An Automated 20–20,000-cps Transmission Measuring Set for Laboratory Use

By G. D. HAYNIE and P. E. ROSENFELD

(Manuscript received February 27, 1963)

An automated loss and phase measuring set has been developed to measure, in a point-by-point fashion, two-port networks over the frequency range from 20 to 20,000 cps with a maximum accuracy of 0.01 db and 0.1 degree. Manual, semiautomatic, and automatic modes of operation are provided. During automatic operation, the interval between successive data points is determined by programmed intervals of frequency, loss, and phase. These quantities are programmed either separately or in combination. Machine selection of the measurement points is controlled by comparing the program with information fed back from the signal oscillator and the measured circuit. At each selected point, frequency is measured, loss and phase are determined by self-balancing standards, and these data are recorded on punched tape.

I. INTRODUCTION

Automatic measurement techniques have been widely and successfully applied in situations where many repetitive measurements of a similar type are required. In many cases the measuring instruments are programmed for a specific measurement or sequence of measurements, and versatility is of minor importance. This paper describes an automatic transmission measuring set that was developed for laboratory use in the 20- to 20,000-cps frequency range to measure two-port networks having a wide variety of characteristics.

2501

The first section of the paper specifies the quantities that are measured and the range and accuracy of the measurements, and lists the functions that the skilled operator of a manual measuring set must perform when making laboratory measurements. In Section II some of the factors involved in data taking are discussed, and the chosen method of automatic data selection is described. Section III describes, in broad terms, the measuring and control circuits used. Operation of the automated set in the manual, semiautomatic, and automatic modes is described in Section IV. A more detailed description of the measuring and control circuits, along with a description of some of the more important subsystems, is given in Section V. In Section VI some of the most important systems considerations which influenced the measuring set design are discussed. Section VII gives the results of tests made to confirm the measuring set accuracy and to evaluate, using selected networks, the operation of the automatic control and data selection functions of the measuring set.

1.1 Required Measurements, Accuracy and Ranges

The basic quantities measured are the insertion loss and phase shift of a two-port network as a function of frequency when measured between 600-ohm unbalanced impedances.[1] The measurement frequency is continuously variable from 20 to 20,000 cps and adjustable to ±0.1 cps. The input power level to the measured network is variable in 5-db steps from +10 dbm to −20 dbm, each step being accurate to ±0.25 db. The range of loss measurement is from 0 to 120 db with an accuracy of ±0.01 db for losses less than 40 db. Phase shift is measured from 0 to 360° with an accuracy of ±0.1° for losses less than 40 db. Measurements can be made at impedances other than 600 ohms by changing a pair of plug-in terminations, but the oscillator output level will not be calibrated, and the loss and phase measurements may not be direct reading.

1.2 Function of the Measuring Set

Measurements are required for adjustment of networks and for obtaining data which can be processed by a computer or recorded in tabular or graphical form. In obtaining data with a manually operated set, the operator must choose and select the measurement frequencies, make the loss and phase measurement, and record the data. Any fully automatic set must also carry out these functions and, in particular, must provide a way of introducing the necessary inputs into the set so that the required data will be obtained.

II. SELECTION OF DATA POINTS

The a priori knowledge of the behavior of the networks being measured will not always be sufficient to enable the network designer to preselect the frequencies at which the measurements should be made in order that no important details be missed. If the frequency interval between successive measurements is made small, all the important details will be detected, but this would increase the over-all time required to make the measurements and give superfluous data. Unneeded data may not be objectionable when recorded in graphical form, but stored digital data should be kept to a minimum. In order to select only meaningful data on a network with uncertain characteristics, some form of feedback is needed when selecting the data points. When making manual measurements, a skilled operator supplies this feedback.

To illustrate several criteria by which a machine can select data points, an arbitrary loss characteristic is illustrated in Fig. 1(a). If a certain frequency interval, Δf, were to be chosen as the interval between measured points, important parts of the characteristic would be missed if a smooth curve were drawn through the points taken as shown in Fig. 1(b). Application of this criterion does not require any feedback from the measured network.

If the same characteristic were to be measured by using a selected increment of loss, ΔL, to determine the measured points, the characteristic would be completely determined within $\pm\Delta L$. This is shown in Fig. 1(c). The loss interval criterion does require feedback. It can be seen that the effect of the frequency "comb" (Δf criterion) is that fine structure may be missed, but coarse structure is reproduced pretty well. The effect of the loss "comb" is to reproduce fine structure well, but large area errors may occur when coarse structure is measured.

Application of the frequency and loss "combs" together results in a much better reproduction in general than either can do alone. Fig. 1(d) shows a smooth curve drawn through the measured points taken by the use of both criteria simultaneously. The same remarks hold for phase measurements in which a selected increment of phase, $\Delta\theta$, is used to select measured points.

In the general measurement case, the criteria used to select data points in one part of the frequency range may not be the desirable criteria for another part of the frequency range. The bandpass filter shown in Fig. 2 may be used to illustrate this. Typically, one would be interested only in the coarse loss structure in the stop bands and what the minimum loss is to within perhaps 3 db. Another item of interest might be the loss slope in the transition regions or fine variations of the loss and phase in the
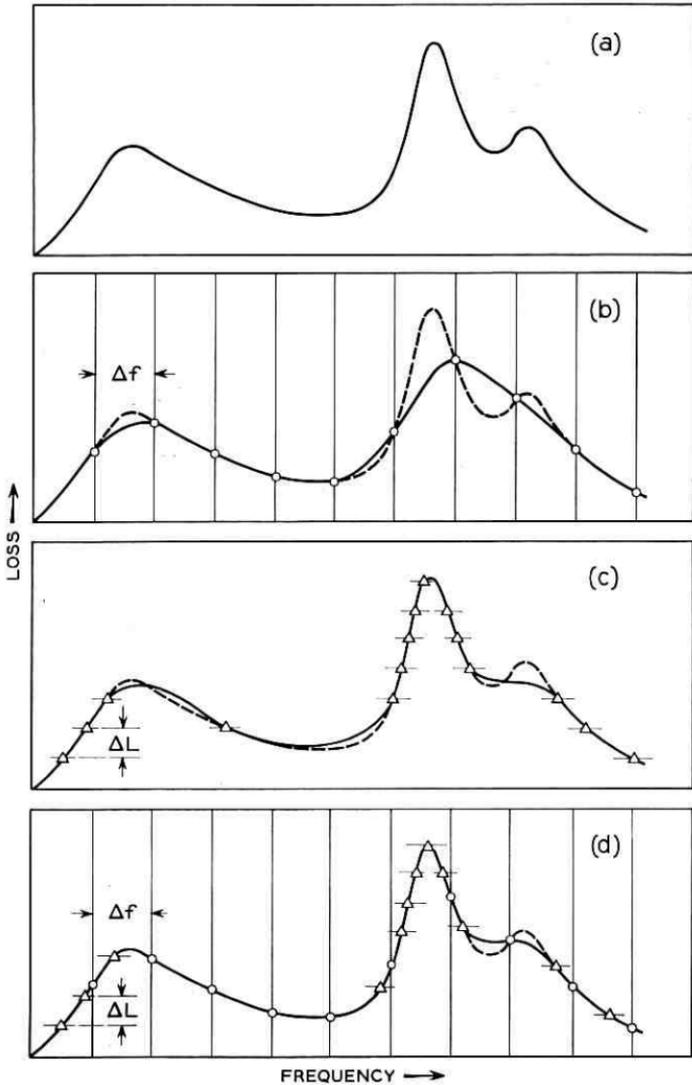
Fig. 1 — Criteria for data selection.

passband, because these characteristics influence carrier transmission. Thus, the choice of which criteria to use and the size of the chosen intervals will vary in different frequency bands.

In the data selection method adopted, a program control is provided which permits Δf, ΔL, and $\Delta \theta$ intervals to be set up independently in each of 5 frequency bands. The Δf intervals are 1000, 300 and 100 cps; the
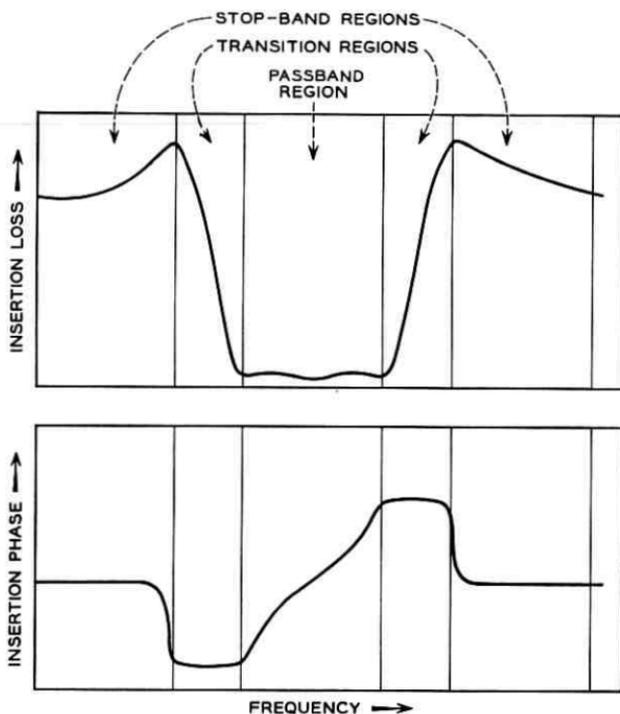
Fig. 2 — Bandpass filter transmission characteristic.

ΔL intervals are 10, 3, 1, 0.3, 0.1 and 0.03 db; and the $\Delta\theta$ intervals are 30, 10, 3, 1 and 0.3 degrees. The edges of the bands can be set to any arbitrary frequency in the 20- to 20,000-cps range to the nearest 10 cps.

III. BLOCK DIAGRAM OF THE TRANSMISSION MEASURING SET

The block diagram of the measuring set can be conveniently divided into two parts, the measuring circuit and the control circuit.

3.1 Measuring Circuit

A simplified block diagram of the measuring circuit is shown in Fig. 3. The signal oscillator, adjusted to the measurement frequency F, is applied to the unknown and to the loss standard. As the sampling switch S_1 rapidly compares the outputs from the two paths, any difference in transmission between the unknown and the loss standard will result in variations of the measurement frequency level and phase at the switching rate.
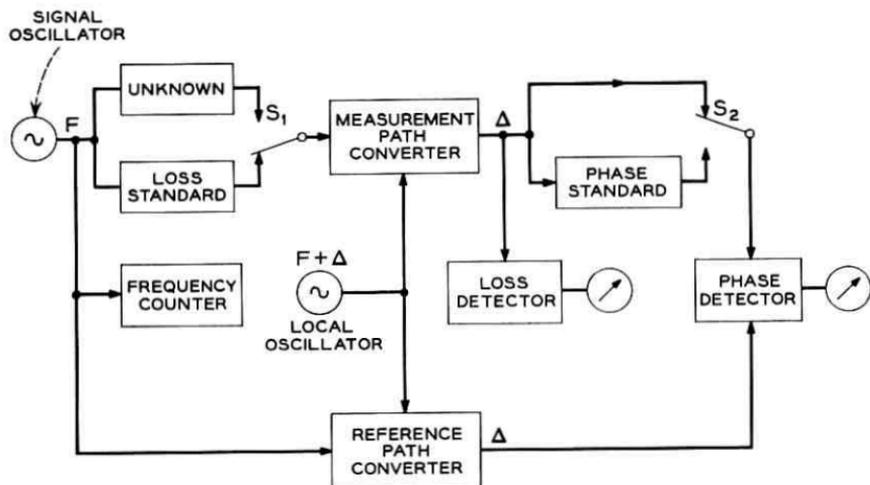
Fig. 3 — Simplified block diagram — measuring circuit.

These level and phase variations at frequency F are linearly translated to level and phase variations at the fixed frequency Δ by the measurement path converter. The level differences at the converter output are detected by the loss detector and indicated by proportional deflections on the loss detector meter. When the loss standard is adjusted to have loss equal to that of the unknown, the loss detector meter will indicate a null. The loss measurement is then completed by recording the setting of the loss standard.

For the phase measurement, a second path is needed to provide a reference signal for the phase detector. The output from the measurement path converter is applied to the phase standard and the strapped path, and switch S_2 is switched synchronously with S_1. The phase standard is adjusted until the phase at S_2 is the same for both switch positions, this condition being indicated by a null on the phase detector meter. If the phase of the loss standard is negligible, the phase of the unknown will then be equal to the phase of the phase standard. The frequency counter measures the signal oscillator frequency, and recording the setting of the phase standard and frequency counter completes the measurement of a point.

3.2 Control Circuit

In Fig. 4 the control circuit is added to the measurement circuit. Self-balancing controls for the loss and phase standards have been added along

with automatic readout. Automation of the set is accomplished by the addition of the over-all control and the oscillator sweep control. The over-all control consists of selector switches for setting up the measurement program; a range control circuit to sense the band in which the oscillator frequency, F, is located; loss and phase detectors to sense the magnitude of the loss and phase changes; and a frequency control circuit used in frequency interval programming. Before discussing the circuit operation in more detail, the three operating modes of the measuring set will be described.

IV. OPERATION OF THE MEASURING SET

The three modes of operation for the set are manual, semiautomatic, and automatic. Regardless of which mode is selected, the operator must connect the network to be measured, adjust the terminal impedances if
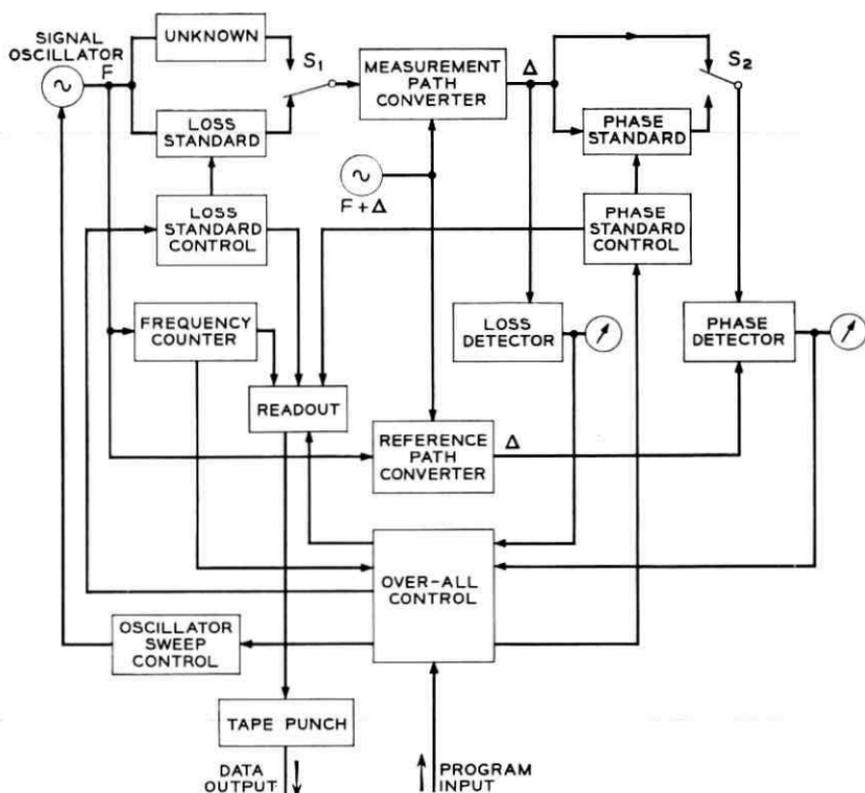

Fig. 4 — Simplified block diagram — measuring and control circuits.

they are to be different from 600 ohms, and select the oscillator output level. Fig. 5 shows the console arrangement of the set.

4.1 *Manual Operation*

In the manual operating mode, the oscillator can be manually adjusted to ±5 cps by means of the coarse tuning control used in conjunction with a film scale. The fine tuning control, when used with the frequency counter, makes it possible to set frequency to ±0.1 cps. The frequency counter is operated by a pair of pushbutton controls to give either a 1-second or 10-second time base.

Relay switching of the IF filters makes two detector bandwidths available in the manual mode, one bandwidth being 10 times the other. The narrower bandwidth is normally used when the measurement frequency is below 500 cps or at other frequencies when the maximum available signal-to-noise ratio at the detector is desired.

Manual adjustment of a set of decade switches is required to balance the loss standard. The loss unbalance indication is provided in analog
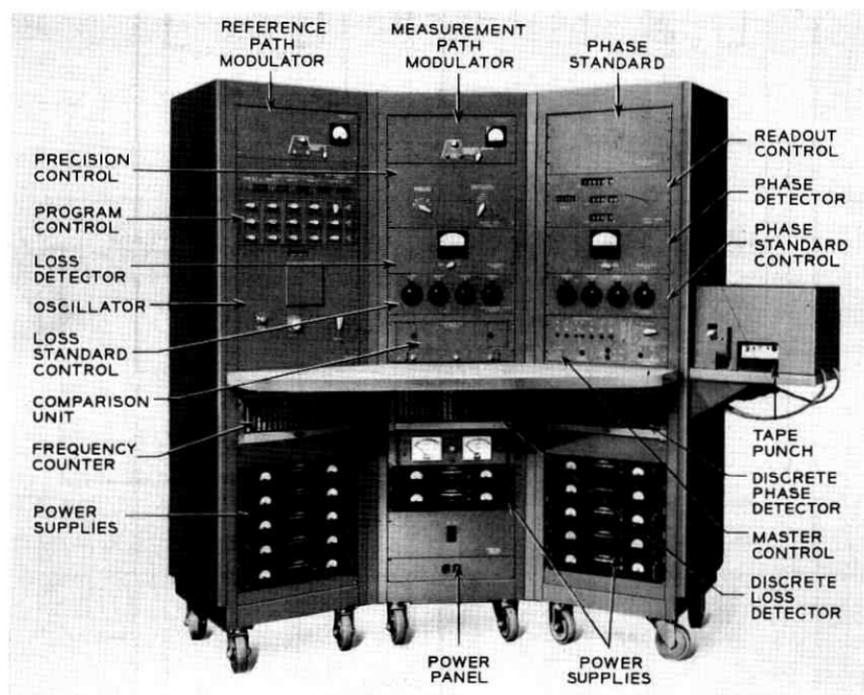

Fig. 5 — Console arrangement of transmission measuring set.

form by a calibrated meter with 3 scales and in digital form by indicating lamps placed just above the decade switches. These lamps indicate which decade switch should be adjusted to obtain a balance. The manual phase balance is obtained in the same way by means of similar indications.

After the frequency has been measured and the loss and phase standards balanced, the measured values appear on the visual readout panel. Data can either be recorded manually or punched on paper tape by operating the print out pushbutton.

The oscillator can be swept in the manual mode by operating the coarse tuning control. As the frequency is changed, indications of the loss and phase variations in the measured network can be obtained from the analog and digital indicators of loss and phase unbalances.

4.2 Semiautomatic Operation

Operation of the set in the semiautomatic mode permits the operator to initiate, by means of pushbutton controls, any step in the sequence which forms the automatic operating cycle. The controls for the separate steps are labeled "balance loss," "balance phase," "read frequency," "display data," "print out," and "sweep frequency." Front panel lamps are placed above each control to indicate operation of the set in each one of these steps. Operation of the balance loss control results in automatic switching of the loss standard pads in an ordered sequence to achieve balance. Automatic balance of the phase standard is initiated by operation of the balance phase control. The loss and phase balances are made to a precision which depends either on the absolute loss or on the setting of a manual precision control.

The read frequency control has the same function as in the manual mode: that is, to initiate a frequency measurement with either a 1-second or 10-second time base.

The function of the display data step is to store on the visual readout panel the readings of the loss standard, phase standard, and frequency counter.

As in the manual mode, operation of the print out control causes the data to be recorded on paper tape.

The choice of the frequency for the next data point may be made by the operator or by the program control. In the latter case, the range control, Δf, ΔL, and $\Delta\theta$ switches are set for the desired program, and operation of the sweep frequency control causes the oscillator to sweep until one of the programmed intervals is exceeded.

It remains only to note that the sequence just described is not rigid, and the various steps can, with a few exceptions, be interchanged in sequence or omitted. For example, if the phase shift of a network is not of interest, the balance phase step can be omitted.

4.3 *Automatic Operation*

In the automatic mode, the operator must manually set the oscillator to the minimum frequency of interest and set up the desired measurement program. The precision of frequency measurement must be set to 0.1 cps, 1 cps, or programmed. In the programmed position, the precision is 0.1 cps at measurement frequencies below 500 cps and 1 cps above 500 cps. After the start control is operated, the set will cycle continuously until all desired data is taken. Each cycle consists of 5 steps: balance loss, balance phase, read frequency, print out, and sweep frequency. Each cycle is timed, and in the event that a cycle is not completed in a prescribed time, an alarm signal occurs. This alarm also signals the end of a run.

V. DESCRIPTION OF SUBSYSTEMS IN THE MEASURING SET

Fig. 6 is a partially simplified block diagram of the measuring set, including the control circuits. Functioning of the over-all system will be more readily understood if some of the subsystems are considered separately, so the oscillator, loss measuring circuit, phase measuring circuit, programming circuit, and some of the other important circuits will be separately described.

5.1 *Oscillator*

A detailed block diagram of the oscillator is shown in Fig. 7. The oscillator is of the heterodyne type, containing one crystal oscillator fixed at 97 kc and one LC oscillator which is variable from 97.02 to 117 kc. The oscillator modulator converts these two frequencies to the measurement frequency, which can be varied from 0.02 to 20 kc in a single range. The stability of the oscillator output frequency is 0.1 cps over the period of a measurement. Two AGC loops are employed in the oscillator to obtain an output flat with frequency to ±0.1 db. The level select pads which follow the power amplifier are used to adjust the level applied to the circuit being measured.

Tuning of the variable oscillator is controlled by three capacitors. Coarse tuning is accomplished with a servo-driven, variable air capacitor geared to a 200-inch film scale. The servo can be operated manually with
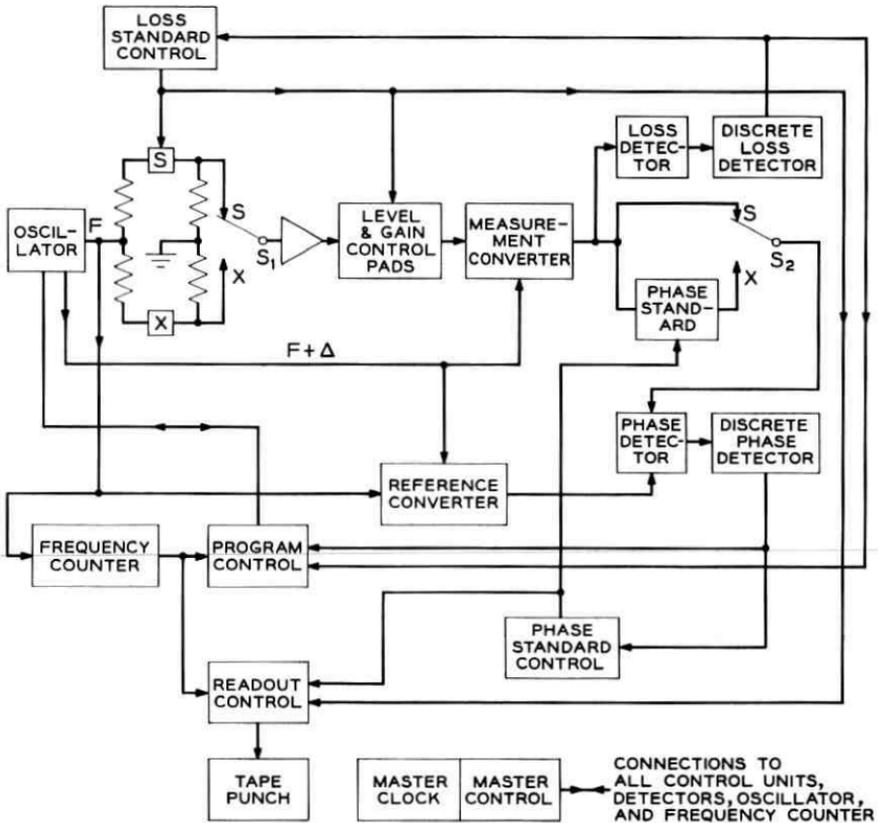
Fig. 6 — Simplified block diagram — measuring and control circuits.

the coarse tuning control or automatically with signals from the program control. A small variable air capacitor mounted on the front panel provides the manual fine tuning control, and the third capacitor is voltage controlled for operation in an AFC loop.

When the AFC relay is operated, the voltage-controlled capacitor is connected to the LC oscillator. The control voltage for the variable capacitor is obtained from a phase detector which has as its inputs the measurement signal and a precise 100-cps frequency. If the open-loop measurement frequency is within ±10 cps of a 100-cps multiple, energizing the AFC relay will cause the measurement frequency to lock to this 100-cps multiple with no frequency error.

A lamp and a photodiode are used with the film scale to detect 100-cps multiples as the measurement frequency is being swept. A decimal
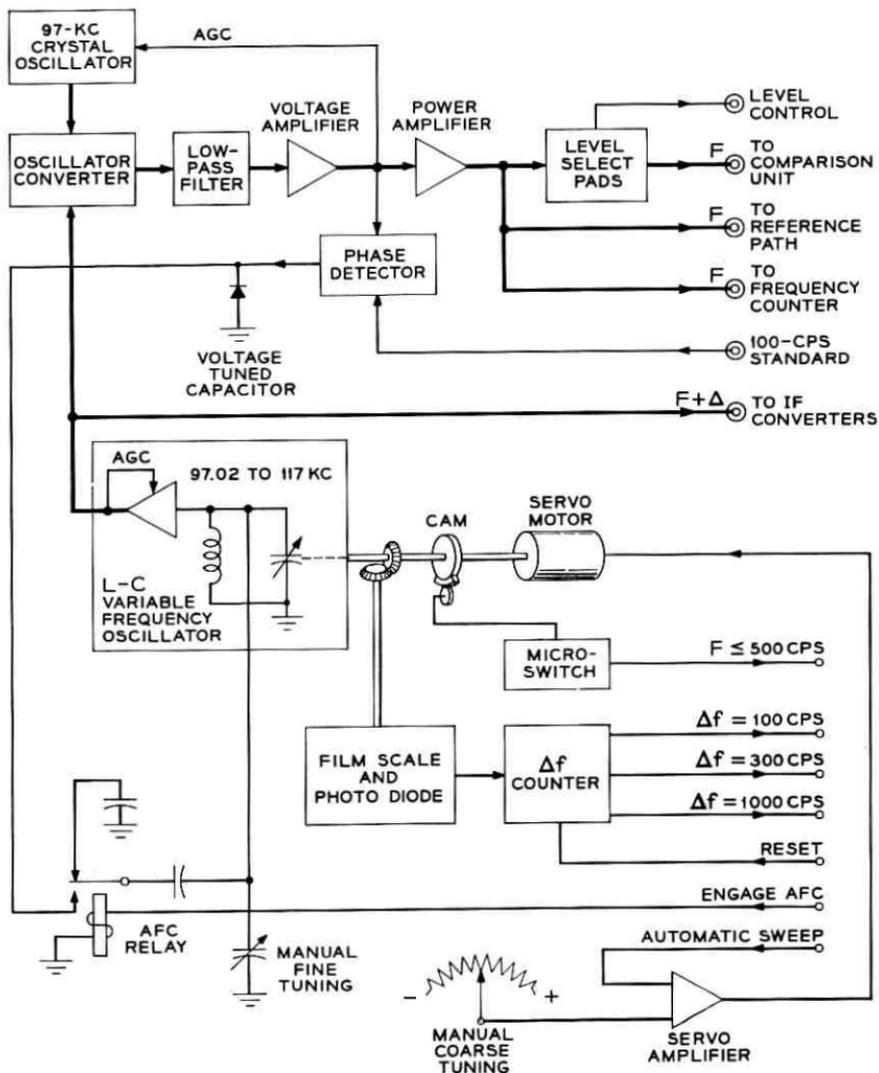
Fig. 7 — Oscillator block diagram.

counter is used to provide indications for frequency intervals of 100, 300, and 1000 cps.

A cam, mounted on the coarse tuning capacitor shaft, operates a micro-switch to select the narrow-band IF filter during automatic operation at measurement frequencies below 500 cps.

5.2 *Loss Measuring Circuit*

Fig. 8 shows the loss measuring circuit together with the self-balancing circuit. The test signal from the oscillator, after being attenuated to the desired level by the output level selector, is applied to the loss standard and the unknown. The loss standard consists of relay-switched attenuator pads, which can be inserted in 0.01-db steps from 0 to 119.99 db.

The outputs from the standard (S path) and unknown (X path) paths are alternately sampled by a mercury relay operating at either 10 cps or 1.6 cps, depending on whether the wideband or narrow-band IF filter is used. The sampling relay is followed by a high-impedance buffer amplifier, so that the input levels and terminating impedances for the standard and unknown paths are essentially constant. This permits the
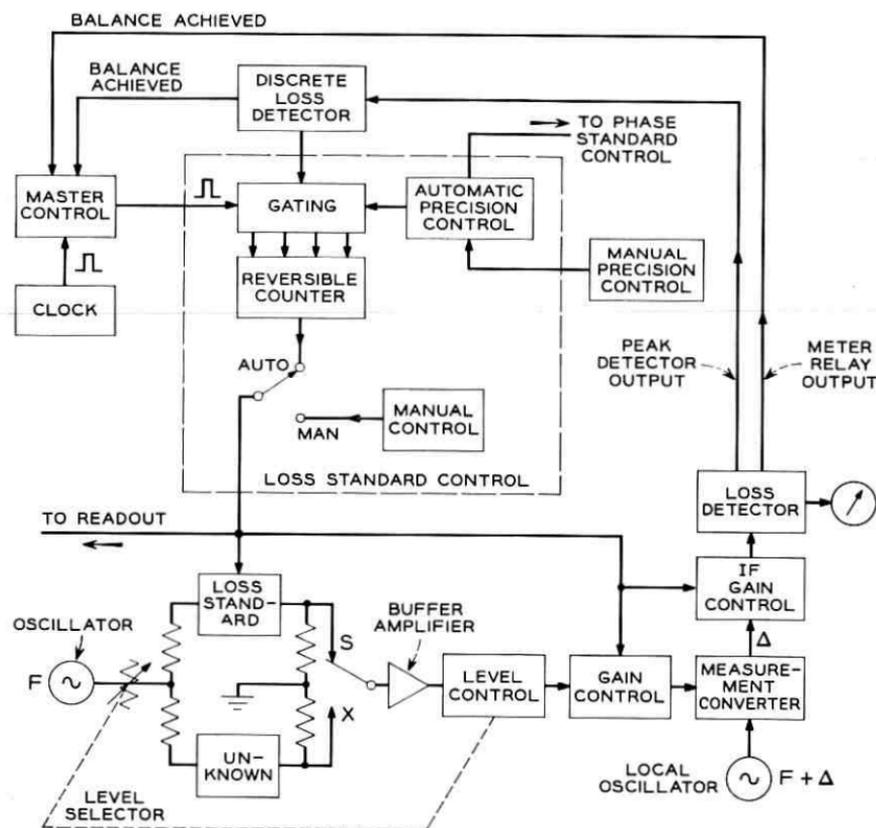

Fig. 8 — Simplified block diagram — loss measuring circuit.

sampling rate to be chosen independently of the characteristics of the unknown network.

The buffer amplifier is followed by a level control circuit, ganged to the oscillator output level selector in such a way that the S path input level into the loss detector is kept constant ±0.1 db as the oscillator output level is varied. This is also the raison d'être for the gain control circuits which are ganged to the loss standard. The gain control circuits provide an S path input level at the loss detector that is constant to ±0.1 db over the full loss measuring range of the set. Maintenance of this constant level makes possible a loss detector sensitivity constant to ±1 per cent. The gain control circuits also provide a constant S path input level into the measurement path converter for losses up to 80 db. This eases the linearity and crosstalk requirements on the converter.

The measurement path converter linearly translates the amplitude of the 20- to 20,000-cps measurement signal to the 97-kc IF. A beam deflection tube is used as the converter. The tube was used because of the high degree of balance which can be maintained for the local oscillator signal in the modulator output. The voltage gain to this input is less than −76 db. This balance eases the IF filter requirements for measurements at low frequencies, since at the lowest measurement frequency of 20 cps, the local oscillator frequency is within 20 cps and the upper sideband is within 40 cps of the desired lower sideband. The narrow-band crystal filter (28-cycle noise bandwidth) is used for automatic measurements at frequencies below 500 cps. In the manual and semiautomatic modes at measurement frequencies above 500 cps and in cases where a maximum signal-to-noise ratio is not required, the wideband filter (280-cycle noise bandwidth) is used. It was the settling time of these filters that determined the maximum switching rates of the sampling relay.

The loss detector compares the amplitudes of the IF signal for the S and X positions of the sampling relay. The difference in amplitudes is proportional to the difference in loss between the loss standard and the unknown. This loss difference, or unbalance, is displayed on a calibrated panel meter with three full-scale ranges, 0.1, 1, and 10 db. A meter relay is connected in series with the panel meter and is used to give a discrete indication to the master control when the loss unbalance is less than 0.015 db.

A discrete loss detector is connected to the loss detector in parallel with the meter circuits to provide additional discrete indications of the loss unbalance. The discrete loss detector is an eight-level analog-digital converter. The eight levels correspond to loss unbalance magnitudes less than 0.01 db or greater than 0.01 db, 0.03 db, 0.1 db, 0.3 db, 1 db, 3 db,

and 10 db, together with the sign. The indications are used for balancing the loss standard and for loss programming.

The loss standard control operates the relays of the loss standard in accordance with the settings of the front panel control knobs in the manual mode, and in accordance with the state of a 4-stage reversible decimal counter in the semiautomatic and automatic modes.

When an automatic loss balance is being made, balancing pulses generated by the master clock are gated into the loss standard control by the master control. The loss standard control then gates the pulses into the counter stage controlling the most significant decade requiring a change. Control for this gating is furnished by the discrete loss detector, and the result is a rapid balancing sequence.

Since the detector signal-to-noise ratio decreases as the measured loss increases, automatic control of the precision of balance is required during automatic measurements. The precision of automatic balance versus measured loss is given in Table I. For cases where precision less than that provided by the automatic control is necessary or desirable, a manual precision control is provided.

When a balance has been achieved, as indicated by signals from the discrete loss detector and the loss detector meter relay, balancing pulses are no longer gated to the loss standard control. If the balance achieved condition is maintained for a period of 4 successive balancing pulses, the loss balance is considered complete and, in the automatic operation mode, the phase balancing step is begun.

5.3 Phase Measuring Circuit

The phase measuring circuit is shown in Fig. 9. Since the phase detector indication is affected by the input levels, the loss balance must precede the phase measurement.

The sampling relay in the comparison unit samples the phase difference in the S and X paths. This phase difference is translated by the measure-

TABLE I

Measured Loss	Precision of Automatic Balance
0–39 db	0.01 db
40–59 db	0.03 db
60–79 db	0.1 db
80–89 db	0.3 db
90–99 db	1 db
100–120 db	3 db
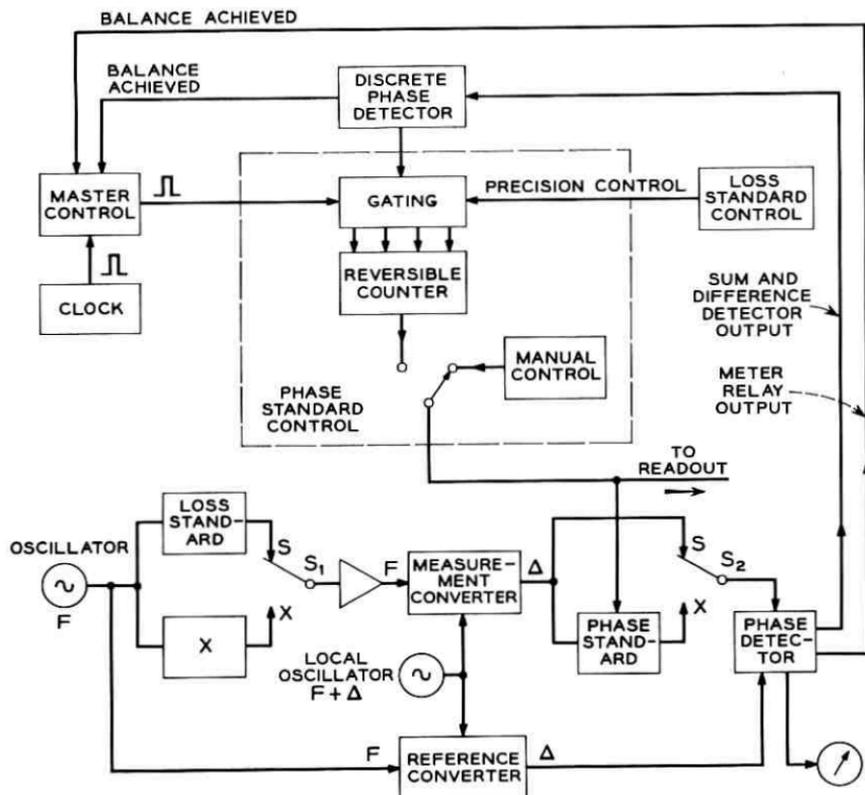
Fig. 9 — Simplified block diagram — phase measuring circuit.

ment path converter to its negative at IF. The phase standard is switched in synchronism with the sampling relay in such a way that the phase variation at the measuring path input to the phase detector is equal to the phase difference between the unknown and the phase standard. This phase is compared to the constant phase of the reference path by the phase detector, and the phase difference between the unknown and the phase standard is displayed on a calibrated panel meter. The meter has three full-scale ranges, 1°, 10°, and 80°. A meter relay is also used here to give a discrete indication when the phase unbalance is less than 0.15°. The deflection sensitivity of the phase detector is maintained constant within ±1 per cent by controlling the input levels to the phase detector and by using a voltage-controlled phase shifter in a feedback loop, so that the output of the phase detector will be zero when the sampling relays are in the S position.

The phase standard is a digital standard which can be relay-switched to vary phase from 0 to 360° in 0.1° steps. The phase standard is balanced in essentially the same way as the loss standard. The precision of the phase balance, which depends on the signal-to-noise ratio, is determined either by the insertion loss of the network being measured or by the setting of the manual precision control.

5.4 *Program Control Circuit*

As shown in Fig. 10, the program control receives inputs from the discrete loss and phase detectors, the frequency counter, and the oscillator Δf counter. When the oscillator is being swept during semiautomatic or automatic operation, the frequency counter is cycled continuously, using a 0.1-second time base. The swept frequency is measured with a maximum error of 22 cps, and the output from the frequency counter is compared to the settings of the range selector switches. This comparison indicates which of the selected ranges the measurement frequency lies in, and also determines which set of Δf, ΔL, and $\Delta\theta$ switches control the program. The settings of the appropriate Δf, ΔL, and $\Delta\theta$ switches are
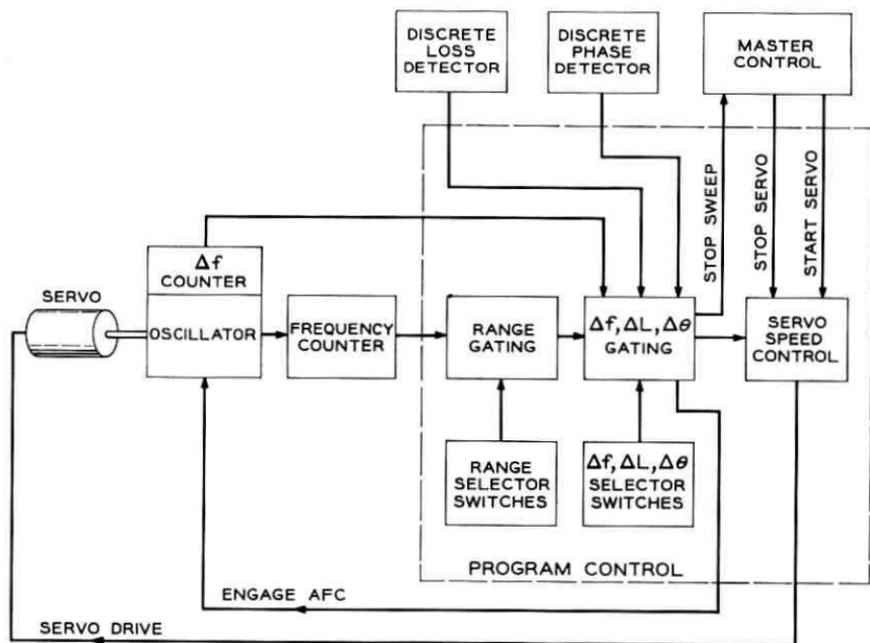
Fig. 10 — Simplified block diagram — programming circuits.

compared to the outputs of the Δf counter, the discrete loss detector, and the discrete phase detector. Coincidence circuits indicate a match when it occurs and stop the oscillator sweep. If the oscillator is stopped because of a Δf interval being equalled, a signal is sent to the oscillator to engage the AFC loop and reset the Δf counter.

The speed of the oscillator sweep is controlled by the program control. This speed is proportional to the programmed ΔL or $\Delta \theta$ interval, whichever requires the slower speed. When the indicated loss or phase unbalance reaches $\frac{1}{3}$ of the programmed interval, the servo speed is reduced. These two controls of the oscillator sweep speed make possible a reasonably short sweep time without, at the same time, causing large errors in the programmed intervals due to overshoot. Section 6.2.3 shows the considerations involved in the choice of sweep speeds.

5.5 Other Important Circuits

In addition to the circuits already discussed, there are several others which deserve a brief description.

5.5.1 Frequency Counter

The frequency counter can be controlled to measure frequency with 3 time bases. When used with the 1-second and 10-second time bases, the counter takes a single measurement, and the reading is used for the visual and punched tape readouts. When the 0.1-second time base is used, the counter takes measurements continuously, and the readings are used to provide inputs to the range control circuit.

5.5.2 Readout Control

The front panel of the readout control contains the in-line readout indicators to display the measured frequency, loss, and phase and 6 selector switches to provide an identification number for the punched tape readout. The readout control unit also contains the required translators for the visual and punched tape readouts and the parallel-to-serial converter for the tape readout. A circuit is also provided to interrupt the recorded readout in the event of a failure in the data parity check.

5.5.3 Master Clock

All timing pulses necessary for the system are generated by the master clock. The clock has two rates, 10 cps and 1.6 cps, for use with the wide- and narrow-band IF filters, respectively. Pulses from the clock go to the

sampling relays, loss detector, discrete loss detector, loss standard control, phase detector, discrete phase detector, and phase standard control.

5.5.4 *Master Control*

Sequence control is the best description of the function of the master control unit. In the automatic mode, it monitors the operation of the balance loss, balance phase, read frequency, print out, and sweep frequency steps, and controls the order in which they are operated. The master control panel contains pushbutton controls to operate each of these steps in the semiautomatic mode. An error control is provided which allows the operator using the manual or semiautomatic mode to designate the most recently punched data point as an error. Finally, the panel contains the all-important control which the sorcerer's apprentice could not find, the one marked "stop."

VI. SYSTEMS CONSIDERATIONS WHICH INFLUENCED TEST SET DESIGN

The most important objectives in the design of the test set were to achieve the specified measurement ranges and accuracies and to provide the means for rapid and accurate data selection based on settings of the program control. Some of the factors which were considered in the test set design will now be discussed.

6.1 *Measurement Errors*

The following sources of error were considered in the test set design:
(a) loss standard imperfections
(b) phase standard imperfections
(c) insertion ratio errors due to misterminations
(d) harmonics and modulation products
(e) IF bandwidth and switching rate of comparison unit
(f) random noise
(g) power frequency pickup
(h) crosstalk and
(i) drift.

6.1.1 *Loss Standard*

It was required that all steps from 0 to 40 db be accurate to ± 0.002 db and that the phase shift of the loss standard be independent of loss to within $\pm 0.02°$ for frequencies from 20 cps to 20 kc. To meet the requirements on attenuation accuracy, stable wire-wound resistors with the

required accuracy were used. To meet the phase requirements, it was necessary that the design impedance of the loss standard be made less than 600 ohms. For convenience in measuring the loss and phase of the standard with existing equipment, the design impedance was made 75 ohms. After compensation, variations of loss setting produced no observable phase changes on a phase detector with 0.01° sensitivity.

6.1.2 *Phase Standard*

All steps of the phase standard were required to be accurate to ±0.02°, with the output level variations between any two steps less than 1 db. The unit consists of 3 stages of relay-switched RC elements. Low reactance, wire-wound resistors and high-Q temperature-compensated silver mica capacitors are used to achieve the necessary component stability of 0.01 per cent and the low output level variation. The phase standard was calibrated on the X75706 TMS (Ref. 1, p. 1515), which has the required phase accuracy of 0.02° at 97 kc. The phase standard is also frequency-sensitive, so a requirement of ±0.01 per cent, or about ±10 cps, was placed in the IF frequency stability.

6.1.3 *Errors Due to Misterminations*

For the comparison circuit shown in Fig. 11, the maximum error, φ, in measuring insertion loss and phase between nonideal terminations is approximately,[2]

$$\varphi = |S_{11}G| + |S_{22}L| + |GL| + |S_{11}'G'| + |S_{22}'L'| + |G'L'|$$

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,
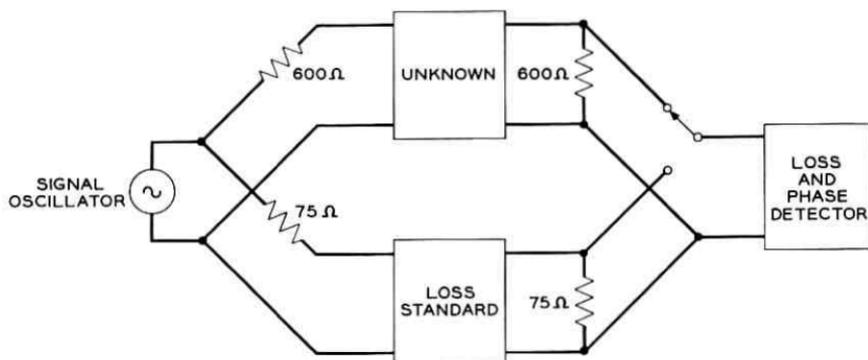

Fig. 11 — Comparison circuit.

and the physical input and output scattering coefficients of the network under test, all taken with respect to 600 ohms. The quantities G', L', S_{11}', and S_{22}' are similarly defined for the loss standard path in the comparison circuit and are taken with respect to 75 ohms.

G, L, G', L', S_{11}', and S_{22}' are under the designer's control. The resistor elements used in the generator and load terminations are within 0.05 per cent of nominal, and in general the scattering parameters S_{11}' and S_{22}' of the loss standard are much smaller than S_{11} and S_{22}. Thus, for an unknown with scattering coefficients as large as 0.2, the maximum error will be $2 \times 0.2 \times 0.00025 = 10^{-4}$. This corresponds to an error of 0.0009 db or 0.006°.

6.1.4 *Harmonics and Modulation Products*

One group of products of special concern are those which produce spurious signals of frequency F at the input to the IF converter and of frequency Δ at the output of the IF converter. These products are: harmonics of F present in the oscillator output, Δ or $F + \Delta$ present in the oscillator output, and F or Δ present in the $F + \Delta$ signal. The requirement on the maximum magnitude of these products depends on the linearity of the gain and level control amplifiers and the IF modulator linearity and gain. Requirements were placed so that all spurious sources of Δ at the IF modulator output are 65 db below the IF signal for losses of 50 db or less and 20 db below the IF signal at 120 db loss.

Other products of special concern are the $F + \Delta$ and $2F + \Delta$ products in the IF converter output as F approaches 20 cps. The IF converter is balanced against $F + \Delta$ and maintains a $\Delta/(F + \Delta)$ ratio of 30 db or greater in the converter output. No balance is obtained against $2F + \Delta$, so the narrow-band IF filter must provide 35 db discrimination at $F + \Delta$ and 65 db discrimination at $2F + \Delta$. This very steep cutoff requirement on the IF filter imposes the limitation on the operating speed of the measuring set.

6.1.5 *IF Bandwidth and Switching Rate of Comparison Unit*

The filter that provides the necessary narrow IF bandwidth will also have a large envelope delay and long settling times. The total filter transient time determines the maximum switching rate of the comparison unit and the maximum rate at which the loss and phase standards can be balanced. Since one of the objectives in automating the set was to operate at high speed, a wide filter bandwidth is desirable. As a compromise, two IF filters are used. During automatic operation from 20 cps to 500

cps, a filter with a 3-db bandwidth of 18 cps is used and from 500 cps to 20 kc, the 3-db bandwidth is 180 cps. The filters have a maximally flat delay characteristic which gives a relatively short transient time.[3] The two switching rates used are 10 cps and 1.6 cps.

6.1.6 *Random Noise*

The presence of random noise in the measuring circuit will cause both fixed and fluctuating errors in the indications of the loss and phase detectors. The fixed errors cannot be detected and thus produce errors in the measurements. The fluctuating errors can be detected and eliminated, providing they are averaged for a sufficiently long period. No averaging is done in the discrete detectors used in automatic balancing of the loss and phase standards, so the precision of automatic balance is reduced as the noise level increases. The precision of automatic balance versus level is given in Section 5.2.

The noise bandwidth of the system is determined by the bandwidth of the IF filters. The noise bandwidth of the narrow filter is 28 cycles and that of the wide filter is 280 cycles. For losses greater than 20 db, the low-level point of the system is at the comparison switch (S_1, Fig. 6). The noise figures of the 0-db buffer amplifier and the level and gain control amplifiers which follow S_1 determine the system noise figure. The system noise figure varies with frequency, being 28 db at frequencies above 1 kc and increasing to 45 db at 20 cps.

For 40-db loss measurements and 0-dbm sending level, the minimum signal is -40 dbm. At frequencies greater than 1 kc, and for a 280-cycle noise bandwidth, the minimum signal-to-noise ratio is 81 db, which is greater than the required for a measurement with 0.01 db precision. For 120 db loss and a 28-cycle noise bandwidth, the minimum signal-to-noise ratio is 11 db. This allows 3-db precision with the discrete detector.

6.1.7 *Power Frequency Pickup*

Harmonics of the power frequency which are coupled into the measurement circuits also cause fixed and fluctuating errors in the loss and phase detector indications. These errors occur whenever the measurement frequency is close enough to one of these harmonic frequencies that the IF filter does not discriminate against it. In the automatic and semiautomatic modes, a blanking circuit is provided to prevent the ΔL and $\Delta \theta$ program controls from stopping the oscillator within ± 15 cps of 60, 180, or 300 cps. Inside these bands, the outputs from the discrete detectors are not reliable for programming and balancing. With the use of the

blanking, measurements may be made over the 20- to 20,000-cps range with a precision of 0.01 db for losses up to 40 db.

6.1.8 *Crosstalk*

Since the signal level into both modulators is kept constant for losses up to 80 db, crosstalk at IF or through the modulators was not a difficult problem.

The main source of crosstalk was in power supply coupling. Decoupling is especially difficult in those units operating at 20 cps, since effective reactive components become very large at that frequency. Some decoupling was effected by using separate power supplies, and in other cases an active filter was used to provide decoupling and attenuation of power supply ripple.

6.1.9 *Drift*

The 1.6-cps switching rate requires that drifts in the common path circuits be less than 0.01 db and 0.1° during the 0.6-second period.

6.2 *Programming Errors*

As discussed in Section II, the measurement programming permits the division of the measurement frequency range into as many as 5 bands, and in each of these bands a selection of Δf, ΔL, and $\Delta\theta$ intervals may be made. A natural question to ask is, how accurate are these intervals when point-by-point measurements are made?

6.2.1 *Range Control Error*

As the input to the frequency counter is varied at a rate of K cps/second over the counting period τ seconds, the maximum error, E, in the counter reading will be,

$$E = \frac{K\tau}{2} + \frac{1}{\tau}.$$

For $\tau = 0.1$ second and with the maximum sweep rate of 240 cps/second, an error of 22 cps can result. This error in counter reading can cause, in the worst case, a range control error of 33 cps.

6.2.2 Δf *Program Accuracy*

The low-inertia servo motor selected for the oscillator sweep drive and the relatively high-speed Δf recognition circuitry make it possible to stop

at any Δf frequency interval with an overshoot of less than 10 cps. When the oscillator drive is stopped, the oscillator frequency is phase locked to a multiple of a 100-cps frequency standard, thus giving exact Δf intervals.

6.2.3 ΔL and $\Delta \theta$ Program Interval Accuracy

Measurements of the ΔL and $\Delta \theta$ program intervals are made with a frequency source which is being swept, on unknowns with finite bandwidths, using a narrow-band detector. In most cases, the delay introduced by the detector will be long compared to the delay present in the measured networks. The inertia of the oscillator drive motor and the detector delay will cause the oscillator to overshoot the frequency where the ΔL or $\Delta \theta$ interval is exactly reached. Referring to Fig. 12, it is seen that the program error E_L, for loss is

$$E_L = n \log_2 \frac{f_2}{f_1}$$

where n = loss slope in db/octave
  f_2 = frequency where oscillation is stopped
  f_1 = frequency where ΔL is reached.
For a delay of τ in recognizing ΔL and a sweep speed of K cps/second

$$E_L = n \log_2 \left(1 + \frac{K\tau}{f_1} \right) \doteq 1.44 \frac{nK\tau}{f}$$

for small E_L and f_1 arbitrary. If the fractional program error P is defined as

$$P = \frac{E_L}{\Delta L}$$

then the acceptable rate of sweep for loss intervals is

$$K = \frac{1}{1.44\tau} P \frac{f\Delta L}{n}.$$

The desirable fractional program error and τ are considered constants, and ΔL depends on the program input. It can be shown that by using the differentiated output of the loss detector fed back to control the oscillator sweep speed, K, a constant fractional program error can be achieved. Similar results can be obtained for phase programming.

The approach used in this set was to sweep the oscillator at a rate proportional to ΔL over the loss range from 0 to $\Delta L/3$ in such a way that a program error of just less than 60 per cent will occur. When $\Delta L/3$ is recognized, the sweep rate is reduced to about $\frac{1}{6}$ of the original rate to
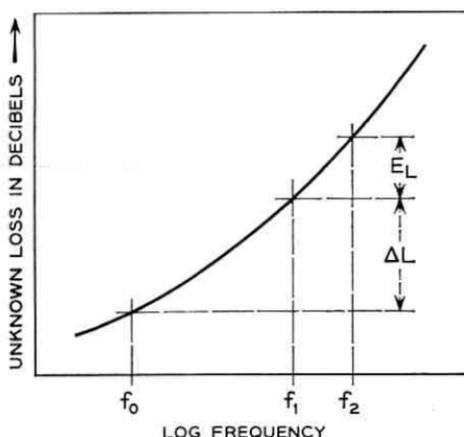
Fig. 12 — ΔL program error.

limit the overshoot to something less than 10 per cent. The sweep for phase intervals is controlled in the same way, and in cases where both ΔL and $\Delta\theta$ are programmed, the slower sweep speed called for is used. The discrete detectors which measure the ΔL and $\Delta\theta$ intervals are accurate to ± 2 per cent.

