technique can be made considerably more flexible so as to encompass a wide variety of practical situations while maintaining its suitability for computer use.

By the term "recursive," we mean that a smoothed coordinate is determined from a previously computed average of past data (one number) and a new observation. Thus the storage requirements are independent of the number of observations and are actually quite modest. We will also use the term "optimum smoothing," which is to be interpreted in the least squares sense, i.e., the weighting of data inversely proportional to

their error variances. The ability to perform optimum smoothing in real time where the variations (not necessarily the magnitudes) in observational error are determinable (e.g., trajectory-dependent errors) is one of the advantages of the method described here. Over long periods of observation, a significant increase in the accuracy of the smoothed coordinates may be achieved by properly weighting the data. In the actual derivation, however, no restrictions are placed on the weighting factors, thus allowing an arbitrary sequence of weights to be placed on the data.

Sections II and III are devoted to the estimation of position and velocity assuming the body under observation is traveling along a straight line with no acceleration. In Section II we consider the case of constant data intervals where some emphasis has been placed on the ability to compensate properly for missing observations. We have also included formulas for determining optimum gate sizes for predicted observations, which are of considerable importance in some tracking systems. In Section III the method is extended to handle the case of arbitrary data intervals.

That the restriction to linear flight does not limit the applicability of this method to many practical cases is shown in Section IV, where methods for including the effects of known accelerations are discussed. In Appendix A we show that the sum of square deviations from the least squares line of regression may be obtained recursively. This result may be useful as a means of real-time error detection or as a way of providing an instantaneous estimate for the average observational error during tracking. Appendix B describes the changes necessary in performing least squares recursive smoothing over a fixed number of prior observations — in contrast to the method in the main text, where the smoothing is effective over all observations. Finally, in Appendix C, the extension of the method to include estimation of acceleration is outlined.

The derivation and discussion of this method of data smoothing are presented in the context of radar observations of moving bodies. However, this technique can be applied to any observable quantity which can be expressed as a linear combination of functions where the expansion coefficients are to be estimated.

II. CONSTANT DATA INTERVALS

The derivation of this method of performing recursive least squares smoothing is based on a technique employed by Kaplan.[2] For simplicity, we initially consider the case of straight-line unaccelerated flight. A sequence of observations $\mathring{x}_i$, $i = 1, 2, \cdots, n$, τ seconds apart, are made by an instrument whose measurement errors are taken to be uncorrelated

and normally distributed with mean zero and variance σ_i^2. We wish to obtain the least squares estimate of the position ($\bar{x}_n$) and velocity ($\bar{v}_n$) as of the nth observation* by suitably combining the latest observation ($\mathring{x}_n$) with a linear combination of the $n-1$ previous observations. To do this, we will first minimize the sum of squared deviations from the least squares line of regression, R_n^2, assuming n observations have been made. We may write this sum as follows:

$$R_n^2 = \sum_{i=1}^{n} \{\mathring{x}_i - [\bar{x}_n - (n-i)\bar{u}_n]\}^2 w_i , \tag{1}$$

where $\bar{u}_n = \bar{v}_n \tau$ and w_i is an arbitrary weighting factor. Differentiating this equation with respect to $\bar{x}_n$ and $\bar{u}_n$ and setting each resulting equation equal to zero yields the normal equations:

$$\begin{aligned} \bar{x}_n \sum_{i=1}^{n} w_i - \bar{u}_n \sum_{i=1}^{n} (n-i) w_i &= \sum_{i=1}^{n} \mathring{x}_i w_i , \\ -\bar{x}_n \sum_{i=1}^{n} (n-i) w_i + \bar{u}_n \sum_{i=1}^{n} (n-i)^2 w_i &= -\sum_{i=1}^{n} \mathring{x}_i (n-i) w_i . \end{aligned} \tag{2}$$

We define the sums F_n, G_n, and H_n as

$$\begin{aligned} F_n &= \sum_{i=1}^{n} w_i , \\ G_n &= \sum_{i=1}^{n} (n-i) w_i , \\ H_n &= \sum_{i=1}^{n} (n-i)^2 w_i , \end{aligned} \tag{3}$$

so that equations (2) take the form

$$F_n \bar{x}_n - G_n \bar{u}_n = \sum_{i=1}^{n} \mathring{x}_i w_i , \tag{4}$$

$$-G_n \bar{x}_n + H_n \bar{u}_n = -\sum_{i=1}^{n} \mathring{x}_i (n-i) w_i . \tag{5}$$

The simultaneous solution of (4) and (5) yields the standard least squares estimates of position and velocity; however, they explicitly involve the presence of all observations $\mathring{x}_i$. To cast this in recursive form,

* Estimation to any other time $(n+p)\tau$ can be included by replacing (1) by

$$\sum_{i=1}^{n} \{\mathring{x}_i - [\bar{x}_{n+p} - (n+p-i)\bar{u}_{n+p}]\}^2 w_i = \min$$

we repeat the process of (1) through (5) omitting the last observation $\mathring{x}_n$. This yields estimates of the $(n-1)$st position and velocity:

$$R_{n-1}^{\,2} = \sum_{i=1}^{n-1} \{\mathring{x}_1 - [\bar{x}_{n-1} - (n-1-i)\bar{u}_{n-1}]\}^2 w_i \,. \tag{6}$$

Differentiating with respect to $\bar{x}_{n-1}$ and $\bar{u}_{n-1}$ and making use of definitions (3), the resulting equations may be written in the form

$$F_n(\bar{x}_{n-1} + \bar{u}_{n-1}) - G_n\bar{u}_{n-1} = \sum_{i=1}^{n-1} \mathring{x}_i w_i + w_n(\bar{x}_{n-1} + \bar{u}_{n-1}), \tag{7}$$

$$-G_n(\bar{x}_{n-1} + \bar{u}_{n-1}) + H_n\bar{u}_{n-1} = -\sum_{i=1}^{n} \mathring{x}_i(n-i)w_i. \tag{8}$$

Subtracting (7) from (4) and (8) from (5), we have

$$\begin{aligned} F_n[\bar{x}_n - (\bar{x}_{n-1} + \bar{u}_{n-1})] - G_n(\bar{u}_n - \bar{u}_{n-1}) \\ = w_n[\mathring{x}_n - (\bar{x}_{n-1} + \bar{u}_{n-1})], \\ -G_n[\bar{x}_n - (\bar{x}_{n-1} + \bar{u}_{n-1})] + H_n(\bar{u}_n - \bar{u}_{n-1}) = 0, \end{aligned} \tag{9}$$

whose solutions are:

$$\bar{x}_n = (\bar{x}_{n-1} + \bar{u}_{n-1}) + \alpha_n[\mathring{x}_n - (\bar{x}_{n-1} + \bar{u}_{n-1})], \tag{10}$$

$$\bar{u}_n = \bar{u}_{n-1} + \beta_n[\mathring{x}_n - (\bar{x}_{n-1} + \bar{u}_{n-1})], \tag{11}$$

where

$$\alpha_n = \frac{w_n H_n}{J_n}, \tag{12}$$

$$\beta_n = \frac{w_n G_n}{J_n}, \tag{13}$$

$$J_n = F_n H_n - G_n^{\,2}. \tag{14}$$

Equations (10) and (11) explicitly indicate that a smoothed coordinate is a linear combination of "old" data with a new observation; α_n and β_n are the position and velocity smoothing coefficients and are calculated for each new observation from (12) and (13) using definitions (3) and (14).

The estimates $(\bar{x}_{n-1} + \bar{u}_{n-1})$ and $\bar{u}_{n-1}$ are merely the predicted nth

position and velocity (times the data interval) based on $n - 1$ observations, so that we may define

$$\hat{x}_n = \bar{x}_{n-1} + \bar{u}_{n-1}, \qquad \hat{u}_n = \bar{u}_{n-1}. \tag{15}$$

The smoothing equations now take the form

$$\bar{x}_n = \hat{x}_n + \alpha_n(\mathring{x}_n - \hat{x}_n) = (1 - \alpha_n)\hat{x}_n + \alpha_n\mathring{x}_n, \tag{16}$$

$$\bar{u}_n = \hat{u}_n + \beta_n(\mathring{x}_n - \hat{x}_n) = (\hat{u}_n - \beta_n\hat{x}_n) + \beta_n\mathring{x}_n, \tag{17}$$

which show that the predicted nth position and velocity may be used to represent all past data.

The ability to optimize the smoothing for varying measurement errors σ_i^2 arises from the fact that the quantities F_n, G_n, H_n, and J_n can themselves be summed recursively:

$$\begin{aligned} F_n &= F_{n-1} + w_n, \\ G_n &= G_{n-1} + F_{n-1}, \\ H_n &= H_{n-1} + 2G_{n-1} + F_{n-1}, \\ J_n &= J_{n-1} + w_nH_n. \end{aligned} \tag{18}$$

Thus, each new observation can be arbitrarily weighted by w_n and, as can be shown by statistical analysis, optimumly weighted by choosing $w_n = 1/\sigma_n^2$. Since J_n is defined by (14), the last recursion relation above may be used as a consistency check. Note that a missing observation may be properly accounted for by choosing its weighting factor equal to zero and cycling the sums as usual. Equations (10) through (13) then set the smoothed coordinates equal to their predicted values, and the future weighting of both old and new data is now altered to compensate for the missing observation. To illustrate this, Tables I and II have been constructed to show how the weighting of data changes for the case of missing second and fifth observations. Here we have chosen $w_i = 1$ for simplicity. Note that the effect of the missing observations on the variance ratios $\sigma_{\bar{x}_n}^2/\sigma_0^2$ and $\sigma_{\bar{u}_n}^2/\sigma_0^2$ (see Table II) diminishes rapidly with the addition of new observations compared to the values of these ratios when all observations are present.

The actual weighting coefficients applied to the observations to yield the smoothed position and velocity coordinates can be determined by solving (4) and (5) simultaneously. If we define

$$\bar{x}_n = \sum_{i=1}^{n} c_i\mathring{x}_i, \qquad \bar{u}_n = \sum_{i=1}^{n} d_i\mathring{x}_i, \tag{19}$$

TABLE I — ALL OBSERVATIONS PRESENT ($w_i = 1$)

n	1	2	3	4	5	6	7	8
F	1	2	3	4	5	6	7	8
G	0	1	3	6	10	15	21	28
H	0	1	5	14	30	55	91	140
J	0	1	6	20	50	105	196	336
α	(1)	1	5/6	7/10	3/5	11/21	13/28	5/12
β	(0)	1	1/2	3/10	1/5	3/21	3/28	1/12
σ_x^2/σ_0^2	1	1	0.833	0.700	0.600	0.524	0.464	0.417
σ_u^2/σ_0^2		2	0.500	0.200	0.100	0.057	0.036	0.024
	$\mathring{x}_1$	$\mathring{x}_2$	$\mathring{x}_3$	$\mathring{x}_4$	$\mathring{x}_5$	$\mathring{x}_6$	$\mathring{x}_7$	$\mathring{x}_8$
$\bar{x}_1$	1							
$\bar{x}_2$	0	1						
$\bar{x}_3$	−1/6	2/6	5/6					
$\bar{x}_4$	−2/10	1/10	4/10	7/10				
$\bar{x}_5$	−1/5	0	1/5	2/5	3/5			
$\bar{x}_6$	−4/21	−1/21	2/21	5/21	8/21	11/21		
$\bar{x}_7$	−5/28	−2/28	1/28	4/28	7/28	10/28	13/28	
$\bar{x}_8$	−2/12	−1/12	0	1/12	2/12	3/12	4/12	5/12
$\bar{u}_1$								
$\bar{u}_2$	−1	1						
$\bar{u}_3$	−1/2	0	1/2					
$\bar{u}_4$	−3/10	−1/10	1/10	3/10				
$\bar{u}_5$	−2/10	−1/10	0	1/10	2/10			
$\bar{u}_6$	−5/35	−3/35	−1/35	1/35	3/35	5/35		
$\bar{u}_7$	−3/28	−2/28	−1/28	0	1/28	2/28	3/28	
$\bar{u}_8$	−7/84	−5/84	−3/84	−1/84	1/84	3/84	5/84	7/84

then the solution of (4) and (5) yields:

$$c_i = \frac{1}{J_n}[H_n - (n - i)G_n]w_i\,, \tag{20}$$

$$d_i = \frac{1}{J_n}[G_n - (n - i)F_n]w_i\,. \tag{21}$$

These coefficients are tabulated in Tables I and II for the special cases considered.

In digital computer applications, this method would require storage of $\hat{x}_n$, $\hat{u}_n$, F_{n-1}, G_{n-1}, H_{n-1} (and J_{n-1}, if desired). Upon receipt of $\mathring{x}_n$, w_n is determined; the sums are updated, (18); α_n and β_n are computed, (12) and (13); $\bar{x}_n$ and $\bar{u}_n$ are determined, (16) and (17); and $\hat{x}_{n+1}$ and $\hat{u}_{n+1}$ are formed, (15), and stored with the current values of the sums. The amount of storage and computation per cycle is independent of the number of observations.

In situations where this type of smoothing can be employed, the method outlined here offers several advantages over conventional

TABLE II — OBSERVATIONS 2 AND 5 MISSED

n	1	2	3	4	5	6	7	8
F	1	1	2	3	3	4	5	6
G	0	1	2	4	7	10	14	19
H	0	1	4	10	21	38	62	95
J	0	0	4	14	14	52	114	209
α	(1)	0	1	5/7	0	19/26	31/57	95/209
β	(0)	0	1/2	2/7	0	5/26	7/57	19/209
σ_x^2/σ_0^2	1		1	0.714	1.500	0.741	0.544	0.455
σ_u^2/σ_0^2			0.500	0.214	0.214	0.077	0.044	0.029
	$\mathring{x}_1$	$\mathring{x}_2$	$\mathring{x}_3$	$\mathring{x}_4$	$\mathring{x}_5$	$\mathring{x}_6$	$\mathring{x}_7$	$\mathring{x}_8$
$\bar{x}_1$	1							
$\bar{x}_2$								
$\bar{x}_3$	0		1					
$\bar{x}_4$	−1/7		3/7	5/7				
$\bar{x}_5$	−1/2		1/2	1				
$\bar{x}_6$	−6/26		4/26	9/26		19/26		
$\bar{x}_7$	−11/57		3/57	10/57		25/57	31/57	
$\bar{x}_8$	−38/209		0	19/209		57/209	76/209	95/209
$\bar{u}_1$								
$\bar{u}_2$								
$\bar{u}_3$	−1/2		1/2					
$\bar{u}_4$	−5/14		1/14	4/14				
$\bar{u}_5$	−5/14		1/14	4/14				
$\bar{u}_6$	−5/26		−1/26	1/26		5/26		
$\bar{u}_7$	−16/114		−6/114	−1/114		9/114	14/114	
$\bar{u}_8$	−23/209		−11/209	−5/209		7/209	13/209	19/209

methods. For example, the weighting factors w_i can be arbitrarily chosen without affecting the exactness condition; i.e., in the absence of errors, true straight-line data are unaltered by the smoothing process. Thus, the w_i's can be chosen to place greater (or lesser) emphasis on new data by increasing (or decreasing) their relative weight with respect to old data. If the measurement errors (σ_i^2) vary in a known way with position, velocity, or time, the w_i can be chosen equal to $1/\sigma_i^2$ so as to optimize the smoothing for these errors. It is important to note here that optimum smoothing requires only a knowledge of the ratio of the errors, not their magnitudes. Of course, the magnitudes must be known in order to evaluate the statistical accuracy of the smoothed coordinates.

The presence of the sums makes it possible to determine the statistical accuracy in the smoothed coordinates at any time (assuming uncorrelated Gaussian measurement errors) for two important cases:

(a) the weighting factors are chosen equal to K/σ_i^2, where K is a normalization constant;

(b) $\sigma_i^2 = \sigma_0^2$ (a constant for all i) and $w_i = 1$ or 0.

In both cases, we may write

$$\sigma_{\bar{x}_n}^2 = \sum_{i=1}^{n} c_i^2 \sigma_i^2, \qquad \sigma_{\bar{u}_n}^2 = \sum_{i=1}^{n} d_i^2 \sigma_i^2, \tag{22}$$

which, by making use of (3) and (14), can be evaluated as

$$\sigma_{\bar{x}_n}^2 = K \frac{H_n}{J_n}, \tag{23}$$

$$\sigma_{\bar{u}_n}^2 = K \frac{F_n}{J_n} \quad \left(\text{or } \sigma_{\bar{v}_n} = \frac{K}{\tau^2} \frac{F_n}{J_n} \right) \tag{24}$$

for case (a), and K replaced with σ_0^2 for case (b).

The accuracy of an arbitrary prediction of p units ($p\tau$ seconds) into the past or future may be determined from

$$\begin{aligned} \sigma_{\hat{x}_{n+p}}^2 &= \sum_{i=1}^{n} (c_i + p d_i)^2 \sigma_i^2, \\ &= \frac{K}{J_n} [H_n + 2pG_n + p^2 F_n]. \end{aligned} \tag{25}$$

Equation (25) is of great value in certain tracking systems where the $(n + 1)$st observation is "captured" by making a prediction, $\hat{x}_{n+1}$ of $\mathring{x}_{n+1}$ and surrounding it with a gate within which the observation must fall. The optimum gate size is obtained by adding the variances of the $(n + 1)$st observation and the prediction. Thus, using (25) with $p = 1$, we find:

$$\text{optimum 1-sigma gate} = \begin{cases} \sqrt{\sigma_{n+1}^2 + K \dfrac{H_{n+1}}{J_n}}, & \text{case (a)}, \\ \sigma_0 \sqrt{1 + \dfrac{H_{n+1}}{J_n}}, & \text{case (b)}. \end{cases} \tag{26}$$

The gate size determined from (26) is automatically increased when observations are missed, and is optimum for any sequence of observations and misses.

For completeness, we present the explicit value of all quantities after n observations for the case of all $w_i = 1$:

$$F_n = n,$$

$$G_n = \frac{n(n-1)}{2},$$

$$H_n = \frac{n(n-1)(2n-1)}{6},$$

$$J_n = \frac{n^2(n^2-1)}{12},$$

$$\alpha_n = \frac{2(2n-1)}{n(n+1)},$$

$$\beta_n = \frac{6}{n(n+1)},$$

$$c_i = \frac{2[(2n-1) - 3(n-i)]}{n(n+1)},$$

$$d_i = \frac{6[(n-1) - 2(n-i)]}{(n-1)n(n+1)},$$

$$\sigma_{\bar{x}_n}^2 = \sigma_0^2 \frac{2(2n-1)}{n(n+1)} = \alpha_n \sigma_0^2,$$

$$\sigma_{\dot{v}_n}^2 = \frac{\sigma_0^2}{\tau^2} \frac{12}{(n-1)n(n+1)},$$

$$\sigma_{\dot{x}_{n+p}}^2 = 2\sigma_0^2 \frac{(n-1)(2n-1) + 6p(n-1) + 6p^2}{(n-1)n(n+1)}.$$

optimum

$$\text{1-sigma gate} = \sigma_0 \sqrt{\frac{(n+1)(n+2)}{n(n-1)}}.$$

III. VARIABLE DATA INTERVALS

In many cases of interest, the time intervals between observations may vary over wide limits. The results of the previous section can easily be extended to include the case of arbitrary observation times $\mathring{t}_i$ by replacing the quantity $(n - i)$ by $(\mathring{t}_n - \mathring{t}_i)$ and replacing u_n by v_n. The sums G_n and H_n are redefined as

$$\begin{aligned} G_n &= \sum_{i=1}^{n} (\mathring{t}_n - \mathring{t}_i) w_i, \\ H_n &= \sum_{i=1}^{n} (\mathring{t}_n - \mathring{t}_i)^2 w_i, \end{aligned} \tag{27}$$

where F_n is unchanged, (3), and the recursion relations become

$$\begin{aligned} F_n &= F_{n-1} + w_n , \\ G_n &= G_{n-1} + (\mathring{t}_n - \mathring{t}_{n-1})F_{n-1} , \\ H_n &= H_{n-1} + 2(\mathring{t}_n - \mathring{t}_{n-1})G_{n-1} + (\mathring{t}_n - \mathring{t}_{n-1})^2 F_{n-1} \\ &= H_{n-1} + (\mathring{t}_n - \mathring{t}_{n-1})(G_n + G_{n-1}), \\ J_n &= J_{n-1} + w_n H_n . \end{aligned} \tag{28}$$

The smoothing equations retain the same form, except that the smoothing coefficient β_n now implicitly contains the time dependence

$$\bar{v}_n = \hat{v}_n + \beta_n(\mathring{x}_n - \hat{x}_n). \tag{29}$$

The predicted $(n + 1)$st observation, $\hat{x}_{n+1}$, is now computed from

$$\hat{x}_{n+1} = \bar{x}_n + (\mathring{t}_{n+1} - \mathring{t}_n)\bar{v}_n \tag{30}$$

where, obviously, one must either know $\mathring{t}_{n+1}$ before receipt of the $(n + 1)$st observation or defer computation of $\hat{x}_{n+1}$ until after the $(n + 1)$st observation is made. For those systems which must predict the $(n + 1)$st observation in order to "capture" it, one must estimate the time of observation $(\hat{t}_{n+1})$ and apply an appropriate gate about $\hat{x}_{n+1}' = \bar{x}_n + (\hat{t}_{n+1} - \mathring{t}_n)\bar{v}_n$, large enough to account for all sources of error. No degradation in the quality of the fit is made if only a poor estimate of $\hat{t}_{n+1}$ can be obtained, since this is merely a device for capturing $\mathring{x}_{n+1}$. Once the observation has been received, $\hat{x}_{n+1}'$ can be corrected to yield $\hat{x}_{n+1}$ as follows:

$$\hat{x}_{n+1} = \hat{x}_{n+1}' + (\mathring{t}_{n+1} - \hat{t}_{n+1})\hat{v}_{n+1} . \tag{31}$$

In the preceding section, we described how a missing observation could be properly accounted for. In the case of variable data intervals, only the time difference between successive observations enters the equations. Thus, an observation that was anticipated ($\mathring{x}_{n+1}$ at $\mathring{t}_{n+1}$) but never made ($w_{n+1} = 0$) has no effect on the equation. The next observation is predicted at time $\mathring{t}_{n+2}$, i.e.,

$$\begin{aligned} \hat{x}_{n+2} &= \hat{x}_{n+1} + (\mathring{t}_{n+2} - \mathring{t}_{n+1})\hat{v}_{n+1} \\ &= \bar{x}_n + (\mathring{t}_{n+2} - \mathring{t}_n)\bar{v}_n , \end{aligned} \tag{32}$$

and the only quantity entering the equations is $(\mathring{t}_{n+2} - \mathring{t}_n)$.

The determination of the optimum gate size is now somewhat more

complicated than before. In general, $\hat{t}_{n+1}$ is determined from some function of present position and velocity, $T(\bar{x}_n, \bar{v}_n)$, so that

$$\hat{x}_{n+1} = \bar{x}_n + \bar{v}_n T(\bar{x}_n, \bar{v}_n).$$

In the approximation for small estimation errors, $\sigma_{\hat{x}_{n+1}}^2$ can be determined from*

$$\sigma_{\hat{x}_{n+1}}^2 = \sigma_{\bar{x}_n}^2 \left(\frac{\partial \hat{x}_{n+1}}{\partial \bar{x}_n}\right)^2 + 2\sigma_{\bar{x}_n \bar{v}_n} \left(\frac{\partial \hat{x}_{n+1}}{\partial \bar{x}_n}\right)\left(\frac{\partial \hat{x}_{n+1}}{\partial \bar{v}_n}\right) + \sigma_{\bar{v}_n}^2 \left(\frac{\partial \hat{x}_{n+1}}{\partial \bar{v}_n}\right)^2, \quad (33)$$

where, using (19) through (22) and definition (14), the covariance may be evaluated as

$$\sigma_{\bar{x}_n \bar{v}_n} = \sum_{i=1}^{n} c_i d_i \sigma_i^2 = K \frac{G_n}{J_n}$$

and the gate size becomes

$$\text{optimum 1-sigma gate} = \sqrt{\sigma_{\hat{x}_{n+1}}^2 + \sigma_{\hat{x}_{n+1}}^2}.$$

For most practical purposes, an adequate approximation to the optimum gate size can be made by first estimating $\hat{t}_{n+1}$ and then evaluating

$$\hat{H}_{n+1} = H_n + 2(\hat{t}_{n+1} - \mathring{t}_n)G_n + (\hat{t}_{n+1} - \mathring{t}_n)^2 F_n, \quad (34)$$

which is inserted in (26).

Note that for this case of variable data intervals, the quantities H_n/J_n and F_n/J_n still determine the instantaneous position and velocity variances, respectively, and (25) (with p now equal to the prediction time in seconds) determines the statistical accuracy of an arbitrary extrapolation or interpolation.

IV. KNOWN ACCELERATION

The previous sections have dealt with the case of zero acceleration, which is but a special case of motion with known acceleration. For the case of known and constant acceleration a, all the preceding results apply if one modification of the smoothing equation is made. The predicted $(n + 1)$st position and velocity should be determined as follows:

$$\begin{aligned} \hat{x}_{n+1} &= \bar{x}_n + \bar{v}_n(\hat{t}_{n+1} - \mathring{t}_n) + \tfrac{1}{2}a(\hat{t}_{n+1} - \mathring{t}_n)^2, \\ \hat{v}_{n+1} &= \bar{v}_n + a(\hat{t}_{n+1} - \mathring{t}_n) \end{aligned} \quad (35)$$

with $(\hat{t}_{n+1} - \mathring{t}_n) = \tau$ for the case of constant data intervals.

* See, for example, Ref. 3, p. 51.

If the acceleration is a known function of the position and velocity $a(x, v)$, one could treat this case as above by evaluating $a(x, v)$ at $(\bar{x}_n , \bar{v}_n)$ and using (35). This procedure works very well if $a(x, v)$ is not a sensitive function of position and velocity and the time interval between observations is not too large. Possible criteria for these restrictions for the case of constant data intervals are

$$\frac{\tau^2}{2}[a(\bar{x}_n \pm \sigma_{\bar{x}_n}, \bar{v}_n \pm \sigma_{\bar{v}_n}) - a(\bar{x}_n , \bar{v}_n)] \ll \sigma_{x_{n+1}}$$

or

$$\tau[a(\bar{x}_n \pm \sigma_{\bar{x}_n}, \bar{v}_n \pm \sigma_{\bar{v}_n}) - a(\bar{x}_n , \bar{v}_n)] \ll \sigma_{\hat{v}_{n+1}},$$

which express the fact that the variation in the acceleration due to the errors in the smoothed data should not contribute significantly to the error in predicted position or velocity. For acceleration functions or smoothing times [i.e., $(n - 1)\tau$] which do not satisfy the above criteria, other methods must be employed to optimize the smoothing.

In order to assess the systematic error in smoothed position and velocity due to a varying component of acceleration not compensated for by the method outlined above, one can replace the varying component by some average value. Then it is easy to show that, for a constant acceleration a, the differences between true position and velocity and the linearly smoothed values are

$$\begin{aligned} x_n{}^T - \bar{x}_n &= \frac{(n-1)(n-2)}{6}\left(\frac{a\tau^2}{2}\right), \\ v_n{}^T - \bar{v}_n &= \frac{n-1}{2}(a\tau). \end{aligned} \tag{36}$$

It is obvious that this method of performing least squares smoothing can be extended to the case of an unknown but *constant* acceleration which is a special case of motion with known "jerk" (rate of change of acceleration). The derivation and some results for this case are given in Appendix C, and a comparison with the estimation errors for linear smoothing is made.

V. ACKNOWLEDGMENTS

To the extent of the author's knowledge, the method of performing least squares smoothing described here is new, but many of the results derived from the method are not. It has not been the author's purpose to present the results with mathematical rigor, but rather to indicate the

flexibility of this technique. For a more complete discussion of least squares smoothing and some interesting variants, the reader is referred to a partial list of references included at the end.[4,5,6,7,8]

The author wishes to thank G. L. Baldwin for a valuable discussion and R. B. Blackman for very helpful advice and comments.

APPENDIX A

*Recursive Summation of Squared Residuals**

In connection with a study of methods to detect errors in the smoothing of data by the technique described in the main text, it was found that the sum of squared deviations from the least squares line of regression could be determined recursively. Insofar as this may be of some interest in certain data processing systems, we will briefly outline the derivation and present the results in this appendix.

The sum of squared deviations from the least squares line of regression (hereafter called squared residuals) is defined by (1), which we repeat here:

$$R_n^2 = \sum_{i=1}^{n} [\mathring{x}_i - \bar{x}_n + (n - i)\bar{u}_n]^2 w_i . \tag{37}$$

The recursion relation may be obtained by replacing all n's by $(n + 1)$'s in (37):

$$R_{n+1}^2 = \sum_{i=1}^{n+1} [\mathring{x}_i - \bar{x}_{n+1} + (n + 1 - i)\bar{u}_{n+1}]^2 w_i , \tag{38}$$

which can be immediately converted to:

$$R_{n+1}^2 = \sum_{i=1}^{n} [\mathring{x}_i - \bar{x}_{n+1} + (n + 1 - i)\bar{u}_{n+1}]^2 w_i + w_{n+1}(\mathring{x}_{n+1} - \bar{x}_{n+1})^2 .$$

By making use of the smoothing and prediction equations, (10), (11), and (15), we can rewrite the last result as

$$\begin{aligned} R_{n+1}^2 = \sum_{i=1}^{n} \big(& [\mathring{x}_i - \bar{x}_n + (n - i)\bar{u}_n] \\ & - \{(\mathring{x}_{n+1} - \hat{x}_{n+1})[\alpha_{n+1} - (n + 1 - i)\beta_{n+1})]\} \big)^2 w_i \\ & + w_{n+1}(\mathring{x}_{n+1} - \bar{x}_{n+1})^2 . \end{aligned} \tag{39}$$

The evaluation of this sum is straightforward but tedious and makes

* See also Ref. 7.

use of the defining relations for the quantities F_n, G_n, H_n, J_n, α_n, and β_n, (3), (12), (13), (14), and the recursion relations, (18). The result may be written in either of two forms:

$$\begin{aligned} R_{n+1}^2 &= R_n^2 + \frac{w_{n+1}}{1 - \alpha_{n+1}} (\mathring{x}_{n+1} - \bar{x}_{n+1})^2 \\ &= R_n^2 + w_{n+1}(1 - \alpha_{n+1})(\mathring{x}_{n+1} - \hat{x}_{n+1})^2. \end{aligned} \tag{40}$$

Thus only one storage slot and a modest amount of computation need be invested to obtain the instantaneous value of the squared residuals.

In order to use this as an error-detecting device, some assumptions regarding the observational error must be made. If the variations in the σ_i^2 are known and compensated for by choosing $w_n = K/\sigma_n^2$, and if the errors are uncorrelated between observations, then the average value of R_n^2 can be shown to be

$$\text{ave } (R_n^2) = K[(\text{number of observations}) - 2]. \tag{41}$$

If the $\sigma_i^2 = \sigma_0^2$ and $w_i = 1$ or 0, then

$$\text{ave } (R_n^2) = \sigma_0^2(F_n - 2) \tag{42}$$

since, for this case, F_n is equal to the number of observations. The fact that the average value is proportional to two less than the number of observations is related to the fact that two degrees of freedom have been used up in determining $\bar{x}_n$ and $\bar{u}_n$. If the variations in σ_i^2 are unknown or too complicated to compensate for, then it is often possible, for a given system, to determine some bounds on the average value of squared residuals.

If we define

$$\bar{R}_n^2 = \frac{R_n^2}{(\text{number of observations}) - 2}$$

then, from above we have

$$\text{ave } (\bar{R}_n^2) = \begin{cases} K & \text{for } w_i = K/\sigma_i^2, \\ \sigma_0^2 & \text{for } w_i = 1 \text{ or } 0. \end{cases} \tag{43}$$

It can be shown from a straightforward application of statistical analysis* that

$$\text{var } (\bar{R}_n^2) = \frac{2K^2}{(\text{number of observations}) - 2} \tag{44}$$

with K^2 replaced by σ_0^4 for the case of $w_i = 1$ or 0.

* See, for example, Ref. 3, p. 103.

If no prior knowledge of the magnitude of the measurement errors is available, the squared residuals can be used to obtain an estimate of the average measurement error if all $w_i = 1$, or of the average *effective* measurement error if an arbitrary weighting sequence is used.

APPENDIX B

Fixed Memory Smoothing[5]

In contrast to the method described in the main text, where the smoothing is designed to include all observations (i.e., variable smoothing time), we can set up a smoothing procedure which fits a least squares line of regression to the latest r observations. The ability to optimize for varying observational errors and missing observations is retained, but we must provide storage for the r observations.

The two quantities to be minimized may be written as

$$\begin{aligned} &\sum_{i=n+1-r}^{n} [\mathring{x}_i - \bar{x}_n + (n - i)\bar{u}_n]^2 w_i , \\ &\sum_{i=n-r}^{n-1} [\mathring{x}_i - \hat{x}_n + (n - i)\hat{u}_n]^2 w_i . \end{aligned} \tag{45}$$

We define the sums F_n, G_n, H_n, and J_n as follows:

$$\begin{aligned} F_n &= \sum_{i=n+1-r}^{n} w_i , \\ G_n &= \sum_{i=n+1-r}^{n} (n - i) w_i , \\ H_n &= \sum_{i=n+1-r}^{n} (n - i)^2 w_i , \\ J_n &= F_n H_n - G_n^2 , \end{aligned} \tag{46}$$

and their recursion relations are

$$\begin{aligned} F_n &= F_{n-1} + w_n - w_{n-r} , \\ G_n &= G_{n-1} + F_{n-1} - r w_{n-r} , \\ H_n &= H_{n-1} + 2G_{n-1} + F_{n-1} - r^2 w_{n-r} , \\ J_n &= J_{n-1} + w_n H_n - w_{n-r}[H_{n-1} - 2(r - 1)G_{n-1} + (r - 1)^2 F_{n-1}]. \end{aligned} \tag{47}$$

The smoothing equations can be shown to be

$$\begin{aligned} \bar{x}_n &= \hat{x}_n + \alpha_n(\mathring{x}_n - \hat{x}_n) + \gamma_n[\mathring{x}_{n-r} - (\hat{x}_n - r\hat{u}_n)], \\ \bar{u}_n &= \hat{u}_n + \beta_n(\mathring{x}_n - \hat{x}_n) + \delta_n[\mathring{x}_{n-r} - (\hat{x}_n - r\hat{u}_n)], \end{aligned} \tag{48}$$

where

$$\begin{aligned} \alpha_n &= \frac{w_n H_n}{J_n}, \qquad \gamma_n = \frac{w_{n-r}(rG_n - H_n)}{J_n}, \\ \beta_n &= \frac{w_n G_n}{J_n}, \qquad \delta_n = \frac{w_{n-r}(rF_n - G_n)}{J_n}, \end{aligned} \tag{49}$$

and the prediction equations are the same as (15).

The quantities $(rG_n - H_n)$ and $(rF_n - G_n)$ may also be written recursively, for if we define

$$A_n = rF_n - G_n,$$

$$B_n = rG_n - H_n,$$

then

$$A_n = A_{n-1} - F_{n-1} + rw_{n-r},$$

$$B_n = B_{n-1} + A_{n-1} - (G_{n-1} + F_{n-1}).$$

The explicit appearance of $\mathring{x}_{n-r}$ in (48) demonstrates the fact that storage must be provided for the last r observations. As in the method described in the main text, the w_i can be chosen completely arbitrarily and if they are chosen equal to $1/\sigma_n^2$, the smoothing is optimized for the varying data error. It may be advantageous to provide storage for the r weighting factors w_i if they are computed from some function. This would remove the necessity for recomputing w_{n-r} when it had already been computed at a previous $n' = n - r$.

In terms of computer operation, the time required to process each new observation is independent of r. Thus significant savings in computation time over other methods (stored coefficients, cascaded simple sums) are achieved only when r is large. However, these other methods smooth data in a predetermined manner and cannot alter the weighting of data to compensate for an arbitrary sequence of observations and misses or optimize for trajectory-dependent measurement errors.

For the case of all $w_i = 1$, the explicit values of all quantities are given at the end of Section II of the main text, with n replaced by r.

APPENDIX C

Estimation of Acceleration

To distinguish between smoothing over observations of linear and parabolic flight, we shall refer to the former as linear smoothing and the latter as quadratic smoothing. For the latter, we want to minimize

$$\sum_{i=1}^{n} \{\mathring{x}_i - [\bar{x}_n - (n-i)\bar{u}_n + (n-i)^2\bar{s}_n]\}^2 w_i , \tag{50}$$

where

$$\bar{u}_n = \bar{v}_n\tau, \qquad \bar{s}_n = \tfrac{1}{2}\bar{a}_n\tau^2.$$

Straightforward differentiation of this expression with respect to the three unknowns yields the normal equations:

$$\begin{aligned} F_n\bar{x}_n - G_n\bar{u}_n + H_n\bar{s}_n &= \sum_{i=1}^{n} \mathring{x}_i w_i , \\ -G_n\bar{x}_n + H_n\bar{u}_n - I_n\bar{s}_n &= -\sum_{i=1}^{n} \mathring{x}_i(n-i)w_i , \\ H_n\bar{x}_n - I_n\bar{u}_n + K_n\bar{s}_n &= \sum_{i=1}^{n} \mathring{x}_i(n-i)^2 w_i , \end{aligned} \tag{51}$$

where F_n , G_n , and H_n are defined by (3) and

$$\begin{aligned} I_n &= \sum_{i=1}^{n} (n-i)^3 w_i , \\ K_n &= \sum_{i=1}^{n} (n-i)^4 w_i . \end{aligned} \tag{52}$$

The recursion relations for I_n and K_n are

$$\begin{aligned} I_n &= I_{n-1} + 3H_{n-1} + 3G_{n-1} + F_{n-1} , \\ K_n &= K_{n-1} + 4I_{n-1} + 6H_{n-1} + 4G_{n-1} + G_{n-1} , \end{aligned}$$

and, as before, J_n is defined as the determinant of the coefficients in the normal equations:

$$J_n = F_nH_nK_n - 2G_nH_nI_n - H_n^3 - F_nI_n^2 - K_nG_n^2. \tag{53}$$

Repeating the procedure for the estimates $\bar{x}_{n-1}$, $\bar{u}_{n-1}$, and $\bar{s}_{n-1}$ and subtracting the resulting equations from the corresponding equations in

(51), we obtain the three-variable equivalent to (9). The solution for quadratic smoothing can be written

$$\begin{aligned} \bar{x}_n &= \hat{x}_n + \alpha_n(\mathring{x}_n - \bar{x}_n), \\ \bar{u}_n &= \hat{u}_n + \beta_n(\mathring{x}_n - \hat{x}_n), \\ \bar{s}_n &= \hat{s}_n + \gamma_n(\mathring{x}_n - \hat{x}_n), \end{aligned} \tag{54}$$

where

$$\begin{aligned} \alpha_n &= w_n \frac{H_n K_n - I_n^2}{J_n}, \\ \beta_n &= w_n \frac{G_n K_n - H_n I_n}{J_n}, \\ \gamma_n &= w_n \frac{G_n I_n - H_n^2}{J_n}. \end{aligned} \tag{55}$$

The auxiliary prediction equations are

$$\begin{aligned} \hat{x}_{n+1} &= \bar{x}_n + \bar{u}_n + \bar{s}_n , \\ \hat{u}_{n+1} &= \bar{u}_n + 2\bar{s}_n , \\ \hat{s}_{n+1} &= \bar{s}_n , \end{aligned} \tag{56}$$

where additional terms may be added to compensate for known "jerk."

If, in addition to the definitions (19), we define

$$s_n = \sum_{i=1}^{n} e_i \mathring{x}_i ,$$

the weighting coefficients c_i , d_i , and e_i applied to the observations to yield smoothed position, velocity and acceleration are:

$$\begin{aligned} c_i &= \frac{1}{J_n}[(H_n K_n - I_n^2) - (G_n K_n - H_n I_n)(n - i) \\ &\quad + (G_n I_n - H_n^2)(n - i)^2] w_i , \\ d_i &= \frac{1}{J_n}[(G_n K_n - H_n I_n) - (F_n K_n - H_n^2)(n - i) \\ &\quad + (F_n I_n - G_n H_n)(n - i)^2] w_i , \\ e_i &= \frac{1}{J_n}[(G_n I_n - H_n^2) - (F_n I_n - G_n H_n)(n - i) \\ &\quad + (F_n H_n - G_n^2)(n - i)^2] w_i . \end{aligned} \tag{57}$$

The variance ratios for smoothed position, velocity, and acceleration (assuming constant observational error) can be evaluated as

$$\begin{aligned}
\sigma_{\bar{x}_n}^2/\sigma_0^2 &= \frac{H_nK_n - I_n^2}{J_n}, \\
\sigma_{\bar{u}_n}^2/\sigma_0^2 &= \frac{F_nK_n - H_n^2}{J_n}, \\
\sigma_{\bar{s}_n}^2/\sigma_0^2 &= \frac{F_nH_n - G_n^2}{J_n}.
\end{aligned} \tag{58}$$

For all $w_i = 1$, these quantities can be explicitly evaluated as functions of n:

$$F_n = n,$$

$$G_n = \frac{n(n-1)}{2},$$

$$H_n = \frac{n(n-1)(2n-1)}{6},$$

$$I_n = \frac{n^2(n-1)^2}{4},$$

$$K_n = \frac{n(n-1)(2n-1)(3n^2-3n-1)}{30},$$

$$J_n = \frac{n^3(n-1)^2(n+1)^2(n-2)(n+2)}{2160},$$

$$\alpha_n = \frac{3(3n^2-3n+2)}{n(n+1)(n+2)},$$

$$\beta_n = \frac{18(2n-1)}{n(n+1)(n+2)},$$

$$\gamma_n = \frac{30}{n(n+1)(n+2)},$$

$$c_i = 3\,\frac{(3n^2-3n+2) - 6(2n-1)(n-i) + 10(n-i)^2}{n(n+1)(n+2)},$$

$$d_i = 6\,\frac{3(2n-1)(n-1)(n-2) - 2(8n-11)(2n-1)(n-i) + 30(n-1)(n-i)^2}{(n-2)(n-1)n(n+1)(n+2)},$$

$$e_i = 30\,\frac{(n-1)(n-2) - 6(n-1)(n-i) + 6(n-i)^2}{(n-2)(n-1)n(n+1)(n+2)},$$

$$\sigma_{\bar{x}_n}^2 = \sigma_0^2\left[\frac{3(3n^2-3n+2)}{n(n+1)(n+2)}\right] = \alpha_n\sigma_0^2,$$

$$\sigma_{\bar{v}_n}^2 = \frac{\sigma_0^2}{\tau^2}\left[\frac{12(8n-11)(2n-1)}{(n-2)(n-1)n(n+1)(n+2)}\right],$$

$$\sigma_{\bar{a}_n}^2 = \frac{4\sigma_0^2}{\tau^4}\left[\frac{180}{(n-2)(n-1)n(n+1)(n+2)}\right].$$

It is interesting to compare the position and velocity variance ratios for linear and quadratic smoothing:

$$\frac{\text{position variance (quadratic)}}{\text{position variance (linear)}} = \frac{3(3n^2-3n+2)}{2(n+2)(2n-1)} \to \frac{9}{4} \text{ for large } n,$$

$$\frac{\text{velocity variance (quadratic)}}{\text{velocity variance (linear)}} = \frac{(8n-11)(2n-1)}{(n-2)(n+2)} \to 16 \text{ for large } n.$$

These ratios indicate the cost in position and velocity accuracy for the inclusion of acceleration estimation.