6.3 *Operating Speed*

The maximum speed of operation of the set is obtained in the automatic mode. When the set takes a data point, it goes through a cycle consisting of sweep frequency, balance loss, balance phase, read frequency, and print out.

As discussed in Section 6.2, the sweep rate is limited by the allowable program error. The sweep period could be minimized if a more complex feedback arrangement were used along with an oscillator drive motor with a 1000:1 speed range. The motor which was used has a speed range of 15:1 and the maximum sweep speed is 240 cps/second.

In the balance loss operation, the IF filter forms part of the control loop. The long envelope delay and settling times associated with the narrow-band filter constitute almost all of the loop delay, with the remainder being the conversion time of the discrete detector and the delay of the balancing pulse gates. The balancing rate is 10 pps with the wideband filter and 1.6 pps with the narrow-band filter. Even though the IF filter is not in the balance phase control loop, for practical reasons the balancing rate and logic are similar to those used in the loss standard. A timeout of 4 pulses is used after each of the balances before the balance is considered achieved.

In the read frequency operation, either a 1-second or 10-second time base is used for the frequency counter.

The speed of the readout is limited by the maximum speed of the recording tape punch and the stepping switch used in the parallel-to-serial data conversion. Without contact protection, the stepping switch is about 3:1 faster than the tape punch. Since 40 characters are punched at each measurement point and the maximum speed of the punch is 20 characters per second, the lower bound for readout is 2 seconds. The actual readout time, with contact protection on the stepping switch, is just under 3 seconds. Section 7.4 gives the breakdown of time in each step of a typical cycle.

VII. RESULTS WITH THE AUTOMATED SET

Tests were made to determine the measurement range and accuracy of the set. Selected networks were used to determine the accuracy of programmed intervals and the operating speeds, and to evaluate to some extent the efficiency of data selection.

7.1 *Measurement Accuracy*

Verification of the loss and phase accuracies of the set was carried out by a process of measuring a number of calibrated loss pads and delay networks individually and in combination. These measurements showed that the objectives of 0.01 db and 0.1 degree accuracy were met for losses up to 40 db. A nominal 120-db loss was measured as 114 db.

7.2 *Program Accuracy*

Measurements were made using selected networks to determine the accuracy of the Δf, ΔL, and $\Delta\theta$ intervals.

7.2.1 Δf *Program Accuracy*

When a programmed Δf limit is reached, the oscillator is then phase locked to a 100-cps frequency standard so that the frequency intervals are, for all practical purposes, exact. Some jitter and 100-cps sidebands do appear at the oscillator output, but the resulting variations in the instantaneous oscillator frequency are less than 0.003 per cent or 0.03 cps, whichever is larger.

7.2.2 ΔL *Program Accuracy*

Each of the ΔL intervals was programmed on a network with the maximum loss slope which can be measured to 0.01 db (0.12 db/cps).

The error in the programmed intervals is 10 per cent or 0.05 db, whichever is larger.

7.2.3 $\Delta\theta$ Program Accuracy

Each of the $\Delta\theta$ intervals was programmed on the maximum phase slope which can be measured to 0.1° (1.44°/cps). The error in the intervals is 15 per cent or 0.8°, whichever is larger. From these results, it can be seen that the minimum frequency change that will occur in the presence of steep loss or phase slopes is about 0.5 cps.

For $\Delta\theta$ intervals programmed on a 25-μsec delay line (slope 0.01°/cps), the program error is 2 per cent or 0.1°, whichever is larger.

7.3 Data Selection

7.3.1 Delay Equalizer

A simple example of automatic data selection is provided by measurements on a delay equalizer. Loss and phase are measured from 500 to 3400 cps, and loss must be known to the nearest 0.1 db. The program used is $\Delta L = 0.1$ db and $\Delta f = 100$ cps. Delay is calculated on a computer from the phase measurements, and the ΔL program characterizes the unknown within the required 0.1 db. The Δf program eliminates any significant uncertainty in the measurement frequencies and thus removes one source of error in the delay measurement. The data obtained are plotted in Fig. 13.

7.3.2 Bandpass Filter

As a second example of automatic data selection, a program was set up to a test a telegraph tone channel filter with the requirements listed in Table II.

Assuming that the band of interest is from 1 to 5 kc and that the phase linearity in the pass band is also of interest, the program was set up as shown in Table III.

The data taken are plotted on Fig. 14. In ranges 1 and 5, all of the points are taken by the $\Delta f = 300$ program, but one has the assurance that the loss did not deviate by more than 10 db between any two adjacent points. The $\Delta L = 3$ db program is used in ranges 2 and 4 where the 16-db and 7.5-db loss requirements are placed. In range 3 the phase change of about 220° is approximately linear with respect to frequency, so the $\Delta\theta = 30°$ program provides 7 points spaced at approximately uni-
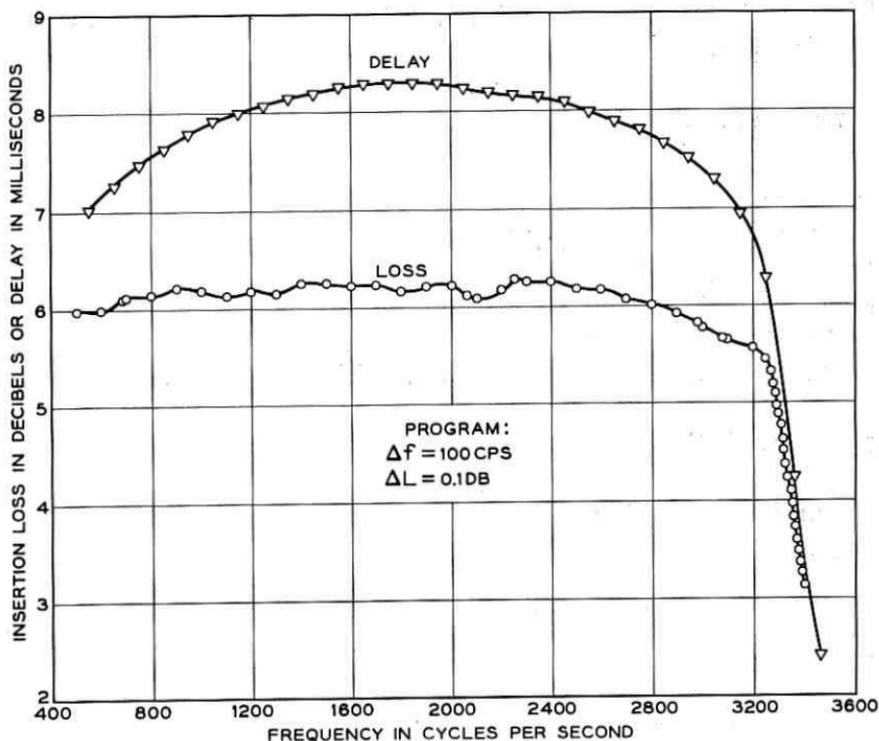
Fig. 13 — Automatic data selection — delay equalizer.

TABLE II

Loss (Relative to 2975-cps Center Frequency)	Frequency
>3 db	2975 ± 42.5 cps
>7.5 db	2890 cps
>16 db	2975 ± 170 cps
>40 db	below 2400 cps
>40 db	above 3700 cps

TABLE III

Range	Lower Frequency (kc)	Upper Frequency (kc)	Program		
			Δf	ΔL	$\Delta \theta$
1	1	2.65	300	10	off
2	2.65	2.89	off	3	off
3	2.89	3.06	off	off	30
4	3.06	3.3	off	3	off
5	3.3	5	300	10	off
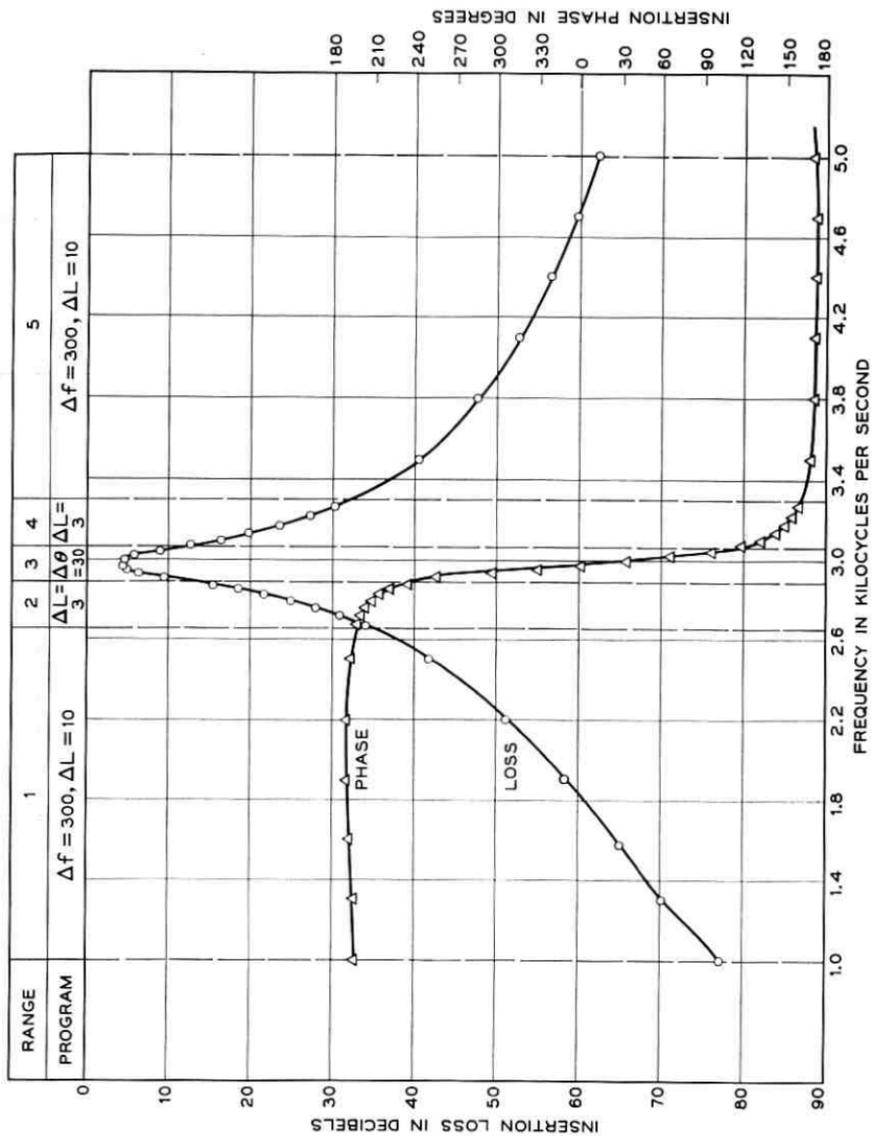
Fig. 14 — Automatic data selection — bandpass filter.

form frequency intervals. The resulting data is similar to what one would expect if it were taken by a capable human operator.

7.4 *Operating Speed*

The program described in Section 7.3.2 required 6.5 minutes to run, with an average of 14 seconds required for each point. This average time per point can be broken down as shown in Table IV.

In general, except for the read-out step and the miscellaneous delays, the time required for each step varies according to the switching rate used, the size of the programmed intervals, the precision of the measurements, and the unknown being measured. However, from the data taken so far, 14 seconds appears to be a good average.

Manual measurements were made on the filter described in Section 7.3.2 at approximately the same points plotted on Fig. 14, and the time required for the manual measurements was more than 5 times that required for the automatic measurements. Experience indicates that the average speed of the automated set is at least 10 times that of manual sets.

7.5 *Conclusions*

An automated laboratory measuring set has been developed which makes loss and phase measurements based on a preselected program and on information fed back from the measured network. The program provides the means for accurate and relevant data selection from simple instructions based on network requirements. Automatic read-out eliminates operator read-out errors, and machine processing of the paper tape output provides flexibility in the form of output data. The possibility of us-

TABLE IV

Step	Approximate Time Required (sec)
Sweep frequency	3
Balance loss	2
Loss time out	0.5
Balance phase	2
Phase time out	0.5
Read frequency	2
Read-out	3
Miscellaneous delays	1
Total	14

ing the set in the semiautomatic and manual modes provides additional flexibility.

The higher speed of the automated set makes it possible to obtain more comprehensive data on networks being measured, with the attendant benefit to both designers and users, and this data is obtained in less time at lower cost.

REFERENCES

1. Elliott, J. S., A High-Precision Direct-Reading Loss and Phase Measuring Set for Carrier Frequencies, B.S.T.J., **41**, Sept., 1962, p. 1494.
2. Leed, D., and Kummer, O., A Loss and Phase Set for Measuring Transistor Parameters and Two-Port Networks between 5 and 250 mc, B.S.T.J., **40**, May, 1961, p. 841.
3. Bangert, J. T., Practical Applications of Time Domain Theory, Bell System Monograph 3404, p. 35.

Signal Distortion in Nonlinear Feedback Systems

By I. W. SANDBERG

This paper reports on some properties of the solutions to the functional equation $s_2(t) = \varphi[\mathbf{C}s_2(t) + s_1(t)]$, where φ is a nonlinear function, the operator $\mathbf{C}$ is a convolution, and s_1 is a known function belonging to a prescribed Banach space.

The equation plays a central role in the theory of signal transmission through a general physical system containing linear time-invariant elements and a single time-variable nonlinear element. In particular we establish conditions under which $s_2(t)$ is the fixed point of a contraction mapping of the Banach space into itself and we discuss some consequences of this result.

As a direct application, we consider the range of validity of two simple cascade flow graphs (i.e., flow graphs without feedback loops) for approximately determining the signal distortion in nonlinear feedback systems when the distortion is small. Our discussion is not restricted to specific types of nonlinear characteristics.

I. INTRODUCTION

This paper reports on some properties of the solutions to the functional equation $s_2(t) = \varphi[\mathbf{C}s_2(t) + s_1(t)]$, where φ is a nonlinear function, the operator $\mathbf{C}$ is a convolution, and s_1 is a known function belonging to a prescribed Banach space.

The equation plays a central role in the theory of signal transmission through a general physical system containing linear time-invariant elements and a single time-variable nonlinear element. In particular we establish conditions under which $s_2(t)$ is the fixed point of a contraction mapping of the Banach space into itself and we discuss some consequences of this result.

As a direct application, we consider the range of validity of a simple cascade flow graph (i.e., a flow graph without feedback loops) for

approximately determining the signal distortion in nonlinear feedback systems when the distortion is small. Our discussion is not restricted to specific types of nonlinear characteristics.

Except in cases in which the nonlinearity is very small, our results establish the utility of the graph only when (in a certain precise sense) the feedback around the nonlinear element is small. However, our results show that the range of validity of this flow graph is very much greater than that indicated by an earlier writer[1] who has considered this question for the case in which the nonlinear characteristic is of the form $x + \epsilon x^{m}$, where ϵ and m are real constants with m an odd positive integer.

As is well known, large amounts of feedback are often present in physical systems. In fact, large amounts of feedback are often used to reduce nonlinear distortion. Consequently the established range of validity of the flow graph mentioned above does not include the most important cases of interest. To deal with such situations we propose an alternative, but very closely related, flow graph for approximately determining the signal distortion when the distortion is small. It appears that the range of validity of this graph includes the vast majority of cases of engineering interest.

Section II considers some mathematical preliminaries. In Section III we describe a model of the physical system under consideration and show the relevance of the functional equation mentioned above. Section IV presents some preliminary results which are concerned with the properties of certain linear operators. In the remaining sections we consider both some properties of the solutions to the functional equation and engineering implications of the results.

II. MATHEMATICAL PRELIMINARIES

Let $\Re = [\Theta, \rho]$ be an arbitrary metric space.[2] A mapping $\mathbf{A}$ of the space $\Re$ into itself is said to be a contraction if there exists a number $k < 1$ such that

$$\rho(\mathbf{A}x, \mathbf{A}y) \leq k\rho(x, y)$$

for any two elements $x, y \ \varepsilon \ \Theta$. The contraction-mapping fixed-point theorem[2] is basic to much of the subsequent discussion. It states that every contraction mapping defined in a complete metric space $\Re$ has one and only one fixed point (i.e., there exists a unique element $z \ \varepsilon \ \Theta$ such that $\mathbf{A}z = z$). Furthermore $z = \lim_{n \to \infty} \mathbf{A}^{n}x_0$, where x_0 is an arbitrary element of Θ.

Throughout the discussion $\mathcal{L}_2$ denotes the space of complex-valued

square-integrable functions defined on the real interval $(-\infty,\infty)$. The norm of $f(t)$ ε $\mathcal{L}_2$ is denoted by $\| f \|_2$ and is defined by

$$\| f \|_2^2 = \int_{-\infty}^{\infty} | f(t) |^2 \, dt.$$

The symbol $\mathcal{L}_1$ denotes the space of absolutely integrable functions defined on $(-\infty,\infty)$. We shall use the symbols $\mathcal{L}_{2R}$ and $\mathcal{L}_{1R}$, respectively, to denote the intersections of the spaces $\mathcal{L}_2$ and $\mathcal{L}_1$ with the set of real-valued functions. It is well known that $\mathcal{L}_2$ and $\mathcal{L}_{2R}$ are Banach spaces.

We take as the definition of the Fourier transform of $f(t)$ ε $\mathcal{L}_2 \cup \mathcal{L}_1$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \, e^{-i\omega t} \, dt, \qquad f \, \varepsilon \, \mathcal{L}_1$$

$$F(\omega) = \text{l.i.m.} \int_{-\infty}^{\infty} f(t) \, e^{-i\omega t} \, dt, \qquad f \, \varepsilon \, (\mathcal{L}_2 - \mathcal{L}_1)$$

and consequently when $f(t)$ ε $\mathcal{L}_2$

$$f(t) = \text{l.i.m.} \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) \, e^{i\omega t} \, d\omega.$$

With this definition, the Plancherel identity reads

$$2\pi \int_{-\infty}^{\infty} f(t)\underline{g}(t) \, dt = \int_{-\infty}^{\infty} F(\omega)\underline{G}(\omega) \, d\omega, \qquad f, g \, \varepsilon \, \mathcal{L}_2 .$$

Throughout the discussion $\mathcal{K}(\Sigma)$ denotes the space of bounded real-valued functions that (i) are defined on the real interval $(-\infty,\infty)$ and (ii) are continuous on $(-\infty,\infty) - \Sigma$, where Σ is an arbitrary fixed finite or infinite set of isolated points.* The norm of f ε $\mathcal{K}(\Sigma)$ is denoted by $\| f \|_\infty$ and is defined by

$$\| f \|_\infty = \sup_t | f(t) | .$$

With this norm $\mathcal{K}(\Sigma)$ is a Banach space.† The norm of a linear operator Q defined on $\mathcal{K}(\Sigma)$ is denoted by $\| Q \|_\infty$ and similarly for the norm of a linear operator defined on $\mathcal{L}_2$.

We shall say that a real-valued function $f(t)$ belongs to $\mathcal{D}$ if and only if there exists a function $\hat{f}(t)$ that agrees with $f(t)$ almost everywhere

---

* Various signals of interest in communication systems such as pulses are not contained in $\mathcal{K}(\Sigma)$ if Σ is the null set.

† Any Cauchy sequence of functions belonging to $\mathcal{K}(\Sigma)$ converges to a bounded continuous function on $(-\infty,\infty) - \Sigma$ (since the sequence converges uniformly on $(-\infty,\infty) - \Sigma$). Since, in addition, the sequence converges at each point of discontinuity, $\mathcal{K}(\Sigma)$ is complete.

and is such that the set of points at which* sign $[\hat{f}(t)]$ is discontinuous is a set of isolated points.

The symbols **I** and **O** are used throughout to denote, respectively, the identity operator and the null operator (i.e., for all f, $\mathbf{O}f = 0$).

III. MATHEMATICAL MODEL OF THE PHYSICAL SYSTEM

Consider a physical system containing linear time-invariant elements and a single time-variable nonlinear element. Let s_1 and s_2, respectively, denote the system's input and output signals and let v and w, respectively, denote the input and output signals associated with the nonlinear device, which is assumed to be characterized by the equation

$$w = \varphi(v,t) = \varphi[v] \tag{1}$$

in which $\varphi(v,t)$ is a real-valued function of the real variables v and t.

We shall consider separately two cases:

(i) $s_1, s_2, v, w \; \varepsilon \; \mathcal{K}(\Sigma)$ for some Σ

(ii) $s_1, s_2, v, w \; \varepsilon \; \mathcal{L}_{2R}$.

It is assumed in each case that there exist well-defined linear operators $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ such that† $v = \mathbf{\Gamma}[s_1, w]$ and $s_2 = \mathbf{\Lambda}[s_1, w]$. It is convenient to define four linear operators **A**, **B**, **C**, and **D** in the following manner

$$\begin{aligned} v &= \mathbf{\Gamma}[s_1, w] = \mathbf{\Gamma}[s_1, 0] + \mathbf{\Gamma}[0, w] \\ &= \mathbf{A}s_1 + \mathbf{C}w \end{aligned} \tag{2}$$

$$\begin{aligned} s_2 &= \mathbf{\Lambda}[s_1, w] = \mathbf{\Lambda}[s_1, 0] + \mathbf{\Lambda}[0, w] \\ &= \mathbf{D}s_1 + \mathbf{B}w. \end{aligned} \tag{3}$$

The relation between s_1, s_2, v, and w is summarized by the flow graph in Fig. 1.

As a very simple illustration of the generality of the graph in Fig. 1, observe that the flow graph of the classical single loop feedback system in Fig. 2(a) in which **E** and **F** are linear operators can readily be reduced to the form shown in Fig. 1. The reduced graph is given in Fig. 2(b).

Our concern is with the influence of the nonlinear element represented by φ. Hence it is sufficient to consider the situation in Fig. 1 in which $\mathbf{D} = \mathbf{O}$ and $\mathbf{A} = \mathbf{B} = \mathbf{I}$. The corresponding graph is shown in Fig. 3. For this graph, using (1), (2), and (3)

$$s_2 = \varphi[\mathbf{C}s_2 + s_1]. \tag{4}$$

---

* Let sign $[\hat{f}(t)] = 1$ when $\hat{f}(t) = 0$.

† This is essentially the same model used by the writer in another study.[3]
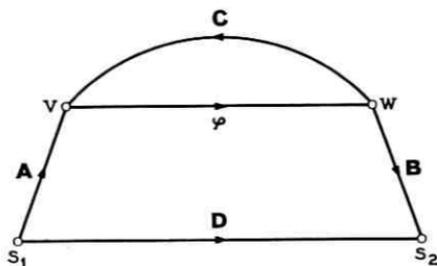
Fig. 1 — Flow-graph representation of a general transmission system containing linear elements and a single time-variable nonlinear element φ.

3.1 The Time-Variable Nonlinear Element and Definition of Signal Distortion

It is assumed throughout that $\varphi[f(t)]$ is measurable whenever f is measurable, that $\varphi(0,t) = 0$ for all t, and that for all t and all $v_1 \geqq v_2$

$$\alpha(v_1 - v_2) \leqq \varphi(v_1,t) - \varphi(v_2,t) \leqq \beta(v_1 - v_2)$$

where α and β are real constants.
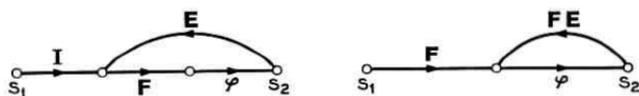
Fig. 2 — Two flow graphs with identical transmission from s_1 to s_2.

In our application of the theorems in Sections V and VII, we shall suppose that $\varphi(v,t) = v + \bar{\varphi}(v,t)$ where, for all t, $\bar{\varphi}(v,t)$ is of order less than v as $v \to 0$. That is,* we shall suppose that for sufficiently small
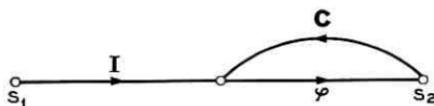
Fig. 3 — Basic flow graph for studying the influence of φ.

input signals the element represented by φ behaves essentially as a unit gain amplifier and hence that, for such signals, the system in Fig. 1 behaves essentially as a linear time-invariant system. Let s_{02} denote the

---

* Note that in the frequently encountered case in which φ is independent of t and φ' is a monotone decreasing function of v, $\beta = 1$.

output signal s_2 in Fig. 1 when $\varphi(v,t) = v$. We shall say that $(s_2 - s_{02})$ is the signal distortion introduced by the departure of $\varphi(v,t)$ from v.

3.2 *The Operator* **C**

Unless stated otherwise, it is assumed that

$$\mathbf{C}f = \int_{-\infty}^{\infty} c(t - \tau)f(\tau)\, d\tau \tag{5}$$

where $c(t)$ is a real-valued function of t. In cases of engineering interest **C** is a causal (i.e., $c(t) = 0, t < 0$). However our mathematical results are not restricted to cases in which **C** is causal.

IV. PRELIMINARY RESULTS

This section is concerned with a proof of

Theorem I: (a) *Let $c(t)$ ε $\mathfrak{L}_{1R}$ and $C(\omega) \neq 1$. Then $(\mathbf{I} - \mathbf{C})$ is a bounded mapping of $\mathfrak{K}(\Sigma)$ into itself that possesses a bounded inverse. In fact, there exists a function $h(t)$ ε $\mathfrak{L}_{1R}$, with Fourier transform $C(\omega)[1 - C(\omega)]^{-1}$, such that*

$$(\mathbf{I} - \mathbf{C})^{-1}g = g + \int_{-\infty}^{\infty} h(t - \tau)g(\tau)\, d\tau$$

for any g ε $\mathfrak{K}(\Sigma)$. If $h(t)$ ε $\mathfrak{D}$,

$$\|(\mathbf{I} - \mathbf{C})^{-1}\|_{\infty} = 1 + \int_{-\infty}^{\infty} |h(t)|\, dt.$$

(b) *Suppose alternatively* that $c(t)$ ε $\mathfrak{L}_{2R}$, ess sup$_{\omega}$ $|C(\omega)| < \infty$, and that inf$_{\omega}$ $|1 - C(\omega)| > 0$. Then $(\mathbf{I} - \mathbf{C})$ is a bounded mapping of $\mathfrak{L}_{2R}$ into itself that possesses a bounded inverse. Moreover*

$$\|(\mathbf{I} - \mathbf{C})^{-1}\|_2 = \text{ess sup}_{\omega} |1 - C(\omega)|^{-1}$$

$$\|\mathbf{C}\|_2 = \text{ess sup}_{\omega} |C(\omega)|.$$

4.1 *Proof of Part (a)*

Since $c(t)$ ε $\mathfrak{L}_{1R}$, the validity of the assertion that $(\mathbf{I} - \mathbf{C})$ is a bounded

---

* The notation ess sup $Q(\omega)$ denotes inf sup $Q(\omega)$ where $\mathfrak{N}$ is an arbitrary zero-measure subset of the real line. In at least almost all cases of engineering interest, the "essential supremum" of the modulus of a Fourier transform is equal to its supremum.

mapping of $\mathcal{K}(\Sigma)$ into itself is obvious. For the remainder of part (a) we need

Lemma I: Let $c(t)$ ε $\mathcal{L}_{1R}$ and $C(\omega) \neq 1$. Then there exists a function $h(t)$ ε $\mathcal{L}_{1R}$, with Fourier transform $C(\omega)$ $[1 - C(\omega)]^{-1}$, such that

$$h(t) - c(t) = \int_{-\infty}^{\infty} h(t - \tau)c(\tau)\, d\tau$$

almost everywhere.

4.1.1 *Proof of Lemma I:*

It is known[4]* that if $c(t)$ ε $\mathcal{L}_{1R}$ and $\inf_{\omega} | 1 - C(\omega) | > 0$, there exists an $f(t)$ of bounded total variation on $(-\infty, \infty)$ such that

$$[1 - C(\omega)]^{-1} = \int_{-\infty}^{\infty} e^{-i\omega t}\, df(t).$$

Under these conditions, it follows that

$$h(t) = \int_{-\infty}^{\infty} c(t - \tau)\, df(\tau)$$

is an element of $\mathcal{L}_{1R}$ which possesses the required Fourier transform. However, since $C(\omega)$ is uniformly continuous and $C(\omega) \to 0$ as $\omega \to \infty$, the inequality $\inf_{\omega} | 1 - C(\omega) | > 0$ is satisfied if and only if $C(\omega) \neq 1$. Thus the assumptions in Lemma I imply the existence of a function $h(t)$ ε $\mathcal{L}_{1R}$ with the stated transform.† Since

$$\frac{C(\omega)}{1 - C(\omega)} - C(\omega) = \frac{C(\omega)}{1 - C(\omega)}\, C(\omega)$$

the Fourier transforms of $[h(t) - c(t)]$ and

$$\int_{-\infty}^{\infty} h(t - \tau)c(\tau)\, d\tau$$

are equal. This establishes the equation stated in the lemma.

Let $g(t)$ denote any element of $\mathcal{K}(\Sigma)$ and assume that there exists an $f(t)$ ε $\mathcal{K}(\Sigma)$ such that

---

* The writer is indebted to V. E. Beneš for directing attention to the result in Ref. 4.

† A moment's reflection will show that when $C(\omega)$ is rational in ω, a proof of this result follows directly from the identification of the terms in its partial-fraction expansion.

$$g(t) = (\mathbf{I} - \mathbf{C})f = f(t) - \int_{-\infty}^{\infty} c(t - \tau) f(\tau) \, d\tau. \tag{6}$$

It is certainly true that (6) implies

$$\int_{-\infty}^{\infty} h(t - \tau) g(\tau) \, d\tau$$

$$= \int_{-\infty}^{\infty} h(t - \tau) f(\tau) \, d\tau - \int_{-\infty}^{\infty} h(t - \tau) \left[\int_{-\infty}^{\infty} c(\tau - u) f(u) \, du \right] d\tau. \tag{7}$$

Since f is bounded and $h, c \ \varepsilon \ \mathcal{L}_{1R}$, Fubini's theorem implies that the last integral can be written as

$$\int_{-\infty}^{\infty} f(u) \left[\int_{-\infty}^{\infty} h(t - \tau) c(\tau - u) \, d\tau \right] du$$

and hence, in accordance with the lemma, as

$$\int_{-\infty}^{\infty} h(t - \tau) f(\tau) \, d\tau - \int_{-\infty}^{\infty} c(t - \tau) f(\tau) \, d\tau. \tag{8}$$

Therefore,

$$\int_{-\infty}^{\infty} h(t - \tau) g(\tau) \, d\tau = \int_{-\infty}^{\infty} c(t - \tau) f(\tau) \, d\tau = f(t) - g(t).$$

Thus, if there exists an $f \ \varepsilon \ \mathcal{K}(\Sigma)$ such that $(\mathbf{I} - \mathbf{C})f = g$,

$$f(t) = g(t) + \int_{-\infty}^{\infty} h(t - \tau) g(\tau) \, d\tau. \tag{9}$$

However, direct substitution and an application of Fubini's theorem show that the right-hand side of (9) is a solution of (6). Hence it is *the* solution.

Since $\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty = \sup\{ \| f \|_\infty : (\mathbf{I} - \mathbf{C})f = g; \ f, g \ \varepsilon \ \mathcal{K}(\Sigma); \ \| g \|_\infty = 1 \}$, it is evident from (9) that

$$\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty \leq 1 + \int_{-\infty}^{\infty} | h(t) | \, dt.$$

We shall next show that if $h(t) \ \varepsilon \ \mathfrak{D}$

$$\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty \geq 1 + \int_{-\infty}^{\infty} | h(t) | \, dt - \delta,$$

where δ is an arbitrary positive number and hence that

$$\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty = 1 + \int_{-\infty}^{\infty} | h(t) | \, dt.$$

Choose a real number t_0 and consider

$$g(t_0) + \int_{-\infty}^{\infty} h(t_0 - \tau)g(\tau)\, d\tau. \tag{10}$$

We may assume that the set of points at which sign $[h(t)]$ is discontinuous is a set of isolated points. Let Ξ denote the union of the closed intervals of length δ_1 centered at t_1 and at each of the discontinuities of sign $[h(t_0 - \tau)]$ regarded as a function of τ. Let Ξ_* denote the complement of Ξ with respect to the real line. For any $\delta > 0$, choose δ_1 such that*

$$\int_{\Xi} |h(t_0 - \tau)|\, d\tau \leq \tfrac{1}{2}\delta.$$

Choose $g(t) \ \varepsilon \ \mathfrak{K}(\Sigma)$ such that $g(t_0) = 1$, $\|g\|_\infty = 1$, and

$$g(\tau) = \text{sign } [h(t_0 - \tau)], \qquad \tau \ \varepsilon \ \Xi_*.$$

Then

$$g(t_c) + \int_{-\infty}^{\infty} h(t_0 - \tau)g(\tau)\, d\tau = 1 + \int_{-\infty}^{\infty} |h(t)|\, dt$$

$$+ \int_{\Xi} h(t_0 - \tau)\{g(\tau) - \text{sign } [h(t_0 - \tau)]\}\, d\tau,$$

and

$$\left| g(t_0) + \int_{-\infty}^{\infty} h(t_0 - \tau)g(\tau)\, d\tau \right| \geq 1 + \int_{-\infty}^{\infty} |h(t)|\, dt - \delta.$$

This completes the proof of the first part of Theorem I.

Of course similar arguments show that if $c(t) \ \varepsilon \ \mathfrak{D}$,

$$\|\mathbf{C}\|_\infty = \int_{-\infty}^{\infty} |c(t)|\, dt.$$

4.2 *Proof of Part* (b):

The proof of this part involves essentially the same arguments presented elsewhere.[3]

Let $f \ \varepsilon \ \mathfrak{L}_{2R}$. Then, using Plancherel's identity,

---

* The integral of $|h(t_0 - \tau)|$ over Ξ does not exceed the sum of the integrals over $|\tau| \geq T$ and $\Xi - [-T,T]_*$ where $[-T,T]_*$ denotes the complement of $[-T,T]$ with respect to the real line. The first integral can be made arbitrarily small by choosing T sufficiently large and with fixed T the second integral can be made arbitrarily small by choosing δ_1 sufficiently small.

$$\| \, \mathbf{C}f \, \|_2^{\,2} \;=\; \frac{1}{2\pi} \int_{-\infty}^{\infty} |\, C(\omega)F(\omega)\,|^2 \, d\omega$$

$$\leqq \; \operatorname*{ess\,sup}_{\omega} \, |\, C(\omega)\,|^2 \, \|\, f \,\|_2^{\,2}.$$

Thus $\mathbf{C}$, and hence $(\mathbf{I} - \mathbf{C})$, are bounded.

Now consider the equation

$$(\mathbf{I} - \mathbf{C})f = g; \qquad g \; \varepsilon \; \mathcal{L}_{2R}, \qquad \|\, g \,\|_2 = 1.$$

Since $\inf_{\omega} |\, 1 - C(\omega)\,| > 0$, there exists a unique solution $f \; \varepsilon \; \mathcal{L}_{2R}$ and $F(\omega) = G(\omega)[1 - C(\omega)]^{-1}$. Again using Plancherel's identity

$$\|\, f \,\|_2^{\,2} = \|\, (\mathbf{I} - \mathbf{C})^{-1}g \,\|_2^{\,2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\, 1 - C(\omega)\,|^{-2} \, |\, G(\omega)\,|^2 \, d\omega \qquad (11)$$

$$\leqq \; \operatorname*{ess\,sup}_{\omega} \, |\, 1 - C(\omega)\,|^{-2} \, \|\, g \,\|_2^{\,2}.$$

Clearly, $\|(\mathbf{I} - \mathbf{C})^{-1}\|_2 \leqq \operatorname*{ess\,sup}_{\omega} |\, 1 - C(\omega)\,|^{-1}$.

According to the definition of the essential supremum of a function, for any $\delta > 0$ there exists a set of values of ω of nonzero measure such that

$$|\, 1 - C(\omega)\,|^{-1} > \operatorname*{ess\,sup}_{\omega} |\, 1 - C(\omega)\,|^{-1} - \delta.$$

Since $|\, G(\omega)\,|$ is permitted to vanish only on the complement of such a set, it is evident from (11) that

$$\|\, (\mathbf{I} - \mathbf{C})^{-1} \,\|_2 \geqq \operatorname*{ess\,sup}_{\omega} |\, 1 - C(\omega)\,|^{-1} - \delta$$

for any $\delta > 0$. Thus, in view of the upper bound on $\|\, (\mathbf{I} - \mathbf{C})^{-1} \,\|_2$,

$$\|\, (\mathbf{I} - \mathbf{C})^{-1} \,\|_2 = \operatorname*{ess\,sup}_{\omega} |\, 1 - C(\omega)\,|^{-1}.$$

A similar argument shows that $\|\, \mathbf{C} \,\|_2 = \operatorname*{ess\,sup}_{\omega} |\, C(\omega)\,|$. This completes the proof of Theorem I.

V. PROPERTIES OF SOLUTIONS TO $s_2 = \varphi[\mathbf{C}s_2 + s_1]$

Theorem II: Let $c(t) \; \varepsilon \; \mathcal{L}_{1R}$, $C(\omega) \neq 1$, *and let* $\varphi[v] = v + \bar{\varphi}[v]$ *be as defined in Section 3.1. Let* $\varphi[f]$ *be continuous with respect to* t *on the complement of* Σ *whenever* $f \; \varepsilon \; \mathcal{K}(\Sigma)$. *Suppose that*

$$\|\, (\mathbf{I} - \mathbf{C})^{-1} \,\|_{\infty} \cdot \|\, \mathbf{C} \,\|_{\infty} \; max \; (|\, 1 - \alpha \,|, |\, \beta - 1 \,|) = r < 1.$$

Then for any $s_1 \; \varepsilon \; \mathcal{K}(\Sigma)$, *there exists a unique* $s_2 \; \varepsilon \; \mathcal{K}(\Sigma)$ *such that* $s_2 =$

$\varphi[\mathbf{C}s_2 + s_1]$. *Furthermore* $s_2 = \lim_{n\to\infty} s_{2n}$ *where*

$$s_{2n} = (\mathbf{I} - \mathbf{C})^{-1}\bar\varphi[\mathbf{C}s_{2(n-1)} + s_1] + (\mathbf{I} - \mathbf{C})^{-1}s_1$$

and s_{20} *is an arbitrary element of* $\mathcal{K}(\Sigma)$. *The nth approximation* s_{2n} *satisfies*

$$\| s_{2n} - s_2 \|_\infty \leqq \frac{r^n}{1 - r} \| s_{21} - s_{20} \|_\infty.$$

Before proceeding to the proof of Theorem II we state the analogous result for the space $\mathcal{L}_{2R}$.

Theorem III: Let $c(t) \; \varepsilon \; \mathcal{L}_{2R}$, $\sup_\omega | C(\omega) | < \infty$, $\inf_\omega | 1 - C(\omega) | > 0$, *and let* $\varphi[v] = v + \bar\varphi[v]$ *be as defined in Section 3.1 Suppose that*

$$\| (\mathbf{I} - \mathbf{C})^{-1} \|_2 \cdot \| \mathbf{C} \|_2 \; max \; (|1 - \alpha |, | \beta - 1 |) = r < 1.$$

Then the conclusion of Theorem II follows with $\mathcal{K}(\Sigma)$ *replaced with* $\mathcal{L}_{2R}$ *and with the* $\mathcal{K}(\Sigma)$ *norm replaced with the* $\mathcal{L}_{2R}$ *norm.*

5.1 *Proof of Theorem II*

We have

$$s_2 = \mathbf{C}s_2 + s_1 + \bar\varphi[\mathbf{C}s_2 + s_1]$$

and hence, in view of the first part of Theorem I and the assumption that $C(\omega) \neq 1$, $s_2 = \mathbf{L}s_2$ where

$$\mathbf{L}s_2 = (\mathbf{I} - \mathbf{C})^{-1}\bar\varphi[\mathbf{C}s_2 + s_1] + (\mathbf{I} - \mathbf{C})^{-1}s_1.$$

It is evident that $\mathbf{L}$ is a mapping of $\mathcal{K}(\Sigma)$ into itself. We shall show that under the conditions stated in the theorem $\mathbf{L}$ is in fact a contraction mapping of $\mathcal{K}(\Sigma)$ into itself. Let $f, g \; \varepsilon \; \mathcal{K}(\Sigma)$ and observe that

$$\| \mathbf{L}f - \mathbf{L}g \|_\infty = \| (\mathbf{I} - \mathbf{C})^{-1}\{\bar\varphi[\mathbf{C}f + s_1] - \bar\varphi[\mathbf{C}g + s_1]\} \|_\infty$$
$$\leqq \| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty \cdot \| \bar\varphi[\mathbf{C}f + s_1] - \bar\varphi[\mathbf{C}g + s_1] \|_\infty$$

and that

$$\| \bar\varphi[\mathbf{C}f + s_1] - \bar\varphi[\mathbf{C}g + s_1] \|_\infty$$
$$= \left\| \left(\frac{\varphi[\mathbf{C}f + s_1] - \varphi[\mathbf{C}g + s_1]}{\mathbf{C}f - \mathbf{C}g} - 1\right) \mathbf{C}(f - g) \right\|_\infty$$
$$\leqq max \; (| 1 - \alpha |, | \beta - 1|) \| \mathbf{C}(f - g) \|_\infty$$
$$\leqq max \; (| 1 - \alpha |, | \beta - 1 |) \| \mathbf{C} \|_\infty \cdot \| f - g \|_\infty.$$

Thus **L** is a contraction when $r < 1$. This proves Theorem II with the exception of the last inequality* which follows directly from the fact that s_2 can be written as

$$s_2 = s_{20} + \sum_{j=0}^{\infty} [s_{2(j+1)} - s_{2j}] \tag{12}$$

in which for all $j \geqq 1$

$$\| s_{2(j+1)} - s_{2j} \|_{\infty} = \| \mathbf{L} s_{2j} - \mathbf{L} s_{2(j-1)} \|_{\infty} \leqq r \| s_{2j} - s_{2(j-1)} \|_{\infty} .$$

5.2 *Proof of Theorem III:*

With obvious modifications the proof of Theorem II suffices.

VI. A CASCADE GRAPH FOR APPROXIMATELY DETERMINING THE SIGNAL s_2 IN FIG. 3

Suppose that the input signal s_1 in Fig. 3 is an element of $\mathcal{K}(\Sigma)$. Then under the assumptions stated in Theorem II, the output signal s_2 is an element of $\mathcal{K}(\Sigma)$ and is given by (12) where s_{20} is an arbitrary element of $\mathcal{K}(\Sigma)$. The key inequality that must be satisfied is† (using the first part of Theorem I and assuming that c, $h \ \varepsilon \ \mathfrak{D}$)

$$\int_{-\infty}^{\infty} | c(t) | \, dt \left(1 + \int_{-\infty}^{\infty} | h(t) | \, dt \right) \tag{13}$$

$$\cdot \max (| 1 - \alpha |, | \beta - 1 |) = r < 1.$$

If we take $s_{20} = (\mathbf{I} - \mathbf{C})^{-1} s_1$, the sum $\sum_{j=0}^{\infty} [s_{2(j+1)} - s_{2j}]$ represents the nonlinear distortion present in the output. The first term in this series is

$$(s_{21} - s_{20}) = (\mathbf{I} - \mathbf{C})^{-1} \bar{\varphi}[(\mathbf{I} - \mathbf{C})^{-1} s_1]$$

and, using the inequality in Theorem II, a bound on the error incurred in ignoring the remainder of the series is given by

$$\left\| (s_{21} - s_{20}) - \sum_{j=0}^{\infty} [s_{2(j+1)} - s_{2j}] \right\|_{\infty} = \| s_{21} - s_2 \|_{\infty}$$

$$\leqq \frac{r}{1 - r} \| s_{21} - s_{20} \|_{\infty} .$$

---

* Note that when $s_{20} = 0$, this inequality implies that

$$\| s_2 \|_{\infty} \leqq (1 - r)^{-1} \| s_{21} \|_{\infty} .$$

† In physical systems both integrands vanish for negative arguments.

That is, the error in ignoring the remainder of the series is at most $r(1 - r)^{-1}$ times the norm of the first term. Thus if r is sufficiently small the function $(s_{21} - s_{20})$ is a good approximation to the distortion component of the signal s_2. Fig. 4 shows the corresponding flow graph for determining s_{21}, the approximation to the output signal in Fig. 3. Theorem III leads to analogous results and the same flow graph for the case in which signals belong to $\mathcal{L}_{2R}$. The essential difference is that in the $\mathcal{L}_{2R}$ case attention is focused on the energies of the signals.

6.1 *Relation of the Graph in Fig. 4 to a Well-Known Engineering Technique*

The flow graph in Fig. 4 characterizes the essence of a well-known engineering technique[5,6,7] for approximately determining the effect of feedback on nonlinear distortion introduced in one stage of an amplifier, when the distortion is "small." In particular, observe that if, as indicated in Fig. 4, $u(t)$ denotes the distortion* produced by the open-loop system
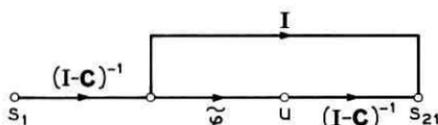
Fig. 4 — Cascade flow graph for approximately determining the output signal in Fig. 3 (s_{21} is the approximation to s_2).

with the same "small-signal transmission" and the same output stage as the feedback system in Fig. 3, then the output distortion in Fig. 3 is approximately $(\mathbf{I} - \mathbf{C})^{-1}u$. In engineering terms, feedback is said[5,6,7] to reduce the nonlinear distortion by the amount of the "return difference" $[1 - C(\omega)]$ [i.e., by the formal frequency domain representation of the operator $(\mathbf{I} - \mathbf{C})$].

6.2 *Comparison with Desoer's Results*

In an interesting paper[1] Desoer has considered the range of validity of the graph in Fig. 4 for the case in which $\varphi(v,t) = v + \epsilon v^m$, where ϵ and m are real constants with m an odd positive integer. In his discussion all signals belong to $\mathcal{K}(\Sigma)$ with Σ the null set. He considers the analysis of a feedback system of the type shown in Fig. 2(a) and argues that if the norm of the input to the amplifier is sufficiently small and if $|\epsilon|$ is sufficiently small,† then the distortion component of the output

---

* The "distortion generator" referred to in the usual engineering arguments produces the signal $u(t)$.

† This writer feels that some additional restriction on $|\epsilon|$ is necessary in order that Desoer's condition (B) be satisfied. It would suffice to assume that $|\epsilon| < 3/16$.

signal is given by the sum of an infinite series in which the norm of each term is less than one-fourth the norm of the preceding term. When the system is characterized by a flow graph of the type shown in Fig. 3 [as indicated earlier, the analysis of the seemingly more complicated situation in Fig. 2(a) can be reduced at once to a consideration of this type of graph], the first term in the series is the distortion determined from Fig. 4.

Desoer focuses attention on simplifications that can be exploited in cases of engineering interest. According to him it is a matter of experience that for a typical low-pass feedback amplifier $\| C \|_\infty = -C(0)$ [observe that in general $\| C \|_\infty \geqq \sup_\omega | C(\omega) |$]. In addition he presents a heuristic argument to support the claim that in such amplifiers $\| (I - C)^{-1} \|_\infty$ is approximately equal to 2. With

$$1 + \int_{-\infty}^{\infty} | h(t) | \, dt = \| (I - C)^{-1} \|_\infty = 2,$$

the condition that r [in (13)] be less than $\frac{1}{4}$ (i.e., the condition corresponding to Desoer's criterion for determining the applicability of the graph in Fig. 4) is

$$\max \left(| 1 - \alpha |, | \beta - 1 | \right) < \frac{1}{8 \, \| C \|_\infty}. \tag{14}$$

It is a routine matter to show that in the high loop-gain case (i.e., the case of principal engineering interest) (14) is a much less stringent condition on the permitted degree of nonlinearity than that implied by Desoer's upper bound on $| \epsilon |$ and his input norm bound. For a loop gain of 100 [i.e., $C(0) = -100$] and $m = 3$, the bound on $| \epsilon |$ is such that (14) permits any deviation from unity of the slope of $\varphi(v)$ [i.e., max $(| 1 - \alpha |, | \beta - 1 |)$] which does not exceed 2,500 times the permitted maximum deviation from unity of the slope of $v + \epsilon v^3$ over the operating range*† implied by Desoer's input norm bound.

6.3 An Extension of Theorem II

Note that when the loop gain is large the permitted amount of nonlinearity in (14) is quite small. Although it is not clear whether the range of validity of the graph in Fig. 4 can be substantially increased, it is a simple matter to show that the iterates s_{2n} defined in Theorem II

---

* In the notation of Ref. 1, the signal input to the nonlinear element is $z - \mu\beta\zeta$ and $\| \zeta \| \leqq \frac{4}{3} \| \zeta_1 \| \leqq \frac{8}{3} | \epsilon |$.
† Desoer's bound on $| \epsilon |$ can be considerably improved in the large loop-gain case by assuming that ζ lies within a ball of much smaller radius (such as $\| \mu\beta \|^{-1}$).

converge to the unique $s_2 \varepsilon \mathcal{K}(\Sigma)$ that satisfies $s_2 = \varphi[\mathbf{C}s_2 + s_1]$ even if $r \geqq 1$, provided that

$$\| \mathbf{C}(\mathbf{I} - \mathbf{C})^{-1} \|_\infty \max (|1 - \alpha|, |\beta - 1|) < 1.$$

This follows from the contraction-mapping fixed-point theorem and the fact that the relation between s_1 and s_2 can be written as $s_3 = s_1 + \tilde{\varphi}[\mathbf{C}(\mathbf{I} - \mathbf{C})^{-1}s_3 + s_1]$ where $s_3 = (\mathbf{I} - \mathbf{C})s_2$.

In the next section we consider an alternative, but closely related, cascade graph for determining the output distortion in Fig. 3. For a given loop gain (assuming it is large) the alternative graph is valid for much larger amounts of nonlinearity than that indicated in (13) or (14).

VII. ADDITIONAL RESULTS RELATING TO THE EQUATION $s_2 = \varphi[\mathbf{C}s_2 + s_1]$

In this section we assume that there exists a function ψ such that $\psi[\varphi(x)] = x$ for all real x and t. Hence $\psi[s_2] = \mathbf{C}s_2 + s_1$. Specifically

Definition: Let $\psi(x,t) = x + \bar{\psi}(x,t)$ be a real-valued function of the real variables x and t such that $\psi(0,t) = 0$ for all t, and that for all t and all $x \geqq y$

$$\gamma(x - y) \leqq \psi(x,t) - \psi(y,t) \leqq \sigma(x - y)$$

where γ and σ are real constants.

Theorem IV: Let $\psi[f]$ be continuous with respect to t on the complement of Σ whenever $f \varepsilon \mathcal{K}(\Sigma)$. Let $c(t) \varepsilon \mathcal{L}_{1R}$, $C(\omega) \neq 1$, and suppose that

$$\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty \max (|1 - \gamma|, |\sigma - 1|) = q < 1.$$

Then for any $s_1 \varepsilon \mathcal{K}(\Sigma)$, there exists a unique $s_2 \varepsilon \mathcal{K}(\Sigma)$ such that $\psi[s_2] = \mathbf{C}s_2 + s_1$. In fact, $s_2 = \lim\limits_{n \to \infty} \bar{s}_{2n}$ where

$$\bar{s}_{2n} = -(\mathbf{I} - \mathbf{C})^{-1}\bar{\psi}[\bar{s}_{2(n-1)}] + (\mathbf{I} - \mathbf{C})^{-1}s_1$$

and $\bar{s}_{20}$ is an arbitrary element of $\mathcal{K}(\Sigma)$. The nth approximation $\bar{s}_{2n}$ satisfies

$$\| s_2 - \bar{s}_{2n} \|_\infty \leqq \frac{q^n}{1 - q} \| \bar{s}_{21} - \bar{s}_{20} \|_\infty.$$

The analogous result for the space $\mathcal{L}_{2R}$ is

Theorem V: Let $c(t) \varepsilon \mathcal{L}_{2R}$, $\sup\limits_{\omega} |C(\omega)| < \infty$, $\inf\limits_{\omega} |1 - C(\omega)| > 0$, and suppose that

$$\|(\mathbf{I} - \mathbf{C})^{-1}\|_2 \max (|1 - \gamma|, |\sigma - 1|) = q < 1.$$

Then the conclusion of Theorem IV follows with $\mathcal{K}(\Sigma)$ replaced with $\mathcal{L}_{2R}$ and with the $\mathcal{K}(\Sigma)$ norm replaced with the $\mathcal{L}_{2R}$ norm.

7.1 *Proof of Theorem IV*

In view of the first part of Theorem I and the assumption that $C(\omega) \neq 1$, the equation $\psi[s_2] = \mathbf{C}s_2 + s_1$ can be written as $s_2 = \mathbf{M}s_2$, where $\mathbf{M}$ is the mapping of $\mathcal{K}(\Sigma)$ into itself defined by

$$\mathbf{M}s_2 = -(\mathbf{I} - \mathbf{C})^{-1}\tilde{\psi}[s_2] + (\mathbf{I} - \mathbf{C})^{-1}s_1.$$

Thus, Theorem IV follows from the contraction-mapping fixed-point theorem and the readily verified fact that

$$\| \mathbf{M}f - \mathbf{M}g \|_{\infty} \leqq q \| f - g \|_{\infty}, \qquad q < 1$$

for any $f, g \; \varepsilon \; \mathcal{K}(\Sigma)$. With obvious modifications this argument suffices to establish Theorem V.

7.2 *A Cascade Graph for Approximately Determining the Signal s_2 in Fig. 3 when $\| \mathbf{C} \|_{\infty}$ Is Large*

As in the discussion of Theorem II, if we set $\bar{s}_{20} = (\mathbf{I} - \mathbf{C})^{-1}s_1$ in Theorem IV,

$$s_2 = \bar{s}_{20} + \sum_{j=0}^{\infty} [\bar{s}_{2(j+1)} - \bar{s}_{2j}],$$

in which $\sum_{j=0}^{\infty} [\bar{s}_{2(j+1)} - \bar{s}_{2j}]$ represents the nonlinear distortion component of the output signal in Fig. 3. The first term in this series is

$$(\bar{s}_{21} - \bar{s}_{20}) = -(\mathbf{I} - \mathbf{C})^{-1}\tilde{\psi}[(\mathbf{I} - \mathbf{C})^{-1}s_1], \tag{15}$$

and the error incurred in ignoring the remainder of the series $\left(\text{i.e., } \left\| (\bar{s}_{21} - \bar{s}_{20}) - \sum_{j=0}^{\infty} [\bar{s}_{2(j+1)} - \bar{s}_{2j}] \right\|_{\infty}\right)$ is, as in Section VI, at most $q(1 - q)^{-1}$ times the norm of the first term [(i.e., of (15)].

Thus if

$$q = \| (\mathbf{I} - \mathbf{C})^{-1} \|_{\infty} \max (| 1 - \gamma |, | \sigma - 1 |) \tag{16}$$

is sufficiently small, $(\bar{s}_{21} - \bar{s}_{20})$ is a good approximation to the distortion component of s_2. Fig. 5 shows the corresponding flow graph* for determining $\bar{s}_{21}$, the approximation to the output signal s_2 in Fig. 3.

* An analogous interpretation of Theorem V leads to the same flow graph.
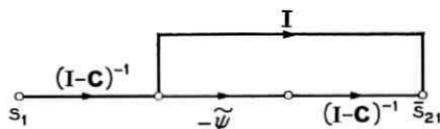
Fig. 5 — Alternative cascade flow graph for approximately determining the output signal in Fig. 3 ($\bar{s}_{21}$ is the approximation to s_2).

Observe that this graph differs* from the one in Fig. 4 only in that $\tilde{\varphi}$ is replaced with $-\tilde{\psi}$. However the expression for q in (16), unlike the corresponding expression for r in Theorem II, does not contain the factor $\| \mathbf{C} \|_\infty$, which is a measure of the amount of feedback present in the system (as indicated earlier, in low-pass feedback amplifiers typically $\| \mathbf{C} \|_\infty = | C(0) |$).

The condition that q in (16) be less than $\frac{1}{4}$ when $\| (\mathbf{I} - \mathbf{C})^{-1} \|_\infty = 2$ (i.e., the condition corresponding to Desoer's practical criterion for determining the applicability of the graph in Fig. 4) is clearly†

$$\max (\,| 1 - \gamma |, | \sigma - 1 |\,) < 1/8.$$

VIII. ACKNOWLEDGMENT

The writer is indebted to V. E. Beneš for reading the draft.

APPENDIX

The following inequality, in which $\mathfrak{N}$ denotes either $\mathfrak{K}(\Sigma)$ or $\mathcal{L}_{2R}$, can be used in some cases to compare the output signals in Figs. 4 and 5

$$\| \mathbf{Q}\tilde{\varphi}[f] + \mathbf{Q}\tilde{\psi}[f] \| \leq \frac{p}{1 - p} \| \mathbf{Q}\tilde{\psi}[f] \| \qquad (17)$$

where f is an arbitrary element of $\mathfrak{N}$, $\| \cdot \|$ denotes the norm for the space $\mathfrak{N}$, and $\mathbf{Q}$ is any linear operator defined on $\mathfrak{N}$ such that

$$\| \mathbf{Q} \| \cdot \| \mathbf{Q}^{-1} \| \max (\,| 1 - \gamma |, | \sigma - 1 |\,) = p < 1.$$

---

* In the Appendix a general inequality is presented which in some cases can be used to bound the norm of the difference between the two approximations to the distortion component of s_2, $(s_{21} - s_{20})$ in Fig. 4 and $(\bar{s}_{21} - \bar{s}_{20})$ in Fig. 5, in terms of the norm of $(\bar{s}_{21} - \bar{s}_{20})$. The inequality is not applicable when $\| \mathbf{C} \|_\infty$ is large unless $\max (\,| 1 - \gamma |, | \sigma - 1 |\,)$ is sufficiently small.

† It is sometimes desirable to consider the unrealistic and very much simpler situation in Fig. 3 when $\mathbf{C}$ represents multiplication by a real constant c. In that case the relation between s_1 and s_2 can be written as $s_2 = c^{-1}\psi[s_2] - c^{-1}s_1$, from which it is evident that, provided $| c |$ is sufficiently large, the contraction-mapping fixed-point theorem is applicable with a small contraction constant even when ψ represents a highly nonlinear element.

Inequality (17) follows from the fact that $g = \mathbf{Q}\varphi[f]$ is the fixed point of the contraction-mapping $\mathbf{N}$ (with contraction constant p) defined by $\mathbf{N}g = \mathbf{Q}f - \mathbf{Q}\check{\psi}[\mathbf{Q}^{-1}g]$.

REFERENCES

1. Desoer, C. A., Nonlinear Distortion in Feedback Amplifiers, I.R.E. Trans. on Circuit Theory, **CT-9,** March, 1962, pp. 2–6.
2. Kolmogorov, A. N. and Fomin, S. V., *Elements of the Theory of Functions and Functional Analysis*, New York, Graylock Press, 1957.
3. Sandberg, I. W., On The Properties of Some Systems That Distort Signals — II, to be published.
4. Pitt, H. R., *Tauberian Theorems*, Bombay, Oxford University Press, 1958, p. 99.
5. Bode, H. W., *Network Analysis and Feedback Amplifier Design*, New York, D. Van Nostrand, 1945.
6. Gray, T. S., *Applied Electronics*, New York, John Wiley and Sons, 1954.
7. Angelo, E. J., Jr., *Electronic Circuits*, New York, McGraw-Hill, 1958.

A Three-Conductor Elementary Clogston Coaxial Transmission Line—Calculation, Fabrication and Experiment

By J. M. MANLEY

Eddy current losses in a conductor of a transmission line can be reduced if the conductor can be divided into two parallel shells and the current divided evenly between these two. According to the Clogston theory, this may be achieved if the dielectric constant of the space between these two shells is less than that of the lines' main dielectric by the right amount.

This paper describes the calculation of the properties of a three-conductor Clogston line, its fabrication, the measurement of its attenuation, and additional calculations relating to the termination of the line.

The calculations indicate that with a line having this structure attenuation can be reduced 24 per cent at 4 mc and about 15 per cent at 1 mc and 10 mc from that of a conventional coaxial line having the same dimensions as the one studied here. A reduction of 18 per cent instead of 24 per cent was realized in the 600-foot length of line which was built. About one-half of this difference is a result of the ideal ratio of dielectric constants not having been attained in this first attempt.

It was found that ideal termination of the line is difficult. However, a simple approximation to an ideal termination yields an insertion loss for the line only a small amount larger than the ideal, provided the length of line is such that its loss is about 20 db.

I. INTRODUCTION

A three-conductor coaxial cable, in which the locations of the two inner conductors are transposed at regular intervals, was built and shown to have about 20 per cent lower attenuation over a certain frequency interval than a two-conductor coaxial of the same size.[1] This is because the eddy current losses are smaller, and this in turn is a result of the normal central conductor having been divided into two and the normal central current having been divided nearly evenly between these two.

It is possible to do this so that the two conductors occupy approximately the same space as the single one did before, since the skin effect causes most of the current to flow in a thin shell at the outer periphery of the inner conductor over a certain frequency interval. Dividing the central conductor provides two shells for current flow.

One way to achieve equal currents in the two inner shells is to transpose the two inner conductors frequently enough in the manner of litz wire and so force the even division. This is the method used in the project described in Ref. 1. Another method is to so choose the dimensions and properties of the two dielectric spaces that the line has a natural mode of propagation with approximately the desired current distribution. This is a special case of the method suggested by Clogston a number of years ago,[2] namely, that the conductors of a transmission line be made of many laminations with proper dielectric separation. Other work has also been done on the three-conductor case.[3] It is also possible to obtain the desired current distribution by having the inductances associated with the several spaces in the proper ratio. This has been done in experimental lines built in Japan.[4] The proper values of inductance are achieved by having spiral gaps of increasing pitch as the center of the cable is approached.

The Clogston method which is used in the experiment described in this paper depends on having the proper ratio of dielectric constants in the several spaces. Fig. 1 is a pictorial representation of the Clogston line built for our experiment. This line is uniform with respect to length, in contrast with the transposed line. Fig. 2 is an enlarged photograph of the central member of the cable.

Following Clogston's work on laminated conductors for transmission lines,[2] an experimental line having 100 laminations was built.[5] Measurements on this cable did not show the expected reduction of attenuation.[5]
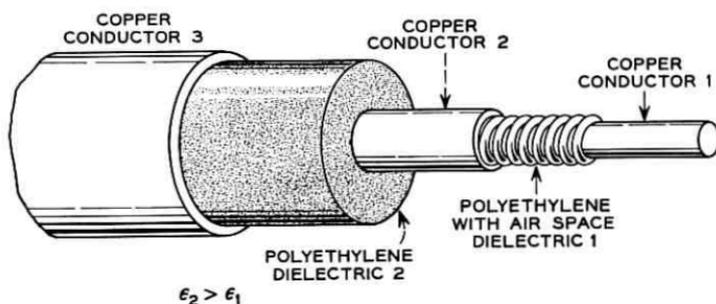

Fig. 1 — Drawing showing construction of three-conductor Clogston transmission line.
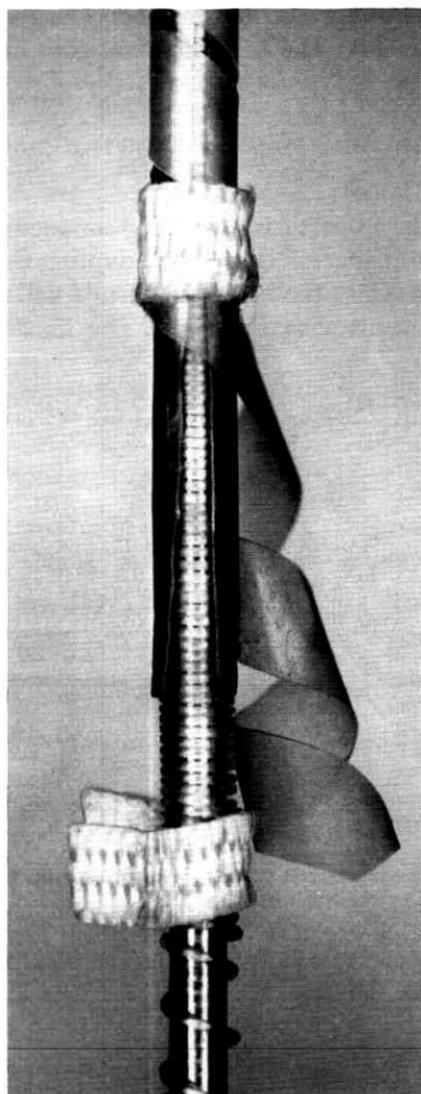
Fig. 2 — Enlarged photograph of two-conductor central member of Clogston line.

It was suspected that nonuniformity in spacing and thickness of laminations had a part in the discrepancy, and some calculations of this were made.[6] But the discrepancy was not resolved until the work[7] by Gordon Raisbeck was carried out. Meanwhile, it was decided to build the simplest laminated line, namely one in which the central member con-

sisted of two conductors instead of one. The first result of this program was the transposed line.[1] The next step was the three-conductor Clogston line project reported in the present paper.

II. CALCULATION OF THE PROPERTIES OF THE THREE-CONDUCTOR CO-AXIAL LINE

The transmission properties of a three-conductor coaxial structure were derived in the paper of Ref. 1 by adapting Schelkunoff's general results[8] to this case. Only the necessary results will be repeated here.

Sections of the line are shown in Fig. 3. This illustration and the nota-
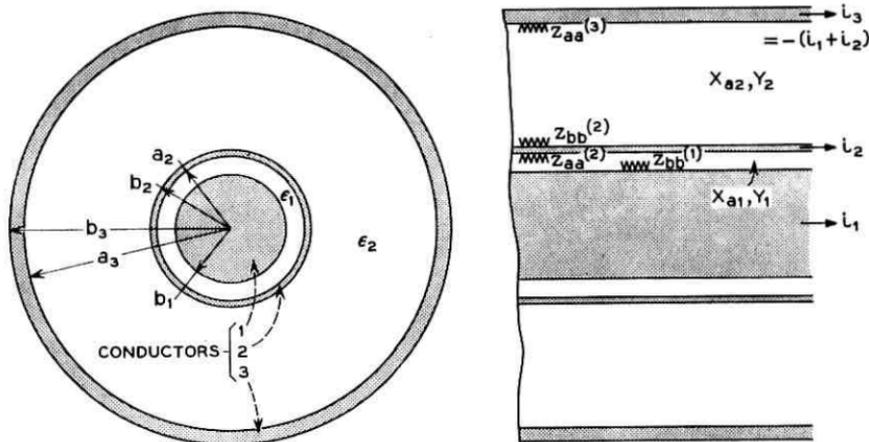
Fig. 3 — The three-conductor Clogston line: cross section and schematic longitudinal section showing surface impedances.

tion used here are the same as in Ref. 1. The conductors are numbered 1, 2, 3 beginning at the innermost. The currents flowing in these conductors and the voltages on them are designated i_1, i_2, i_3; v_1, v_2, v_3 in same order. It is seen that

$$-i_3 = i_1 + i_2 \quad \text{and}$$

$$v_3 = 0.$$

The behavior of the line is described by the two homogeneous equations

$$[Z_{11} - \gamma^2/Y_1]i_1 - Z_{ab}(i_1 + i_2) = 0$$
$$-Z_{ab}i_1 + [Z_{22} - \gamma^2/Y_2](i_1 + i_2) = 0 \tag{1}$$

and two others

$$v_2 - v_1 = -\gamma i_1/Y_1$$
$$v_2 = \gamma(i_1 + i_2)/Y_2 \tag{2}$$

where

$$Z_{11} = Z_{bb}^{(1)} + Z_{aa}^{(2)} + jX_{a_1} \tag{3}$$

is the effective series impedance of the simple coaxial line consisting of conductors 1 and 2 only and where

$$Z_{22} = Z_{bb}^{(2)} + Z_{aa}^{(3)} + jX_{a_2} \tag{4}$$

is the effective series impedance of the simple coaxial line consisting of conductors 2 and 3 only. In these expressions $Z_{bb}^{(m)}$ and $Z_{aa}^{(m)}$ are the outer and inner surface impedances, respectively, of conductor m, using Schelkunoff's concept and notation described in Ref. 8. The surface impedances are fictitious coefficients which gather the impedance effects of a conductor at its surfaces. $Z_{ab}^{(2)}$ is the surface *transfer* impedance of conductor 2; γ is the propagation constant. Also

$$jX_{a_1} = j\mu \frac{\omega}{2\pi} \log (a_2/b_1)$$

$$Y_1 = j \frac{2\pi\omega\epsilon_1}{\log (a_2/b_1)}$$

and

$$jX_{a_2} = j\mu \frac{\omega}{2\pi} \log (a_3/b_2)$$

$$Y_2 = j \frac{2\pi\omega\epsilon_2}{\log (a_3/b_2)} \tag{5}$$

are the series impedances and shunt admittances of the two dielectric spaces between conductors 1 and 2 and between conductors 2 and 3, respectively. All these coefficients are per unit length of line.

The propagation of voltages and currents in the natural modes of the line are governed by the four values of γ which alone permit the two homogeneous equations (1) to be satisfied. These values of γ are roots of the characteristic equation

$$\gamma^4 - \gamma^2(Y_1Z_{11} + Y_2Z_{22}) + Y_1Y_2(Z_{11}Z_{22} - Z_{ab}^2) = 0 \tag{6}$$

obtained by setting the determinant of coefficients of the homogeneous

equations equal to zero. The roots occur in pairs, so that we have

$$\gamma_1 = -\gamma_2$$
$$\gamma_3 = -\gamma_4.$$
(7)

For the dimensions considered here, γ_1 and γ_3 are associated with low- and high-loss modes respectively.

If it is assumed a priori that the currents i_1 and i_2 in the two laminations are equal, and if the dimensions and impedances of the three-conductor cable used for the transposed line are taken, then it is found that attenuation of the low-loss mode is reduced about 30 per cent below that of a comparable two-conductor coaxial. However, these assumptions require the ratio of dielectric constants to be complex, viz.

$$\epsilon_2/\epsilon_1 = 1.23 - j\,0.03$$
(8)

which implies dissipation in at least one of the spaces. This would introduce additional loss which has not been considered.

Instead of doing further analytical work along these lines, we resorted to numerical computation of the line properties, using again the dimensions of the uniform three-conductor cable made for the transposed line (OD about 0.18 inch), but with real values for the ratio ϵ_1/ϵ_2. The values of attenuation constant versus ratio ϵ_1/ϵ_2 of dielectric constants for several values of spacing between conductors 1 and 2 found in these calculations are plotted in Fig. 4. With these real values for the ratio ϵ_1/ϵ_2, there is an appreciable (about 24 per cent) reduction of attenuation at the minimum. Near this minimum, the ratio of currents in conductors 1 and 2 is nearer 1.2, as may be seen in Fig. 5, where the ratio of voltages is plotted also. It will be noticed in Fig. 4 that the curve of attenuation versus ϵ_1/ϵ_2 is considerably sharper where the lamination spacing is 15 mils instead of 5 mils. In Fig. 6 the attenuation constants α_1 and α_3 of the low- and high-loss modes, respectively, and also that of a reference two-conductor coaxial, are plotted versus frequency. This reference coaxial has the same size outer conductor as that of the special cable and a solid inner conductor with an OD very nearly the same as that of conductor 2; the dielectric is specified by ϵ_2. The shape of the α_1 curve is similar to that for the transposed line, except that attenuation at low frequencies approaches that of the reference two-conductor coaxial, whereas attenuation in the transposed line was somewhat higher. The phase distortion of the three-conductor Clogston and the two-conductor reference line are plotted in Fig. 7. For further comparison of the Clogston line, attenuation and phase distortion of the two-conductor
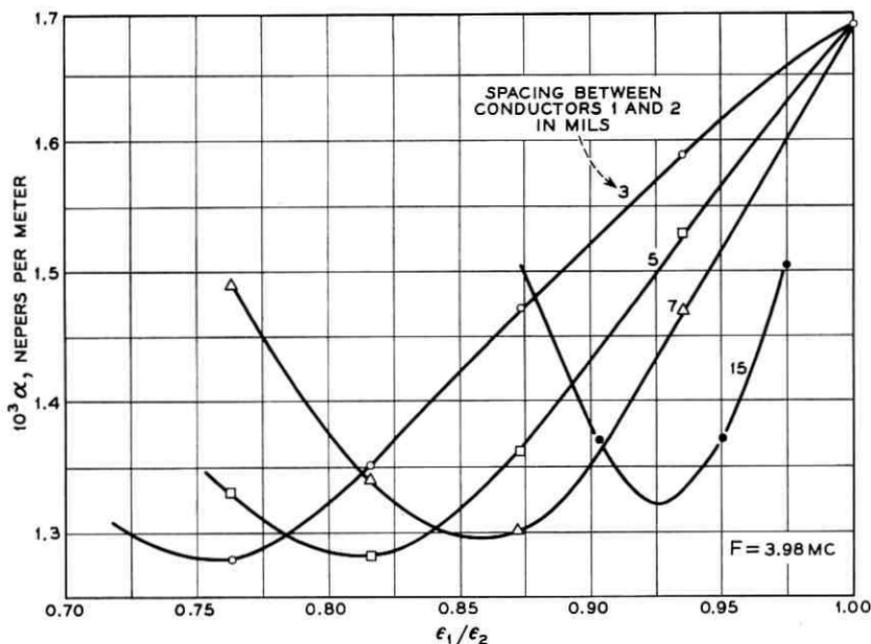
Fig. 4 — Calculated attenuation of three-conductor Clogston line.

coaxial consisting of conductors 2 and 3 — that is, two concentric tubes — are shown in Figs. 6 and 7, respectively.

Other properties of the line are shown in Figs. 8, 9, 10 and 11. Z_1 and Z_2 are the characteristic impedances between conductors 1 and 3 and between 2 and 3, respectively, for the low-loss mode with propagation constant γ_1. For the high-loss mode, the corresponding impedances are z_1 and z_2. These are the ratios of voltage and current amplitude for each of the natural traveling waves which can appear on the line. This may be seen clearly when the total line currents and voltages are expressed in terms of the natural traveling waves which are the solutions of the line equations (1)

$$i_1(x) = i_{11}e^{\gamma_1 x} + i_{21}e^{-\gamma_1 x} + i_{31}e^{\gamma_3 x} + i_{41}e^{-\gamma_3 x}$$
$$i_2(x) = i_{12}e^{\gamma_1 x} + i_{22}e^{-\gamma_1 x} + i_{32}e^{\gamma_3 x} + i_{42}e^{-\gamma_3 x}$$
$$v_1(x) = v_{11}e^{\gamma_1 x} + v_{21}e^{-\gamma_1 x} + v_{31}e^{\gamma_3 x} + v_{41}e^{-\gamma_3 x}$$
$$v_2(x) = v_{12}e^{\gamma_1 x} + v_{22}e^{-\gamma_1 x} + v_{32}e^{\gamma_3 x} + v_{42}e^{-\gamma_3 x}.$$

(9)
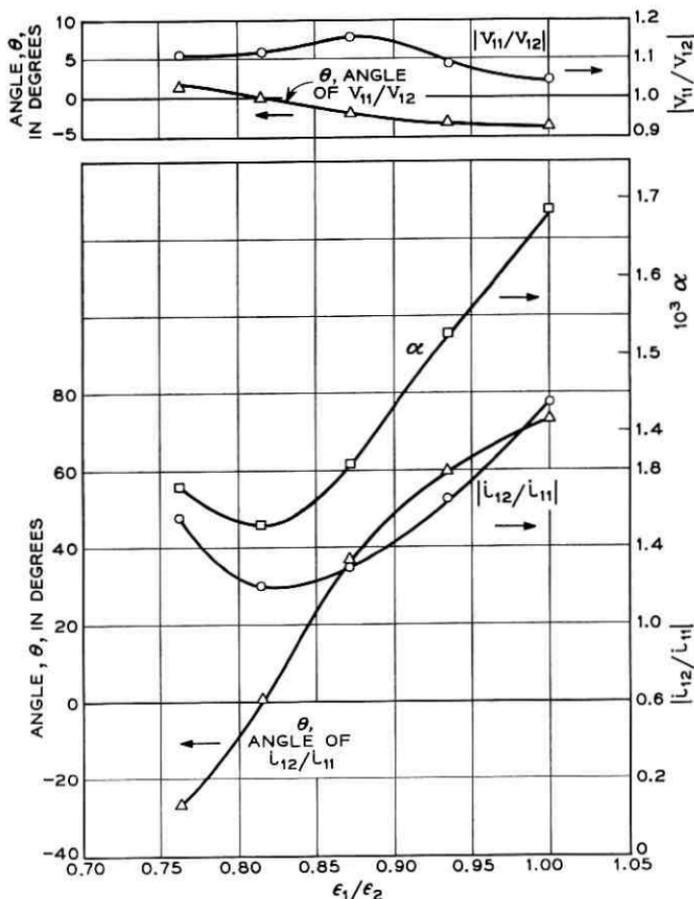
Fig. 5 — Calculated current and voltage ratios in three-conductor Clogston line.

The first subscript refers to the mode and the second to the conductor. For example, i_{11} is the current associated with wave of mode 1 in conductor 1, and i_{21} is that for the wave of mode 2 in conductor 1. Mode 2, as noted in (7), is the backward traveling wave with propagation constant $-\gamma_1$. Thus, modes 1 and 2 are the forward and backward waves having low attenuation γ_1 and modes 3 and 4 are the forward and backward waves having high attenuation γ_3. From these facts, we have the following relations

$$Z_1 = \frac{v_{11}}{i_{11}} = -\frac{v_{21}}{i_{21}}$$

$$Z_2 = \frac{v_{12}}{i_{12}} = -\frac{v_{22}}{i_{22}}$$

$$z_1 = \frac{v_{31}}{i_{31}} = -\frac{v_{41}}{i_{41}}$$

$$z_2 = \frac{v_{32}}{i_{32}} = -\frac{v_{42}}{i_{42}}$$

$$h_1 = \frac{v_{12}}{v_{11}} = \frac{v_{22}}{v_{21}}$$

$$h_3 = \frac{v_{32}}{v_{31}} = \frac{v_{42}}{v_{41}}.$$

(10)
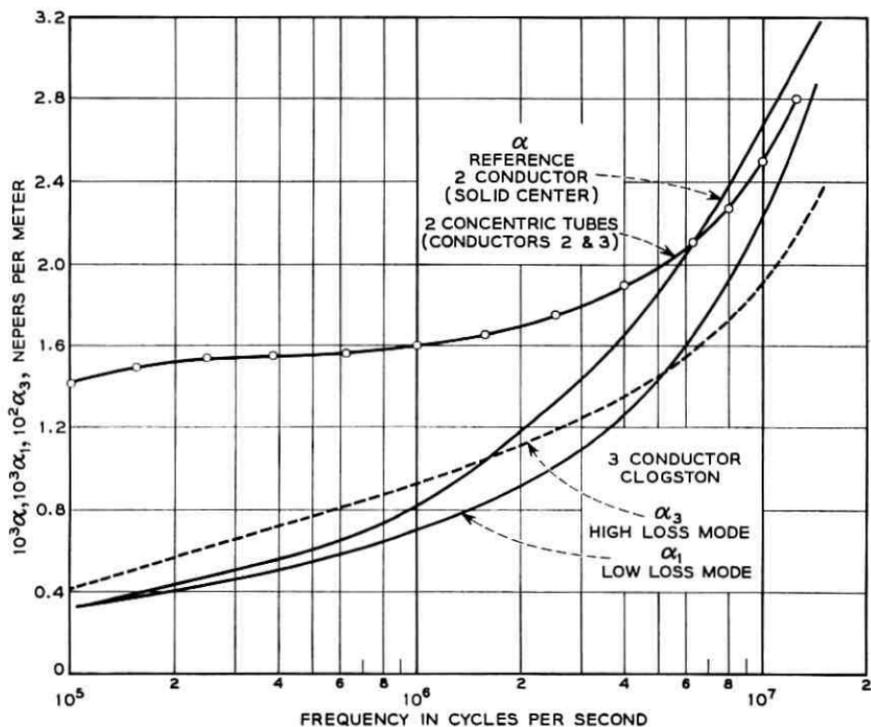
Fig. 6 — Attenuation constants of three-conductor Clogston line compared to those of two two-conductor lines of the same dimensions.
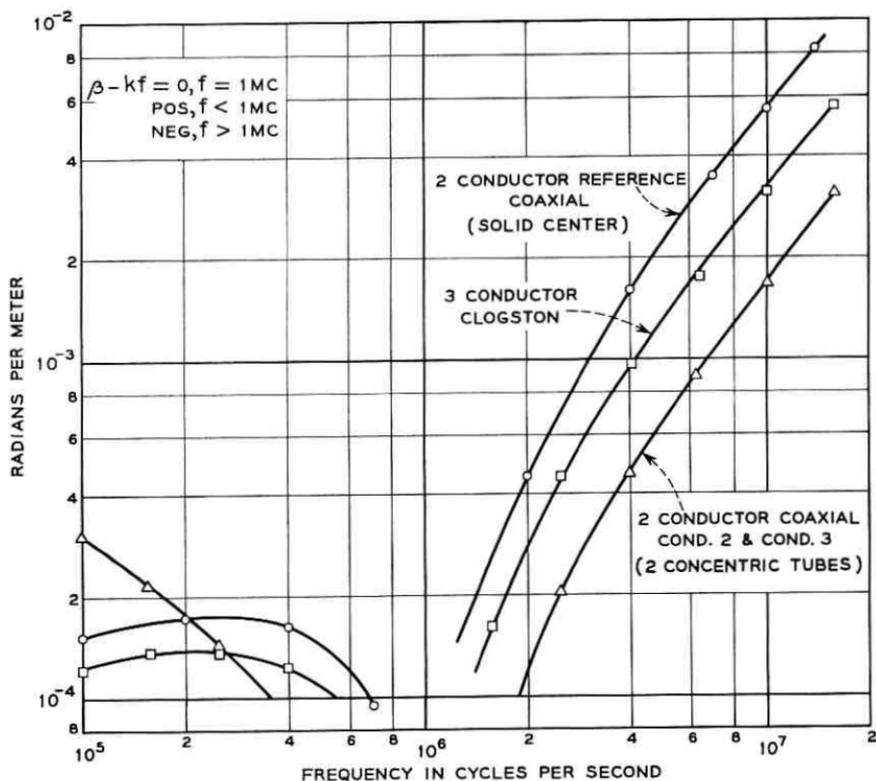
Fig. 7 — Phase distortion of three-conductor Clogston line compared with those of two-conductor coaxials of same dimensions.

The natural ratios of voltages on the two conductors for any one traveling wave are not the same as the ratios of impedances, e.g., $v_{11}/v_{12} \neq Z_1/Z_2$. The voltage and current ratios were calculated by substituting the values of γ_1 and γ_3 into (1) and (2). They are plotted in Figs. 9 and 11. From these figures it is seen that the current associated with the high-loss mode (propagation constant γ_3) in conductor 2 is nearly equal to that in conductor 1 but flows in the *opposite* direction, so that very little flows in the outer conductor 3. Thus, the current that flows in the traveling waves of this mode does not divide more or less evenly between the two inner conductors as it flows down the line, as required for reduction of eddy current losses, but uses these two conductors mainly as go and return paths. In other words, looking at this Clogston line as a two-conductor structure, current in this mode is essentially a circulating current in one of the conductors. It is also seen that below about 2 mc

the power at this mode has a fairly large reactive component. The high-loss mode is thus undesirable for other reasons than its order-of-magnitude higher loss.

The relations and figures of this section will be referred to again in Section IV when the subjects of termination and measurement are taken up.

III. MAKING THE THREE-CONDUCTOR CABLE

After considering a number of ideas and making several trials, it was decided to form the inner dielectric by spiraling a polyethylene thread around the inner solid wire as shown in Fig. 2. This method, suggested by Gordon Raisbeck, seemed to be the most suitable for our experiment. If the thread has a circular cross section, calculations showed that the desired effective dielectric constant is achieved if the threads nearly touch.
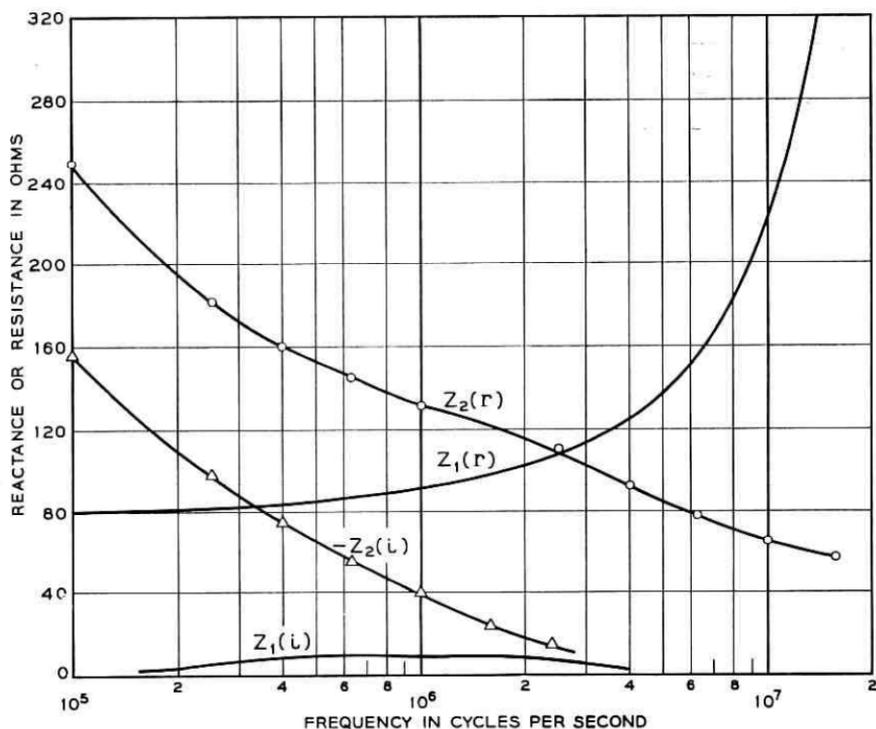

Fig. 8 — Components of characteristic impedances of conductors 1 and 2 (with conductor 3 as return) of three-conductor Clogston line for low-loss mode.
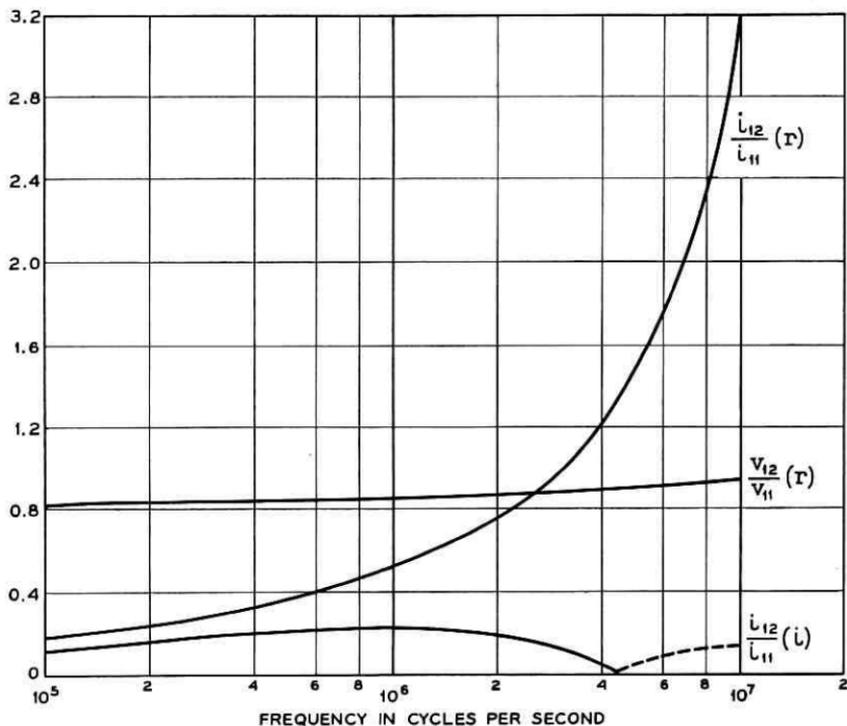
Fig. 9 — Current and voltage ratios in three-conductor Clogston line for low-loss mode.

One problem here was the effect on this insulation of the heat necessarily present during the extrusion of the main dielectric over the composite consisting of conductors 1 and 2 and the dielectric between them. This problem was solved by a modification of the normal extrusion process worked out by members of the chemical research laboratory who did this work for us on their experimental extruder.

The two-conductor composite central member of the cable was made in two trips through a small-scale cable making machine. In the first, 5-mil polyethylene thread was wound around a 19-gauge copper wire as it moved through the machine. In the second trip, 1-mil copper tape and the insulated wire were pulled through a forming die at the center of the spinner so that the copper was wrapped around the insulated wire, leaving a longitudinal gap about 4 mils wide. The copper was held in place by a spiral wrapping of 2-mil thick polyethylene tape. The correct spacing of thread was determined by measuring capacitance and calcu-

lating dielectric constant for several pieces a few feet long. A linear puller was used instead of a circular capstan to avoid bending this composite conductor. The puller consisted of two rubber-toothed belts and a suitable mounting and speed control.

The copper tape used for conductor 2 was inspected and cleaned between Teflon blocks. Even so, a few tiny bits of copper came off and lodged between turns of thread. This was detected by an alarm system rigged to give an indication as soon as a low resistance developed between conductors 1 and 2 during the forming process, so that the machine could be stopped and the trouble sought and removed.

An enlarged view (about $8\times$) of the composite two-conductor center is shown in Fig. 2. The finished form of the structure is at the right, where the polyethylene thread may be seen through the gap between the edges of the copper tape. To the left of this, where the unwrapped polyethylene tape no longer binds it, the gap in the copper tape is
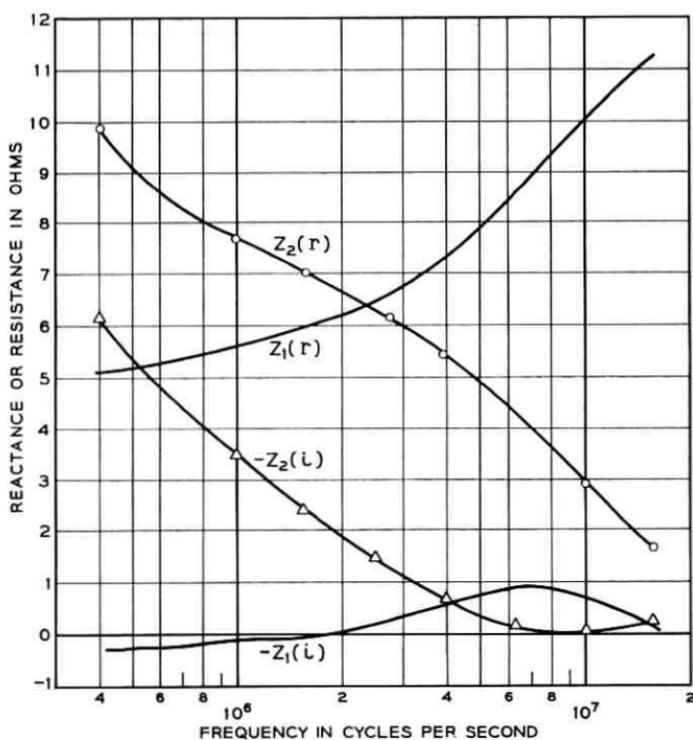

Fig. 10 — Components of characteristic impedances of conductors 1 and 2 (with conductor 3 as return) of three-conductor Clogston line for high-loss mode.
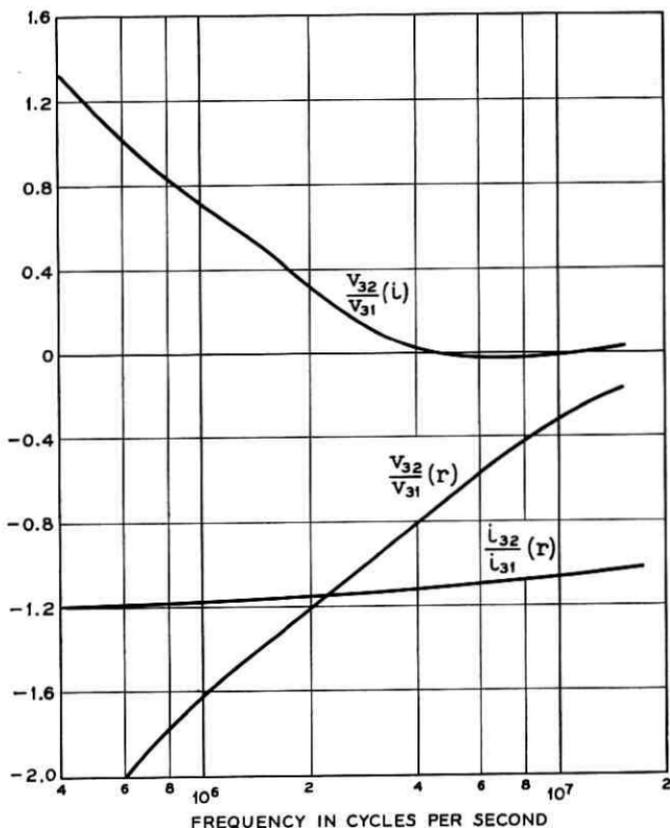
Fig. 11 — Current and voltage ratios in three-conductor Clogston line for high-loss mode.

larger. Small pieces of pressure-sensitive fabric tape have been used to keep the polyethylene thread and tape from unraveling.

Calculations of the dielectric constant of the space between conductors 1 and 2 of this composite were made from capacitance and dimensional measurements; they indicated a value of 0.76 for ϵ_1/ϵ_2 instead of the intended 0.81. The principal reason for the difference appears to be insufficient control of thread spacing and copper tape binding tightness throughout the whole length of the conductors. Weighed against the difficulties of, and the time required for, making a new central element, the difference between attenuation for 0.76 and that for 0.81 seemed small enough to warrant going ahead with the experiment.

After some preliminary runs, the full length of the central member (about 600 feet) was put through the chemistry laboratory's extruder to

make the insulation between conductors 2 and 3 of the cable. This was not entirely successful, since examination showed that the outside diameter of the polyethylene, which was supposed to be 170 mils, actually varied from 150 to 190 mils. After due consideration it was concluded that this was an unavoidable consequence of the modified extrusion process. In order to avoid the uncertain effects of this diameter variation, a machine was made which shaved off excessive polyethylene as the line was pulled along so that the outside diameter now is 151 mils within less than 1 mil. Examination of the inner part of the line at this point showed that no polyethylene had been forced into the space between conductors 1 and 2 during the extrusion process.

The outer conductor of the cable was made by forming 4-mil copper tape over the extruded polyethylene. This conductor is held closed with a small overlap along the longitudinal seam by a spiral wrapping of pressure-sensitive fabric tape.

The length of the cable was carefully measured to be 539.8 feet, or 164.3 meters. The dimensions of the composite structure — conductors 1 and 2 — are difficult to obtain accurately. The best estimates of these dimensions of the line as built, using the terminology of Fig. 3, are (in inches)

$$2a_2 = 0.0466 \qquad 2a_3 = 0.1515$$
$$2b_1 = 0.036$$
$$2b_2 = 0.049 \qquad 2b_3 = 0.1595.$$

Using these dimensions and the measured values of capacitance of the two dielectrics, we find that

$$\epsilon_{r2} = 2.2, \qquad \epsilon_{r1} = 1.67, \qquad \epsilon_1/\epsilon_2 = 0.76.$$

IV. TERMINATION

Because the answers to the question "How do we terminate the Clogston line?" are complicated, let us try to see first just what is meant by proper termination.

If a two-conductor line with its two natural traveling waves, or modes, is terminated with the characteristic impedance for the forward mode (i.e. $Z_1 = v_{11}/i_{11}$), only the forward traveling wave exists on the line and the attenuation constant of the forward natural wave is the attenuation constant of the transmission line. If the termination is other than Z_1, some of the backward natural wave is also present, a situation which may be described otherwise as involving reflection or nonuniform flow of energy.

In the three-conductor line, there are four possible natural traveling

waves, and the total currents and voltages on the line are made up of various amounts of these, as indicated by (9). Consideration of these equations along with (10) shows that if the inner conductors 1 and 2 are terminated with the characteristic impedances Z_1 and Z_2 respectively of the low-loss forward traveling waves and the voltages on these two conductors established in the ratio natural for these waves, then the total currents and voltages in the line will consist of these waves alone. The attenuation of the line will then be that of the low-loss modes, as shown in Figs. 4 and 6. This is the situation we desire, but which was found difficult to achieve experimentally.

In the case of the transposed line, the voltages between the two inner conductors and the outer are the same for the low-loss mode and opposite in sign for the high-loss mode. Thus, the proper way to launch and terminate the line for propagation at the low-loss mode while eliminating the high-loss mode is also the simplest procedure: i.e., to connect together the two inner conductors at both ends of the line.

The situation is not as simple for the elementary Clogston line. The results of calculations plotted on Figs. 8, 9, 10 and 11 show that the voltages of the two inner conductors with respect to the outer are not the same for the low-loss traveling wave. Further, they show that the ratio of these voltages is not the same as the ratio of the characteristic impedances of the two inner conductors when operating in the low-loss mode.

Terminating the Clogston line by connecting together conductors 1 and 2 at the ends and using an average characteristic impedance introduces an unknown proportion of current at the high-loss mode; but the procedure is simple, so it was tried. The attenuation at 1 mc was about 50 per cent higher than expected and about equal to the expected value around 10 mc. Further, there were bumps in the attenuation curve at the low-frequency end, indicating the presence of reflections or some other combination of natural modes. This method is taken up again in the next section, where the results of a search for terminating methods for practical use of the line are given.

The present purpose, however, is to see if the line performs as predicted by theory, and since the most interesting aspect of the theory is the low-loss mode, we tried to establish transmission via this mode alone.

Finally, the method used to drive and terminate the Clogston line shown in Fig. 12 was arrived at. It will be noticed that at the receiving end of the line the two inner conductors are terminated separately. The two-terminal ends of the networks could not be paralleled at both sending and receiving ends of the line without the use of some sort of hybrid
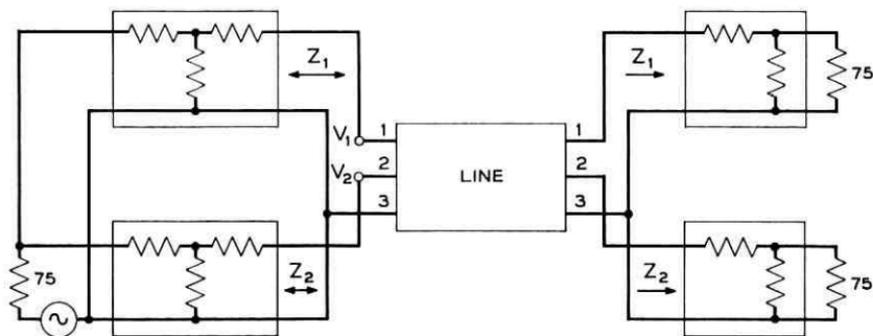
Fig. 12 — Circuit arrangement for terminating and measuring attenuation of three-conductor Clogston line.

or bridge structure; otherwise a transmission loop is formed which makes it very difficult if not out of the question to satisfy all the terminating requirements.

This arrangement, while measuring the attenuation through conductors 1 and 2 separately (which of course should be the same), was adequate for the purpose of testing line performance.

The T and L networks of Fig. 12 were designed to match the impedances Z_1 and Z_2 of the line and to supply the line input so that the ratio of voltages on conductors 1 and 2 was v_{11}/v_{12}. No attempt was made to obtain a design which would provide a match over the frequency band of interest following the data of Figs. 8 and 9. Instead, the networks were made resistive and a separate set was made for each of about eight frequencies. Also, values for Z_1, Z_2 and v_{11}/v_{12} modified from those shown in Figs. 8 and 9 were used. The latter values were for the expected outer diameter (170 mils) of the cable. When the uneven outer polyethylene had to be shaved down to 151 mils in order to obtain a uniform OD, a complete new calculation was not made; only the change in a few parameters was found, and this used as a basis for estimating the others. The impedances were verified by using a pulse signal and observing for reflections as terminations were varied.

Using launching and terminating networks designed in this way and connected to the line as shown in Fig. 12, the attenuation in both inner conductors was measured at several frequencies surrounding the design frequency. The quantity actually measured was the insertion loss of the line, using a precise transmission measuring set. Some of the data are plotted in Fig. 13, where it may be seen that the two curves of each set intersect near the network design frequency. They cross and diverge in
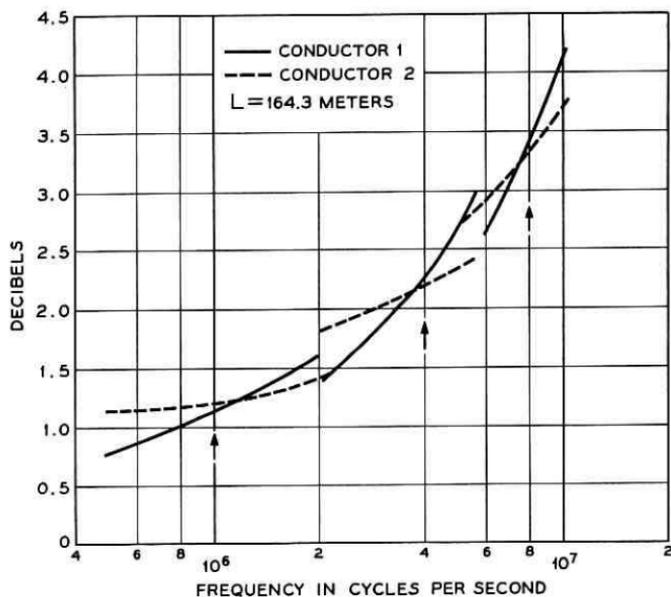
Fig. 13 — Attenuation measurements of three-conductor Clogston line with arrangement of Fig. 12.

both directions beyond the design frequency because the networks provide launching properties only at this frequency. However, they do give the proper value of attenuation in the line at this point. The attenuation at each of the intersections is plotted against frequency in Fig. 14 along with calculated attenuations for the elementary Clogston line and for the reference two-conductor line. The calculated values were obtained by adjusting those shown in Fig. 6 as described previously.

The curves of Fig. 14 show general agreement between measured and predicted attenuation of the Clogston line. Near 4 mc the reduction of attenuation is about 18 per cent instead of the predicted 24 per cent. Part of the reason for this is that, as pointed out in Section III, the ratio of dielectric constants is 0.76 and not 0.815, as required for a minimum. This reduces the predicted change to 21 per cent.

It seems reasonable to conclude that an elementary Clogston line consisting of three coaxial conductors can be built to have a ratio of dielectric constants reasonably near the minimum point and to have an attenuation close to that predicted.

V. FURTHER STUDY OF THE TERMINATING PROBLEM

The method of termination just described would, of course, be completely unsuitable for practical operation of the line. For this reason, a

theoretical study of the problem of coupling the three-conductor line to two-terminal source and load in a practical operation of the line was made by Gordon Raisbeck. This unpublished work led him to a reconsideration of transmission properties of the line when the two inner conductors are connected together at the ends of the line.

Briefly, his conclusions are that for a line long enough so that its total loss is several nepers, the attenuation is very close to that calculated for the low-loss mode of the properly terminated line. Also, the characteristic impedance of the line approaches independence of length as length increases and varies only a small amount with frequency. These facts show that for a reasonable length of line the predicted lower attenuation may be achieved with a simple resistance termination.

Using his formulas, the insertion gain of the 164-meter line when connected between two 50-ohm resistances was computed. This is in fairly good agreement with the corresponding measured loss shown in Fig. 15 when the difference in outside diameters is taken into account. Evidently this is too small a length for the method to be suitable.

Computation from the same formulas of the insertion loss of a 1600-meter length of line (total loss about 2 nepers at 4 mc) was made also and is plotted in nepers per meter in Fig. 16 along with attenuation of
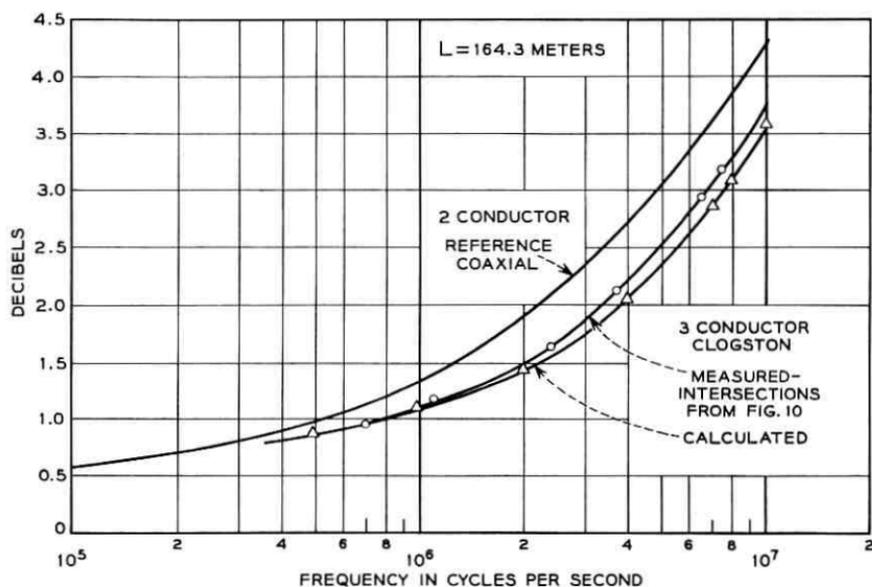

Fig. 14 — Attenuation of three-conductor Clogston transmission line compared with a two-conductor line of like dimensions.

the low-loss mode alone from Fig. 6. This shows that a line of this length with simple termination does have attenuation very close to that of the properly terminated line operating in the low-loss mode. The effects of imperfect launching and impedance matching for the desired mode and the presence of some current in the high-loss mode are thus seen to be small and to occur mainly at the low frequencies. While there is no 1600-meter length of real line available for measurement, the calculated results appear to be reasonable.

A brief resumé of Raisbeck's analysis follows. The method used by him in this work was to start with a length l of the six-terminal, three-conductor line whose transmission properties were calculated and described in Section II and then to short together conductors 1 and 2 at both ends. The matrix describing this new four-terminal structure was then worked out and its transmission properties calculated in terms of two possible traveling waves and characteristic impedances or, in mathematical language, the eigenvalues and eigenvectors of the matrix. The derivation of the matrix of the new line and the calculation of its properties are fairly involved and will not be repeated here.
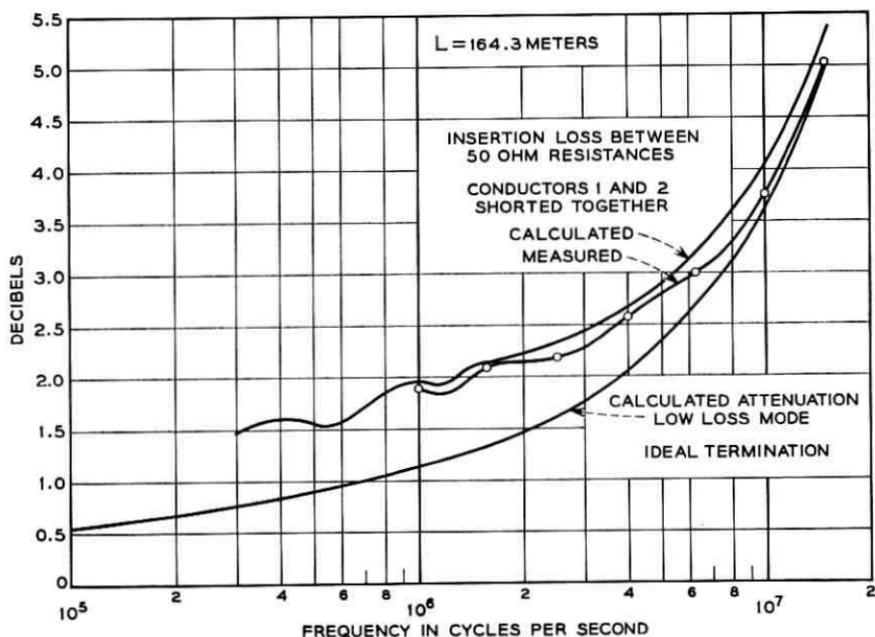

Fig. 15 — Attenuation of short experimental three-conductor Clogston line with simplified termination.
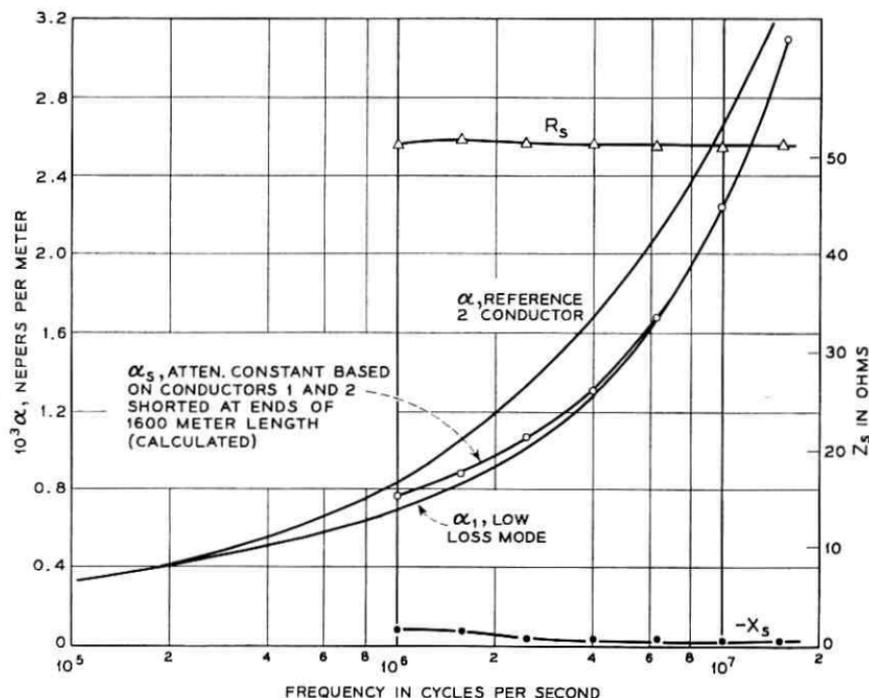
Fig. 16 — Attenuation and characteristic impedance of relatively long three-conductor Clogston line with simplified termination.