REFERENCES

1. Shapiro, H. S., Least-Square Smoothing and Prediction of Noisy Linear and Polynomial Data, unpublished memorandum.
2. Kaplan, E. L., Recurrent Prediction Assuming Known Acceleration, unpublished memorandum.
3. Bennett, C. A., and Franklin, N. L., *Statistical Analysis*, John Wiley & Sons, New York, 1954.
4. Howard, W. A., Least Squares Curve Fitting, unpublished memorandum.
5. Blum, M., Fixed Memory Least Squares Filters Using Recursion Methods, I.R.E. Trans., **IT-3,** 1957, p. 178.
6. Blum, M., On the Mean Square Noise Power of an Optimum Linear Discrete Filter Operating on Polynomial Plus White Noise Input, I.R.E. Trans., **IT-3,** 1957, p. 225.
7. Rose, M. E., The Analysis of Angular Correlation and Angular Distribution Data, Phys. Rev., **91,** 1953, p. 610.
8. Blackman, R. B., Discrete Data Smoothing by Cascaded Simple Sums, unpublished memorandum.

A Loss and Phase Set for Measuring Transistor Parameters and Two-Port Networks Between 5 and 250 mc*

By D. LEED and O. KUMMER

(Manuscript received September 14, 1960)

An insertion loss and phase measuring set has been developed for making small-signal measurements on transistors and general two-port networks with maximum inaccuracy of 0.1 db and 0.5 degree over a frequency range from 5 to 250 mc. In order to realize accuracy substantially independent of test frequency, the measurement information is heterodyned to a fixed intermediate frequency, where detection is performed with the aid of adjustable loss and phase-shift standards. Use of a rapid sampling technique to compare the unknown with a high-frequency standard eliminates errors from circuit drifts and also reduces the magnitude of the "instrument-zero line" to a small value. Besides discussing the over-all operation of the new test set, the paper presents the design approaches used in solving problems related to purity of terminations as seen from the unknown, automatic control of beat oscillator frequency, conversion, signal-to-noise, and design of loss and phase standards. Particular attention is given to the features of the set which especially adapt it to the measurement of transistor parameters.

I. INTRODUCTION

In the early stages of development of specific transistor types, compromise performance targets are worked out that are both desirable from a circuit use standpoint and feasible from a fabrication viewpoint. Fairly simple electrical measuring instruments are sufficient to guide the development during this early period. However, past performance has shown that the device designer converges rather rapidly on the agreed-upon targets, and device reproducibility quickly reaches a point where

* This work was supported in part by Task V of Joint Military Services Contract DA-36-039 sc-64618.

more precise knowledge of the device characteristics would permit greater sophistication in circuit applications. At this point, accurate measurements must be made of the frequency characteristics of the device and of the circuits in which it is used, in order to sustain further progress in circuitry development. This paper is concerned with an instrument designed to meet such measurement needs in the frequency range from 5 to 250 mc.

The first portion of the paper reviews the special problems involved in making accurate broadband measurements on transistors. It is noted that the chief limitation stems from the difficulty of providing a known circuit environment around the unit under test. At high frequencies, this problem is aggravated by selecting excessively large or excessively small termination impedances. For this reason, an impedance level of 50 ohms has been chosen for both source and load in the measurement apparatus that is described. The data required for completely characterizing a transistor on a two-port basis are obtained by making four independent insertion loss and phase measurements, with the transistor inserted in four different configurations in relation to the 50-ohm terminations. General two-port networks may be measured as well as transistors.

Section II summarizes the basic measurement method and indicates measurement accuracies and ranges. This is followed by a discussion of the transformations from the measurement parameters to other matrix sets used in design and analysis. This section also lists the principal objectives, other than those related to accuracy, that specifically influenced the measurement set design. Chief among these were the demands for small-signal measurements, for built-in facilities for expediting measurements on transistors, and for a considerable amount of automatic operation to promote ease-of-use and minimize operator error.

Sections III and IV describe the measuring circuit in progressively finer detail, starting from a block diagram description. Questions relating to both system and circuit design are discussed, together with the approaches used in solution.

Section V deals with the aspects of the set which especially adapt it to transistor measurement. This includes the technique for bringing a 50-ohm measurement plane up to the base of the transistor header and the method of biasing the transistor without introducing signal path reflections.

Equipment design considerations are covered in Section VI, with special emphasis on the features for relieving the amount of manual operator labor during long runs of measurements.

Tests made to validate the measurement accuracy are considered in

Section VII. Special self-calibrating procedures are described for absorbing secondary sources of test error.

Measurement results are presented in Section VIII, showing the method of obtaining an h-parameter characterization for a broadband diffused-base transistor.

1.1 *Optimum Two-Port Measurement Parameters for Transistors*

The principal source of inaccuracy in measuring impedances or transmission quantities at high frequencies stems from the difficulty of providing a known circuit environment around the object under test. Unlike the problem of providing adequate loss or phase standards, which may be relieved by heterodyning the measurement information to a constant detection frequency, there is no evasive solution to the termination control question. This problem is not relieved by the introduction of a converter; it must be dealt with directly at the frequency of test signal applied to the unknown.

There are several reasons why the provision of well-defined source and load terminations becomes more difficult in transistor measurements. First of all, the transistors intended for circuit uses in the VHF range are not, at present, coaxially encapsulated. As a result, a "jig" is necessary for making the transition from the noncoaxial basing of the transistor to the usual coaxial geometry of the ports of the measurement set. Since the residual parameters of the jig contribute to the circuit environment around the transistor, it is necessary to produce either a jig whose parasitic elements are insignificant, or one whose parasitic elements are evaluable.

Another unique difficulty in providing known terminations during transistor measurement arises from the need to energize the device with dc in order to make it active. Unavoidably, the circuitry used to connect biasing currents and voltages to the transistor tends also to introduce impedance and transmission vagaries in the signal path.

At low frequencies, ac residuals due to jigs and to dc activation are generally negligible, and this makes it possible to realize a wide gamut of known termination impedances. "Shorts" and "opens" can be closely approximated, so that h, y, or z parameter sets may be measured directly. However, at high frequencies, the range of attainable sources and receiver impedances is limited by the residual parameters of the jig and bias connection circuitry, and by the extreme difficulty of realizing broadband open- and short-circuit impedances.

The measurement set described in the present paper deals with the

high-frequency termination problem by recognizing that attempts to realize the extreme impedances required for direct measurement of h, y, or z parameters would not prove fruitful. Instead, the design effort was aimed at synthesizing a 50-ohm termination level. At this impedance level, stray reactances that would ordinarily make the realization of very large or very small impedances impracticable can be compensated so that they produce only small reflections.

With 50 ohms established as the source and load impedance, a set of four measurements completely characterizes the transistor at each frequency of test. This set of measurements consists of the insertion loss and phase in each direction of transmission through the transistor, and the insertion loss and phase due to bridging each of the transistor ports across the 50-ohm signal path. During the bridging measurements the remote port is terminated in 50 ohms. These four measurements are closely related to the set of finite termination parameters defined by I.R.E. Standards.[1] They may be transformed to any of the other two-port characterization frameworks or to equivalent circuit representations.

1.2 *Interest in Two-Port Measurements*

A measuring instrument which yields sufficient data to characterize active devices completely is equally useful for characterizing general two-port networks. This is an important consideration, since the circuit designer not only needs to characterize devices from a circuit-use standpoint but also must assess the performance of complete circuits using these devices.

Indeed, the need for the most precise measurement ordinarily occurs in the design of "analog-type" communication systems, such as L3 carrier[2] or the TD-2 microwave relay system.[3] These cases demand the highest accuracy because of the opportunity for systematic pile-up of distortion through long chains of tandem-connected apparatus. For this reason, highly accurate laboratory instrumentation is needed for measuring loss, phase, and impedance of linear two-ports, so that effectual steps may be taken to solve the equalization problem during the laboratory phase of the system design.

In the development of regenerative PCM systems, the accuracy called for in measuring frequency characteristics is generally less stringent than that required for analog systems. The relaxation of accuracy tolerances is possible because of the absence of frequency distortion pile-up through successive repeaters. However, it is still desirable to supplement heavily used time-domain measurement procedures with frequency-characteristic measurements on the essentially linear parts of such systems.

II. MEASUREMENT PRINCIPLES AND PERFORMANCE

2.1 *Method of Measurement*

The basic quantities measured are insertion loss and phase shift; the general method of measurement is illustrated in Fig. 1(a).

Vibrating relays s_1 and s_2 sequentially interpose the unknown and a coaxial strap between a 50-ohm source and load. The interchange is made 60 times a second. During the time interval when the path is completed through the strap, the instrument essentially measures the amplitude of the output voltage, $|E_s|$, and stores the measured value. When the path through x is completed, $|E_x|$ is measured and stored. The differ-

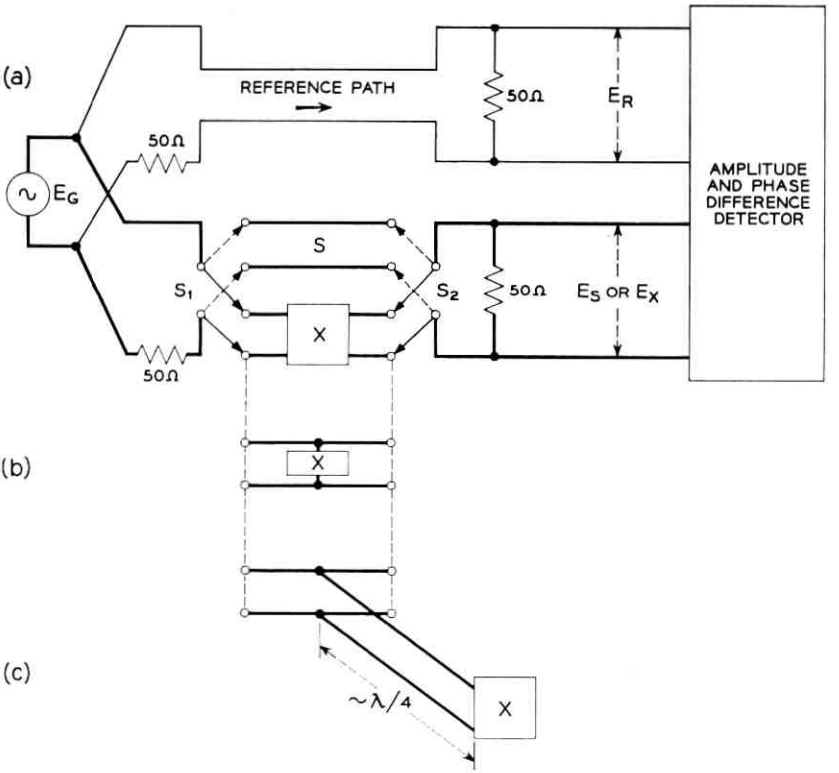


Fig. 1 — (a) Rapid comparison principle of measurement: alternatively interposing X and S between generator and load prevents errors from variations of generator level, detection sensitivity, or shift of detector operating point. (b) Impedance measurement by bridging. (c) Inversion of high impedance by λ/4 line transformer.

ence between the two stored values is read out on a null balanced attenuator standard as the insertion loss,

$$20 \log_{10} \left| \frac{E_S}{E_X} \right|.$$

Insertion phase shift is the difference of phase between E_S and E_X. Since E_S and E_X appear in time sequence, their phase difference may not be obtained directly. Instead, each of these two signals must be phase-compared in succession with the constant output, E_R, from a supplementary reference path. First, the phase difference between E_S and E_R is measured and stored; this is followed by measurement and storage of the phase angle between E_X and E_R. The insertion phase shift is just the difference between the two successive phase measurements, and this difference is read out on a null-balanced calibrated phase shifter.

The comparison of s with x is so rapid that the measurement results are unaffected by slow wanders of source level or by drifts of detector operating point or of gain. This arrangement also obviates error from shifts in source level or detection sensitivity with frequency, since the source and detector are the same for both s and x. To prevent errors in the comparison of the unknown with the standard, the transmission paths through the switch must be well matched to the nominal termination level, the two paths must transmit equally, and the crosstalk from the open to the transmitting path must be small.

Impedance values may be inferred by measuring the insertion loss and phase caused by bridging the unknown impedance across the x path, as suggested in Fig. 1(b). If the measurement yields an insertion loss and phase factor, $e^{\alpha+j\beta}$, where α is the insertion loss in nepers and β is the insertion phase shift in radians, then the impedance, Z, is computable from the relationship

$$W = e^{\alpha+j\beta} = 1 + \frac{1}{2}\left(\frac{50}{Z}\right). \tag{1}$$

A chart technique has been worked out for graphically converting from W to Z, based on the fact that the two are bilinearly related. The bilinear relation simplifies the mapping problem, because circular loci in the W plane transform to circular loci in the Z plane. Chart designs of this type are covered in earlier work.[4]

Impedance measurement sensitivity decays rapidly as $|Z|$ increases; a resistor of 25 ohms causes 6 db loss, but the loss for 100 ohms is only 2 db. The drop in insertion loss decreases measuring accuracy. Consequently, high impedances are initially transformed to lower values by the use of a quarter-wave coaxial line transformer, as shown in Fig. 1(c).

The transformer consists of a 39-foot length of 50-ohm air dielectric coaxial line, and is therefore useful as an impedance inverter only in the vicinity of odd multiples of 5 mc. The need for *a priori* knowledge of the constants of the line may be avoided by the self-calibrating technique of measurement described later in Section 7.3. This technique also removes the necessity for adjusting the signal frequency so that the line is an exact odd multiple of quarter wave length.

2.2 *Measurement Accuracy and Ranges*

Signal Source. 5 to 250 mc in two bands; ±3 per cent scale calibration accuracy. The stability and precision of setting are sufficient for adjusting to specifically desired frequencies with a tolerance exceeding 10 kc, using high-accuracy commercial counters for frequency measurement.

Source and Load Terminations. 50 ohms for transistor measurement; 50 or 75 ohms for coaxially terminated networks.

Insertion Loss. Range: 60 db loss to 30 db gain. Accuracy: ±0.1 db from 30 db gain to 40 db loss; ±0.3 db from 40 db to 60 db loss.

Insertion Phase. Range: 360 degrees. Accuracy: ±0.5 degree from 30 db gain to 40 db loss; ±1.5 degrees from 40 to 60 db loss.

These accuracies apply to the measurement of coaxially terminated networks whose image impedances are well matched to either the 50- or 75-ohm termination levels. Further error results from interaction between the impedances of the unknown and the residual impurity of the terminations. Mistermination errors are estimated in Section 7.2 for the case of passive, reciprocal unknowns.

The listing of return losses in Table I is based on measurements made directly at the test set ports into which the unknown is plugged.

TABLE I

Frequency up to	Return Loss Relative to 50 Ohms, in db	
	Source Termination	Load Termination
Transistor Measurements		
100 mc	>35	>31
200 mc	>30	>25
250 mc	>28	>25
General Two-Port Measurements (50 Ohms)		
100 mc	>35	>35
200 mc	>30	>30
250 mc	>28	>28

These data show that the reflections from the jig and biasing circuitry are quite small, as evidenced by the fact that return losses for transistor measurement are at most 5 db poorer than for coaxially terminated unknowns. The design of the elements that determine the return losses presented during transistor measurement is discussed in Section V.

2.3 *Relationship of Measured Quantities to Other Two-Port Descriptions*

Transistors are characterized by making the following four measurements:

(a) Insertion loss and phase shift in both directions of transmission between 50-ohm impedances.

(b) Insertion loss and phase shift obtained when first one port of the unknown and then the other is bridged across the 50-ohm transmission path. The remote port of the transistor is terminated in 50 ohms during these bridging measurements.

If the index number 1 is assigned to one of the ports of the unknown and 2 to the other, then the two measurements of (a) yield the data $e^{\varphi_{12}}$ and $e^{\varphi_{21}}$, and (b) yields $e^{\varphi_{11}}$ and $e^{\varphi_{22}}$. These four parameters define a matrix set, P, which completely characterizes the device under test:

$$P = \begin{pmatrix} e^{\varphi_{11}} & e^{\varphi_{12}} \\ e^{\varphi_{21}} & e^{\varphi_{22}} \end{pmatrix}.$$

P may be transformed to the finite termination parameter matrix by replacing $e^{\varphi_{11}}$ and $e^{\varphi_{22}}$ with the corresponding impedance values computed from (1) in Section 2.1.

The conversion to the scattering matrix, using the appropriate termination resistance as normalizing number, i.e. 50 or 75 ohms, is relatively direct. Defining the scattering matrix S in the conventional way,

$$S = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix},$$

it is shown in Appendix A that the coefficients of P and S are related such that

$$s_{11} = \frac{3 - 2e^{\varphi_{11}}}{-1 + 2e^{\varphi_{11}}},$$

$$s_{12} = e^{-\varphi_{21}},$$

$$s_{21} = e^{-\varphi_{12}},$$

$$s_{22} = \frac{3 - 2e^{\varphi_{22}}}{-1 + 2e^{\varphi_{22}}}.$$

The relation of the elements of P to those of the h, y, and z matrices is obtainable from previous work.[4]

2.4 *Objectives*

A number of objectives influenced the design of the measurement set, other than those relating specifically to accuracies and ranges of the measured quantities.

The broadest object was to provide a measuring facility capable of completely characterizing either passive or active linear two-ports. This was accomplished by incorporating bridging loss as well as insertion loss measurement features.

Secondly, the instrument was to aid in studying small-signal parameters of transistors and their variation with shift of dc operating point. To make such measurements possible, unknowns are excited very lightly in relation to bias power magnitudes. For example, in measuring between 30 db gain and 19.9 db loss, the power delivered to the input port is always less than -30 dbm. The light drive, when combined with transistor loss, decreases the signal voltage available for detection. Consequently, signal-to-noise questions were dominant in the instrument design.

Finally, it was recognized that a great many measurements would be necessary to adequately characterize an unknown over the full 5 to 250 mc frequency range. For this reason, the design aimed at cutting down, to a reasonable minimum, the amount of decision making, knob manipulation, patching changes, and other operator activities required for obtaining measurement answers. This not only has the effect of speeding up the measurements, but also reduces the chances for human error. A high price was paid in the form of "behind-the-panel" complexity to partially substitute machine logic, machine programming, and mechanisms for operator thought and motor activity.

III. DESIGN OF THE MEASURING SET

3.1 *Introduction*

The development work naturally divided into two parts: the development of an insertion loss and phase measurement set covering the 5 to 250 mc range, and the development of jig, auxiliary fixtures, biasing facilities, and programming arrangements to adapt the basic set to transistor measurement.

In order to obtain measurement accuracy substantially independent of frequency over the $5\frac{1}{2}$-octave range of test signal, it was necessary to

heterodyne the measurement information to a fixed intermediate frequency, where detection was actually performed with the aid of precisely calibrated loss and phase shift standards. This not only relieved the problem of standards design, but also provided the opportunity to reduce thermal and tube noise by use of a narrow intermediate frequency bandwidth. The improvement in signal-to-noise ratio was an important factor because of the low level of transistor excitation. On the debit side, the narrow IF band made it necessary to control the frequency of the beat oscillator automatically, because the required precision in the setting of the intermediate frequency could not reasonably be obtained by manual tuning.

Besides possible error due to mistermination, the test set is subject to errors from miscalibration of standards, crosstalk, and noise. The contribution from crosstalk and noise is closely tied up with the heterodyne aspect of the instrument design, and it is of interest to examine these error sources briefly, before proceeding with further details.

The heterodyne outline of the set is shown in Fig. 2 stripped of the rapid comparison and null-balancing features. It will be noted that a crosstalk path exists for spurious transmission of signal frequency, F, from the reference path to the unknown path via the common beat oscillator connection, and vice versa. The damage done by this unwanted

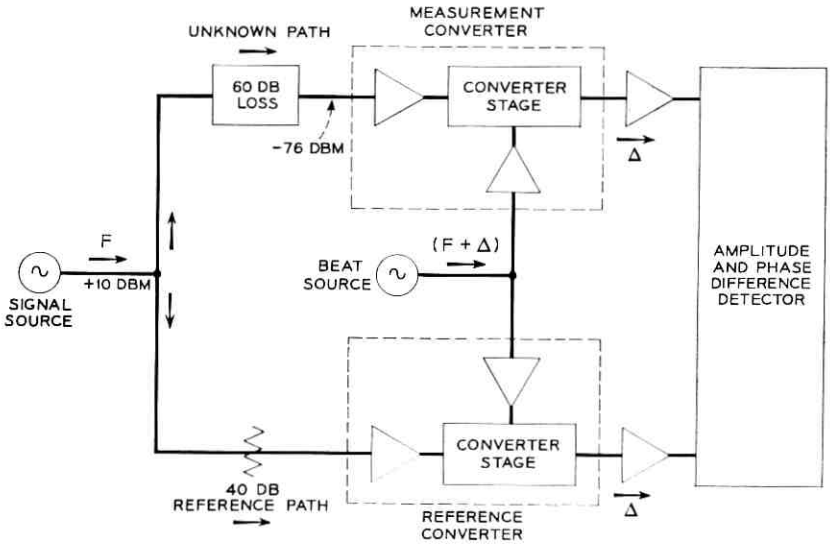


Fig. 2 — Heterodyne framework of test set at 60-db loss level.

transmission is a function of the unknown's loss. In the present set, the input level to the measurement converter is −76 dbm when measuring at the maximum loss level of 60 db. As a result, any stray component of frequency F, equivalent in its effect to a signal of −116 dbm at the measurement converter input, could lead to maximum errors of 0.1 db or 0.5 degree, depending on the phase of the spurious signal. In addition to crosstalk, leak from the signal oscillator to low-level points in the signal paths is an equally potent source of error.

These potential sources of inaccuracy were controlled in a number of ways. First of all, buffer amplifiers were introduced in the signal and beat frequency channels preceding the conversion stages in each converter. All RF and IF circuitry was carefully shielded to guard against radiation and air path couplings. It was also helpful to reduce the maximum possible level difference between signal inputs to the two converters by adding a 40-db loss pad in the reference path. The object of these efforts was to keep errors from crosstalk and pick-up below 0.1 db when measuring 60-db loss.

Similarly, a 0.1-db limit was established on error due to noise at 60-db loss level. Study showed that this objective could be met by relatively straightforward approaches to converter, IF amplifier, and detector design. In fact, with a 20 kc IF bandwidth and simple linear detection, a converter noise figure of 30 db proved tolerable, The relatively lenient demand on noise figure made it possible to use a vacuum tube conversion stage, operating in a square law mode, with the attendant low power requirement on beat oscillator drive. Moreover, the modest noise figure could be realized without introducing gain in the buffer circuitry preceding the conversion stages. The design of the over-all converter is covered in Section 4.3.

The noise created by IF circuitry is small compared to the contribution of the converter.

3.1.1 *Loss Measurement*

The measurement system is shown in block form in Fig. 3 for the case of 50-ohm insertion loss and phase measurements. The basic source is a signal oscillator, tuning from 5 to 250 mc. It delivers energy through a 50-ohm transmission path to a pair of vibrating relays that sequentially interpose the unknown and a standard between the signal source and measurement converter. A pair of solenoid-operated coaxial switches allows any one of three unknown paths to be selected, depending upon whether 50-ohm, 75-ohm or transistor measurements are to be made.

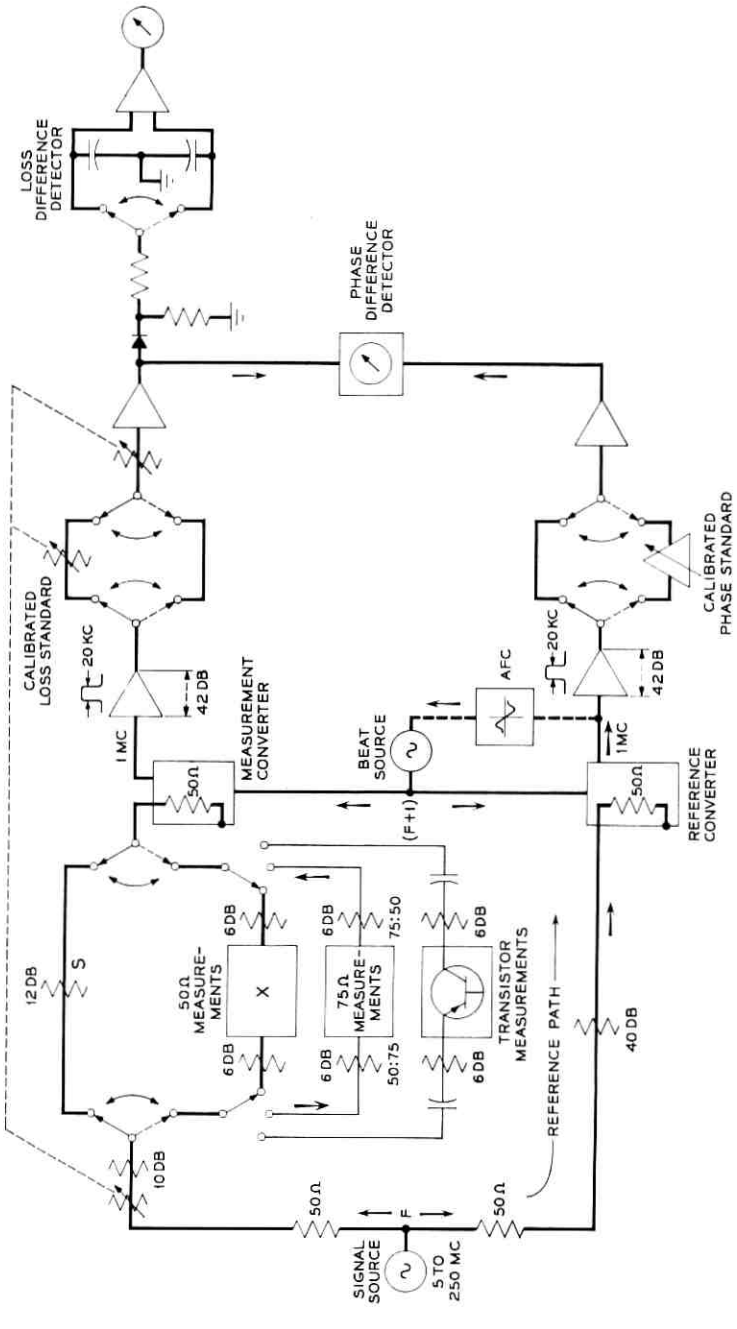


Fig. 3 — Block diagram of measurement set. Parameters of unknown are heterodyned to a constant detection frequency, where their magnitudes are read out on null-balanced, calibrated loss and phase standards.

To reduce the frequency of taking "zeroes," the electrical lengths of the three unknown paths have been adjusted to equal the length of the standard path. Over the full 250 mc frequency range, the instrument "zero-line" is less than 0.1 db and 2 degrees for all three measurement modes.

An auxiliary oscillator, operating always one megacycle higher in frequency than the signal oscillator, beats the measurement information down to a fixed 1 mc intermediate frequency. In the interest of obtaining maximum S/N ratio at the detection point, the IF bandwidth was made as narrow as possible without, at the same time, imposing unreasonably severe tolerances on the settability or frequency stability of the continuously tuned source and beat oscillators. A bandwidth of 20 kc was chosen; the band shaping is introduced by the IF amplifier following the converter. After passing through the amplifier, the signal proceeds through another vibrating relay path and ends up finally at an amplitude-sensing detector.

The relays provide the means for rapidly comparing the transmission of the unknown and standard paths. All moving contacts are synchronized at a 60-cycle rate. When the upper paths are closed, as indicated in the figure, the circuit is completed through the standard channel at both test-signal and intermediate frequencies. The detector then measures the level it receives and stores the measured value in the form of a voltage across a capacitor. When the lower path is completed, the unknown is energized, and the detector measures and stores the transmission through the unknown path on a second capacitor. Using the indication of the difference between the two items of stored data as presented on the magnitude meter, the operator adjusts the loss standard for a null. At the null point, the loss of the unknown must exactly equal the loss in the calibrated attenuation standard, provided the converter accurately transposes changes of level from the signal to the intermediate frequency. The converters are so lightly excited that the nonlinearity errors are less than 0.05 db over the full range of loss measurement.

In order to minimize the dynamic range over which the loss detector must operate, common-path attenuation is ganged to the loss standard in such a way that the sum of the common-path attenuation plus that of the standard is approximately constant. Consequently, any variation of detector input level with change of test frequency occurring at the point of null balance must be due only to the frequency characteristic of conversion loss. The conversion loss does increase about 8 db between 50 and 250 mc, but this droop is prevented from introducing a sensitivity variation by an AGC circuit located in the detector.

In measuring at the higher losses, a greater portion of the common-path attenuation is assigned to the attenuator section located at intermediate frequency. Although this does have the somewhat undesirable effect of increasing the signal drive on transistors when measuring at high loss, it prevents the input level to the converter from dropping dangerously close to noise. The way in which the pattern of attenuation is worked out assures sufficient signal for an S/N ratio* of 22 db at the detection point during 60-db loss measurements at 5 mc. At 250 mc, the S/N ratio is only 14 db when measuring 60-db loss, but, even under these circumstances, the error contribution due to noise is less than 0.1 db. These matters are discussed more fully in Section 4.2.1.

Increasing the level at the unknown in proportion to its loss has the further desirable effect of reducing the dynamic range of the signal applied to the converter. The total transmission measurement range of 90 db (60 db loss to 30 db gain) subjects the converter to an input level change of only 60 db.

When making gain measurements, the loss standard is inserted in series with the unknown, by transposing it from the s to the x channel. The necessary changes in the common attenuation ganging are made automatically at the same time.

3.1.2 *Phase Measurement*

The technique of phase measurement is analogous to that of loss measurement. The output from the reference converter provides a 1-mc phase-reference signal. During the period when the upper path is completed through all the relays, the phase detector measures and stores the value of the phase difference between the outputs of the standard and reference paths. When the relays change state, the detector makes another measurement of the phase difference between the output of the unknown path and the output of the reference path as modified by the calibrated phase standard. The difference between these two stored measurements, which is displayed continuously on the phase-indicating meter, is adjusted to zero by operating the phase standard for a null. At the null point, the phase shift through the standard exactly offsets the high-frequency phase shift through the unknown.

3.2 *Attributes of the Measurement System*

The measurement circuit combines features of rapid comparison and null balancing of standards with heterodyne detection.

* S/N ratio, as used in this paper, refers to the quotient of rms carrier voltage to rms noise voltage in the 20-kc IF bandwidth.

Use of rapid comparison and null balancing makes the measured data depend solely on the difference between the properties of the unknown and standard. Drift of source level, or gain and phase drift in the IF amplifiers and detectors cause no error, as long as the drift is small over the $\frac{1}{60}$ second interval required to complete the comparison between the unknown and standard. Also, the rapid comparison feature greatly relieves requirements on tracking of conversion phase and loss between the two converters, since the converters, too, are common to the channels being compared.

The heterodyne technique has several conspicuous advantages. First of all, locating the loss and phase standards at the intermediate frequency avoids the problem of developing broadband standards. It would be especially difficult to design and construct a variable phase standard with an accuracy of 0.2 degree that had a frequency-insensitive calibration between 5 and 250 mc.

Secondly, noise power is reduced before detection by the narrow band IF amplifiers; so error of measurement due to system noise is reduced. This question is dealt with quantitatively in Section 4.2.1.

Furthermore, the heterodyning permits loss of the unknown to be made up by single-frequency gain at the intermediate frequency. This is advantageous, since a large amount of relatively flat broadband gain is difficult to provide.

Finally, the heterodyne system introduces the advantage of selective detection. It provides immunity from errors due to harmonics and stray signals in the output of the unknown.

IV. SUBSYSTEMS OF THE MEASUREMENT SET

4.1 *Automatic Frequency Control of Beat Oscillator*

The heterodyne technique of measurement introduces the need for a tunable beat oscillator to translate measurement information to the 1 mc IF. The starting point for the design of the beat source involves the requirements on settability of its frequency and this, in turn, is based on the allowed frequency slip of the IF from the nominal 1 mc. Two principal factors limit the permissible deviation: the IF bandwidth (20 kc) and the sensitivity of the calibration of the phase standard to frequency shifts (0.1 degree per kilocycle deviation from 1 mc). Since the magnitude of the calibration error due to frequency shift is cyclic with phase-shifter angle, it cannot be eliminated with a simple phase-slope network.

The contribution of the IF amplifiers must also be considered. It was not possible to eliminate completely the residue of mistracking between

the phase slopes of the two amplifiers, and what is left amounts to approximately one-half degree per kilocycle. Shifts of the IF occurring in less than $\frac{1}{60}$ second could therefore cause jitter of the phase indication.

Taking all of these factors into consideration, it may be shown that the difference between beat and signal frequencies must be set with a precision of one kilocycle. To meet this requirement at 250 mc, the frequency of the beat oscillator must be settable to one part in 250,000. The severity of this requirement, combined with the obvious inconvenience of having to tune the beat oscillator manually to proper frequency every time the signal oscillator is changed, made it a virtual necessity to control automatically the frequency of the beat oscillator from a sampling of its frequency difference with respect to the signal oscillator.

A discriminator, shown in Fig. 3, centered at 1 mc senses the error of intermediate frequency and delivers an error voltage that actuates two modes of automatic frequency control. The first of these is an electromechanical servo which achieves a coarse correction by motor-tuning the frequency of the beat oscillator for minimum IF error. While the motor control is able to keep the beat frequency approximately in step with the signal frequency over the full 250 mc range, it does allow a residual error because of backlash in gearing. Consequently, the mechanical control is supplemented by an all-electronic frequency control which is very narrow in its range but very crisp in its action, and is capable of eliminating the residue of error due to backlash in the mechanical loop.

The automatic control system is shown in some greater detail in Fig. 4. The source of beat signal is a General Radio Company unit oscillator modified for electronic and motor tuning. The full frequency range is covered in two bands. A conventional Foster Seeley discriminator with a sensitivity of 0.2 volt/kc provides the error voltage. After 40 db of direct coupled gain, the error voltage exerts an electronic correction by controlling the dc operating point of an inverse-biased silicon diode network connected across the oscillator tank. The biased diode was selected as the voltage-sensitive reactance because its comparatively low base capacitance permitted it to be most readily integrated into the existing oscillator. Only a small range of electronic control was required, because of the wide tuning capabilities of the parallel-acting servo control.

The prime mover in the servo loop is a two-phase induction motor coupled to the tuning shaft of the oscillator through intermediate gearing; ac error information for driving the motor is obtained from a conventional servo modulator fed by the dc error voltage.

The electronic loop exerts a stabilizing action on the mechanical loop. In fact, the design gain of the mechanical loop is great enough to cause self-oscillation if the electronic loop is disabled. Hence, the electronic

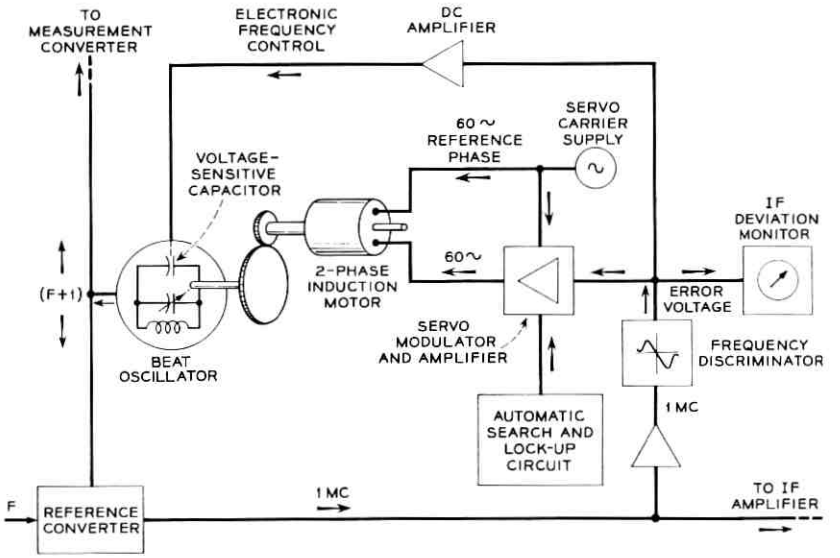


Fig. 4 — Automatic frequency control for beat oscillator, combining servo loop of wide tuning range but limited response rate with narrow-range, fast-acting electronic correction.

loop not only absorbs the remnant of error due to backlash in gearing, but also, by virtue of its greater response rate, introduces the type of stabilization of the mechanical loop that would normally be realized by the use of phase-lead corrective networks.

Neither the electronic nor the mechanical loop gain is constant over the 5 to 250 mc range of the beat source. In the case of the mechanical control, the variation arises because the frequency of the beat oscillator varies logarithmically with the angle of the driving shaft. This enhances the sensitivity of the motor tuning at the higher frequencies of operation. However, the sensitivity of the electronic tuning also intrinsically increases with frequency. Consequently, the electronic loop is capable of providing effective "dynamic braking" over the full 5 to 250 mc frequency range.

An automatic scan circuit causes the beat oscillator to hunt for lock-up when the instrument is first turned on, or whenever synchronization is lost.

4.2 *Sampling System and Detection*

The block diagram of Fig. 3 and the accompanying description in Section III are intended to clarify the basic principles of the measuring

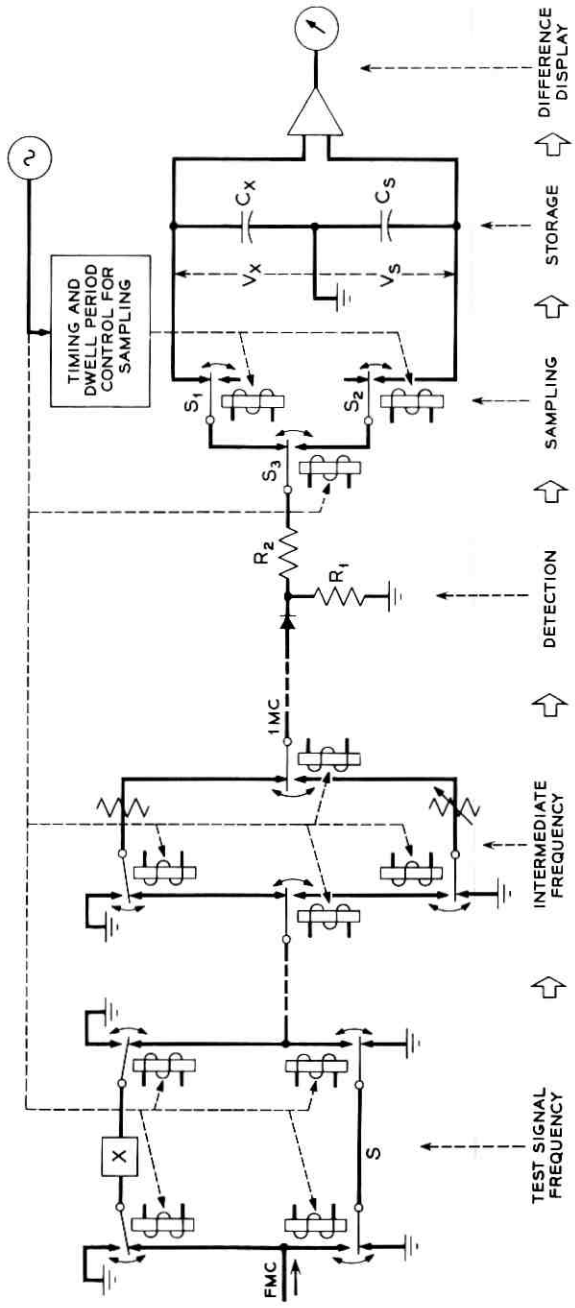


Fig. 5 — Details of rapid-comparison switching in loss measurement.

system. Certain of the significant details of the rapid comparison switching feature as applied to loss measurements will now be taken up. The situation with regard to phase measurements is strictly analogous.

One of the first realities that had to be faced was capacitance between open and transmitting paths in the comparison switches. Even though this capacitance was only 0.3 micromicrofarad in the actual relays used, the crosstalk from this cause could produce large errors when high losses were measured. For this reason, a second double-pole double-throw switch configuration was added at test-signal frequency for the purpose of grounding the open path at both its input and output ends. The complete switching complex is illustrated in Fig. 5.

It is clear that a transmission error results in the comparison of x with s if the contact resistances of the switches are different in their two states of dwell. The problem of contact resistance symmetry is most severe in the case of the switch array operating at test-signal frequency between 50-ohm impedances. In 50-ohm transmission circuits, a resistance asymmetry of 0.5 ohm in just one of the relays would produce an error of 0.1 db. It would be quite easy to obtain far smaller resistance asymmetries than 0.5 ohm in a nonvibratory relay designed for use in low-frequency circuits. But in a relay intended for high-speed repetitive operation both contact area and contact pressure are necessarily limited. With these factors in mind, and considering the extent to which the situation is aggravated by the 250 mc operation, an electromagnetically driven wetted-mercury contact relay element was selected to perform the basic switching function. The resistance asymmetry of this element is estimated to be less than 0.05 ohm at 250 mc. It takes approximately 1 millisecond for the relay to change state. The type of relay element used, in which the mercury is confined by encapsulating the entire relay in a sealed glass envelope, has been described previously.[5]

Coaxial transmission paths are built up by putting the relays inside concentric brass cavities. This serves to minimize the discontinuity when the relays are inserted in 50-ohm coaxial circuits. The cavities are constructed in the form of two mating pieces, which assemble around the relay capsules, as shown in Fig. 6. Each cavity block holds a pair of relays; one block takes care of input switching, the other handles the output. Fig. 6 shows only signal paths through the switch; the coils for driving the movable contacts were intentionally omitted, as well as the shields which prevent coupling between the driving winding and the signal circuits.

An effort was made to maintain 50-ohm geometry by dimensioning the cavity so that the highest possible return losses were realized looking

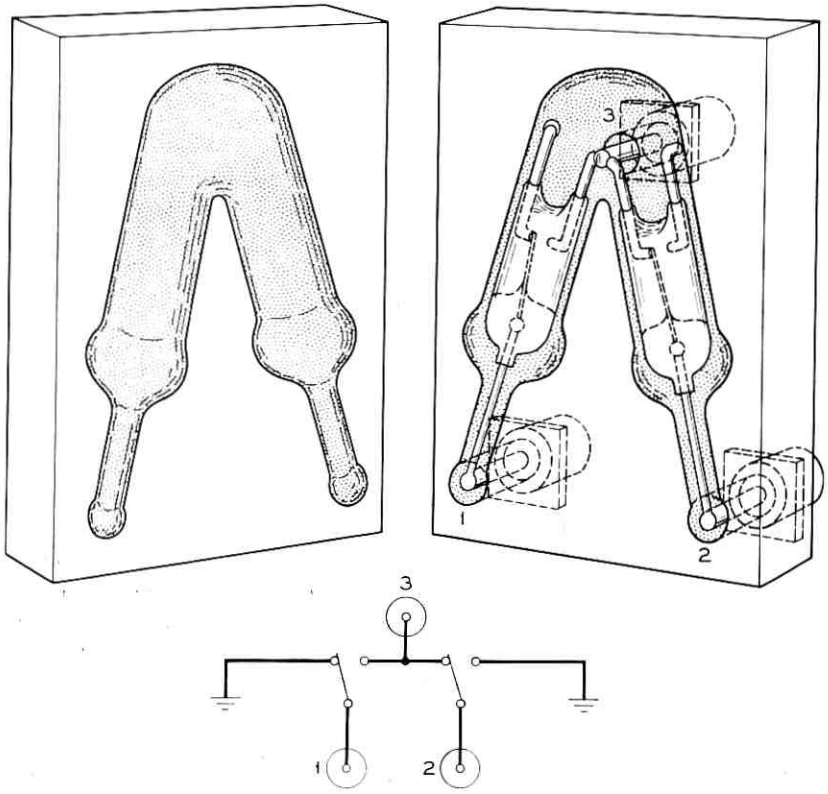


Fig. 6 — 5 to 250 mc comparison switch unit. Glass-encapsulated mercury relays are enclosed within coaxial cavities to provide 50-ohm transmission paths.

into either the 1–3 or 2–3 transmission paths with a match in port 3. Return loss exceeding 35 db was obtained up to 100 mc; between 100 and 250 mc there was a drop to 28 db.