The two traveling waves are described by the roots λ_s and λ_s^{-1} of the characteristic equation of the new matrix. These roots are given by

$$\lambda_s = \frac{1}{2}\frac{a}{b} \pm \frac{1}{2}\sqrt{\frac{a^2}{b^2} - 4} \tag{11}$$

and the characteristic impedance is

$$Z_s = \frac{\pm c/a}{\sqrt{1 - 4b^2/a^2}}, \qquad R_e(Z_s) > 0 \tag{12}$$

where the definitions

$$a = (\lambda_3 - \lambda_3^{-1})(\lambda_1 + \lambda_1^{-1})(-1 + h_3)\left(\frac{1}{Z_1} + \frac{h_1}{Z_2}\right)$$

$$- (\lambda_1 - \lambda_1^{-1})(\lambda_3 + \lambda_3^{-1})(-1 + h_1)\left(\frac{1}{z_1} + \frac{h_3}{z_2}\right)$$

$$b = (\lambda_3 - \lambda_3^{-1})(-1 + h_3)\left(\frac{1}{Z_1} + \frac{h_1}{Z_2}\right)$$

$$- (\lambda_1 - \lambda_1^{-1})(-1 + h_1)\left(\frac{1}{z_1} + \frac{h_3}{z_2}\right)$$

$$c = (\lambda_1 - \lambda_1^{-1})(\lambda_3 - \lambda_3^{-1})(h_1 - h_3)$$

relate the parameters of the four-terminal line to those of the six-terminal line given in (10). In these expressions

$$\lambda_3 = e^{\gamma_3 l}, \qquad \lambda_1 = e^{\gamma_1 l}, \qquad \lambda_s = e^{\gamma_s l} \tag{14}$$

all the propagation constants γ having negative real parts.

After studying these relations, Raisbeck observed that for a line long enough so that the attenuation was several nepers, λ_s is nearly equal to λ_1 ; i.e., the attenuation of the line with conductors 1 and 2 shorted at the ends is approximately the same as that of the low-loss mode in the properly terminated three-conductor line.

To see how this comes about, let l be large enough so that

$$\lambda_3 <<< 1, \qquad \lambda_3^{-1} >>> 1$$
$$\lambda_1 << 1, \qquad \lambda_1^{-1} >> 1 \tag{15}$$

and then neglect λ_3 and λ_1 and divide through by λ_3^{-1}. We get

$$\frac{a}{b} = \lambda_1^{-1}\left[1 - \frac{(-1 + h_1)\left(\dfrac{1}{z_1} + \dfrac{h_3}{z_2}\right)}{(-1 + h_3)\left(\dfrac{1}{Z_1} + \dfrac{h_1}{Z_2}\right)}\right] \tag{16}$$

$$= \lambda_1^{-1}(1 - d)$$

and

$$\frac{c}{a} = \left(\frac{Z_1 Z_2}{Z_2 + h_1 Z_1}\right)\left(\frac{h_1 - h_3}{1 - h_3}\bigg/(1 - d)\right). \tag{17}$$

The quantity d does not depend on length of line, but on the voltage ratios and impedances of the natural traveling waves of the line. It is small, being 0.1 at 1 mc, 0.056 at 4 mc, and 0.016 at 10 mc, using the data of Figs. 8, 9, 10 and 11. Thus a/b is large, and we have very nearly

$$\lambda_s^{-1} \doteq \frac{a}{b} \doteq \lambda_1^{-1}(1 - d)$$

$$\lambda_s \doteq \frac{b}{a} \doteq \lambda_1(1 + d) \tag{18}$$

or

$$e^{\gamma_s l} \doteq e^{\gamma_1 l}(1 + d).$$

This shows that the attenuation of the line having conductors 1 and 2 shorted at the ends of a sufficiently long section is nearly the same as that of the ideally terminated line. The characteristic impedance of the line with conductors 1 and 2 shorted is given by

$$Z_s \doteq \frac{c}{a} \doteq \left(\frac{Z_1 Z_2}{Z_2 + h_1 Z_1}\right)\left(\frac{h_1 - h_3}{1 - h_3}\right)(1 + d). \tag{19}$$

This expression for Z_s is nearly equal to the impedance of Z_1 and Z_2 in parallel.

While the values of α_s and Z_s plotted in Fig. 16 were calculated from the exact values of a, b, and c, the calculations show that for the 1600-meter length, the above approximations are quite close to the correct values.

The formula for insertion gain between two resistances R, as used in the calculations, is

$$G = \frac{4RZ_s}{\lambda_s^{-1}(Z_s + R)^2 - \lambda_s(Z_s - R)^2}$$

$$\doteq \frac{4R/Z_s}{(1 + R/Z_s)^2} e^{-\gamma_s l}. \tag{20}$$

VI. SUMMARY

Early in the paper, the properties of an elementary Clogston line were discussed, and the process of building such a line for experiment described. In testing the line's performance, it was found that it performs about as predicted and that terminating the line ideally is a difficult problem. In Section V, it has been seen that Raisbeck's analysis leads to a simple way of coupling the line to source and load with resistances which, while it does not provide ideal termination, does allow performance which is quite close to ideal when the length of line section involved has a loss of around 20 db or more.

Thus we conclude that it is possible to build a uniform three-conductor line which has about 20 per cent less attenuation than a comparable two-conductor line over a considerable frequency band. Since this is almost the same as the reduction obtained with the transposed line, this line is an improvement in not requiring transpositions.

While a line made of two concentric tubular conductors may be the best for pulse transmission because of its low phase distortion, the use

of this three-conductor Clogston line may be thought of as a good compromise. This is because it provides less phase distortion than a conventional two-conductor line, although more than the tubular line (see Fig. 7), and in addition its solid center conductor permits a low-loss path for sending power down the line, and lowers the transmission loss considerably in the low-frequency range (see Fig. 6).

VII. ACKNOWLEDGMENTS

REFERENCES

1. Raisbeck, G., and Manley, J. M., Transmission Characteristics of a Three-Conductor Coaxial Transmission Line with Transpositions, B.S.T.J., **37**, July, 1958, pp. 835–876.
2. Clogston, A. M., Reduction of Skin Effect Losses by the Use of Laminated Conductors, B.S.T.J., **30**, July, 1951, p. 491.
3. Black, H. S., and Morgan, S. P., U. S. Patent 2,924,795, filed December 31, 1954, issued February 9, 1960; also Morgan, S. P., unpublished work.
4. Sugi, M., and Kawazoe, C., Multilayer Stranded Coaxial Submarine Cable, Sumitomo Electric Industries, Ltd.
5. King, Mrs. R. A., An Experimental Clogston 2 Transmission Line, B.S.T.J., **38**, March, 1959, p. 517.
6. Morgan, S. P., Mathematical Theory of Laminated Transmission Lines, B.S.T.J., **31**, September, 1952, p. 883 and Nov., 1952, p. 1121.
7. Raisbeck, G., Nonuniformities in Laminated Transmission Lines, B.S.T.J., **38**, March, 1959, p. 477.
8. Schelkunoff, S. A., The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields, B.S.T.J., **13**, October, 1934, p. 532.

Linear Time-Varying Circuits—Matrix Manipulations, Power Relations, and Some Bounds on Stability

By SIDNEY DARLINGTON

This paper is concerned with general circuits of linear, time-varying, positive, two-terminal components. It describes methods of manipulating corresponding matrix (vector) differential equations. It uses the manipulations to derive equations for power and bounds on stability. The bounds apply to the exponential factors associated with the basis functions of periodically varying circuits.

I. INTRODUCTION

Linear time-varying circuits of special kinds have been designed and analyzed with notable success. On the other hand, theoretical techniques suitable for more general linear time-varying circuits have been developing much more slowly. The development of more general techniques can be approached in various ways. One can seek to specialize the pure mathematics of linear differential equations, in order to discover the properties of those equations which can actually correspond to physical circuits. Alternatively, one can seek to apply the classical analysis of general dynamical systems. As still another alternative, one can seek to generalize, for time-varying circuits, concepts, principles, and techniques which have long been applied to fixed circuits.

This paper illustrates the circuit theory approach. After formulating matrix (vector) differential equations corresponding to circuits of linear, time-varying, two-terminal components, it describes some general methods of manipulation. These apply to combinations of time-varying matrices and the differentiation operator, and are time-varying counterparts of manipulations applied to constant matrices in the theory of fixed circuits. Thereafter, the paper uses the manipulations to derive formulas for power, and some bounds on stability.

The power equations are conventional and reflect the well known fact

that time-varying capacitors and inductors can supply energy to a circuit. However, they can be derived in a way which illustrates manipulation of time-varying matrices in particularly simple terms. Furthermore, the form of the power equations suggests a starting point which leads eventually to the bounds on stability.

It is assumed throughout the paper that all circuit components (fixed or varying) are positive. The stability bounds (as derived) assume that the circuits vary in a periodic manner. Then the basis functions can be arranged as a set of exponentials, each multiplied by a periodically varying coefficient (except in singular cases which are the time-varying counterparts of fixed circuits whose frequency functions have multiple poles). The basis functions are the counterparts of the "natural modes" of fixed networks, and they play an equally important role.

The signs of the real parts of the exponents, in the basis functions, determine whether the functions grow indefinitely, or die out. The stability bounds derived herein are upper and lower bounds on the real parts of the exponents. The specific bounds depend on the composition of the circuit — whether it is composed exclusively of resistors and capacitors, resistors and inductors, or capacitors and inductors, or includes all three kinds of components. For each composition, there are two pairs of bounds, corresponding respectively to the node equation and mesh equation (which will be defined in Section II).

The form of the bounds is illustrated by the following: Let G and C be the node matrices of the conductances and capacitances of a periodically varying circuit of resistors and capacitors. Consider the zeros in terms of the scalar variable λ of the determinant of $(G + \frac{1}{2}\dot{C} + \lambda C)$. If the capacitances are positive at all times, the matrices can be so defined that C is positive definite. Then the zeros of the determinant are real, and they vary periodically with time. It will be shown that the time averages of the instantaneous maximum and minimum over the set of zeros are upper and lower bounds on the real parts of the exponents in the basis functions.

The mesh equations lead to similar bounds in terms of the zeros of the determinant of $(K - \frac{1}{2}\dot{R} + \lambda R)$, in which K and R are the mesh matrices of the stiffnesses and resistances of the capacitors and resistors. It is well known, on energy grounds, that a circuit of varying positive resistors and fixed positive capacitors cannot be unstable. These bounds show that, likewise, a circuit of fixed resistors and varying capacitors cannot be unstable. It is true even though the varying capacitors can give power gain.

Similar bounds can be obtained for circuits of resistors and inductors only, by a simple transformation of current and voltage variables.

The bounds for "RC" and "RL" circuits are at least reminiscent of known bounds on the characteristic roots of a dissymmetrical constant matrix. Let A be a real matrix. The largest and the smallest of the characteristic roots of $\frac{1}{2}(A + A')$ are upper and lower bounds on the real parts of the characteristic roots of A itself.

For a circuit of inductors and capacitors only, let S be the node matrix of the reciprocals of the inductances, and let C be again the node matrix of the capacitances. Consider the zeros of the determinants of the *two* matrices $(\frac{1}{2}\dot{C} + \lambda C)$ and $(-\frac{1}{2}\dot{S} + \lambda S)$. Treat the two sets of zeros as a single set of numbers. Then the time average of the maximum over the set is an upper bound on the real parts of the exponents in the basis functions, and the time average of the minimum over the set is a lower bound. The mesh analysis leads to similar bounds, except that the pertinent matrices are now $(\frac{1}{2}\dot{L} + \lambda L)$ and $(-\frac{1}{2}\dot{K} + \lambda K)$, where L and K are the mesh matrices of inductances and stiffnesses.

In some ways, the bounds for circuits of inductors and capacitors are less satisfactory than those for circuits of resistors and capacitors. When the inductors and capacitors are fixed (and positive) the damping is necessarily zero. When they are varying the damping may or may not be zero. The bounds derived here generally bracket zero damping. Thus they do not say whether or not a time-varying circuit of inductors and capacitors has any basis functions with damping different from zero. They merely say that if the damping is different from zero it is at least within the bounds.

Similar bounds are easily established for circuits of all three kinds of components. However, they tend to be weaker than the bounds for circuits of two kinds of components only, in ways which will be explained.

Some of the derivations have not been completed to the extent of proving validity for all singular, as well as normal, situations. What is reported here is the result of exploratory studies of time-varying circuits, which still leave some details to be filled in.

The author is indebted to I. W. Sandberg, whom he has consulted freely concerning properties of matrices in general and of positive definite and semidefinite matrices in particular. The circuit and power equations are formulated more carefully by C. A. Desoer and A. Paige.[1] However, their analysis tends more to pure mathematics and less to the theory of fixed circuits. They do not include the circuit theory type of manipulations or the stability bounds, which are the primary concerns of this

paper. H. E. Meadows has noted somewhat similar, but weaker bounds.[2] R. A. Rohrer[3] has derived independently the same stability bounds. However, he has done so in terms of classical dynamics (generalization of the equations of Hamilton and Lagrange) rather than in the circuit theory terms used here.

The material in this paper is organized as follows: All aspects of circuits of resistors and capacitors are considered first — formulation of circuit equations, power relations, and stability conditions. Then the same analysis is shown to apply to circuits of resistors and inductors, by changes of variables. Thereafter, circuits including both inductors and capacitors are analyzed in a similar way. This appears to be less confusing than trying to develop the properties of all the kinds of circuits simultaneously.

II. CIRCUITS OF RESISTORS AND CAPACITORS

2.1 *Formulation of Circuit Equations*

Circuit equations can be formulated for linear time-varying components in almost exactly the same way as for fixed components. Basically, there are two parts to the formulation. The first defines the behavior of individual components; the second applies rules for combining the effects of the various components in a circuit.

The components with which we are concerned are of the two-terminal or one-port type. Fig. 1 represents a typical component, with terminals j and k. Voltages E_j and E_k are associated with the terminals. Current I_{kj} enters the component through terminal k and leaves it through terminal j. (Then $I_{jk} = -I_{kj}$.)

The role of the component is to perform an operation which interrelates the voltage difference $E_k - E_j$ and the current I_{kj}. The components with which we are concerned here perform only linear operations. We shall be interested sometimes in the operation which transforms
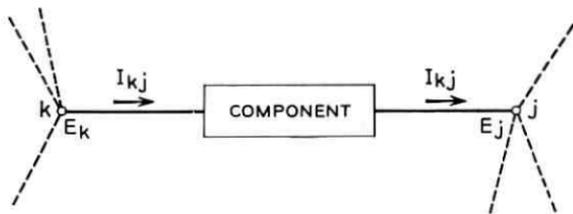

Fig. 1 — A two-terminal component.

$(E_k - E_j)$ into I_{kj} and sometimes in the inverse operation. Thus we may write

$$I_{kj} = \Lambda_{kj}(E_k - E_j)$$
$$E_k - E_j = \Lambda_{kj}^{-1} I_{kj} \tag{1}$$

in which Λ_{kj} is a linear operation and Λ_{kj}^{-1} is its inverse.

Sometimes it is more convenient to consider the charge Q_{kj}, which is of course related to the current by $I_{kj} = \dot{Q}_{kj}$. We shall make extensive use of the symbol p to indicate differentiation:

$$p = \frac{d}{dt}. \tag{2}$$

Then the charge-current relation becomes

$$I_{kj} = pQ_{kj} \tag{3}$$

or, inversely

$$Q_{kj} = \int I_{kj}\, dt. \tag{4}$$

The operations performed by linear resistors and capacitors are displayed in Table I. For each of the two kinds of components, the operations are stated in the two inversely related forms, and in terms of both I_{kj} and Q_{kj}. In a time-varying circuit, the resistance R_{kj} and conductance G_{kj} of the resistor and the capacitance C_{kj} and stiffness K_{kj} of the capacitor may be functions of time.

When the coefficients vary with time they must be written in proper

TABLE I — LINEAR OPERATIONS PERFORMED BY RESISTORS AND CAPACITORS

Notation: $E_k - E_j$ = voltage across component
I_{kj} = current through component
Q_{kj} = charge delivered to component
$I_{kj} = pQ_{kj}$

Resistors: G_{kj} = conductance, R_{kj} = resistance

$$R_{kj} = G_{kj}^{-1}$$
$$I_{kj} = G_{kj}(E_k - E_j), \qquad Q_{kj} = \int G_{kj}(E_k - E_j)dt$$
$$(E_k - E_j) = R_{kj}I_{kj}, \qquad (E_k - E_j) = R_{kj}pQ_{kj}$$

Capacitors: C_{kj} = capacitance, K_{kj} = stiffness

$$C_{kj} = K_{kj}^{-1}$$
$$Q_{kj} = C_{kj}(E_k - E_j), \qquad I_{kj} = pC_{kj}(E_k - E_j)$$
$$E_k - E_j = K_{kj}Q_{kj}, \qquad E_k - E_j = K_{kj}\int I_{kj}dt$$

relationship to the differentiation operation p. For example

$$pC_{kj}(E_k - E_j) = \frac{d}{dt}[C_{kj}(E_k - E_j)] \neq C_{kj}p(E_k - E_j). \qquad (5)$$

Throughout the paper we shall be critically concerned with the "non-commutability of p and time-varying coefficients."

A circuit is formed by interconnecting a number of components, for example as illustrated in Fig. 2(a). The interconnections may be represented by the corresponding linear graph, as in Fig. 2(b). The interactions between the various components are determined by Kirchoff's two laws. The voltage law says that the same voltage E_k can be assigned to each node (graph vertex) k, in forming the voltage differences for all components connected to k. Then the sum of the voltage differences around any mesh (graph cycle) must be zero. The current law states that the sum of the currents into any node must be zero. These remarks are exactly the same whether or not the components vary with time.

We shall consider separately the two different forms of circuit equa-
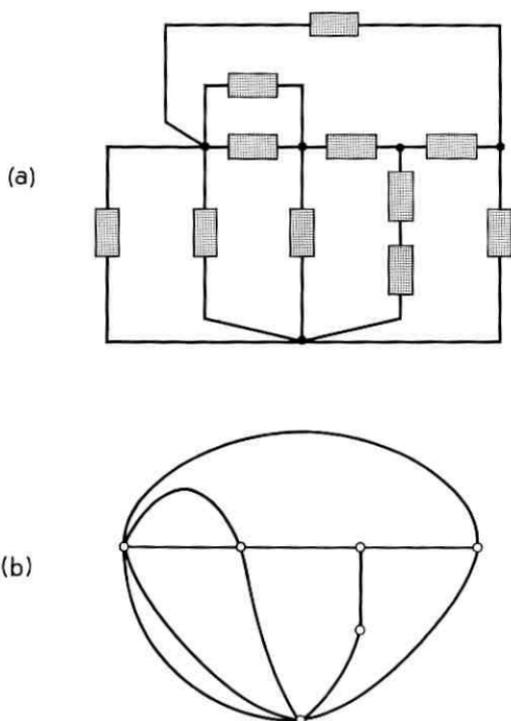
Fig. 2 — A typical circuit: (a) circuit diagram; (b) linear graph.

tions which are commonly used — the node equations and the mesh equations. For the node equations one node is chosen as datum, and the excitation (forcing function) is described as currents fed into other nodes as in Fig. 3(a). Then the node voltages (relative to the datum) are

(a)

NODE 0 = DATUM $(I_{N0} = -\Sigma I_{NK}, E_0 = 0)$

(b)
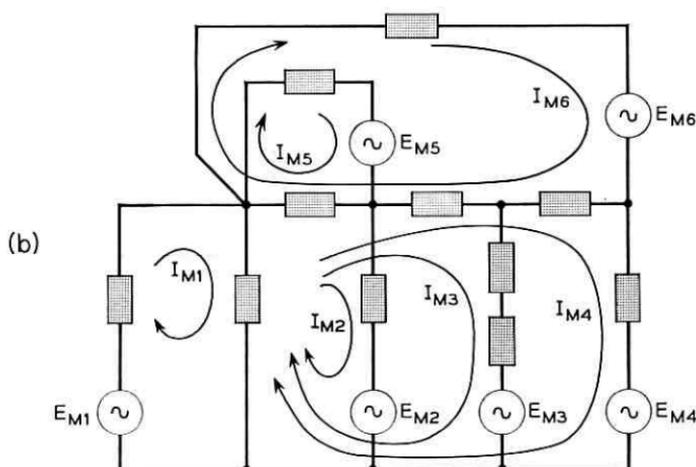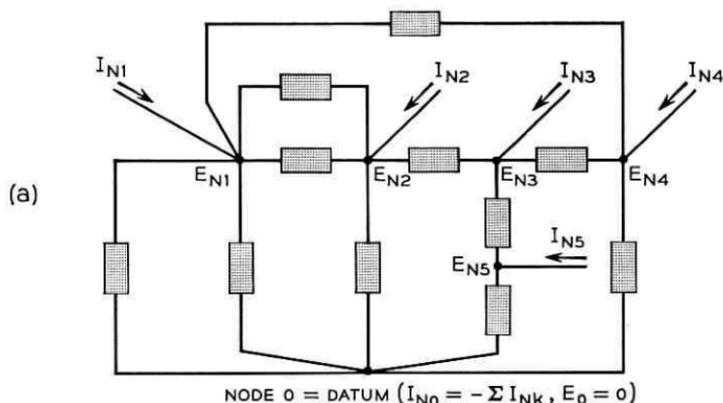
Fig. 3 — Node and mesh currents and voltages: (a) node analysis; (b) mesh analysis.

related to the currents by a vector differential equation. The procedure is exactly the same for time-varying circuits as for fixed circuits, provided one is careful to preserve the correct order of the differentiation operator p and time-varying coefficients.

The vector equation may be written

$$I_N = (G + pC)E_N . \tag{6}$$

Here E_N and I_N are column matrices, or vectors, whose elements are the voltages and excitation currents associated with the various nodes. If only certain of the nodes are externally accessible, for excitation currents, the elements of I corresponding to the other ("internal") nodes are simply constrained to be zero.

G and C are square matrices defining the specific relation between I_N and E_N. Their elements are easily determined from the instantaneous conductances and capacitances of the resistors and capacitors of the circuit, and the usual relations of elementary circuit theory apply even though the conductances and capacitances vary with time.

It follows that G and C, at each instant of time, have the properties of the matrices usually associated with fixed networks. Thus G and C are *symmetrical*, and if the conductances and capacitances of the components are nonnegative, G and C are *positive definite or semidefinite*. The symmetry and the nonnegative character of the matrices, together, lead to an important part of the specialization in circuit theory, relative to the usual pure mathematics of differential equations.

When the differentiation operator p appears in front of a matrix, it signifies differentiation of each element of the matrix. When it is followed by a matrix product, it signifies differentiation of each element in the single matrix equal to the product of matrices. Thus (6) is merely a compact way of writing

$$I_N = GE_N + \frac{d}{dt} Q_N$$

$$Q_N = CE_n = (q_{ij}) \qquad (7)$$

$$\frac{d}{dt} Q_N = \left(\frac{d}{dt} q_{ij} \right).$$

For the mesh equations, meshes are chosen in a somewhat arbitrary way, and the excitation is described as voltage generators inserted in the meshes as in Fig. 3(b). The meshes correspond to cycles in the linear graph, and the number of meshes is the maximum number of independent cycles permitted by the topology of the graph. The mesh currents are related to the excitation voltages by a vector differential equation. As before, the procedure is the same for time-varying as for fixed circuits, provided p and time-varying coefficients are written in the proper order.

For a network of resistors and capacitors, it is more convenient to use charges in place of currents. Then

$$E_M = (K + Rp)Q_M$$
$$I_M = pQ_M.$$

(8)

Here E_M, Q_M, I_M are column matrices or vectors. E_M and I_M are *not* the same as E_N and I_N, although they are related to them in a quite complicated way.

R and K are square matrices defining the specific relation between E_M and Q_M. Their elements are determined from the instantaneous resistances and stiffnesses of the resistors and capacitors of the circuit, and the usual relations of elementary circuit theory again apply. R and K are related to G and C of (6), but in a quite complicated way. The elements of R and K are *not* simply reciprocals of elements of G and C nor are matrices R and K the inverses of G and C.

Like G and C, R and K have properties of matrices usually associated with fixed networks. They are *symmetrical*, and if the resistances and stiffnesses of the components are nonnegative, the matrices are *positive definite or semidefinite*.

A further point should be mentioned. If (8) is to be valid for all circuits, the constants of integration implicit in the Q_{kj}, as defined in (4), must be consistent in the following sense: They must be such that the indefinite integral in (4) can be replaced by a definite integral, say

$$\int_{t_0}^{t} I_{kj}\, dt,$$

with t_0 the same for all kj. This is the same as requiring that all Q_{kj} must be zero at some one time t_0.

When the condition is not met, the superposition theorem can be invoked to express the complete circuit relations in two parts. The first assumes that all capacitors are completely discharged at time t_0, and relates the charges accumulated at $t > t_0$ to the generator (mesh) voltages at $t > t_0$. The second starts with the actual charges at $t = t_0$ and determines their later values in the absence of generator voltages. When the capacitances vary with time, the initial charges may have a much more important effect than is possible in a fixed circuit. This will be discussed in physical terms in Section 2.4.1.

2.2 *An Algebra of p and Matrix Coefficients*

This section introduces manipulations of a sort which we shall use extensively in this and later papers. In the manipulations it is convenient

to use both p and a dot over a symbol to indicate differentiation with respect to time:

$$px = \dot{x} = \frac{d}{dt}x. \tag{9}$$

Generally (but not quite always) the dot will be used for rates of change of coefficients and p for differentiation of primary variables (such as voltages or currents) or products of coefficients and primary variables.

By way of introduction, consider the following scalar expression:

$$pax = \frac{d}{dt}(ax) = \dot{a}x + apx = (\dot{a} + ap)x. \tag{10}$$

Suppose x is a principal variable and a is a time-varying coefficient. Then a may be regarded as a linear operator which multiplies x by a function of time. In the same way, p is an operator which differentiates x, pa is an operator which multiplies x by a and differentiates the product, and ap is an operator which differentiates x and multiplies the derivative by a.

Equation (10) may be said to state a commutation rule, which may may also be stated as an "operator identity" (both sides of which are operators):

$$pa = \dot{a} + ap. \tag{11}$$

Thus p and a commute without change if and only if a is constant, so that $\dot{a} = 0$.

The concept is easily extended to more complicated combinations, for example

$$(a + p)(b + p) = ab + \dot{b} + (a + b)p + p^2. \tag{12}$$

An algebra of this sort, in p and scalar coefficients, is useful for the manipulation of scalar differential equations. It is a principal tool in, for example, Ref. 4.

For present purposes we need to extend the concept to an algebra of p and *matrix* coefficients, as a tool for manipulating vector differential equations. Suppose X is a matrix variable and A is a matrix coefficient. As a first example, it is easily established that

$$pAX = \dot{A}X + ApX = (\dot{A} + Ap)X. \tag{13}$$

The relation follows at once from the fact that each element of AX is a sum of terms, each of which is a product of one element from A and one from X. Equation (11) can be applied term by term and the results can then be sorted into contributions to $\dot{A}X$ and ApX.

Corresponding to (13) is the "operator identity"

$$pA = \dot{A} + Ap. \tag{14}$$

The concept is quickly extended to more complicated operations. Some "operators equations" are collected in Table II, together with some familiar algebraic matrix identities. In the table

$$A^t = \text{transpose of } A$$

$$A^{-1} = \text{inverse of } A \tag{15}$$

$$U = \text{unit or identity matrix.}$$

Constant scalar and matrix coefficients commute with p without change. Time-varying coefficients do not. (But of course matrix factors, constant or not, generally do not commute with each other.) This may be regarded as the most important distinction between the theories of linear time-varying and fixed circuits. If it were not for the difference in the commutation rules, most of the familiar techniques applied to fixed circuits would apply directly to time-varying circuits. As it is in fact, the more complicated commutation rules lead to numerous complications, as we shall see.

2.2.1 Matrices of Order One

Much of this paper is concerned with scalar quantities (for example, net input power) derived from vector circuit equations. For some pur-

TABLE II — SOME MATRIX RELATIONS

A, B, X, Y = matrices, φ = a scalar

In relations which involve inverses, pertinent matrices are assumed to be square and nonsingular.

SOME ALGEBRAIC IDENTITIES

$$X(A + B)Y = XAY + XBY$$
$$\varphi XY = X\varphi Y = XY\varphi$$
$$(XY)^t = Y^t X^t, \quad (XY)^{-1} = Y^{-1}X^{-1}$$

SOME OPERATOR IDENTITIES

$$pA = \dot{A} + Ap, \quad p^2 A = \ddot{A} + 2\dot{A}p + Ap^2$$
$$pAp = \dot{A}p + Ap^2 = -p\dot{A} + p^2 A$$
$$pA\varphi = \dot{\varphi}A + \varphi pA, \quad pAe^\varphi = e^\varphi(\dot{\varphi}A + pA)$$

DERIVATIVES OF SOME MATRIX FUNCTIONS

$$pX^2 = \dot{X}X + X\dot{X} \neq 2\dot{X}X$$
$$pXYZ = \dot{X}YZ + X\dot{Y}Z + XY\dot{Z}$$
$$\dot{X}p(X^{-1}) = -\dot{X}X^{-1}, \quad pX^{-1} = -X^{-1}\dot{X}X^{-1} \neq -(X^{-1})^2\dot{X}$$
$$pX^t = (pX)^t$$

poses, scalars may be represented by matrices of order one, and the scalars in question will be derived in that form. Certain operations will be used repeatedly in this connection.

Suppose W and V are $n \times 1$ column matrices and Y is an $n \times n$ square matrix. Then

$$
\begin{aligned}
YV &= \text{a column matrix} \\
W'YV &= \text{a matrix of order one.}
\end{aligned}
\tag{16}
$$

A matrix of order one is always symmetrical, for there are no off-diagonal terms to interchange in forming the transpose. Thus, using a transpose rule from Table II,

$$
W'YV = (W'YV)^t = V'Y'W. \tag{17}
$$

Certain special cases are particularly important.

If $Y^t = Y$ and W, V are column matrices,

$$
W'YV = V'YW. \tag{18}
$$

The differentiation rules in Table II require

$$
p(V'YV) = \dot{V}'YV + V'Y\dot{V} + V'\dot{Y}V. \tag{19}
$$

Applying (18) gives

if $Y^t = Y$ and V is a column matrix

$$
\dot{V}'YV = V'Y\dot{V} = \tfrac{1}{2}\, p(V'YV) - \tfrac{1}{2}\, V'\dot{Y}V. \tag{20}
$$

Finally, suppose the transpose of Y is the negative of Y. Then (17) requires

if $Y^t = -Y$ and V is a column matrix

$$
V'YV = V'Y'V = -V'YV = 0. \tag{21}
$$

(If a quantity is equal to its negative it must be zero.)

2.3 *Power and Stability in Terms of the Node Equation*

2.3.1 *Instantaneous and Average Powers*

We return now to the circuit equation of the nodal analysis:

$$
I_N = (G + pC)E_N. \tag{22}
$$

The power P supplied to the circuit by the excitation currents is

$$P = \sum_k E_{Nk}I_{Nk} = E_N{}^tI_N. \tag{23}$$

Multiplying the circuit equation by E^t,

$$P = E_N{}^t \, (G + pC)E_N. \tag{24}$$

Expanding in terms of identities in Table II gives

$$P = E_N{}'GE_N + E_N{}'\dot{C}E_N + E_N{}'C\dot{E}_N. \tag{25}$$

Applying (20) to the last term leaves

$$P = E_N{}^t(G + \tfrac{1}{2}\dot{C})E_N + \tfrac{1}{2}p(E_N{}'CE_N). \tag{26}$$

The power relation confirms, in circuit equation terms, what must be expected on physical arguments. Thus $\tfrac{1}{2}p(E_N{}'CE_N)$ is the rate at which energy is being stored electrically in the capacitors. Then $E_N{}'GE_N$ is the rate at which energy is being dissipated in the resistors, and $\tfrac{1}{2}E_N{}^t\dot{C}E_N$ is the rate at which energy is being removed from the circuit by whatever means are used to vary the capacitances. (Recall that *increasing* a capacitance *decreases* the stored energy per unit charge.)

The average power $\bar{P}$ is frequently of interest as well as the instantaneous power P. For the average over a *finite* interval, say t_1 to t_2, integration of (26) gives

$$\operatorname*{Ave}_{t_2 \text{ to } t_1} P = \operatorname*{Ave}_{t_2 \text{ to } t_1} [E_N{}^t(G + \tfrac{1}{2}\dot{C})E_N] + \frac{(E_N{}'CE_N)\,|_{t_1}^{t_2}}{2(t_2 - t_1)}. \tag{27}$$

When $E_N{}'CE_N$ is bounded at all times, the last term approaches zero as $(t_2 - t_1)$ approaches infinity. Thus, for long time averages,

if $E_N{}'CE_N$ is bounded at all times

$$\bar{P} = \operatorname{Ave} [E_N{}'(G + \tfrac{1}{2}\dot{C})E_N]. \tag{28}$$

2.3.2 *Linear Transformations on E_N and I_N*

The power equation suggests rearranging the current equation so as to emphasize the matrix $(G + \tfrac{1}{2}\dot{C})$. Operator identities in Table II yield

$$I_N = [(G + \tfrac{1}{2}\dot{C}) + \tfrac{1}{2}(pC + Cp)]E_N. \tag{29}$$

It is now time to introduce a linear transformation of a sort which will be used extensively in this and later papers. In particular, let

$$E_N = N\hat{E}$$
$$\hat{I} = N'I_N. \tag{30}$$

Linear transformations of this sort are of course standard means for diagonalizing matrices. They are also well known as means for generating equivalent circuits, when components are fixed.[5] Their appropriateness for time-varying circuits is by no means obvious, and they do in fact lead to serious complications not encountered in connection with fixed circuits. As usual, the complications stem from the commutation rules.

Multiplying (29) by N^t and using (30) gives

$$\hat{I} = [N^t(G + \tfrac{1}{2}\dot{C})N + \tfrac{1}{2}(pN^tCN + N^tCNp) + \tfrac{1}{2}(N^tC\dot{N} - \dot{N}^tCN)]\hat{E}. \tag{31}$$

Note that

$$(N^tC\dot{N} - \dot{N}^tCN)^t = -(N^tC\dot{N} - \dot{N}^tCN). \tag{32}$$

When C is positive definite, there is a transformation N such that:

$$\begin{aligned} N^tCN &= U \\ N^t(G + \tfrac{1}{2}\dot{C})N &= -D. \end{aligned} \tag{33}$$

Here U is the unit matrix of suitable order and D is a diagonal matrix. Defining D with negative sign simplifies the later discussion. The existence of a suitable transformation matrix, at each instant, follows from elementary circuit and matrix theory (for positive components). When C is only positive semidefinite, (6) can be transformed into a new equation, in fewer dimensions, with a positive *definite* C. One such procedure is outlined in the Appendix.

With the negative sign in the second equation of (33), the (diagonal) elements d_{kk} of D are identical with the zeros λ_k of the determinant of matrix $(G + \tfrac{1}{2}\dot{C} + \lambda C)$.

$$\begin{aligned} d_{kk} &= \lambda_k \\ \det (G + \tfrac{1}{2}\dot{C} + \lambda_k C) &= 0. \end{aligned} \tag{34}$$

In our applications, because G and C are functions of time, the λ_k are functions of time. If the circuit components are always nonnegative, the λ_k are all real.

When the transformation N is fixed by (33), (31) and (32) become

$$\begin{aligned} \hat{I} &= (-D + pU + J)\hat{E} \\ J &= \tfrac{1}{2}[N^{-1}\dot{N} - (N^{-1}\dot{N})^t] \\ J^t &= -J. \end{aligned} \tag{35}$$

The power equation (26) becomes

$$P = E'I = \hat{E}'N'I = \hat{E}'\hat{I} = -E'D\hat{E} + \tfrac{1}{2}p(\hat{E}'\hat{E}). \qquad (36)$$

(The product $\hat{E}'J\hat{E} = 0$ because $J' = -J$ and $\hat{E}$ is a column matrix.) The power equation could have been transformed directly, but we shall find it informative to have also the transformed current equation (35).

2.3.3 *Some Bounds on the Basis Functions*

Carrying the analysis only a little further establishes some interesting bounds on the basis functions of circuits which vary periodically. This subsection outlines a derivation, but simplifies the argument by means of some somewhat restrictive assumptions. The next subsection reviews the derivation and removes most of the restrictions.

The basis functions are counterparts, for time-varying circuits, of the familiar natural modes of fixed circuits. In node terms, they are a set of linearly independent solutions of (22) for E_N with $I_N = 0$. Thus, if E_σ is a vector basis function,

$$0 = (G + pC)E_\sigma. \qquad (37)$$

When $I_N = 0$, $P = 0$ and (26) becomes

$$0 = E_\sigma'(G + \tfrac{1}{2}\dot{C})E_\sigma + \tfrac{1}{2}p(E_\sigma'CE_\sigma). \qquad (38)$$

Also, when $I_N = 0$, $\hat{I} = 0$ and the transformed equations (35) and (36) become

$$0 = (-D + Up + J)\hat{E}_\sigma$$

$$0 = \hat{E}_\sigma'D\hat{E}_\sigma + \tfrac{1}{2}p(\hat{E}_\sigma'\hat{E}_\sigma) \qquad (39)$$

$$E_\sigma = N\hat{E}_\sigma.$$

If the circuit has n degrees of freedom there are n basis functions in the set. They may be chosen in many different ways, but each choice is a linear transformation on every other choice.

If the circuit varies periodically, G and C in (37) and (38) vary periodically [and also D and J in (39)]. It is well known that the basis functions of a linear differential equation with periodically varying coefficients can usually be so chosen that they are exponentials with periodic coefficients.* Any exceptions are singular cases which we can best take care of in the next subsection.

---

* See the discussion of the Floquet-Poincaré theorem in a text on differential equations, such as pages 78–81 in Coddington and Levinson.[6]

Thus for a periodically varying network, we can use

$$E_\sigma = H_\sigma \exp (s_\sigma t). \tag{40}$$

Here s_σ is a constant, the exponential is a scalar factor, and H_σ is a periodically varying vector. For the purposes of this subsection, we can best restrict ourselves to circuits for which s_σ and H_σ are real. The restriction will be removed in the next subsection.

We shall find it convenient to replace H_σ by another periodic vector F_σ, related to H_σ by

$$H_\sigma = F_\sigma e^\theta \tag{41}$$

in which θ is an arbitrary periodic real function of t (with the same period as the time-varying circuit components). In terms of F_σ,

$$E_\sigma = F_\sigma \exp (s_\sigma t + \theta). \tag{42}$$

Because the exponential is simply a scalar factor, (42) and the last equation of (39) require

$$\begin{aligned}
\hat{E}_\sigma &= \hat{F}_\sigma \exp (s_\sigma t + \theta) \\
F_\sigma &= N\hat{F}_\sigma .
\end{aligned} \tag{43}$$

For any matrix A,

$$pAe^{(s_\sigma t+\theta)} = [\exp (s_\sigma t + \theta)][(s_\sigma + \dot{\theta})A + pA]. \tag{44}$$

Using this relation in (39) and then cancelling out the exponential factor gives

$$\begin{aligned}
0 &= [(s_\sigma + \dot{\theta})U - D + J]\hat{F}_\sigma + p\hat{F}_\sigma \\
0 &= \hat{F}_\sigma'[(s_\sigma + \dot{\theta})U - D]\hat{F}_\sigma + \tfrac{1}{2}p(\hat{F}_\sigma'\hat{F}_\sigma).
\end{aligned} \tag{45}$$

Now $\hat{E}_\sigma$ does not remain bounded over an infinite time interval. Hence averaging the second equation of (39) does not eliminate the second term. On the other hand, $\hat{F}_\sigma$ is bounded at all times, and thus the second equation of (45) implies:

$$0 = \text{Ave } \hat{F}_\sigma'[(s_\sigma + \dot{\theta})U - D]\hat{F}_\sigma . \tag{46}$$

Since $\hat{F}_\sigma$, D and $\dot{\theta}$ are all periodic, while s_σ is constant, the long time average is the same as the average over any one period.

The matrix $(s_\sigma + \dot{\theta})U - D$ is diagonal. That is, all its elements are zero except on the main diagonal, where the typical element is

$$a_{kk} = s_\sigma + \dot{\theta} - \lambda_k \tag{47}$$

$$\text{Det } (G + \tfrac{1}{2}\dot{C} + \lambda_k C) = 0.$$

Then multiplying out the matrix product in (46) gives

$$0 = \text{Ave} \sum_k (s_\sigma + \dot\theta - \lambda_k)\hat{F}_{\sigma_k}^{\ 2} \qquad (48)$$

in which $\hat{F}_{\sigma_k}$ is the kth element in the column matrix $\hat{F}_\sigma$.

When all the quantities are real (as assumed), condition (48) cannot be true if all the coefficients $(s_\sigma + \dot\theta - \lambda_k)$ are positive at all times, or if all are negative at all times. We can use this circumstance to set bounds on the exponent s_σ of the basis function. The arbitrary function θ has been introduced as a means of strengthening these bounds, as explained below.

Let us assume temporarily that the λ_k are at all times distinct. Fig. 4(a) illustrates a plot of the λ_k over one period of their periodic variations. Let λ_M be the largest λ_k, and let θ_M be a choice of θ such that

$$\lambda_M - \dot\theta_M = \text{constant.} \qquad (49)$$

Fig. 4(b) illustrates a plot of the corresponding $\lambda_k - \dot\theta_M$. Every coefficient $(s_\sigma + \dot\theta - \lambda_k)$ in (48) will now be positive unless s_σ is no greater than $\lambda_M - \dot\theta_M$.

Now $\dot\theta_M$ is actually determined uniquely by (49). Because θ_M is periodic, the average of $\dot\theta_M$ over any period is zero, and hence also the long-time average. Then, when $\lambda_M - \dot\theta_M$ is constant

$$\text{Ave} (\lambda_M - \dot\theta_M) = \lambda_M - \dot\theta_M = \text{Ave } \lambda_M$$
$$\theta_M = \lambda_M - \text{Ave } \lambda_M. \qquad (50)$$

Thus all terms in (48) will have positive coefficients $(s_\sigma + \dot\theta_M - \lambda_k)$ unless

$$s_\sigma \leqq \text{Ave } \lambda_M. \qquad (51)$$

Exactly the same sort of procedure leads also to a lower bound on s_σ. Let λ_m be the smallest λ_k, and θ_m a choice of θ which makes $\lambda_m - \dot\theta_m$ a constant. Fig. 4(c) illustrates the corresponding $\lambda_k - \dot\theta_m$. The bound on s_σ can be written

$$s_\sigma \geqq \text{Ave } \lambda_m. \qquad (52)$$

2.3.4 Recapitulation, Discussion, and Removal of Restrictions

This subsection reviews the derivation of the stability bounds, and removes most of the restrictions.

Our diagonalization of the matrices $G + \frac{1}{2}\dot{C}$ and C assumed that C is positive definite. The Appendix describes how an equation with a positive semidefinite C can be transformed into one of lower dimensionality
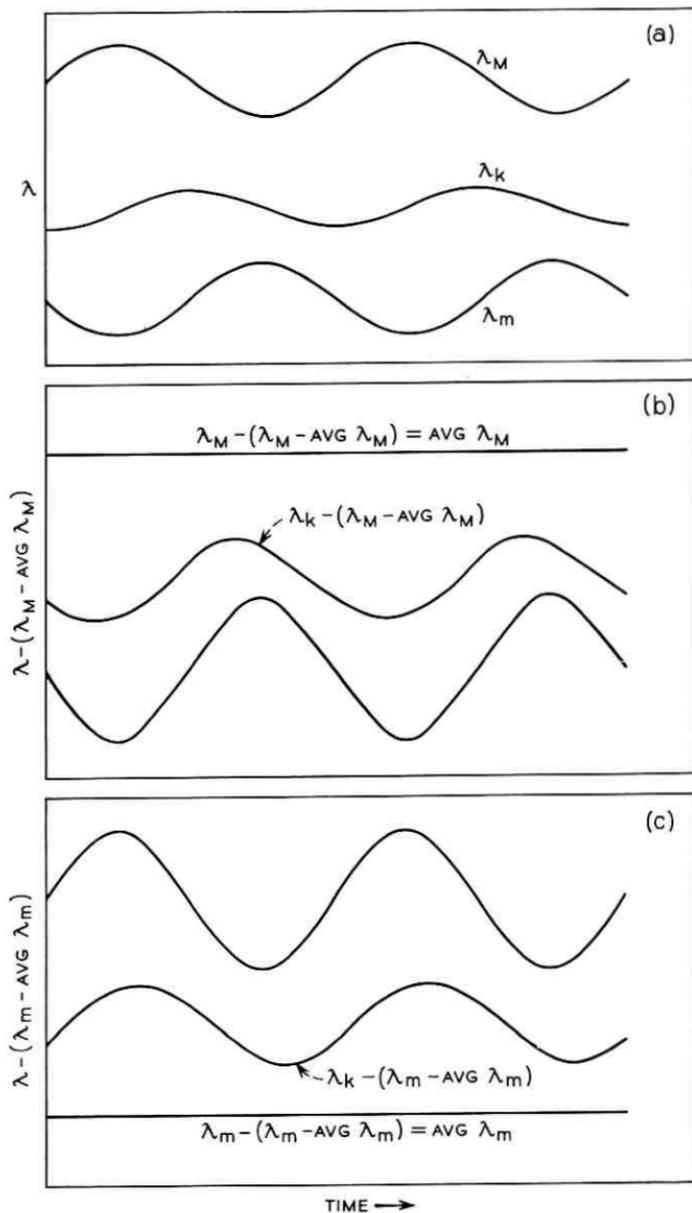
Fig. 4 — Characteristic roots: (a) λ; (b) $\lambda - \dot{\theta}$, with $\dot{\theta}$ to make maximum = constant; (c) $\lambda - \dot{\theta}$, with $\dot{\theta}$ to make minimum = constant.

with a positive definite C. More exactly, the transformation is such that the components of the transformed current and voltage vectors may be divided into two autonomous parts. If n and m are the order and rank of the original matrix C, m components of the transformed current and voltage vectors are related by a differential equation of our standard form, in m dimensions, with a positive definite C matrix, and m linearly independent basis function subject to our stability bounds. The remaining $n - m$ components of the current and voltage vectors are related by a purely algebraic equation, which raises no questions of stability.

In our derivation of the bounds on s_k, we restricted ourselves to periodically varying circuits such that all the basis functions (corresponding to a particular choice) are exponentials with periodic coefficients. In singular instances such a choice is not possible. However, for periodically varying circuits, the basis functions can all be exponentials with coefficients which are at most polynomials in t with periodic coefficients. Furthermore, the coefficients are polynomials in t only when more than one basis function has the same coefficient s_k in the exponential factor exp $(s_k t)$. Out of the set of basis functions corresponding to a single s_k, one can always be assigned a periodic coefficient H_k. This is sufficient to establish our bound on s_k, without regard for the possible existence of other basis function with the same s_k.

We also assumed that the constant s_k and the periodic function H_k are real. The assumption is not actually necessary. With real circuit components, complex basis functions can be chosen in conjugate pairs. Then equally valid choices are the real and imaginary parts of the complex functions. Thus we can write, for a complex exponential basis function,

$$(X_\sigma + iY_\sigma)e^{(s_\sigma + i\omega_\sigma)t} = Z_{1\sigma} \exp (s_\sigma t) + iZ_{2\sigma} \exp (s_\sigma t)$$

$$Z_{1\sigma} = X_\sigma \cos \omega_\sigma t - Y_\sigma \sin \omega_\sigma t \tag{53}$$

$$Z_{2\sigma} = X_\sigma \sin \omega_\sigma t + Y_\sigma \cos \omega_\sigma t.$$

We can now use either $Z_{1\sigma} \exp (s_\sigma t)$ or $Z_{2\sigma} \exp (s_\sigma t)$ as a basis function, in place of $H_\sigma \exp (s_\sigma t)$. The only difference is that the Z_σ's are not periodic, as is H_σ. However, the Z_σ's *are* bounded at all times, and so are F_σ and $\hat{F}_\sigma$, which are now defined by

$$Z_{1\sigma} \text{ or } Z_{2\sigma} = F_\sigma e^\theta$$

$$F_\sigma = N\hat{F}_\sigma. \tag{54}$$

Furthermore, the *boundedness* of $\hat{F}_\sigma$ is all that is required, for passing from (45) to (46), provided the average in (46) is a long-time average. Thus the real part s_σ, of a complex coefficient $s_\sigma + i\omega_\sigma$, also obeys our bounds.

An alternate proof uses the complex basis function itself, replaces $E_\sigma{}^t$ by its conjugate in (38) and (39) and sorts out real and imaginary parts of the equations.

The exponents corresponding to circuits of constant resistors and capacitors are necessarily real. However, when both the resistors and capacitors vary with time, the exponents may be complex. A number of specific examples are known, including examples cited by Desoer and Paige[1] and by Meadows.[2]

Finally, we assumed that the zeros λ_k of the determinant of $(G + \frac{1}{2}\dot{C} + \lambda C)$ were distinct at all times. This is not necessary, provided we make a simple change in the statement of our bounds on s_σ. Suppose the various λ_k's crisscross, as in Fig. 5 (which may be contrasted with Fig. 4a). The variables λ_M and λ_m are now defined as the instantaneous maxima and minima over the set of variables $\lambda_1, \cdots, \lambda_n$. Otherwise the bounds are the same as before.

Thus, for general circuits of periodically varying positive resistors and capacitors, if s_σ is the real part of the coefficient in the exponential factor associated with a basis function,
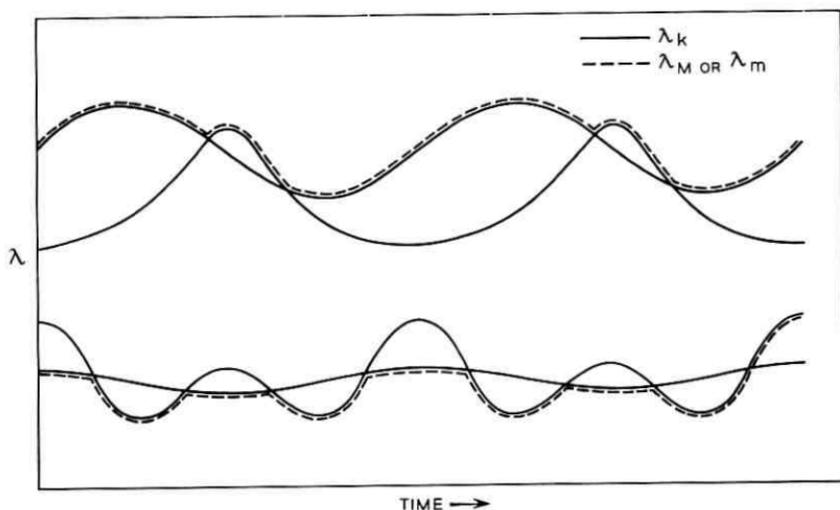

Fig. 5 — λ_M and λ_m when λ_k's crisscross.

$$\text{Ave } \lambda_m \leqq s_\sigma \leqq \text{Ave } \lambda_M \qquad (55)$$

$$\lambda_m, \lambda_M = \text{instantaneous min. and max.}$$

$$\text{over the set } \lambda_1, \cdots, \lambda_n.$$

$$\text{Det } [G + \tfrac{1}{2}\dot{C} + \lambda_k C] = 0.$$

When the circuit components are constants, the λ_k are constant, and also each s_σ is exactly equal to one of the λ_k. Why do the averages of the λ_k only furnish bounds when the components vary with time?

The answer lies in the joint implications of the vector and scalar equations

$$
\begin{aligned}
0 &= [(s_\sigma + \dot{\theta})U - D + J]\hat{F}_\sigma + p\hat{F}_\sigma \\
0 &= \text{Ave } \{\hat{F}_\sigma{}'[(s_\sigma + \dot{\theta})U - D]\hat{F}_\sigma\} \\
J &= N^{-1}\dot{N} - (N^{-1}\dot{N})^t \\
J^t &= -J
\end{aligned}
\qquad (56)
$$

[equations (45) and (46) of Section 2.3.3]. It will be sufficient to consider only the second-order case, for which (56) represents the following collection of equations

$$
\begin{aligned}
0 &= (s_\sigma + \dot{\theta} - \lambda_1)\hat{F}_{\sigma 1} + \dot{\hat{F}}_{\sigma 1} + J_{12}\hat{F}_{\sigma 2} \\
0 &= (s_\sigma + \dot{\theta} - \lambda_2)\hat{F}_{\sigma 2} + \dot{\hat{F}}_{\sigma 2} - J_{12}\hat{F}_{\sigma 1} \\
0 &= \text{Ave } [(s_\sigma + \dot{\theta} - \lambda_1)\hat{F}_{\sigma 1}{}^2 + (s_\sigma + \dot{\theta} - \lambda_2)\hat{F}_{\sigma 2}{}^2].
\end{aligned}
\qquad (57)
$$

($J_{11} = J_{22} = 0$ and $J_{21} = -J_{12}$ because $J^t = -J$.)