The grounding leads seen in Fig. 6 have sufficiently low impedance to hold crosstalk between ports 1 and 2 to a level of −50 db at 250 mc with a 50-ohm termination inserted in port 3. Thus, the error from crosstalk is less than 0.1 db even when 60-db loss is measured, since switching is done at both the input and output of the unknown.

The design of the comparison switches for use at the 1 mc intermediate frequency posed only minor crosstalk and impedance match problems. Mercury relays were used here, also.

Amplitude detection is accomplished with the linear rectifier shown in

Fig. 5. The system of switches (s_1, s_2, s_3) located just beyond the rectifier is instrumental in charging the storage capacitors, c_s and c_x, to voltages proportional to x and s path transmissions. The charging path is through resistor R_2; the time constant of R_2C_s and R_2C_x is in the order of 2 seconds. In connecting c_s and c_x during the charging intervals, it is necessary to allow for the physical impossibility of perfectly synchronizing all the relays in the measuring set with respect to instant of contact transfer and uniformity of dwell time. Moreover, short-term transients are initiated at the change of state from x to s and vice versa. For these reasons, a specially timed pair of sampling relays, s_1 and s_2, is provided to close the charging path through R_2 a short time after the instants of contact transfer in the previous relays.

4.2.1 *Signal-to-Noise Factors*

The points of the measuring circuit at which random noise is sensed are the inputs to the loss and phase difference detectors in the block diagram of Fig. 3. The noise at these points is essentially confined to a 20 kc band by the IF selectivity. There are two disturbing effects due to the noise:

(a) It contributes error by creating a small dc component in the output of the detectors. The amount of this component will be different for the x and s sampling.

(b) It gives rise to fluctuations of the detector outputs. The frequency components of this fluctuation lying within the display bandwidth of the instrument, i.e., within the bandwidth of circuitry following the detectors, contribute jitter on the indicating meters. This affects resolving power.

The present set uses a linear rectifier for the detection of loss, and a "sum-and-difference" type of phase detector[6] constructed from transformers and linear rectifiers.

The guiding object in setting up level patterns was to have enough signal at the detector inputs to keep errors due to noise less than 0.1 db and 0.5 degree in 60-db loss measurements at 250 mc. Rather than operate at S/N levels significantly higher than necessary to meet this criterion, any extra margin was traded off in the form of lighter signal excitation of unknowns and extra impedance masking with loss pads.

For the linear detector used, it may be shown that the dc offset due to noise leads to an error of 0.1 db at an input S/N ratio of approximately 14 db.[7] This fact, considered together with converter noise figure and impedance masking requirements, resulted in the following allocation

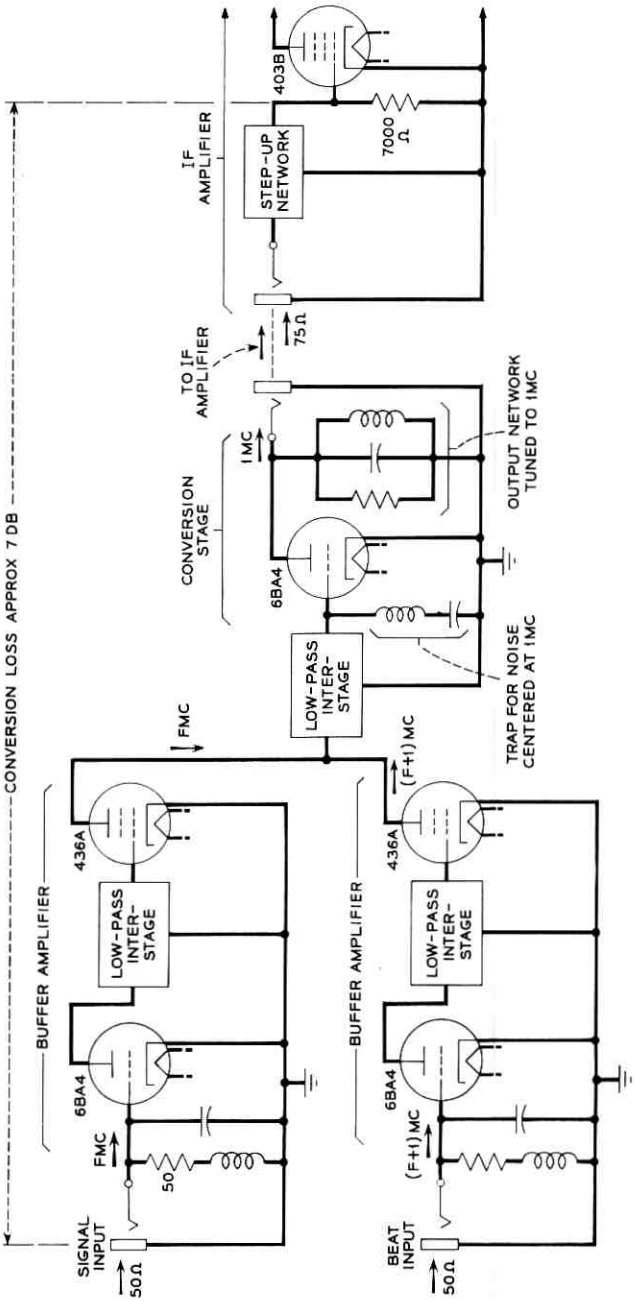


Fig. 7 — Converter circuit.

of signal power to unknowns as a function of loss level:

Loss Level	*Drive to Matched Unknown (dbm)*
0	−40
10	−30
20	−20
30	−10
40	−10
50	−10

As already noted in Section 4.1.1, the power to the unknown is automatically varied by ganging the IF loss standard to common-path attenuation inserted at the test signal frequency.

The contribution of input noise to mean voltage at the output of the sum-and-difference phase detector is dependent upon the phase difference between the two input carriers. When the carriers are 90 degrees out of phase, there is no dc offset due to noise. Consequently, it is most advantageous to operate at quadrature. However, because of the residue of conversion-phase mistracking between the two converters, it was not possible to establish a 90-degree operating point over the full 250 mc span of test-signal frequency. At null balance, the maximum variation from quadrature is about 30 degrees, and this leads to a shift in dc output corresponding to 0.3 degree at the lowest input S/N ratio of 14 db.

4.3 *Conversion*

The measurement converter must faithfully transpose amplitude and phase data from the test-signal frequency to the 1-mc intermediate frequency. It must, in addition, have reasonable noise figure and be well matched to 50 ohms at the signal input. Both converters must preserve high isolation between their signal and beat frequency inputs.

The converter design which grew out of these considerations is shown schematically in Fig. 7. Buffer stages provide the requisite high loss between signal and beat frequency inputs. Working back to either input from the point of joining at the grid of the converter tube, the reverse loss is about 60 db at 250 mc. This figure is set by tube and stray capacity.

To preserve the accuracies cited in Section 2.2, it was necessary to compensate the grid circuits of the input amplifiers so that a relatively pure 50-ohm impedance was presented to the driving source. This was mandatory at the signal input, and was also done at the beat-frequency

input. The choice of a 6BK4 for the input stage relieved the problem of impedance control. Because this tube, with its low cathode lead inductance, imposes less electronic loading on the driving circuit than does a conventionally based tube. By introducing a small amount of inductance in series with the 50-ohm grid resistor, it was possible to absorb the predominately capacitative grid-cathode loading. This method of compensation yielded an input return loss of 30 db up to 250 mc. Actually, it was unnecessary to add external inductance. The pig-tail leads of the grid resistor were trimmed to provide the required reactance.

All interstage networks are conventional. The gain from either input to the grid of the converter tube is approximately 0 db at 5 mc and drops 4 db at 250 mc. The 436A stage, operated at a G_m of 30,000 micromhos, is helpful in overcoming the loss of the first tube. A resonant trap at the converter grid prevents the transmission of noise power centered at one megacycle, thereby improving over-all noise figure. This trap is vital because noise centered at one megacycle would actually be amplified by a single-ended converter.

The level of beat-frequency voltage at the converter grid is such that the converter stage operates in a square law mode, rather than as a switch-type modulator. Signal-frequency excitation at the converter grid never exceeds 0.03 volt, with the result that the nonlinearity error, even for the largest signals, is negligible. The conversion loss, as measured from signal input jack to the grid of the first stage of the IF amplifier, varies from 3.5 db at 5 mc to 12 db at 250 mc. At 5 mc, the noise figure is approximately 28 db, decaying to 36 db at 250 mc. The principal factor determining the noise performance is the relatively high ratio of amplification to conversion gain in the converter stage. With the noise figure data just quoted, it was possible to meet both the accuracy and measurement range objectives.

The output network of the conversion stage and the input network of the IF amplifier are designed to provide specific impedance and transmission characteristics. First of all, the impedance presented to the plate at the intermediate frequency is made small compared with the tube's dynamic resistance, to assure linearity when it acts as a converter. Secondly, to avoid plate remodulation, the output network presents a short circuit to the modulating carriers. And finally, leak of the carriers to the grid of the IF amplifier is small enough so that the spurious 1-mc component created by modulation in the first IF stage is negligible.

4.4 *IF Amplification*

The frequency shaping of the IF band is provided by 42 db gain amplifiers of conventional design following the converters in the over-all

block diagram of Fig. 3. The variation in amplifier gain is less than 0.1 db over a 2-kc band centered at 1 mc; the bandwidth between 3-db points is 20 kc. Two 403B (6AK5) vacuum tube stages provide the necessary gain. Input and output impedances are approximately 75 ohms. A schematic of the amplifier is given in Fig. 8.

In addition to the gain shape requirements, a tolerance was imposed on phase-slope tracking between the two amplifiers. This was necessary to avoid jitter of the phase indication due to incidental, low-order flutters of the intermediate frequency not removed by the AFC. Over a frequency shift of ±500 cycles, centered at 1 mc, the difference of insertion phase shift between the two amplifiers is less than 0.5 degree.

4.5 *Loss Standard*

The loss standard is a precisely calibrated attenuator operating at IF. At null balance, its loss may be equated to loss or gain of the high-frequency unknown. In addition to high calibration accuracy and low aging, the standard must have one other extremely important property—its phase shift must be independent of loss setting. This last attribute is very necessary, since the phase-difference detector cannot distinguish phase change of the loss standard from phase change of the unknown.

The standard is built up of four variable resistive attenuators connected in series. The impedance level is 75 ohms. The standard has discrete calibrated steps of 10, 1, and 0.1 db, and a continuously variable vernier covering a range of 0.2 db. The total range covered is 61.1 db.

Carbon film resistors are used throughout, in order to meet the constant-phase requirement with minimum effort. The resistors are of good

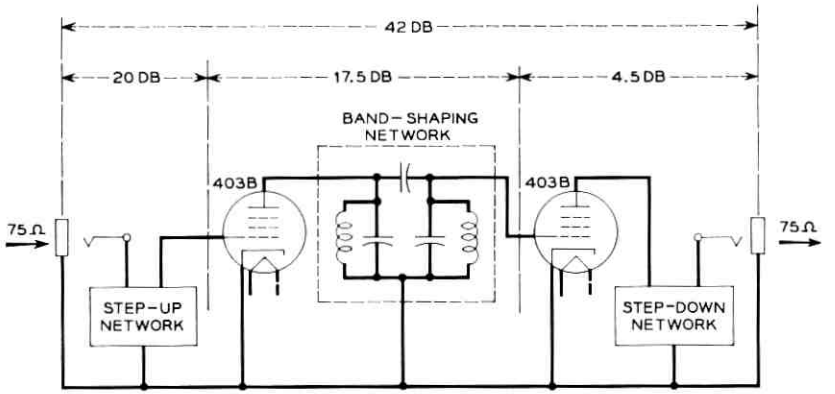


Fig. 8 — Schematic of IF amplifier.

quality and have high long-term stability, but the initial accuracy is only to within 2 per cent. Resistance errors of this order are capable of causing calibration errors of several tenths of a db on the "tens" db decade, where the attenuation sensitivity to resistor inaccuracy is greatest.

To avoid possible error due to initial tolerance of the resistors, the standard is calibrated precisely at dc. With the aid of resistive computing circuitry ganged to the attenuator decades, the calibration error at each setting of the standard is then projected as a correction factor on an auxiliary meter. This arrangement endows the loss standard with a net accuracy of 0.03 db. The long-term permanence of attenuators of this type has been measured, and it has been found that the calibration drift is less than 0.05 db in several years. Temperature coefficient problems are minor since the measurement set is normally operated in a temperature-controlled environment.

4.6 *Phase-Shift Standard*

At null balance, the phase-shift standard is direct reading in terms of the unknown's insertion phase shift. It must have a range of 360 degrees, and a calibration accuracy to 0.2 degree. Since the phase standard does not affect the loss difference detector, its gain may be permitted to vary slightly with phase-shift angle.

The heart of the phase standard is a continuously variable four-quadrant sine condenser of high quality and permanence.[8] It has two linearly subdivided scales: a coarse 0 to 360 degree calibration on a cylinder connected directly to the rotor shaft of the condenser, and a 0 to 10 degree vernier dial connected to the rotor shaft through reduction gearing. Both of these dials are coupled to their respective shafts by friction clutches, so they may be arbitrarily set. This allows the operator to set up a "phantom zero," by slipping the dials to indicate zero after null-balancing the standard. A considerable amount of arithmetic is saved during relative insertion phase measurements when this technique is used to establish a dummy "zero" at the reference frequency.

The difference of insertion phase shift between two different test signal frequencies is read out as a difference between two dial settings. Therefore, an error arises if the actual difference of phase shift introduced by the standard is not precisely equal to the nominal phase difference read from the calibrated dials. Owing to the linearity error of the sine condenser used, the discrepancy between actual phase change and indicated phase change could range up to 2 degrees.

The imperfection of the sine condenser is absorbed by transcribing its

error curve to the periphery of a cam driven by the rotor shaft. A follower, contacting the cam periphery, then automatically adjusts an incremental phase shifter by the amount necessary to reduce the scale error of the standard to less than ± 0.2 degree. Fig. 9 illustrates the principle of the correction.

V. TRANSISTOR MEASUREMENT

5.1 *Coaxial Jig and Fixtures*

The first problem to be solved in making broadband transistor measurements has to do with providing a suitable transition from the round coaxial geometry of the test set ports to the pig-tail lead geometry of the transistor. The object is to bring the 50-ohm measurement plane right up to the base of the transistor header. As suggested in Fig. 10, this may be done by forming short sections of 50-ohm transmission line right under the header, using the active terminal wires of the transistor as inner conductors.

Fig. 11 shows the actual jig worked out using this principle. Each of the two cylindrical holes in the top of the jig forms a 50-ohm transmission line with one of the terminal wires of the transistor under test. The rudimentary coaxial line runs for about one-half inch before it merges with a short section of 50-ohm strip line consisting of a narrow rectangular conductor between ground planes. A 50-ohm type N coaxial con-

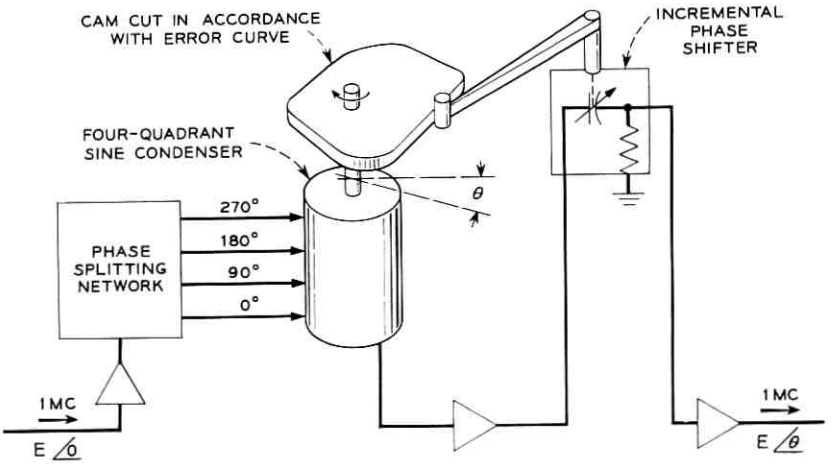


Fig. 9 — Block diagram of phase standard, showing how nonlinearity error of the sine condenser is cancelled by introducing phase corrections with an incremental phase shifter actuated from the error curve.

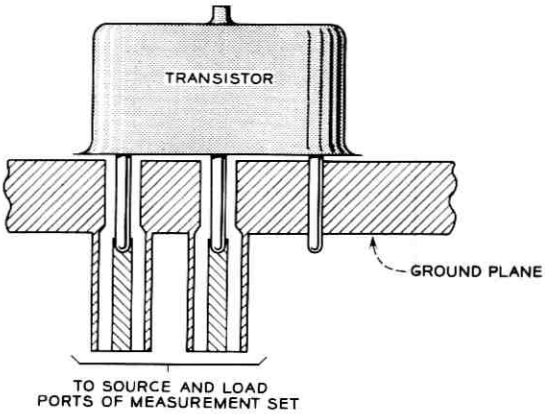


Fig. 10 — Principle of extending 50-ohm geometry to base of header by using transistor lead wires as inner conductors of 50-ohm transmission lines.

nector, inserted in the base of the jig, completes the transmission path. In spite of the discontinuities from joining of geometrically dissimilar lines, measurements indicate that the path through the jig introduces a return loss of 34 db at 250 mc.

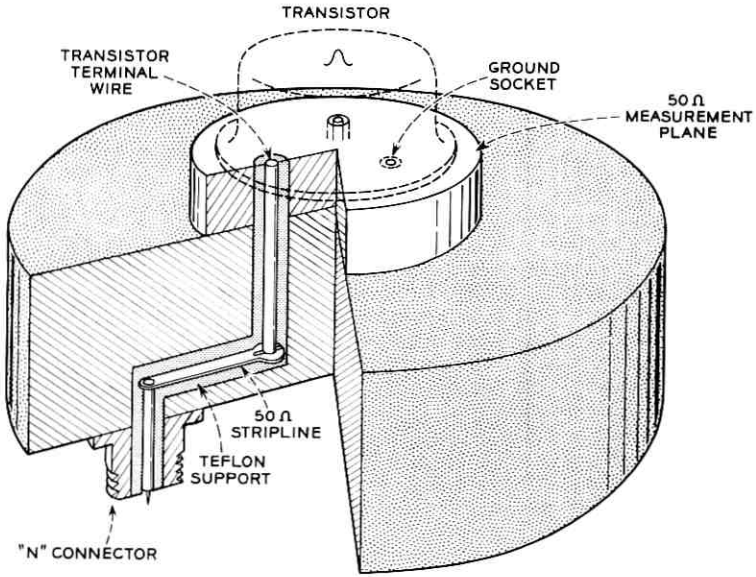


Fig. 11 — Sectional view through actual transistor test jig, showing how 50-ohm transmission geometry is preserved.

The next step is to introduce the biasing currents and voltages to make the transistor active without adding impedance discontinuities. This is done with the aid of the specially designed coaxial pads shown in Fig. 12. The shunt arms of these pads are well by-passed to ground for signal frequencies between 5 and 250 mc, but they are conductively isolated from ground. It is therefore possible to introduce dc biasing and monitoring facilities at the by-passed nodes of the shunt arms without impairing the transmission purity of the signal path. Identical biasing arrangements are provided for both input and output electrodes.

The pad is composed of the elements shown in Fig. 13. A rod resistor forms the series arm, and disc resistors are used for the shunt arms. The by-pass capacitors are physically small but large enough in capacitance value to provide negligible impedance down to 5 mc. All elements are assembled in a coaxial structure. The cavities in the coaxial housing are machined to the optimum diameters for minimum reflection from the terminated pad. With this type of structure, it was possible to obtain return losses of 44 db at 125 mc and 36 db at 250 mc.

This arrangement would have doubtful value unless the activating currents from the bias supplies were confined to the transistor. None of the energizing current may be allowed to flow back into the generator impedance or into the load impedance in an indeterminate way. For this reason, blocking capacitors are used to isolate the test set from the bias

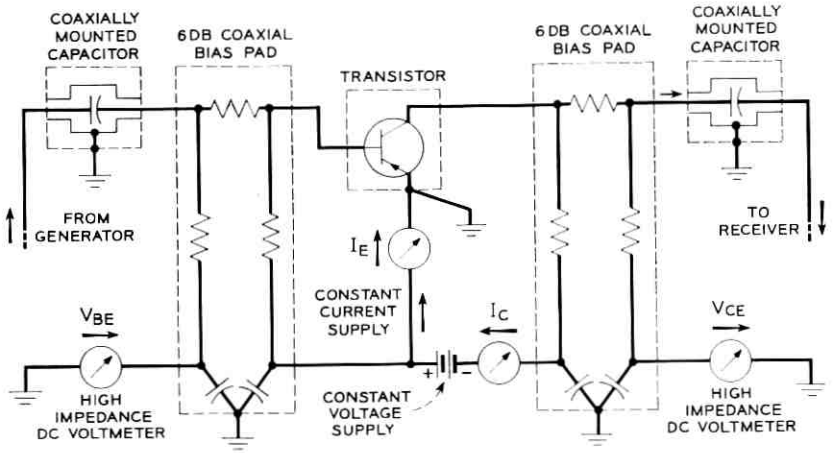


Fig. 12 — Signal and biasing paths during grounded-emitter measurements on a p-n-p transistor. Discontinuity due to biasing is minimized by feeding activating current through shunt legs of well-matched 50-ohm attenuation pad. Blocking capacitors prevent flow of biasing current back into measurement set.

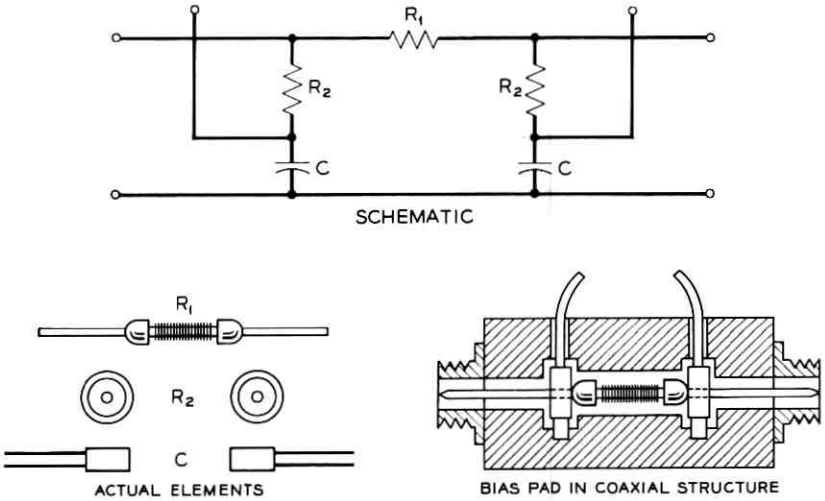


Fig. 13 — Bias pads coaxially constructed to preserve 50-ohm geometry. The shunt capacitance in each leg actually consists of three miniature capacitors inserted in parallel at points 120 degrees apart along periphery of shunt resistors R_2.

pads. Like the bias pads, the capacitors must be designed for minimum reflection, since they lie directly in the transmission path.

The capacitors are therefore enclosed in special coaxial housings to make them appear like short lengths of 50-ohm coaxial line. Fig. 14 illustrates the construction. The basic capacitance of 0.7 microfarad is provided by a tubular condenser whose shell is one of the electrodes. This large capacitance is necessary because phase shift-free transmission is required down to 5 mc. To permit the condenser to be inserted in a 50-ohm structure with minimum discontinuity, a pair of conically tapered electrodes are fitted to its terminals. This assembly forms the inner conductor of a short section of coaxial line. The corresponding outer conductor, which is shown disassembled, consists of three pieces, specially tapered to match the cigar-shaped inner conductor. A capacitative ring may be seen in the center piece of the outer conductor. The ring was introduced for the purpose of absorbing the remnant of reflection from the condenser's inductance and from the tapers. Measurements indicate a net return loss of 44 db at 125 mc and 34 db at 250 mc.

5.2 *DC Biasing and Monitoring Facilities*

Since the set was intended to be capable of measuring the variation of small-signal parameters with shift of biasing levels, it was important

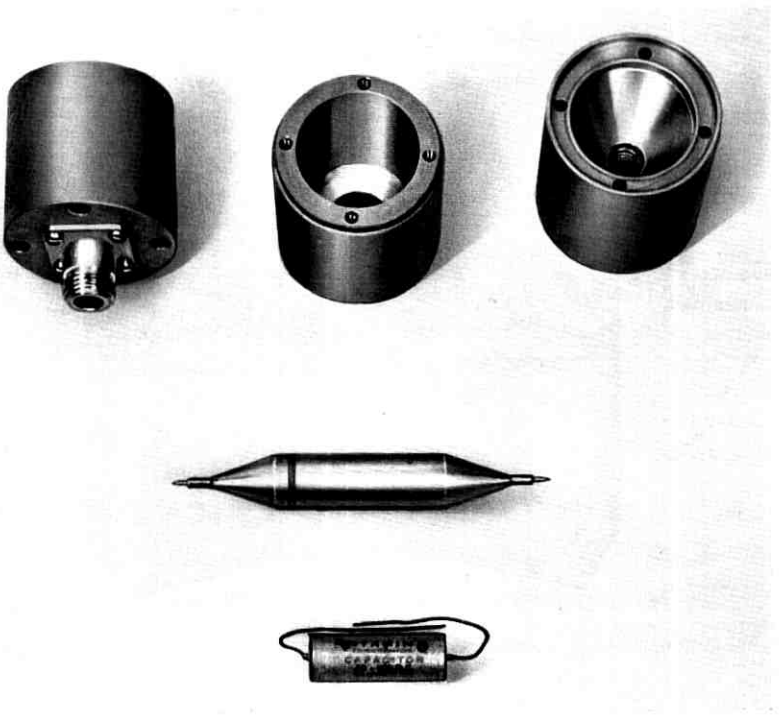


Fig. 14 — Disassembled view of blocking condenser: 0.7-microfarad capacitor is mounted in coaxial structure to make it look like a short section of 50-ohm transmission line.

to provide means for adjusting to a wide range of operating points. Three power supplies are provided for this purpose. Each is capable of being operated as a constant voltage or as a constant current supply. The third supply is used for tetrode measurements.

A system of monitoring operates in combination with the power supplies to facilitate setting of operating points.

VI. EQUIPMENT ASPECTS

6.1 *Over-all Layout*

The instrument is housed in a three-bay cabinet with the two end bays slanted inward to make it easier for a centrally located operator to read dials and manipulate knobs. This arrangement is shown in Fig. 15.

Just above table level in the central bay is the programming center. By means of pushbutton controls on the programmer, the operator can

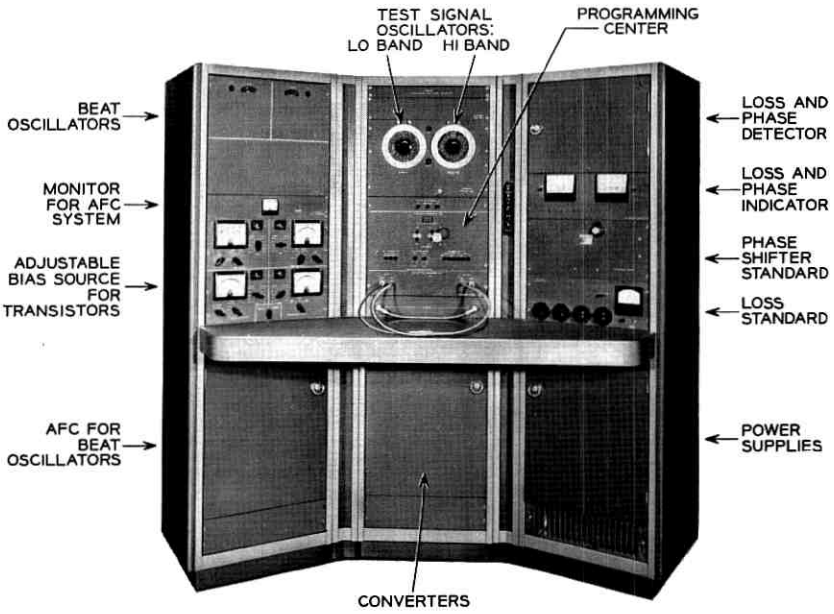


Fig. 15 — 5 to 250 mc phase and transmission measuring set.

set the machine up for 50- or 75-ohm insertion measurements on coaxially terminated networks, or for transistor measurement in grounded base, collector, or emitter configurations. The programmer includes jack appearances for inserting networks or transistors to be tested. Above the programmer is the test-signal source with its two large dials for selecting test frequency on either the low band (5 to 50 mc) or high band of operation (50 to 250 mc). The converters which translate the measurement data to the 1 mc detection frequency are located in the central bay (behind hinged panel) just below table level.

The beat-frequency source for the converters is mounted in the left hand bay. Automatic control circuitry in this bay maintains the beat frequency at the required 1 mc offset with respect to the frequency of test signal. Below the beat source are the biasing supplies for providing known dc energizing currents and voltages to transistors being tested.

The right-hand bay houses the loss and phase detector circuit, loss- and phase-indicating meters, and calibrated phase shifter and loss standards.

Power supplies and the remainder of the test set components are located below table level in the front and in mounting space available in the back.

6.2 *Programming Features*

A considerable measure of automatic operation has been introduced to make manual patching changes unnecessary when transferring among the various measurement modes. The necessary signal path changes from 50 ohms to 75 ohms or to transistor measurement are all effected by solenoid-actuated coaxial relays that receive their command signals from pushbuttons located on the central programmer unit. These may be seen in Fig. 16. The upper row of pushbuttons selects one of the three broad measurement categories, 50 ohms, 75 ohms, or transistor. If 50- or 75-ohm measurement is chosen, the unknown is connected to the appropriate ports in the lower part of the programmer.

Coaxial relays are also used to set up automatically the internal signal paths for each of the possible transistor measurements. By pushing the appropriate buttons, the operator may program the machine for six different measurements, with any terminal of the transistor being

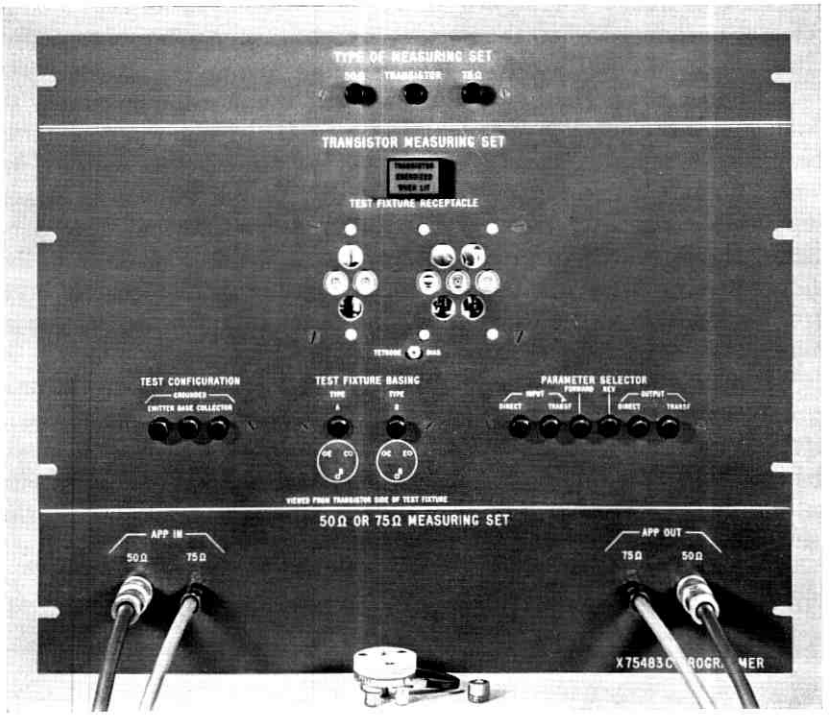


Fig. 16 — Programming center: solenoid-operated coaxial switches automatically set up signal paths for 50-ohm, 75-ohm, or transistor measurement configurations on command from front panel pushbuttons.

taken as common. These six measurements cover forward and reverse transmission and input and output bridging loss, with or without impedance inversion by the quarter-wave line. The unknown path is completed externally by plugging the jig into the centrally located jack field labeled "Test Fixture Receptacle" in Fig. 16. It is evident that special keying information must be given, since there are six ways to insert the jig, corresponding to the six possible measurements. The keying is done with the aid of six illuminable apertures located around the boundary of the test fixture receptacle. A particular one of the apertures automatically lights up in response to the specific combination of buttons that the operator has pushed. The jig is then oriented, before insertion, so that the grounded electrode of the transistor adjoins the illuminated aperture.

The biasing and programming circuitry is coordinated to keep the dc operating point unchanged during a shift from one measurement parameter to another. For example, if the operating point for grounded-emitter operation is established during measurement of forward loss, it remains unchanged when the buttons are pushed for reverse loss or for either of the two bridging measurements.

A number of safeguards were introduced to reduce the possibility of damaging transistors by operator errors. The most important of these provides for initial setup of approximate operating points with an internal dummy resistor network substituted for the transistor. Power supplies and auxiliary circuitry have been designed to minimize dangerous transient currents and voltages when transfer is made from the dummy network to the transistor. Moreover, special care has been exerted to insure that all biases are transferred at the same time, with minimum relative lag.

When the transistor is energized, a warning light appears on the programmer panel to remind the operator not to withdraw the jig without first switching off the biases. This precaution is taken because it is not generally possible to pull the jig out in such a way that all biases are interrupted at precisely the same instant. Conversely, the operator is cautioned against inserting the jig when the warning light is on.

VII. ACCURACY CONSIDERATIONS

7.1 *Validation of Accuracy*

The accuracy of the set was validated by a technique which did not require VHF standards of loss or phase shift. As a first step, the insertion loss and phase of each of four coaxial pads were measured. A meas-

TABLE II

Frequency (mc)	Arithmetic sum of loss and phase measurements on four 10-db pads		Measured loss and phase of cascade connection of the four 10-db pads		Magnitude of the difference	
	Loss (db)	Phase (deg)	Loss (db)	Phase (deg)	Loss (db)	Phase (deg)
5	39.99	2.4	39.87	1.9	0.12	0.5
50	39.91	20.7	39.87	20.4	0.04	0.3
115	39.99	47.3	39.94	46.4	0.05	0.9
225	40.02	93.4	39.98	93.0	0.04	0.4

urement was then made of the over-all loss and phase through the four pads connected in tandem, and the result compared with the arithmetic sum of the measurements on the individual pads. This comparison provides a measure of error which is independent of the loss of the pads or of their separate phase shifts, provided the pads are sufficiently well matched to 50 ohms to insure that interaction and mistermination effects are negligible. For the purpose of this test, pads of 35 db return loss were adequate. Measurements were made at a number of widely separated frequencies in order to uncover any latent pickup or crosstalk errors. Results are shown in Table II.

A similar check performed with two 10 db pads gave the results of Table III.

A further confirmation of the phase measurement accuracy was obtained by measuring the insertion phase shift of a section of precision 50-ohm coaxial line at a large number of frequencies. The line had a phase slope of approximately 0.5 degree per megacycle. Nowhere in the 5 to 250 mc band did the measured phase shift deviate by more than 0.3 degree from the linear phase characteristic drawn through the measured points.

TABLE III

Frequency (mc)	Arithmetic sum of loss and phase measurements on two 10-db pads		Measured loss and phase of cascade connection of the two 10-db pads		Magnitude of the difference	
	Loss (db)	Phase (deg)	Loss (db)	Phase (deg)	Loss (db)	Phase (deg)
5	20.01	1.2	19.98	1.1	0.03	0.1
50	19.97	10.5	20.01	10.3	0.04	0.2
115	20.02	23.5	20.01	23.2	0.01	0.3
225	20.03	46.3	20.03	46.1	0.00	0.2

7.2 *Residual Errors*

The principal remaining errors are those due to residual misterminations and to calibration errors of the loss and phase standards.

It is a fairly simple matter to compute loss and phase errors due to misterminations for measurements in the 50-ohm mode, because a high degree of symmetry exists, in this case, between the standard and unknown high-frequency channels. As was seen in Fig. 3, the only physical asymmetry is introduced by the mode-selecting relays in the unknown channel. However, measurements have shown that the path through these relays creates a reflection of not more than 0.01, even at 250 mc. Consequently, it is reasonable to consider that the effect of the relays is to introduce a flat time delay, which may be balanced by an equivalent length of coaxial cable added to the standard channel. The reflection from the attenuation pads is also less than 0.01, and therefore negligible.

Since both the s and x channels are essentially free of lumped discontinuities, it is possible to adjust cable lengths and arrange components symmetrically so that the impedances seen looking toward source or load from the mid-plane of the 12 db pad in the standard channel match those presented to the input and output ports of the unknown apparatus. Moreover, symmetry is sufficient to insure equal Thevenin generators at the mid plane of the 12 db pad and at the input port of the unknown. Symmetry also leads to equal transmissions from the reference plane in the s channel, and from the output port of the unknown, to the common point of convergence at the comparison switch preceding the measurement converter.

Under these circumstances, Appendix B shows that the loss and phase measurement error due to modest misterminations, when measuring passive, bilateral unknowns, is contained in the expression

$$\varphi \approx -s_{11}G - s_{22}L + GL, \tag{2}$$

where φ is in nepers and radians. The quantities G, L, s_{11}, and s_{22} are the reflection coefficients of the generator termination, load termination, and physical input and output scattering coefficients of the network under test, all taken with respect to the nominal impedance level (50 or 75 ohms).

If the phase angles of all quantities add up in the most pessimistic way to produce maximum loss error, (2) indicates that the error in loss or phase measurement may be as great as

$$\varphi_{\max} = |s_{11}G| + |s_{22}L| + |GL|. \tag{3}$$

For example, consider the situation at 250 mc in the present set.

Generator and load terminations, as seen from the unknown in the 50-ohm measurement mode, exhibit reflection coefficient magnitudes close to 0.04. Therefore, in measuring networks whose input and output scattering coefficients are, let us say, 0.1 with respect to nominal, the error in the loss measurement may be as much as

$$2(0.1)(0.04) + (0.04)^2 = 0.0096 \text{ neper},$$

or approximately 0.08 db. If the reflections were phased to produce greatest phase measurement error, (2) leads to the conclusion that the inaccuracy, in this example, could be as large as 0.01 radian, or 0.56 degree.

It is of interest to observe the dependence of the error magnitudes on the network's own imperfections. Assuming network reflections of 0.04 instead of 0.1, the errors drop to 0.04 db and 0.27 degree, without postulating any improvement in the source and load termination.

7.3 *Elimination of Errors Due to Misterminations During Impedance Measurements*

Impedances are determined by inference from bridging loss and phase measurements. When seeking highest measuring accuracy with this method, a number of secondary imperfections in the set must somehow be accounted for:

(a) impedance deviation of source and receiver terminations from the nominal value;

(b) existence of a transmission and phase "zero-line," amounting to 0.1 db and 2 degrees between 5 and 250 mc;

(c) difficulties encountered at high frequencies in precisely defining the location of the measurement plane;

(d) finite loss of the quarter-wave transformer and deviations from precise quarter-wave length when measuring high impedances;

(e) inherent transmission and phase-measurement errors of the loss and phase standards in the set; this also includes nonlinear effects in the modulator.

These factors may easily lead to measuring errors of the order of 10 per cent, especially at the higher frequencies of operation.

If the object is maximum accuracy, without regard to volume of computation, it is possible to calibrate the set so that the errors listed in (a) through (d) are largely eliminated. Computation may be only a minor factor if a digital computer is available to handle large amounts of calculation. The self-calibration procedure is based on the observation that, during impedance measurement, the set essentially measures

the complex transmission from an input port to an output port as a function of the value of an unknown impedance, Z, connected across a third port. It follows that the measured insertion loss and phase shift associated with Z must be of the general form

$$e^{-\theta} = \frac{a + bZ}{1 + cZ}, \tag{4}$$

where a, b, c are constants dependent only on the interior network of the set.[9] These would in general vary with frequency. Their values, however, may be obtained by measuring insertion loss and phase shift for three known values of Z. All residuals except item (e) are then automatically included in the a, b, c factors.

The practicability of this procedure was checked in the following way. First, the a, b, c constants were evaluated at approximately 10 frequencies, by measuring a coaxial short ($Z = 0$), a coaxial open ($Z = -j\infty$), and a high-quality termination ($Z = 50$ ohms). The planes of the "open" and of the "short" were designed to coincide, and they actually did so within a tolerance of 1 millimeter. Finally, a fourth impedance of known properties — a 30-centimeter length of precision 50-ohm line shorted at the far end — was inserted, and its impedance was computed from the relationship in (4), using the measured loss and phase data. All differences noted between the theoretical impedance value and the measured value were accounted for by the set's intrinsic loss and phase measurement inaccuracy [item (e) above].

VIII. TYPICAL MEASUREMENT RESULTS

One of the major uses of the set has been to provide measurement data for computing transistor h parameters up to 250 mc. A typical case was the characterization of broadband diffused-base germanium transistors in the grounded emitter configuration.

In measuring the two bridging parameters, the technique described in the last section was used. This consisted in making three preliminary insertion loss and phase measurements with three known impedances successively bridged across the 50-ohm transmission path: a "short," an "open," and a high-quality 50-ohm termination. A fourth measurement was then made looking into the transistor-loaded jig with the remote port terminated in a 50-ohm standard. Using the data of the first three measurements, it was possible to correct the fourth measurement for secondary sources of error contributed by the test set.

An error would normally result if the calibration impedances did not effectively lie in the plane defined by the base of the transistor header.

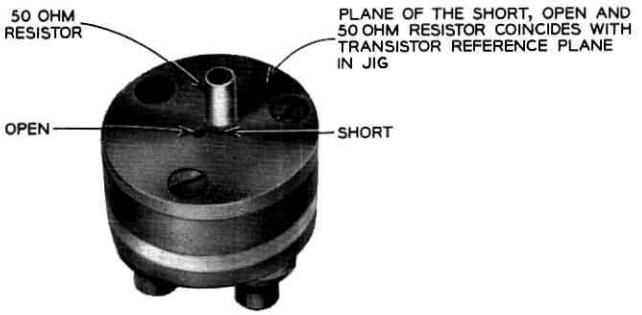


Fig. 17 — Calibrating fixture for use during transistor impedance measurements.

Fortunately, a ready-made solution to this problem was already available. By using the techniques employed in designing the transistor jig, it was possible to construct a companion fixture for unambiguously locating the three calibrating standards in the desired plane.

The calibrating fixture, which is shown in Fig. 17, has three internal 50-ohm transmission paths between the ports in the base and the top surface. One path terminates in a "short," another in an "open," and the third in a miniature 50-ohm film resistor enclosed in a conducting hood. Measurements indicate that the return loss presented by the 50-ohm resistor is greater than 34 db up to 250 mc. Since the path lengths have been made equal to those used in the jig, the plane of the standards coincides with that of the unknown. The calibrating fixture is successively inserted into the set in three different orientations, to present, in sequence, the "open," the "short," and the 50-ohm standard to the test set port that senses the transistor impedance.

The work in transforming from the measured data to h parameters was performed by a computer. A total of seven measurements had to be assimilated at each frequency: forward and reverse insertion loss and phase, the three calibrating measurements for correcting bridging loss and phase data, and, lastly, the input and output bridging measurements. A program already existed for converting from the scattering to the hybrid parameter matrix, so the computer first transformed the measurement data to an s matrix. The final output consisted of the frequency characteristics of the four h parameters. By way of example, Fig. 18 shows the variation of the magnitude and angle of h_{21} for one of the transistors tested. Notice that the phase of h_{21} is asymptotic to -90 degrees in the vicinity of 250 mc, which is consistent with the 6 db per octave slope in the magnitude characteristic starting from about 50 mc.

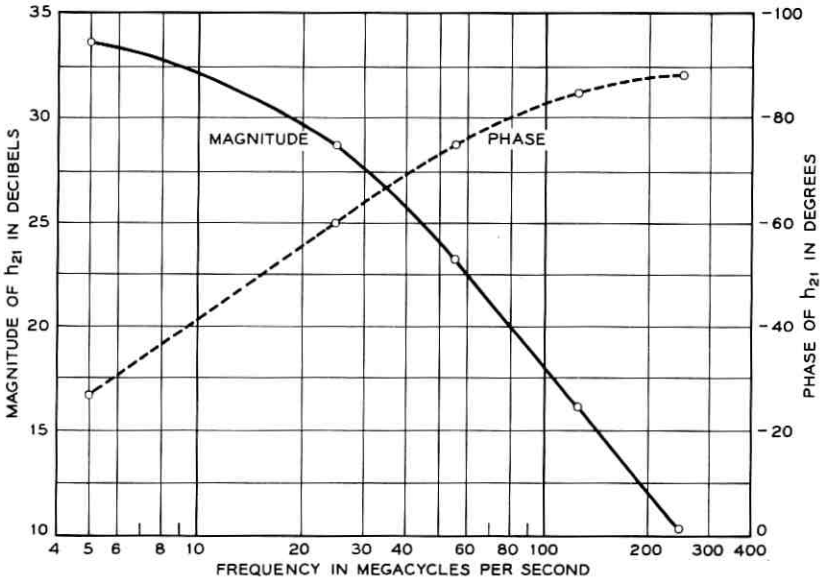


Fig. 18 — The h_{21} parameter for an A2104 transistor; grounded emitter, I_E = +10 milliamperes; V_{CE} = −10 volts.

IX. ACKNOWLEDGMENTS

The design of the automatic frequency control system for the beat oscillator is due to H. G. Follingstad. E. Widmann was responsible for design of the transistor jig and measurement features. The over-all mechanical design is the result of the efforts of R. P. Wells. Particular credit must be given to W. J. Fischer, who tested the circuits and cooperated in the development of various sections of the set. The construction of the transistor and calibrating fixtures and the transistor measurements were under the direction of J. Sevick. The mechanical assembly of the phase standard was due to J. Pasiecznik, while L. Howson and W. G. Hammett provided the design and calibration of the error correcting incremental phase shifter.

The authors are especially indebted to S. Doba, Jr., for much helpful criticism and advice.

APPENDIX A

Relation of Measured Quantities to Scattering Parameters

The insertion ratio of a network is defined as E_S/E_X in Fig. 19. The input-output scattering coefficient of a two-port network is the reciprocal

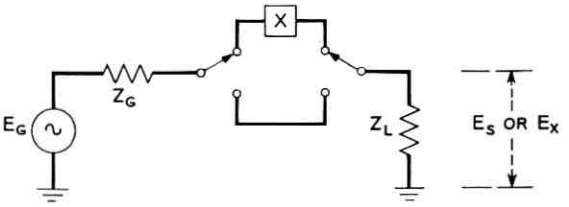


Fig. 19 — Network with insertion ratio E_S/E_X .

of the measured insertion ratio, provided the measurement is made between terminations equal to the normalizing numbers for the scattering matrix.[10] If $e^{\varphi_{12}}$ and $e^{\varphi_{21}}$ are the insertion ratios for the two directions of transmission, then

$$s_{12} = e^{-\varphi_{21}}$$

and

$$s_{21} = e^{-\varphi_{12}}.$$

In the present set, the source and load terminations are equal. It is therefore only necessary to show how s_{22} and s_{11} relate to measured insertion ratio during bridging measurements, when the remote port of the unknown is terminated in the common load resistance. Under these conditions, the impedance, Z, in equation (1) of the text, is the impedance presented by the network with the remote port terminated in its normalizing number. Thus, if port 1 is bridged,

$$\frac{Z_{11}}{50} = \frac{1 + s_{11}}{1 - s_{11}}$$

and, for port 2,

$$\frac{Z_{22}}{50} = \frac{1 + s_{22}}{1 - s_{22}}.$$

These two equations, together with (1), lead immediately to

$$s_{11} = \frac{3 - 2e^{\varphi_{11}}}{2e^{\varphi_{11}} - 1}$$

and

$$s_{22} = \frac{3 - 2e^{\varphi_{22}}}{2e^{\varphi_{22}} - 1}.$$

APPENDIX B

Insertion Ratio Measurement Errors Due to Mistermination

With the aid of Fig. 19, we define insertion ratio as the quantity E_S/E_X, where E_S is the voltage across Z_L when a path of zero loss and negligible length is connected between source and load, and E_X is the load voltage when the unknown is inserted. $|E_S/E_X|$ is insertion loss expressed as a numeric, and $\measuredangle(E_S/E_X)$ is the insertion phase shift.

We deal here with the common situation in which the nominal or "design" value of source and load is some impedance, Z_0. In the actual measurement situation the source impedance, Z_G, and load impedance, Z_L, are slightly different from Z_0. The mistermination error, ϵ, is contained in the quotient

$$\epsilon = \frac{\text{measured insertion ratio}}{\text{insertion ratio between design terminations}}. \tag{5}$$

It is desirable to obtain an expression for ϵ in terms of the scattering parameters of the unknown with respect to Z_0 terminations. This is preferred because of the close connection between the coefficients of the scattering matrix and readily measured quantities at high frequencies. Such a relationship is easily obtained by expressing the circuit properties of Z_G, the intermediate network, and Z_L in terms of their wave transmission matrices. It will be recalled that this matrix relates incident and reflected voltage waves at the ports in the manner

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix} \begin{pmatrix} b_2 \\ a_2 \end{pmatrix}. \tag{6}$$

The a's and b's are incident and reflected waves traveling on an impedance level of Z_0 ohms. These two equations may be solved for the coefficients of the "T" matrix in terms of scattering parameters:

$$\begin{aligned} T_{11} &= \frac{1}{s_{21}}, \\ T_{12} &= -\frac{s_{22}}{s_{21}}, \\ T_{21} &= \frac{s_{11}}{s_{21}}, \\ T_{22} &= s_{12} - \frac{s_{11}s_{22}}{s_{21}}. \end{aligned} \tag{7}$$

The T matrices of source impedance, network, and load impedance are multiplied to obtain the over-all matrix applicable from the generator terminals (E_G) to the load. The load voltage may then be shown to equal the generator voltage, E_G, divided by the sum of the four coefficients of the over-all T matrix. If the elements of the component matrices are expressed in terms of the scattering parameters [e.g., (7)], the load voltage will involve only the scattering coefficients of the network and reflection coefficients of source and load. This procedure results in the load voltages:

$$E_X = \frac{E_G}{2} \frac{s_{21}(1 - G)(1 + L)}{1 - s_{22}L - s_{11}G - GL(s_{12}s_{21} - s_{11}s_{22})}, \tag{8}$$

$$E_S = \frac{E_G}{2} \frac{(1 - G)(1 + L)}{1 - GL}. \tag{9}$$

The s parameters refer to the unknown network; Z_0 is both the input and output design impedance; G and L are the reflection coefficients of source and load impedance with respect to Z_0.

Hence the measured insertion ratio is

$$\frac{E_S}{E_X} = \frac{1 - s_{22}L - s_{11}G - GL(s_{12}s_{21} - s_{11}s_{22})}{s_{21}(1 - GL)}. \tag{10}$$

From (10), we determine the insertion ratio between design terminations (by setting $G = L = 0$) to be simply $1/s_{21}$. Hence we have

$$\epsilon = \frac{1 - s_{22}L - s_{11}G - GL(s_{12}s_{21} - s_{11}s_{22})}{1 - GL}. \tag{11}$$

If $|G|$ and $|L|$ are each much less than unity, and if s_{11}, s_{22}, and s_{21} are reasonably small, as is ordinarily the case, then

$$\epsilon = 1 + \varphi, \tag{12}$$

$$\varphi \sim -s_{11}G - s_{22}L + GL, \tag{13}$$

where φ is in nepers and radians.

REFERENCES

1. I.R.E. Standards on Solid State Devices: Methods of Testing Transistors, Proc. I.R.E., **44**, 1956, p. 1542.
2. Ketchledge, R. W., and Finch, T. R., L3 Coaxial System — Equalization and Regulation, B.S.T.J., **32**, 1953, p. 833.
3. Roetken, A. A., Smith, K. D., and Friis, R. W., TD-2 Microwave Radio Relay System, B.S.T.J., **30**, 1951, p. 1041.
4. Follingstad, H. G., Complete Linear Characterization of Transistors from Low Through Very High Frequencies, I.R.E. Trans., **I-6**, 1957, p. 49.

5. Slonczewski, T., Precise Measurement of Repeater Transmission, Elect. Engg., **73,** 1954, p. 346.
6. Alsberg, D. A., and Leed, D., Precise Direct Reading Phase and Transmission Measuring System for Video Frequencies, B.S.T.J., **28,** 1949, p. 221.
7. Goldman, S., *Frequency Analysis, Modulation and Noise,* McGraw-Hill, New York, 1948, p. 246, eq. (114).
8. Blackburn, J. F., *Components Handbook,* M.I.T. Radiation Laboratory Series, Vol. 17, McGraw-Hill, New York, 1949, ch. 9.
9. Bode, H. W., *Network Analysis and Feedback Amplifier Design,* D. Van Nostrand Co., New York, 1945, p. 223.
10. Carlin, H. J., The Scattering Matrix in Network Theory, I.R.E. Trans., **CT-3,** 1956, p. 88.

An AC Bridge for Semiconductor Resistivity Measurements Using a Four-Point Probe

By M. A. LOGAN

(Manuscript received September 19, 1960)

A new direct-reading ac bridge circuit has been developed to measure semiconductor bulk and sheet resistivity, using a four-point (or other appropriate) probe. The range of resistivity which can be measured is from 0.001 to 10,000 ohm-cm. Resistivity is read directly from resistance decades and a ratio multiplier, eliminating voltmeter and ammeter errors — the final reading being the result of a bridge-balancing operation for each measurement. Stability and sensitivity provide better than 0.5 per cent electrical accuracy, with mechanical point spacing being the controlling limitation on the over-all accuracy of the measurement.

The use of ac eliminates the influence of rectification, thermal, or contact potentials on the measurements, and also provides sensitivity more readily than with dc. The four-point probe and test specimen are the only nongrounded elements.

An Appendix compiles four-point probe conversion factors for thin circular and rectangular slices of material. New tables are presented for slices having a continuous diffused skin all over, and thus also conducting across the back.

I. INTRODUCTION

A basic measurement made on a semiconductor material is its resistivity. This is a measure of the impurity content, and determines the suitability of the material for a particular application and the necessary process parameters for subsequent operations. This measurement also determines whether a process step has been performed satisfactorily. Present methods for making this measurement usually are variations of the basic voltmeter-ammeter circuit, using direct-current power supplies and instruments. Such direct-current methods have many causes for error, several of which, while known to exist, are difficult to evaluate.

An ac measuring circuit has been developed for measurement of resistivity, retaining the four-point probe, but eliminating or minimizing to a negligible amount the errors inherent in the former dc systems. Every component of the new system is ac operated and grounded, except the test specimen and the four-point probe. By using an ac bridge, neither current nor voltage is measured. Rather, only their ratio is read from an accurate resistance standard, which really is what is required. The circuit can also be used for dc, but then many of its advantages are lost.

The new electrical ac system is accurate to 0.5 per cent when a test current that develops at least 1 millivolt between the two voltage probes is used. It includes built-in calibration for a system check at any time. It does not correct errors in point spacing of the four-point probe, but can provide a presetting for any given spacing, so that the resistance ratio dials read directly the bulk resistivity in terms of a semi-infinite body. Also, because it is a bridge system, the balance can be servo-controlled and plotted continuously on a recorder, or read out and recorded digitally.

II. PRINCIPLES OF A FOUR-POINT PROBE MEASURING CIRCUIT

The problem can be introduced by a description of a common dc system. For a resistivity evaluation of semiconductor material, electrical connections are made by pressing four needle points against the surface of the specimen. A convenient geometry is to space equally the four points along a straight line. Measuring current is then passed through the two outer points, called the current probes, and the voltage developed thereby between the two inner points, called the voltage probes, is measured. A common dc circuit is shown in Fig. 1.

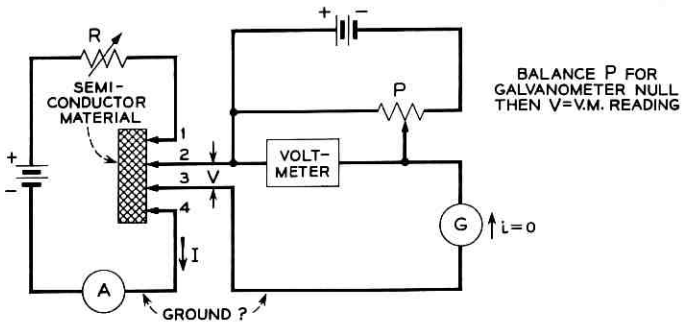


Fig. 1 — Basic four-point probe measuring circuit.

Each point introduces an unknown constriction resistance into the circuit, the minimum value of which is determined by the point pressure and the bulk resistivity of the specimen, as will be shown. In Fig. 2 is shown the equivalent resistance network of the semiconductor body and the four needle-point connections.

The unknown resistances of each of the points are shown as w, x, y, and z, respectively. The magnitudes of w and z can be compensated for by decreasing the current controlling resistor R of Fig. 1, or by using a constant-current generator instead. Thus by any of several means the current through the points 1 and 4 can be set to a specified value and measured by means of the current meter A, regardless of the magnitudes of w and z.

The voltage determination makes use of a balancing arrangement to eliminate current through points 2 and 3 with their unknown contact resistances x and y. By adjusting the potentiometer P, until the brush location is found for the condition of no current through galvanometer G, the voltage read by the voltmeter is that which opposes and exactly equals the unknown voltage V. Of course, if a high enough resistance dc voltmeter is used, the balance method can be avoided. "High" means of the order of 1000 times higher than whatever the unknown resistances x and y might be. Such voltmeters are power-line operated and require a ground on one terminal to bypass parasitic power line leakage currents away from the device being measured. One such ground can be connected to the circuit, such as on the galvanometer side. If this is done, then

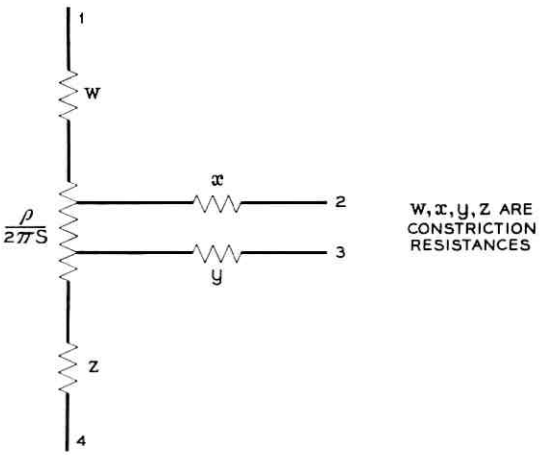


Fig. 2 — Equivalent circuit of semiconductor and four needle points.

the current supply cannot be grounded with this circuit, since points 3 and 4 would then be shorted and the current distribution in the semiconductor would no longer be known. Hence, floating batteries must be used with a power-line operated voltmeter.

Mathematical statements of the relationships involved in a resistivity determination using a four-point probe have been given by Valdes[1] and Smits.[2] For the case of a semi-infinite body and equal point spacings, the expression relating bulk resistivity, ρ, current, I, voltage, V, and point spacing, S, is

$$\rho = \left(\frac{V}{I}\right)(2\pi S).$$

Other expressions are derived for other than uniform spacing and proximity effects of nearby boundaries, all of which only alter the $(2\pi S)$ factor. In every case, the ratio of V/I appears explicitly. Thus, a determination of the voltage difference between the two inner points caused by a current flow through the two outer points is an indirect approach for a resistivity determination. Neither V nor I is really wanted, but rather their ratio. A direct measurement of this ratio is one of the features of the new circuit.

III. BLOCK DIAGRAM AND CIRCUIT FEATURES

The new bridge circuit and the use of ac rather than dc are shown in Fig. 3. The basic circuit is straightforward and inherently accurate. An oscillator in series with a current-limiting resistor sends alternating current through the two current points, the test specimen, and a grounded decade resistor. The voltages to ground of the two voltage points and part of the decade resistor are connected to three high-input impedance, nonphase-reversing amplifiers. The largest voltage, V_1, is next reversed in phase and added to the other two. The decade potentiometer is then adjusted until the sum is zero. A preamplifier, band pass filter, and an ac null detector provide the indication for this condition. The derivation of the equality of (V/I) in the test specimen to the decade resistor and potentiometer setting is given in the figure.

A summary of the advantages is as follows:

1. All components except the four-point probe and specimen are grounded.
2. The circuit is entirely ac operated.
3. Voltmeter terminals are two "open grids."
4. There are no meters, only a frequency selective null detector.

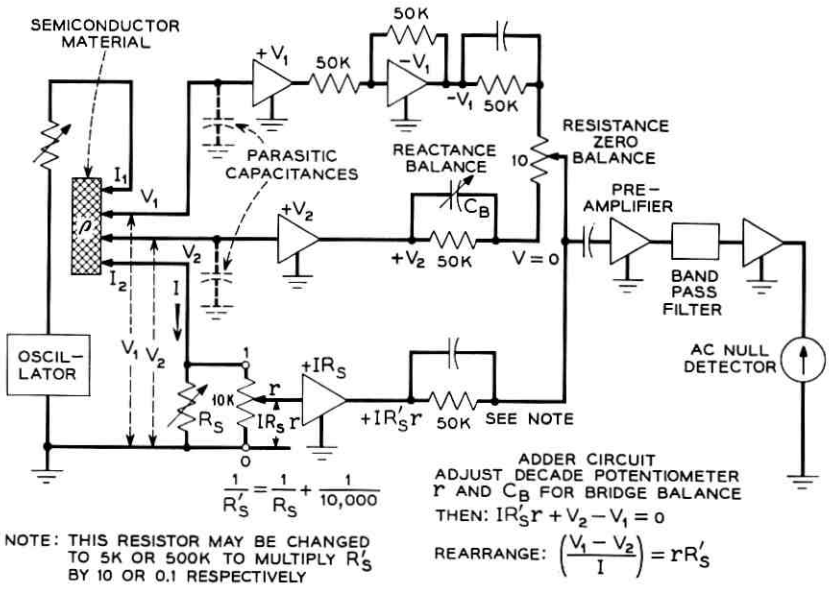


Fig. 3 — Diagram of measuring circuit.

5. Single direct reading of (V/I) from precision decade resistor and potentiometer is possible.

6. A constant-current source is unnecessary.

7. The current-resistor network can include a factor for actual probe-point spacing and read resistivity directly.

8. Servo balance of bridge can be applied for automatic measurement.

9. The circuit is independent of frequency.

The ac method also affords the following additional features:

10. Rectifying point contacts, barrier potentials and thermal potentials do not affect the fundamental ac voltage-current ratios, either in the test specimen or in the measuring circuits.

11. Superimposed DC can be used for incremental resistivity measurements, or to forward-bias a rectifying-current probe point.

IV. CONVERSION TO AC AND ELIMINATION OF METERS

For an ac system the current supply becomes an oscillator, which will be grounded. We now must measure the ac voltage between terminals 2 and 3 of Fig. 1, both of which are off ground. This suggests the use of a "differential" voltmeter. The voltages V_1 and V_2 of Fig. 3 are

ac voltages to ground, and their difference is the desired quantity. However, the voltage V_2 to ground is many times greater than the difference between V_1 and V_2, because at least that same difference exists between points 3 and 4 plus the added voltage drop in the still-present point-contact resistance z of Fig. 2 [which voltage is from 100 to 1000 times greater than $(V_2 - V_1)$], plus the drop in the decade resistor R_S. The use of part of the voltage drop across R_S, which will be found is made equal to $(V_1 - V_2)$ by adjustment, is the feature which eliminates meter readings. The three voltages are summed in an adder circuit, and when they are equal to zero as shown by the preamplifier, bandpass filter and null-detector, they yield directly the desired result:

$$\frac{V_1 - V_2}{I} = rR_S'.$$

Another important advantage is that a constant-current source no longer is needed. As the current changes, so do the voltages, and the bridge remains balanced.

The success of the circuit of Fig. 3 depends upon the high input impedance and relative linearity of the voltage amplifiers designated as $+V_1$, $-V_1$, and $+V_2$. Conventional differential voltmeters usually do not have adequate common-mode voltage suppression. Instead, precision amplifiers and a voltage adder circuit have been adopted, using the techniques of analog computers. The amplifier requirements and designs which satisfy them will be presented.

In return for ac power operation and grounding of all components except the four-point probe and test specimen, only the two indicated parasitic capacitances, in conjunction with the point-contact constriction resistances, x, y, and z of Fig. 2, contribute third-order errors.

The input impedance of each of the probes including the wire connections and "open grid" is essentially a capacitance of the order of 16 mmf. Each capacitance between each probe point and ground introduces a second-order quadrature voltage. Their difference is balanced out during each measurement by use of the reactance balance. There remains only a third-order error in the resistivity reading itself, which places a limit on the test frequency-resistivity product, as will be shown. Because of this, material below 100 ohm-cm resistivity is measured with a four-point probe and 390-cycle test current. Up to 500 ohm-cm resistivity, still with a four-point probe, the test frequency must be lowered to 85 cycles, which then also requires the use of a 4-cycle bandwidth wave analyzer or equivalent, for null indication. For material from 500 to 10,000 ohm-cm resistivity the two-point probe method, with end-

plated current connections and 85 cycles must be used. This is because surface states produce, in addition to capacitance-current effects, a nonhomogenous structure, and the curvilinear current flow no longer can be defined.

V. SUMMARY OF COMPONENT REQUIREMENTS

A description has been given in general terms of a semiconductor ac bridge resistivity measuring circuit. Aside from the oscillator, test specimen, and four-point probe, it consists of components having the following requirements:

1. A decade resistor and potentiometer network.
2. High input impedance precision voltage amplifiers, with (a) required input resistance of the order of 50,000 megohms and (b) relative linearity of one part per million.
3. A voltage adder, with adjustment for a stable zero balance of one part per million.
4. A high-gain selective null detector, which will (a) reject 60 cps and its harmonics; (b) reject harmonics of the testing frequency, primarily the second; and (c) have sensitivity of 0.5 microvolt.

The following sections of this paper will consider each of these parts in detail, to arrive at the component requirements in terms of the measurement accuracy objective. The procedure is to make each part capable of 0.1 per cent accuracy. Then a 0.5 per cent over-all accuracy will be realized.

VI. DECADE RESISTOR NETWORK

A slightly more detailed current-resistor network circuit is shown in Fig. 4. A precision 10,000-ohm Kelvin-Varley decade potentiometer calibrated as a ratio from 0 to 1 is in parallel with a preset fixed resistor decade. Once the decade has been set, the circuit balancing by the potentiometer will not alter the ac through the specimen.

For the decade resistors, the smallest step is 0.01 ohm and the largest is 100 ohms. With the low-resistance steps used with low-resistivity material, the dial-indicated resistance is not very accurate, primarily because of resistance in the wiring connecting it to the circuit. This wiring is fixed and can be measured, and a calibration chart can be made for each test set. This value is further subdivided into 10,000 parts by the decade potentiometer. Equal accuracy thus obtains for any decade resistor setting.

The resistance decades are connected in parallel with the ratio decades.

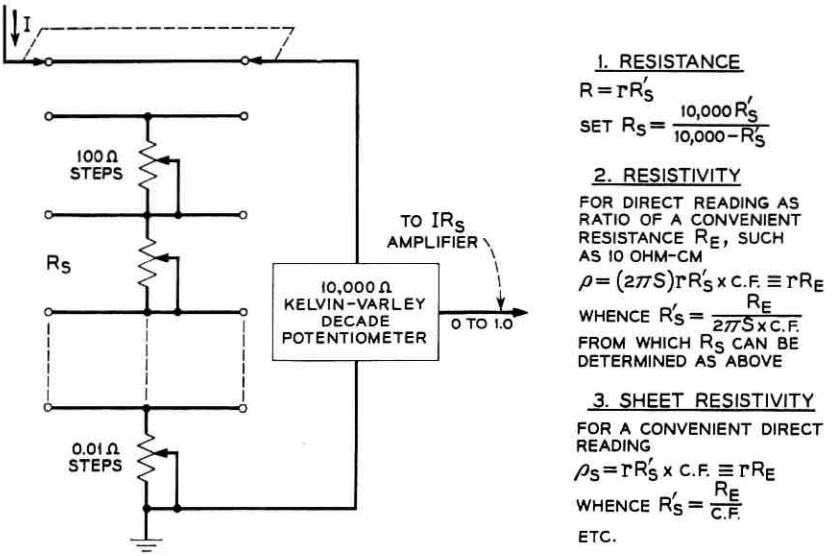


Fig. 4 — Decade resistor and potentiometer network.

A computation has to be made to determine the decade resistor setting to provide the desired parallel resistance. For instance, if 1000 ohms were wanted, the decade resistors would be set at 1111.1 ohms; if 10,000 ohms were wanted, they would be disconnected.

The resistor decade setting can be chosen to have the potentiometer read:

(a) resistance directly;
(b) body resistivity directly, including the actual point spacing; or
(c) sheet resistivity directly, including the actual point spacing.

The determination of R_S for these methods of operation is shown in Fig. 4.

VII. SEMICONDUCTOR EQUIVALENT NETWORK AND DIFFERENTIAL AMPLIFIER REQUIREMENTS

The requirements for the differential voltmeter are determined by the semiconductor material itself. To show this, an equivalent network of the system consisting of the material with the four needle points in contact, is needed.

Referring to Fig. 2, we need to know the order of magnitude of constriction resistances w, x, y, and z. For a given specimen, they all will be somewhat alike, but we cannot make the assumption that they are

even known, or equal. An important way in which the resistance z enters can be seen by the following.

There is a voltage drop at each current probe caused by the current constriction, the effect appearing as though there were resistances w and z as in Fig. 2. Similarly, there appear resistances x and y in the voltage probes, even if there is no current flowing. The resistance z is of first concern, though compensated for as regards current flow by the external circuit, because a voltage across z must be suppressed by the differential voltmeter. It will next be shown that this voltage drop to ground is of the order of 100 to 1000 times higher than the wanted voltage difference between the voltmeter probes 2 and 3.

Holm[3] has developed an expression for the constriction resistance between two materials in contact, one of them yielding plastically,* as

$$R = 0.445\rho \left(\frac{P_y}{F}\right)^{\frac{1}{2}} \text{ ohms},$$

where

ρ = resistivity in ohm-cm of the higher resistivity material,

P_y = yield or tensile strength in grams/cm^2,

F = contact force in grams.

Experimentally it has been found that silicon undergoes only elastic deformation before fracturing. In the 111 plane the fracture strength is about 20×10^6 grams/cm^2. No plastic yield strength data for osmium, the probe point material, have been determined, but they are estimated to be of the same order as those of steel, about 14×10^6 grams/cm^2. As this latter figure is lower, its plastic flow can be assumed to control the contact area. Thus, if each needle point has a force of 25 grams, then

$$\frac{R}{\rho \text{cm}^{-1}} \doteq 300 .$$

This ratio will change inversely as the square root of the applied contact force.

This is the minimum effect that has to be anticipated. Films may increase the ratio further, and rectification will introduce other erratic effects which will act to increase the effective resistance as well as to

* From Equation (14.08) of p. 75, of Ref. 3, combined with Equation (15.01), p. 79. Note that Holm's expression is twice the above, because his case is for like materials, whereas the resistivity of the osmium point used in the probe is small compared to that of a semiconductor.

introduce "noise" into the measurement. Thus, we see that an optimistic equivalent network will be as shown in Fig. 5. For a wanted voltage of one millivolt, there will be a voltage to ground on each differential voltmeter input of from 0.2 to 1 volt. An error in suppressing one volt, by one part in a million, can cause up to a 0.1 per cent error in the measurement. This requirement applies regardless of the resistivity of the material—whether it is 0.001 or 10,000 ohm-cm—as long as a four-point probe is employed, because about one millivolt will always be needed for the wanted signal. This will be discussed later when thermal noise limitations are considered.

The meeting of the linearity requirement is easily checked. The IrR_s input of Fig. 3 is made zero by setting r to zero. The V_1 and V_2 input leads are connected together and to 2 volts ac. They thus have identical voltages to ground and zero voltage difference. For this test condition, there must be less than about one microvolt output into the null detector. This suppression is set up by a micro-adjustment of one adder low-resistance potentiometer called the *resistance zero balance*. Then the oscillator output is decreased toward zero voltage and no output voltage should appear to upset the null balance. A switch is provided for conveniently making this bridge balance and testing for linearity. Input

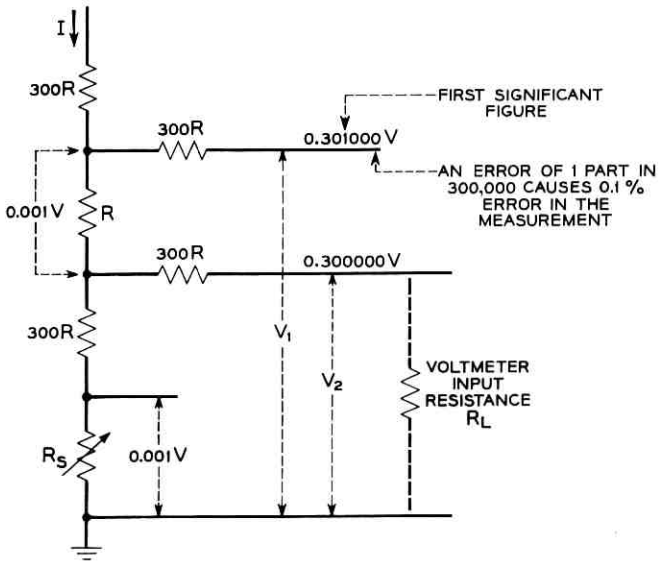


Fig. 5 — Differential amplifier linearity and input resistance effects.

tubes for the amplifiers have to be selected to meet this test. About one-third of those tested meet this requirement.

A second effect is the input resistance component of the differential voltmeter, including the vacuum tube socket, the grid and grid current, switches, wire insulation, leakage of the probe points to the supporting frame, etc. The requirement here is a function of the resistivity being measured. On Fig. 5, a spurious resistance is shown as R_L, connected between probe V_2 and ground. Suppose, for convenience, that a silicon sample of 500 ohm-cm resistivity is being measured and $2\pi S = 1$. Each probe contact will be of the order of 150,000 ohms. To first order, the 0.3 volt to ground at the inaccessible internal junction will be reduced at the accessible probe terminal V_2 by the amount

$$\Delta V_2 = \frac{0.15 \times 10^6}{0.15 \times 10^6 + R_L} \times 0.3 \text{ volt}$$

by ordinary potentiometer action. This is a direct error voltage and for 0.1 per cent accuracy must be less than 10^{-6} volt. For example, 10^{-3} volt is as large as the voltage we are attempting to measure. For 500 ohm-cm material the above equation shows that we must have

$$R_L > 45{,}000 \text{ megohms.}$$

For lower resistivity material, of course, this number becomes lower, but even for 1 ohm-cm material it is 10^8 ohms. It is obvious that only an "open grid" differential amplifier will be adequate. This, and the use of short grid leads, polyvinylchloride insulation, ceramic high-insulation switches, clean thermo-setting plastic mounting for the probe points, and point-to-point wiring, realize the requirement.

Thus, the second requirement for the differential amplifier is that it must have an input resistance greater than 50,000 megohms. This has been achieved.

However, we are using ac, and the reactance of the parasitic capacitance to ground of the differential volt-meter leads may be much less than the above. This places a limit on the resistivity frequency product which can be used. This limit will be developed in a later section. Basically, we can anticipate the results by observing that such a reactance primarily will produce 90° phase shift currents. Voltage drops due to such currents can be identified and balanced out, leaving only the wanted in-phase voltage.

VIII. NULL DETECTOR SENSITIVITY AND BANDWIDTH

The required voltage sensitivity of the bridge amplifier is a problem in signal to noise ratio. If the voltage probes have 1 millivolt difference,

then the voltage from the IrR_s amplifier is also 1 millivolt at the balance condition. An error of 0.1 per cent in setting the balancing voltage develops 1 microvolt change, which appears at the input of the null amplifier as about $\frac{1}{3}$ microvolt, because of the potentiometer action of the adder network. The presence of such a voltage must be recognized for an over-all accuracy of 0.5 per cent.

An ultimate limitation is thermal resistance and tube noise. The lowest signal voltage point in the circuit is the input to the null amplifier. With receiver type tubes, the circuit impedances can be kept to about 20,000 ohms. The average thermal resistance noise voltage is given by the formula:

$$E = \sqrt{4RKT\Delta F},$$

where

R = resistance in ohms,

K = Boltzman's constant, 1.38×10^{-23},

T = temperature in degrees Kelvin,

ΔF = bandwidth in cycles per second.

Assuming a 100 cps bandwidth selective filter, 20,000 ohms resistance, and a temperature of 310°K, the average input noise voltage to the bridge amplifier will be

$$E = \sqrt{4 \times 2 \times 10^4 \times 1.38 \times 10^{-23} \times 310 \times 100}$$
$$= 0.2 \text{ microvolt.}$$

Generally, the first tube plate contributes noise of the same order referred to the input, and other sources must be allowed for. These will be random. With careful design we can anticipate an over-all average noise of about 0.5 microvolt. Adding an error signal of this same amount should give an easily perceptible signal, unless the random nature of the circuit noise and its high peak factor cause too much instability in the no-signal indication. Even this can be alleviated to a considerable extent by damping the dc meter winding with a very large capacitor. The wanted signal is a steady sine wave so, after rectification, damping will not affect it but will suppress the occasional noise peaks to their time average value.

Most of the common laboratory ac voltmeters have a full scale sensitivity of 10 millivolts or better. One tenth of this, or 1 millivolt, can readily be discerned. The null preamplifier gain, from input to filter

output, therefore must be about 2000. The filter midband frequency, which has not yet been mentioned, will be arrived at in a following section, analyzing the effect of probe parasitic capacitance; 390 cycles has been adopted for all doped device material, but 85 cycles must be used for floating zone refined silicon.

Another source of thermal resistance noise is the constriction resistance at the probe points V_1 and V_2. These become controlling with resistivity material of about 50 ohm-cm or higher. The solution for this situation is to use a 4-cycle bandwidth for the null detector or a larger test current to develop 10 to 100 millivolts between the voltage probes. With high resistivity material, the power dissipated is negligible. Both of these means can be used.

IX. BRIDGE REACTANCE ZERO BALANCE

The circuit of Fig. 3 shows the adder network to consist of three equal resistors. Two of them, the $-V_1$ and the $+V_2$ are matched to one part in a million through the use of the resistance zero balance. Second, a reactance adjustment is also necessary to mop up for parasitic capacitance effects in the wiring of the amplifiers and adder circuit, even though the resistivity of the test specimen is quite low, so the probe capacitances are unimportant. The additional considerations for high resistivity material will be covered in Section X. For an extreme adjustment range of ± 50 mmf, it can be shown that an error of less than 0.07 per cent is introduced when a test frequency of 390 cycles is used.

X. ERROR DUE TO PARASITIC PROBE CAPACITANCE

As mentioned earlier, with high-resistivity material the quadrature currents through the voltage probe parasitic capacitances to ground introduce voltages to ground which are larger than the wanted voltage difference, and sometimes too large for compensation by use of the bridge reactance balance. Fortunately, just as the adder network subtracts the two in-phase ground voltages to develop the wanted voltage difference, so does the adder network subtract the two quadrature voltages. Variable trimmer capacitors can be added across each voltage probe to ground, one of which being used during a measurement to increase the smaller of the two quadrature voltages. However, even though the second-order reactance component is balanced, there is a third-order error term remaining which places a limit on the resistivity-frequency product for a specified error.

An appropriate equivalent circuit for analysis is shown in Fig. 6.

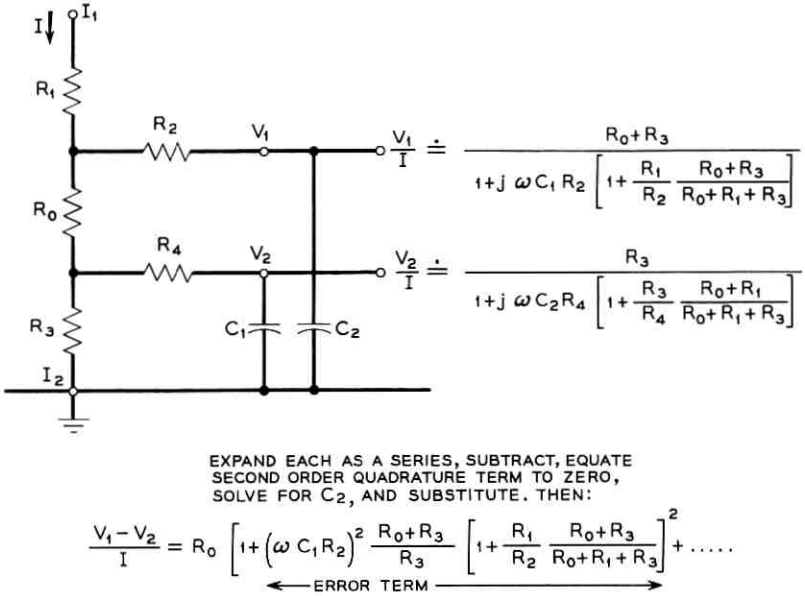


Fig. 6 — Error expression for probe parasitic ground capacitance.

Assuming no interaction between the second-order quadrature currents, expressions for the error term can be derived for two cases, first for a four-point probe (which up to here has been the only one mentioned) and second, a two-point probe where the current connections are made through plated terminals on a bar of material, and only the voltage probes are placed on the sample. The latter is present practice with floating-zone-refined silicon, and there are several reasons to continue this practice. In effect, this arrangement makes R_3 (and R_1) the same order as R_0 and lowers the voltages V_1 and V_2 to ground by a factor of about 0.01. This eases considerably the performance required of the amplifiers. A second effect is that rectification at the current probes is diminished. Thirdly, with high-resistivity material surface-state effects become increasingly important and cause the structure to be nonhomogeneous, invalidating an assumption made in deriving the conversion factors for resistivity expressions. At least, with current plates on the bar end the potential gradient is unaffected by surface states. There remains an error in computing current density. Near the surface it will differ from that in the interior, increasing in effect as the resistivity increases.

To attack the test frequency problem, the error expression derived

on Fig. 6 provides the answer. The easiest approach is to consider the two-point probe first. In this case, the values of R_0 and R_3 are of the same order and two orders of magnitude smaller than R_2 or R_4. For 10,000 ohm-cm material and 10 grams point force, a test frequency of 85 cps will cause an error, due only to frequency, of one per cent. Thus it will be necessary with such material to increase the point force to about 100 grams. For 3000 ohm-cm with 10 grams, the error reduces to 0.1 per cent. Thus it is clear that a frequency of this order is necessary, as well as being desirable for other reasons. The second-order error term on Fig. 6 does not indicate how accurately the trimmer condensers must be set for the quadrature voltage balancing—the lowest frequency possible is desirable for ease with this adjustment. Further, it is necessary to exclude 60 and 120 cps pickup and still detect 0.5 microvolt of wanted signal. We have shown earlier that a very narrow bandwidth is also necessary to keep the thermal resistance noise within bounds.

For device work, the bulk resistivity seldom will exceed 100 ohm-cm, permitting a much higher testing frequency and a four-point probe. For instance, assuming R_1, R_2, R_3, and R_4 are equal and 1000 times larger than R_0, substitution in the equation of Fig. 6 shows that, with 390 cycles, a resistivity of 500 ohm-cm is tolerable. However, difficulty in setting the quadrature balance and instability in the constriction resistances R_2 and R_4 have indicated a maximum of about 100 ohm-cm for this arrangement.

For zone-refined silicon, the four-point probe and 85 cycles can be used up to the order of 500 ohm-cm material, if surface-state effects are known to be unimportant. In case of doubt, the two-point probe method can always be used.

To summarize the discussion of measuring frequency and resistivity:

Four-point probe

up to 100 ohm-cm	390 cycles
up to 500 ohm-cm	85 cycles

Two-point probe

up to 500 ohm-cm	390 cycles
up to 10,000 ohm-cm	85 cycles

XI. RECTIFICATION

The connection to the semiconductor material, through the four pressure points, results in rectifying contacts. The test current flows

through the I_1 and I_2 contacts, but always in opposite sense. Thus, first one and then the other is backward biased, on alternate portions of the ac testing current. The total resistance of the series circuit, however, tends to remain constant as first one point and then the other is high resistance, and this effect is further aided by use of a relatively high series resistance in the oscillator output. The purpose of this resistor is to maintain substantially a sine wave testing current. Minor distortion is unimportant, since the wave filter in the null detector selects the fundamental frequency and excludes the distortion products.

The voltage to ground at each of the two voltage probes, however, does show the full rectification effects of only the I_2 probe point. The two rectified voltages are not quite equal because of the semiconductor body voltage drop, and this difference is the wanted sine wave for comparison to the IR_s voltage.

After transmission of the two probe voltages through the amplifiers and to the adder circuit, the subtraction which takes place in the latter largely balances the rectification components. Perfect balance to the distortion is not attempted, nor is it necessary. The wave filter selects the fundamental and rejects the remainder. Thus, in addition to noise suppression, the wave filter performs a second and equally important function. This is the reason an oscilloscope is sometimes not adequate as a null detector. As the balance is approached, the residual rectification products and noise become relatively large. Small changes in the fundamental cannot be identified.

XII. SUPERIMPOSED DIRECT CURRENT

As mentioned earlier, there is a dc path through the test specimen and oscillator circuit. Thus superimposed dc can be used. One such use is to eliminate a rectified waveform from appearing at the voltage probes. By connecting a battery or dc power supply in series with the oscillator having a magnitude slightly greater than the peak-to-peak ac voltage, the total test current through the I_1 and I_2 probe points will not reverse during a complete cycle of ac. The polarity of the dc, of course, must be in the direction to cause the I_2 probe point to be forward biased.

More direct current than the minimum can be used and incremental resistivity data taken to as high a current as desired, within the heating limitations of the semiconductor material. Most of the heat is generated directly under the back-biased probe point.

XIII. AMPLIFIERS

Each of the four precision amplifiers shown in block diagram form in Fig. 3 is a three-stage feedback amplifier. The three "diverted current"

nonreversing amplifiers are of one design and use series-input-shunt-output negative feedback. A schematic is shown in Fig. 7. The fourth, a reversing amplifier, uses shunt-input-shunt-output feedback. The internal amplification for each is 100,000, or 100 db, but externally they behave like a unity gain circuit. The internal full-gain bandwidth is from 40 to 400 cps. The internal frequency cutoff for the high frequencies starts at 700 cps and provides a constant 30° phase margin, including the effect of 120 mmf capacitance between the conductor and inner shield of the four-foot cable, and 2000 mmf between the inner and outer shields. Gain crossover occurs at about one-half megacycle. For the low-frequency end, somewhat reduced amplification extends to dc. This avoids the need for low-frequency-cutoff coupling networks. Such a circuit also carries the full probe voltage signals through to the adder network, including the dc component when rectification is present. Only after the subtraction is the residual dc error suppressed.

The input tube for all amplifiers is the 6AK5, having a gold-plated grid. The dc grid current measures less than one millimicroampere. A grid circuit is inherently a constant current circuit—hence its ac resistance is much higher than a simple dc voltage-current computation would indicate. A low reverse-current silicon diffused-junction diode is back-biased by a few volts, and connected to the grid of the second stage. This provides symmetrical overload conduction at that point, the grid itself for positive polarity and the diode for negative polarity. This arrangement prevents a Nyquist stable oscillation from occurring when the input grid is opened, causing the amplifier to go into overload, and then reconnected.

XIV. OSCILLATOR AND CURRENT SUPPLY

The oscillator should be of the bridge-stabilized type for good waveform, frequency, and amplitude stability.