With constant coefficients, $J_{12} = 0$, $\hat{F}_\sigma$ is constant, and one suitable solution is

$$
\begin{aligned}
(s_\sigma + \dot{\theta} - \lambda_1) &= 0, & \text{to satisfy the 1st eq.} \\
\hat{F}_{\sigma 2} &= 0, & \text{to satisfy the 2nd eq.} \quad (58) \\
(s_\sigma + \dot{\theta} - \lambda_1) \text{ and } \hat{F}_{\sigma 2} &= 0, & \text{to satisfy the 3rd eq.}
\end{aligned}
$$

With time-varying components, making $(s_\sigma + \dot{\theta} - \lambda_1)$ zero in (57) leaves

$$
\begin{aligned}
0 &= \dot{\hat{F}}_{\sigma 1} + J_{12}\hat{F}_{\sigma 2} \\
0 &= (\lambda_1 - \lambda_2)\hat{F}_{\sigma 2} + \dot{\hat{F}}_{\sigma 2} - J_{12}F_{\sigma 1} \\
0 &= \text{Ave } [(\lambda_1 - \lambda_2)\hat{F}_{\sigma 2}{}^2].
\end{aligned}
\qquad (59)
$$

These equations are (at least usually) incompatible when $J_{12} \neq 0$. Thus the nondiagonal matrix J, which appears only when components are time-varying, is what destroys the identity of the exponents s_σ and the determinant zeros λ_k (as here defined). Recall that J stems from the commutation rules applied to combinations of E^t, p, C, and E.

When the resistors are time-varying but the capacitors are fixed, $\dot{C} = 0$ and $(G + \frac{1}{2}\dot{C} + \lambda C)$ becomes $(G + \lambda C)$, which has been studied extensively in connection with fixed networks. When C is positive definite and G is positive definite or semidefinite, none of the zeros λ_k can be positive, and the same is true of their averages when they are time-varying. Then our bounds exclude any positive s_σ in the exponential factors exp $(s_\sigma t)$ associated with our basis functions. But an unstable basis function (defined as one which grows indefinitely) requires a positive s_σ. Thus our bounds confirm the stability of circuits in which only the resistors vary.

2.3.5 *Comparison with a Known Property of Constant Matrices*

The bounds on s_σ are at least reminiscent of known bounds on the characteristic roots of a dissymmetrical constant matrix. Consider the roots s_σ of

$$\text{Det } (A + s_\sigma U) = 0 \tag{60}$$

in which A is a dissymmetrical matrix. For the closest parallel to our analysis assume that the characteristic roots are all real. The equation may be rewritten as follows:

$$\text{Det } (S + J + s_\sigma U) = 0$$
$$S = \tfrac{1}{2}(A + A^t), \qquad S^t = S \tag{61}$$
$$J = \tfrac{1}{2}(A - A^t), \qquad J^t = -J.$$

When the determinant is zero, there are nonzero vectors X such that:

$$(S + J + s_\sigma U)X = 0. \tag{62}$$

Because $J^t = -J$, this implies the scalar equation

$$X^t(S + s_\sigma U)X = 0. \tag{63}$$

Diagonalizing S leaves

$$\sum_k (s_\sigma - \lambda_k)\hat{X}_k^2 = 0$$
$$\text{Det } (S + \lambda_k U) = 0. \tag{64}$$

The J term in (62) excludes (in general) an $\hat{X}$ in which all elements

are zero except one. Then $s_\sigma \neq \lambda_k$, but the largest and smallest λ_k are upper and lower bounds on s_σ. Simple changes in the analysis establish the same bounds for the real parts of complex characteristic roots.

2.4 Stability in Terms of the Mesh Equation

Section 2.3 dealt exclusively with the node equation. However, the mesh equation can be manipulated in almost exactly the same way. Recall the mesh equation [(8) of Section 2.1],

$$E_M = (K + Rp)Q_M . \tag{65}$$

Note the order of operations Rp here, and contrast it with pC in the node equation.

Define a scalar M by

$$M = \sum_k Q_{Mk}E_{Mk} = Q_M{}^t E_M . \tag{66}$$

While M does not have the dimensions of power, it has many of the mathematical properties of the power function P which we associated with the node equations. Multiplying E_M by $Q_M{}^t$ and applying operator identities gives

$$M = Q_M{}^t(K - \tfrac{1}{2}\dot{R})Q_M + \tfrac{1}{2}p(Q_M{}^t R Q_M). \tag{67}$$

Compare this with the power equation (26) in Section 2.3. The appearance of $-\tfrac{1}{2}\dot{R}$ in M, as opposed to $+\tfrac{1}{2}\dot{C}$ in P, reflects the difference of the order of operations in Rp and pC. The quantity $\tfrac{1}{2}(Q_M{}^t R Q_M)$ does not represent stored energy, but it is mathematically similar to the stored energy function $\tfrac{1}{2}E_N{}^t C E_n$. Both these quantities may be regarded as Lyaponov functions.

Proceeding from here exactly as in the analysis of the node equations leads eventually to similar, but not identical, bounds. The basis functions are now mesh charges Q_σ instead of node voltages E_σ. When circuits vary periodically they can be chosen as exponentials with periodic coefficients:

$$Q_\sigma = H_\sigma \exp (s_\sigma t). \tag{68}$$

The bounds on the s_σ (or the real parts when complex) may be written

$$\text{Ave } \lambda_m \leqq s_\sigma \leqq \text{Ave } \lambda_M$$

$$\lambda_m , \lambda_M = \text{instantaneous min. and max.}$$

$$\text{over the set } \lambda_1 , \cdots , \lambda_n \tag{69}$$

$$\text{Det } [K - \tfrac{1}{2}\dot{R} + \lambda_k R] = 0.$$

Now the coefficients s_σ must be identical in the exponentials associated with the node and mesh analyses. This is because the ratios of charges and voltages remain bounded. Thus the two sets of bounds apply to the same set of s_σ. Then one can form a single pair of bounds by choosing the lesser of the two upper bounds and the greater of the two lower bounds.

When the capacitors are time varying but the resistors are fixed, $\dot{R} = 0$ and $(K - \frac{1}{2}\dot{R} + \lambda_k R)$ becomes the familiar $(K + \lambda_k R)$ of the theory of fixed circuits. Then the bounds require that such a circuit cannot be unstable (with positive circuit components) even though the time-varying capacitors give power gain.

2.4.1 A Complication in Degenerate Special Cases

In the discussion of the mesh formulation we have ignored a degenerate special situation, which complicates a more nearly general analysis. The complications arise when there is a node within the network to which two or more capacitors are connected, but no resistors.

Consider first the simplest example, as illustrated in Fig. 6(a). Here, two capacitors are connected in series between nodes j and k, and no other components are connected to their common node c. If the capacitances are constant, one can simply replace the two capacitors by a single equivalent. When they are time varying, the substitution may have to be more complicated.

The reason is briefly as follows: Suppose there is a positive charge on

(a)

(b)
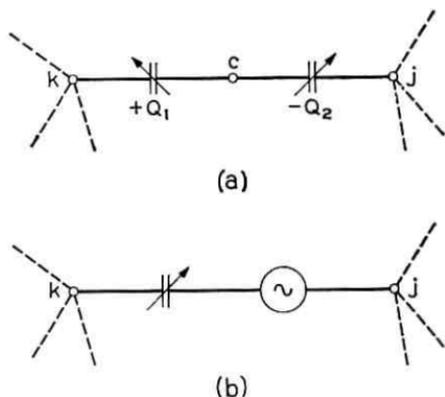
Fig. 6 — Illustration of a degenerate special case: (a) time-varying capacitors with a common node to which no resistors are connected; (b) the Thevenin equivalent.

one capacitor and a negative charge on the other. Then there will be a ratio of charges (at any one time) which will produce a zero voltage difference between terminals j and k. If the ratio of capacitances varies with time, the required ratio of charges varies with time. Conversely, the only combination of constant charges which yields zero voltage difference $(E_k - E_j)$ at all times is zero charge on each capacitor. Furthermore, if there is initially a positive charge on one and a negative charge on the other, both charges cannot be reduced to zero through nodes j and k alone. On the other hand, when the capacitances vary periodically there is a ratio of constant charges such that the voltage difference varies periodically and with zero average.

For the mesh analysis one can invoke Thevenin's theorem, and replace the two capacitors by an equivalent capacitor and a periodic, zero average voltage source, as illustrated in Fig. 6(b). When there are several capacitors and no resistors to a single node, or a combination of resistorless nodes, several Thevenin voltage generators are called for. A general characteristic of circuits which can lead to this sort of complication is a resistance matrix R which is singular (positive semidefinite instead of positive definite).

It follows that the initial charges may be much more important in time-varying than in fixed circuits. For example, suppose there is only one resistorless node and only two capacitors connected thereto. In a fixed circuit, the effects of initial charges will eventually die out except for the voltage at the single resistorless node. In a time-varying circuit, however, the varying Thevenin voltage may produce voltages at all nodes forever.

For the node analysis, one can ignore these complications by simply not eliminating the resistorless nodes. When there are resistorless nodes, the conductance matrix G is positive semidefinite. When G has order n and rank m (and C is positive definite), $n - m$ of the basis functions have zero damping. This can be proved by applying to G, with only minor alterations, the analysis applied to a semidefinite C as outlined in the Appendix. The Thevenin voltages of the mesh analysis set up undamped oscillations within the circuit ($s_\sigma = 0$) and correspond to the undamped modes of the node analysis. The validity of the derivation of our stability bounds, which does not preclude a singular G matrix, implies that the bounds will include zero when G is positive semidefinite.

One way to avoid these complications is to avoid ideal capacitors. If a resistor is connected across each capacitor, to represent the leakage through any actual component, there are no resistorless nodes.

III. CIRCUITS OF RESISTORS AND INDUCTORS

The theory of circuits of resistors and capacitors can be applied to circuits of resistors and inductors by interchanging currents and voltages and also the mesh and node formulations. Since currents and voltages are interchanged, one needs variables which are related to voltages in the same way that charges are related to currents. Corresponding to node voltage E_k, let

$$E_k = \dot{\Phi}_k \tag{70}$$

or, inversely

$$\Phi_k = \int E_k \, dt. \tag{71}$$

If an inductor is connected between nodes j and k, $E_k - E_j$ is the voltage across the inductor, and $\Phi_k - \Phi_j$ is proportional to the flux linkages within it. The fact that the definition of Φ_k in (70) or (71) leaves undetermined a constant of integration reflects the fact that constant flux linkages produce no voltage across an inductor.

Table III states the linear operations performed by inductors, in two inversely related forms, and in terms of both $E_k - E_j$ and $\Phi_k - \Phi_j$. The inductance L_{kj} and its reciprocal S_{kj} can vary with time, provided the order of the differentiation symbol p and the coefficients is preserved.

For resistors and inductors, the *mesh* equation corresponds to the *node* equation (6) of our previous circuits, and is

$$E_M = (R + pL)I_M. \tag{72}$$

The mesh voltage and current vectors E_M and I_M are defined as before, and also the matrix R of the mesh resistances. Then L is the mesh matrix of the inductances of the inductors. If one properly chooses constants of integration stemming from (71), the *node* equation is

TABLE III — LINEAR OPERATIONS PERFORMED BY INDUCTORS

$$I_{kj} = \text{current through component}$$
$$E_k - E_j = \text{voltage across component}$$
$$\Phi_k - \Phi_j = \text{variable related to flux}$$
$$E_k = p\Phi_k$$

$$L_{kj} = \text{inductance}, \quad S_{kj} = \text{reciprocal inductance}$$

$$S_{kj} = L_{kj}^{-1}$$

$$\Phi_k - \Phi_j = L_{kj}I_{kj}, \qquad E_k - E_j = pL_{kj}I_{kj}$$
$$I_{kj} = S_{kj}(\Phi_k - \Phi_j), \qquad I_{kj} = S_{kj}\int(E_k - E_j)dt$$

$$I_N = (S + Gp)\Phi_N$$
$$E_N = p\Phi_N \tag{73}$$

and corresponds to the old *mesh* equation (8). Here vector I_N represents the node currents as before, and G is again the matrix of node conductances. The elements of vector Φ_N are the Φ_k and S is the node matrix of the reciprocals of the inductances.

The formulation of (73) requires an assumption regarding constants of integration, implicit in the definition of the Φ_k, exactly like the assumptions regarding the Q_{kj} in the formulation of (8).

One can now proceed exactly as before to obtain power equations and bounds on the exponential factors associated with the basis functions of periodically varying circuits.

Section 2.4.1 described degeneracies which occur in the *mesh* analysis of circuits of resistors and capacitors which have resistorless *nodes*. The counterparts for circuits of resistors and inductors occur in the *node* analysis of circuits with resistorless *meshes*.

IV. CIRCUITS WHICH CONTAIN BOTH CAPACITORS AND INDUCTORS

4.1 *Circuits of Capacitors and Inductors Only*

For circuits which contain both capacitors and inductors, but no resistors, the node equation contains the capacitor and inductor terms from our previous equations (6) and (73). Thus

$$I_N = S\Phi_N + pCE_N$$
$$E_N = p\Phi_N \tag{74}$$

or, replacing E_N by $p\Phi_N$,

$$I_N = (S + pCp)\Phi_N . \tag{75}$$

As before, S is the node matrix of the reciprocals of the inductances, and C is the node matrix of the capacitances.

To obtain the input power P from the excitation, multiply by $E_N{}^t$, which is the same as $\dot{\Phi}_N{}^t$

$$P = \dot{\Phi}_N{}'(S + pCp)\Phi_N . \tag{76}$$

Manipulating in terms of operator identities, again identifying $\dot{\Phi}_N$ with E_N, and using the symmetry of matrices S and C gives

$$P = -\tfrac{1}{2}\Phi_N'\dot{S}\Phi_N + \tfrac{1}{2}E_N'\dot{C}E_N + \tfrac{1}{2}p(\Phi_N'S\Phi_N) + \tfrac{1}{2}p(E_N'CE_N). \quad (77)$$

The first two terms are the rates at which energy is removed from the system by the means used to vary the inductances and capacitances. The last two terms are the rates of increase of the electrical energy stored in the inductors and capacitors.

The basis functions Φ_σ are solutions of

$$0 = (S + pCp)\Phi_\sigma. \quad (78)$$

The corresponding power equation is

$$0 = -\tfrac{1}{2}\Phi_\sigma'\dot{S}\Phi_\sigma + \tfrac{1}{2}E_\sigma'\dot{C}E_\sigma + \tfrac{1}{2}p(\Phi_\sigma'S\Phi_\sigma) + \tfrac{1}{2}p(E_\sigma'CE_\sigma) \quad (79)$$

$$E_\sigma = p\Phi_\sigma.$$

If the circuit components vary periodically, we can define a bounded function F_σ by

$$\Phi_\sigma = F_\sigma \exp(s_\sigma t + \theta). \quad (80)$$

There will be a corresponding function F_σ' defined by

$$E_\sigma = F_\sigma' \exp(s_\sigma t + \theta). \quad (81)$$

Because $E_\sigma = p\Phi_\sigma$, F_σ' is related to F_σ by

$$F_\sigma' = (s_\sigma + \dot{\theta})F_\sigma + \dot{F}_\sigma. \quad (82)$$

Under reasonable circuit conditions (which exclude, for example, discontinuous changes in inductances), F_σ' is bounded, as well as F_σ. Using (80) and (81) in (79), eliminating the exponentials and averaging gives

$$0 = \text{Ave } F_\sigma'[(s_\sigma + \dot{\theta})S - \tfrac{1}{2}\dot{S}]F_\sigma + \text{Ave } F_\sigma't[(s_\sigma + \dot{\theta})C + \tfrac{1}{2}\dot{C}]F_\sigma'. \quad (83)$$

If both S and C are positive definite, all matrices can be diagonalized. This merely requires a different transformation on F_σ and F_σ'. Thus let

$$F_\sigma = N_\sigma\hat{F}_\sigma, \qquad F_\sigma' = N_\sigma'\hat{F}_\sigma' \quad (84)$$

and choose N_σ and N_σ' in such a way that

$$N_\sigma'SN_\sigma = U, \qquad N_\sigma''CN_\sigma' = U \quad (85)$$

$$N_\sigma'\dot{S}N_\sigma = 2D_\sigma, \qquad N_\sigma''\dot{C}N_\sigma' = -2D_\sigma'.$$

Here U is again the unit matrix and D_σ and D_σ' are diagonal matrices.

After transformation, (83) can be written

$$0 = \text{Ave } \sum_k \{(s_\sigma + \dot{\theta} - \lambda_b)\hat{F}_{\sigma k}^2 + (s_\sigma + \dot{\theta} - \lambda_k)\hat{F}_{\sigma k}'^2\}. \quad (86)$$

Here, $\hat{F}_{\sigma k}$ and $\hat{F}_{\sigma k}'$ are elements of the column matrices $\hat{F}_\sigma$ and $\hat{F}_\sigma'$, and λ_k and λ_k' are the corresponding (like-rowed) diagonal elements of D_σ and D_σ'. As before, all quantities can be made real (provided $\hat{F}_{\sigma k}$, $\hat{F}_{\sigma k}'$ are only required to be bounded, not necessarily periodic). Then s_σ cannot be so large that all terms in the sum are positive, or so small that all terms in the sum are negative.

If one makes no use of the implicit relation between $\hat{F}_\sigma$ and $\hat{F}_\sigma'$, one can consider the two sets of constants λ_k and λ_k' as simply two parts of a single set. Then bounds on s_σ can be obtained exactly as before. Since the λ_k depend only on C and the λ_k' only on S, while C and S represent different circuit components, the two sets of bounds are likely to crisscross as in Fig. 7. Thus, including also the relation of λ_k and λ_k' to S and C [defined by (85)], the bounds may be written

$$\text{Ave (min. over } \lambda) \leqq s_\sigma \leqq \text{Ave (max. over } \lambda)$$

$$\text{set } \lambda = \lambda_k\text{'s and } \lambda_k'\text{'s}$$

$$\text{Det } (-\tfrac{1}{2}\dot{S} + \lambda_k S) = 0$$

$$\text{Det } (\tfrac{1}{2}\dot{C} + \lambda_k'C) = 0. \tag{87}$$

The mesh analysis differs only in the specific quantities involved. The mesh circuit equation is

$$E_M = (K + pLp)Q_M$$

$$I_M = pQ_M. \tag{88}$$

As before, K is the mesh matrix of stiffnesses and L is the mesh matrix of inductances.

The input power from the excitation is

$$P = -\tfrac{1}{2}Q_M'\dot{K}Q_M + \tfrac{1}{2}I_M'\dot{L}I_M + \tfrac{1}{2}p(Q_M'KQ_M) + \tfrac{1}{2}p(I_M'LI_M). \tag{89}$$

The bounds on s_σ may be written

$$\text{Ave (min. over } \lambda) \leqq s_\sigma \leqq \text{Ave (max. over } \lambda)$$

$$\text{set } \lambda = \lambda_k\text{'s and } \lambda_k'\text{'s}$$

$$\text{Det } (-\tfrac{1}{2}\dot{K} + \lambda_k K) = 0$$

$$\text{Det } (\tfrac{1}{2}\dot{L} + \lambda_k L) = 0. \tag{90}$$

4.2 *Circuits of all Three Kinds of Components*

For circuits which contain capacitors, inductors and resistors, the
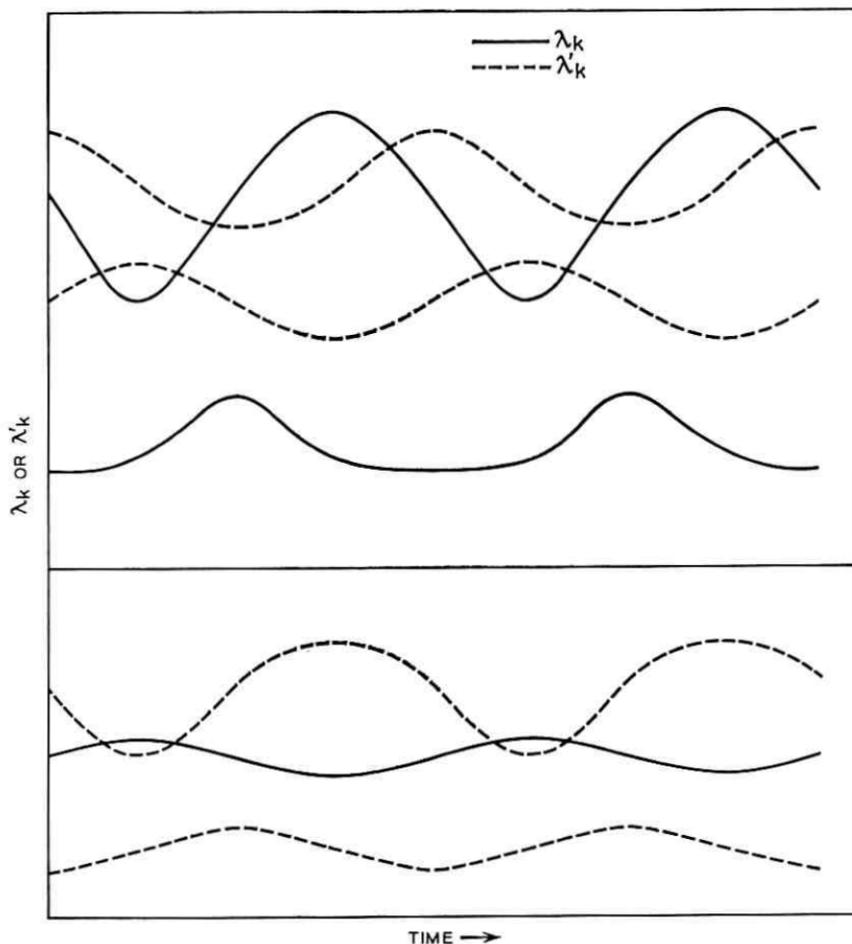
Fig. 7 — Characteristic roots of $(-\frac{1}{2}\dot{S} + \lambda S)$ and $(\frac{1}{2}\dot{C} + \lambda C)$.

node equation may be written

$$
\begin{aligned}
I_N &= (S + Gp + pCp)\Phi_N \\
E_N &= p\Phi_N .
\end{aligned}
\tag{91}
$$

The corresponding power equation may be written

$$
\begin{aligned}
P = -\tfrac{1}{2}\Phi_N{}^t\dot{S}\Phi_N + \tfrac{1}{2}E_N{}^tGE_N + \tfrac{1}{2}E_N{}^t\dot{C}E_N \\
+ \tfrac{1}{2}p(\Phi_N{}^tS\Phi_N) + \tfrac{1}{2}p(E_N{}^tCE_N)
\end{aligned}
\tag{92}
$$

in which the new term, $\frac{1}{2}E_N'GE_N$, represents the power dissipated in the resistors.

The stability bounds may be written

$$\text{Ave (min. over } \lambda) \leqq s_\sigma \leqq \text{Ave (max. over } \lambda)$$

$$\text{set } \lambda = \lambda_k\text{'s and } \lambda_k''\text{'s} \tag{93}$$

$$\text{Det } (-\tfrac{1}{2}\dot{S} + \lambda_k S) = 0$$

$$\text{Det } (G + \tfrac{1}{2}\dot{C} + \lambda_k'C) = 0.$$

Note that λ_k' is the same as λ_k' for a circuit of capacitors and resistors only [see (55)]. On the other hand, the present λ_k do not depend at all on the resistors.

The corresponding mesh equation is

$$E_M = (K + Rp + pLp)Q_M$$

$$I_M = pQ_M. \tag{94}$$

Then the power equation may be written

$$P = -\tfrac{1}{2}Q_M'\dot{K}Q_M + \tfrac{1}{2}I_M'RI_M + \tfrac{1}{2}I_M'\dot{L}I_M$$

$$+ \tfrac{1}{2}p(Q_M'KQ_M) + \tfrac{1}{2}p(I_M'LI_M). \tag{95}$$

Finally, the stability conditions may be written

$$\text{Ave (min. over } \lambda) \leqq s_\sigma \leqq \text{Ave (max. over } \lambda)$$

$$\text{set } \lambda = \lambda_k\text{'s and } \lambda_k''\text{'s} \tag{96}$$

$$\text{Det } (-\tfrac{1}{2}\dot{K} + \lambda_k K) = 0$$

$$\text{Det } (R + \tfrac{1}{2}\dot{L} + \lambda_k'L) = 0.$$

4.3 *Discussion*

Let us return temporarily to circuits of capacitors and inductors only. Suppose all the components are constant (and positive). Then it is well known that the basis functions (natural modes) are undamped. Since $\dot{S}$ and $\dot{C} = 0$, etc., when the components are constant, all λ_k and λ_k' in (88) or (90) are zero. Then the upper and lower bounds must be zero. Thus, in the special case of fixed components, our bounds collapse onto the actual s_σ.

Suppose the inductors are fixed but the capacitors are time-varying. Then, for example, in the bounds (88), $\lambda_1 = 0$ but generally $\lambda_k' \neq 0$ in

The zero coefficients λ_k will affect the maximum or minimum of the combined set λ at any times when all λ_k' have the same sign.

When components vary periodically, C and S, etc., vary periodically and $\dot{C}$ and $\dot{S}$, etc., have zero averages. Then any nonzero elements in $\dot{C}$ and $\dot{S}$, etc., must be positive some of the time and negative some of the time. This suggests (but does not prove) that the bounds on s_σ include zero. It is confirmed by some quite different analysis, which is outlined in a short paper in this issue of the B.S.T.J.[7] The paper points out that the s_σ's of nondissipative circuits occur in pairs of the form $+s_\sigma$, $-s_\sigma$. Thus when any s_σ has a nonzero real part, another s_σ will have the negative of that real part, and any true bounds are consistent therewith. It also follows that the bounds described above can be tightened (for nondissipative networks) by using the bound closer to zero and its negative.

The basis functions of a circuit of capacitors and inductors may all have zero damping even though the components are periodically time-varying. This remark is supported by well known properties of the Mathieu-Hill equation, which corresponds to the one-dimensional special case of our vector equations. A weakness of our bounds is that they do not give meaningful sufficient conditions for basis functions with (positive or negative) damping different from zero. The bounds will (at least almost) always be different from zero, but by bracketing zero they will not exclude it.

Suppose now the circuit includes resistors as well as inductors and capacitors. Consider first the bounds (93) derived from the node equation. If the node conductances (represented by matrix G) are sufficiently large, every characteristic root λ_k' will be more negative than every characteristic root λ_k (which does not involve G). Then the lower (more negative) bound on s_σ will be set by the λ_k' and the upper bound by the λ_k. If the conductances are further increased, the lower bound will become more negative, but the upper bound will remain unchanged. On the other hand, increasing the conductances will usually increase the damping of all the basis functions.

The same remarks apply to the bounds derived from the mesh equation except that mesh resistances replace node conductances. Thus when resistors add substantial damping in circuits containing both capacitors and inductors, our *upper* (more positive) bounds may be too weak to have much significance.

Our node analysis assumes that both S and C are positive definite, although the initial formulation may make one or both positive semi-definite. As before, however, the procedure can easily be modified for positive semidefinite matrices. Recall that two different transformations

are used to diagonalize S and C, in the derivation of our bounds. When both are positive semidefinite, they can be separately transformed to positive definite submatrices bordered by zeros. Because *constant* transformations can be used (in accordance with the Appendix), the transformed $\dot{S}$ and $\dot{C}$ retain the derivative relationship to the transformed S and C. Then the number of characteristic roots λ_k, λ_k' depends on the ranks of S and C. When only capacitors are connected to some nodes, the mesh analysis will involve some zero damped Thevenin voltages; when only inductors appear in some meshes, the node analysis will involve some Thevenin currents.

APPENDIX

Positive Semidefinite Capacitance Matrices

When C is only positive semidefinite, one can transform equation (6) as follows: The rank m of C is now less than the order n. There exist transformation matrices N_1 such that

$$H_1'CN_1 = \begin{vmatrix} 0 & 0 \\ 0 & \hat{C} \end{vmatrix}. \tag{97}$$

The zeros represent submatrices of zero elements, bordering the complete matrix with $n - m$ rows and $n - m$ columns of zeros. Then $\hat{C}$ is an $m \times m$ positive *definite* submatrix.

These remarks apply to time-varying as well as to fixed circuits. Furthermore, the matrix N_1 can be so chosen that it is *constant*, provided the time-varying capacitances are always positive (>0, not $\geqq 0$). In fact, N_1 can be chosen as a matrix of 0's and 1's only. A detailed demonstration is beyond the scope of this paper; briefly it derives from the fact that the rank of C is determined by the topology of the capacitor part of the circuit, without regard to the values of the capacitances (provided they remain >0, so that the topology cannot be a function of time).

Because $\dot{N}_1 = 0$, $N_1'p = pN_1'$, and (6) can be transformed (as for fixed circuits) into

$$\hat{I}_1 = \left[\hat{G}_1 + p \begin{vmatrix} 0 & 0 \\ 0 & \hat{C} \end{vmatrix} \right] \hat{E}_1 \tag{98}$$

$$E = N_1\hat{E}_1, \qquad \hat{I}_1 = N_1'I.$$

A further transformation, N_2, is now easily found, of the form:

$$\begin{vmatrix} U & 0 \\ N_{12}{}^t & U \end{vmatrix} \begin{vmatrix} \hat{G}_1 \end{vmatrix} \begin{vmatrix} U & N_{12} \\ 0 & U \end{vmatrix} = \begin{vmatrix} G_g & 0 \\ 0 & \hat{G} \end{vmatrix}. \tag{99}$$

Here U represents a unit, or identity submatrix of appropriate order, and N_{12} is an $m \times (n - m)$ submatrix. Because of the zero submatrix in the upper right-hand corner of the complete matrix,

$$\begin{vmatrix} U & 0 \\ N_{12}{}^t & U \end{vmatrix} \begin{vmatrix} 0 & 0 \\ 0 & \hat{C} \end{vmatrix} \begin{vmatrix} U & N_{12} \\ 0 & 0 \end{vmatrix} = \begin{vmatrix} 0 & 0 \\ 0 & \hat{C} \end{vmatrix}. \tag{100}$$

Thus the capacitance matrix is unaffected by the second transformation.

The elements of the N_{12} portion of N_2 may be time-varying. However, because the time-varying elements combine with no nonzero elements of $\hat{C}$, $N_2{}^t p\hat{C} = pN_2{}^t\hat{C}$. Thus the second transformation yields

$$\hat{I} = \left[\begin{vmatrix} G_g & 0 \\ 0 & \hat{G} \end{vmatrix} + p \begin{vmatrix} 0 & 0 \\ 0 & \hat{C} \end{vmatrix} \right] \hat{E}_2. \tag{101}$$

We can now divide the components of the transformed current and voltage vectors into two parts, as follows:

$$\begin{aligned} \hat{I}_g &= G_g\hat{E}_g, & \text{in } n - m \text{ dimensions} \\ \hat{I} &= (\hat{G} + p\hat{C})\hat{E}, & \text{in } m \text{ dimensions.} \end{aligned} \tag{102}$$

The first equation is algebraic, and its diagonalization is routine. The second equation is in our standard form (6), with $\hat{C}$ positive *definite*, and it may be transformed further in accordance with (29) to (36).

We arrived at this formulation by starting with a one-to-one correspondence between modes and components of the current and voltage vectors, and then transformed the corresponding differential equation. Desoer and Paige[1] arrive at the same conclusion by *starting* with more general choices of the current and voltage components, taking account of the circuit topology.

REFERENCES

1. Desoer, C. A., and Paige, A., Linear Time-Varying Networks: Stable and Unstable, IEEE International Convention, March, 1963.
2. Meadows, H. E., Analysis of Linear Time-Variable RC Networks, IEEE International Convention, March, 1963.
3. Rohrer, R. A., Stability of Linear, Time-Varying Networks from Energy Function Analysis, Dissertation, Department Of Electrical Engineering, University of California, Berkeley, California.
4. Darlington, S., Nonstationary, Smoothing and Prediction Using Network Theory Concepts, I.R.E. Trans. on Circuit Theory, CT-6, May, 1959, pp. 1–13.
5. Howitt, N., Equivalent Electrical Networks, Proc. I.R.E., 20, June, 1932, pp. 1042–1051.
6. Coddington, E. A., and Levinson, N., *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
7. Darlington, S., A Relation between the Basis Functions of Periodically Varying Nondissipative Circuits, B.S.T.J., this issue p. 2969.

Mode Selection in an Aperture-Limited Concentric Maser Interferometer

By TINGYE LI

The concentric interferometer with a limiting aperture in its mid-plane is analyzed for its mode-selective properties. Two of the lowest-order transverse modes and their losses for the infinite-strip geometry are computed by solving the associated integral equations by the method of successive approximations. The apertured concentric interferometer is found to be more mode-selective than the apertureless concentric interferometer or the Fabry-Perot interferometer with parallel plane mirrors. Computed results indicate that the optimum aperture size for maximum mode selectivity is approximately the size of the major lobe of the diffraction pattern of the dominant mode at the aperture plane. However, the maximum selectivity attainable does not exceed that of the "comparable" confocal system. The latter system is not very practical because it requires either very long resonator lengths or very small mirrors.

I. INTRODUCTION

Interferometer-type resonators that are used in optical masers are inherently multimode devices. The resonant modes that can exist in such devices may be classified as longitudinal and transverse modes. The longitudinal mode order is determined by the number of field variations along the axis of the interferometer, while the transverse mode order is determined by the number of field variations in the plane of the mirrors. For each longitudinal mode order, there exists a set of transverse modes. The number of modes that can partake in the oscillations of an optical maser is dependent on the geometry and the losses of the resonator, the width of the atomic resonance of the active material and the degree of population inversion. Practically, an optical maser will oscillate in several modes simultaneously unless special steps are taken to suppress the unwanted ones. For applications such as optical communication it is desirable from the standpoint of noise, coherence, spectral purity, etc. to suppress all but one mode in a maser oscillator. There-

fore, mode-selection schemes are very important if the optical maser is to be a useful source of coherent and monochromatic radiation.

Aside from losses of a random nature such as those due to inhomogeneities in the medium and mirror imperfections, the resonant modes in a maser interferometer also suffer from losses due to diffraction around the mirrors.[1,2] The diffraction loss varies very little with longitudinal mode order, but increases very rapidly with increasing transverse mode order. Thus interferometer-type resonators are inherently mode-selective with respect to transverse modes. By using a long, thin configuration (small mirrors and large mirror separation) it is possible to suppress all but the dominant transverse mode.[3] Also, by operating just above the oscillation threshold[4] or by using a short resonator[5] it is possible to restrict the oscillations to a single longitudinal mode. The output power in these cases is somewhat limited.

Since it is desirable to pump the active medium strongly so as to obtain as much power as possible from a maser oscillator, methods for providing additional mode selection are important. In general, mode selection involves the introduction of loss to the resonator in some prescribed manner. Kleinman and Kisliuk[6] have proposed the use of an additional Fabry-Perot interferometer to discriminate against unwanted longitudinal modes. Kogelnik and Patel[7] obtained essentially a single-frequency output from a gaseous maser using three concave mirrors. Collins and White[8] used two tilted Fabry-Perot etalons within the resonator of a ruby maser to select longitudinal as well as transverse modes. Schemes for suppressing unwanted higher-order transverse modes by employing a limiting aperture in the focal plane of a suitable optical system within the resonator of a ruby maser were tried by Burch,[9] Baker and Peters,[10] and Skinner and Geusic.[11] These latter schemes are essentially equivalent to an interferometer system consisting of a pair of spherical mirrors spaced concentrically and having a limiting aperture in its mid-plane. Since the field distribution over the mid-plane of a concentric system is essentially the far-field pattern (Fourier transform) of the field over the mirrors, the higher-order transverse modes will have a wider lateral spread than the lower-order ones. Therefore, a suitably chosen aperture will introduce very little loss to the dominant mode but will introduce appreciable loss to the higher-order modes, making the apertured system quite mode-selective to transverse orders.

In order to study the mode-selective property of an apertured concentric system, we have set up the appropriate integral equations for the infinite-strip geometry and solved them iteratively on a digital computer for the intensity distributions and the losses of the two lowest-

order transverse modes. The computed results indicate that there is an optimum aperture for maximum mode selectivity,* but that this maximum never exceeds that of the "comparable" confocal system. In fact, as the aperture is made smaller and smaller the behavior of the apertured concentric system approaches that of the confocal system. However, the comparable confocal system is not very practical because it requires either very small mirrors or very large mirror spacing.
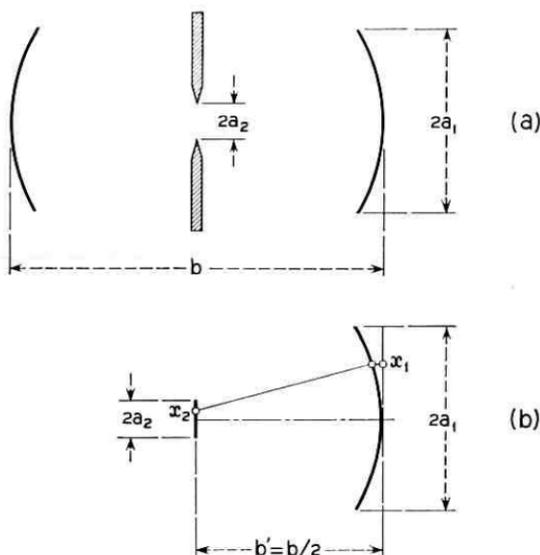
Fig. 1 — Geometry of the aperture-limited concentric maser interferometer. (a) Full-concentric system with aperture. (b) Equivalent half-concentric system.

II. INTEGRAL EQUATIONS OF THE SYSTEM

Since the full-concentric system with a mid-plane aperture is equivalent to a half-concentric system with a plane mirror of the same size as the aperture,[12] it is convenient to use the half-concentric model for the formulation of the integral equations (Fig. 1). Also, it is convenient to use the two-dimensional model of infinite-strip mirrors, since the three-dimensional problem of rectangular mirrors can be reduced to a two-dimensional one.[1] The behavior of systems with circular mirrors is expected to be similar to that of square-mirror systems.

Fig. 1(a) shows the geometry of an apertured concentric interferometer.

* The loss of the higher-order mode relative to the loss of the dominant mode may be regarded as a measure of mode selectivity.

The infinite-strip concentric mirrors are of width $2a_1$ and are separated by b, which is twice the radius of curvature of the mirrors. The aperture is of width $2a_2$ and is in the mid-plane of the interferometer. Fig. 1(b) shows the equivalent half-concentric model which is to be used for the analysis. The integral equations defining the modes can be derived from our previous analysis on interferometers with curved mirrors by setting $g_1 = 0$ and $g_2 = 1$ in equations (11) and (12) of Ref. 12. The resulting equations are

$$\gamma^{(1)}\psi^{(1)}(x_1) = \frac{e^{j\pi/4}}{\sqrt{\lambda b'}} \int_{-a_2}^{a_2} \exp\left[jk(2x_1x_2 - x_2^2)/2b'\right]\psi^{(2)}(x_2)\, dx_2 \quad (1)$$

and

$$\gamma^{(2)}\psi^{(2)}(x_2) = \frac{e^{j\pi/4}}{\sqrt{\lambda b'}} \int_{-a_1}^{a_1} \exp\left[jk(2x_1x_2 - x_2^2)/2b'\right]\psi^{(1)}(x_1)\, dx_1 \quad (2)$$

where the subscripts and superscripts one and two denote the curved mirror and the plane mirror, respectively, as shown in Fig. 1(b). The ψ's are the eigenfunctions that describe the relative field distributions over the mirrors, and the γ's are the corresponding eigenvalues that specify the loss and the phase shift the wave suffers during each transit. The propagation constant k is equal to $2\pi/\lambda$, where λ is the wavelength in the medium. The mirror separation b' is equal to $b/2$.

Equations (1) and (2) are single-transit equations which can be combined to form two round-trip equations. They are

$$\gamma\psi^{(1)}(x_1) = \int_{-a_1}^{a_1} K_1(x_1, \bar{x}_1)\psi^{(1)}(\bar{x}_1)\, d\bar{x}_1 \quad (3)$$

and

$$\gamma\psi^{(2)}(x_2) = \int_{-a_2}^{a_2} K_2(x_2, \bar{x}_2)\psi^{(2)}(\bar{x}_2)\, dx_2 \quad (4)$$

where the kernels K_1 and K_2 are given by

$$K_1(x_1, \bar{x}_1) = \frac{j}{\lambda b'} \int_{-a_2}^{a_2} \exp\left[jk\{x_2(x_1 + \bar{x}_1) - x_2^2\}/b'\right] dx_2 \quad (5)$$

$$K_2(x_2, \bar{x}_2) = \frac{j}{\lambda b'} \int_{-a_1}^{a_1} \exp\left[jk\{x_1(x_2 + \bar{x}_2) - (x_2^2 + \bar{x}_2^2)/2\}/b'\right] dx_1 \quad (6)$$

and the eigenvalue γ is equal to $\gamma^{(1)}\gamma^{(2)}$. Since one round trip in the half-concentric system is equivalent to a single transit in the full-concentric system, (3) may be regarded as the integral equation defining

the modes of the aperture-limited concentric interferometer while (4) gives the field distribution across the aperture.

The kernels K_1 and K_2 as defined by (5) and (6) are symmetric; that is, $K_1(x_1, \bar{x}_1) = K_1(\bar{x}_1, x_i)$ and $K_2(x_2, \bar{x}_2) = K_2(\bar{x}_2, x_2)$. Therefore, the eigenfunctions $\psi_n^{(1)}$ and $\psi_n^{(2)}$ corresponding to distinct eigenvalues γ_n are orthogonal[13] in the sense that

$$\int_{-a_1}^{a_1} \psi_n^{(1)}(x_1)\psi_m^{(1)}(x_1)\, dx_1 = 0, \qquad m \neq n \qquad (7)$$

$$\int_{-a_2}^{a_2} \psi_n^{(2)}(x_2)\psi_m^{(2)}(x_2)\, dx_2 = 0, \qquad m \neq n. \qquad (8)$$

As in the general case of curved mirrors,[12] the eigenfunctions are complex and are orthogonal in the non-Hermitian sense.

The integral equations can be solved numerically using iterative techniques. However, it is possible to extract the asymptotic behavior of the solutions from these equations for very small apertures. Thus for $(a_2/a_1) \ll 1$ and $a_2^2/b\lambda \ll 1$, the terms involving x_2^2 and $\bar{x}_2^2$ can be neglected and the integral equations reduce to those for the asymmetric confocal configuration (with unequal mirrors). It has been shown by Boyd and Kogelnik[14] that the modes and the corresponding losses of an asymmetric confocal system are the same as those of a symmetric system with equal mirrors of width $2a$ where $a = \sqrt{a_1 a_2}$. Therefore, the behavior of the aperture-limited concentric system approaches asymptotically that of the confocal system as the aperture is made smaller and smaller. We designate the confocal system with Fresnel number $N_c = a_1 a_2/b'\lambda$ as the *comparable* confocal system.

III. COMPUTED RESULTS AND DISCUSSION

The two lowest-order modes and their eigenvalues for an aperture-limited infinite-strip concentric interferometer were computed using the method of successive approximations[1] on an IBM 7090 computer. One hundred increments were used in the numerical integration of (1) and (2). Curves for power loss per transit for the two lowest-order modes and for different values of Fresnel number ($N = a_1^2/b\lambda$) are given in Fig. 2. The abscissa is the half width of the aperture normalized to a_1/N, where both a_1 and N should be regarded as constants. The dashed curves give the loss of the comparable confocal interferometer as functions of its half Fresnel number, which is equal to $a_1 a_2/b\lambda$. The losses for the limiting case of infinitely wide aperture are given on the column on the right side of the figure.
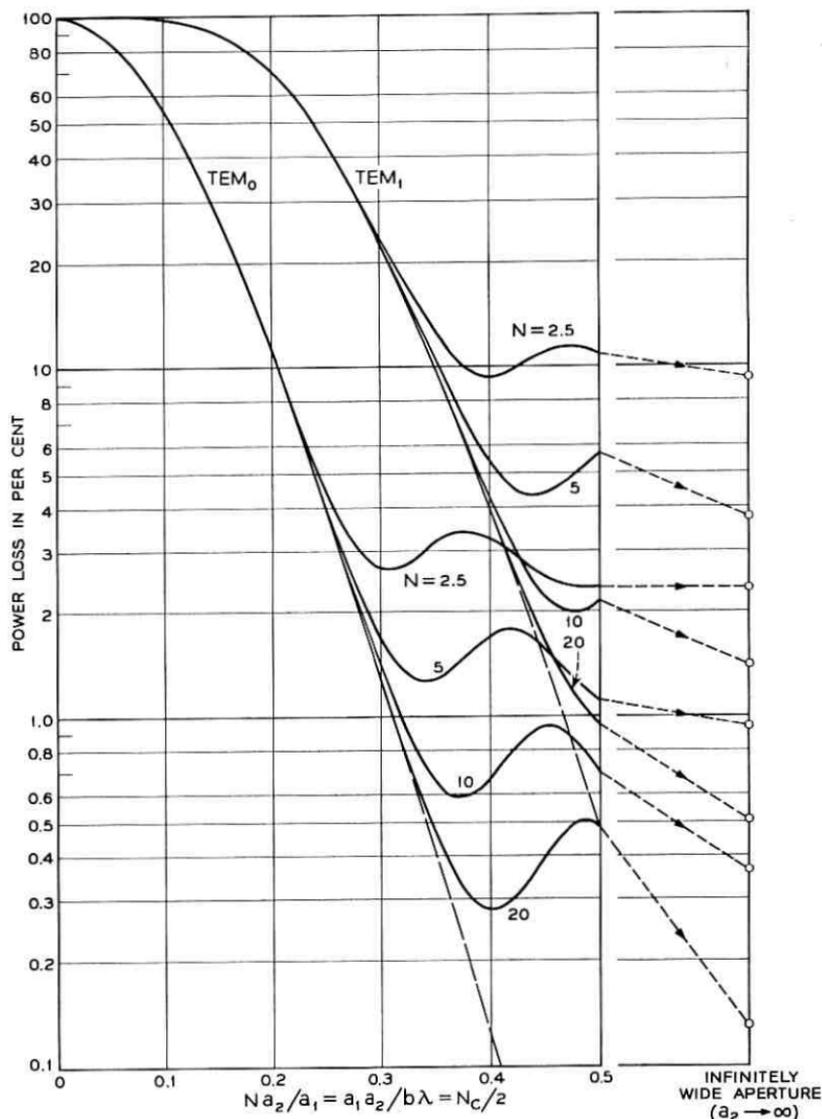
Fig. 2 — Power loss per transit for the two lowest-order modes of an aperture-limited concentric interferometer. The abscissa is the half width of the aperture normalized to a_1/N. (Both a_1 and N should be regarded as constants.) The dashed curves are for the "comparable" confocal interferometer.

Figs. 3 and 4 show the relative amplitude and phase distributions of the field intensity for the two lowest-order modes. Curves A and B are for the aperture-limited concentric interferometer with $N = 10$ and $a_2/a_1 = 0.0325$; A is the field distribution over the mirror while B is the field distribution over the aperture. Curve C is for the comparable confocal interferometer with a Fresnel number (N_c) of 0.65. Curve D is for an apertureless $(a_2 \to \infty)$ concentric interferometer with $N = 10$ and is the same as for the parallel plane configuration except for the reversal of sign in the phase distribution.[1,12]

The ratio of the loss of TEM_1 mode to the loss of TEM_0 mode, which is a measure of mode selectivity, is plotted in Fig. 5 as a function of the normalized aperture half width. The dotted curve represents the same ratio for the comparable confocal system plotted as a function of its half Fresnel number, which is equal to $a_1 a_2/b\lambda$. The short slant lines represent segments of the loci of constant loss for the TEM_0 mode. The ratio of the losses for the limiting case of infinitely wide aperture is approximately four* for large N $(N > 1)$ and is given on the column on the right side of the figure.

That a suitably chosen limiting aperture placed in the mid-plane of a concentric interferometer should be mode-selective can be surmised by considering the field distribution over its mid-plane. This field distribution, as given by (2), is, except for a quadratic phase factor, the Fourier transform of the mode pattern over the mirrors. It resembles very closely the far-field pattern of the equivalent parallel plane system[1,12] and therefore has a lateral spread which increases with mode order. Consequently, a limiting aperture having the width of the major lobe of the dominant mode would have a small effect on that mode but would increase quite significantly the losses of the higher-order modes. The optimum aperture width (for a given N) corresponding to maximum mode selectivity as shown in Fig. 5 is approximately equal to the major-lobe width. As the aperture is made larger, more and more of the minor lobes are uncovered and they interfere either constructively or destructively at the mirrors, producing the oscillations in the loss curves in Fig. 2 and in the relative loss curves in Fig. 5.

Fig. 5 shows that a suitably chosen aperture can increase the relative loss of the TEM_1 mode several times its apertureless value. The effect of the aperture on other higher-order modes is expected to be even greater. By observing the number of iterations required to produce a

---

* For large Fresnel numbers $(N > 1)$ the ratio (loss of TEM_n mode)/(loss of TEM_0 mode) is approximately $(1 + n)^2$ for parallel plane or concentric configurations.
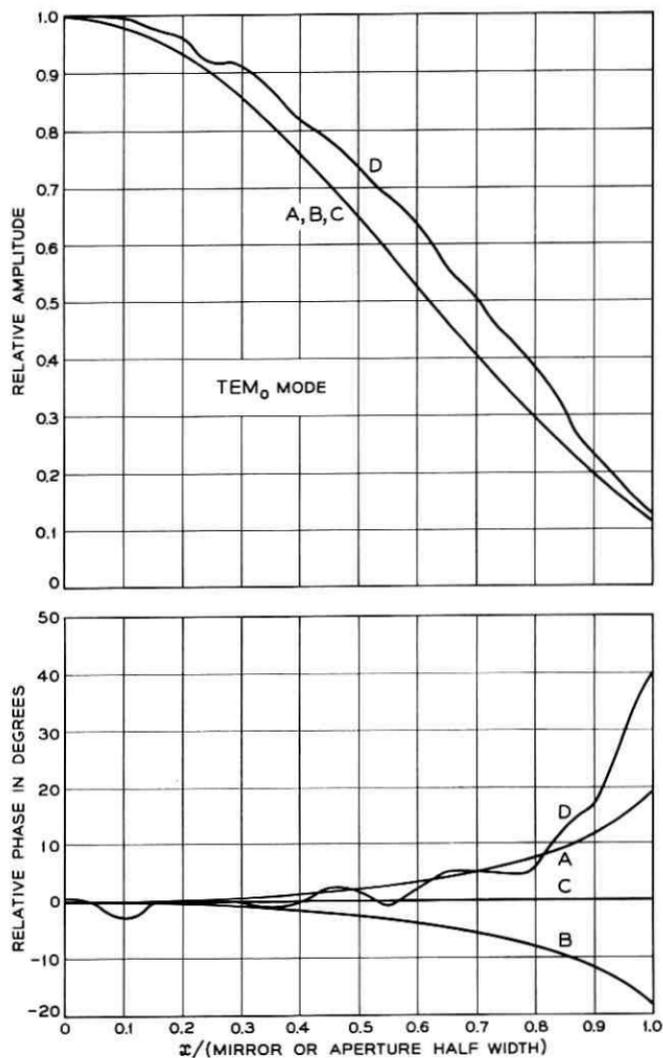
Fig. 3 — Relative amplitude and phase distributions of the lowest-order (TEM$_0$) mode. Both A and B are for an aperture-limited concentric interferometer with $N = 10$ and $a_2/a_1 = 0.0325$; A is the field distribution over the mirror and B is the field distribution over the aperture. C is for the "comparable" confocal interferometer with a Fresnel number of 0.65. D is for a concentric interferometer with $N = 10$ and $a_2 \to \infty$.
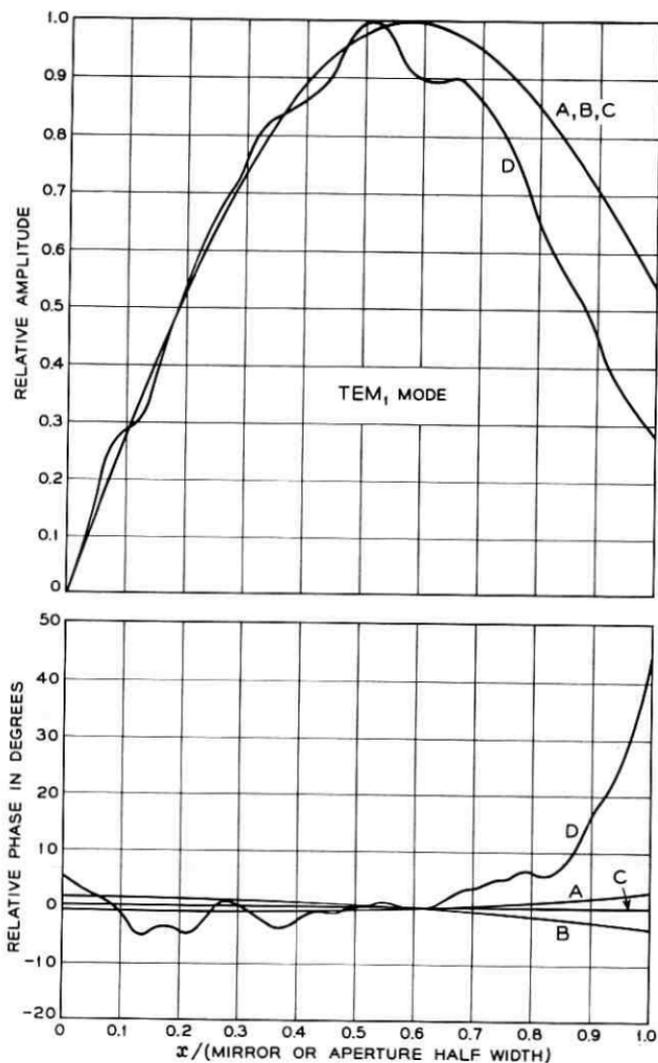
Fig. 4 — Relative amplitude and phase distributions of the second lowest-order (TEM₁) mode. The notation for the curves is the same as for Fig. 3.

steady-state solution from an arbitrary initial trial function, it is possible to infer the relative loss of other higher-order modes. For example, the number of iterations required for the dominant (TEM$_0$) mode of an apertureless concentric (or parallel plane) system with $N = 10$ is about 800, whereas with an aperture such that $a_2/a_1 = 0.0325$ the num-

ber required reduces to about 25, which indicates that the relative losses
of the higher-order even-symmetric modes are now very much higher.

The behavior of the apertured concentric system was found analyt-
ically to approach that of the confocal system in the limit of very small
apertures. The computed results confirm this and show further that the
mode selectivity (in terms of relative loss) of the apertured concentric
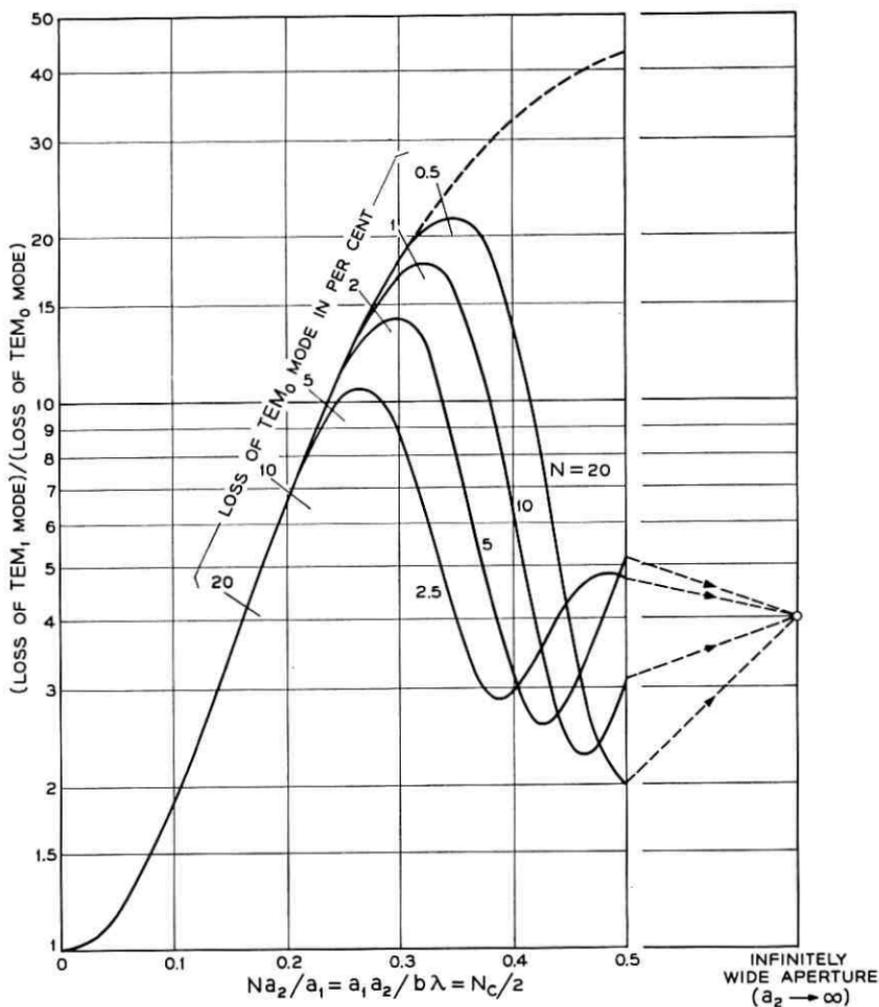
Fig. 5 — Ratio of the losses of the two lowest-order modes versus normalized
aperture half width. The dashed curve is the ratio for the "comparable" confocal
interferometer. The short slant lines are segments of the loci of constant loss for
TEM_0 mode.

system can never be greater than that of the confocal system. However, the effective suppression of unwanted higher-order modes in a maser oscillator requires not only that the relative losses of the higher-order modes be high but also that their absolute losses be greater than the gain of the active medium. To satisfy the second condition a confocal system would have to operate with a very small Fresnel number, whereas a reasonably large Fresnel number can be used for the apertured concentric system. For example, if a maser having a gain of 20 per cent per pass is required to produce a single transverse-mode output, an apertured concentric configuration (with square mirrors) having $N = 20$ and $a_2/a_1 = 0.018$ can be used, but a confocal configuration would need an N_c of about 0.7, which means either very long resonator lengths or very small mirrors. Such configurations are generally undesirable because lengthening the resonator tends to increase the number of longitudinal modes that can oscillate and decreasing the size of the mirrors tends to diminish the power output capability.

IV. ACKNOWLEDGMENT

REFERENCES

1. Fox, A. G., and Li, Tingye, Resonant Modes in a Maser Interferometer B.S.T.J., **40**, March, 1961, p. 453.
2. Boyd, G. D., and Gordon, J. P., Confocal Multimode Resonator for Millimeter through Optical Wavelength Masers, B.S.T.J., **40**, March, 1961, p. 489.
3. Rigrod, W. W., Kogelnik, H., Brangaccio, D. J., and Herriott, D. R., Gaseous Optical Masers with External Concave Mirrors, J. Appl. Phys., **33**, February, 1962, p. 743.
4. Javan, A., Ballik, E. A., and Bond, W. L., Frequency Characteristics of a Continuous-Wave He-Ne Optical Maser, J. Opt. Soc. Am., **52**, January, 1962, p. 96.
5. Haisma, J., and DeLang, H., Mode Patterns Obtained by Tuning a Small Gas Laser, Phys. Letters, **3**, January, 1963, p. 240.
6. Kleinman, D. A., and Kisliuk, P. P., Discrimination against Unwanted Orders in the Fabry-Perot Resonator, B.S.T.J., **41**, March, 1962, p. 453.
7. Kogelnik, H., and Patel, C. K. N., Mode Suppression and Single Frequency Operation in Gaseous Optical Masers, Proc. I.R.E., **50**, November, 1962, p. 2365.
8. Collins, S. A., and White, G. R., Interferometer Laser Mode Selector, Appl. Opt., **2**, April, 1963, p. 448.
9. Burch, J. M., Ruby Masers with Afocal Resonators, J. Opt. Soc. Am., **52**, May, 1962, p. 602.
10. Baker, J. A., and Peters, C. W., Mode Selection and Enhancement with a Ruby Laser, Appl. Opt., **1**, September, 1962, p. 674.

11. Skinner, J. G., and Geusic, J. E., Diffraction-Limited Ruby Oscillator, J. Opt. Soc. Am., **52**, November, 1962, p. 1319.
12. Fox, A. G., and Li, Tingye, Modes in a Maser Interferometer with Curved and Tilted Mirrors, Proc. IEEE, **51**, January, 1963, p. 80.
13. Hildebrand, F. B., *Methods of Applied Mathematics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1952, p. 412.
14. Boyd, G. D., and Kogelnik, H., Generalized Confocal Resonator Theory, B.S.T.J., **41**, July, 1962, p. 1347.

Broadband Electro-Optic Traveling-Wave Light Modulators

By M. DiDOMENICO, JR. and L. K. ANDERSON

(Manuscript received June 13, 1963)

A detailed analysis has been carried out of a broadband traveling-wave electro-optic light modulator using single crystals of currently available materials. The structure analyzed is that first proposed by Rigrod and Kaminow in which a light beam is reflected back and forth across a microwave transmission line with the angle between optical and microwave phase velocity vectors chosen so that the component of the optical phase velocity vector in the direction of microwave propagation is equal to the microwave phase velocity. For generality both amplitude and phase modulation are considered in single crystals belonging to three different symmetry groups, T_d and $D_{2d}(V_d)$, (linear electro-optic effect) and O_h (quadratic effect), for two different orientations of the modulating field.

The inclusion of such practical factors as microwave and optical loss results in an optimum design in which the modulator dimensions and operating temperature are uniquely determined by the optical and microwave dielectric properties of the modulating medium. The main conclusions of the analysis are:

(i) Either cuprous chloride or suitably biased strontium titanate may be used to produce 50 per cent linear modulation over a bandwidth of 10 gc with less than 5 watts of modulating power, using structures a few centimeters long that have practical manufacturing and operating tolerances.

(ii) The upper modulation frequency limit is set by a cutoff frequency which arises from the finite width of the optical beam.

(iii) The presence of appreciable crystalline strain in the modulating medium requires that the modulating system consist of a simple phase modulator followed by some form of frequency-selective discriminator. Such a system requires monochromatic but not necessarily coherent light.

I. INTRODUCTION

Broadband microwave modulation of light requires the use of an extended traveling-wave structure. In order to obtain a cumulative inter-

action between the microwave and optical traveling waves, some form of velocity matching is necessary. In a structure of finite length a practical velocity matching criterion is that the component of the optical phase velocity vector taken along the microwave wave vector be equal to the microwave phase velocity. Several possible arrangements immediately suggest themselves for realizing this type of interaction. For example, we could use true traveling-wave structures[1-4] or iterative structures[5] where the optical and modulating signal are adjusted to interact intermittently with the proper phase. Broadband modulation further requires that the modulating structure have little dispersion, which implies propagation in a TEM or TEM-like mode. Thus, since the dielectric constants of usefully transparent modulating media are much less at optical frequencies than at microwave frequencies, broadband microwave modulation of light can be achieved by using some form of optical slow-wave structure, of which the simplest is that first proposed by Rigrod and Kaminow.[1] In this structure, shown schematically in Fig. 1, an optical plane wave zigzags back and forth in an electro-optic medium between plane mirrors while the interacting microwave modulating signal propagates longitudinally down the structure. The object of this paper is to analyze in detail the operating characteristics of this form of modulator, to present a systematic and quantitative procedure for designing modulators that require a minimum of modulating power, and finally to present numerical results for a number of presently available electro-optic materials.

The kinds of materials considered here fall into two classes. The first of these comprises materials which exhibit a linear electro-optic effect, and the second includes materials which possess a quadratic or Kerr electro-optic effect. The specific materials considered in the first class are:

(i) cubic crystals of symmetry T_d, such as cuprous chloride (CuCl) and zinc sulfide (ZnS),

(ii) KDP and its isomorphs of symmetry $D_{2d}(V_d)$.

The second class of materials consists of centrosymmetric cubic crystals of symmetry O_h such as the titanates (SrTiO$_3$ and BaTiO$_3$) operated in their paraelectric phase.

All electro-optic modulators using transparent media are basically phase modulators in that the modulating fields act directly to change the optical phase velocity of the medium. The phase modulation introduced on a light beam can then be used to transmit intelligence, or if desired it may be converted into intensity modulation. This latter process may in a sense be viewed as a homodyne process[6] in which we beat two phase-modulated normal modes of a birefringent modulating medium
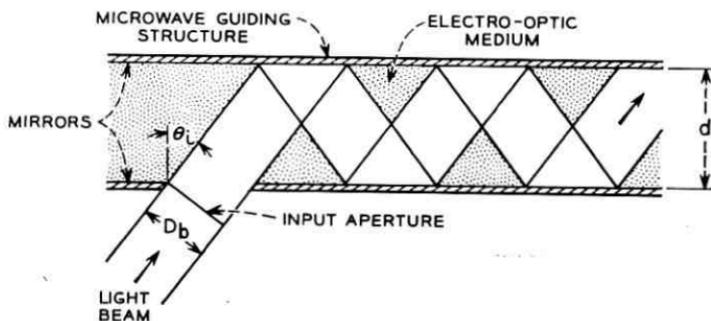
Fig. 1 — Basic zigzag modulator configuration (after Rigrod and Kaminow, Ref. 1).

(e.g., an ordinary and extraordinary wave), the mixing taking place in a polarization analyzer (e.g., a Nicol prism) following the modulator. Most amplitude modulators to date have been of this form.

A systematic comparison of the suitability of the aforementioned materials in a zigzag modulator is complicated by the fact that the optimum modulator configuration depends upon the form of modulation desired (e.g., AM, PM, polarization modulation, etc.) as well as on the demodulation system used (e.g., simple quantum counter, heterodyne, etc.). The comparison is further complicated by the necessity of taking into account such practical problems as crystalline strains, losses (both microwave and optical) and various operational and manufacturing tolerances. Of these effects the existence of strain, particularly random strain, is of paramount importance in determining the ultimate form the modulator will take. The effect of random strain is to partially destroy the spatial coherence of the modulated wave without affecting the temporal coherence. As a result the performance of any modulation-detection system which involves the mixing of two optical signals will be degraded, and we show in fact that the only system largely unaffected by strain is a phase modulator followed by an optical discriminator and quantum counter.

The organization of the remainder of the paper is as follows: In Section II we consider in general the basic phase modulation process in the zigzag modulator, obtaining expressions for the time-varying optical phase shifts which include the effects of microwave loss and lack of synchronism between optical and microwave signals. In Section III we use these results to obtain expressions for the modulation sensitivity of practical phase modulators. In Section IV the same thing is done for amplitude modulators. The practical implications of such effects as diffraction,

strain, optical losses, etc. are discussed in Section V. Finally, in Section VI we outline a procedure for selecting the optimum modulator dimensions and operating temperature so as to obtain the most efficient transfer of information onto the optical beam. In this same section we consider the question of modulator bandwidth.

We conclude with a tabulation of the performance to be expected from presently available materials in the zigzag modulator configuration. Our results show that it should be possible to build a modulator using either the quadratic effect in strontium titanate or the linear effect in cuprous chloride that will provide 50 per cent linear modulation over a 10 gc bandwidth for modulating powers less than about 5 watts. Two factors contribute to the high modulation sensitivity in strontium titanate. These are: (*i*) use of a dc bias results in a large effective microwave electro-optic coefficient, and (*ii*) a long optical path is obtained in a small volume of material, resulting in a very efficient use of microwave power.

II. GENERAL THEORY

The configuration analyzed is shown in Fig. 1. The light is assumed to propagate in the yz plane, between mirrors located in the xy plane, and to have its electric vector polarized either perpendicular to or parallel to the yz plane. The modulating wave propagates in the y direction, along the axis of the modulator.* We may distinguish two cases. In the first (see Fig. 2a) the electric vector of the modulating field lies in the optical plane of incidence (this we call the 0° modulator). In the second (see Fig. 2b) the modulating electric field is perpendicular to the optical plane of incidence (90° modulator). A TEM modulating wave can be propagated in the medium with either polarity if the mirrors are made of nonconducting dielectric multilayers.

In the presence of the modulating field and any dc bias, the optical index ellipsoid[7] is assumed to take the form shown in Fig. 3(a),† with the crystalline axes of the material aligned so that the principal axes of the ellipsoid coincide with the coordinate axes.‡ The significance of the

---

* For convenience, we choose to take the xz plane as the transverse plane of the modulator and the y axis as the direction of propagation or modulator axis. By aligning this Cartesian system with the principal axes of the index ellipsoid of the modulating medium, it can be shown that the only effect of changing the direction of propagation from, say, the y to the x axis is to cause a sign reversal in the retardation of the optical wave.

† By this assumption, we exclude cases in which the electro-optic effect takes the form of a field-dependent orientation of the index ellipsoid. It can be shown that this mode of operation is much less effective than the one in which the orientation of the index ellipsoid remains fixed, and only the lengths of the axes of the ellipsoid change.

‡ The crystal cut necessary to achieve this alignment for each material is given in Appendix A.
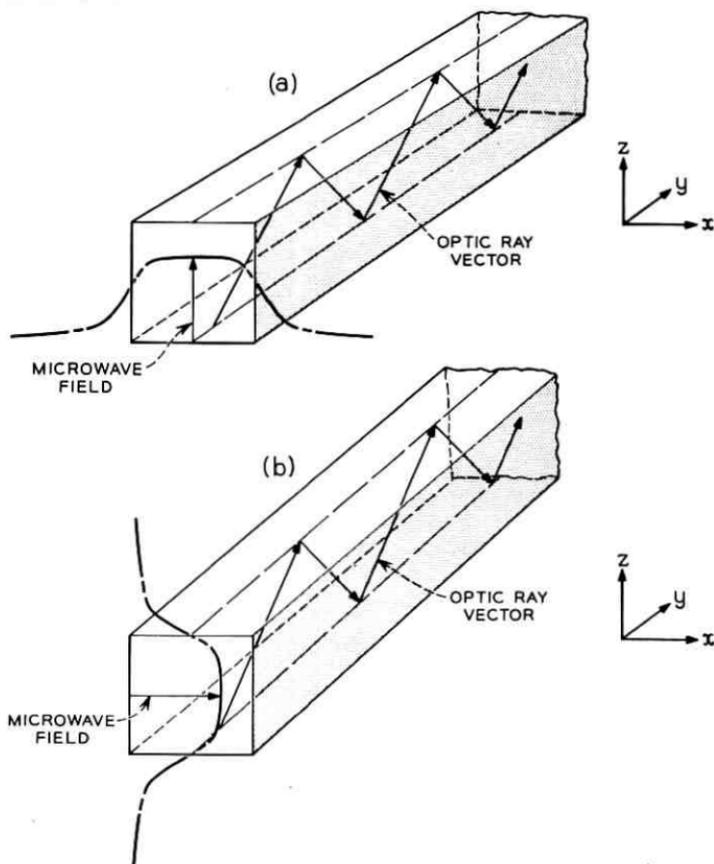
Fig. 2 — Relative orientation of modulating electric field and optical plane of incidence in the two zigzag modulator configurations: (a) 0° modulator; (b) 90° modulator.

index ellipsoid is that it defines the wave vectors and planes of polarization of the two normal modes which can propagate through the medium in any given direction. A plane perpendicular to the wave vector passing through the center of the ellipsoid cuts it in an ellipse whose major and minor axes define the directions of the two orthogonal planes of polarization. The lengths of these axes are equal to the reciprocals of the two indices of refraction. In the configuration shown in Fig. 1, the two normal modes have their electric vectors polarized respectively parallel and perpendicular to the plane of incidence. The two indices of refraction we denote by $n_\parallel$ and $n_\perp$; from Fig. 3(b) it is evident that

$$1/n_\perp^2 = 1/n_x^2 \tag{1a}$$

and

$$\frac{1}{n_\parallel^2} = \frac{\cos^2 \theta_i}{n_y^2} + \frac{\sin^2 \theta_i}{n_z^2}. \tag{1b}$$

As an optical plane wave traverses the modulating medium, its phase velocity varies continuously with time. A convenient way to deduce the time dependence of the over-all phase shift, and hence the phase modulation, is to follow the trajectory of a fixed point on the wave front as it moves through the medium. This is similar to the method of characteristics used to treat nonlinear and parametric wave propagation problems and is valid whenever the phase of the wave or the eikonal of geometrical
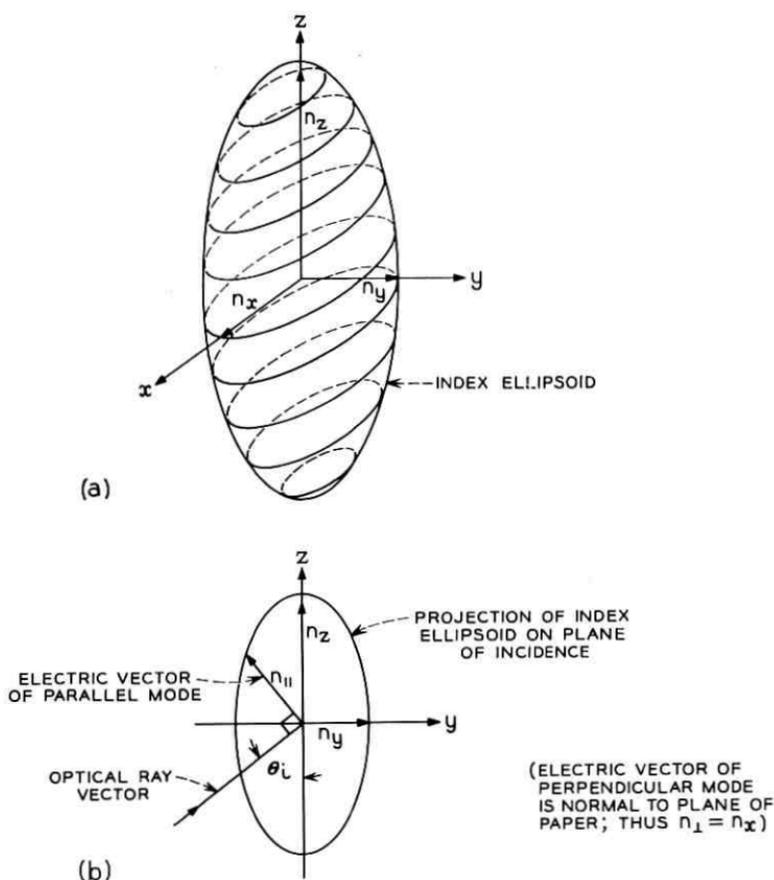
Fig. 3 — (a) Orientation of index ellipsoid with respect to coordinates of modulator; (b) projection of index ellipsoid on optical plane of incidence.

optics is large. When this condition is satisfied the wave propagates approximately as a plane wave. This can be demonstrated by expanding the eikonal in a power series over small space regions and time intervals.[8] It is then correct to say that the time taken for a point on the wavefront to advance a distance y along the modulator axis is given by

$$(t - t_0) = \frac{n_0 y}{c \sin \theta_i} \tag{2}$$

where c is the velocity of light in vacuo, n_0 is the refractive index in the absence of any electric field, t_0 is the starting time at $y = 0$, and θ_i is the angle between the light ray and the z axis. This simple expression replaces a complicated integral equation and is correct to first order in the fractional change $\Delta n/n$ in the index of refraction, and hence is valid only when $|\Delta n/n| \ll 1$, as is usually the case.

The value of the wave phase at any point in space is given by the integral

$$\varphi(t) = \omega \int_{\zeta_1}^{\zeta_2} \frac{d\zeta}{v_p(\zeta,t)}$$

where $\omega/2\pi$ is the optical frequency, and ζ is measured in the ray direction. This equation can be written in terms of the component of the phase velocity vector in the y-direction* as

$$\varphi(t) = \frac{\omega}{\sin \theta_i} \int_0^S \frac{dy}{v_p(y,t)} = \frac{2\pi}{\lambda_0 \sin \theta_i} \int_0^S n(y,t) \, dy \tag{3}$$

where λ_0 is the free-space optical wavelength and S the modulator length. Equation (3) applies separately to both normal modes of the electro-optic medium when the appropriate index of refraction n is used. How n depends on the applied electric field for the various materials and configurations of interest is derived in Appendix A.

By substituting (2) into (3) and utilizing the expressions derived in Appendix A for $n(y,t)$, one can calculate the time dependence of the wave phase correct to first order in the fractional change $\Delta n/n$. As long as the modulator length is bounded, as it must be if diffraction effects, etc., are to be minimized (see Section V), this calculation should give accurate results since $|\Delta n/n| \ll 1$.

In the remainder of this section we make use of (2) and (3) to obtain expressions for the wave phase as a function of time for the two normal modes of the modulating medium. These expressions describe the basic

---

* Note that when dispersion can be neglected the y component of the optical phase velocity vector is just the group velocity of the optical energy flow along the modulator axis.

modulation process within the medium. In subsequent sections we shall go on to describe the practical considerations involved in the actual realization of phase and amplitude modulators.

2.1 Linear Electro-Optic Effect in Cubic Materials

2.1.1 0° Modulator

For this case the two modes have refractive indices given by (1) and (104)

$$\frac{1}{n_\perp} = \frac{1}{n_0} (1 + n_0^2 r_{41} E_z)^{\frac{1}{2}} \tag{4}$$

$$\frac{1}{n_\parallel} = \frac{1}{n_0} (1 - n_0^2 r_{41} E_z \cos^2 \theta_i)^{\frac{1}{2}} \tag{5}$$

where r_{41} is the linear electro-optic coefficient (see Appendix A). For an attenuated modulating wave of the form

$$E_z(y,t) = E_m e^{-\alpha_m y} \cos (\omega_m t - \beta_m y) \tag{6}$$

the phase of the wave polarized perpendicular to the plane of incidence is approximately given by

$$\varphi_\perp(t_0)$$

$$\approx \frac{2\pi}{\lambda_0} \frac{n_0}{\sin \theta_i} \int_0^S \{1 - \tfrac{1}{2} n_0^2 r_{41} E_m \, e^{-\alpha_m y} \cos [\omega_m t_0 + (\beta_e - \beta_m) y]\} \, dy \tag{7}$$

where for convenience we have introduced the notation

$$\frac{n_0 \omega_m}{c \sin \theta_i} \equiv \beta_e . \tag{8}$$

The approximation holds when $| n_0^2 r_{41} E_m | \ll 1$. Equation (7) can be integrated directly, giving

$$\varphi_\perp(t) = \frac{2\pi}{\lambda_0} \frac{n_0}{\sin \theta_i} S - \frac{2\pi}{\lambda_0} \frac{n_0^3}{\sin \theta_i} r_{41} E_m \, e^{-\alpha_m S/2}$$

$$\times \left(\frac{\sinh^2 \frac{\alpha_m S}{2} + \sin^2 \frac{\beta S}{2}}{\alpha_m^2 + \beta^2} \right)^{\frac{1}{2}} \cos (\omega_m t + \psi) \tag{9}$$

where β is defined by

$$\beta = \beta_e - \beta_m \tag{10}$$

and ψ is a time-independent phase factor. In the absence of microwave losses this reduces to the result obtained earlier by Kaminow and Liu.[2] It is evident that the phase modulation described by (9) is maximized by the choice $\beta = \beta_e - \beta_m = 0$, corresponding to the matching of the y-components of the microwave and optical phase velocity vectors. Thus, it is clear that microwave losses do not affect the basic velocity matching condition. Under this condition (9) reduces to

$$\varphi_\perp(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} [1 - \tfrac{1}{2} n_0^2 r_{41} E_m f(\alpha_m S/2) \cos \omega_m t] \quad (11)$$

where for the sake of brevity we have defined the function $f(\alpha_m S/2)$ by

$$\exp\left(-\frac{\alpha_m S}{2}\right) \left[\frac{\sinh (\alpha_m S/2)}{(\alpha_m S/2)}\right] \equiv f\left(\frac{\alpha_m S}{2}\right). \quad (12)$$

For the other wave, whose plane of polarization lies in the plane of incidence, the corresponding expression is

$$\varphi_\parallel(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left[1 + \tfrac{1}{2} n_0^2 r_{41} E_m \cos^2 \theta_i f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t\right]. \quad (13)$$

2.1.2 90° Modulator

The analysis for the 90° modulator is formally identical, except that (105) is used for the principal indices of refraction, in place of (104), so that

$$1/n_\perp = 1/n_0 \quad (14a)$$

and

$$\frac{1}{n_\parallel} = \frac{1}{n_0} (1 - n_0^2 r_{41} E_x \cos 2\theta_i)^{\frac{1}{2}}. \quad (14b)$$

The results for a modulating wave of the form given by (6) are

$$\varphi_\perp(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} = \text{constant} \quad (15)$$

and

$$\varphi_\parallel(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left[1 + \tfrac{1}{2} n_0^2 r_{41} E_m \cos 2\theta_i f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t\right]. \quad (16)$$

2.2 Quadratic Electro-Optic Effect in Cubic and/or Amorphous Materials

The analysis for the quadratic effect in the presence of a large dc bias is very similar to that for the linear effect. The situation is complicated

somewhat by the presence of two independent electro-optic coefficients in the case of the quadratic effect in cubic crystals (see Appendix A). We consider the two configurations discussed before, in which the modulating field lies parallel (0° modulator) or perpendicular (90° modulator) to the optical plane of incidence.

2.2.1 0° Modulator

For this case the relevant refractive indices are [refer to (110)]

$$\frac{1}{n_\perp} = \frac{1}{n_0} (1 + n_0^2 \rho_2 E_z^2)^{\frac{1}{2}} \tag{17}$$

and

$$\frac{1}{n_\parallel} = \frac{1}{n_0} \left[1 + n_0^2 \rho_1 E_z^2 \left(\sin^2 \theta_i + \frac{\rho_2}{\rho_1} \cos^2 \theta_i \right) \right]^{\frac{1}{2}} \tag{18}$$

where ρ_1 and ρ_2 are quadratic electro-optic coefficients (see Appendix A). In order to achieve approximately linear modulation, as well as to enhance the modulation sensitivity, the applied field should consist of a large static bias and a smaller microwave field, so that

$$E_z(y,t) = E_0 + E_m e^{-\alpha_m y} \cos (\omega_m t - \beta_m y). \tag{19}$$

Then to first order in E_m/E_0

$$E_z^2 \approx E_0^2 + 2E_0 E_m e^{-\alpha_m y} \cos (\omega_m t - \beta_m y)$$

and the expressions for $1/n_\perp$ and $1/n_\parallel$ become linear in E_m. The integrations needed to obtain $\varphi_\perp$ and $\varphi_\parallel$ are similar to those in the preceding section. Again the maximum ac phase shift occurs with velocity matching, for which case we obtain

$$\varphi_\perp(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left\{ 1 - \tfrac{1}{2} n_0^2 \rho_2 E_0 \left[E_0 + 2E_m f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right] \right\} \tag{20}$$

and

$$\varphi_\parallel(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left\{ 1 - \tfrac{1}{2} n_0^2 \rho_1 E_0 \left(\sin^2 \theta_i + \frac{\rho_2}{\rho_1} \cos^2 \theta_i \right) \right.$$
$$\left. \times \left[E_0 + 2E_m f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right] \right\}. \tag{21}$$

In deriving (20) and (21) we note that because the dc bias makes the material birefringent even in the absence of a microwave modulating field, the ordinary ($\perp$) and extraordinary ($\parallel$) modes cannot, strictly

speaking, be simultaneously matched to the modulating wave. We assume, however, that the bias is small ($|n_0^2 \rho E_0^2| \ll 1$) so that the distinction between ordinary and extraordinary phase velocities becomes unimportant. For known materials this condition is met for fields as large as 10^5 v/cm — i.e., for any field that the material can sustain without dielectric breakdown.

2.2.2 90° Modulator

The analysis here is similar to the 0° case, except that the indices of refraction are simply [refer to (111)]

$$\frac{1}{n_\perp} = \frac{1}{n_0} (1 + n_0^2 \rho_1 E_x^2)^{\frac{1}{2}} \tag{22a}$$

and

$$\frac{1}{n_\parallel} = \frac{1}{n_0} (1 + n_0^2 \rho_2 E_x^2)^{\frac{1}{2}} \tag{22b}$$

so that the resulting wave phases are

$$\varphi_\perp(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left\{ 1 - \tfrac{1}{2} n_0^2 \rho_1 E_0 \left[E_0 + 2E_m f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right] \right\} \tag{23}$$

and

$$\varphi_\parallel(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left\{ 1 - \tfrac{1}{2} n_0^2 \rho_2 E_0 \left[E_0 + 2E_m f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right] \right\}. \tag{24}$$

2.3 Linear Electro-Optic Effect in Uniaxial Materials

The results here are similar to those obtained earlier for the effect in cubic materials.

2.3.1 0° Modulator

Equations (1) and (113) apply here, so that

$$\begin{aligned} \frac{1}{n_\perp} &= \frac{1}{n_0} (1 + n_0^2 r_{63} E_z)^{\frac{1}{2}} \\ \frac{1}{n_\parallel} &\approx \frac{1}{n_0} \left(1 - n_0^2 r_{63} E_z \cos^2 \theta_i - \frac{2\Delta n}{n_0} \sin^2 \theta_i \right)^{\frac{1}{2}} \end{aligned} \tag{25}$$

where we have set $\Delta n \equiv n_e - n_0$, which we assume to be much less than unity (n_e refers to the extraordinary refractive index). The wave phases for the two modes are then given by

$$\varphi_{\perp}(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left[1 + \tfrac{1}{2} n_0^2 r_{63} E_m f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right] \quad (26)$$

and

$$\varphi_{\parallel}(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i}$$
$$\times \left[1 - \frac{2\Delta n}{n_0} \sin^2 \theta_i - \tfrac{1}{2} n_0^2 r_{63} E_m \cos^2 \theta_i f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right]. \quad (27)$$

2.3.2 *90° Modulator*

From (1) and (114) we obtain

$$1/n_{\perp} = 1/n_e$$

and

$$\frac{1}{n_{\parallel}} = \frac{1}{n_0} \left(1 - n_0^2 r_{63} E_x \cos^2 \theta_i \right)^{\frac{1}{2}} \quad (28)$$

so that the wave phases are

$$\varphi_{\perp}(t) = \frac{2\pi}{\lambda_0} \frac{n_e S}{\sin \theta_i} = \text{constant} \quad (29)$$

and

$$\varphi_{\parallel}(t) = \frac{2\pi}{\lambda_0} \frac{n_0 S}{\sin \theta_i} \left[1 - \tfrac{1}{2} n_0^2 r_{63} E_m \cos 2\theta_i f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \right]. \quad (30)$$

III. PHASE MODULATORS

The preceding results are immediately applicable to the phase modulator. The maximum modulation sensitivity is obtained when the beam is linearly polarized in the direction ($\perp$ or $\parallel$) showing the greatest effect, although unpolarized light can also be used.

For all the cases discussed in the previous section, the ac component of the wave phase can be written in the form

$$\varphi_{ac}(t) = \frac{2\pi n_0^3}{\lambda_0} A S g(\theta_i) f\left(\frac{\alpha_m S}{2}\right) E_m \cos \omega_m t, \quad (31)$$

where A is proportional to electro-optic coefficient and $g(\theta_i)$ depends on the angle of incidence. The quantity of most immediate interest is the modulation sensitivity, which we may define as the ratio of the

peak phase shift to the square root of the power input to the modulator.* To make the problem definite, we consider a parallel-plate line of spacing h and width w, filled with a medium of microwave dielectric constant κ'. If we ignore fringing fields,† the power flow is given in terms of the electric field by

$$P = \frac{1}{2} \frac{\sqrt{\kappa'}}{\eta} E_m^2 hw, \tag{32}$$

where $\eta = 377$ ohms is the impedance of free space. The modulation sensitivity is then

$$\left| \frac{M}{\sqrt{P}} \right| = \frac{2\pi n_0^3}{\lambda_0} A Sg(\theta_i) f\left(\frac{\alpha_m S}{2}\right) \sqrt{\frac{2\eta}{\sqrt{\kappa'}\,hw}} \tag{33}$$

where M is the peak ac phase shift. Since the angle of incidence is determined in terms of material parameters by the velocity matching condition $\beta_e = \beta_m$, which reduces to $\sin\theta_i = n_0/\sqrt{\kappa'}$ for the present geometry, (33) can also be written as

$$\left| \frac{M}{\sqrt{P}} \right| = \frac{2\pi n_0^3}{\lambda_0} A Sp\left(\frac{n_0}{\sqrt{\kappa'}}\right) f\left(\frac{\alpha_m S}{2}\right) \sqrt{\frac{2\eta}{\sqrt{\kappa'}\,hw}} \tag{34}$$

where $p(n_0/\sqrt{\kappa'})$ is a function of $(n_0/\sqrt{\kappa'})$ which replaces $g(\theta_i)$. Values of A, $g(\theta_i)$ and $p(n_0/\sqrt{\kappa'})$ for the cases of interest here are given in Table I. In this table, the linear electro-optic coefficient which we write simply as r_{ij} is to be taken to be r_{41} in the case of cubic materials, and r_{63} for uniaxial crystals.

Numerical results for the power required in practical cases are given in Section VII.

IV. AMPLITUDE MODULATORS

In this section we treat the problem of obtaining broadband amplitude-modulated light by suitably combining the two simultaneously phase-modulated normal modes of a transparent electro-optic medium. In the ensuing treatment we will consider only the intensity variations in the modulated beam. In so doing we are ignoring the phase information that would be recovered, for example, by an optical heterodyne re-

---

* The advantage of defining a modulation sensitivity in this way is that it is a function of the microwave-optical properties of the electro-optic modulating medium and is expressed in terms of power which is easily measured.

† The effects of fringing fields and the problems associated with non-TEM mode propagation in this structure are considered in Section 5.1.

TABLE I — FUNCTIONS APPEARING IN (33) AND (34) FOR
VARIOUS PHASE MODULATOR CONFIGURATIONS

Modulator	Mode	A	$g(\theta_i)$	$p(n_0/\sqrt{\kappa'})$
Linear, 0°	$\perp$	$\frac{1}{2}r_{ij}$	$1/\sin\theta_i$	$\sqrt{\kappa'}/n_0$
	$\parallel$	$\frac{1}{2}r_{ij}$	$\dfrac{\cos^2\theta_i}{\sin\theta_i}$	$\dfrac{\sqrt{\kappa'}}{n_0}\left(1 - \dfrac{n_0^2}{\kappa'}\right)$
Linear, 90°	$\perp$	0	—	—
	$\parallel$	$\frac{1}{2}r_{ij}$	$\dfrac{\cos 2\theta_i}{\sin\theta_i}$	$\dfrac{\sqrt{\kappa'}}{n_0}\left(1 - \dfrac{2n_0^2}{\kappa'}\right)$
Quadratic, 0°	$\perp$	$\rho_2 E_0$	$1/\sin\theta_i$	$\sqrt{\kappa'}/n_0$
	$\parallel$	$\rho_1 E_0$	$\dfrac{\sin^2\theta_i + \dfrac{\rho_2}{\rho_1}\cos^2\theta_i}{\sin\theta_i}$	$\dfrac{\dfrac{n_0^2}{\kappa'} + \dfrac{\rho_2}{\rho_1}\left(1 - \dfrac{n_0^2}{\kappa'}\right)}{\dfrac{n_0}{\sqrt{\kappa'}}}$
Quadratic, 90°	$\perp$ $\parallel$	$\rho_1 E_0$ $\rho_2 E_0$	$1/\sin\theta_i$	$\sqrt{\kappa'}/n_0$

ceiver, and in effect are assuming that the receiver is a simple quantum counter. Thus we look for modulation of the form

$$I(t) = I_0(1 + M \cos \omega_m t) \tag{35}$$

where I_0 is the average light intensity and M is the modulation index.

Cubic materials are of particular interest as amplitude modulators since they are isotropic in the absence of an applied field and become uniaxial only upon application of a field. This is in contrast to what takes place in materials with lower degrees of symmetry where naturally occurring birefringence exists. The significance of starting with an isotropic crystal, as opposed to a uniaxial or even biaxial crystal, is that in the former no static retardation is accumulated between the two normal modes. Hence any ac difference in phase produced between these modes can easily be converted into useful intensity modulation. The presence of a static contribution to the differential phase shift can cause the induced ac birefringence to be masked out due to the angular spread of the beam, constructional uncertainties or finite source linewidth. (See Section 4.3.)

There are two simple ways in which a time-varying phase shift be-

tween two orthogonally polarized light beams may be converted into approximately linear intensity modulation. One of these is most useful when there is no dc retardation between the two beams, and the other when there is a controllable dc retardation, as in the biased quadratic effect modulator.

4.1 *Amplitude Modulation Using the Linear Electro-Optic Effect in Cubic Materials*

The first method of producing amplitude modulation is shown in Fig. 4. Circularly polarized light is produced by passing the linearly polarized incident beam through a quarter-wave plate in the manner shown.

The light emerging from the quarter-wave plate may be decomposed into two orthogonal normal modes, one aligned with the fast or ζ axis of the quarter-wave plate and the other aligned with the slow or ξ axis (see Fig. 4). Both normal modes are of equal amplitude but separated in phase by 90°. In order to obtain the desired modulation, the plane of transmission of the analyzer must make an angle of 45° with the plane of incidence.

For the arrangement given in Fig. 4 one can easily show that the ratio of transmitted to incident light intensity is given by

$$I_t/I_0 = \tfrac{1}{2}[1 + \sin(\varphi_\perp - \varphi_\parallel)]. \tag{36}$$
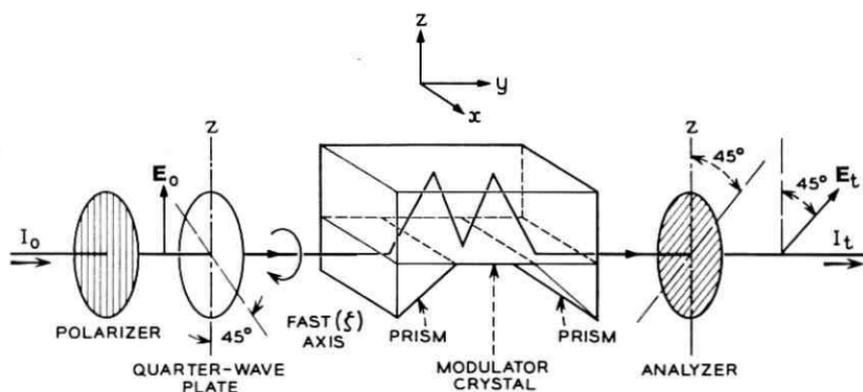
Fig. 4 — Arrangement for obtaining linear amplitude modulation for cubic crystals exhibiting a linear electro-optic effect. This arrangement can also be used with uniaxial crystals if the quarter-wave plate is replaced by a suitable optical compensator.

For the modulators using the linear electro-optic effect in cubic crystals, (11), (13), (15), and (16) yield

$$\varphi_\perp - \varphi_\parallel = - \frac{\pi n_0^3 r_{41}}{\lambda_0} E_m S g(\theta_i) f\left(\frac{\alpha_m S}{2}\right) \cos \omega_m t \qquad (37)$$

where

$$g(\theta_i) = \begin{cases} \dfrac{1 + \cos^2 \theta_i}{\sin \theta_i} & (0° \text{ modulator}) \\[3mm] \dfrac{\cos 2\theta_i}{\sin \theta_i} & (90° \text{ modulator}). \end{cases} \qquad (38)$$

To obtain a specific expression for the modulation sensitivity, which we define as $|M/\sqrt{P}|$ as before, we may again consider the simple parallel plate geometry discussed in Section III. It is evident from (36) and (37) that the modulated signal will in general consist of an infinite number of sidebands whose amplitudes are determined by Bessel functions. For simplicity we limit ourselves to the case of small modulation index (viz., $|\varphi_\perp - \varphi_\parallel| < \pi/6$) so that only the first sideband is important. We then obtain for the modulation sensitivity

$$\left|\frac{M}{\sqrt{P}}\right| = \frac{\pi \sqrt{\eta/2}}{V_{\frac{1}{2}}} \frac{S}{\sqrt{hw}} \frac{1}{\kappa'^{\frac{1}{4}}} p\left(\frac{n_0}{\sqrt{\kappa'}}\right) f\left(\frac{\alpha_m S}{2}\right), \qquad (39)$$

with

$$p\left(\frac{n_0}{\sqrt{\kappa'}}\right) = \begin{cases} \dfrac{2 - n_0^2/\kappa'}{n_0/\sqrt{\kappa'}} & (0° \text{ modulator}) \\[3mm] \dfrac{1 - 2n_0^2/\kappa'}{n_0/\sqrt{\kappa'}} & (90° \text{ modulator}). \end{cases} \qquad (40)$$

In this expression we have again used the velocity matching condition $\sin \theta_i = n_0/\sqrt{\kappa'}$ to replace $g(\theta_i)$ by $p(n_0/\sqrt{\kappa'})$, and have replaced the electro-optic coefficient with the more generally useful half-wave retardation voltage, $V_{\frac{1}{2}}$, defined by*

$$V_{\frac{1}{2}} = \frac{\lambda_0}{2n_0^3 r_{41}}. \qquad (41)$$

Once again we defer computing actual numbers until Section VII.

---

* Note that when applied to the transverse electro-optic effect $V_{\frac{1}{2}}$ need not be the actual voltage applied to the crystal to obtain π phase retardation, which depends on the geometry of the crystal, but is to be viewed as a convenient collection of constants.

4.2 *Amplitude Modulation Using the Quadratic Electro-Optic Effect*

Use of the quadratic effect in conjunction with a dc bias to provide amplitude modulation differs from the linear case just considered in that there is now a bias-field dependent static phase difference produced between the ordinary and extraordinary rays. Approximately linear intensity modulation can be obtained in this case by placing the active medium between crossed polarizer and analyzer, as shown in Fig. 5, provided that the dc bias is adjusted so that operation occurs about the point of half-transmission. For the arrangement of Fig. 5, the ratio of transmitted to incident light intensity is given by

$$I_t/I_0 = \tfrac{1}{2}[1 + \cos(\varphi_\perp - \varphi_\parallel)]. \qquad (42)$$

According to (20), (21), (23), and (24), the phase difference for the quadratic effect can be written as

$$|\varphi_\perp - \varphi_\parallel| = \frac{\pi n_0^3}{\lambda_0}(\rho_1 - \rho_2)E_0 S g(\theta_i)\left[E_0 + 2E_m f\left(\frac{\alpha_m S}{2}\right)\cos\omega_m t\right] \qquad (43)$$

where

$$g(\theta_i) = \begin{cases} \sin\theta_i & (0° \text{ modulator}) \\ 1/\sin\theta_i & (90° \text{ modulator}). \end{cases} \qquad (44)$$

Equation (43) is in the form $\varphi_\perp - \varphi_\parallel = \gamma_{dc} + \gamma_{ac}$; hence (42) can be written:

$$I_t/I_0 = \tfrac{1}{2}[1 + \cos\gamma_{dc}\cos\gamma_{ac} - \sin\gamma_{dc}\sin\gamma_{ac}]. \qquad (45)$$
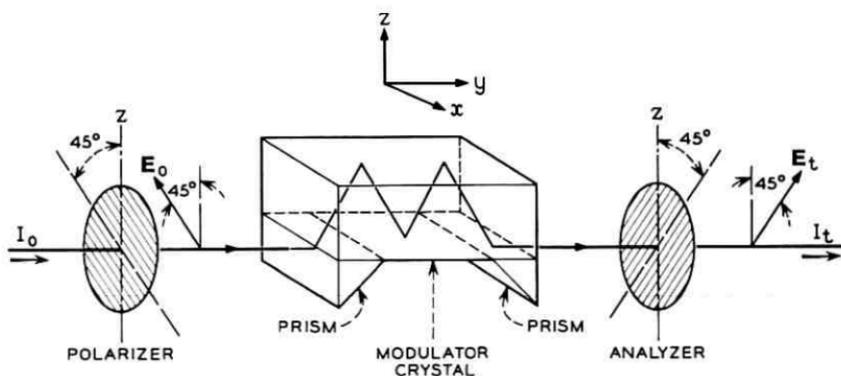

Fig. 5 — Arrangement for obtaining linear amplitude modulation in materials exhibiting a quadratic electro-optic effect.

For $\gamma_{ac} < \pi/6$ (small modulation) this becomes, approximately

$$I_t/I_0 \approx \tfrac{1}{2}(1 - \gamma_{ac} \sin \gamma_{dc}). \tag{46}$$

Maximum modulation sensitivity results when the operating point is chosen so that*

$$\gamma_{dc} \equiv \frac{\pi n_0^3 (\rho_1 - \rho_2) E_0^2 S g(\theta_i)}{\lambda_0} = (2m + 1)\frac{\pi}{2}, \qquad m = 0, 1, 2, \cdots. \tag{47}$$

The point that should be noted here is that for the linear Kerr effect modulator there are modes of operation and that these modes, which are designated by the index m, constitute an independent operating variable.

For the operating points defined by (47), the modulation sensitivity for the parallel-plate geometry can be written in the form

$$\left| \frac{M}{\sqrt{P}} \right| = 4\pi \sqrt{2\eta} \, \frac{S}{\sqrt{hw}} \frac{1}{\kappa'^{\frac{1}{4}}} \, p\left(\frac{n_0}{\sqrt{\kappa'}} \right) K E_0 f\left(\frac{\alpha_m S}{2} \right), \tag{48}$$

where

$$p\left(\frac{n_0}{\sqrt{\kappa'}} \right) = \begin{cases} \dfrac{n_0}{\sqrt{\kappa'}} & (0° \text{ modulator}) \\[2ex] \dfrac{\sqrt{\kappa'}}{n_0} & (90° \text{ modulator}). \end{cases} \tag{49}$$

In deriving this equation we have, as before, made use of the synchronism condition $\sin \theta_i = n_0/\sqrt{\kappa'}$, and have replaced the electro-optic coefficient $(\rho_1 - \rho_2)$ by the more familiar Kerr constant

$$K = \frac{n_0^3 (\rho_1 - \rho_2)}{2\lambda_0}. \tag{50}$$

It should be noted that the dc field is not an independent variable, but, because it determines the operating point [see (47)], is limited to the discrete values

$$E_0 = \frac{1}{2} \sqrt{\frac{(2m + 1)}{KSp\left(\dfrac{n_0}{\sqrt{\kappa'}} \right)}}. \tag{51}$$

In practice E_0 would be made as large as the dielectric strength of the material would permit, subject to (50). We should point out that the

---

* It is obvious that the arrangement shown in Fig. 4 could also be utilized here if desired; the only difference would be that the quiescent operating point would be chosen to make the dc part of $\varphi_\perp - \varphi_\parallel$ in (43) an odd multiple of π.

effects of electrostriction at high dc fields will undoubtedly result in electrostrictively induced birefringence. To the extent that the electrostriction is uniform, it is equivalent to a change in the Kerr constant, and can be compensated for by tuning E_0. On the other hand, nonuniform electrostriction can degrade the modulator performance in the same way as nonuniform strain. (See Section 5.6.)

4.3 *Amplitude Modulation Using the Transverse Linear Electro-Optic Effect in Uniaxial Crystals*

Use of the zigzag configuration to produce amplitude modulation with noncubic materials is complicated somewhat by the presence of a large static phase retardation due to the natural birefringence of the material. When the modulator is used with well-collimated monochromatic light, however, the effects of the natural birefringence can in principle be eliminated by means of an optical compensator, thereby making it worthwhile to consider this case as well.

The physical arrangement used in conjunction with the linear effect in cubic crystals, and shown in Fig. 4, is also applicable here if we replace the quarter-wave plate by a suitable compensator.[10] The compensator allows the phase of one optical mode of the modulator to be varied relative to that of the other.

For this arrangement the ratio of transmitted to incident intensity is given by

$$I_t/I_0 = \tfrac{1}{2}[1 + \cos{(\varphi_\perp - \varphi_\parallel + \varphi_c)}] \tag{52}$$

where φ_c is the relative phase shift introduced by the compensator. The phase retardation in the modulator is given by (26) to (30), and has the form

$$\varphi_\perp - \varphi_\parallel = -\frac{\pi n_0^3 r_{63}}{\lambda_0} E_m S g(\theta_i) f\left(\frac{\alpha_m S}{2}\right) \cos{(\omega_m t)} + \varphi_s \tag{53}$$

where $g(\theta_i)$ is given by (38) of Section 4.2 and φ_s is the static retardation, given by

$$\varphi_s = \begin{cases} \dfrac{4\pi \Delta n S}{\lambda_0} \sin{\theta_i} & (0° \text{ modulator}) \\[2ex] 2\pi \dfrac{\Delta n S}{\lambda_0} \dfrac{1}{\sin{\theta_i}} & (90° \text{ modulator}). \end{cases} \tag{54}$$

It is evident from (52) and (53) that maximum linear amplitude modulation is obtained when the compensator phase shift is adjusted so that

$$\varphi_c + \varphi_s = (2k + 1)\pi/2, \qquad k = 0, 1, 2, 3, \cdots, \tag{55}$$

for which case (52) becomes

$$I_t/I_0 = \tfrac{1}{2}[1 - (-1)^{(k+1)} \sin (\varphi_\perp - \varphi_\parallel)_{\text{ac}}]. \tag{56}$$

With this adjustment of φ_c, uniaxial and cubic materials are optically equivalent as far as modulator operation is concerned [compare (53) and (56) with (36) and (37)], so that (39) to (41) for the modulation sensitivity can be applied here as well by changing r_{41} to r_{63}.

The question arises, of course, as to the practicality of compensating for the natural birefringence over a finite aperture; if the modulator is to function properly, the static retardation φ_s must be the same, within a fraction of a radian, for all rays comprising the beam. Three factors are involved here:

   (*i*) monochromaticity of the optical source,
   (*ii*) dimensional tolerances, and
   (*iii*) angular spread in the incident beam.

To illustrate the problem we consider the 0° modulator in KDP. In this configuration the static retardation is given by

$$\varphi_s = 4\pi \frac{\Delta n S}{\lambda_0} \sin \theta_i. \tag{57}$$

If we neglect dispersion in Δn, a range of wavelengths $\Delta \lambda_0$ in the source leads to a spread in φ_s of

$$|\Delta \varphi_s| \approx \varphi_s \left| \frac{\Delta \lambda_0}{\lambda_0} \right| \tag{58}$$

while a variation ΔS in the modulator length results in a spread

$$\Delta \varphi_s \approx \varphi_s \frac{\Delta S}{S}. \tag{59}$$

Under typical circumstances ($\Delta n = 0.04$, $S = 4$ cm, $\theta_i = 19°$, $\lambda_0 = 0.6 \, \mu$; see Section VII) φ_s is of order 10^4 radians, so that the requirement $\Delta \varphi_s < 1$ implies that the source monochromaticity and modulator dimensions be maintained within a few parts in 10^5. Although the foregoing tolerances are tight, they are attainable.

A more fundamental problem concerns the angular spread in the incident beam as it traverses the modulating medium. For example, let us suppose that the beam divergence can be characterized by a spread $\Delta \theta_i$ in the angle of incidence. This will produce a variation in φ_s given by

$$\Delta \varphi_s \approx 4\pi \frac{\Delta n S}{\lambda_0} \cos \theta_i \Delta \theta_i \tag{60}$$

so that the inequality $\Delta\varphi_s < 1$ requires that

$$\Delta\theta_i < \frac{\lambda_0}{4\pi\Delta n S \cos\theta_i}. \tag{61}$$

For the example presently under consideration, (61) gives $\Delta\theta_i < 3.2 \times 10^{-5}$ radian $= 6.5$ sec of arc. Whether or not this requirement can be met depends on the details of the optical system. If, for example, the modulator were operated in the far field of an optical maser of output aperture D which is smaller than the modulator aperture, diffraction alone would produce a spread $\Delta\theta_i \approx 1.22\lambda_0/D$, so that the inequality could be satisfied only if the beam diameter were greater than 2.3 cm, which is clearly unreasonable. On the other hand, if the modulator were placed in the near field of the optical maser, (which extends for a distance D^2/λ_0, which is of order 1 m for a 1-mm diameter beam), diffraction could be negligible, and strain refraction would then set a lower limit to the minimum realizable effective $\Delta\theta_i$, which would in all likelihood still be considerably in excess of the limit set by (61).

In view of the foregoing practical considerations it would seem unreasonable to attempt the construction of an amplitude modulator in the zigzag configuration using a material such as KDP. Moreover, as will be shown in connection with the KDP phase modulator, this material typically requires at least an order of magnitude more power than other materials under consideration.

4.4 Summary of Results

The results for the various forms of amplitude modulator may be summarized in a single equation, of the same form as that obtained for the phase modulators

$$\left| \frac{M}{\sqrt{P}} \right| = \frac{2\pi n_0^3}{\lambda_0} A S p \left(\frac{n_0}{\sqrt{\kappa'}} \right) f \left(\frac{\alpha_m S}{2} \right) \sqrt{\frac{2\eta}{\sqrt{\kappa'}\, hw}}. \tag{62}$$

Values of A and $p(n_0/\sqrt{\kappa'})$ are given in Table II.