The current output requirements cover a wide range. The load requires constant voltage, but the range of resistivity is expected to vary from 10,000 ohms to 0.001 ohm and the current from 0.1 microampere to 300 milliamperes. The load power in the semiconductor will be less than 1 milliwatt.

An open-circuit voltage of the order of 2 volts and an internal impedance (without feedback) designed for about a 5 ohms load is required. A series resistor of that amount will limit the current for the lowest resistivity condition and maintain good waveform. For any higher resistivity, simply increasing the series resistance reduces the test current as needed, keeping the open circuit voltage constant.

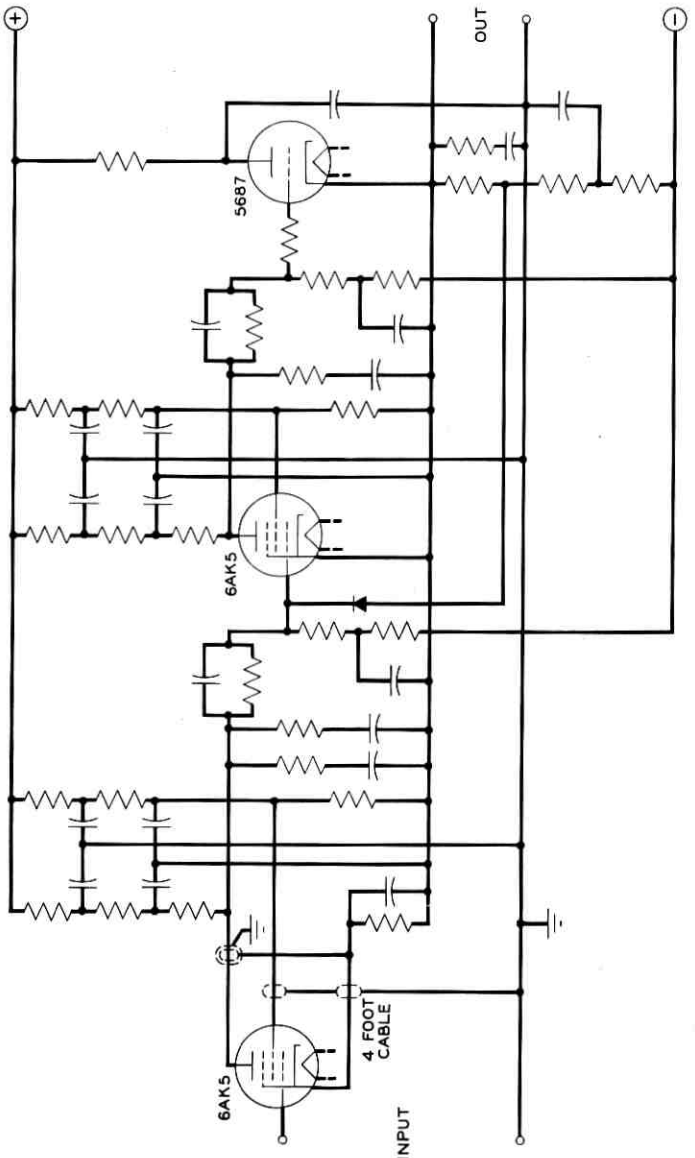


Fig. 7 — Schematic of series input one-to-one amplifier.

XV. ACCURACY AND PRECISION OF AC VERSUS DC

A statistical analysis of repeated measurements on the same specimen can establish the precision of a measurement but not the accuracy. The verification of the electrical circuit design accuracy has been established by the use of precision resistor networks such as the simulation shown in Fig. 2, including relative variations of resistances x and y by factors of 3 or 4 to 1. One of these is permanently wired to a switch, and can be used at any time. Other easily made ac tests using semiconductor material, such as varying the test current and frequency or measuring with the surface lapped or optically polished, showed slight effect using ac, validating the principles of the design, and indicate that an electrical accuracy of 0.5 per cent has been realized.

15.1 *Two-Point Probe*

The first two sets of measurements presented here have all been made using a two-point probe with the current connections made to plated end-surfaces of rectilinear bar samples. This type of measurement has no point-spacing error, and exhibits only electrical measurement errors. The two voltage points are cast in a rigid thermosetting plastic bar and pressed against the test specimen by an inverted pivot midway between the two points, assuring equal and independent point forces.

Measurements made by C. L. Paulnack and W. J. Thierfelder are shown in Figs. 8 and 9 for a "21 ohm-cm" p-type silicon bar. Note the expanded scale used for the ac chart, compared to the dc chart.

By a switching arrangement designed to avoid any mechanical motion, a forward and reverse measurement was made with the dc test set; then an ac measurement was made. The points were then lifted and reset on the surface before the next set of ac and dc measurements was made, repeating the procedure until a series of five measurements was completed each day.

After seven days of measurements, the surface of the silicon became pitted, and the points were moved 0.015 inch to a new position. The precision of the ac measurement clearly distinguished the nonuniformity of the specimen, whereas the dc test set was unable to. The dc resistivity indication was 5 per cent higher than was the ac.

The maximum expected experimental error for ac measurements is 0.19 per cent, compared to 0.82 per cent for dc, based on these results.

A second set of measurements made by C. L. Paulnack and S. J. Silverman is shown in Fig. 10, for a floating-zone silicon rectilinear crystal, approaching intrinsic resistivity of 11,000 ohm-cm. This 11,000

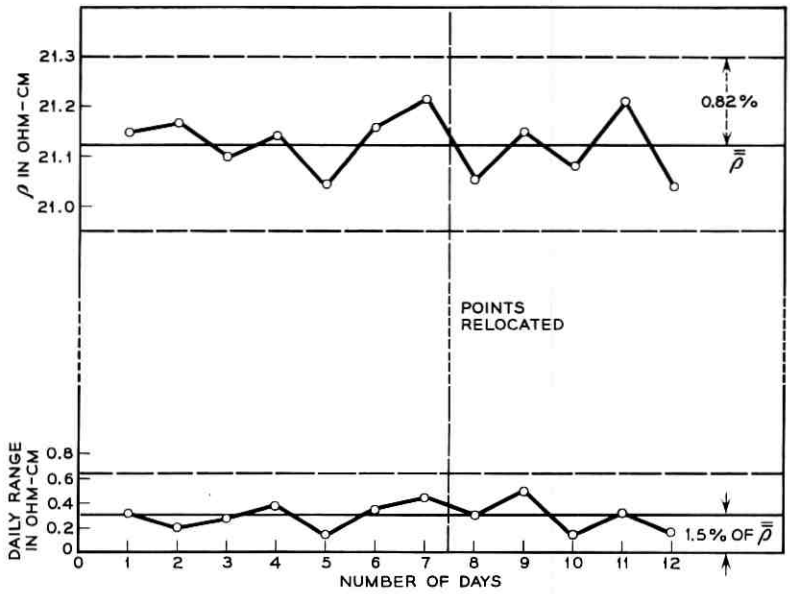


Fig. 8 — Two-point dc resistivity measurements of silicon.

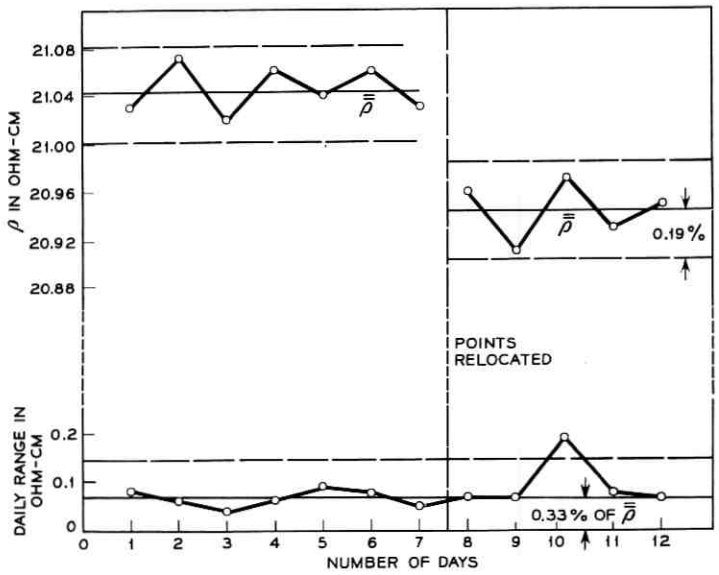


Fig. 9 — Two-point ac resistivity measurements of silicon.

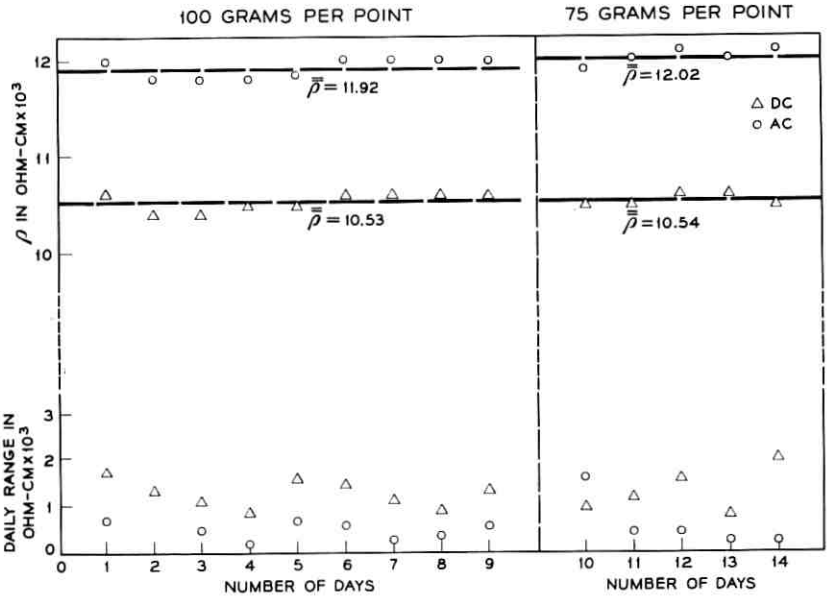


Fig. 10 — Resistivity of floating-zone-refined silicon.

ohm-cm material is approaching the upper design limit for the ac set, but the range is still about one-third that for the dc set. The dc resistivity indication was 14 per cent lower than was the ac.

Independence of sheet resistivity to point force is not true for very thin diffused junctions. With elastic deformation, the number of intrinsic carriers present is increased, the effect diminishing with distance from the needle point. These intrinsic carriers subtract from the effect of the impurities present, which formed the junction. If the junction is close enough to the surface, the effect of increasing the force is to observe an apparent abrupt decrease in sheet resistivity, as deformation of the junction brings it to the surface and conduction through the body then also obtains. This effect is reversible.

15.2 *Four-Point Probe*

Fig. 11 shows measurements on slices from an n-type silicon ingot using a very stable four-point probe. The probe design is based on a development due to N. J. Chaplin. Each slice was measured four times, with the points being raised and lowered between each measurement. These measurements include two causes for error, electrical and mechani-

cal. The relative position of the voltage needle points to the current needle points cannot absolutely be fixed, because of the requirement of independent and equal force application to all four points. The pairs are suspended independently, which permits relative pair position changes. However, any error is minimized, because half-way along a line between the two current points the voltage gradient has a broad minimum. Hence, the percentage error in voltage difference is less than the percentage error in mechanical displacement from the ideal location.

From Fig. 11, the short-time average range for four measurements is 0.65 per cent of the resistivity, including both electrical and point-spacing variations. From this, the maximum experimental error is 0.47 per cent. It can be verified independently that most of the variations are due to point spacing, by making repeated impressions of the points on a polished lead sheet and then measuring the actual spacings with a machinist's microscope. Over a longer time, the point-spacing variations can be expected to increase further, reducing the over-all precision. Point wear is one effect, but a more important cause is permanent relative changes in spacing due to handling of the probe assembly. In view of the results of the preceding section, it is evident that the electrical accuracy is superior to the point-spacing accuracy, even with the very stable four-point probe used in these tests. Hence, an evaluation of only electrical precision with a four-point probe cannot be achieved, only the over-all precision, as shown. Even so, the expected experimental error will be less than 0.5 per cent.

In general, ac measurements using device material of less than 500

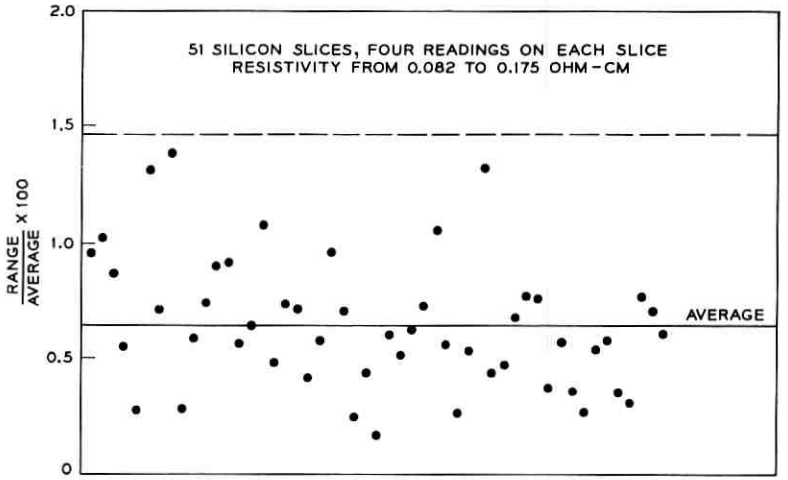


Fig. 11 — Four-point probe measurements.

ohm-cm resistivity are 5 to 15 per cent lower in resistivity than dc measurements. This consistent difference has not been accounted for.

XVI. EQUIPMENT

A photograph of the test set is shown in Fig. 12. The lower panel contains the resistance decades, the decade potentiometer, the bridge

Fig. 12 — Front view of semiconductor resistivity bridge.

and its zero balancing controls, the first stage of the preamplifier, and the four precision amplifiers. The center panel contains the second stage of the preamplifier, a 390-cycle bandpass filter, an ac millivoltmeter, and the regulated dc heater supply for the input tubes of the precision amplifiers. The upper panel includes the power supplies, the ac oscillator and test current controls, and a dc voltmeter for monitoring several circuits. The oscillator may be operated at either 390 cycles or 85 cycles. For all ordinary device work, a frequency of 390 cycles can be used, the bridge is self-contained, and only a probe is needed as auxiliary apparatus. When a frequency of 85 cycles is necessary, an external 4-cycle bandpass wave analyzer must be provided.

In Fig. 13 the probe amplifier and control circuit is shown affixed to the side of a four-point probe. The input vacuum tubes of the $+V_1$ and $+V_2$ amplifiers are underneath. Terminals are on top to connect direct wires to the needle points. One switch selects for measurement, bridge balance or calibrate, using a simulating resistor network described earlier. The other selects for use of a two- or four-point probe. Four-foot cables connect this circuit to the back of the resistivity bridge chassis.

XVII. CONCLUSION

This report has described the principles used in the development of a new general-purpose semiconductor resistivity measuring set using a four- or two-point probe and having an over-all electrical accuracy of better than 0.5 per cent. It covers the range from 0.001 to 10,000 ohm-cm material. It is an ac bridge, and every component is grounded except the specimen and the four-point probe.

The precision and accuracy, compared to a DC measurement, result from the fact that, while surface and point-contact potential effects cannot be separated from a dc measurement, they cannot affect the fundamental frequency ac voltage-current ratio. This principle is the basis for the measuring set design. In utilizing this principle, and at the same time grounding all components to permit ac power operation, a bridge circuit has been developed. The successful realization of the precision and accuracy required the development of an "open grid" differential amplifier that is both linear and has a common mode suppression of one part in a million.

APPENDIX

A.1 *Sheet Resistivity Conversion Factors*

When measurements are made with a four-point probe on thin homogeneous slices or surfaces having a one-sided diffused junction, the

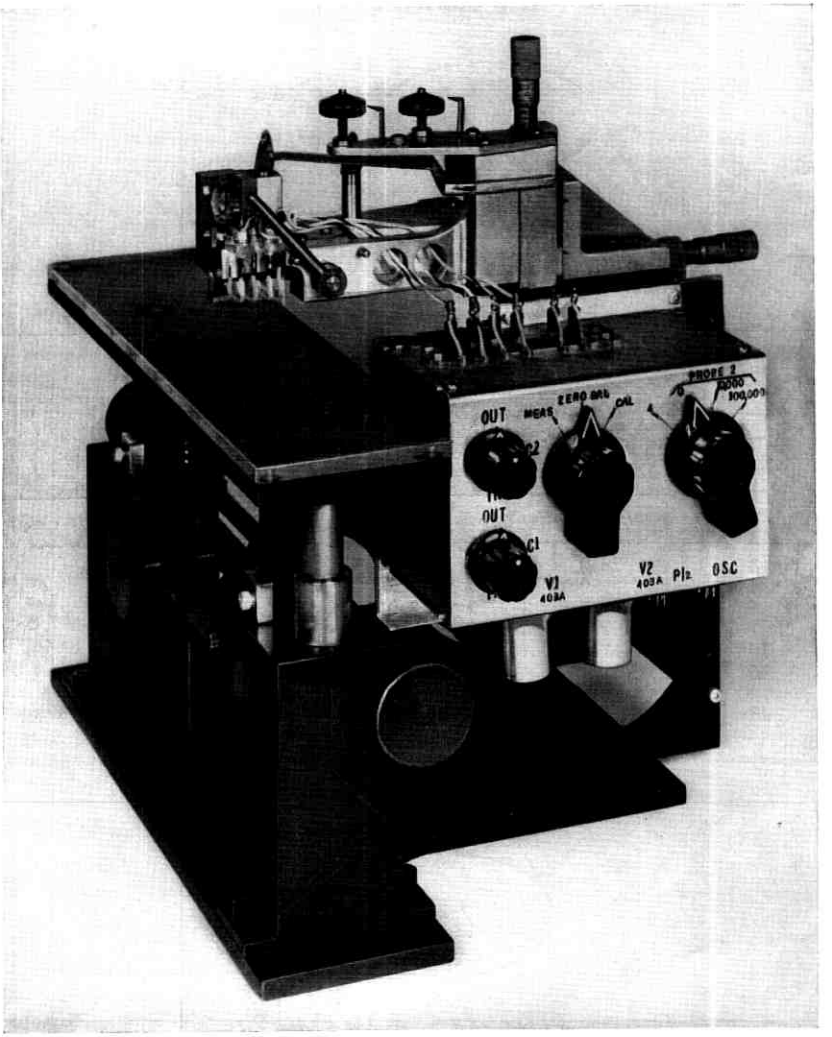


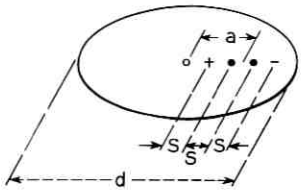
Fig. 13 — Four-point probe with amplifier attached.

resistivity reading V/I is altered, compared to an infinite sheet, by the finite size of the sample. Even for an infinite sheet, a conversion factor $\pi/(\ln 2)$ applies for equally spaced points on a line. Multiplying a measured V/I reading by the appropriate conversion factor places all measurements on a common, infinite-sheet basis. That is:

$$\rho_s = \frac{V}{I} \times \text{conversion factor.}$$

Table I — Conversion Factors for Sheet Resistivity Measurements of Circular Sample Using Four-Point Probe

Sheet with Insulated Edges



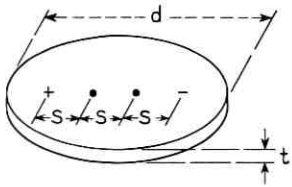
$0 \le a \le \left(\frac{d}{2} - \frac{3S}{2}\right)$

POINTS ON DIAMETER

$\rho_S = \frac{V}{I}$ C.F.; $\rho \doteq \frac{V}{I}$ C.F.W; $W/S < 0.5$

d/S	Conversion Factor = $\frac{\pi}{\ln 2 + \frac{1}{2}\ln \frac{\left[1-\left(\frac{2a}{d}+\frac{S}{d}\right)\left(\frac{2a}{d}-\frac{3S}{d}\right)\right]\left[1-\left(\frac{2a}{d}-\frac{S}{d}\right)\left(\frac{2a}{d}+\frac{3S}{d}\right)\right]}{\left[1-\left(\frac{2a}{d}-\frac{S}{d}\right)\left(\frac{2a}{d}-\frac{3S}{d}\right)\right]\left[1-\left(\frac{2a}{d}+\frac{S}{d}\right)\left(\frac{2a}{d}+\frac{3S}{d}\right)\right]}}$					
	$a/d = 0$	$a/d = 0.1$	$a/d = 0.2$	$a/d = 0.3$	$a/d = 0.4$	$a/d = 0.45$
3	2.2662					
4	2.9289					
5	3.3625	3.2719	2.9176			
7.5	3.9273	3.8780	3.6903	3.1123		
10	4.1716	4.1415	4.0263	3.6754		
15	4.3646	4.3504	4.2957	4.1284	3.2635	
20	4.4364	4.4282	4.3967	4.3000	3.8038	
40	4.5076	4.5059	4.4977	4.4730	4.3451	3.8568
∞	4.5324	4.5324	4.5324	4.5324	4.5324	4.5324

Conducting Sheet on All Surfaces



$2W < t \ll d$

W = JUNCTION DEPTH

$\frac{d+t}{S}$	Conversion Factor = $\frac{\pi}{\ln 2}$ (also independent of location of sample)
3	4.5324
4	4.5324
5	4.5324
7.5	4.5324
10	4.5324
15	4.5324
20	4.5324
40	4.5324
∞	4.5324

When slices have a continuous diffused skin all over, however, conversion terms more nearly approaching $\pi/(\ln 2)$ (the infinite sheet case) apply. Values for some important practical cases have been calculated, and tabulated along with Smits' earlier figures,[2] for ready reference.

A.2 *Circular Sample*

For a one-sided diffusion, only one image is necessary for each current point, to fulfill the boundary condition. Furthermore, the points need not be symmetrically placed with respect to the center. An extension of Smits' table for the circle, for the case of the four equally spaced points lying along a diameter, has been calculated and is shown in Table I.

For a two-sided diffusion the presence of conduction across the back surface acts for a circle exactly as if the front were part of a continuous infinite sheet. Hence, the conversion factor for this important case is always the same no matter where the points are placed or arranged.

The proof for the two circular conducting sheets, connected at the circumference, can be obtained by using conformal transformations. First cut the circumference, unfold, and place in the coordinate system as shown in Fig. 14. Then transform the upper circle into the upper half

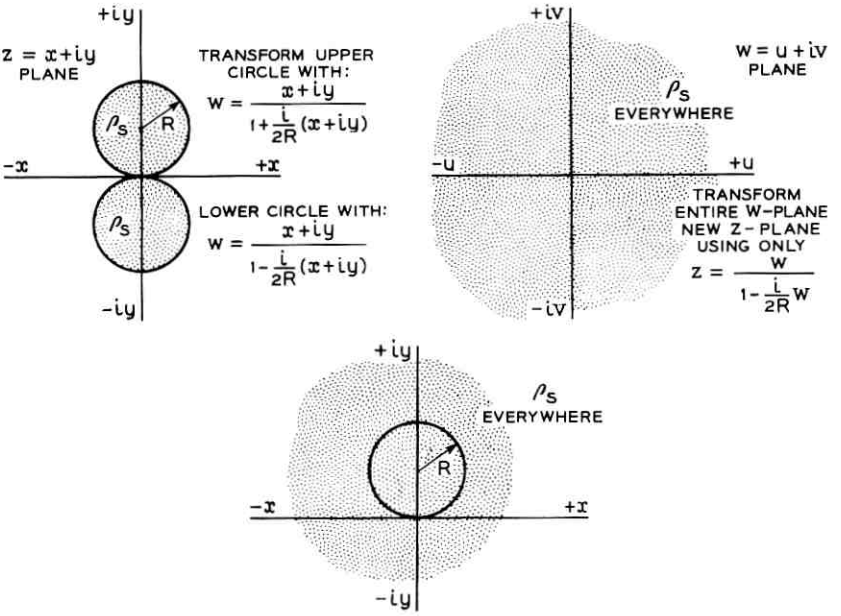
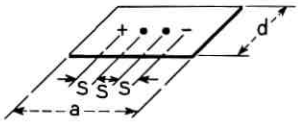


Fig. 14 — Conformal transformations for double-surfaced circle.

TABLE II — CONVERSION FACTORS FOR SHEET RESISTIVITY MEASUREMENTS OF RECTANGULAR SAMPLE USING FOUR-POINT PROBE

Sheet with Insulated Edges

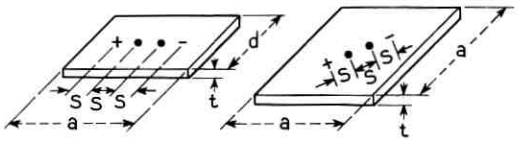


$$\rho_s = \frac{V}{I} C.F.; \quad \rho \doteq \frac{V}{I} C.F.W; \quad W/S < 0.5$$

W = JUNCTION DEPTH

d/S	$a/d = 1$	$a/d = 2$	$a/d = 3$	$a/d \geqq 4$	
1.0			0.9988	0.9994	
1.25			1.2467	1.2248	
1.5		1.4788	1.4893	1.4893	
1.75		1.7196	1.7238	1.7238	
2.0		1.9454	1.9475	1.9475	
2.5		2.3532	2.3541	2.3541	
3.0	2.4575	2.7000	2.7005	2.7005	
4.0	3.1137	3.2246	3.2248	3.2248	
5.0	3.5098	3.5749	3.5750	3.5750	
7.5	4.0095	4.0361	4.0362	4.0362	
10.0	4.2209	4.2357	4.2357	4.2357	
15.0	4.3882	4.3947	4.3947	4.3947	
20.0	4.4516	4.4553	4.4553	4.4553	
40.0	4.5120	4.5129	4.5129	4.5129	
∞	4.5324	4.5324	4.5324	4.5324	

Conducting Sheet on All Surfaces



$$\rho_s = \frac{V}{I} C.F.; \quad W < t/2$$

$\frac{d+t}{S}$	$\frac{a+t}{d+t} = 1$	$\frac{a+t}{d+t} = 2$	$\frac{a+t}{d+t} = 3$	$\frac{a+t}{d+t} \geqq 4$	$\frac{a+t}{S}$	Square Measure on Diagonal
1.0			1.9976	1.9497		
1.25			2.3741	2.3550		
1.5		2.9575	2.7113	2.7010		
1.75		3.1596	2.9953	2.9887		
2.0		3.3381	3.2295	3.2248	2.0	3.4700
2.5		3.6408	3.5778	3.5751	2.5	3.8696
3.0	4.9124	3.8543	3.8127	3.8109	3.0	4.1943
4.0	4.6477	4.1118	4.0899	4.0888	4.0	4.4212
5.0	4.5790	4.2504	4.2362	4.2356	5.0	4.4865
7.5	4.5415	4.4008	4.3946	4.3943	7.5	4.5233
10.0	4.5353	4.4571	4.4536	4.4535	10.0	4.5295
15.0	4.5329	4.4985	4.4969	4.4969	15.0	4.5318
20.0	4.5326	4.5132	4.5124	4.5124	20.0	4.5322
40.0	4.5325	4.5275	4.5273	4.5273	40.0	4.5323
∞	4.5324	4.5324	4.5324	4.5324	∞	4.5324

of a w plane and the lower circle into the lower half, reconnecting the two, now semi-infinite, surfaces along the x axis. The transformations shown are chosen to keep the origin in the center, and to place $i2R$ of the upper circle and $-i2R$ of the lower circle at infinity on the real axis. The associated overlapping insulating sheets formed by the areas outside the circles are discarded. Finally, transform the entire w-plane back into a new z-plane using everywhere the inverse transformation which restores the upper half of the w-plane to the original upper circle. The lower half of the w-plane fills the entire new z-plane outside the circle, and is connected to it at the circumference. Thus, in the new z-plane, any measurement inside the circle, which is merely a part of an infinite sheet, is identical with the same measurement on the original surface.

A.3 *Rectangular Sample*

For a one-sided diffusion, as shown by Smits and included in Table II, a doubly infinite array of image point pairs, one pair in each of the identical rectangles covering an infinite sheet, represents the system. For a two-sided diffusion covering all surfaces of the slice, image point pairs of the equivalent infinite sheet appear in only half the rectangles, checkerboard fashion, to represent the system. The rectangles void of image points represent the conducting under side.

The image point array for a rectangle with a two-sided diffusion is shown in Fig. 15. All points contribute to the voltage between points 1 and 2.

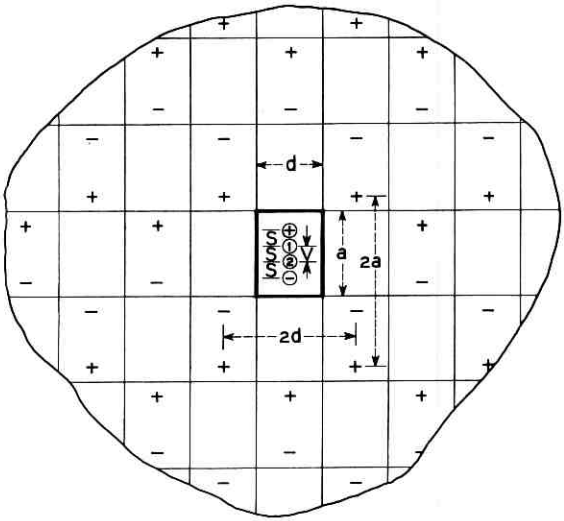
Each horizontal infinite line of equally spaced current sources, a distance $2d$ apart, as shown by Ollendorff[4] and diagrammed by Smits, causes a potential of

$$\varphi - \varphi_0 = -\frac{I\rho_s}{2\pi} \ln\left(2 \sinh \frac{\pi y}{2d}\right)$$

when a perpendicular to the line passes both through the voltage point and a current source, and

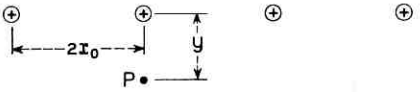
$$\varphi - \varphi_0 = -\frac{I\rho_s}{2\pi} \ln\left(2 \cosh \frac{\pi s}{2d}\right)$$

when a perpendicular to the line passes through the voltage point but is half-way between two of the current sources. For both cases, the perpendicular distance y is measured from the line of current sources to the voltage point being evaluated. For a current sink, the sign is reversed.



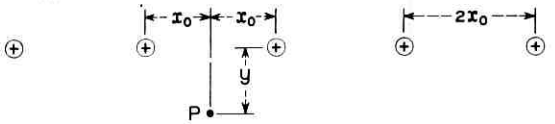
$$V = \frac{I\rho_s}{\pi}\left[\ln 2\cosh\frac{\pi S}{2d} + \sum_{n=1}^{\infty}\frac{2(-1)^n}{n\left[e^{n\pi a/d}+(-1)^n\right]}\left[\cosh 2n\pi\frac{S}{d} - \cosh n\pi\frac{S}{d}\right]\right]$$

A. POTENTIAL DUE TO LINE ARRAY WITH ALTERNATE GAPS, WITH POINT ON LINE OF ONE CURRENT SOURCE.



$$\varphi - \varphi_0 = -\frac{I\rho_s}{2\pi}\ln 2\sinh\frac{\pi y}{2x_0}$$

B. POTENTIAL DUE TO LINE ARRAY WITH ALTERNATE GAPS, WITH POINT ON LINE HALFWAY BETWEEN CURRENT SOURCES.



$$\varphi - \varphi_0 = -\frac{I\rho_s}{2\pi}\ln 2\cosh\frac{\pi y}{2x_0}$$

Fig. 15 — Image point array for rectangle with two-sided diffusion.

These expressions reduce the present problem to a summation of potentials, each due to a line of current sources, in only one direction.

Taking separately the two lines of current sources which include the two real current sources, their contribution to the desired total is

$$\Delta V = \frac{I\rho_s}{\pi} \ln\left(2 \cosh \frac{\pi s}{2d}\right).$$

Now, counting up all the other lines, eight infinite series are formed:

$$\Delta V = -\frac{I\rho_s}{\pi} \sum_{n=1}^{\infty} \left[\begin{array}{l} +\ln\left(2 \cosh \frac{\pi}{2d}\right)[(2n-1)a - 2s] \\ +\ln\left(2 \cosh \frac{\pi}{2d}\right)[(2n-1)a + 2s] \\ +\ln\left(2 \sinh \frac{\pi}{2d}\right)(2na + s) \\ +\ln\left(2 \sinh \frac{\pi}{2d}\right)(2na - s) \\ -\ln\left(2 \cosh \frac{\pi}{2d}\right)[(2n-1)a - s] \\ -\ln\left(2 \cosh \frac{\pi}{2d}\right)[(2n-1)a + s] \\ -\ln\left(2 \sinh \frac{\pi}{2d}\right)(2na + 2s) \\ -\ln\left(2 \sinh \frac{\pi}{2d}\right)(2na - 2s) \end{array}\right].$$

The hyperbolic terms are changed to the exponential forms and the logarithms of the products separated to the sum of two logarithms. The first term above thus becomes:

$$\frac{\pi}{2d}[(2n-1)a - 2s] + \ln[1 + e^{-\pi[(2n-1)(a/d)-2(s/d)]}].$$

The sum of the first terms of the eight expressions thus formed cancels. The eight logarithms are placed in series form, the first one being

$$e^{-\pi[(2n-1)(a/d)-2(s/d)]} - \tfrac{1}{2}(e^{-\pi[(2n-1)(a/d)-2(s/d)]})^2 + \tfrac{1}{3}(\cdots.$$

The first terms of the eight series now are summed, then the second terms, etc. The condensed form for the first term is:

$$2(e^{-\pi(2n-1)(a/d)} + e^{-\pi 2n(a/d)})\left(\cosh 2\pi \frac{s}{d} - \cosh \pi \frac{s}{d}\right).$$

The numbers $n = 1, 2, 3, 4, \cdots$ are next substituted in each such expression, which is factored, and the geometric series is identified and summed, yielding one final series:

$$\Delta V = \frac{I\rho_s}{\pi} \sum_{n=1}^{\infty} \frac{2(-1)^n}{n[e^{n\pi(a/d)} + (-1)^n]} \left(\cosh 2n\pi \frac{s}{d} - \cosh n\pi \frac{s}{d}\right).$$

Adding the ΔV term for the two separately evaluated lines and rearranging:

$$\rho_s = \frac{V}{I} \times \text{conversion factor (C.F.)},$$

where

$$\text{C.F.} = \frac{\pi}{\ln\left(2\cosh\frac{\pi s}{2d}\right) + \sum_{n=1}^{\infty} \frac{2(-1)^n}{n[e^{n\pi(a/d)} + (-1)^n]} \left(\cosh 2n\pi \frac{s}{d} - \cosh n\pi \frac{s}{d}\right)}.$$

A.4 *Double-Sided Square Sample With Points on a Diagonal*

A square sample with points on a diagonal can be cut to form an equivalent single sheet, as shown in Fig. 16. First observe that the horizontal diagonal, both front and back, is an equipotential line. Any point on it is equidistant from the current source and sink. Second, no current crosses the vertical diagonal, both front and back, because of symmetry. Cut these two diagonals on the back, unfold, and make the edges parallel to the line of points conducting to maintain the equipotential condition, as in Fig. 16. The other two edges are nonconducting.

This problem also can be solved by the method of images. The doubly infinite array which represents this problem is shown on Fig. 17. Each horizontal infinite line of equally spaced but alternate current sources

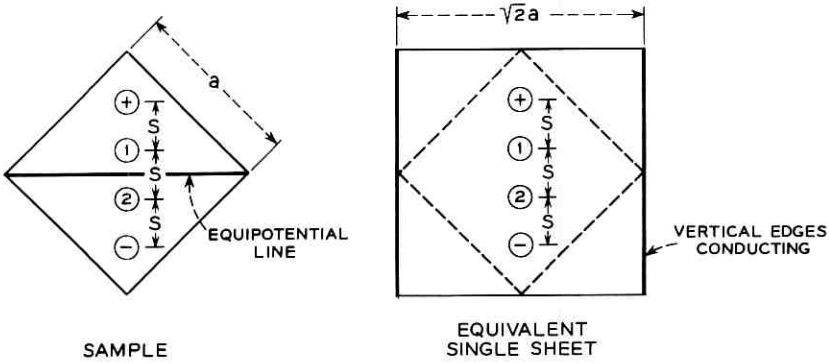
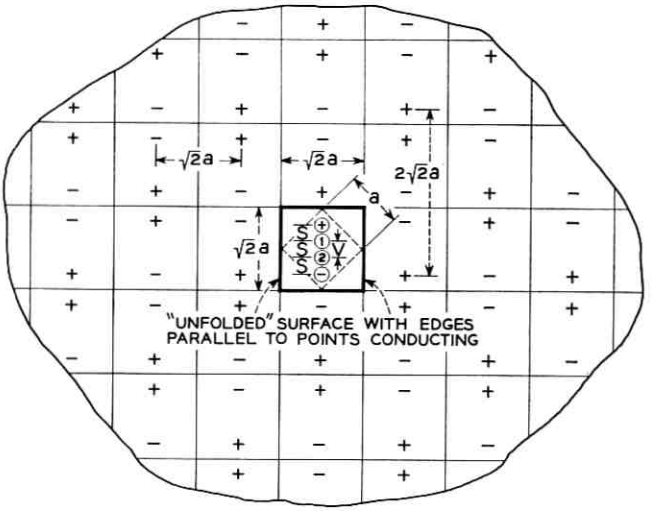
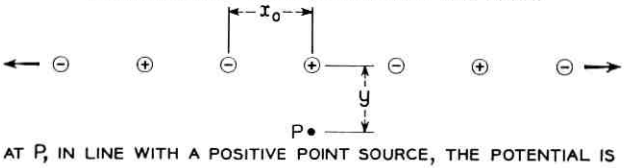


Fig. 16 — Square with front and back conduction and points on a diagonal.



$$V=\frac{I\rho_s}{\pi}\left[\ln\left(1+\frac{1}{\cosh\frac{\pi S}{\sqrt{2}a}}\right)+\sum_{n=1}^{\infty}\frac{4}{(2n-1)(1+e^{(2n-1)\pi})}\left[\cosh\frac{2(2n-1)\pi S}{\sqrt{2}a}-\cosh\frac{(2n-1)\pi S}{\sqrt{2}a}\right]\right]$$

POTENTIAL DUE TO AN ALTERNATING LINE ARRAY



AT P, IN LINE WITH A POSITIVE POINT SOURCE, THE POTENTIAL IS

$$\varphi-\varphi_0=-\frac{I\rho_s}{2\pi}\ln\tanh\frac{\pi y}{2x_0}$$

Fig. 17 — Image point array for square with two-sided diffusion and points on a diagonal.

and sinks, a distance $\sqrt{2}a$ apart, using a combination of the two line potential equations of the previous example, causes a potential of

$$\varphi-\varphi_0=-\frac{I\rho_s}{2\pi}\ln\tanh\frac{\pi y}{2\sqrt{2}a}$$

at a point a perpendicular distance y from a source. When in line with a sink, the sign is reversed.

As before, this expression reduces the problem to a summation of potentials, each due to a line of current sources and sinks, in only one direction.

Taking separately the two lines that include the real current source and sink, their contribution to the desired total is

$$\Delta V = \frac{I\rho_s}{\pi} \ln \left[1 + \frac{1}{\cosh \dfrac{\pi s}{\sqrt{2}a}} \right].$$

Counting up all the other lines, eight infinite series are again found:

$$\Delta V = \frac{I\rho_s}{\pi} \sum_{n=1}^{\infty} \left[\begin{array}{l} +\ln \dfrac{1}{\tanh \pi \left(n + \dfrac{s}{2\sqrt{2}a} \right)} \\ +\ln \dfrac{1}{\tanh \pi \left(n - \dfrac{s}{2\sqrt{2}a} \right)} \\ -\ln \dfrac{1}{\tanh \pi \left(n + \dfrac{s}{\sqrt{2}a} \right)} \\ -\ln \dfrac{1}{\tanh \pi \left(n - \dfrac{s}{\sqrt{2}a} \right)} \\ +\ln \dfrac{1}{\tanh \pi \left[\left(n - \dfrac{1}{2} \right) + \dfrac{s}{\sqrt{2}a} \right]} \\ +\ln \dfrac{1}{\tanh \pi \left[\left(n - \dfrac{1}{2} \right) - \dfrac{s}{\sqrt{2}a} \right]} \\ -\ln \dfrac{1}{\tanh \pi \left[\left(n - \dfrac{1}{2} \right) + \dfrac{s}{2\sqrt{2}a} \right]} \\ -\ln \dfrac{1}{\tanh \pi \left[\left(n - \dfrac{1}{2} \right) - \dfrac{s}{2\sqrt{2}a} \right]} \end{array} \right].$$

Convert these to exponential form using

$$\frac{1}{\tanh x} = \frac{1 + e^{-2x}}{1 - e^{-2x}}$$

and then expand the logarithm into a series. The first term becomes:

$$2[e^{-2\pi[n+(s/2\sqrt{2}a)]} + \tfrac{1}{3}(e^{-2\pi[n+(s/2\sqrt{2}a)]})^3 + \tfrac{1}{5}(\cdots].$$

Add the first terms of the first four, the first terms of the second four; then the second terms of the first four, the second terms of the second four; etc. Factor out the common terms in each group of four, rearrange and add the pairs, ending with a single series. The first term is

$$4(e^{-\pi(2n-1)} - e^{-2\pi n})\left(\cosh\frac{2\pi s}{\sqrt{2}a} - \cosh\frac{\pi s}{\sqrt{2}a}\right).$$

Substitute $n = 1, 2, 3$, etc., into the first bracket, factor out the common terms, identify the geometric series, and sum, yielding the final series:

$$\Delta V = \frac{I\rho_s}{\pi}\sum_{n=1}^{\infty}\frac{4}{(2n-1)(1+e^{(2n-1)\pi})} \cdot\left[\cosh\frac{2(2n-1)\pi s}{\sqrt{2}a} - \cosh\frac{(2n-1)\pi s}{\sqrt{2}a}\right].$$

Adding the ΔV term for the two separately evaluated lines and rearranging:

$$\rho_s = \frac{V}{I} \times \text{C. F.},$$

where

$$\text{C.F.} = \frac{\pi}{\ln\left(1+\dfrac{1}{\cosh\dfrac{\pi s}{\sqrt{2}a}}\right) + \displaystyle\sum_{n=1}^{\infty}\frac{4}{(2n-1)(1+e^{(2n-1)\pi})} \cdot\left[\cosh\frac{2(2n-1)\pi s}{\sqrt{2}a} - \cosh\frac{(2n-1)\pi s}{\sqrt{2}a}\right]}.$$

REFERENCES

1. Valdes, L. B., Resistivity Measurements on Germanium for Transistors, Proc. I.R.E., **42**, 1954, p. 420.
2. Smits, F. M., Measurement of Sheet Resistivities with the Four-Point Probe, B.S.T.J., **37**, 1958, p. 699.
3. Holm, R., *Electric Contacts* (Eng. trans.), Hugo Gebers Förlag, Stockholm, 1946.
4. Ollandorff, F., *Potentialfelder der Elektrotechnik*, Springer, Berlin, 1932.

Errors in Detection of RF Pulses Embedded in Time Crosstalk, Frequency Crosstalk, and Noise

By E. A. MARCATILI

(Manuscript received September 21, 1960)

The probability of error in the detection of RF pulses embedded in a combination of Gaussian noise, time crosstalk from the tails of two neighboring pulses, and frequency crosstalk from an adjacent channel, is calculated.

It is shown that for a given probability of error it is possible to maximize the pulse repetition frequency and simultaneously to minimize the channel spacing and signal-to-thermal noise by operating the system at a signal-to-thermal noise level close to the level of the combined time and frequency crosstalk.