V. FACTORS AFFECTING PRACTICAL MODULATOR PERFORMANCE

It is appropriate before discussing the actual design and performance of zigzag modulators to consider some of the practical factors which affect the operation of the device. In particular we consider the effects of

(i) fringing fields
(ii) deviations from the correct angle of incidence
(iii) diffraction

TABLE II — FUNCTIONS APPEARING IN (62) FOR VARIOUS
AMPLITUDE MODULATOR CONFIGURATIONS

Modulator	A	$g(\theta_i)$	$p(n_0/\sqrt{\kappa'})$
Linear, 0°	$\frac{1}{2}r_{ij}$	$\dfrac{1 + \cos^2 \theta_i}{\sin \theta_i}$	$\dfrac{2 - n_0^2/\kappa'}{n_0/\sqrt{\kappa'}}$
Linear, 90°	$\frac{1}{2}r_{ij}$	$\dfrac{\cos 2\theta_i}{\sin \theta_i}$	$\dfrac{1 - 2n_0^2/\kappa'}{n_0/\sqrt{\kappa'}}$
Quadratic, 0°	$(\rho_1 - \rho_2)E_0$	$\sin \theta_i$	$n_0/\sqrt{\kappa'}$
Quadratic, 90°	$(\rho_1 - \rho_2)E_0$	$1/\sin \theta_i$	$\sqrt{\kappa'}/n_0$

(iv) optical losses

(v) lack of parallelism in the reflecting surfaces, and

(vi) strain.

5.1 *Effect of Fringing Fields*

The calculations made thus far are strictly correct only in the limit of a uniform microwave field across the active cross section of the modulator. Whether this condition is satisfied or not depends upon the ratio of the width of the electro-optic modulating medium, w, to the width, H, of the parallel plate conductor (see Fig. 6). This ratio, given by

$$\xi = w/H \tag{63}$$

determines the filling factor of the dielectric loaded parallel plate guide. It can be shown from the analysis of Kaminow and Liu[2] that when $\xi = 1$ the field distribution in the modulating medium is uniform and propagates as a TEM wave. This result is, however, valid only when open-circuit boundary conditions prevail at the ends of the waveguide cross section.[2] These boundary conditions will apply physically only to the extent that fringing fields can be neglected. This condition can be expected to be satisfied when the change in transverse impedance in going from the dielectric region to air is large, or alternatively, when the energy stored in the dielectric is large compared to the energy stored in the fringing fields. As long as the dielectric constant of the electro-optic medium is in excess of 10, fringing fields should be negligible and therefore have relatively little effect on the foregoing calculations.

It is important to note that it is generally difficult to propagate TEM or TEM-like waves over a broad band of frequencies in partially loaded

two-conductor transmission lines of the form shown in Fig. 6. This is because dielectric waveguide modes can be supported in such structures at high frequencies. The relative importance of the dielectric modes depends on the filling factor ξ. For filling factors less than unity, three propagating regions can be distinguished.[2] The first occurs at low frequencies where

$$\lambda_m \gg \sqrt{\kappa'}w. \tag{64}$$

In this region the field distribution in the dielectric region is uniform and TEM-like mode propagation results. If, on the other hand, we have

$$\lambda_m \ll \sqrt{\kappa'}w \tag{65}$$

then, as shown in Ref. 2, the microwave energy resides primarily within the dielectric with a cosinusoidal field distribution, and once again propagates as a TEM-like wave. For those cases where $\lambda_m \approx \sqrt{\kappa'}w$ we find that a symmetrical TE_{10}-like waveguide mode propagates with the dispersion characteristic of such a mode.

What is of primary significance here is that for values of the filling factor between

$$0.5 < \xi \leqq 1$$

the dispersion in the microwave phase velocity is small (from the calculations given in Ref. 2 a 30–40 per cent variation in phase velocity results for $\xi = 0.5$). Moreover, as the filling factor approaches unity, not only does the dispersion vanish but the field distribution also becomes uniform. Based upon these considerations it is evident that a parallel plate guide completely filled with a moderately high dielectric constant elec-
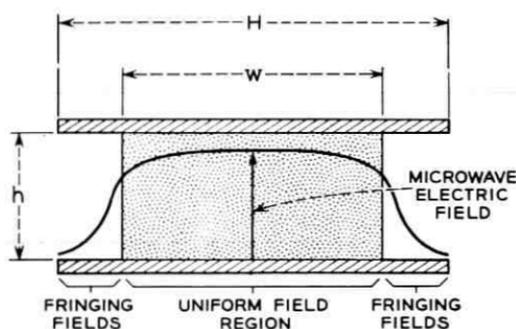

Fig. 6 — Representative microwave electric field distribution in the transverse plane of the modulating medium.

tro-optic material is a suitable practical broadband microwave guiding structure.

5.2 *Sensitivity of the Modulator Performance to Small Changes in the Angle of Incidence*

The velocity matching condition requires that the angle of incidence be adjusted so that $\sin \theta_i = n_0/\sqrt{\kappa'}$. To determine the effect of small deviations from this value, we return to the general expressions for the basic phase modulation processes [(9) and following] and observe that in the lossless case ($\alpha_m = 0$) the modulation sensitivity is always of the form

$$\left| \frac{M}{\sqrt{P}} \right| = C_1 g(\theta_i) \frac{\sin \frac{\beta S}{2}}{\frac{\beta S}{2}} \tag{66}$$

where $g(\theta_i)$ is a slowly varying function of θ_i, and the parameter β is given by [see (8) and (10)]

$$\beta = \beta_m \left(\frac{n_0/\sqrt{\kappa'}}{\sin \theta_i} - 1 \right).$$

Neglecting any variation of $g(\theta_i)$, we find that a variation in θ_i of

$$| \Delta \theta_i | \approx 0.6 \frac{\lambda_m}{S} \frac{n_0}{\kappa'} \left(1 - \frac{n_0^2}{\kappa'} \right)^{-\frac{1}{2}} \tag{67}$$

from the value for perfect velocity matching will reduce the modulation sensitivity by a factor of 2.

5.3 *Diffraction Effects*

Diffraction can limit the performance of the zigzag modulator by producing a divergence in the beam as it passes through the modulator. This is equivalent to a spread in the angle of incidence which, for satisfactory modulator operation, must be less than the value of $\Delta \theta_i$ given by (67). Whether or not this requirement can be met depends primarily on whether the modulator is operating in the far or near field of the optical source. The near field of an aperture of diameter D_b illuminated by a coherent plane wave extends a distance of order D_b^2/λ_0. In this region, which is about 1 m for a 1-mm diameter beam, there is little spread. At distances appreciably greater than this, the beam propagates according

to the simple theory of Fraunhofer diffraction, with the main lobe having an angular width[11]

$$\delta\theta \approx \frac{1.22\lambda_0}{D_b}.$$ (68)

Thus, in the worst possible case diffraction could result in an effective spread in the angle of incidence of order $1.22\lambda_0/D_b$, and hence limit the allowable value of the ratio of modulator length to beamwidth, according to (67), to

$$\frac{S}{D_d} \leq \frac{1}{2} \frac{\lambda_m}{\lambda_0} \frac{n_0}{\kappa'} \left(1 - \frac{n_0^2}{\kappa'}\right)^{-\frac{1}{2}}.$$ (69)

The aperture D_b is further restricted, in terms of the geometry of the modulator, by the requirement that the beam remain in the modulating medium after the first bounce. From Fig. 1 it is evident that this requires that

$$D_b \leq 2d \sin \theta_i,$$ (70)

where $d = h$ in the 0° modulator, and $d = w$ in the 90° modulator.

Equations (69) and (70) together require that

$$\frac{S}{d} \leq \frac{\lambda_m}{\lambda_0} \frac{n_0^2}{\kappa'^{\frac{3}{2}}} \frac{1}{(1 - n_0^2/\kappa')^{\frac{1}{2}}}.$$ (71)

5.4 Effect of Optical Losses

We shall distinguish between two kinds of optical loss. The first arises from scattering due to crystalline inhomogeneities and absorption within the bulk material, and is described by an absorption coefficient α_S. The second is the reflection loss at the mirrors, which are assumed to have a reflectivity R. The optical transmissivity can then be written as

$$I/I_0 = \exp\left[-\alpha_S S \csc \theta_i + (S/d) \cot \theta_i \ln R\right]$$ (72)

when $(S/d) \cot \theta_i$, the number of bounces, is large. The relative importance of reflection and absorption losses can be deduced by looking at the ratio, χ, of the quantities in the exponent in (72), i.e.

$$\chi = \frac{1}{\alpha_S d} \cos \theta_i \,|\ln R\,|.$$ (73)

For $\chi > 1$ reflection losses dominate, whereas for $\chi < 1$ the absorption losses are relatively more important.

If we require a certain minimum transmissivity, (72) and (73) fix an

upper limit to the linear dimensions of the modulator in terms of a loss parameter. Thus if $\chi > 1$ and we set $R^{(S/d) \cot \theta_i} = \frac{1}{2}$ in (72), corresponding to a reflection loss of 50 per cent, we obtain

$$\frac{S}{d} \leqq \frac{1}{1 - R} \frac{n_0/\sqrt{\kappa'}}{(1 - n_0^2/\kappa')^{\frac{1}{2}}} \ln 2 \qquad (74)$$

when $1 - R \ll 1$. By comparing (71) and (74) we can establish the relative importance of diffraction effects and reflection losses in determining the maximum allowed value of the ratio S/d.* We find that the reflection losses are the important factor whenever the reflectivity is small enough that

$$1 - R > 0.17 \frac{\lambda_0}{\lambda_m} \frac{\kappa'}{n_0} \qquad (75)$$

and diffraction effects dominate when the inequality is reversed. For $\lambda_0 = 1 \ \mu$, $\lambda_m = 3$ cm (X-band), the quantity $0.17 \ \lambda_0/\lambda_m$ is of order 5×10^{-6}. Since even with the best of dielectric multilayers[12] it is difficult to exceed $R = 0.999$ ($1 - R = 10^{-3}$), it is evident that reflection losses are the important factor in materials for which $\kappa'/n_0 \leqq 100$, as is normally the case. An important exception is a ferroelectric operated not far above its Curie point, for which the ratio κ'/n_0 can be very large.

If we now have the situation where $\chi < 1$ and we set $\exp (-\alpha_S S \csc \theta_i) = \frac{1}{2}$ in (72), corresponding to an absorption loss of 50 per cent, we obtain

$$S < \frac{1}{\alpha_S} \frac{n_0}{\sqrt{\kappa'}} \ln 2 \qquad (76)$$

with no restriction on d. This equation then sets an upper limit on the modulator length when reflection losses are negligible. For materials having good optical quality where α_S is typically of the order of $0.05 \ \text{cm}^{-1}$† such a case might correspond to having $R = 0.999$ and $d \approx 0.2$ mm.

5.5 *Effect of Lack of Parallelism in the Reflecting Surfaces*

A lack of parallelism in the mirrors causes the light and modulating waves to progressively slip out of synchronism, and so degrades the modulator performance. The effect may be estimated as follows.

---

* By comparing (71) with (74) we are in effect determining the minimum allowed value of the ratio S/d for a fixed sideband amplitude. Such a comparison is of significance since the sidebands carry the modulation information and since their amplitude will determine the signal-to-noise ratio at the detector.
† Accurate values of the absorption coefficients of highly transparent materials are not readily available and are liable to depend on the details of individual sample preparation. A value $\alpha_S \approx 0.05 \ \text{cm}^{-1}$ is a conservative estimate (cf. Ref. 13).

The time T taken for the optical signal to travel the length of the modulator is given by

$$T = \int_0^S \frac{dy}{v_y(y)} \tag{77}$$

where $v_y = (c/n_0) \sin \theta(y)$ is the y component of the optical phase velocity vector (Fig. 7). If the two reflecting surfaces are inclined at an angle ψ, the angle of incidence changes by ψ with each successive bounce. When the total number of bounces N is large, we may approximate $\theta(y)$ by the continuous function $\theta(y) = \theta_i + (N\psi/S)y$, so that the total optical transit time is given by

$$T = \frac{n_0}{c} \int_0^S \sin\left(\theta_i + \frac{N\psi}{S} y\right) dy$$

$$\approx \frac{n_0}{c} S \sin \theta_i + \frac{n_0}{c} \frac{N\psi S}{2} \cos \theta_i, \quad \text{for} \quad N\psi \ll 1. \tag{78}$$

If the optical and microwave phase velocities are matched at the modulator input $(\sin \theta_i = n_0/\sqrt{\kappa'})$, the first term entering in (78) is just the microwave transit time. Thus at the output of the modulator the optical and microwave signals have shifted, in terms of the microwave phase, by an amount

$$\Delta\varphi = \omega_m \Delta T = \frac{n_0 \omega_m}{c} \frac{S^2}{2d} \frac{\cos^2 \theta_i}{\sin \theta_i} \psi.$$

With the synchronism condition $\sin \theta_i = n_0/\sqrt{\kappa'}$, this becomes

$$\Delta\varphi = \frac{\pi \sqrt{\kappa'} S^2}{d\lambda_m} (1 - n_0^2/\kappa')\psi, \tag{79}$$
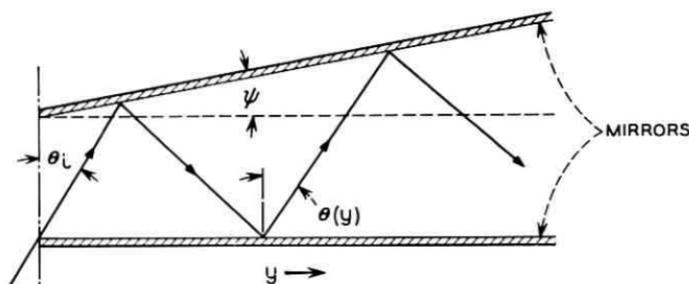

Fig. 7 — Model used in calculating the effect of nonparallelism of the reflecting surfaces.

where λ_m is the free-space modulating wavelength. This phase shift must be less than π if we are to avoid stripping modulation off the beam, since the effect of the modulating field will reverse as the phase shift exceeds π. This leads to a condition on the parallelism of the mirrors

$$\psi \; < \frac{d\lambda_m}{\sqrt{\kappa'}S^2} \frac{1}{(1 \; - \; n_0{}^2/\kappa')} . \tag{80}$$

5.6 *Effect of Strains*

A complete evaluation of the effect of strain on the modulator performance is very difficult; hence we shall not attempt a complete treatment here. Nonetheless, since strains can have a profound effect on the operation of the modulator, even an approximate treatment such as presented herein can be of value.

In the present discussion it is convenient to consider two broad classes of strain, uniform strain and random strain. The effect of a strain uniform throughout the crystal can be described in terms of a photoelastic matrix[14] which must then be combined with the electro-optic matrix in order to completely describe the optical properties of the crystal. The net effect is to add a static birefringence to that induced electrically.[14] This problem can in principle be solved analytically, although a large strain produces an additional complication since it will in general lower the crystal symmetry and hence change the form of the electro-optic tensor. Although cases of quasiuniform strain sometimes occur,* we shall consider here only the case of random strain, since this imposes a more fundamental limitation on device performance.

In a randomly strained crystal where the strain-induced variations occur over distances large compared with an optical wavelength, a ray in progressing through the crystal will see a continuously varying index ellipsoid. If there is little correlation between the shape and orientation of the index ellipsoid at two adjacent points, then a statistical model may be used to study the effect of the strains. Such a calculation is now being carried out and will be reported separately. We confine ourselves here to some general remarks, and a simplified analysis which is valid in the limit of very severe strain.

The general effect of random strain is to partially destroy the spatial coherence of a plane wave traversing the medium, without affecting its

---

* For example, cubic crystals drawn from a melt are sometimes observed to be uniaxial, with the growth axis as the optic axis.

temporal coherence.* The wave phase at the output of the modulator may then be written as the sum of a coherent and incoherent term, i.e.

$$\Phi(\mathbf{r},t) = \varphi(t) + \delta(\mathbf{r}),$$

where $\delta(\mathbf{r})$ is a random variable, and $\mathbf{r}$ is some suitably defined transverse position vector. Random strains will also mix the two normal modes of the unstrained crystal. We shall not consider the details of this mixing here, but simply assume that there exist at the output of the modulator two orthogonally polarized modes whose wave phases are given by

$$\Phi_\perp(\mathbf{r},t) = \varphi_\perp(t) + \delta_\perp(\mathbf{r}) \tag{81}$$

and

$$\Phi_\parallel(\mathbf{r},t) = \varphi_\parallel(t) + \delta_\parallel(\mathbf{r}) \tag{82}$$

where $\varphi_\perp$ and $\varphi_\parallel$ are the wave phases in the unstrained crystal as previously calculated, and $\delta_\perp(\mathbf{r})$ and $\delta_\parallel(\mathbf{r})$ vary more or less randomly across the surface of the beam. In this model the effect of strain on the modulator performance depends on the autocorrelation functions of $\delta_\perp$ and $\delta_\parallel$, and the cross-correlation between $\delta_\perp$ and $\delta_\parallel$, as the subsequent analysis will make clear. These correlations will depend on the state of strain of the crystal, and on the length of the optical path, becoming less as the path length increases.

The electric vectors at the output of the modulator can be written in the form

$$e_\perp(\mathbf{r},t) = E_\perp \exp\{j[\omega t + \varphi_\perp(t)] + j\delta_\perp(\mathbf{r})\} \tag{83}$$

and

$$e_\parallel(\mathbf{r},t) = E_\parallel \exp\{j[\omega t + \varphi_\parallel(t)] + j\delta_\parallel(\mathbf{r})\} \tag{84}$$

where $E_\perp$ and $E_\parallel$ are the wave amplitudes. From these equations we can show that although strains do not affect the basic phase modulation process, they do affect the over-all modulator-detector system performance. Details depend on the system used, but the following general statement can be made: Any optical communication system in which the modulation or detection process is based upon the heterodyning or homodyning of two optical signals requires that the two signals have relative spatial coherence. Any process that tends to destroy that coherence, such as random strain, will degrade the system performance markedly. It is

---

* The fact that in general strain will also degrade the collimation of the beam is of importance in some applications. See Appendix B.3.

therefore necessary to be able to evaluate which system will allow us to recover the modulating intelligence with maximum certainty. All modulation-detection schemes presently under consideration fall into three classes

(i) amplitude modulators followed by quantum counters,

(ii) phase or amplitude modulators followed by optical heterodyne or homodyne detectors, and

(iii) phase modulators followed by optical frequency discriminators and then by quantum counters.

Starting with (83) and (84), we analyze these three classes in Appendix B. The results are summarized below.

5.6.1 *Amplitude Modulators followed by Quantum Counters*

As discussed in Section IV, the amplitude modulator consists of a system in which two phase-modulated waves of the form given by (83) and (84) are combined in an analyzer. As shown in Appendix B, the time-varying intensity transmitted by the analyzer is proportional to an integral of the form [see (118)]

$$\int_s \frac{\sin}{\cos} \left\{ \delta_\perp(\mathbf{r}) - \delta_\parallel(\mathbf{r}) \right\} da \tag{85}$$

the integration being taken over the cross section of the beam. In the case of severe random strain, when $\delta_\perp$ and $\delta_\parallel$ can *independently* take on all values (i.e., no cross-correlation between $\delta_\perp$ and $\delta_\parallel$), the integral (85) vanishes and there is no amplitude modulation. More moderate strain would result in a nonzero depth of modulation which would, however, be less than that attainable in a strain-free crystal.

5.6.2 *Heterodyne or Homodyne Operation*

In heterodyne or homodyne operation the modulated signal is mixed with a second locally generated optical signal to obtain a microwave intermediate frequency which is then detected conventionally. Assuming a square-law detector and a strong local oscillator having both spatial and temporal coherence, the output signal at the difference frequency is proportional to [Appendix B, (121)]

$$\int_s \frac{\cos}{\sin} \left\{ \delta_{\perp,\,\parallel}(\mathbf{r}) \right\} da \tag{86}$$

which vanishes in the limit of very bad strain where $\delta_\perp$ and $\delta_\parallel$ can take on all values.

Thus in this mode of operation, too, strains result in a reduced detector output. The effect here is likely to be more severe than in the case of the simple amplitude modulator-quantum counter system, since for a given degree of strain the autocorrelation of $\delta_\perp$ or $\delta_\parallel$ is expected to be less than the cross-correlation of $\delta_\perp$ and $\delta_\parallel$. Thus, in a given crystal the integral (85) is liable to have a larger value than the integral (86).

5.6.3 *Phase Modulators Followed by Frequency Discriminators*

In the third system under consideration phase modulation is directly converted to amplitude modulation in some frequency-selective optical circuit[15,16] of which a Fabry-Perot etalon is perhaps the simplest example. Spatial coherence per se plays no essential role here. Analytically, as is shown in Appendix B, the integral analogous to the ones given by (85) and (86) which arises in this case has the form [see 124]

$$\int_s |\exp [j\delta_{\perp,\parallel}(\mathbf{r})]|^2 \, da. \tag{87}$$

This of course has a constant value equal to the beam area and can therefore not influence the system response. However, strain can still adversely effect this system to the extent that the discriminator performance is degraded by loss of collimation in the transmitted beam.

Thus strain will *always* be a problem in optical modulators. In fact the preceding considerations show that the presence of severe random strain will result in no useful output from those modulation-detection schemes which rely on the *interference* of two beams. In this case only the phase modulator-frequency discriminator system has any real chance of success.

VI. OPTIMUM DESIGN PROCEDURE

6.1 *Choice of Modulator Dimensions*

For the purpose of discussing the optimization of modulator dimensions as well as the choice of operating temperature, a useful figure of merit is the product of the modulation sensitivity and the optical transmissivity, which for *all* modulators can be written as

$$\mathfrak{F} \equiv \left| \frac{M}{\sqrt{P}} \right| \frac{I}{I_0} = \frac{2\pi}{\lambda_0} n_0^3 S \sqrt{\frac{2\eta}{\sqrt{\kappa'} dl}} \begin{Bmatrix} \frac{1}{2} r_{41} \\ \text{or} \\ \rho E_0 \end{Bmatrix} p \left(\frac{n_0}{\sqrt{\kappa'}} \right) f \left(\frac{\alpha_m S}{2} \right)$$
$$\times \exp \left[-\frac{\sqrt{\kappa'}}{n_0} S \left(\alpha_s - \frac{1}{d} \ln R \sqrt{1 - \frac{n_0^2}{\kappa'}} \right) \right]. \tag{88}$$

Here d is the distance between the reflecting surfaces (either h or w, depending on whether the $0°$ mode or the $90°$ mode is used), and l is the width of the reflecting surfaces; recall that S is the modulator length. The remaining functions $f(\alpha_m S/2)$ and $p(n_0/\sqrt{\kappa'})$ have been defined above in (12) and Tables I and II respectively. Equation (88) applies when we have velocity matching.

The optimum modulator dimensions can now be found by maximizing the figure of merit. It is evident that the dimension l should be made as small as possible. If the beam diameter D_b is specified, this then determines the minimum value of l, $l_{\min} = D_b$. The mirror spacing d may next be chosen so as to maximize (88), the appropriate value being

$$d_{\mathrm{opt}} = -2S \ln R \, \frac{\sqrt{\kappa'}}{n_0} \left(1 - \frac{n_0^2}{\kappa'} \right)^{\frac{1}{2}}. \tag{89}$$

The ratio S/d obtained from this equation corresponds to a reflection loss of roughly 50 per cent. It is worth pointing out that for practically attainable reflectivities, this optimum value of S/d is less than the maximum value set by diffraction effects [see (77)], and so represents a reasonable design value.

When the optimum value of d given by (89) is substituted back into (88), there results an equation for $\mathfrak{F}$ which depends only on S; S may then be chosen so as to maximize this figure of merit. The optimum value, S_{opt}, is defined by the transcendental equation

$$\frac{1}{\exp \alpha_m S_{\mathrm{opt}} - 1} - \frac{1}{2\alpha_m S_{\mathrm{opt}}} = \frac{\sqrt{\kappa'}}{n_0} \frac{\alpha_S}{\alpha_m}. \tag{90}$$

In the absence of optical absorption ($\alpha_S = 0$), the optimum length is given by $\alpha_m S_{\mathrm{opt}} = 1.25$. As α_S increases, the optimum length decreases, as shown in Fig. 8. When the inequality $(n_0/\sqrt{\kappa'})(\alpha_m/\alpha_S) \ll 1$ is satisfied, we find that the optimum length is determined solely by the optical attenuation, and is given by

$$S_{\mathrm{opt}} = \frac{n_0}{2\sqrt{\kappa'}\alpha_S}. \tag{91}$$

The dimensions of the modulator are thus completely determined: l by the beamwidth, d by the optical reflection losses and S by the microwave and optical attenuation constants. For materials having sufficiently low microwave and optical losses, the optimum length given by (90) may be greater than any practically realizable crystal length. Under these circumstances one should make S as large as possible, and then choose d to satisfy (89).
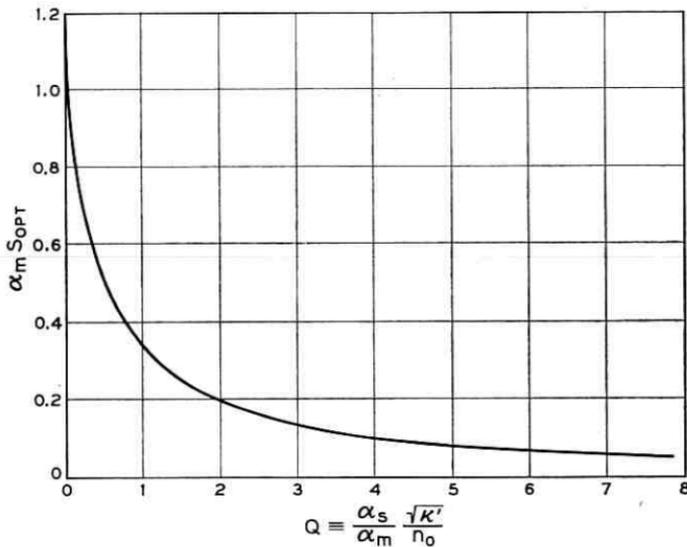
Fig. 8 — Dependence of optimum modulator length on optical and microwave loss parameters.

6.2 Choice of Operating Temperature

The dielectric and electro-optic constants of useful modulating materials are strongly temperature-dependent. In ferroelectrics, for example, both the electro-optic coefficients and the microwave losses increase dramatically near the Curie point. We should thus like to establish whether there exists a preferred operating temperature which maximizes the figure of merit [see (88)] independently of the modulator dimensions.

We begin by assuming that the modulator dimensions have been optimally chosen according to the results of the preceding section. With this choice of dimensions, (88) becomes

$$
\mathfrak{F}_{opt} = \frac{2\pi}{\lambda_0} n_0{}^3 \sqrt{ -\frac{\eta}{\sqrt{\kappa'} D_b \ln R} \left\{ \begin{matrix} \frac{1}{2} r_{41} \\ \text{or} \\ \rho E_0 \end{matrix} \right\} p\left(\frac{n_0}{\sqrt{\kappa'}}\right) \left[\frac{n_0/\sqrt{\kappa'}}{\alpha_m (1 - n_0{}^2/\kappa')^{\frac{1}{2}}} \right] }
$$

$$
\times\, e^{-\frac{1}{2}} \left[\sqrt{u} e^{-Qu} f\left(\frac{u}{2}\right) \right] \tag{92}
$$

where $u = \alpha_m S_{opt}$ is found from (90) and where we introduce the notation $(\alpha_s/\alpha_m)(\sqrt{\kappa'}/n_0) \equiv Q$. The function $g(u) = \sqrt{u} e^{-Qu} f(u/2)$ is, through (90), an implicit function of the independent variable Q. The function may be approximated as

$$g(Q) \approx 0.214 \frac{\left[1 - \exp\left(-\frac{3}{2}\sqrt{Q} \right) \right]}{\sqrt{Q}}.$$

For a TEM-like wave, the microwave attenuation constant is given in terms of the dielectric constant and loss tangent of the medium by

$$\alpha_m = \frac{\pi}{\lambda_m} \sqrt{\kappa'} \tan \delta. \tag{93}$$

The figure of merit can now be approximated solely in terms of material parameters as

$$\mathfrak{F}_{opt} = 15.8 \frac{n_0^4}{\lambda_0} \frac{1}{\sqrt{-\alpha_s D_b \ln R}} \frac{1}{\kappa'^{\frac{3}{2}}} \begin{Bmatrix} \frac{1}{2}r_{41} \\ \text{or} \\ \rho E_0 \end{Bmatrix} p\left(\frac{n_0}{\sqrt{\kappa'}} \right)$$

$$\times \left\{ \frac{1 - \exp\left[-\frac{3}{2}\left(\frac{\alpha_s \lambda_m}{\pi n_0 \tan \delta} \right)^{\frac{1}{2}} \right]}{(1 - n_0^2/\kappa')^{\frac{1}{4}}} \right\} \tag{94}$$

where all the numerical constants, including the impedance of free space, $\eta = 377$ ohms, have been combined into a single factor.

Once the temperature dependence of the electro-optic coefficient, dielectric constant and loss tangent are known,* (94) can be used, at least in principle, to select an optimum operating temperature.

Further analytical progress requires some assumption of the temperature dependence of these coefficients as well as a choice of the mode of operation. To be definite we select the mode for which $p(n_0/\sqrt{\kappa'}) = \sqrt{\kappa'}/n_0$. As may be seen from Tables I and II, this is the most favorable value that can be realized and applies, for example, to the linear, 0° phase modulator.

The temperature coefficient of the linear electro-optic coefficient is expected to follow that of the dielectric constant.[17] The reason for this is as follows: The electro-optically induced birefringence arises fundamentally from a change in the polarization, rather than the applied field, and in fact Pockel's[18] original formulation of the electro-optic effect was in terms of constants e_{ij} relating the coefficients of the index ellipsoid to the induced polarization. The r_{ij} coefficients used here are related to Pockel's e_{ij} by

$$r_{ij} = \epsilon_0(\kappa' - 1)e_{ij}.$$

---

* The index of refraction, n_0, to the extent to which it is determined by the electronic polarizability, should be almost independent of temperature.

Thus for materials having a high dielectric constant we expect r_{ij} to be approximately proportional to κ', and to the extent that the coefficients e_{ij} are temperature independent,[17] to have the same temperature dependence as κ'.

A similar argument in the case of the quadratic effect leads to the conclusion that ρ should be proportional to κ'^2. Here, however, the situation is complicated by the fact that the bias field E_0, which for optimum operation should be as large as the breakdown field strength of the material permits, will be temperature-dependent whenever this breakdown field strength is temperature-dependent. In many solids we expect dielectric breakdown to occur at some definite value of internal field. In materials of high dielectric constant the internal field is determined by the polarization, so that the breakdown field strength, and hence the allowed dc bias, will be inversely proportional to κ'. The net result is that the product of ρE_0 may be taken proportional to the first power of κ', and the quadratic and linear effects become formally identical.

The preceding assumptions allow us to write the figure of merit for a given material wholly in terms of its dc and optical dielectric parameters, so that $\mathfrak{F}$ becomes proportional to

$$\frac{\kappa'^3}{(1 - n_0^2/\kappa')^{\frac{1}{4}}} \left[1 - \exp\left(-\frac{3\lambda_m \alpha_S}{2\pi n_0 \tan \delta} \right)^{\frac{1}{2}} \right]. \tag{95}$$

We must now distinguish between low and high microwave losses, the criterion being the magnitude of the ratio $3\lambda_m \alpha_S / 2\pi n_0 \tan \delta$. For sufficiently low microwave losses, the exponential factor in (95) will disappear, so that the figure of merit is proportional simply to $\kappa'^{3/2}/(1 - n_0^2/\kappa')^{1/4}$. Since the ratio n_0^2/κ' will always be substantially less than unity, the optimum operating temperature is that which maximizes κ'. For ferroelectric materials this will mean lowering the temperature to the vicinity of the Curie point, T_c. Of course the losses will normally increase as the temperature is lowered, so that at some point the approximation of low microwave loss breaks down. For sufficiently high loss we may expand the exponential term in (95), obtaining a figure of merit proportional to

$$\frac{1}{(1 - n_0^2/\kappa')^{\frac{1}{4}}} \sqrt{\frac{\kappa'^{\frac{3}{2}}}{\tan \delta}}. \tag{96}$$

For ferroelectrics such as KDP[19,20] or SrTiO$_3$[21,22] in which both κ' and $\tan \delta$ are dominated by a $(T - T_c)^{-1}$ dependence, T_c being the Curie temperature, this result says that operation near the Curie point is

favored (although not strongly) even in the presence of substantial microwave loss.

6.3 *Bandwidth Limitations: Design of Optical Coupler*

The question of the bandwidth of the zigzag modulator has two aspects — microwave bandwidth and optical bandwidth. Since both the optical and modulating signals propagate as TEM-like waves, it might at first be thought that the bandwidth would be limited only by the dispersion in the material parameters of the modulating medium. For most materials this would imply a microwave bandwidth as large as 10 or 20 gc, and since the fractional optical bandwidth required to accommodate even microwave frequency sidebands is very small (of order 10^{-5}), optical dispersion should never be a problem.

A more discriminating analysis of the question of bandwidth reveals, however, that the physical dimensions of the modulator limit the realizable bandwidth, especially when these dimensions are optimized so as to obtain the ultimate in low-power performance. In an optimum design the length S_{opt} and width (or height) d_{opt} are uniquely determined by the electrical properties of the modulating medium both at microwave and optical frequencies as well as the loss properties of the medium [see (89) through (91) and the attendant discussion]. For highly transparent low-loss microwave materials, d_{opt} computed from (89) is nominally in the order of a few tenths of a millimeter and S_{opt} is usually several centimeters (see Table IV in Section VII for typical results). The one dimension which remains unspecified is the beam diameter D_b. As remarked above in Section 6.1, D_b should be chosen as small as possible and yet remain within the diffraction limits of the device over the length S_{opt}. Accordingly, we find that, based upon certain operational practicalities and diffraction limits, D_b must normally be restricted to about a millimeter. Thus, as the aperture of the device is typically an order of magnitude larger than the spacing between mirrors, it is impossible to satisfy (70); i.e., the beam bounces back out of the modulator, and hence cannot be confined to the active area of the device by means of a simple aperture arrangement. In such cases the beam can be confined to the interior of the modulator only if we include an appropriate optical coupler at the input (and output) of the modulator.

It is shown in Appendix C that a suitable coupler which allows all of the incident optical energy to be transferred into the active medium of the modulator is a Fabry-Perot resonator of the form shown either in Fig. 10(a) or Fig. 11. The physical operation of the coupler is easily understood when one observes that because of its finite length energy must

leak off to the right into the interior of the modulator. Then by properly adjusting the spacing between reflectors as well as the reflectance of the lower reflector it is possible to completely cancel the reflected wave, and have the energy trapped in the coupler section walk off into the modulator. Expressions for the required spacing and reflectance are derived in Appendix C. Here we simply point out that such a coupler can be realized in a typical case with reflectivities in the neighborhood of 99 per cent, which is easily attained, and with dimensions L_c and l_c of the same order as D_b and d_{opt}, respectively.

We are now in a position to examine the bandwidth limitations imposed by the coupler and modulator dimensions. The bandwidth is generally determined by three factors. In order of importance these are

   (i) the spacing l_c and loaded Q of the coupler,
   (ii) the size of D_b compared with the microwave modulating wavelength, and
   (iii) the microwave impedance of the structure.

The last factor, microwave impedance, is of practical significance because the impedance of the modulator will inevitably be low, and it becomes progressively more difficult to obtain a broadband match into a transmission line as its impedance is lowered. However, a number of standard microwave techniques are available to overcome this difficulty.

The second factor, the beam diameter, affects the bandwidth in the following way. In order that the microwave modulation be detectable in a direct simple way it is necessary that the modulation frequency phase be approximately constant across a given section of the beam. Hence it is necessary that the projection of the beam cross section along the axis of the structure be appreciably less than half the microwave wavelength in the medium. Mathematically

$$D_b \ll \frac{\lambda_m}{2 \cos \theta_i \sqrt{\kappa'}} \tag{97}$$

where λ_m is the free-space wavelength.

Finally, with reference to the first factor, the coupler spacing and loaded Q, we note that in order to avoid microwave impedance discontinuities in the low-impedance modulator line it is mandatory that l_c and d be very nearly equal. If the spacing d is then kept small so as to obtain the lowest possible modulating power, the input coupler will limit the bandwidth in the following way. The Fabry-Perot coupler will be resonant at a series of frequencies spaced by

$$\Delta f = \frac{c}{2n_0 l_c \cos \theta_i}. \tag{98}$$

The base bandwidth δf of the coupler will be determined by the Q of an individual mode, and can be shown from Appendix C to be

$$\delta f = \frac{1}{\pi} \frac{c}{D_b} \frac{\sin \theta_i}{n_0} \begin{cases} \sqrt{1 - n_0^2 \sin^2 \theta_i} & \text{for} \quad n_0^2 < \sqrt{\kappa'} \\ \cos \theta_i & \text{for} \quad n_0^2 > \sqrt{\kappa'}. \end{cases} \quad (99)$$

Since this bandwidth is available in each of the Fabry-Perot modes, microwave subcarrier operation at frequencies above this baseband is allowed[23] provided that the subcarrier frequency (or optical sideband) is some multiple of the mode spacings given by (98). The problem, of course, is that the mode spacings become very large when the coupler reflector spacing l_c is small, and since l_c must be comparable with the modulator thickness d, it follows that the lowest allowed microwave sub-carrier will be at much too high a frequency when d is adjusted to its optimum value of a fraction of a millimeter (e.g., $d = 0.2$ mm corresponds to $\Delta f \approx 300$ gc). It is evident then that the minimum power configuration is restricted to baseband operation with a bandwidth determined either by (97) or (99), whichever is smaller. For $D_b = 1$ mm, (97) gives $f_m \lesssim 10$ gc when the modulating medium is a ferroelectric such as strontium titanate, while (99) gives $\delta f \approx 5$ gc.

Greater bandwidth can be obtained, at the expense of increased modulation power, by making the thickness d considerably greater than optimum. If d is made sufficiently large (e.g., 3.6 mm in the case of the strontium titanate modulator), a simple input aperture can be used in place of the Fabry-Perot coupler.* This removes the frequency limitations set by (98) and (99), leaving only that imposed by the finite beam width (97) and material dispersion. The frequency limitation imposed by the beam width is quite fundamental. In fact, when the projection of the beam diameter along the axis of the modulator is exactly equal to a full microwave wavelength in the material there results no net modulation. We therefore define a cutoff frequency for the modulator as

$$f_c = \frac{c}{D_b \cos \theta_i \sqrt{\kappa'}}. \quad (100)$$

In summary, then, we find that in order to fully exploit the low modulation power requirements of the zigzag modulator it is necessary to restrict the frequency of operation for most substances to a few gc at baseband,† and to use a relatively sophisticated coupler to get the optical

---

* If $n_0^2 > \sqrt{\kappa'}$ it will be necessary to use a more complex aperture analogous to the one shown in Fig. 11.

† For CuCl the operating frequency can be as high as about 20 gc (see Table IV in Section VII).

energy into the device. By relaxing the power requirements the upper frequency limit can be extended above 10 gc in a relatively simple structure. Examples of both kinds of design are given in the following section.

VII. DISCUSSION AND CONCLUSIONS

We have analyzed in detail the behavior of zigzag phase and amplitude light modulators (Sections III and IV) as well as the basic underlying theory for this class of modulators (Section II). Our treatment differs from the work of other investigators[1-3] along these lines in a number of crucial ways. We have included microwave losses in the initial calculation of the optical retardation and have considered the 90° mode of operation in addition to the originally proposed 0° mode for several different classes of electro-optic media. The analysis has been detailed enough and has treated enough of the important practical factors which limit the operation of the device (Section V), so that a realistic optimum design procedure can be presented (Section VI). In essence, this procedure requires that the dimensions of the modulator be so chosen that all of the incident microwave power be confined to as small a volume of the electro-optic material as the microwave, optical, and reflection losses will permit. As a result, large modulating field strengths are obtained with moderate amounts of power.

Based upon our treatment we find that there are three main aspects to broadband modulator design. The first is the velocity synchronism condition; here we find that the basic phase velocity matching criterion is unaffected by microwave loss. The second is the effect of modulator dimensions and aperture on bandwidth. These problems have been discussed in Section 6.3, where it was shown that the aperture limits the highest modulating frequency in a fundamental way [see (100)] and that the other dimensions in effect determine the requisite coupling scheme (either Fabry-Perot or simple aperture) and thence the bandwidth. The final criterion of significance is the strain properties of the modulator crystal (Section 5.6). Crystalline strain is a serious problem in all forms of electro-optic amplitude modulators, but is particularly deleterious in the traveling-wave modulator because of the long optical path.

We are now in a position to present some actual designs of zigzag phase and amplitude light modulators as well as the performance to be expected of these devices. Before doing so, however, we first summarize in Table III the pertinent microwave and optical material parameters, to the extent that they are known, for a number of electro-optic modulating crystals which are currently available. With regard to the data presented in Table III, much of which is tentative, two points should be noted.

TABLE III — MATERIAL PARAMETERS FOR SOME ELECTRO-OPTIC CRYSTALS

Material	T_c	T	$(p_1 - p_2)$*	r_{41}†	r_{63}†	n_0	κ'	$\tan \delta$	Remarks
BaTiO$_3$	393°K	393°K 403°K	23×10^{-12} 6.5×10^{-12}			2.4 2.4	6000 4500	0.5 0.4	Electro-optic data apply at $\lambda_0 = 5021$ Å and $f \leqq 200$ kc (Ref. 24). Microwave data apply at 24 gc (Ref. 25).
SrTiO$_3$	33°K	77°K 293°K	0.31×10^{-12} 0.194×10^{-13}			2.409 2.409	1500 300	1.25×10^{-3} 1.5×10^{-3}	Electro-optic data apply at $\lambda_0 = 6328$ Å and $f < 1$ kc. Data obtained from Ref. 26. Value for n_0 applies at $\lambda = 5891$ Å. Microwave data apply at 20 gc (Refs. 21, 27, 28).
KTaO$_3$	≈ 1°K	4°K 77°K	4.3×10^{-12} 0.1×10^{-12}			2.5 2.5	4400 800	0.72×10^{-4} 0.2×10^{-3}	Electro-optic data apply at $\lambda = 6328$ Å and $f < 1$ kc (Ref. 29). Microwave data apply for frequencies below 20 gc (Ref. 29).
CuCl		293°K		6×10^{-10}		1.93	8.3	1×10^{-3}	Electro-optic data measured at dc (Ref. 30). Microwave data measured at 5 gc (Ref. 31).
ZnS		293°K		2.4×10^{-10} 1.2×10^{-10}		2.368	10.25	2×10^{-3}	Electro-optic data measured at dc (2.4×10^{-10}) and 3 gc (1.2×10^{-10}). Microwave data measured at 5 gc.
KDP	120°K	173°K 293°K			6.8×10^{-10}	1.468 1.468	60 20.2	17.5×10^{-3} 7.5×10^{-3}	Electro-optic data apply at high frequencies, $f_0 > 100$ kc, and for $4000 < \lambda_0 < 7000$ Å (Ref. 32). Microwave data apply at 9.2 gc (Ref. 33).
ADP	148°K	173°K 293°K			5.5×10^{-10}	1.479 1.479	16 14	6.5×10^{-3} 6×10^{-3}	T_c refers to antiferroelectric phase transition (Ref. 34). Electro-optic data apply at high frequencies, $f > 100$ kc, and for $4000 < \lambda_0 < 7000$ Å (Ref. 32). Microwave data apply at 9.2 gc (Ref. 33).

* Expressed in units of cm²/V².
† Expressed in units of cm/V.

The first is that the microwave properties for the different materials listed have not all been measured at the same frequency. For simplicity we shall assume that there is negligible dispersion in the microwave parameters. The same assumption is made for the optical data. The second point to be noted is that the data available for the electro-optic coefficients apply in most cases at very low frequencies. Hence the values for the electro-optic coefficients represent the unclamped values and consequently overestimate in many cases the desired microwave or clamped values for these coefficients. Thus this table should be viewed simply as a guide to the material parameters. From the data given in Table III we select three materials, CuCl, $SrTiO_3$ and KDP, as being the best in their respective crystalline classes. Each of these materials combines good microwave properties (low loss with high intrinsic impedance) with a large electro-optic coefficient. With reference to $SrTiO_3$ it should be noted that the quadratic effect exhibited by this material is anomalously high, as it is with $BaTiO_3$ and $KTaO_3$, the effect being 10^5–10^7 times higher than in the best glasses. When one couples this together with the fact that the "effective linear coefficient" for quadratic materials is ρE_0, one immediately sees from the data listed in Table III that for $E_0 > 1000$ v/cm, quadratic materials are potentially as good as if not better than linear materials.

It is instructive to carry out two designs. One, based on the optimum design procedures of Section 6.1, results in a modulator requiring an absolute minimum of modulating power. However this design requires the use of an input optical coupler and is inherently narrow-band. In another, more practical design, the mirror spacing d is chosen to allow the use of a simple aperture. In this design we obtain increased bandwidth at the expense of higher modulating power. Tables IV and V summarize the design parameters, expected performance, and several of the practical limitations for each of these designs. Note that we do not include an amplitude modulator design for KDP. The reason, of course, is the difficulty encountered in trying to compensate for the static retardation in such a material (See Section 3.3.)

In order to carry out the calculations several design parameters were chosen independently as a matter of convenience. Thus the optical wavelength was taken as the visible red line of the He-Ne gas laser at 6328 Å, and in the absence of experimental data the optical absorption coefficient α_S was assumed to have the same value as that found in good optical glasses, viz., $\alpha_S \approx 0.05$ cm^{-1}. The multilayer reflectivity was taken as 0.999, about as high as is practically achievable.[12] A reflectivity this high is essential if the ultimate in low-power modulation is to be obtained in

the optimum design, but the requirements can be relaxed considerably in the practical design. For example, a decrease in reflectivity to 0.982 increases the optical insertion loss of the practical $SrTiO_3$ modulator by only 1.7 db.

In the absence of any really complete temperature data we have taken room temperature, 300°K, as an operating temperature. Although it is true that for the ferroelectrics, $SrTiO_3$ in particular, the temperature analysis of Section VI shows that the modulating power is reduced somewhat as we approach the Curie point, the concomitant increase in κ' reduces θ_i below the practical minimum and at the same time reduces the available bandwidth by decreasing the microwave wavelength in the material [see (97) and accompanying discussion]. Thus 300°K is in fact close to the practical optimum operating temperature for $SrTiO_3$.

The operating frequencies given in Tables IV and V have been arrived at by the following considerations. A practically realizable aperture which remains within diffraction limits is nominally 1 mm. The *optimum* design therefore requires the use of a coupler with restricted bandwidth. The data of Table IV are thus calculated for a modulating frequency of 5 gc, which is within the passband of all modulators. In Table V the operating frequency has been chosen as 10 gc and is in all cases less than the cutoff frequency. What should be noted here is that by ignoring the bandwidth problem, we obtain in the optimum design shown in Table IV the irreducible minimum microwave modulating power. In the case of the practical design listed in Table V the spacing between mirrors d has been increased to allow the use of a simple aperture. The price we pay, of course, is in increased modulating power. We should also make special note of the fact that from a practical viewpoint the impedance level of the design given in Table V is much more favorable than for the one given in Table IV. The restriction on the maximum attainable modulating frequency in the practical design is the aperture size [see (97)].

The column headed "configuration" in Tables IV and V deserves some comment. For each material considered, the configuration chosen was the one for which the function $p(n_0/\sqrt{\kappa'})$ had its maximum value. In the case of $SrTiO_3$ there are three possible configurations for the phase modulator. Table I shows that the 0° phase modulator (PM) in the $\perp$-mode and the 90° PM in either the $\perp$ or $\parallel$-modes are all equally suitable choices. We should bear in mind at this point that for $SrTiO_3$ (as well as $BaTiO_3$ and $KTaO_3$), it is only the difference in the electro-optic coefficients $|\rho_1 - \rho_2|$ that is known (see Table III) and not ρ_1 or ρ_2 alone. However, since in an amorphous material ρ_2 is identically zero (Appendix A) we expect that in a cubic material $|\rho_2|$ will still be

TABLE IV — OPTIMUM MODULATOR PERFORMANCE FOR SEVERAL ELECTRO-OPTIC MATERIALS. MODULATING FREQUENCY, 5 GC; OPTICAL WAVELENGTH, 6328 Å; TEMPERATURE, 300°K

Material	Configuration	θ_i	$|\Delta\theta_i|$	ψ_{max}	D_b	d_{opt}	S_{opt}	δf	IL*	$P(M=\frac{1}{2})$	Remarks
CuCl	0°-PM; ⊥-mode	42.1°	9.5°	0.072°	1 mm	0.15 mm	6.7 cm	24.7 gc	4.3 db	0.315 w	S_{opt} fixed by optical losses. Dimensions are within limits set by diffraction.
	0°-AM	42.1°	9.5°	0.072°	1 mm	0.15 mm	6.7 cm	24.7 gc	4.3 db	0.13 w	
SrTiO$_3$	0°-PM; ⊥-mode	8°	1.18°	0.2°	1 mm	0.20 mm	1.4 cm	5.2 gc	4.3 db	0.17 w	S_{opt} fixed by optical losses. Dimensions are within limits set by diffraction. Results apply for $E_0 =$ 30 kv/cm. Computations for phase modulator treat material as amorphous.
	90°-AM†	8°	1.18°	0.2°	1 mm	0.20 mm	1.4 cm	5.2 gc	4.3 db	0.17 w	
KDP	0°-PM; ⊥-mode	19.1°	4.84°	0.15°	1 mm	0.19 mm	3.3 cm	18.7 gc	4.3 db	2.8 w	S_{opt} fixed by optical losses. Dimensions are within limits set by diffraction.

* The insertion loss does not take reflection losses into account. These can be eliminated in any one of a number of standard ways.
† In the case of the amplitude modulator, a bias field of 30 kv/cm results in operating at a mode number of $m \approx 600$.

TABLE V — PRACTICAL MODULATOR PERFORMANCE FOR SEVERAL ELECTRO-OPTIC MATERIALS. MODULATING FREQUENCY, 10 GC; OPTICAL WAVELENGTH, 6328 Å; TEMPERATURE, 300°K

Material	Configuration	θ_i	$|\Delta\theta_i|$	ψ_{max}	D_b	d	S	f_c	IL^*	$P(M=\frac{1}{2})$	Remarks
CuCl	0°-PM; ⊥-mode 0°-AM	42.1° 42.1°	4.8° 4.8°	0.18° 0.18°	1 mm 1 mm	0.75 mm 0.75 mm	6.7 cm 6.7 cm	140 gc 140 gc	2.6 db 2.6 db	1.6 w 0.66 w	S fixed by optical losses. Dimensions are within limits set by diffraction.
SrTiO₃	0°-PM; ⊥-mode 90°-AM†	8° 8°	0.59° 0.59°	1.8° 1.8°	1 mm 1 mm	3.6 mm 3.6 mm	1.4 cm 1.4 cm	17.5 gc 17.5 gc	2.6 db 2.6 db	3.1 w 3.1 w	S fixed by optical losses. Dimensions are within limits set by diffraction. Results apply for $E_0 = 30$ kv/cm. Computations for phase modulator treat material as amorphous.
KDP	0°-PM; ⊥-mode	19.1°	2.42°	0.61°	1 mm	1.53 mm	3.3 cm	70.7 gc	2.6 db	23 w	S fixed by optical losses. Dimensions are within limits set by diffraction.

* The insertion loss does not take reflection losses into account. These can be eliminated in any one of a number of standard ways.

† In the case of the amplitude modulator a bias field of 30 kv/cm results in operating at a mode number $m \approx 600$.

substantially less than $|\rho_1|$, so that we may take the measured value of $|\rho_1 - \rho_2|$ as an approximate measure of $|\rho_1|$ alone. This has been done wherever necessary.

The remaining parameters listed in Tables IV and V require little or no further comment. The angle of incidence θ_i necessary to achieve velocity synchronism is given together with the deviation in θ_i, $|\Delta\theta_i|$, which results in a 3-db degradation in modulation sensitivity. The angle ψ_{max} is the maximum permissible angular deviation in the parallelicity of the mirrors if the performance is to be degraded by less than 3 db. The modulator dimensions are contained in D_b, d_{opt} or d, and S_{opt} or S depending upon which design is being considered. The column headed IL in the tables refers to the optical insertion loss suffered in traversing the modulator. Finally, the quantity $P(M = \frac{1}{2})$ is the microwave modulating power required to achieve a modulation index of 0.5.

It is worth mentioning at this point that Tables IV and V implicitly contain several idealizations. The first is that the crystals are strain-free. For KDP this assumption is approximately justified, but for CuCl and SrTiO$_3$ it is far from reality and, as noted previously, could seriously degrade the performance of the amplitude modulator. The second idealization comes from the assumption that the modulating field is uniform over the entire cross section of the crystal. In any practical modulator this will not be strictly true. However, to the extent that fringing fields can be neglected in a parallel plate line completely filled with electro-optic material, this assumption is realizable (see Section 5.1). In consequence, the calculations made for the required modulating power should be reasonably accurate. The primary factor which would cause an increase in modulating power is a deviation from the correct angle of incidence. However, the tolerances in θ_i given in Tables IV and V are sufficiently large that maintaining the required angle should present no problem. Finally we should point out that we have assumed that a dc bias as large as 30 kv/cm can be applied to the SrTiO$_3$ sample without dielectric or interfacial breakdown, and that the electro-optic effect does not saturate at these high fields. It is clear that a high dc bias field is desirable with the SrTiO$_3$ modulator, since this increases the effective linear electro-optic coefficient ρE_0.

The results given in Tables IV and V demonstrate that broadband microwave modulation of light can be effected with very little modulating power. In a practically realizable design, we find that a 1 mm $\times$ 3.6 mm $\times$ 1.4 cm rectangular slab of SrTiO$_3$ biased with 30 kv/cm should produce 50 per cent linear modulation of light with less than 5 watts of power at any modulating frequency from dc to 10 gc. A CuCl slab of

dimensions 1 mm $\times$ 0.75 mm $\times$ 6.7 cm should require still less power ($\approx$1–2 w) without the need of any dc bias. At the expense of reduced bandwidth and increased optical complexity, this power can be reduced to less than a watt by a correct choice of the modulator dimensions (optimum deisgn). In the zigzag configuration, CuCl appears to be slightly better than $SrTiO_3$, in the sense of achieving moderately broadband performance with low modulating power. Cuprous chloride has the advantage of extreme bandwidth, of not requiring any bias, and of not needing reflecting mirrors since θ_i exceeds the critical angle for total reflection. Compared with cavity-type modulators using KDP[35] or ADP[36] the CuCl zigzag modulator typically requires three orders of magnitude less drive power and has up to two orders of magnitude greater bandwidth. Compared with other traveling-wave systems,[3] or with those iterative structures proposed to date,[5] it is substantially superior from the viewpoints of size and power.