I. INTRODUCTION

Consider many PCM messages occupying adjacent frequency bands in the same transmission medium, as, for example, in the proposed long distance waveguide communication system.[1] At some point the messages must be separated and read; these operations are performed by the receivers. Each receiver will be considered to consist of a filter and an envelope detector that takes periodic instantaneous samples and decides if the level of the signal is above or below a threshold.

Suppose that at a certain sampling time there is no pulse to be detected. Nevertheless, the received signal will be composed of the summation of three types of interference: time crosstalk or intersymbol interference, frequency crosstalk, and noise.*

Time crosstalk is measured by the envelope of the message at the sampling time in the absence of other messages and noise; it is due to the trailing and leading edges of the other pulses that make the message. It is known that if the sampling is instantaneous the time crosstalk can be reduced to zero by proper choice of filters and input signal,[2] but in a

* Throughout this paper we understand "noise" to be thermal noise.

real system the sampling time is not zero, and consequently the intersymbol interference varies during that time. The actual description of how this varying crosstalk influences the detected signal is a very complicated problem that involves a detailed knowledge, not only of the input pulses and transfer characteristics of transmitters and receivers, but also of the detector. We by-pass this problem by assuming conservatively a fictitious system that indeed has instantaneous sampling, but with the time crosstalk being the maximum value achieved by the time crosstalk in the real system during the finite sampling time.

Frequency crosstalk is measured by the envelope at the sampling time in the absence of the wanted message and the noise; it is due to the fact that the other messages have spectrums that overlap with the transfer characteristic of the receiving filter of the channel under consideration.

Finally, noise is measured by the envelope at the sampling time in the absence of all the messages; it comes essentially from the first amplifier in the receiver.

If the envelope of the three interferences is bigger than the slicing level, the detector decides that a pulse exists in that time slot, and an error is made. Similarly, the detector makes another error if a pulse should be detected but is shadowed by the interferences in such a way that the envelope of the received signal is smaller than the slicing level.

It is the purpose of this paper first to determine the relationship between the amplitudes of the wanted signal, time crosstalk, frequency crosstalk, noise, and slicing level; and second to establish in some sense the most efficient design of a system for a given probability of error.

II. DENSITY DISTRIBUTION OF SIGNAL, TIME CROSSTALK, FREQUENCY CROSSTALK, AND GAUSSIAN NOISE

Simplifying assumptions:

(a) Time crosstalk is represented by the sum of two sine waves of the same amplitude and arbitrary phases. The implications are: First, only the trailing edge of the preceding pulse and the leading edge of the following one are important. Second, each received pulse is symmetrical. This is rigorously true if the input pulse is symmetrical and the system has no phase distortion. Third, the phases of the pulses are uncorrelated, which is true if the pulses have passed through several partially regenerative repeaters.

(b) Frequency crosstalk is represented by a sine wave of arbitrary phase. The implications are: First, only one neighboring channel feeds

non-negligible power into the wanted channel. This is shown to be a reasonable assumption in Appendix A. Second, the pulses in different channels are synchronized. If they were not, the amplitude of the frequency crosstalk would vary between the two extreme values that can be obtained with the best and the worst interleaving of pulses.

(c) The noise is assumed to be Gaussian.

(d) The detector measures instantaneously if the envelope of the received signal is above or below a threshold. This is probably the crudest approximation, because in a real system the detector is not ideal and, what is even worse, the signal passes through repeaters with only partial regeneration.

The vector representing the signal to be detected is

$$S = A + \rho_T e^{i\theta_1} + \rho_T e^{i\theta_2} + \rho_F e^{i\theta_3} + \text{Gaussian noise}, \tag{1}$$

where A is the amplitude of the RF of the wanted pulse; its value is one if there is a pulse to be detected, and zero if there is no pulse; its phase is taken as reference. The second and third term represent the time crosstalk; they are vectors of the same modulus ρ_T, but arbitrary phases θ_1 and θ_2. The fourth term represents the frequency crosstalk of modulus ρ_F and arbitrary phase θ_3. Each one of these three last vectors, being originated from binary pulses, has a 50-50 chance of being present or not. The bivariate density distribution,* calculated in Appendix B, (52) is

$$\begin{aligned} p(x,y) = \frac{e^{-r^2/2\sigma^2}}{16\pi\sigma^2}\Bigg[1 &+ 2e^{-\rho_T{}^2/2\sigma^2} I_0\left(\frac{\rho_T r}{\sigma^2}\right) + e^{-2\rho_T{}^2/\sigma^2} I_0{}^2\left(\frac{\rho_T r}{\sigma^2}\right) \\ &+ e^{-\rho_F{}^2/2\sigma^2} I_0\left(\frac{\rho_F r}{\sigma^2}\right) + 2e^{-(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\frac{\rho_T r}{\sigma^2}\right) I_0\left(\frac{\rho_F r}{\sigma^2}\right) \\ &+ e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0{}^2\left(\frac{\rho_T r}{\sigma^2}\right) I_0\left(\frac{\rho_F r}{\sigma^2}\right)\Bigg], \end{aligned} \tag{2}$$

where x and y are the coordinates of the terminus of S, the vector representing the signal to be detected; $r = \sqrt{(x-A)^2 + y^2}$; σ^2 is the mean noise power; and I_0 is the modified Bessel function of first kind of order zero. The density distribution (2) is only valid for the tail of the distribution; that is,

$$\left.\begin{matrix}\rho_T \\ \rho_F\end{matrix}\right\} \ll r. \tag{3}$$

* For tutorial background see, for example, Bennett.[3]

It is possible to interpret the meaning of each term in (2). The first one is the contribution to $p(x,y)$ when only noise is present; on the average, this combination happens once each eight detections. The second term is the contribution when noise and only one of the two time crosstalk tails are present; on the average, this combination occurs once each four detections. The third term is the contribution when noise and both time crosstalk tails are present; on the average, this combination occurs once each eight detections. The fourth term is the contribution when noise and frequency crosstalk are present; on the average, this combination occurs once each eight detections. The fifth term is the contribution when noise, frequency crosstalk, and one time crosstalk tail are present; on the average, this combination occurs once each four detections. The sixth term is the contribution when noise, the two time crosstalk, and frequency crosstalk are present; on the average, this combination happens once each eight detections.

If there is a pulse to be detected (pulse on), A is equal to one and the density distribution (2) is

$$p(x,y) = p_1(x,y) \qquad (A = 1). \tag{4}$$

If there is no pulse to be detected (pulse off), A is zero and the density distribution is

$$p(x,y) = p_2(x,y) \qquad (A = 0). \tag{5}$$

Both functions, $p_1(x,y)$ and $p_2(x,y)$, schematically plotted as Figs. 1(a) and 1(b), have the same bell shape and circular symmetry around their respective axes located at $x = 1$, $y = 0$, and at $x = y = 0$.

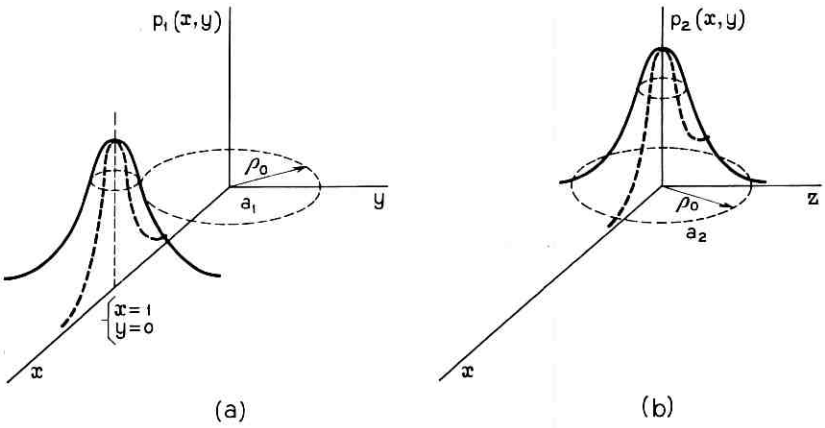


Fig. 1 — Density distribution for (a) pulse on, (b) pulse off.

III. PROBABILITY OF ERROR

In general, the volume defined by

$$P = \int_a p(x,y)\, dx\, dy \tag{6}$$

measures the probability that the signal S be a vector originating at the origin of coordinates and terminating at any point within the area of integration a.

The quantity $(1 - P)$ measures the probability that S is a vector with the terminus outside the area a. The detector decides whether the terminus is inside or outside of a.

Suppose that the signal free of interference A has its terminus outside of the area a; then if the received signal S is also outside of a the detector makes a correct decision, but if S falls inside of a the detector makes an error. Since the probability of finding S inside of a is given by P, this integral measures the probability of error and $(1 - P)$ measures the probability of making a correct decision.

The detector we use is one capable of deciding if the envelope of the received signal is bigger or smaller than a threshold ρ_0.

The probability of error in the "on pulse" condition, Fig. 1(a), is the probability that $|S| < \rho_0$:

$$P_1 = \int_{a_1} p_1(x,y)\, dx\, dy, \tag{7}$$

where $p_1(x,y)$ is derived from (2) by setting $A = 1$, and a_1 is the circle of radius ρ_0 and center at the origin of coordinates. The integration performed in Appendix C yields (70):

$$\begin{aligned} P_1 = \frac{K_0\left[\frac{(1-\rho_0)^2}{2\sigma^2}\right]}{16\pi}\Bigg[1 &+ 2e^{-\rho_T^2/2\sigma^2} I_0\left(\rho_T \frac{1-\rho_0}{\sigma^2}\right) \\ &+ e^{-2\rho_T^2/\sigma^2} I_0^2\left(\rho_T \frac{1-\rho_0}{\sigma^2}\right) + e^{-\rho_F^2/2\sigma^2} I_0\left(\rho_F \frac{1-\rho_0}{\sigma^2}\right) \\ &+ 2e^{-(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\rho_T \frac{1-\rho_0}{\sigma^2}\right) I_0\left(\rho_F \frac{1-\rho_0}{\sigma^2}\right) \\ &+ e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0^2\left(\rho_T \frac{1-\rho_0}{\sigma^2}\right) I_0\left(\rho_F \frac{1-\rho_0}{\sigma^2}\right)\Bigg], \end{aligned} \tag{8}$$

where I_0 is the modified Bessel function of the first kind of order zero and K_0 is the modified Bessel function of the second kind of order zero.

The probability of error in the "off pulse" condition, Fig. 1(b), is the probability that $| S | > \rho_0$:

$$P_2 = \int_{a_2} p_2(x,y)\, dx\, dy, \tag{9}$$

where $p_2(x,y)$ is obtained from (2) by making $A = 0$, and a_2 is the surface outside a_1. The integration performed in Appendix C yields (77):

$$\begin{aligned} P_2 = \frac{e^{-\rho_0^2/2\sigma^2}}{8}\Bigg[1 &+ 2e^{-\rho_T^2/2\sigma^2} I_0\left(\frac{\rho_T\rho_0}{\sigma^2}\right) + e^{-2\rho_T^2/\sigma^2} I_0^2\left(\frac{\rho_T\rho_0}{\sigma^2}\right) \\ &+ e^{-\rho_F^2/2\sigma^2} I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right) + 2e^{-(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\frac{\rho_T\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right) \\ &+ e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0^2\left(\frac{\rho_T\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right)\Bigg], \end{aligned} \tag{10}$$

with I_0 and K_0 being the modified Bessel functions of the first and second kinds.

The six terms appearing in expressions (8) and (10) have the same physical interpretation as that given for the six terms appearing in (2).

Since the "on" and "off" pulses are equally likely, the probability of error of the message is

$$P = \tfrac{1}{2}(P_1 + P_2). \tag{11}$$

The probability of error of the message P can be calculated for any combination of time and frequency crosstalk ρ_T and ρ_F, but it is possible to relate these two values by demanding that, according to some rule, both are equally damaging to the system. The rule we adopt is given by the following equations:

$$\begin{aligned} P_1(\rho_F = 0, \quad \rho_0 = 0.5) &= P_1(\rho_T = 0, \quad \rho_0 = 0.5); \\ P_2(\rho_F = 0, \quad \rho_0 = 0.5) &= P_2(\rho_T = 0, \quad \rho_0 = 0.5). \end{aligned} \tag{12}$$

For any signal-to-noise level and a slicing level equal to half the pulse amplitude ($\rho_0 = 0.5$), the probability of error in the "on" or "off" pulse condition due to noise and only time crosstalk is equal to that due to noise and only frequency crosstalk.

Substituting (8) and (10) in equation (12), we get

$$2e^{-\rho_T^2/2\sigma^2} I_0\left(\frac{\rho_T}{2\sigma^2}\right) + e^{-2\rho_T^2/\sigma^2} I_0^2\left(\frac{\rho_T}{2\sigma^2}\right) - 1 = 2e^{-\rho_F^2/2\sigma^2} I_0\left(\frac{\rho_F}{2\sigma^2}\right). \tag{13}$$

Frequency crosstalk ρ_F has been plotted against time crosstalk ρ_T, for different signal-to-noise levels $1/\sqrt{2}\sigma$ in Fig. 2.

A line defined by the following equation

$$20 \log \frac{1}{\rho_T} = 20 \log \frac{1}{\rho_F} + 3$$

has been included in the same figure (dotted line) for comparison purposes. Either from (13) or from Fig. 2 it can be deduced that for

$$\frac{\rho_T}{2\sigma^2} \ll 1, \qquad \rho_F \cong \sqrt{2}\rho_T \tag{14}$$

and for

$$\frac{\rho_T}{2\sigma^2} \gg 1, \qquad \rho_F \cong 2\rho_T. \tag{15}$$

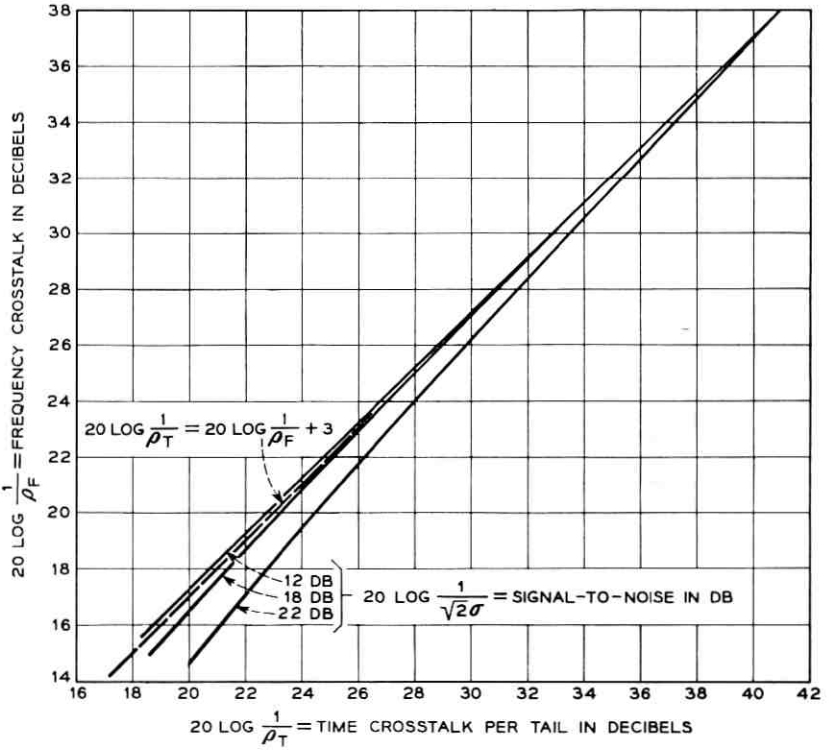


Fig. 2 — Equally damaging time and frequency crosstalk.

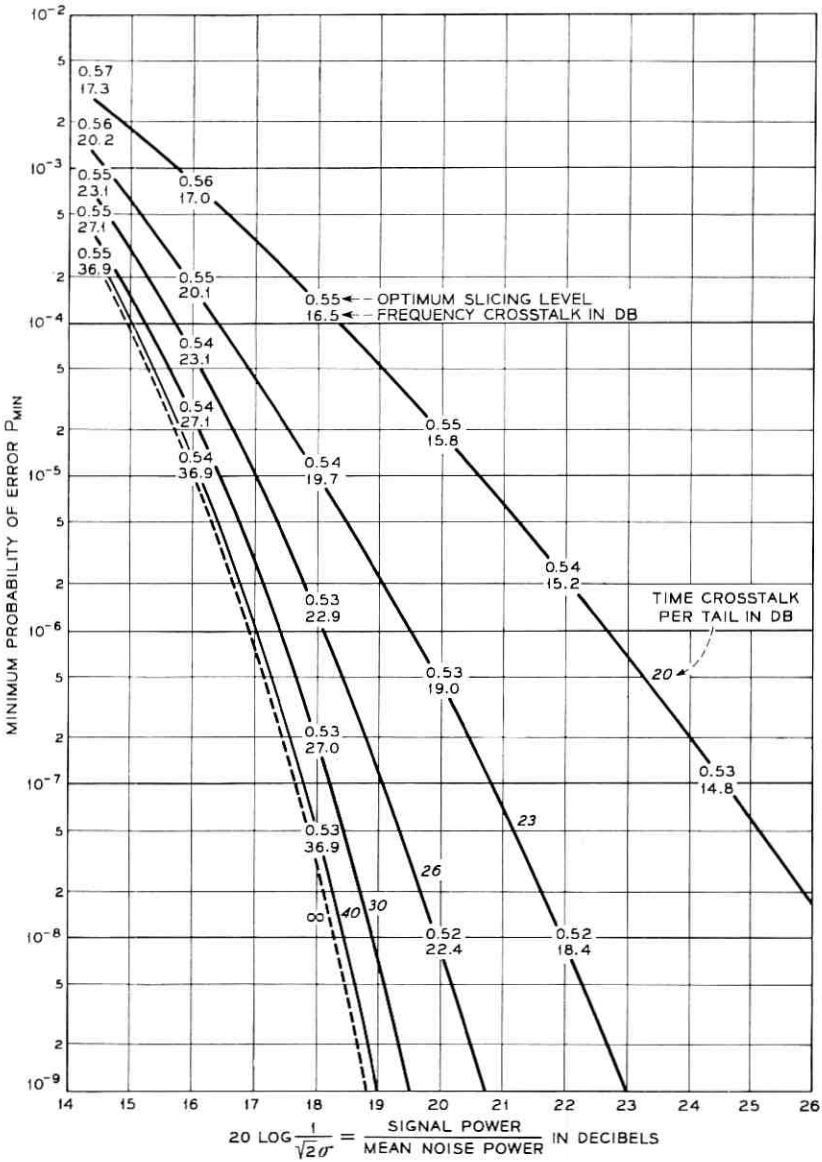


Fig. 3 — Probability of error of a message in the presence of crosstalk and noise.

For a given signal-to-noise ratio, if the normalized time crosstalk intensity per tail, ρ_T, is small compared to the normalized mean noise power $2\sigma^2$, two time-crosstalking tails introduce by themselves as many errors as does one frequency crosstalk 3 db above the level of each tail. But if $\rho_T \gg 2\sigma^2$ the time-crosstalking tails introduce as many errors as does one frequency crosstalk 6 db above the level of each tail.

For each set of values σ, ρ_T, and ρ_F that satisfies (13) we calculate from (11) the optimum slicing level ρ_0 that minimizes the probability of error, and $P_{\min}$, the value of that minimum. Fig. 3 contains this information. The probability of error, $P_{\min}$, is plotted as a function of signal-to-noise level for different values, ρ_T, of time crosstalk per tail. Each set of pairs of numbers on these curves indicates the local optimum slicing level ρ_0 and the frequency crosstalk ρ_F.

The dashed line (no crosstalk) almost coincides with that derived by Bennett.[3] The small difference stems from the fact that Bennett calculates the probability of error of the message for equal contributions of errors from the "on" and "off" pulse condition, while we calculate the minimum probability of error of the message.

IV. OPTIMUM DESIGN REGION

Suppose that we want to design a system with a given probability of error. Is there only one combination of values of crosstalk and signal-to-noise capable of satisfying the demanded probability of error? The answer is no. In Fig. 7 the given probability of error will be an ordinate obtainable with an infinite number of combinations of signal-to-noise level and crosstalk. We will develop two criteria for making a reasonable choice, and for that purpose we need some intermediate steps.

As a first step we redraw the part of Fig. 3 for low probability of error in Fig. 4, using time crosstalk per tail as the abscissa, signal-to-noise as the ordinate, and probability of error as parameter. The frequency crosstalk and optimum slicing level change slightly from point to point, but their exact values have not been written down.

As a second step we derive Fig. 5 from Figs. 6 and 7, which, together with Fig. 8, a sheet of definition of symbols, have been taken from the companion paper.[4] We shall see later how the derivation takes place, but first let us get acquainted with Figs. 6 and 7. In both these figures, the ordinates are proportional to time spacing between successive pulses, τ, times frequency spacing between adjacent channels, $|f_1 - f_2|$. The smaller this product the better, because time and frequency occupancies are proportional to the product. The different coefficients of

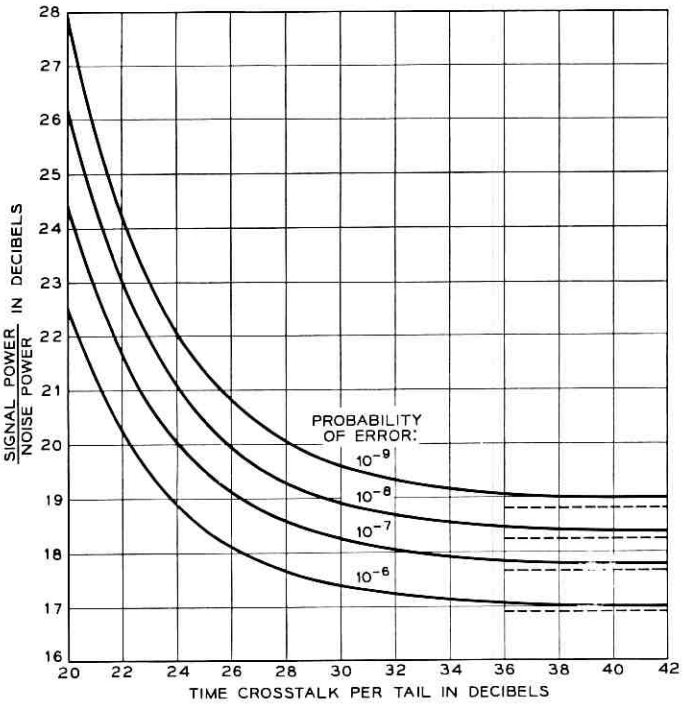


Fig. 4 — Reproduction of a section of Fig. 3, using new coordinates.

proportionality in the figures have to do with the input pulse width $2T$ and sampling time $2T_r$. The abscissas measure the ratio between bandwidth of the sending filter, $2F_1$, and the bandwidth of the receiving filter, $2F_2$. Eack figure contains three sets of curves, each corresponding to different time crosstalk per tail and different frequency crosstalk. Finally, the curves in each set correspond to different combinations of transfer characteristics of the transmitting and receiving filters.

The upper and lower dashed lines in Fig. 5 are applicable to systems with sending and receiving filters, each approximately maximally flat (three cavities); they have been derived from the dotted lines in Figs. 6 and 7, respectively. The upper and lower solid lines in Fig. 5 are applicable to systems with Gaussian sending filter and receiving filter approximately maximally flat (three cavities); they have been derived from the full lines in Figs. 6 and 7, respectively. The ordinates in Fig. 5 are the ordinates of the minimums of Figs. 6 and 7, and the abscissas

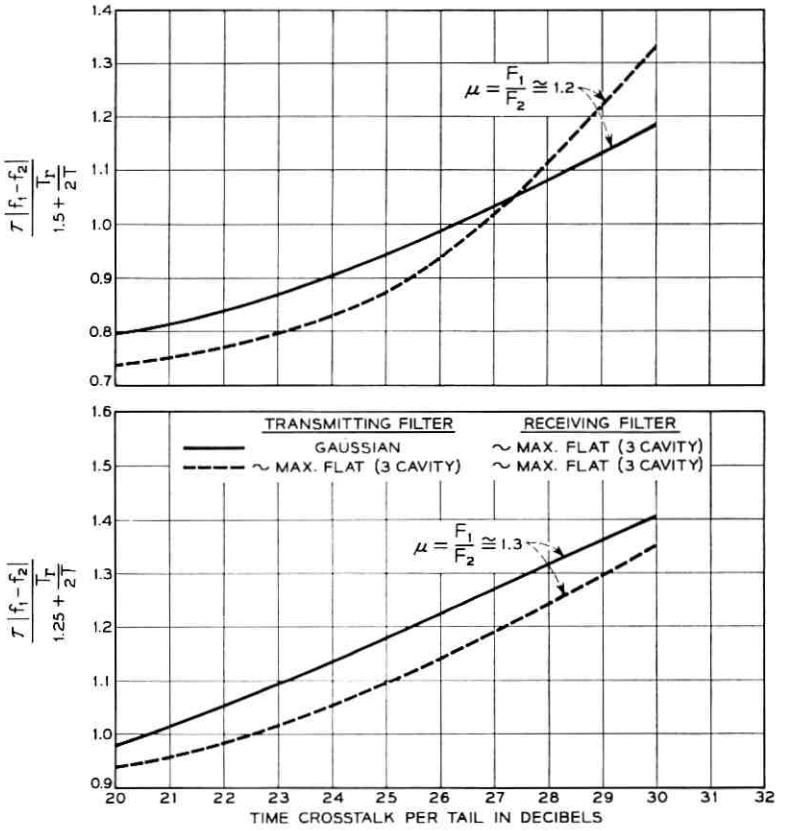


Fig. 5 — Minimum $\tau|f_1 - f_2|$.

in Fig. 5 are the different time crosstalks per tail corresponding to each set of curves in Figs. 6 and 7.

It is important to bear in mind that the ordinates of Fig. 5 are proportional to the minimum time spacing, τ, times channel frequency spacing, $|f_1 - f_2|$, which corresponds to maximum rate of information transmission.

As a third step we compare Fig. 4 with Fig. 5. For the same value of the abscissa both figures have ordinates that measure properties of the system we want to be as small as possible, but, since the slopes in the two figures are of different sign, a system operating at high time crosstalk per tail (small abscissa) will have (Fig. 5) a desirable low value $\tau|f_1 - f_2|$ but a large and unwanted signal-to-noise level. Conversely, a

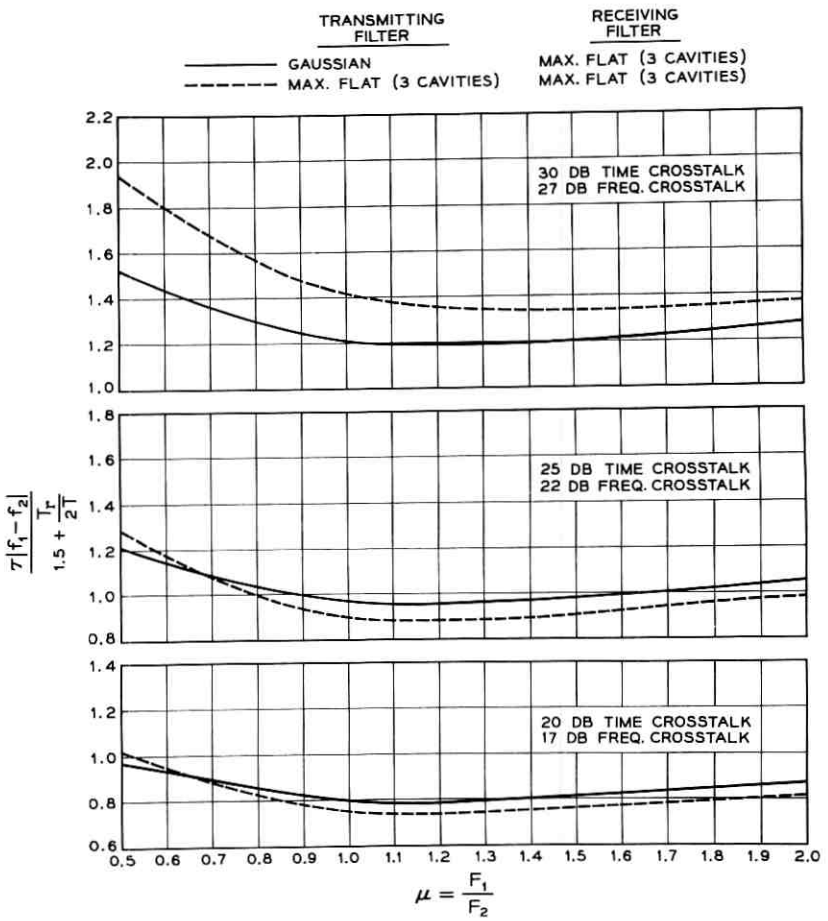


Fig. 6 — $\tau \mid f_1 - f_2 \mid$ curves for $\tau/2T = 1.5 + T_r/2T$.

system operating at low time crosstalk per tail will have an undesirably large $\tau|f_1 - f_2|$ and a wanted low signal-to-noise level. This suggests the existence of an intermediate optimum, and the question now is what function we want to minimize.

The answer is elusive, because what we really want is to minimize the price of a system that handles a certain rate of information with a given probability of error. That cost must be a function of time spacing, channel spacing, signal-to-noise ratio, and perhaps other variables. We don't know that function — at least not now — and because of lack of better knowledge we propose the minimization of two simple functions in which the signal-to-noise ratio is weighted differently:

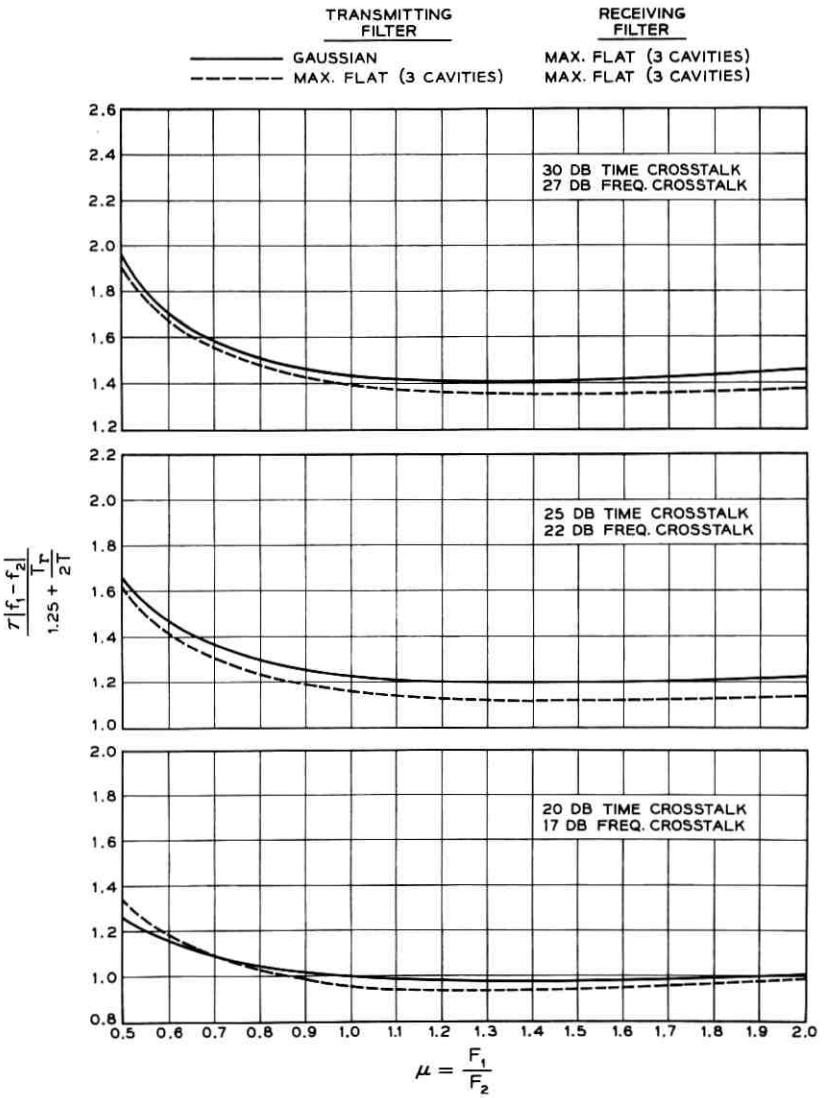


Fig. 7 — $\tau\,|f_1 - f_2|$ curves for $\tau/2T = 1.25 + T_r/2T$.

$$G_1 \propto \tau|f_1 - f_2|20 \log \frac{1}{\sqrt{2}\sigma}$$

and

$$G_2 \propto \tau|f_1 - f_2| \frac{1}{\sqrt{2}\sigma},$$

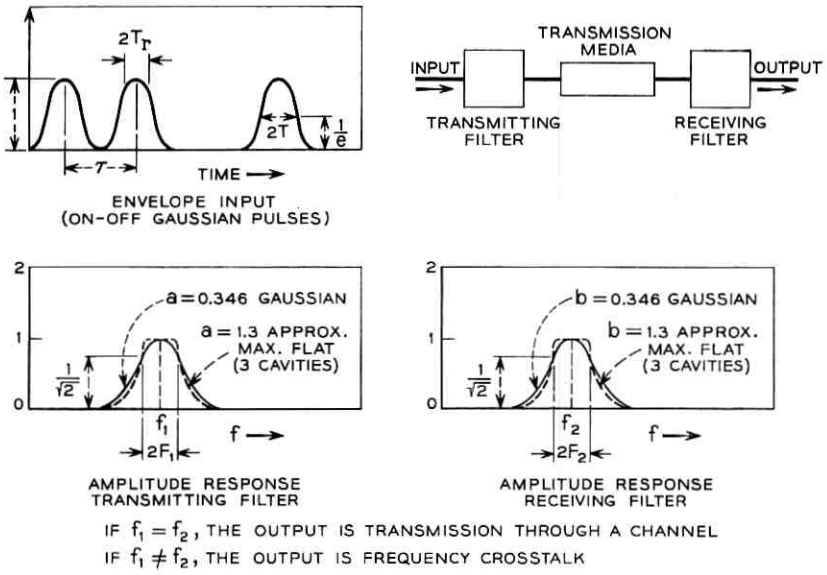


Fig. 8 — Definitions of symbols.

where 20 log $(1/\sqrt{2}\sigma)$ is the signal-to-noise in db and $1/\sqrt{2}\sigma$ is the ratio of rms signal and rms noise.

The different weighting functions were selected in order to introduce some idea about the influence of distance between successive repeaters. Since the amplitude of the received signal decays exponentially with the distance between terminals, for fixed transmitter and receiver G_1 decreases linearly with distance and G_2 decreases exponentially with distance.

Functions G_1 and G_2, obtained by multiplying the ordinates of each curve in Fig. 5 by the properly weighted ordinates of Fig. 4, have been plotted in Figs. 9, 10, 11, and 12. Each figure contains two sets of curves, and in each set the three curves exhibit minimums individualized by the coordinates probability of error and time crosstalk per tail. Those coordinates identify three points of an optimization curve that could be plotted in Fig. 3. For clarity, part of Fig. 3 has been reproduced in Fig. 13, omitting the detailed information on frequency crosstalk and optimum slicing level. In Fig. 13 we have plotted the lines joining each set of three points rather than the points themselves. Since there are eight sets of curves in Figs. 9 through 12, we get eight lines of optimum design in Fig. 13. Four of them correspond to the minimization of G_1 (signal-

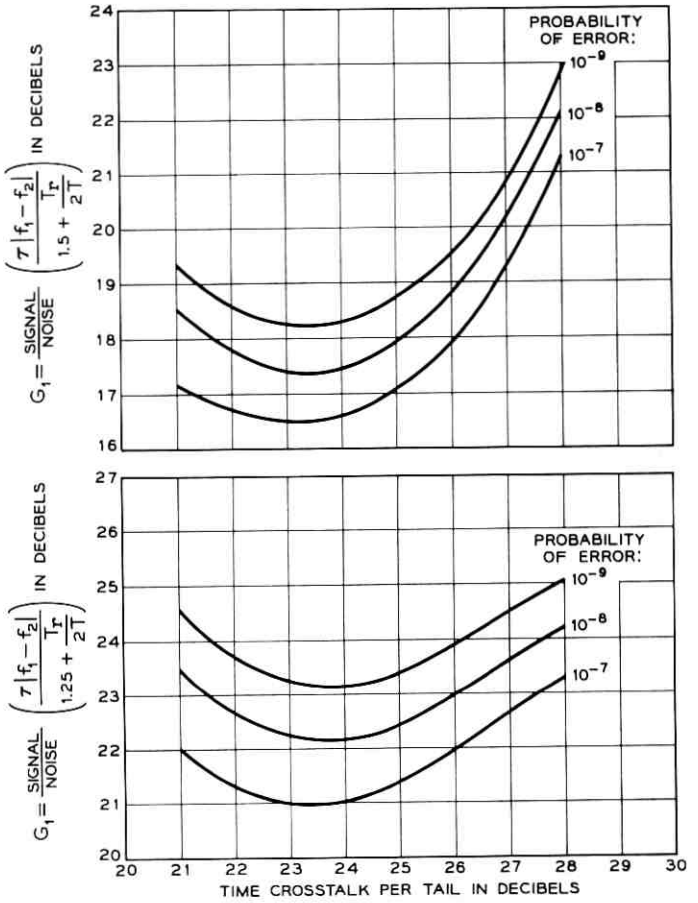


Fig. 9 — Minimization of G_1 , transmitting and receiving filters approximately maximally flat (three cavities).

to-noise in db) and are clustered close to the line defined by the parameter time crosstalk per tail 24 db; the other four lines of optimum design correspond to the minimization of G_2 (rms signal to rms noise) and are close to the line defined by the parameter time crosstalk per tail 26 db. In each cluster, the two solid lines are optimization curves for two systems, both with maximally flat (three cavities) transmitting and receiving filters but with different input pulse width $2T$ and sampling time $2T_r$; the two dashed lines are optimization curves for two systems both with Gaussian transmitting filters and maximally flat (three

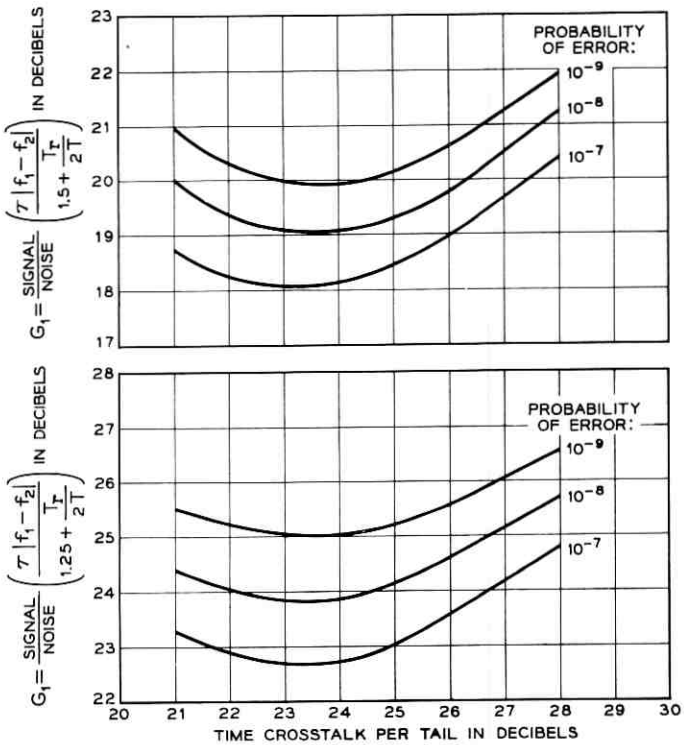


Fig. 10 — Minimization of G_1, transmitting filter Gaussian, receiving filter approximately maximally flat (three cavities).

cavities) receiving filters, but with different input pulse width $2T$ and sampling time $2T_r$.

In spite of the different dependence of signal-to-noise in G_1 and G_2, the different shapes of transmitting and receiving transfer characteristics, and the different input pulse widths and sampling times, all curves of optimum design are rather close to each other, and they are essentially located in the region where rms noise and rms crosstalk are comparable.

The optimum design lines are in general slightly steeper than the constant time crosstalk per tail lines, and, in particular, the two extremes of each of these eight design lines correspond to a change of 100 in the probability of error, around 1.5 db in signal-to-noise and only a few tenths of a db in time crosstalk per tail. This means that, once an optimum system has been built, the probability of error can be changed substantially by modifying only the signal-to-noise ratio (which is easy to do) and, in spite of this change, the system will remain close to the optimum design.

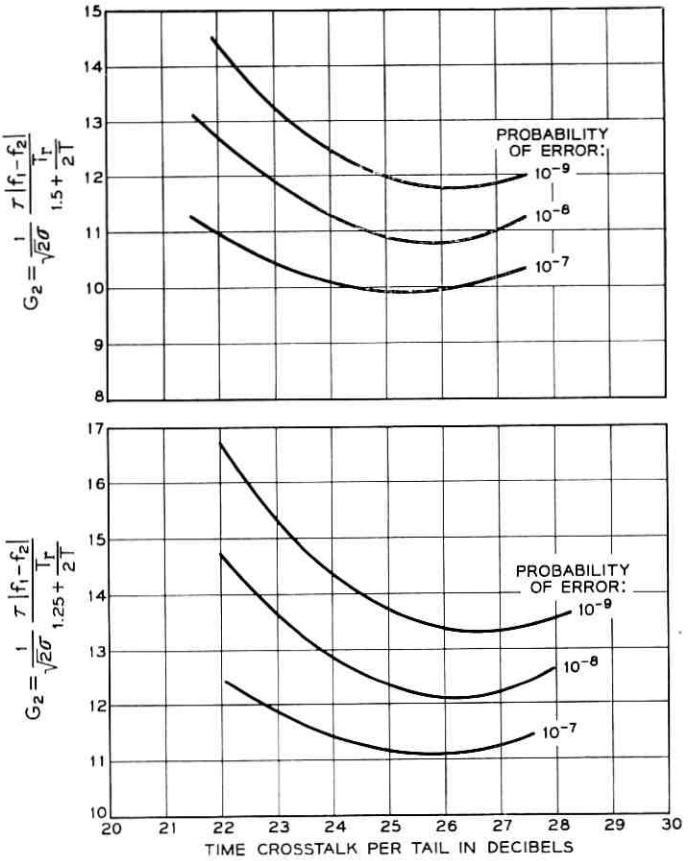


Fig. 11 — Minimization of G_2, transmitting filter Gaussian, receiving filter approximately maximally flat (three cavities).

V. CONCLUSIONS

The probability of error in the envelope detection of an RF signal embedded in Gaussian noise, time crosstalk from two neighboring pulses, and frequency crosstalk from an adjacent channel has been calculated and plotted in Fig. 3.

Also, two kinds of optimum operating conditions have been postulated which yield the results shown in Fig. 13. These conditions allow one to design a system in such a way that some minimization of time spacing between successive pulses, frequency spacing between adjacent channels, and signal-to-noise ratio is achieved.

An example of design is this: Suppose we want an optimally designed

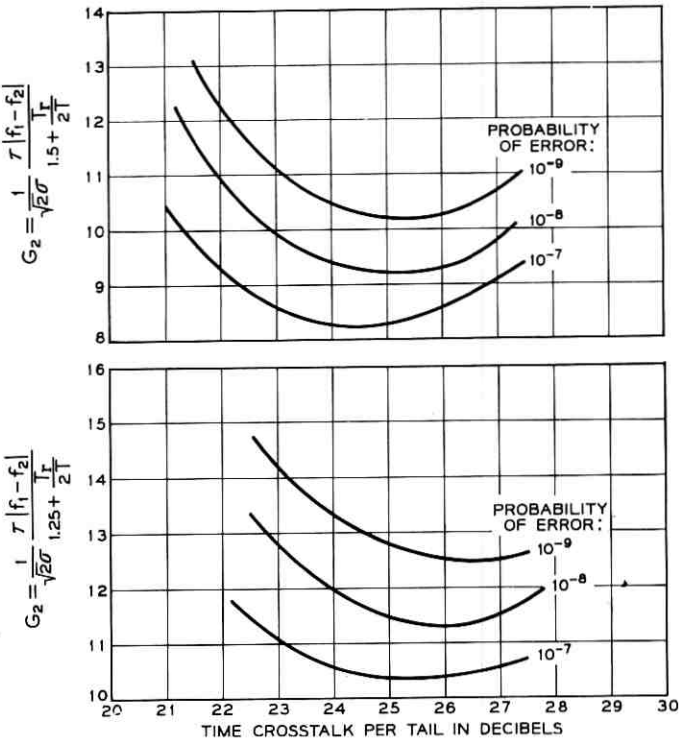


Fig. 12 — Minimization of G_2, transmitting and receiving filters approximately maximally flat (three cavities).

system that has a probability of error of 10^{-8}. We don't know which of the two criteria of optimization developed in this paper is closer to reality, and, because of that lack of knowledge, we adopt the middle of the road for the example. In Fig. 13, the ordinate 10^{-8} and the middle of the optimum design region establish that the system should have a signal-to-noise level of about 20.6 db and a time crosstalk per tail of 25 db. This last datum is enough to enter in the companion paper[4] and to complete the design of the system.

VI. ACKNOWLEDGMENT

I am indebted to Mrs. C. L. Beattie for carrying out the calculations necessary for Fig. 3.

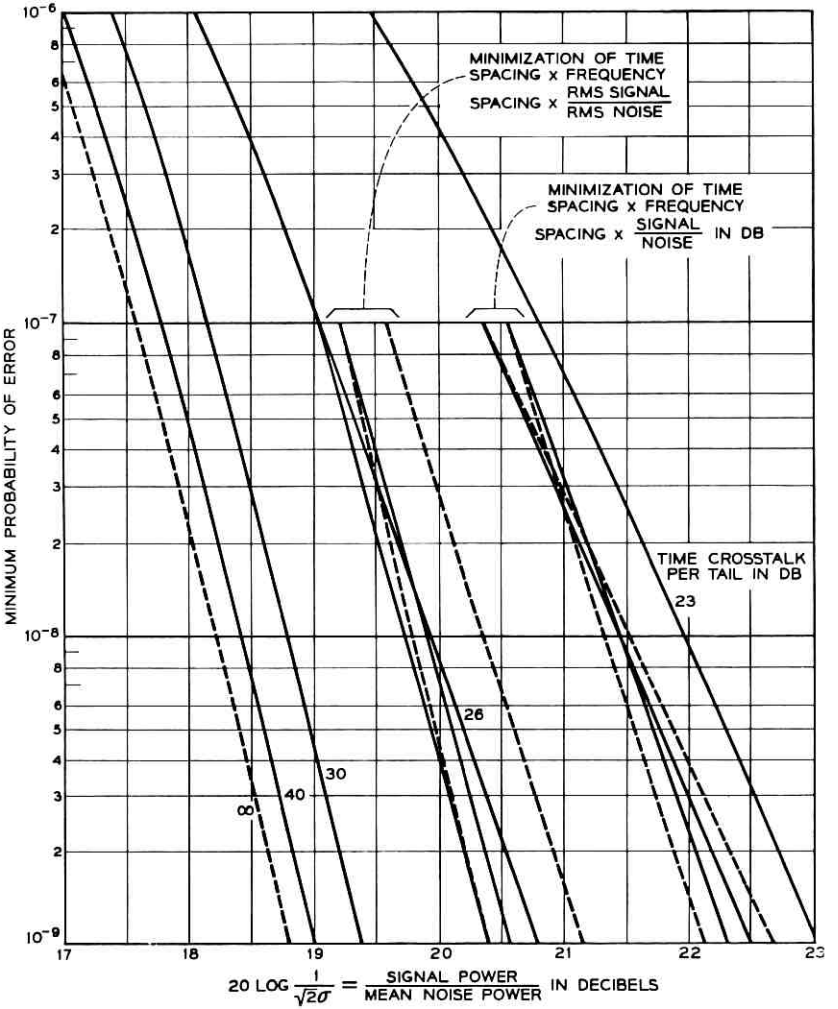


Fig. 13 — Optimum design region: solid curves — transmitting and receiving filters approximately maximally flat; dashed curves — transmitting filter Gaussian, receiving filter approximately maximally flat.

APPENDIX A

Crosstalk Between Adjacent Frequency Channels

We want to determine the frequency crosstalk between adjacent channels in order to find what arrangement of filters is the most favorable.

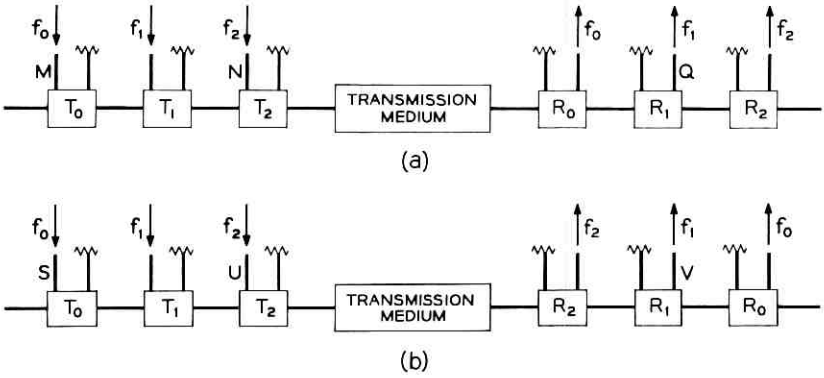


Fig. 14 — Arrangements of transmitting and receiving filters in a system.

The systems we shall deal with, shown in Figs. 14(a) and 14(b), differ in the order in which the bands are dropped. Each system consists of many transmitting and receiving filters, of which only three transmitting filters — T_0, T_1, and T_2 — and three receiving filters — R_0, R_1, and R_2 — are drawn, because we assume that the crosstalk in a receiver (R_1), comes essentially from the immediately neighboring channels.

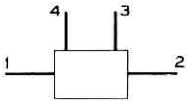


Fig. 15 — Transmitting or receiving filter.

For simplicity, we assume that, except for the frequency at which they are tuned, all the filters are similar to that shown in Fig. 15. They are constant-resistance, symmetrical, and reciprocal, and the transfer functions between terminals are given by the scattering matrix

$$S = \begin{vmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \\ S_{41} & S_{42} & S_{43} & S_{44} \end{vmatrix}$$

$$= \begin{vmatrix} 0 & i\sqrt{1-Y^2} & Y & 0 \\ i\sqrt{1-Y^2} & 0 & 0 & Y \\ Y & 0 & 0 & i\sqrt{1-Y^2} \\ 0 & Y & i\sqrt{1-Y^2} & 0 \end{vmatrix}. \tag{16}$$

Now we can calculate the maximum intensity of the crosstalk C_{MQ} between M and Q in Fig. 14(a) due to a pulse entering at M.

Assuming for simplicity that the system is phase equalized and that the input pulse has a $(\sin x)/x$ shape (rectangular spectrum), the maximum intensity of the crosstalk is given by

$$C_{MQ} = K \int_0^\infty G_0 Y_0 \sqrt{1 - Y_1^2} \sqrt{1 - Y_2^2} \sqrt{1 - Y_0^2}\, Y_1\, df, \quad (17)$$

where K is a constant of proportionality, G_0 is the rectangular spectrum of the input pulse centered at f_0, and $Y_0\sqrt{1 - Y_1^2}\sqrt{1 - Y_0}\, Y_1$ derived from (16), and Fig. 14(a) is the transfer function between M and Q. The subindices 0 and 1 refer to the center frequencies f_0 and f_1 of each scattering coefficient.

Following arguments similar to the preceding one, the maximum crosstalk intensities between N and Q in Fig. 14(a) and between S and V and between U and V in Fig. 14(b), are

$$C_{NQ} = K \int_0^\infty G_2 Y_2 \sqrt{1 - Y_0^2}\, Y_1\, df, \quad (18)$$

$$C_{SV} = K \int_0^\infty G_0 Y_0 \sqrt{1 - Y_1^2}\,(1 - Y_2^2) Y_1\, df, \quad (19)$$

$$C_{UV} = K \int_0^\infty G_2 Y_2 \sqrt{1 - Y_2^2}\, Y_1\, df. \quad (20)$$

The factors involved in each integrand of (17) through (20) have been plotted in Figs. 16(a), (b), (c), and (d). The integrands of (17) through (20) are obtained by multiplying the curves in Fig. 16(a) through Fig. 16(d) respectively. The results which happen to be the output spectra are plotted in Fig. 17(a) through Fig. 17(d).

The integration of these curves with respect to frequency, that is, the areas between the curves and the frequency axes, are, because of (17) through (20), proportional to the maximum intensities of the crosstalks.

Comparing these areas we deduce

$$C_{MQ} = C_{UV}, \quad (21)$$

$$C_{NQ} = C_{SV}, \quad (22)$$

$$C_{MQ} \ll C_{NQ}, \quad (23)$$

$$C_{UV} \ll C_{SV}. \quad (24)$$

The first two equations show that the total crosstalk in the system of

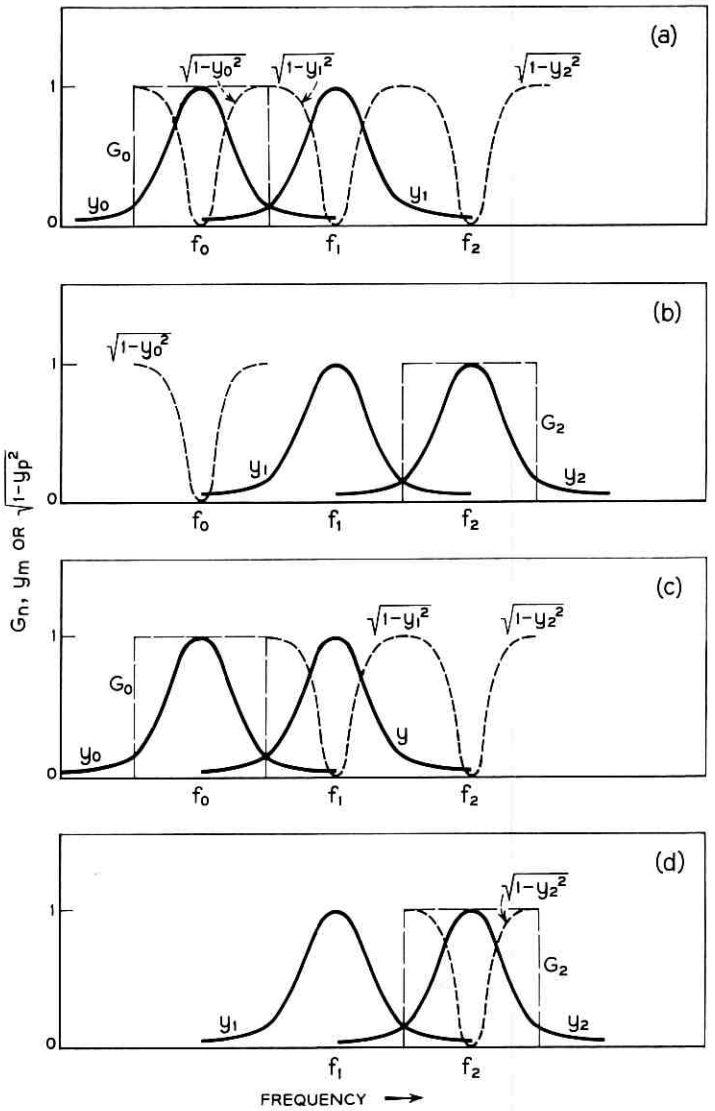


Fig. 16 — Factors involved in integrands of (17) through (20); G_n = input signal spectra; Y_m , $\sqrt{1 - Y_p^2}$ = scattering coefficients.

Fig. 14(a) is the same as the total crosstalk in the system of Fig. 14(b). Furthermore, from (23) and (24) we deduce that the total crosstalk in either system comes from the superposition of two signals, of which one is negligible compared to the other.

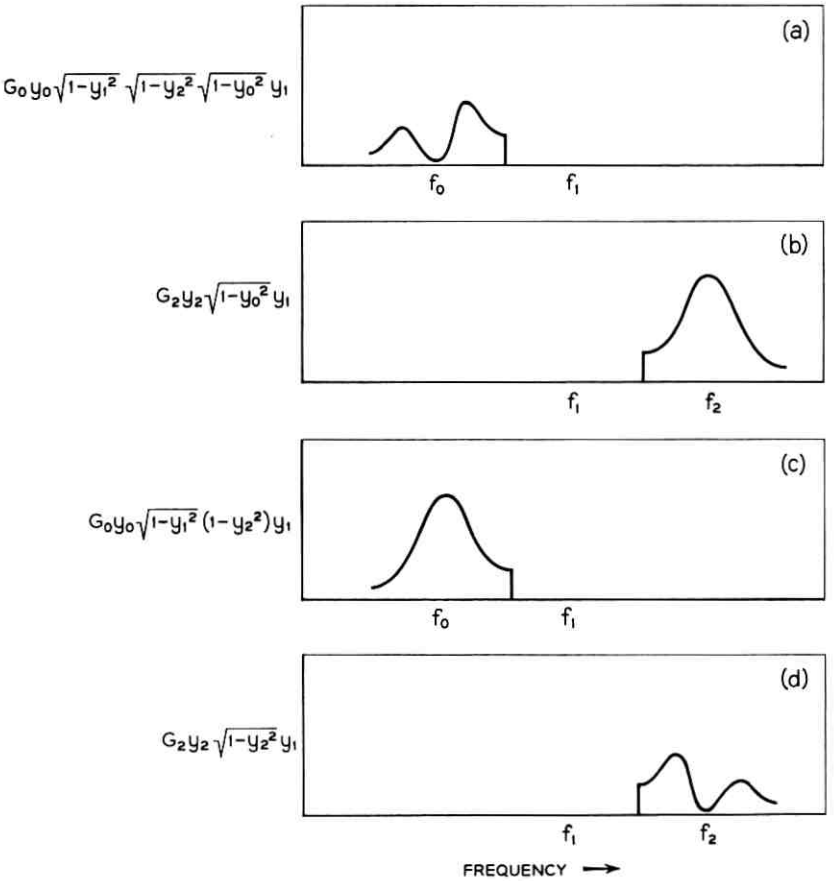


Fig. 17 — Spectra of crosstalk signals.

APPENDIX B

Bivariate Density Distribution

An on-off pulse embedded in unwanted crosstalk and noise is represented vectorially by

$$S = A + \rho_T e^{i\theta_1} + \rho_T e^{i\theta_2} + \rho_F e^{i\theta_3} + \text{Gaussian noise.} \qquad (25)$$

The amplitude A of the RF pulse is unity if the pulse is on, and zero if the pulse is off; the phase of this vector is selected zero, as reference. Time crosstalk is represented by two vectors of the same modulus ρ_T and arbitrary phases θ_1 and θ_2. Frequency crosstalk is represented by a

vector of modulus ρ_F and arbitrary phase θ_3. Each one of these three vectors, since it originated from binary pulses, has equal probability of being present or not. The phases θ_1, θ_2, and θ_3 have a constant probability of acquiring any value between zero and 2π.

We want to calculate the density distribution of S, and we know the density distribution of each one of its five uncorrelated terms:

$$p_1(x,y) = \delta(x - A)\delta(y), \tag{26}$$

$$p_2(x,y) = p_3(x,y) = \frac{\delta(\sqrt{x^2 + y^2})}{4\pi\sqrt{x^2 + y^2}} + \frac{\delta(\sqrt{x^2 + y^2} - \rho_T)}{4\pi\,\rho_T}, \tag{27}$$

$$p_4(x,y) = \frac{\delta(\sqrt{x^2 + y^2})}{4\pi\sqrt{x^2 + y^2}} + \frac{\delta(\sqrt{x^2 + y^2} - \rho_F)}{4\pi\,\rho_F}, \tag{28}$$

$$p_5(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}, \tag{29}$$

where $\delta(z)$ is the Dirac delta function and σ^2 is the variance, which in this particular problem measures the mean noise power.

It is known that the distribution $p(x,y)$ of the sum S of independent terms is equal to the inverse transform of the product of the double Fourier transform of the density distribution of each term of the sum.[3]

The double Fourier transform of a distribution $p_n(x,y)$ is, by definition

$$C_n = \iint_{-\infty}^{\infty} e^{i(\xi x+\eta y)} p_n(x,y)\, dx\, dy. \tag{30}$$

Replacing $p_n(x,y)$ by (26) through (29) and integrating,

$$C_1 = e^{i\xi A}, \tag{31}$$

$$C_2 = C_3 = \tfrac{1}{2}[1 + J_0(\rho_T\sqrt{\xi^2 + \eta^2})], \tag{32}$$

$$C_4 = \tfrac{1}{2}[1 + J_0(\rho_F\sqrt{\xi^2 + \eta^2})], \tag{33}$$

$$C_5 = e^{-(\sigma^2/2)(\xi^2+\eta^2)}, \tag{34}$$

where J_0 is the Bessel function of first order and kind.

The inverse transform of the product of these functions is the distribution $p(x,y)$ of the signal S we were looking for:

$$p(x,y) = \frac{1}{32\pi^2} \iint_{-\infty}^{\infty} e^{-i[\xi(x-A)+\eta y]-(\sigma^2/2)(\xi^2+\eta^2)}[1 + J_0(\rho_T\sqrt{\xi^2 + \eta^2})]^2 \cdot[1 + J_0(\rho_F\sqrt{\xi^2 + \eta^2})]\, d\xi\, d\eta. \tag{35}$$

The distribution $p(x,y)$ is the summation of several double integrals to be evaluated. The most general of them is

$$W = \iint_{-\infty}^{\infty} e^{-i[\xi(x-A)+\eta y]-(\sigma^2/2)(\xi^2+\eta^2)} J_0(\rho_1\sqrt{\xi^2+\eta^2}) \cdot J_0(\rho_2\sqrt{\xi^2+\eta^2}) J_0(\rho_3\sqrt{\xi^2+\eta^2})\, d\xi\, d\eta. \tag{36}$$

Replacing each Bessel function by an integral expression,[5]

$$W = \frac{1}{(2\pi)^3} \iiint_0^{2\pi} d\alpha\, d\beta\, d\gamma \iint_{-\infty}^{\infty} e^{-i\xi[(x-A)-\rho_1\cos\alpha-\rho_2\cos\beta-\rho_3\cos\gamma]} \cdot e^{-i\eta(y-\rho_1\sin\alpha-\rho_2\sin\beta-\rho_3\sin\gamma)-(\sigma^2/2)(\xi^2+\eta^2)}\, d\xi\, d\eta. \tag{37}$$

The two integrations from $-\infty$ to ∞ are known Fourier transforms, and (37) becomes

$$W = \frac{1}{(2\pi\sigma)^2} \cdot \iiint_0^{2\pi} e^{[(x-A-\rho_1\cos\alpha-\rho_2\cos\beta-\rho_3\cos\gamma)^2-(y-\rho_1\sin\alpha-\rho_2\sin\beta-\rho_3\sin\gamma)^2]/2\sigma^2} d\alpha\, d\beta\, d\gamma. \tag{38}$$

By changing variables,

$$\begin{aligned} x - A &= r\cos\varphi, \\ y &= r\sin\varphi \end{aligned} \tag{39}$$

the exponent can be rearranged:

$$W = \frac{e^{-(r^2+\rho_1^2+\rho_2^2+\rho_3^2)/2\sigma^2}}{(2\pi\sigma)^2} \cdot \iiint_0^{2\pi} e^{[r\rho_1\cos(\alpha-\varphi)+r\rho_2\cos(\beta-\varphi)+r\rho_3\cos(\gamma-\varphi) -\rho_1\rho_2\cos(\alpha-\beta)-\rho_1\rho_3\cos(\alpha-\gamma)-\rho_2\rho_3\cos(\beta-\gamma)]/\sigma^2} d\alpha\, d\beta\, d\gamma. \tag{40}$$

We start integrating with respect to α. The integral to be solved is essentially

$$W_\alpha = \int_0^{2\pi} e^{\rho_1[r\cos(\alpha-\varphi)-\rho_2\cos(\alpha-\beta)-\rho_3\cos(\alpha-\gamma)]/\sigma^2} d\alpha. \tag{41}$$

The exact result is[3]

$$W_\alpha = 2\pi I_0\left(\frac{\rho_1}{\sigma^2}\sqrt{\begin{matrix}[r-\rho_2\cos(\varphi-\beta)-\rho_3\cos(\varphi-\gamma)]^2 \\ +[\rho_2\sin(\varphi-\beta)+\rho_3\sin(\varphi-\gamma)]^2\end{matrix}}\right), \quad (42)$$

but if we carry this expression to (40) the integration with respect to β and γ becomes extremely complicated.

A substantial simplification can be obtained if we consider first that we are interested only in the tails of the distributions, and consequently

$$r \gg \begin{Bmatrix}\rho_1 \\ \rho_2 \\ \rho_3\end{Bmatrix}. \quad (43)$$

Second, for

$$\left.\begin{matrix}\frac{r\rho_1}{\sigma^2} \\ \frac{r\rho_2}{\sigma^2} \\ \frac{r\rho_3}{\sigma^2}\end{matrix}\right\} \gg 1, \quad (44)$$

which is the only nontrivial case, the main contribution to the triple integral (40) comes from values of the integrating variables

$$\left.\begin{matrix}\alpha \\ \beta \\ \gamma\end{matrix}\right\} \cong \varphi. \quad (45)$$

Because of (43) and (45), expression (41) can be reduced to

$$W_\alpha \cong e^{-\rho_1[(\rho_2+\rho_3)/\sigma^2]}\int_0^{2\pi} e^{(\rho_1 r/\sigma^2)\cos(\alpha-\varphi)}\,d\varphi \quad (46)$$

and, after performing the integration,

$$W_\alpha \cong 2\pi e^{-\rho_1[(\rho_2+\rho_3)/\sigma^2]} I_0\left(\frac{\rho_1 r}{\sigma^2}\right). \quad (47)$$

The reader may also derive this result from (42), (43), and (45). Substituting this result of the integration on α, in (40),

$$W = \frac{e^{(r^2+\rho_1{}^2+\rho_2{}^2+\rho_3{}^2+2\rho_1\rho_2+2\rho_1\rho_3)/2\sigma^2}}{2\pi\sigma^2} I_0\left(\frac{\rho_1 r}{\sigma^2}\right) \cdot \int_0^{2\pi}\int e^{[r\rho_2 \cos(\beta-\varphi)+r\rho_3 \cos(\gamma-\varphi)-\rho_2\rho_3 \cos(\beta-\gamma)]/\sigma^2}\, d\beta\, d\gamma. \tag{48}$$

Now we perform the integration on β. The integral to be solved is essentially

$$W_\beta = \int_0^{2\pi} e^{(\rho_2/\sigma^2)[r\cos(\beta-\varphi)-\rho_3\cos(\beta-\gamma)]}\, d\beta. \tag{49}$$

Following the same reasoning used to integrate W_α, in (41), the approximate result is

$$W_\beta \cong 2\pi e^{-\rho_2\rho_3/\sigma^2} I_0\left(\frac{\rho_2 r}{\sigma^2}\right). \tag{50}$$

After substituting in (48) and performing the integration on γ, W is

$$W = \frac{2\pi}{\sigma^2} e^{-[r^2+(\rho_1+\rho_2+\rho_3)^2]/2\sigma^2} I_0\left(\frac{\rho_1 r}{\sigma^2}\right) I_0\left(\frac{\rho_2 r}{\sigma^2}\right) I_0\left(\frac{\rho_3 r}{\sigma^2}\right). \tag{51}$$

Substituting this generic result in (35), the density distribution of the signal S is obtained:

$$p(x,y) = \frac{e^{-r^2/2\sigma^2}}{16\pi\sigma^2}\left[1 + 2e^{-\rho_T{}^2/2\sigma^2} I_0\left(\frac{\rho_T r}{\sigma^2}\right) + e^{-2\rho_T{}^2/\sigma^2} I_0{}^2\left(\frac{\rho_T r}{\sigma^2}\right) + e^{-\rho_F{}^2/2\sigma^2} I_0\left(\frac{\rho_F r}{\sigma^2}\right) + 2e^{-(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\frac{\rho_T r}{\sigma^2}\right) I_0\left(\frac{\rho_F r}{\sigma^2}\right) + e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0{}^2\left(\frac{\rho_T r}{\sigma^2}\right) I_0\left(\frac{\rho_F r}{\sigma^2}\right)\right]. \tag{52}$$

APPENDIX C

Evaluation of Probabilities of Error

Case 1: "Pulse On"

We want to evaluate the integral

$$P_1 = \int_{a_1} p_1(x,y)\, dx\, dy, \tag{53}$$

where a_1 is a circle of radius $\rho_0 \cong \frac{1}{2}$ with center at the origin of coordinates, and $p_1(x,y)$ is obtained from (52) by making

$$A = 1 \tag{54}$$

in the expression

$$r = \sqrt{(x - A)^2 + y^2}. \tag{55}$$

Adopting the following change of variables:

$$\begin{aligned} x &= \rho \cos \psi, \\ y &= \rho \sin \psi, \end{aligned} \tag{56}$$

P_1 becomes

$$P_1 = \int_{-\pi}^{\pi} \int_0^{\rho_0} p_1(\rho \cos \psi, \rho \sin \psi)\, d\rho\, d\psi. \tag{57}$$

The most general term of the integration is proportional to

$$U_1 = \int_{-\pi}^{\pi} \int_0^{\rho_0} e^{-r^2/2\sigma^2} I_0\left(\frac{\rho_1 r}{\sigma^2}\right) I_0\left(\frac{\rho_2 r}{\sigma^2}\right) I_0\left(\frac{\rho_3 r}{\sigma^2}\right) \rho\, d\psi\, d\rho, \tag{58}$$

where

$$r = \sqrt{1 + \rho^2 - 2\rho \cos \psi}. \tag{59}$$

We simplify the integrand. Notice first that, since

$$\rho_0 \cong \tfrac{1}{2}, \tag{60}$$

we deduce, from (59),

$$r > \tfrac{1}{2}$$

independently of ψ; second, the range of interest for σ is

$$\sigma \ll \rho_0 . \tag{61}$$

Therefore, the exponent in (58) is

$$\frac{r^2}{2\sigma^2} \gg 1. \tag{62}$$

Because of this inequality and because, for a small variation of r, the exponential in (58) varies much faster than the modified Bessel functions, most of the contribution to the integral comes from values of the variable close to those that minimize r,

$$\begin{aligned} \psi &= 0, \\ \rho &= \rho_0 , \end{aligned}$$

and (58) becomes

$$U_1 \cong I_0\left(\rho_1 \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_2 \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_3 \frac{1 - \rho_0}{\sigma^2}\right) D, \tag{63}$$

where

$$D = \int_{-\pi}^{\pi} \int_0^{\rho_0} e^{-(1+\rho^2-2\rho \cos \psi)/2\sigma^2} \rho\, d\psi\, d\rho. \tag{64}$$

Integrating with respect to ψ,

$$D = 2\pi \int_0^{\rho_0} e^{-(1+\rho^2)/2\sigma^2} I_0\left(\frac{\rho}{\sigma^2}\right) \rho \, d\rho. \tag{65}$$

Since the exponential varies much faster than the rest of the integrand, most of the contribution to the integral comes from $\rho \cong \rho_0$, and, because of (60) and (61), $I_0(\rho/\sigma^2)$ can be replaced by its asymptotic expansion. Consequently,

$$D \cong \sqrt{2\pi\rho_0}\, \sigma \int_0^{\rho_0} e^{-(1-\rho)^2/2\sigma^2} \, d\rho. \tag{66}$$

Integrating,

$$D = \sqrt{2\pi\rho_0}\, \frac{\sigma^3}{1 - \rho_0} e^{-(1-\rho_0)^2/2\sigma^2}; \tag{67}$$

for compactness, this can be rewritten

$$D \cong \sigma^2 K_0\left[\frac{(1 - \rho_0)^2}{2\sigma^2}\right], \tag{68}$$

where K_0 is the modified Bessel function of the second kind.

Substituting (68) in (63), the general term of the integration (57) is

$$U_1 = \sigma^2 K_0\left[\frac{(1 - \rho_0)^2}{2\sigma^2}\right] I_0\left(\rho_1 \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_2 \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_3 \frac{1 - \rho_0}{\sigma^2}\right) \tag{69}$$

and the probability of error for the "pulse on" condition is

$$\begin{aligned} P_1 = \frac{K_0\left[\frac{(1 - \rho_0)^2}{2\sigma^2}\right]}{16\pi} \Bigg[& 1 + 2e^{-\rho_T^2/2\sigma^2} I_0\left(\rho_T \frac{1 - \rho_0}{\sigma^2}\right) \\ & + e^{-2\rho_T^2/\sigma^2} I_0^2\left(\rho_T \frac{1 - \rho_0}{\sigma^2}\right) + e^{-\rho_F^2/2\sigma^2} I_0\left(\rho_F \frac{1 - \rho_0}{\sigma^2}\right) \\ & + 2e^{(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\rho_T \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_F \frac{1 - \rho_0}{\sigma^2}\right) \\ & + e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0^2\left(\rho_T \frac{1 - \rho_0}{\sigma^2}\right) I_0\left(\rho_F \frac{1 - \rho_0}{\sigma^2}\right)\Bigg]. \end{aligned} \tag{70}$$

Case 2: "Pulse Off"

We want to evaluate the integral

$$P_2 = \int_{a_2} p_2(x,y) \, dx \, dy, \tag{71}$$

where a_2 is the surface outside of a circle of radius ρ_0 with center at the origin of coordinates, and $p_2(x,y)$ is obtained from (52) by making

$$A = 0 \tag{72}$$

in the expression

$$r = \sqrt{(x - A)^2 + y^2}.$$

After changing the variables according to (56), the ψ dependence disappears from $p_2(x,y)$ and the probability of error (71) is

$$P_2 = 2\pi \int_{\rho_0}^{\infty} p_2(r) r \, dr. \tag{73}$$

The most general term of this integral is proportional to

$$U_2 = \int_{\rho_0}^{\infty} e^{-r^2/2\sigma^2} I_0\left(\frac{\rho_1 r}{\sigma^2}\right) I_0\left(\frac{\rho_2 r}{\sigma^2}\right) I_0\left(\frac{\rho_3 r}{\sigma^2}\right) r \, dr. \tag{74}$$

Over the range of integration,

$$\frac{r^2}{2\sigma^2} \gg 1; \tag{75}$$

also, for a small variation of r, the exponential varies much faster than the modified Bessel functions. Consequently,

$$\begin{aligned} U_2 &\cong I_0\left(\frac{\rho_1\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_2\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_3\rho_0}{\sigma^2}\right) \int_{\rho_0}^{\infty} e^{-r^2/2\sigma^2} r \, dr \\ &\cong \sigma^2 e^{-{\rho_0}^2/2\sigma^2} I_0\left(\frac{\rho_1\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_2\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_3\rho_0}{\sigma^2}\right). \end{aligned} \tag{76}$$

Substituting this result in (73), we get

$$\begin{aligned} P_2 = \frac{e^{-{\rho_0}^2/2\sigma^2}}{8} \Bigg[& 1 + 2e^{-{\rho_T}^2/2\sigma^2} I_0\left(\frac{\rho_T\rho_0}{\sigma^2}\right) + e^{-2{\rho_T}^2/\sigma^2} I_0^2\left(\frac{\rho_T\rho_0}{\sigma^2}\right) \\ & + e^{-{\rho_F}^2/2\sigma^2} I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right) + 2e^{-(\rho_T+\rho_F)^2/2\sigma^2} I_0\left(\frac{\rho_T\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right) \\ & + e^{-(2\rho_T+\rho_F)^2/2\sigma^2} I_0^2\left(\frac{\rho_T\rho_0}{\sigma^2}\right) I_0\left(\frac{\rho_F\rho_0}{\sigma^2}\right) \Bigg]. \end{aligned} \tag{77}$$

REFERENCES

1. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, 1954, p. 1209.
2. Sunde, E. D., Ideal Binary Pulse Transmission by AM and FM, B.S.T.J., **37**, 1959, p. 1357.
3. Bennett, W. R., Methods of Solving Noise Problems, Proc. I.R.E., **44**, 1956, p. 609.
4. Marcatili, E. A., Time and Frequency Crosstalk in Pulse-Modulated Systems, this issue, p. 951.
5. Jahnke, E., and Emde, F., *Tables of Functions*, Dover Publications, New York, 1943, p. 149.

Time and Frequency Crosstalk in Pulse-Modulated Systems

By E. A. MARCATILI

(Manuscript received November 2, 1960)

The time and frequency crosstalk between Gaussian RF pulses sent via adjacent frequency channels over the same transmission medium is calculated. Shapes of the transfer characteristics of the transmitting and receiving filters vary from Gaussian to approximately that of a third-order maximally flat filter. The results permit one to design the transmitting and receiving transfer characteristics of adjacent PCM channels in such a way that the product of pulse spacing and channel spacing is minimized.

I. INTRODUCTION

Consider a transmission medium in which many simultaneous messages travel in one single direction. Each message, consisting of coded on-off RF pulses (PCM), has its own carrier and occupies a separate frequency channel. This occurs, for example, in the proposed long distance waveguide communication system.[1] The transmitter is considered as a filter through which the pulses of a message are fed to the transmission medium and the receiver as a filter that selectively couples the transmission medium to a detector.

The problem is to design these filters in such a way that the communication medium handles information at the highest possible rate. This means that the channels must be close to each other, providing high frequency occupancy, and that each message must be made of pulses close to each other, providing high time occupancy. In other words, we want to minimize the product of channel spacing and pulse spacing. What prevents us from making this product arbitrarily small is that, in general, a reduction of pulse and channel spacings implies an increase of time and frequency crosstalk, and these values are fixed by other considerations: the signal-to-noise level and the probability of errors allowed in the system. We shall see how they enter the picture.

The detector of each receiver reconstructs a message by deciding

whether or not a pulse is in the assigned time slot. For that purpose the detector operates only during sampling times that occur at the pulse repetition rate. Suppose that at a given sampling time there is no pulse to be detected; the detector nevertheless receives a signal which is the superposition of three types of interferences: trailing and leading edges of neighboring pulses, or time crosstalk; leakage from pulses in neighboring channels, or frequency crosstalk; and, of course, the main offender, thermal noise. If this signal is bigger than the slicing level the detector decides that there is a pulse in that time slot, and an error is made. Similarly, suppose that there is a pulse to be detected, but that superposed on it are time and frequency crosstalk and noise. If the total amplitude is smaller than the slicing level, the detector decides that a pulse does not exist in that time slot, and another error results.

Quantitative relations between the probability of errors of a system and the three interferences, thermal noise, time crosstalk, and frequency crosstalk were established in the companion paper.[2] The system considered there was such that time crosstalk came from the trailing edge of the pulse in the preceding time slot and from the leading edge of the pulse in the following time slot, while frequency crosstalk came from a single pulse of one of the neighboring channels.

The main result derived from that paper was that for a given probability of error there is a particular set of noise, time crosstalk, and frequency crosstalk levels that simultaneously minimizes time occupancy, frequency occupancy, and signal-to-noise ratio. In order to obtain this result, two conditions were imposed on the crosstalks:

(a) Each crosstalk must contribute with equal weight to the probability of errors; for that purpose, the time crosstalk per tail must be approximately 3 db below the frequency crosstalk.

(b) Time and frequency crosstalks must minimize by themselves the time and frequency occupancy of the system.

How do we design a system capable of satisfying both conditions? Our objective in this paper is to answer that question.

The variables in the system at our disposal to fulfill the required time and frequency crosstalks are:

shape, width, and time spacing of the input pulses;

sampling time;

synchronization of pulses of neighboring channels;

shape, width, and frequency spacing of the transfer characteristics of the sending and receiving filters;

transfer characteristic of the transmission medium.

These are too many variables to include in the problem simultaneously, so we assume that:

(a) The input pulses are Gaussian. This is not critical if the shape of the output pulse is determined essentially by the filtering characteristic of the system, as it should be.

(b) Sampling times in all the channels occur simultaneously. This type of synchronization is pessimistic because it introduces the maximum frequency crosstalk. The most favorable condition is obtained in general by maintaining synchronization but displacing by half a pulse spacing the messages, and consequently the sampling time, of every other one of the successive channels. If no synchronization among channels exists, the frequency crosstalk has a constant probability of acquiring any value between those of the two previous cases.

(c) The transfer characteristic of each filter is the normalized sum (maximum amplitude equal to one) of two Gaussian curves displaced from each other, as shown in Appendix A. Displacing the Gaussian curves yields a transfer characteristic which varies from that of a Gaussian filter to approximately that of a maximally flat filter with three resonant cavities, Fig. 1. Since the input signal is Gaussian and the transfer characteristic is a product of Gaussian functions, the mathematics involved in the calculations is simple. The filters have been idealized in the sense that they have linear phase characteristics. This introduces a constant delay between the input and output signal that is ignored altogether, since it only represents displacement of the time origin. The effect of small departures from linear phase can be evaluated using perturbation techniques.[3]

(d) The transfer characteristic of the transmission medium is unity. This implies that time crosstalk due to multipath transmission and to imperfect phase equalization is negligible compared to that derived from the filters. In a real system it may be desirable to have these two contributions to time crosstalk be of the same order of magnitude.

II. SYSTEM ANALYSIS

Definitions of symbols (see Fig. 2):

$2T$ = width of input Gaussian pulse measured at $1/e$ of the maximum amplitude (8.686 db down),

$1/\tau$ = pulse repetition frequency,

$2T_r$ = sampling time,

$2F_1$ = bandwidth of transmitting filter measured at half power,

$2F_2$ = bandwidth of receiving filter measured at half power,

a = parameter defining shape of transmitting filter, which can be selected from Gaussian to approximately maximally flat (three resonant branches, maximally flat filter),

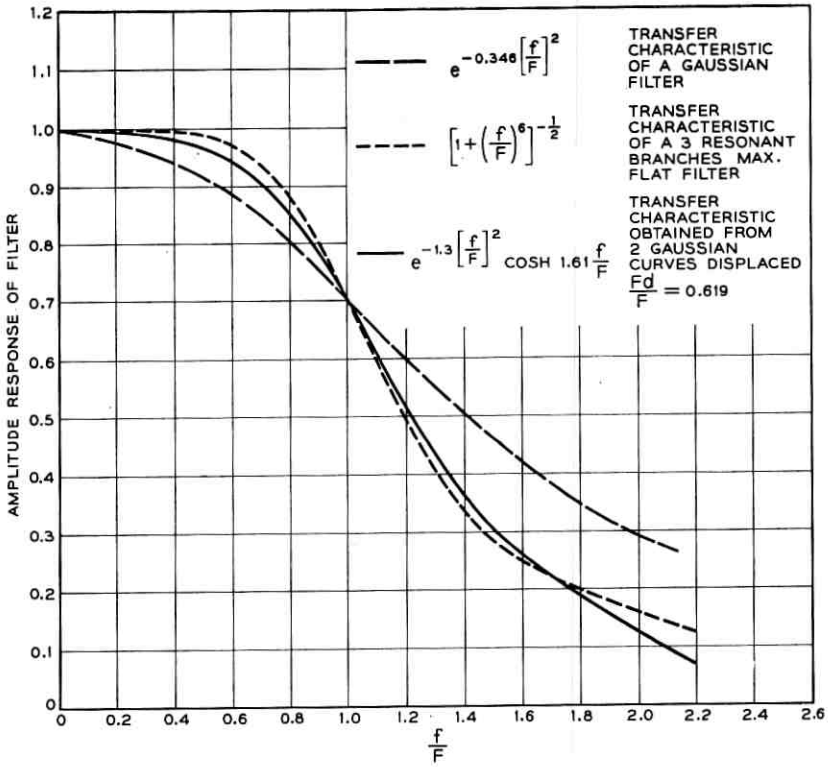


Fig. 1 — Transfer characteristics of different filters.

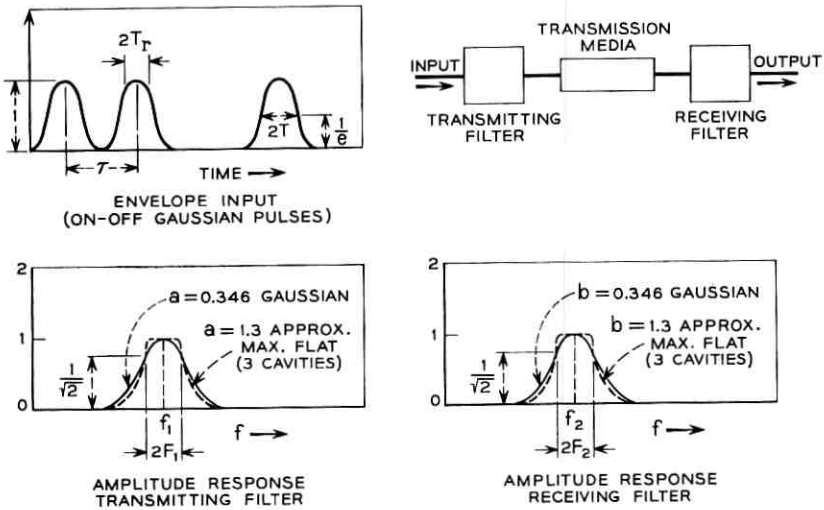


Fig. 2 — Definitions of symbols.

b = parameter defining shape of receiving filter,
f_1 = center frequency of transmitting filter,
f_2 = center frequency of receiving filter,
$\mu = F_1/F_2$ = ratio of transmitting to receiving bandwidth.
$\rho = |f_1 - f_2|/2F_2$ = ratio of channel spacing to the bandwidth of the receiving filter.

Finally, a useful parameter throughout our calculation is

$$k = \frac{4TF_1F_2}{\sqrt{F_1^2 + F_2^2}}.$$

If sending and receiving filters are Gaussian, k measures the pulse width $2T$ times the bandwidth of the system.

The envelope of the transient of a pulse through a channel and the maximum frequency crosstalk between two channels have been derived in Appendix B, equations (24) and (25). From them it is possible to calculate, in a way that will be described later, three functions that determine the best choice of transfer characteristics of the filters and channel spacing for a specified time and frequency crosstalk. Using the width of the input pulse $2T$ as a normalizing factor, those three functions are:

i. Normalized band spacing,

$$2T\,|f_1 - f_2| = \frac{\tau\,|f_1 - f_2|}{\theta + \dfrac{T_r}{2T}};$$

ii. Normalized receiver bandwidth,

$$4TF_2;$$

iii. Their ratio,

$$\rho = \frac{|f_1 - f_2|}{2F_2};$$

where

$$\theta = \frac{\tau - T_r}{2T}.$$

In practical cases, the sampling time $2T_r$ is small compared to the pulse spacing τ, and θ becomes the normalized pulse spacing.

The three functions i, ii, and iii are plotted in Figs. 3, 4, and 5 for $\theta = 1.25$, and in Figs. 6, 7, and 8 for $\theta = 1.5$. They are derived as follows:

(a) We plot, (24), the transient of a pulse through a channel, and the

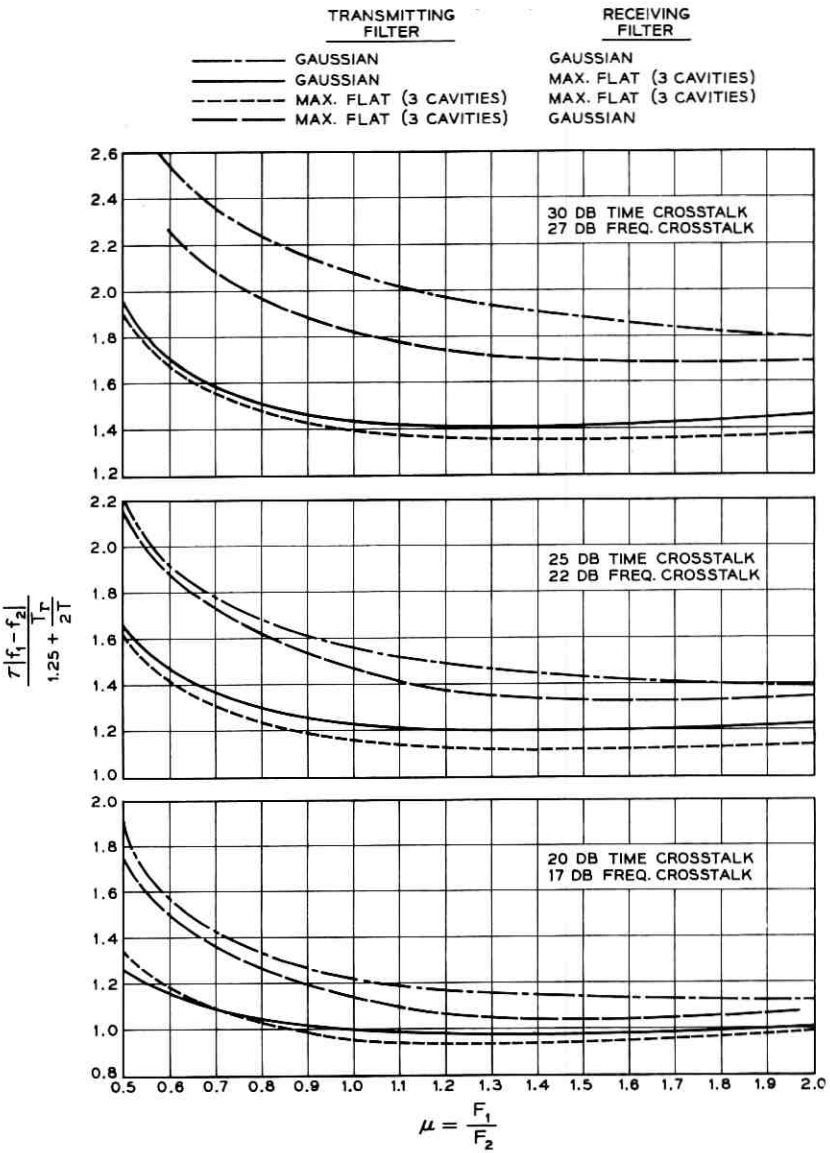


Fig. 3 — Normalized $\tau\,|f_1 - f_2|$ for $\theta = 1.25$.

maximum frequency crosstalk between neighboring channels, (25), for each possible combination of transfer characteristics of transmitting and receiving filters, and the ratio μ between sending and receiving bandwidths. Only one pair of these plots, Figs. 9 and 10, is included in this

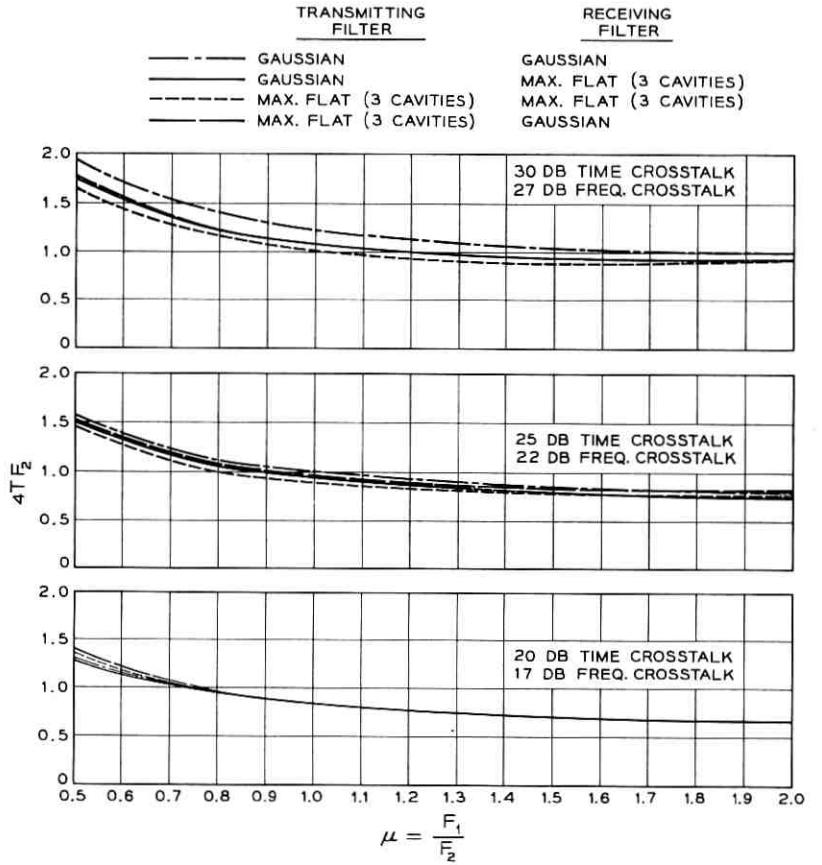


Fig. 4 — Normalized receiver bandwidth for $\theta = 1.25$.

paper to illustrate one example. The selected system has a Gaussian transmitting filter ($a = 0.346$, $m = 0$), approximately maximally flat receiving filter ($b = 1.6$, $n = 1.61$), and bandwidth ratio $\mu = 1$. Fig. 9 depicts the time response to a Gaussian pulse $2T$ wide at 8.686 db, through the two filters with variable over-all bandwidth of the system. The parameter

$$k = \frac{4TF_1}{\sqrt{1 + \mu^2}}$$

is proportional to that over-all bandwidth. Fig. 10 gives the maximum frequency crosstalk for the same Gaussian pulse through two filters

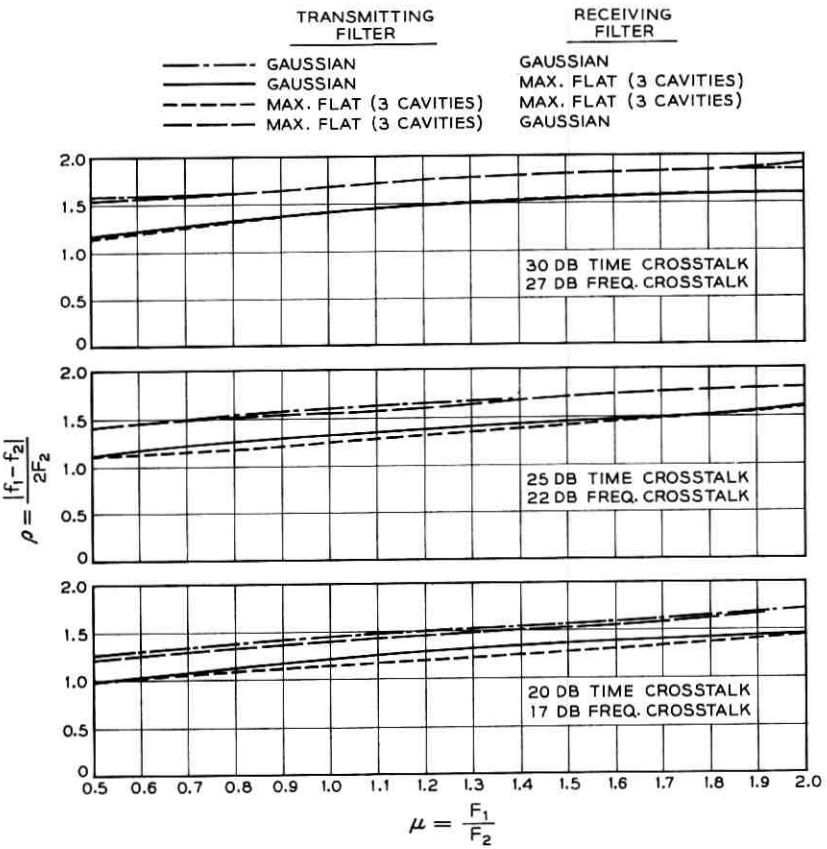


Fig. 5 — Ratio ρ for $\theta = 1.25$.

with fixed transfer characteristic shapes and bandwidth ratio. The abscissa

$$\rho = \frac{|f_1 - f_2|}{2F_2}$$

measures the channel spacing related to the receiver bandwidth, and the parameter is again k.

(b) From Fig. 9 we determine the smallest value of k, (smallest bandwidth of the system), compatible with given values of pulse spacing τ, sampling time $2T$, , and allowed time crosstalk, by following these steps:

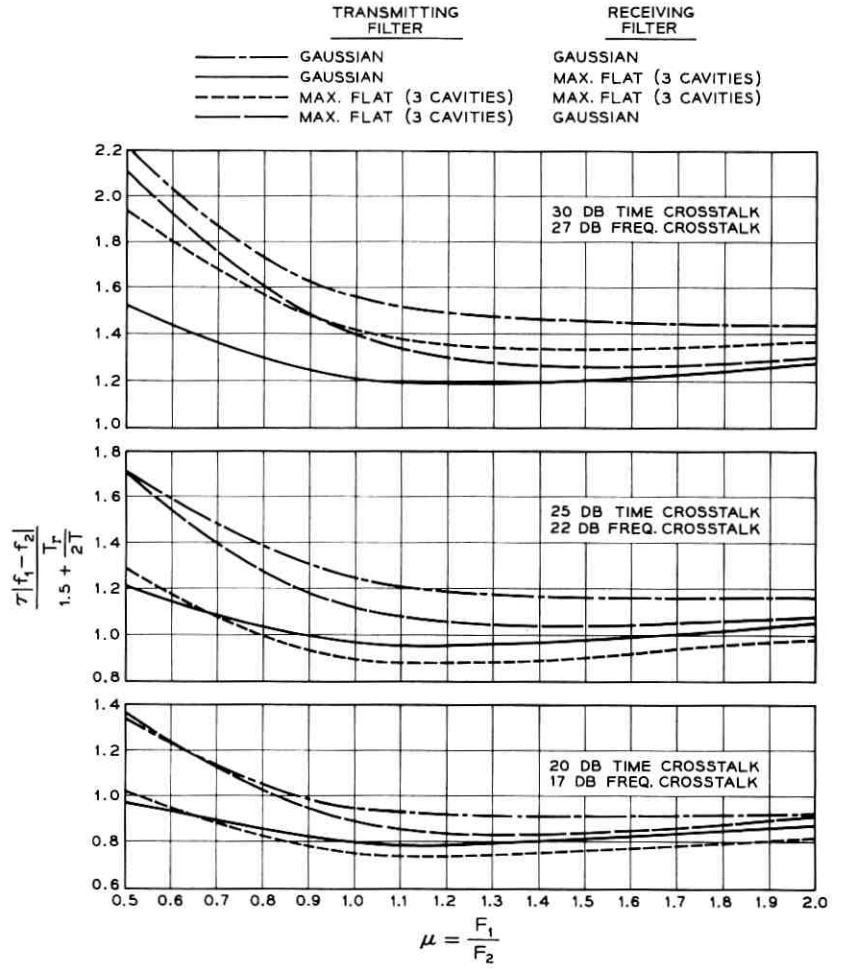


Fig. 6 — Normalized $\tau\,|f_1 - f_2|$ for $\theta = 1.5$.

1. Locate the normalized sampling time. This period, during which time crosstalk takes place, falls between the abscissa values

$$\frac{t}{2T} = \frac{\tau - T_r}{2T} \quad \text{and} \quad \frac{t}{2T} = \frac{\tau + T_r}{2T}.$$

2. Determine the ordinate that measures the allowed time crosstalk level (20-, 25-, and 30-db levels are indicated by dashed lines).

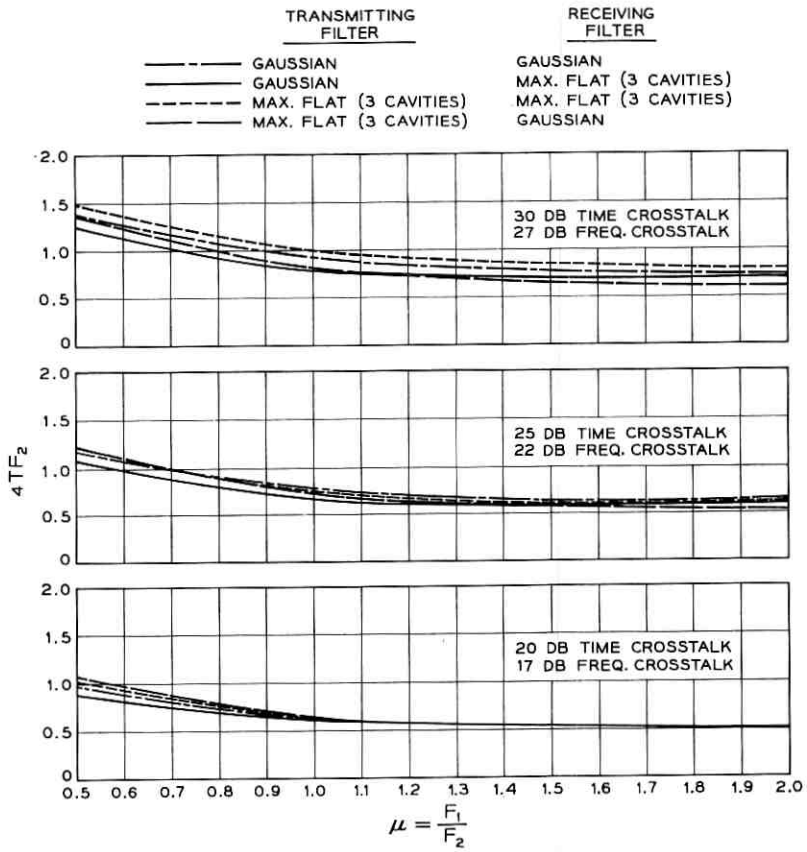


Fig. 7 — Normalized receiver bandwidth for $\theta = 1.5$.

3. Pick the curve with smallest k which, during the sampling time, is always below the allowed time crosstalk level.

(c) From Fig. 10 we determine the value of

$$\rho = \frac{|f_1 - f_2|}{2F_2}$$

by reading the abscissa of the point defined by the allowed frequency crosstalk ordinate (17-, 22-, and 27-db levels are indicated with dashed lines), and the curve characterized by the value of k deduced previously.

(d) The functions i, ii, and iii are derived after some arithmetic from μ, k, and ρ.

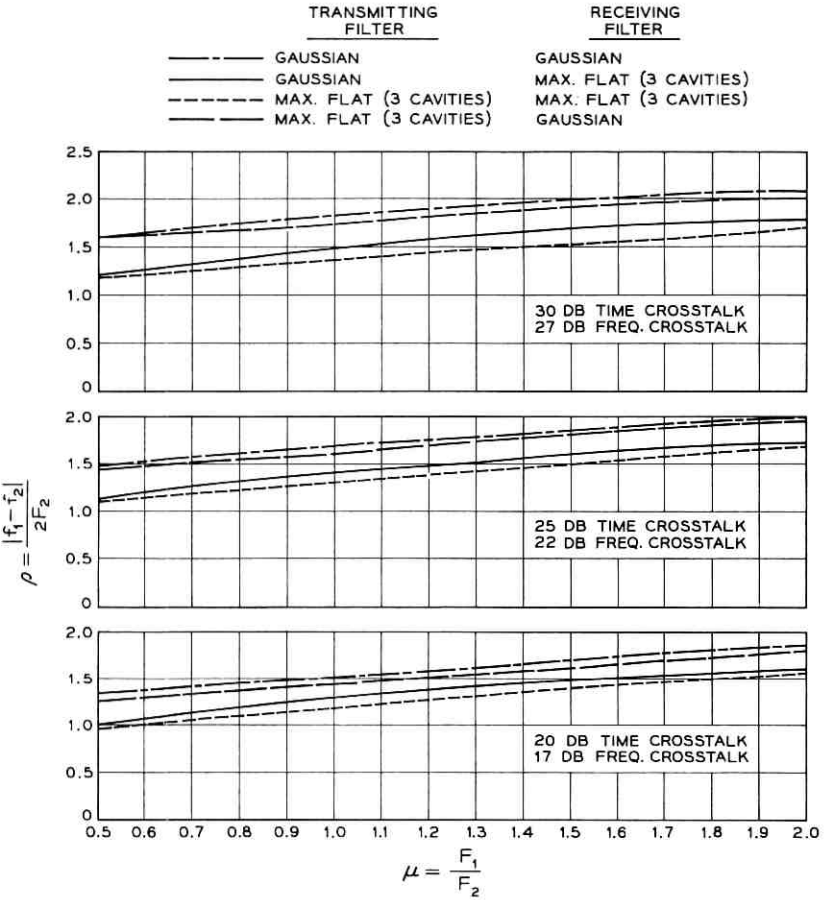


Fig. 8 — Ratio ρ for $\theta = 1.5$.

III. DISCUSSION OF RESULTS

Consider Fig. 3, in which the normalized pulse spacing

$$\frac{\tau}{2T} = 1.25 + \frac{T_r}{2T}$$

has been obtained assuming $\theta = 1.25$. The minimization of the product of pulse spacing times channel spacing $\tau \mid f_1 - f_2 \mid$ is achieved by making the sampling time T_r as short as possible, using maximally flat sending and receiving filters, and dividing the filtering in such a way that

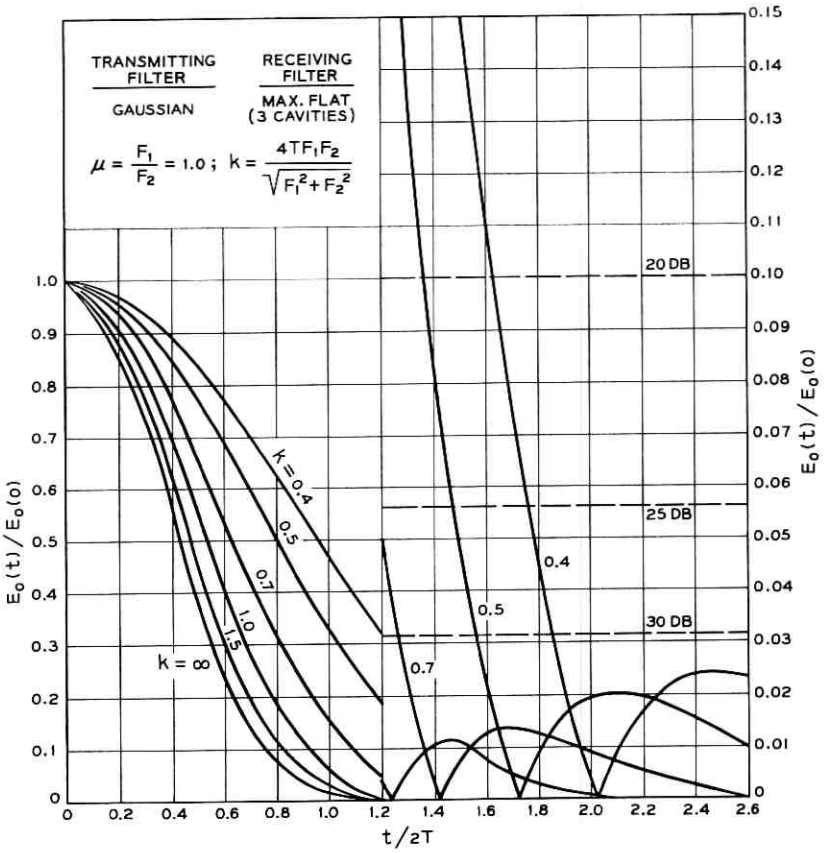


Fig. 9 — Typical plot of transient response to Gaussian input pulse.

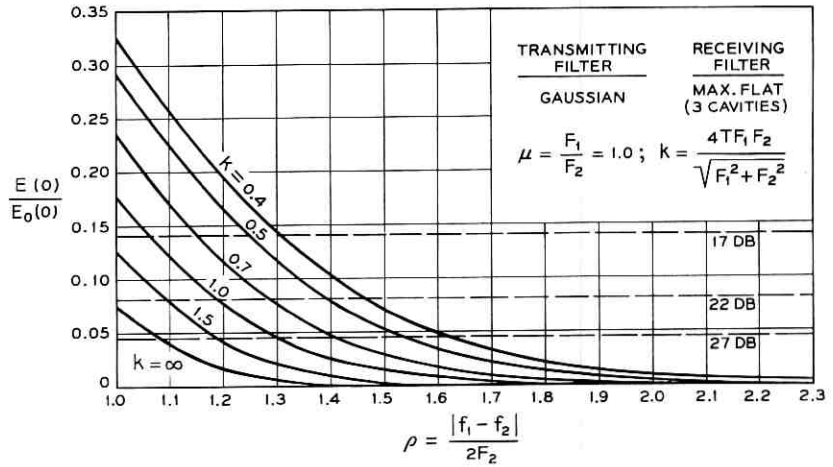


Fig. 10 — Typical plot of maximum crosstalk between neighboring channels.

$$\mu = \frac{F_1}{F_2} \cong 1.3.$$

The three sets of curves in Fig. 3 show that in each set the curves corresponding to maximally flat filters at the receiving end are very similar. Thus, as long as the receiving filter is maximally flat, there is no big advantage in using Gaussian or maximally flat filters at the transmitting end. Nevertheless, using shorter input pulses ($\theta = 1.5$), the normalized pulse spacing becomes

$$\frac{\tau}{2T} = 1.5 + \frac{T_r}{2T},$$

and it can be seen from Fig. 6 that, for low levels of interference (30 db time crosstalk and 27 db frequency crosstalk), the tails of the pulses become so important that there is a strong advantage in using a Gaussian filter at the sending end.

In order to reduce frequency crosstalk, each system should have filters with steep sides, and, in order to reduce time crosstalk, the transfer characteristics should have sloping sides. Figs. 3 and 6 analyzed previously, verify that a good compromise is obtained with a steep sided characteristic at the receiving end and a sloping one at the transmitting end.

Now, the minimums in the curves of Figs. 3 and 6 are very broad. We should select the ratio of bandwidths, μ, as large as possible, because the narrow band at the receiving end reduces the noise level. This must not be carried to extremes, because if μ is large enough the bandwidth of the sending filter may be broader than the channel spacing, and a pulse launched in one filter may waste a lot of power in a neighboring transmitting filter before reaching the receiver. This effect turns out to be of paramount importance when the transmitter characteristic is achieved by staggering filters at RF and IF; since, in this case, the RF filter may have even a wider band than that of the transmitter.

That problem, as well as the design of the receiving filter to have low noise level and the influence of the transmitter's peak power limitation on the filtering, will be discussed in another paper.[4]

We conclude with some design examples, using the following data:

(a) pulse repetition frequency $1/\tau = 160$ mc;

(b) allowed time crosstalk = 30 db;

(c) allowed frequency crosstalk = 27 db;

(d) if the pulses are narrow, say $\theta = 1.5$, then pulse width $2T$, pulse spacing τ, and sampling time $2T_r$ are related by the expression

$$2T = \frac{\tau - T_r}{1.5}.$$

Data (b), (c), and (d) locate the design curves as those in the upper group of Figs. 6, 7, and 8. The upper group of curves in Fig. 6 shows that the lowest value of the product of pulse spacing τ times channel spacing $|f_1 - f_2|$ (maximum rate of transmitted information) is the smallest minimum for the full line. This defines the shape of the filters as

transmitting filter: Gaussian,

receiving filter: approximately maximally flat (three cavities).

The abscissa and the ordinate of that minimum are

$$\mu = \frac{F_1}{F_2} = 1.3,$$

$$2T\,|f_1 - f_2| = \frac{\tau\,|f_1 - f_2|}{1.5 + \dfrac{T_r}{2T}} = 1.18.$$

In the upper-group curves of Figs. 7 and 8, the solid lines' ordinates corresponding to the abscissa $\mu = 1.3$ are

$$4TF_2 = 0.74,$$

$$\frac{|f_1 - f_2|}{2F_2} = 1.6.$$

Solving the last three formulas for sampling times zero and half pulse width, we obtain

$2T_r$ (mμsec)	$2T$ (mμsec)	$2F_1$ (mc)	$2F_2$ (mc)	$\lvert f_1 - f_2 \rvert$ (mc)
0	4.16	231	178	284
1.78	3.57	269	207	331

The input pulse widths $2T$ are different because for the narrow pulses considered

$$2T = \frac{\tau - T_r}{1.5}$$

varies with the sampling time $2T_r$.

Now we shall see what happens when datum (d) of the previous example is changed from narrow pulses ($\theta = 1.5$) to broad pulses ($\theta = 1.25$). Then,

$$2T = \frac{6.25 - T_r}{1.25}$$

and the design answers are derived as in the previous example. The dashed lines of the upper-group curves in Figs. 3, 4, and 5 yield

transmitting filter: approximately maximally flat (three cavities),
receiving filter: approximately maximally flat (three cavities),

$$\mu = 1.4,$$

$$2T\,|f_1 - f_2| = \frac{\tau\,|f_1 - f_2|}{1.25 + \dfrac{T_r}{2T}} = 1.35,$$

$$4TF_2 = 0.9,$$

$$\frac{|f_1 - f_2|}{2F_2} = 1.5.$$

Solving these equations for sampling time zero and half the input pulse width, we obtain

$2T_r$ (mμsec)	$2T$ (mμsec)	$2F_1$ (mc)	$2F_2$ (mc)	$\lvert f_1 - f_2 \rvert$ (mc)
0	5	252	180	270
2.08	4.16	302	216	325

The dotted and the solid lines in the upper group of curves of Figs. 3, 4, and 5 are very close to each other, and consequently there is no big advantage in using either a Gaussian or an approximately maximally flat filter (three cavities) at the transmitting end, while in the example of the narrow input pulse, the use of a transmitting Gaussian filter was definitely advantageous.

In the last table of results we notice that the transmitting filter bandwidth $2F_1$ is close to the channel spacing $|f_1 - f_2|$, and consequently the power fed from one sending filter to a neighboring sending filter may be too large. To reduce this waste of power it is advisable to redesign the system, adopting a value of μ different from the "optimum," for example, $\mu = 1.1$. We shall see that because of the flatness of dashed line in Fig. 3, the increase in channel spacing is small.

Following the instructions of the first example,

$$2T\,|f_1 - f_2| = \frac{\tau\,|f_1 - f_2|}{1.25 + (T_r/2T)} = 1.37,$$

$$4TF_2 = 0.98,$$

$$\frac{|f_1 - f_2|}{2F_2} = 1.4,$$

and

$2T_r$ (mμsec)	$2T$ (mμsec)	$2F_1$ (mc)	$2F_2$ (mc)	$\lvert f_1 - f_2 \rvert$ (mc)
0	5	216	196	274
2.08	4.16	260	236	329

Comparing this table of results with the previous one, we notice that for all sampling times the sending filter bandwidth has been substantially reduced, by approximately 16 per cent, at the expense of an increase in the receiving bandwidth of 9 per cent, and a very small increase of channel spacing of 1 per cent.

The bad influence of long sampling time can be appreciated by comparing the results in the first line in the table of the first example with the last line in the table of the last example. Both systems have equal input pulse widths of 4.16 millimicroseconds, but they have different sampling times, zero and 2.08 respectively. The differences between these two systems can be qualitatively justified by analyzing the necessary changes to pass from the first to the second. By increasing the sampling time, the time crosstalk increases, and, in order to maintain it at 30 db, the sending and receiving filters must be broadened 12 and 33 per cent, respectively. This bandwidth broadening increases the overlapping of transfer characteristics of neighboring channels, and therefore the frequency crosstalk goes up. To reduce it to the original level, 27 db, the channel spacing must increase 16 per cent.

IV. CONCLUSIONS

Considering only time and frequency crosstalk, Figs. 3 through 8 allow one to design the transmitting and receiving filters of adjacent PCM channels capable of minimizing the product of time occupancy and frequency occupancy.

In general, the transmitting and receiving filters should be Gaussian and approximately maximally flat (three cavities) respectively. The sloping side of the first filter contribute towards high pulse-repetition frequency, and the steep sides of the second filter contributes toward narrow channel spacing.

If the input pulses are broad, it is slightly advantageous to use approximately maximally flat filters (three cavities) in the transmitting end also. Naturally, sampling time should be as short as possible.

One set of typical results is for

pulse repetition frequency = 160 mc,
input Gaussian pulse width (at 8.686 db) = 4 mμsec,
sampling time = 1 mμsec,
Gaussian transmitting filter ~250 mc wide, at 3 db,
approximately maximally flat (three-cavity) receiving filter
~200 mc wide, at 3 db,
channel spacing = ~300 mc.

APPENDIX A

Summation of Two Displaced Gaussian Functions

The summation of two equal Gaussian curves $2F_G$ wide at 8.686 db and displaced F_d and $-F_d$ from the origin is

$$e^{-[(f-F_d)/F_G]^2} + e^{[(f+F_d)/F_G]^2}.$$

Normalizing the ordinate at $f = 0$ to unity, the summation becomes

$$Y = \frac{e^{-[(f-F_d)/F_G]^2} + e^{-[(f+F_d)/F_G]^2}}{2e^{-(F_d/F_G)^2}},$$

which can be rewritten,

$$Y = e^{-a(f/F)^2} \cosh mf/F, \tag{1}$$

where $\pm F$ are the values of f at which $Y = 1/\sqrt{2}$ (3 db),

$$a = \left(\frac{F}{F_G}\right)^2, \tag{2}$$

$$m = \frac{2F F_d}{F_G}. \tag{3}$$

Then a and m are related by the equation

$$\sqrt{2}\, e^{-a} \cosh m = 1. \tag{4}$$

From (1) and (4) it follows that once $2F$, the 3db width of the curve Y is given, the shape of it depends exclusively in the parameter a. For

$$a = 0.346 \tag{5}$$

the Gaussian function, plotted in Fig. 1 as

$$Y = e^{-0.346(f/F)^2}, \tag{6}$$

is obtained.

For the particular value

$$a = 1.3, \tag{7}$$

$$Y = e^{-1.3(f/F)^2} \cosh 1.61\left(\frac{f}{F}\right). \tag{8}$$

This function is also plotted in Fig. 1, together with the amplitude response

$$\left[1 + \left(\frac{f}{F}\right)^6\right]^{-\frac{1}{2}}$$

of a third-order maximally flat filter,[5] $2F$ wide at 3 db and centered at $f = 0$. These two curves are very similar except for the argument $f/F > 2$ (ordinates below 20 db), and consequently filters with these transfer characteristics are interchangeable as long as the tails are not important.

APPENDIX B

Gaussian Pulse Through Two Filters

Assume a Gaussian RF pulse of duration $2T$ measured at 8.686 db, and carrier f_1,

$$i(t) = \frac{1}{\sqrt{\pi}T} e^{-(t/T)^2} \cos 2\pi f_1 t. \tag{9}$$

Its Fourier transform is

$$g(f) = \int_{-\infty}^{\infty} e^{-i2\pi ft} i(t)\, dt = \tfrac{1}{2}(e^{-[\pi T(f-f_1)]^2} + e^{-[\pi T(f+f_1)]^2}). \tag{10}$$

Passing this pulse through two filters with transfer frequency characteristics $Y_1(|f| - f_1)$ and $Y_2(|f| - f_2)$, the output signal is

$$e(t) = \int_{-\infty}^{\infty} e^{i2\pi ft} Y_1(|f| - f_1) Y_2(|f| - f_2) g(f)\, df. \tag{11}$$

Substituting $g(f)$ from (10) into (11) and changing variables, one obtains

$$\begin{aligned} e(t) = {} & \frac{e^{i2\pi f_1 t}}{2} \int_{-\infty}^{\infty} Y_1(|f + f_1| - f_1) Y_2(|f + f_1| - f_2) \\ & \cdot e^{-(\pi fT)^2 + i2\pi ft}\, df \\ & + \frac{e^{-i2\pi f_1 t}}{2} \int_{-\infty}^{\infty} Y_1(|f - f_1| - f_1) Y_2(|f - f_1| - f_2) \\ & \cdot e^{-(\pi fT)^2 + i2\pi ft}\, df, \end{aligned} \tag{12}$$

and since the terms are complex conjugate,

$$e(t) = \operatorname{Re} e^{i2\pi f_1 t} \int_{-\infty}^{\infty} Y_1(\,|f+f_1|-f_1) \cdot Y_2(\,|f+f_1|-f_2) e^{-(\pi f T)^2 + i2\pi f t}\, df. \tag{13}$$

The envelope is

$$E(t) = \left| \int_{-f_1}^{\infty} Y_1(f) Y_2(f+f_1-f_2) e^{-(\pi f T)^2 + i2\pi f t}\, df + \int_{-\infty}^{-f_1} Y_1(-f-2f_1) Y_2(-f-f_1-f_2) e^{-(\pi f T)^2 + i2\pi f t}\, df \right|. \tag{14}$$

Since there are many RF cycles in the pulse $f_1 T \gg 1$, the second integral is negligible and

$$E(t) \cong \left| \int_{-\infty}^{\infty} Y_1(f) Y_2(f+f_1-f_2) e^{-(\pi f T)^2 + i2\pi f t}\, df \right|. \tag{15}$$

Furthermore, if the transfer characteristics of the filters are displaced Gaussians, (1), the envelope of the output pulse becomes

$$E(t) = \left| \int_{-\infty}^{\infty} e^{-a(f/F_1)^2 - b[(f+f_1-f_2)/F_2] - (\pi f T)^2 + i2\pi f t} \cosh \frac{mf}{F_1} \cdot \cosh n\left(\frac{f+f_1-f_2}{F_2}\right) df \right|, \tag{16}$$

which can be normalized to $E_0(0)$, the output at the instant $t = 0$ through two filters centered at the same frequency, $f_1 = f_2$. Then,

$$\frac{E(t)}{E_0(0)} = \frac{\left| \int_{-\infty}^{\infty} e^{-a(f/F_1)^2 - b[(f+f_1-f_2)/F_2]^2 - (\pi f T)^2 + i2\pi f t} \cosh \frac{mf}{F_1} \cosh n\left(\frac{f+f_1-f_2}{F_2}\right) df \right|}{\int_{-\infty}^{\infty} e^{-a(f/F_1)^2 - b(f/F_2)^2 - (\pi f T)^2} \cosh \frac{mf}{F_1} \cosh \frac{nf}{F_2}\, df}. \tag{17}$$

Performing the integrations,

$$\frac{E(t)}{E_0(0)} = \frac{e^{-4\rho^2 bB(1+aA^2) - B(t/T)^2}}{4 \cosh \frac{1}{2} mn\mu A^2 B} \cdot \left| e^{B[2n\rho(1+aA^2) - 2mb\rho A^2 + i(m+n\mu) A(t/T) + mn\mu A^2/2]} + e^{B[-2n\rho(1+aA^2) - 2mb\rho A^2 + i(m-n\mu) A(t/T) - mn\mu A^2/2]} + e^{B[-2n\rho(1+aA^2) + 2mb\rho A^2 - i(m+n\mu) A(t/T) + mn\mu A^2/2]} + e^{B[+2n\rho(1+aA^2) + 2mb\rho A^2 - i(m-n\mu) A(t/T) - mn\mu A^2/2]} \right|, \tag{18}$$

in which,

$$\rho = \frac{|f_1 - f_2|}{2F_2}, \tag{19}$$

$$\mu = \frac{F_1}{F_2}, \tag{20}$$

$$A = \frac{1}{\pi F_1 T} = \frac{4}{\pi k\sqrt{1 + \mu^2}}, \tag{21}$$

$$k = \frac{4TF_1F_2}{\sqrt{F_1^2 + F_2^2}}, \tag{22}$$

$$B = \frac{1}{1 + A^2(a + b\mu^2)}. \tag{23}$$

If both filters are centered at the same frequency, $\rho = 0$, the normalized output pulse (18) becomes

$$\frac{E_0(t)}{E_0(0)} = \frac{e^{-B(t/T)^2}}{2\cosh \frac{1}{2}mn\mu A^2B} \cdot \left[e^{\frac{1}{2}mn\mu A^2B}\cos\,(m + n\mu)AB\,\frac{t}{T} + e^{\frac{1}{2}mn\mu A^2B}\cos\,(m - n\mu)AB\,\frac{t}{T}\right]. \tag{24}$$

It also follows from (17) that the maximum amplitude transmitted through the two filters centered at different frequencies occurs at $t = 0$, and that its normalized value at $t = 0$ derived from (18) is

$$\frac{E(0)}{E_0(0)} = \frac{e^{-4\rho^2bB(1+aA^2)}}{2\cosh \frac{1}{2}mn\mu A^2B} \cdot \{e^{\frac{1}{2}mn\mu A^2B}\cosh 2\rho B[n + A^2(na - mb)] + e^{-\frac{1}{2}mn\mu A^2B} \cdot \cosh 2\rho B[n + A^2(na + mb)]\}. \tag{25}$$

REFERENCES

1. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, 1954, p. 1209.
2. Marcatili, E. A., Errors in Detection of RF Pulses Embedded in Time Crosstalk, Frequency Crosstalk, and Noise, this issue, p. 921.
3. Wheeler, H. A., The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, Proc. I.R.E., **27**, 1939, p. 359.
4. Marcatili, E. A., Compression Filtering and Signal-to-Noise Ratio in Pulse-Modulated Systems, to be published.
5. Bennett, W. R., U. S. Patent No. 1,849,656, March 15, 1932.

Contributors to This Issue

WILLIAM F. CLEMENCY, E.E., 1934, Polytechnic Institute of Brooklyn; Western Electric Co., 1923–26; Bell Telephone Laboratories, 1926—. His early work was concerned with manufacturing processes for carbon for transmitters and development of carbon transmitters. During World War II he was concerned with the design of transmitters and earphones for use in military communication systems under high ambient noise. Later he worked on development of acoustical instruments and measurements. Since 1953 he has been concerned with development of telephone sets.

H. G. COOPER, B.S., 1949, M.S., 1950, and Ph.D., 1954, University of Illinois; Bell Telephone Laboratories, 1954—. His work has been in the field of cathode ray beam devices, particularly studies of electron lenses, deflection systems and electron guns. He heads a group concerned with storage tubes and beam-deflection devices. Member American Physical Society, Pi Mu Epsilon, Sigma Tau, Sigma Xi, Tau Beta Pi.

WALTER D. GOODALE, JR., E.E., 1928, Lehigh University; M.E.E., 1937, Polytechnic Institute of Brooklyn; American Telephone & Telegraph Co., 1928–34; Bell Telephone Laboratories, 1934—. For a number of years he was concerned with transmission studies of operators' and station telephone sets and with transmission problems related to central office room noise. He has also worked on coin collector development and, more recently, speakerphone design. Member I.R.E.

JOHN L. KELLY, JR., B.A., 1950, M.A., 1952, and Ph.D., 1953, University of Texas; Bell Telephone Laboratories, 1953—. He has been engaged in studies of television and the application of information theory to television, and in studies of information processing. He recently became head of a subdepartment engaged in information coding research.

OSCAR KUMMER, B.E.E., 1940, Cooper Union; graduate studies, 1940–41, Stevens Institute of Technology; Bell Telephone Laboratories, 1934—. His first work was in the area of communication transformer design, followed by development of oscillators and detectors. During the war he was concerned with defense projects for the Navy and later he

turned to design of measuring equipment for coaxial cable systems. He is now engaged in development of transmission and phase measurement techniques over the range 250 to 4000 mc.

KANEYUKI KUROKAWA, B.S., 1951, and Dr. of Eng., 1958, University of Tokyo; Bell Telephone Laboratories, 1959—. Mr. Kurokawa is on leave of absence from his position as assistant professor at the University of Tokyo. He has been engaged in research on parametric amplifiers. Member I.R.E., Institute of Electrical Engineers (Japan), Institute of Electrical Communication Engineers (Japan).

DANIEL LEED, B.S., 1941, College of the City of New York; M.E.E., 1957, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1946—. He heads a group concerned with the development of systems for measuring the frequency characteristics of parameters significant in network design, including insertion phase shift, loss, envelope delay, and reflection coefficient.

NATHAN LEVINE, B.S., 1952, Massachusetts Institute of Technology; M.S., 1954, and Ph.D., 1957, University of Illinois; Bell Telephone Laboratories, 1957—. He has been engaged in the design and simulation of radar data processing systems in connection with the Nike-Zeus AICBM project. He is currently working on new digital data smoothing techniques. Member American Physical Society.

CAROL C. LOCHBAUM, B.A., Douglass College, 1958; Bell Telephone Laboratories, 1958—. She has been engaged in computer programming for visual and acoustics research problems. Member Phi Beta Kappa.

MASON A. LOGAN, B.S., 1927, California Institute of Technology; M.A., 1933, Columbia University; Bell Telephone Laboratories, 1927—. His early work included transmission design problems of local manual and dial circuits and circuit research on alternating current methods of signaling. During and immediately after the war he worked on mine fire-control systems, proximity fuses, Nike-Ajax, and other military projects. Later he was engaged in development of electromagnets and relays, followed by development of instrumentation for semiconductor device process control and evaluation. At present he is engaged in design and development of data transmission terminals.

E. A. MARCATILI, Aeronautical Engineer, 1947, and E.E., 1948, University of Cordoba (Argentina); Research staff, University of Cordoba, 1947–54; Bell Telephone Laboratories, 1954—. He has been engaged in theory and design of filters in multimode waveguides. More recently he has concentrated on waveguide systems research. Member I.R.E., Physical Association of Argentina.

MAX V. MATHEWS, B.S., 1950, California Institute of Technology; M.S., 1952, and Sc.D., 1954, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1955—. He has specialized in acoustics research in speech transmission and has been especially concerned with stimulation of speech experiments on a digital computer. He recently became head of a subdepartment engaged in human information processing research. Member Acoustical Society of America, I.R.E., Sigma Xi.

JOHN RIORDAN, B.S., 1923, Yale University; American Telephone and Telegraph Co., 1926–34; Bell Telephone Laboratories, 1934—. For a number of years he concentrated on studies of the distribution of currents in railway networks and tracks and in the ground, and the effects of these currents on telephone circuits. Since 1940 he has been engaged in mathematical studies, including Boolean algebra in switching, number theory in cable splicing, combinatorial analysis and probability studies of traffic. Member American Association for the Advancement of Science, American Mathematical Society, Institute of Mathematical Statistics, Mathematical Association of America.

IRWIN W. SANDBERG, B.E.E., 1955, M.E.E., 1956, and D.E.E., 1958, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1958—. He has been concerned with analysis of military systems, particularly radar systems, and with synthesis and analysis of active and time-varying networks. Recently he transferred to a group engaged in research on communications fundamentals. Member I.R.E., Eta Kappa Nu, Sigma Xi, Tau Beta Pi.

LAJOS F. TAKÁCS, Doctor's Degree, 1948, University of Technical and Economical Sciences, Budapest; Doctor of Mathematical Sciences, 1957, Hungarian Academy of Sciences; Tungsram Research Laboratory (Telecommunications Research Institute), Budapest, 1945–55; Research Institute for Mathematics of the Hungarian Academy of Sciences, 1950–58; Roland Eötvös University, Budapest, 1953–58; Columbia University,

1959—; consultant, Bell Telephone Laboratories, 1959—. At present he is teaching probability theory and stochastic processes, and is engaged in research in the mathematical theory of telephone traffic. Member American Mathematical Society, Mathematical Association of America, Society for Industrial and Applied Mathematics, Institute of Mathematical Statistics, Sigma Xi.

Michiyuki Uenohara, B.E., 1949, Nihon University (Japan); M.S., 1953, and Ph.D., 1956, Ohio State University; D.E., Tohoku University (Japan), 1958; Bell Telephone Laboratories, 1957—. He has been engaged in exploratory studies of microwave variable reactance amplifiers and microwave tubes. He was also engaged in microwave tube research at Nihon University from 1949 to 1952, and taught there in 1957. Member American Physical Society, I.R.E., Institute of Electrical Communication Engineers (Japan), Eta Kappa Nu, Pi Mu Epsilon, Sigma Xi, RESA.

V. A. Vyssotsky, A. B., 1950, and M.S., 1956, University of Chicago; Bell Telephone Laboratories, 1956—. He has been studying problems in speech compression and phoneme recognition.