As far as the practical realization of these zigzag modulators is concerned, we believe that there are no insurmountable constructional problems. However, it should be noted that at the present time single crystals of suitable cubic electro-optic materials of the required optical quality and size are not readily available. Continuing materials research, however, indicates that this situation should improve.

We should point out that it appears that high dielectric constant centrosymmetric paraelectrics are particularly well suited to the zigzag configuration. This stems from the fact that the dielectric constant of these materials is intermediate between those of conventional dielectrics and those of ferroelectrics, and their quadratic electro-optic effect is large. The dielectric constant is large enough in these materials that the angle required by the synchronism condition is reduced to the smallest practical value. This results in a long optical path in a crystal of modest physical size, and hence in very efficient use of modulatng power. The existence of a large quadratic electro-optic effect allows us to obtain a very large effective linear coefficient by using a suitable dc bias. It is of interest to note in this connection that the performance of the $SrTiO_3$ modulator can be improved somewhat by going to lower temperatures.

In closing, we observe that the volume of active material required in the zigzag modulator can be cut in half by using the double-pass arrangement shown in Fig. 9. Here both the light and the modulating signal are reflected from the end of the modulator, with velocity synchronism being maintained for both incident and reflected waves. In this way the physical length of the crystal required to achieve a given electrical length can be halved.
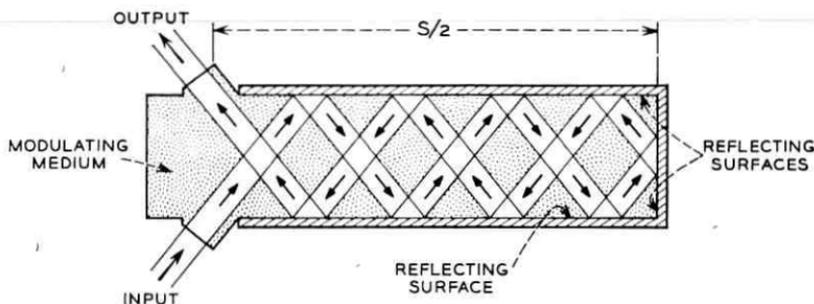
Fig. 9 — Double-pass zigzag modulator configuration.

VIII. ACKNOWLEDGMENTS

APPENDIX A

In this appendix we derive expressions for the electric field dependence of the refractive indices of the two normal modes used in the zigzag modulator.

A.1 Linear Electro-Optic Effect in Cubic Materials of Symmetry T_d

For this class of materials, to which cuprous chloride belongs, the equation of the index ellipsoid relative to the crystalline axes x', y', z' is[14]

$$\frac{1}{n_0^2}(x'^2 + y'^2 + z'^2) + 2r_{41}(E_{x'}y'z' + E_{y'}z'x' + E_{z'}x'y') = 1 \quad (101)$$

where n_0 is the index of refraction in the absence of the electric fields $E_{x'}$, $E_{y'}$, and $E_{z'}$, and r_{41} is the electro-optic coefficient. For an electric field applied along the z' axis ($0°$ modulator) in a crystallographic [001] direction, the equation reduces to

$$\frac{1}{n_0^2}(x'^2 + y'^2 + z'^2) + 2r_{41}E_{z'}x'y' = 1 \quad (102)$$

which is the equation of an ellipsoid having one of its principal axes along the z' axis, and the remaining two in the $x'y'$ plane at $45°$ to the x' and

y' axes.* In a coordinate system x,y,z coincident with these axes the index ellipsoid has the normal form

$$\left(\frac{1}{n_0^2} + r_{41}E_z\right)x^2 + \left(\frac{1}{n_0^2} - r_{41}E_z\right)y^2 + \frac{z^2}{n_0^2} = 1. \qquad (103)$$

The principal indices of refraction and their respective crystallographic directions are thus

$$n_x^2 = \frac{n_0^2}{1 + n_0^2 r_{41}E_z} \qquad \text{along [110]}$$

$$\text{0° case} \qquad n_y^2 = \frac{n_0^2}{1 - n_0^2 r_{41}E_z} \qquad \text{along [$\bar{1}$10]} \qquad (104)$$

$$n_z^2 = n_0^2 \qquad \text{along [001].}$$

In the 90° modulator, the field is applied in the x-direction. For this case one principal axis of the index ellipsoid lies along x', and the remaining two in the $y'z'$ plane, at 45° to y' and z'. Again, taking these principal axes as coordinate axes, we find for the principal indices of refraction

$$n_x^2 = n_0^2 \qquad \text{along [100]}$$

$$\text{90° case} \qquad n_y^2 = \frac{n_0^2}{1 - n_0^2 r_{41}E_x} \qquad \text{along [011]} \qquad (105)$$

$$n_z^2 = \frac{n_0^2}{1 + n_0^2 r_{41}E_x} \qquad \text{along [0$\bar{1}$1].}$$

A.2 *The Quadratic Electro-Optic Effect in Cubic and/or Amorphous Materials*

All substances (solids, liquids, and gases) which have a center of symmetry exhibit the quadratic Kerr effect. For most amorphous substances the electro-optic Kerr effect may be described by the empirical relation[37]

$$\Delta n = K\lambda_0 E^2. \qquad (106)$$

In this relation Δn is the induced birefringence, K is the Kerr constant,

---

* It is not essential that the electric field be applied in a [001] direction. In the general case, however, a principal axis of the index ellipsoid will not be in the direction of the applied field, as we assumed in our analysis in Section II. This complicates both the analysis and operation of the modulator, and results in a decreased electro-optic effect. The basic approach used here would nonetheless remain valid, since for cubic materials the orientation of the index ellipsoid depends only on the *direction* of the applied field, and not on its magnitude. Thus, there is no "wobbling" of the index ellipsoid, and the two normal modes are not mixed in the modulation process.

λ_0 is the free-space optical wavelength, and E is the applied field. Thus in the presence of an electric field isotropic materials effectively become uniaxial and will therefore display the phenomenon of double refraction.

For crystalline solids the effect is more complicated, and must in general be described in terms of a matrix of coefficients. Thus in the general equation for the index ellipsoid

$$a_{11}x'^2 + a_{22}y'^2 + a_{33}z'^2 + 2a_{12}x'y' + 2a_{23}y'z' + 2a_{31}z'x' = 1$$

the coefficients have the field dependence

$$a_{11} - \frac{1}{n_x^2} = \rho_{11}E_{x'}^2 + \rho_{12}E_{y'}^2 + \rho_{13}E_{z'}^2 + \rho_{14}E_{x'}E_{y'} + \rho_{15}E_{y'}E_{z'} \\ + \rho_{16}E_{z'}E_{x'} \quad \text{etc.} \tag{107}$$

leading to a 6×6 matrix of 36 independent electro-optic coefficients ρ_{ij}. As in the linear case, however, crystal symmetry limits the number of nonzero coefficients, and in fact for cubic materials having the highest symmetry (O_h) the quadratic electro-optic effect matrix contains only three independent coefficients and is of the form

$$\rho_{ij} = \begin{bmatrix} \rho_1 & \rho_2 & \rho_2 & 0 & 0 & 0 \\ \rho_2 & \rho_1 & \rho_2 & 0 & 0 & 0 \\ \rho_2 & \rho_1 & \rho_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \rho_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_3 \end{bmatrix}. \tag{108}$$

Note that for amorphous materials $\rho_2 = 0$, and $\rho_1 = \rho_3$.

If the field is now applied along a cube edge, e.g., the z' axis of a crystal of symmetry O_h, the equation for the index ellipsoid reduces to

$$\frac{1}{n_0^2}(x'^2 + y'^2 + z'^2) + E_{z'}^2[\rho_2(x'^2 + y'^2) + \rho_1 z'^2] = 1. \tag{109}$$

This equation is in normal form as it stands and shows that the field induces a change in the refractive index isotropically in the $x'y'$ plane as well as along the field direction or z' axis. The principal indices of refraction determined from (109) are

$$n_x^2 = n_y^2 = \frac{n_0^2}{1 + n_0^2 \rho_2 E_z^2} \quad \text{along [100], [010]}$$

0° case $\tag{110}$

$$n_z^2 = \frac{n_0^2}{1 + n_0^2 \rho_1 E_z^2} \quad \text{along [001].}$$

For a field applied in the x direction (90° modulator) the corresponding results are

$$n_x^{\ 2} = \frac{n_0^{\ 2}}{1 + n_0^2 \rho_1 E_x^{\ 2}} \qquad \text{along [100]}$$

90° case   (111)

$$n_y^{\ 2} = n_z^{\ 2} = \frac{n_0^{\ 2}}{1 + n_0^2 \rho_2 E_x^{\ 2}} \qquad \text{along [010], [001].}$$

A.3 *Linear Electro-Optic Effect in Uniaxial Materials of Symmetry* D_{2d}

For an electric field applied along the c axis of a uniaxial crystal (taken to be the z' axis, so that we have the 0° case), the equation for the index ellipsoid is[38]

$$\frac{1}{n_0^{\ 2}} (x'^2 + y'^2) + \frac{1}{n_e^{\ 2}} z'^2 + 2r_{63} E_{z'} x' y' = 1 \qquad (112)$$

where n_0 and n_e are respectively the ordinary and extraordinary refractive indices, and r_{63} is the electro-optic coefficient of interest. As was the case for cubic materials, one of the principal axes of this ellipsoid lies in the crystallographic [001] direction, the other two in the [110] and [$\bar{1}$10], so that in a coordinate system x,y,z coincident with these axes the principal indices of refraction are

$$n_x^{\ 2} = \frac{n_0^{\ 2}}{1 + n_0^2 r_{63} E_z} \qquad \text{along [110]}$$

0° case $$n_y^{\ 2} = \frac{n_0^{\ 2}}{1 - n_0^2 r_{63} E_z} \qquad \text{along [}\bar{1}\text{10]}$$ (113)

$$n_z^{\ 2} = n_e^{\ 2} \qquad \text{along [001].}$$

For the 90° case the modulating field is still applied along the c axis, but this is taken in the x direction so that the principal indices of refraction are

$$n_x^{\ 2} = n_e^{\ 2} \qquad \text{along [100]}$$

90° case $$n_y^{\ 2} = \frac{n_0^{\ 2}}{1 - n_0^2 r_{63} E_x} \qquad \text{along [011]}$$ (114)

$$n_z^{\ 2} = \frac{n_0^{\ 2}}{1 + n_0^2 r_{63} E_x} \qquad \text{along [0}\bar{1}\text{0].}$$

APPENDIX B

In this appendix we develop some expressions for the output of the three modulator-detector systems enumerated in Section 5.6, using (83) and (84) to account for the effects of random strain. These equations for the wave amplitudes at the output of the modulator were

$$e_\perp(\mathbf{r},t) = E_\perp \exp\{j[\omega t + \varphi_\perp(t)]\} \exp[j\delta_\perp(\mathbf{r})] \tag{115}$$

and

$$e_\|(\mathbf{r},t) = E_\| \exp\{j[\omega t + \varphi_\|(t)]\} \exp[j\delta_\|(\mathbf{r})]. \tag{116}$$

B.1 *Amplitude Modulators*

When the waves given by (115) and (116) are incident on an analyzer whose axes are inclined at an angle θ to $e_\perp$, the transmitted wave is

$$e_t(\mathbf{r},t) = e_\perp(\mathbf{r},t) \cos\theta + e_\|(\mathbf{r},t) \sin\theta. \tag{117}$$

The total transmitted power is proportional to

$$\int |e_t(\mathbf{r},t)|^2 da,$$

the integration extending over the beam cross section, and contains the ac terms or information carrying output terms

$$P(t)\big|_{ac} \propto E_\perp E_\| \cos\theta \sin\theta \left[\cos\varphi(t) \int_s \cos\delta(\mathbf{r})\, da \right.$$
$$\left. - \sin\varphi(t) \int_s \sin\delta(\mathbf{r})\, da \right] \tag{118}$$

where $\varphi(t) = \varphi_\perp(t) - \varphi_\|(t)$ and $\delta(\mathbf{r}) = \delta_\perp(\mathbf{r}) - \delta_\|(\mathbf{r})$.

B.2 *Heterodyne Detection*

In this system we beat in a square-law mixer a phase-modulated signal with a temporally and spatially coherent local oscillator, obtaining an output proportional to the product of the local oscillator and signal amplitudes. For a local oscillator wave of the form

$$e_{LO}(\mathbf{r},t) = E_{LO} \exp[j(\omega_{LO}t + \psi)] \tag{119}$$

the wave amplitude of the output difference frequency is given by

$$e_d(\mathbf{r},t) \propto E_{\perp,\|} E_{LO} \cos[(\omega - \omega_{LO})t - \psi + \varphi_{\perp,\|}(t) + \delta_{\perp,\|}(\mathbf{r})]. \tag{120}$$

This represents the wave amplitude produced at each point on the surface of the mixer. What is of significance beyond this point is the average amplitude $\langle e_d(t) \rangle$, since this determines the response of the IF system.* Averaging over the mixer cross section A_m gives

$$\langle e_d(t) \rangle = E_{\perp, \parallel} E_{LO} \left\{ \cos \left[(\omega - \omega_{LO})t - \psi + \varphi_{\perp, \parallel}(t) \right] \right.$$

$$\times \frac{1}{A_m} \int_S \cos \delta_{\perp, \parallel}(\mathbf{r}) \, da - \sin \left[(\omega - \omega_{LO})t - \psi \right. \quad (121)$$

$$\left. + \varphi_{\perp, \parallel}(t) \right] \frac{1}{A_m} \int \sin \varphi_{\perp, \parallel}(\mathbf{r}) \, da \bigg\} .$$

B.3 *Phase Modulator and Optical Discriminator*

To be specific, we consider an elementary discriminator formed by operating on the side of the resonance of a Fabry-Perot etalon. If we operate at the half-transmission point, the output wave is to first order given by

$$e_0(\mathbf{r},t) = \tfrac{1}{2}[1 + M(\omega_{inst} - \omega)] e_{\perp, \parallel}(\mathbf{r},t) \quad (122)$$

where M is the slope of the transmission curve at ω, and ω_{inst} is the instantaneous frequency

$$\omega_{inst} = \frac{\partial}{\partial t} [\omega t + \varphi_{\perp, \parallel}(t) + \delta_{\perp, \parallel}(\mathbf{r})]$$

$$= \omega + \frac{\partial \varphi_{\perp, \parallel}(t)}{\partial t} . \quad (123)$$

The total transmitted power is proportional to

$$\int | e_0(\mathbf{r},t) |^2 \, da ,$$

and for small frequency deviations the output power is therefore given by

$$P(t) \propto \left[1 + M \frac{\partial \varphi_{\perp, \parallel}(t)}{\partial t} \right]^2 E_{\perp, \parallel}^2 \int_S | \exp [j\delta_{\perp, \parallel}(\mathbf{r})] |^2 \, da . \quad (124)$$

---

* Regardless of the form of detector (photodiode,[39,40] photoconductor[41] or photoemitter[42,43]) the optical mixing occurs over a finite area large compared with an optical wavelength, but small compared to the different frequency wavelength. Thus we consider that the total difference frequency output of the mixer is a sum of the independent contributions from each elemental portion of the active area, i.e., it is proportional to the value of e_d averaged over the photo surface.

It should be borne in mind that (124) is strictly valid only when the discriminator is illuminated with perfectly collimated light. Strain has the additional complicating effect of producing a spread in the beam incident on the discriminator. Such a distribution of incident angles is equivalent to a broadening of the discriminator frequency response, and hence reduces the FM-to-AM conversion efficiency.

APPENDIX C

The coupling structure analyzed in this appendix is the Fabry-Perot resonator shown schematically in Fig. 10(a). The dimensions of the resonator L_c (length) and l_c (spacing) are both assumed to be much larger than the optical wavelength. The structure can therefore be analyzed in terms of plane waves bouncing back and forth between plates. Since the length of the coupler is finite, we assume that energy
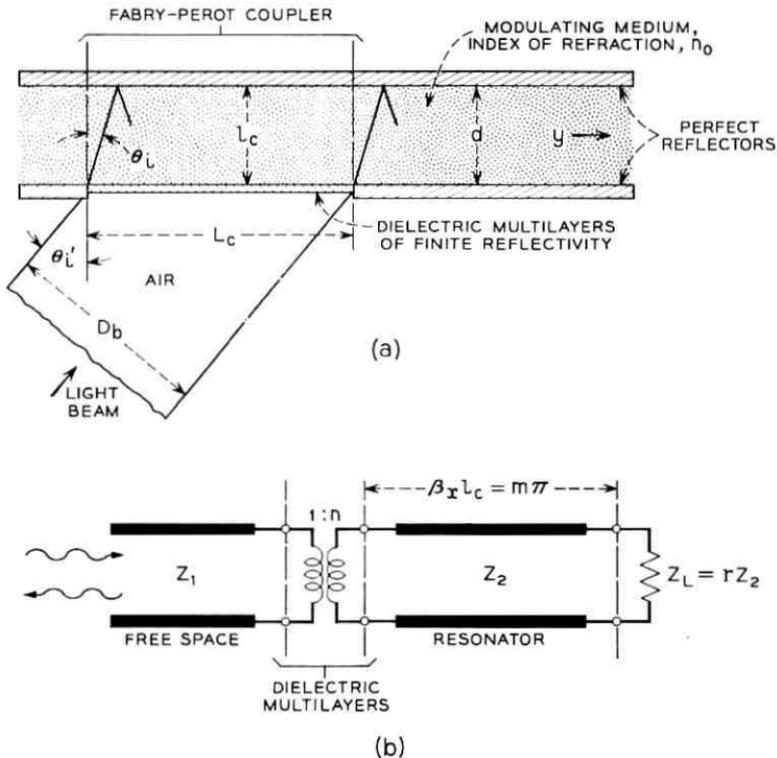

Fig. 10 — Fabry-Perot coupler for optically matching to the modulator when the aperture exceeds the reflector spacing: (a) geometrical arrangement; (b) equivalent circuit.

leaks off to the right in the y direction with group velocity $v_g = (c/n_0)$ $\sin \theta_i$. Neither diffraction losses nor losses within the electro-optic medium are included in this treatment.

We take as a model for the Fabry-Perot resonator a resonant transmission line system with equivalent circuit shown in Fig. 10(b). The resonator is represented by a transmission line which is an integral number of half wavelengths long in the transverse or x direction. The "walk-off" losses are represented by assigning a small resistance, $Z_L = rZ_2$, to the terminating short circuit. Losses can be introduced in a variety of ways, but for high-Q resonators the results are independent of the way the losses are introduced; for the present treatment the representation chosen is particularly convenient. The finite reflectivity dielectric multilayers are represented by an ideal transformer. Strictly speaking, the transformer turns ratio n will depend upon the angle of incidence and details of the multilayer construction. Experimentally, n is deduced by measuring the reflectivity of the multilayers when the resonator is terminated so as to eliminate reflections from the back surface (i.e., the perfect reflector). Finally, the characteristic impedances Z_1 and Z_2 refer, respectively, to the wave impedances of free space and the electro-optic material in the x direction. They depend upon the index of refraction, angle of incidence, and direction of polarization of the optical signal. In order to illustrate the design principles for such a coupling system, we note that for waves polarized perpendicular to the plane of incidence in Fig. 10(a), Z_1 and Z_2 are given by[9]

$$Z_1 = \eta \sec \theta_i' \qquad (125)$$

and

$$Z_2 = \frac{\eta}{n_0} \sec \theta_i \qquad (126)$$

where $\eta = 377$ ohms is the intrinsic impedance of free space and n_0 is the refractive index of the modulator material.

The objective of the analysis which is to follow is basically to calculate the dimensions of the Fabry-Perot coupler and the required reflectivity for 100 per cent transmission of light, in terms of known parameters. We begin by observing that for a given aperture D_b the coupler length L_c is given either by

$$L_c = D_b(1 - n_0^2 \sin^2 \theta_i)^{-\frac{1}{2}} \qquad (127)$$

where use has been made of Snell's law in relating θ_i to θ_i', or by

$$L_c = D_b/\cos \theta_i. \qquad (128)$$

Two equations are required to specify L_c, since it is possible for the angle of incidence required to achieve velocity synchronism to exceed the maximum achievable incident angle Θ_i determined by the limit angle relation

$$\sin \Theta_i = 1/n_0 .$$

It is found in fact that in order to achieve velocity matching the external angle of incidence must satisfy the relation

$$\sin \theta_i' = \frac{n_0^2}{\sqrt{\kappa'}}$$

and for the case $n_0^2 > \sqrt{\kappa'}$, (e.g., CuCl) we cannot employ the arrangement of Fig. 10(a) for which (127) applies. In that case it is necessary to use a more complex coupler such as the one given in Fig. 11 and use (128) in place of (127) to calculate L_c. The spacing l_c can next be found from the resonance condition $\beta_x l_c = m\pi$, which yields

$$l_c = \frac{m\lambda_0}{2n_0 \cos \theta_i}, \qquad \text{for} \quad m = 0, 1, 2, \cdots . \tag{129}$$

The calculation of the requisite reflectivity for 100 per cent light transmission into the modulator can be found in the following manner. We can couple all of the incident energy into the modulator by choosing the turns ratio n (i.e., the reflectivity of the multilayers) to obtain critical coupling, in which case the loaded Q, Q_L, becomes half of the
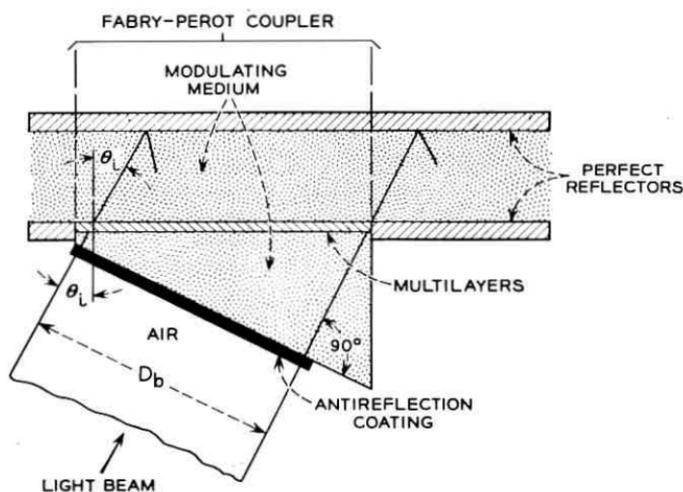

Fig. 11 — Modified coupling arrangement used when $n_0^2/\sqrt{\kappa'} > 1$.

unloaded Q, Q_0, of the resonator. Since the resonator is an integral number of half wavelengths long, the condition for critical coupling is equivalent to choosing n to satisfy the impedance matching relation

$$n^2 = r \frac{Z_2}{Z_1}. \tag{130}$$

Disregarding for the moment the determination of r, which, as will be shown subsequently, can be expressed in terms of the unloaded resonator Q, we find that the multilayer reflectivity, or square of the input reflection coefficient when transmission line Z_2 is terminated [see Fig. 10(b)], must be chosen so that

$$R_c = \left(\frac{1 - r}{1 + r}\right)^2. \tag{131}$$

Equation (130) has been used in obtaining this result.

There only remains now the problem of determining the dimensionless constant r. The unloaded Q of the resonator can be found by observing that energy must leak off in the y-direction [Fig. 10(a)] with group velocity $v_g = (c/n_0) \sin \theta_i$. The power flow to the right then is

$$P = \frac{W_s}{L_c} v_g$$

where W_s/L_c is the optical energy stored per unit length in the Fabry-Perot, and so the resonator $Q(Q_0 = \omega W_s/P)$ is simply

$$Q_0 = 2\pi \frac{L_c}{\lambda_0} n_0 \csc \theta_i. \tag{132}$$

On the other hand, for the equivalent circuit of Fig. 10(b) the unloaded Q is given by

$$Q_0 = m\pi/2r. \tag{133}$$

The constant r can now be found by equating (132) and (133). Then by using (127) or (128) and (129) to eliminate L_c and the integer m, we finally arrive at the formula

$$r = \frac{1}{2} \frac{l_c}{D_b} \sin \theta_i \cos \theta_i \begin{cases} \sqrt{1 - n_0^2 \sin^2 \theta_i} & \text{for} \quad n_0^2 < \sqrt{\kappa'} \\ \cos \theta_i & \text{for} \quad n_0^2 > \sqrt{\kappa'}. \end{cases} \tag{134}$$

It should be noted that the final results drawn from this analysis are valid only when the light is polarized normal to the plane of incidence. The method of analysis outlined applies quite generally, however, and

can be used for any plane of polarization provided that the character-istic impedances Z_1 and Z_2 can be determined. For design purposes the expressions of particular significance are (127), (128) and (129), which fix the dimensions of the coupler, and (131) and (134), which give the necessary reflectivity for transferring 100 per cent of the light into the modulator (neglecting, of course, diffraction losses).

One final comment is in order concerning the output coupler. It is evident almost by inspection that the output coupler could be of pre-cisely the same form as the input coupler but would transmit only half of the light. As an alternative, we suggest using a scheme whereby the beam is simply allowed to diverge at the output of the modulator. If the beam divergence as determined by (49) is not excessive, then in principle all of the light could be collected by a lens, thus making the output coupler 100 per cent efficient.

REFERENCES

1. Rigrod, W. W., and Kaminow, I. P., Wide-Band Microwave Light Modulation, Proc. IEEE, **51**, Jan., 1963, p. 137.
2. Kaminow, I. P., and Liu, J., Propagation Characteristics of Partially Loaded Two-Conductor Transmission Line for Broadband Light Modulators, Proc. IEEE, **51**, Jan., 1963, p. 133.
3. Peters, C. J., Gigacycle Bandwidth Coherent Light Traveling-Wave Phase Modulator, Proc. IEEE, **51**, Jan., 1963, p. 147.
4. Kaminow, I. P., Kompfner, R., and Louisell, W. H., Improvements in Light Modulators of the Traveling-Wave Type, I.R.E. Trans. Microwave Theory and Techniques, **MTT-10**, Sept., 1962, p. 311.
5. White, R. M., and Enderby, C. E., Electro-Optical Modulators Employing Intermittent Interaction, Proc. IEEE, **51**, Jan., 1963, p. 214.
6. See, for instance, Rabinowitz, P., Jacobs, S., Targ, R., and Gould, G., Homo-dyne Detection of Phase-Modulated Light, Proc. IEEE, **50**, Nov., 1962, p. 2365.
7. Born, M., and Wolf, E., *Principles of Optics*, Pergamon Press, New York, 1959, Chap. XIV.
8. Landau, L. D., and Lifshitz, E. M., *The Classical Theory of Fields*, Addison-Wesley Publishing Co., Inc., Reading, Mass., 1951, Chap. 7.
9. Ramo, S., and Whinnery, J. R., *Fields and Waves in Modern Radio*, second ed., John Wiley and Sons, Inc., New York, 1953, p. 298.
10. Jenkins, F. A., and White, H. E., *Fundamentals of Optics*, McGraw-Hill, New York, 1957, 3rd edition, Sec. 27.4.
11. See Ref. 7, Section 8.5.2.
12. Heavens, O. S., *Optical Properties of Thin Solid Films*, Academic Press, New York, 1955.
13. *American Institute of Physics Handbook*, edited by Gray, D. E., McGraw-Hill, New York, 1957, Sec. 6C.
14. Mason, W. P., Optical Properties and the Electro-Optic and Photoelastic Effects in Crystals Expressed in Tensor Form, B.S.T.J., **29**, April, 1950, p. 161.
15. Harris, S. E., and Siegman, A. E., A Proposed FM Phototube for Demodulating Microwave-Frequency-Modulated Light Signals, I.R.E. Trans. Electron Devices, **ED-9**, July, 1962, p. 322; see also Harris, S. E., Conversion of FM Light to AM Light Using Birefringent Crystals, Appl. Phys. Letters, **2**, Feb., 1963, p. 47.
16. Kaminow, I. P., Splitting of Fabry-Perot Rings by Microwave Modulation of Light, Appl. Phys. Letters, **2**, January, 1963, p. 41.

17. Kanzig, W., *Solid State Physics*, ed. Seitz, F., and Turnbull, D., Academic Press, New York, 1957, pp. 88–91.
18. Pöckels, F., *Lehrbuck Der Kristalloptik*, B. Teubner, Leipzig, 1906.
19. Kaminow, I. P., and Harding, G. O., Complex Dielectric Constant of KH_2PO_4 at 9.2 Gc/sec, Phys. Rev., **129**, Feb., 1963, p. 1562.
20. Landauer, R., Loss Tangent in Ferroelectric Above an Order-Disorder Transition, Bull. Am. Phys. Soc., **8**, Jan., 1963, p. 61.
21. Rupprecht, G., and Bell, R. O., Microwave Losses in Strontium Titanate Above the Phase Transition, Phys. Rev., **125**, March, 1962, p. 1915.
22. Silverman, B. D., Microwave Absorption in Cubic Strontium Titanate, Phys. Rev., **125**, March, 1962, p. 1921.
23. Gordon, E. I., and Rigden, J. D., The Fabry-Perot Electrooptic Modulator, B.S.T.J., **42**, Jan., 1963, p. 155.
24. Hornig, A. W., Doctoral Dissertation, Stanford University, Stanford, California, 1955. Available from University Microfilms, Ann Arbor, Mich., Publ. No. 13,268.
25. Benedict, T. S., and Durand, J. L., Dielectric Properties of Single Domain Crystals of $BaTiO_3$ at Microwave Frequencies, Phys. Rev., **109**, Feb., 1958, p. 1091.
26. Kurtz, S. K., private communication.
27. Bell, R. O., and Rupprecht, G., Measurement of Small Dielectric Losses in Material with a Large Dielectric Constant at Microwave Frequencies, I.R.E. Trans. Microwave Theory and Techniques, **MTT-9**, May, 1961, p. 239.
28. Rupprecht, G., Bell, R. O., and Silverman, B. D., Nonlinearity in Microwave Losses in Cubic Strontium Titanate, Phys. Rev., **123**, July, 1961, p. 97.
29. Geusic, J. E., Kurtz, S. K., Nelson, T. J., and Wemple, S. H., Nonlinear Dielectric Properties of $KTaO_3$ near Its Curie Point, Appl. Phys. Letters, **2**, May 15, 1963, p. 185.
30. West, C. W., Electro-Optic and Related Properties of Crystals with the Zinc Blende Structure, J. Opt. Soc. Am., **43**, April, 1953, p. 335.
31. Sterzer, F., Blattner, D. J., Johnson, H. C., and Miniter, S., 1963 International Solid State Circuits Conf. Digest of Technical Papers, p. 112.
32. Carpenter, R. O'B., The Electro-Optic Effect in Uniaxial Crystals of the Dihydrogen Phosphate Type, J. Opt. Soc. Am., **40**, April, 1950, p. 225.
33. Kaminow, I. P., Paper presented at Quantum Electronics Conference, Paris, February, 1963.
34. See Ref. 17, p. 124.
35. Kaminow, I. P., Microwave Modulation of the Electro-Optic Effect in KH_2-PO_4, Phys. Rev. Letters, **6**, May, 1961, p. 528.
36. Blumenthal, R. H., Design of a Microwave Frequency Light Modulator, Proc. I.R.E., **50**, April, 1962, p. 452.
37. Born, M., *Optik*, J. Springer, Berlin, 1933, Chap. 80
38. Billings, B. H., The Electro-Optic Effect in Uniaxial Crystals of the Type $[XH_2PO_4]$, J. Opt. Soc. Am., **39**, Oct., 1949, p. 797.
39. Reisz, R. P., High Speed Semiconductor Photodiodes, Rev. Sci. Instr., **33**, Sept., 1962, p. 994.
40. Anderson, L. K., 1963 International Solid State Circuits Conf. Digest of Technical Papers, p. 114.
41. DiDomenico, M., Jr., Pantell, R. H., Svelto, O., and Weaver, J. N., Optical Frequency Mixing in Bulk Semiconductors, Appl. Phys. Letters, **1**, Dec., 1962, p. 77.
42. McMurtry, B. J., and Siegman, A. E., Photomixing Experiments with a Ruby Optical Maser and Traveling-Wave Microwave Phototube, Appl. Opt., **1**, January, 1962, p. 51.
43. Chemelli, R. G., unpublished work.
44. See Ref. 9, p. 298.

Systematic Jitter in a Chain of Digital Regenerators

By C. J. BYRNE,* B. J. KARAFIN and D. B. ROBINSON, JR.

Digital pattern variations are a major source of timing jitter in self-timed digital regenerators. In a chain of similar regenerators, the same jitter is introduced at each one. Therefore, the jitter accumulates systematically.

A model is introduced to predict the accumulation of jitter in a chain of identical regenerators with single tuned timing filters. Simple expressions have been derived from the model for the response of the timing phase to sudden pattern transitions and to random patterns. These results show that the jitter grows without bound in a long chain, but its form is such that the tendency to cause errors will not increase as the chain becomes longer.

Measurements on a chain of 84 regenerators in a field installation are in excellent agreement with the theory.

I. INTRODUCTION

An ideal digital regenerator must restore the shape and timing of a received signal. A self-timed regenerator extracts the timing information from the received pulse train by means of a narrow bandpass filter tuned to the repetition rate. Imperfections in the regenerator and noise in the channel disturb the timing information and cause the phase of the pulse train to vary. This variation is called jitter.

Jitter in self-timed repeaters has been studied for some time. Sunde[1] and DeLange[2] concluded that jitter due to random noise would not be serious. Attention then turned to jitter introduced in the timing tanks by pattern variations. Bennett[3] studied a single mistuned regenerator. Rowe[3] analyzed a long chain of tuned regenerators. S. O. Rice studied jitter due to a random pattern in a long chain of systematically detuned regenerators.

The first fully developed digital regenerator intended for use in long chains became available in 1961. This was the repeater for the T1 PCM

---

* Bellcomm, Inc.

Carrier System. Investigations of jitter in this repeater by R. C. Chapman, Jr. and later by the present authors clearly indicated that the predominant jitter in chains was caused by systematic sources related to the pattern. No one cause of jitter was predominant, but rather several contributed significantly. Mistuning did not appear to be a major contributor.

As a result of these observations, it was felt that a simple, approximate theory was needed instead of the more complex, precise attack which had been used with mistuning. Should each of the several sources of jitter require its own precise analysis, the complete analysis, with all sources, might be extremely cumbersome. Therefore, Chapman proposed a simple model which depends on measurement to give the phase shift generated in each repeater. He used the model to predict how jitter would build up in a long chain of repeaters for a transition from one repetitive pattern to another. In this paper we will adopt Chapman's model and derive from it the behavior of a long chain of repeaters for arbitrary input patterns. The theory derives the dynamic behavior of a long chain of repeaters from static measurements of a short chain of repeaters using simple repetitive patterns.

The theory has been tested by measurements on a chain of 84 repeaters, installed in the field. Agreement with the theory is very good, and is best for the repeaters near the end of the chain. There is no theoretical reason or experimental evidence to suspect that the theory is not valid for much longer chains.

II. A MODEL OF JITTER ACCUMULATION

In a self-timed forward acting regenerator[5] (see Fig. 1) a timing wave is extracted from the received digital signal by a narrow bandpass filter. This timing wave will be phase modulated by whatever jitter has accumulated in the signal. The filter will smooth or reduce the amount of the phase modulation.
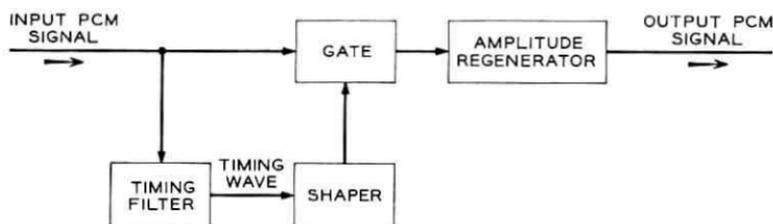

Fig. 1 — Block diagram of a self-timed forward acting regenerator.

The filtered timing wave is then used to retime the digital pattern. With ideal retiming a sharp spike is formed. This spike determines the time at which a decision is made by the amplitude regenerator and thus determines the leading edge of each pulse. Therefore, the digital pattern carries the phase of the timing wave as pulse position information. Some high-frequency, fine structure is lost, since only digital marks carry timing information, but the high-frequency jitter is not significant in long chains.

In the process of extracting, reshaping, and using the timing wave, the regenerators distort the timing and inject jitter. This jitter adds to that in the incoming signal and is processed with it in all regenerators further down the chain. The total jitter tends to increase as the chain grows longer. Our purpose is to describe the manner in which this jitter accumulates.

The model originally proposed by Chapman is shown in Fig. 2. The digital signal is not represented, since it acts only as a carrier for the timing wave for present purposes. The delays between regenerators are also eliminated, since they do not affect the manner of jitter accumulation.

A number of assumptions have been made to make the model tractable. These are:

1. The same jitter is injected at each repeater. This assumption precludes causes such as random noise and neglects variations in regenerator parameters. This assumption seems reasonable, since systematic jitter accumulates more rapidly than the nonsystematic type, and therefore the jitter at the end of a sufficiently long chain will be predominantly due to systematic causes. The most likely sources of systematic jitter are those which are related to the digital pattern itself, which is the same at each regenerator.

2. All significant jitter sources can be represented by an equivalent
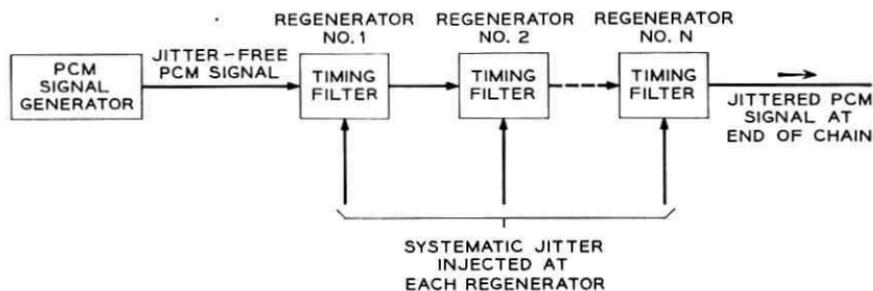

Fig. 2 — Block diagram of Chapman's model for jitter accumulation.

jitter at the input of each timing tank. This apparently arbitrary assumption will be discussed in the section on jitter sources.

3. Jitter adds linearly from repeater to repeater.

4. The timing tank is a simple single-tuned circuit, tuned to the pulse repetition frequency.

5. If the rate of change of phase jitter is small, then, as shown in Appendix A, as far as phase modulation is concerned, the timing tank is equivalent to a single-pole low-pass filter. The pole corresponds to the half bandwidth of the tuned circuit.

Therefore,

$$\Theta_o = \frac{1}{1 + (s/B)} \Theta_i \tag{1}$$

where

θ_i is the input jitter

θ_o is the output jitter

$B = \omega_o/2Q$ is the half bandwidth of the tank.

Inspection of (1) shows that very low-frequency jitter is unchanged by the timing filter, while higher-frequency jitter is attenuated and phase shifted.

The jitter introduced in the first repeater will be modified by the jitter transfer function of the timing filter and passed on to the second repeater. Here the same jitter will be introduced again and added to the jitter received from the first. The sum will be operated on by the timing tank of the second repeater and the result passed on to the third. This sequence will be repeated to the end of the chain, accumulating jitter all the way.

The jitter at the end of a chain of N repeaters will be equal to the sum of the jitter introduced in the last repeater operated on by one tank, the jitter introduced in the next-to-last repeater operated on by two tanks in cascade, and so on back to the first repeater. Since the jitter waveforms introduced in each repeater are assumed to be identical, we can express the sum as

$$\Theta_N(s) = \sum_{n=1}^{N} \Theta(s) \left(\frac{1}{1 + (s/B)}\right)^n \tag{2}$$

where $\Theta(s)$ is the transform of the pattern jitter introduced at each repeater.

The right-hand side of (2) is the sum of a geometric series and can easily be shown to be

$$\Theta_N(s) = \Theta(s) \frac{B}{s} \left[1 - \left(\frac{1}{1 + (s/B)} \right)^N \right]. \tag{3}$$

Equation (3) has an interesting approximation. If we neglect short-time (high-frequency) phenomena, we can restrict our attention to the region where the magnitude of s is much less than B. Then for large N,

$$\left(\frac{1}{1 + (s/B)} \right)^N \approx \exp\left(-N \frac{s}{B} \right). \tag{4}$$

Substituting this into (3) we have

$$\Theta_N(s) \approx \Theta(s) \frac{B}{s} \left[1 - \exp\left(-\frac{N}{B} s \right) \right]. \tag{5}$$

In the time domain, (5) becomes

$$\theta_N(t) \approx B \left[\int_{-\infty}^{t} \theta(t) \, dt - \int_{-\infty}^{t} \theta\left(t - \frac{N}{B} \right) dt \right] \tag{6}$$

$$\theta_N(t) \approx B \int_{t-(N/B)}^{t} \theta(t) \, dt. \tag{7}$$

Random jitter is more suitably represented by its power density than its Laplace transform. The term "power" is misleading here, because the square of an angle has no relation to power. However, we can define angular power density in direct analogy to a voltage power density: the angular power density at a frequency f is the mean squared value of the spectral components of the angle in a band one cycle per second wide centered at f.

The power density of jitter at the end of a chain of N repeaters can be found by replacing s by $j\omega$ in (3) and finding the square of the magnitude of the transfer function;

$$\Phi_N = \left(\frac{B}{\omega} \right)^2 \left| 1 - \left(\frac{1}{1 + j(\omega/B)} \right)^N \right|^2 \Phi \tag{8}$$

where Φ is the power density of the jitter injected in each repeater.

A simpler approximation to (8) can be derived from (5)

$$\Phi_N \approx N^2 \left[\frac{\sin \dfrac{N}{2B} \omega}{\dfrac{N}{2B} \omega} \right]^2 \Phi. \tag{9}$$

The exact transfer function given in (8) has been computed and is plotted in Fig. 3 for a range of values of N (the number of repeaters in the chain). At very low frequencies the transfer function is flat, and the power gain is proportional to the square of N. For higher frequencies, the power gain falls off as the inverse square of frequency. As a result of this spectrum shaping, the bulk of the jitter gain for large N is at low frequency. The strong ripples are interesting, but of little practical significance.

We will return to a study of the properties of (3) and (8), but first we will study the sources of jitter in each repeater.

III. SOURCES OF JITTER

Systematic jitter must be caused by a disturbance which arrives at successive regenerators in step with the pattern. The most likely cause of such a disturbance is variation in the pattern itself. In an ideal regenerator the pattern would have no effect on the phase of the timing wave. Unfortunately, in actual repeaters there are many imperfections which produce phase variations correlated with the pattern.

Causes of pattern jitter include intersymbol interference, finite pulse width effects, amplitude-to-phase conversion, and mistuning of the timing filter. These sources are extremely difficult to analyze in detail; furthermore, a careful analysis of one regenerator design would have to be completely redone for another.

As we shall show, the complexity of these jitter sources occurs principally in the detailed high-frequency fine structure of the jitter. The low-frequency behavior of jitter (averaged over many pulses) is much more tractable. In fact, the low-frequency jitter due to an arbitrary pulse pattern can be predicted from measurements with a small set of test signals. This permits us to completely ignore the detailed analysis of the jitter sources, using very simple measurements of a few sample regenerators instead.

The methods that will be used in this section lack rigor and the assumptions are not fully justified. This situation cannot be remedied without doing just what we wish to avoid: performing exhaustive analysis of a particular repeater design. Therefore, we proceed without rigor and rely on the experimental evidence to validate the results.

The first step is to show that the major sources of jitter can be represented by an equivalent jitter introduced before the timing filter, and that this jitter at a particular instant depends only on the last few bits of the pattern. In other words, there are no long memory mechanisms in the equivalent jitter sources.
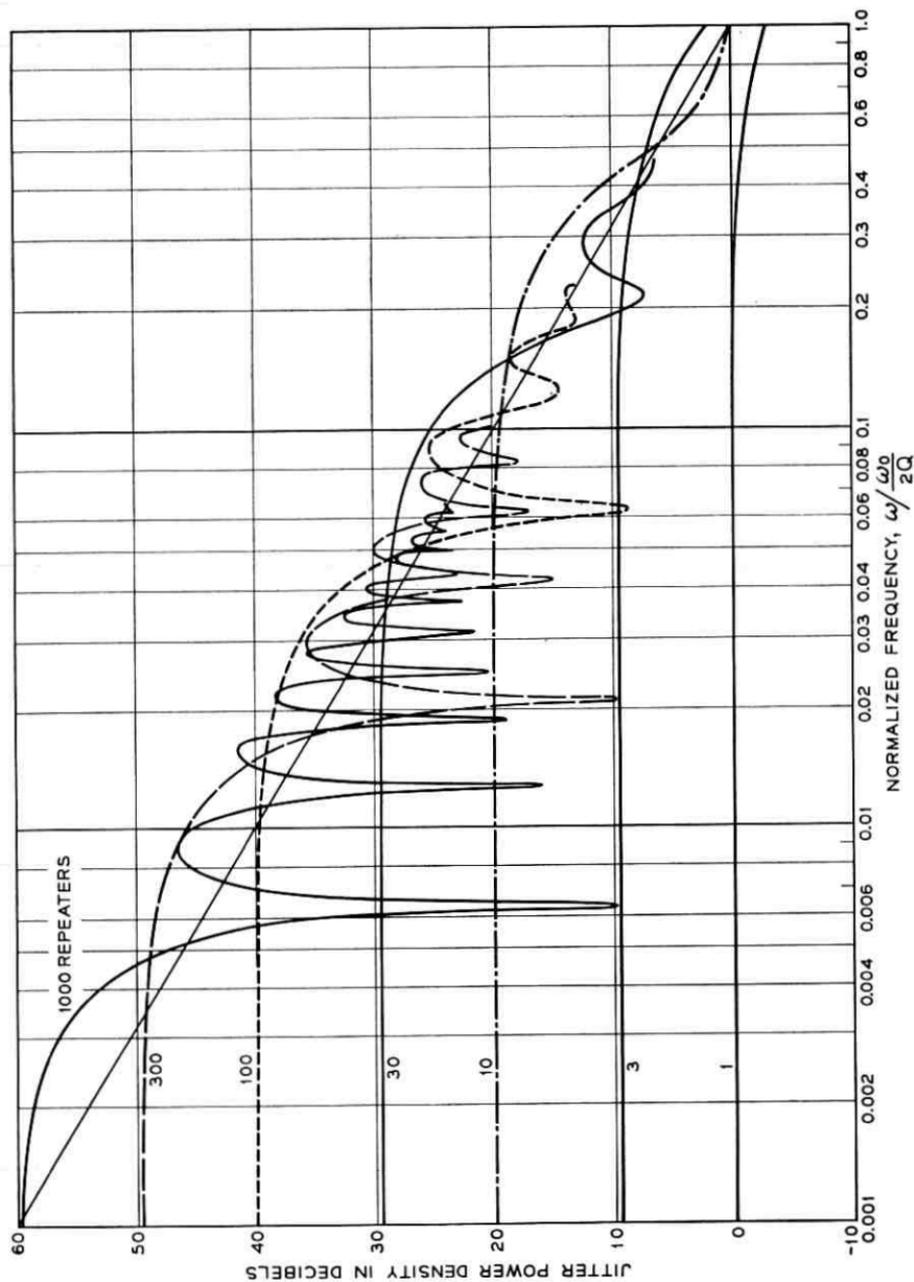
Fig. 3 — Jitter spectrum due to a random pattern (computed from the model).

The first source to be considered is intersymbol interference. Imperfect equalization of the transmission transfer function results in smearing each pulse so that adjacent pulses overlap to some extent. When clipping is used to reduce noise, the effective center of a pulse depends on whether or not pulses are adjacent to it. This effect occurs in front of the timing filter and the interaction of the pulses is effective over only a few bit intervals. Therefore, interpulse interference has the properties we assumed for jitter sources.

Finite pulse width effects[4] and amplitude-to-phase conversion are similar in that they both depend upon the amplitude of the timing wave in the tank circuit. This amplitude varies with the density of the pattern. Finite pulse width effects are due to the disturbance of the timing wave when a pulse is actually present. The amount of the phase disturbance depends upon the amplitude of the timing wave. Amplitude-to-phase conversion occurs in detecting zero crossings of the timing wave in order to form spikes for ideal retiming. If the threshold is not exactly at the zero level of the timing wave, a phase shift will be introduced which depends upon the timing wave amplitude.

At first sight, the jitter due to the timing wave amplitude variations appears to contradict the assumptions about jitter sources, since the variations are in a signal which has been operated on by the timing filter. However, we can show that these effects are equivalent to phase shifts before the filter.

The digital bit pattern can be regarded as a high-frequency amplitude modulation of a carrier at the bit repetition frequency. This signal, which is also phase modulated by jitter, is passed through the timing filter. It is shown in Appendix A that the output of the timing filter (the timing wave) will be amplitude and phase modulated, and if $A_i(s)$ is the transform of the input amplitude modulating signal, and if this signal is small, then

$$A_o(s) = A_i(s) \frac{1}{1 + (s/B)} \tag{10}$$

where $A_o(s)$ is the transform of the timing wave amplitude modulating signal and B is the half bandwidth of the timing filter.

If there is an amplitude-to-phase conversion mechanism with a coefficient C, the timing wave jitter will be

$$\Theta_o(s) = CA_o(s) = CA_i(s) \frac{1}{1 + (s/B)}. \tag{11}$$

But it is also shown in Appendix A that a phase modulation $\Theta(s)$ be-

fore the filter will cause a timing wave modulation given by

$$\Theta_o(s) = \Theta_i(s) \frac{1}{1 + (s/B)}. \tag{12}$$

By comparing (11) with (12) it can be seen that a jitter source operating on the amplitude modulation of the timing wave after the filter can be replaced by an equivalent jitter source operating on the amplitude modulation of the bit pattern before the filter. The equivalent jitter source is given by

$$\Theta_i(s) = CA_i(s). \tag{13}$$

Note that there are no long-term repeater memory effects in this jitter. Thus the timing wave amplitude sources have the properties we assumed.

The remaining source of jitter is mistuning of the timing filter. No simple argument can be presented to show that a proper equivalent exists for this. However, S. O. Rice has made a careful study of the systematic mistuning jitter. He has found that such jitter is equal to the sum of two terms: a high-frequency term which is predominant for short chains, and a low-frequency term which is predominant for long chains. The latter term is identical to the result we obtain using our model. Therefore, although mistuning does not directly fit our model, the results for this special source agree with those predicted by our model. In the case of the particular regenerators used in the experiments reported below, mistuning did not appear to be a major source of jitter.

As we have said, the above demonstration is not absolutely rigorous, and in the end we depend on the experimental evidence to justify our assumptions. This demonstration shows, however, that the model is not without some theoretical justification. So now, considering that all the important jitter sources can be represented by equivalent sources in front of the timing filter and that there are no long-term repeater memory mechanisms associated with these equivalent sources, we proceed to show how the jitter caused by an arbitrary bit pattern can be predicted from measurements with a set of test signals.

A sample from a bit sequence is shown in Fig. 4(a). Consider this as a sequence of blocks of 8 bits each. The factors involved in choosing a block length of 8 will be discussed later. The jitter introduced during the time span of a particular block will be chiefly caused by the bits within the block and only slightly influenced by bits outside of the block. This is a consequence of the assumption that there are no long memory mechanisms involved.

Therefore, it is reasonable to assign a particular value to the phase shift

(a)

100% DUTY CYCLE UNIPOLAR PULSES SHOWN

(b)
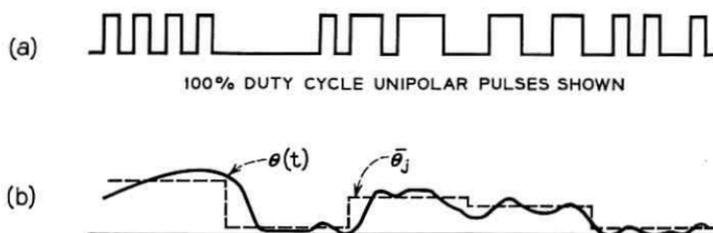
Fig. 4 — (a) Sample from a bit sequence; (b) actual and approximate jitter waveforms.

of each 8-bit block. This value is the average value of the actual phase shift over the interval. The resultant approximate jitter waveform is shown in Fig. 4(b).

The average value of phase shift for a particular 8-bit block may be determined by repeating the block over and over again with a suitable test signal generator and measuring the dc phase shift that results. By this means, an average phase shift $\bar{\theta}_j$ can be measured for each possible pattern j. There are 36 patterns 8 bits long which are distinct when repeated. These patterns are shown in Table I.

The block length used for measurement must be a compromise; it should be as long as possible to minimize end effects, but if it is too long there will be so many distinct patterns that the measurements would be impractical. The length 8 was chosen as a reasonable value.

We are now able to represent the jitter $\theta(t)$ injected due to an arbitrary pattern in terms of measurements with a few test signals. The resultant representation is not valid for high frequencies because of the artificial structure of the 8-bit blocks. However, it should be approximately valid for the low-frequency content of the jitter. Since it has been shown that low-frequency jitter accumulates much more strongly than high-frequency jitter (see Fig. 3), we are mainly interested in the low-frequency jitter. The agreement of this model with the actual low-frequency jitter will be only approximate because overlap and end effects of the blocks have been neglected.

In the next two sections this model will be applied to special types of patterns to predict the nature of jitter produced by them.

IV. REPETITIVE PATTERN TRANSITIONS

The worst possible jitter occurs when a repetitive pattern which causes an extreme phase lag is suddenly changed to a pattern which causes an extreme phase lead or vice versa. Because this is the worst case of jitter it is worth careful study.

TABLE I — STATIC PHASE SHIFTS OF REPETITIVE 8-BIT PATTERNS

j Pattern Number	Pattern	$\bar{\theta}_j$ Phase Lead Per Repeater (Degrees)	Weight, True Random	Weight, Test Random
1	10000000	−0.7	8	1
2	11000000	+5.9	8	1
3	10100000	+1.3	8	2
4	10010000	−1.8	8	2
5	10001000	−1.8	4	1
6	11100000	+2.0	8	1
7	11010000	+2.2	8	2
8	11001000	−2.4	8	2
9	11000100	−3.1	8	2
10	11000010	−0.2	8	2
11	10101000	+1.8	8	3
12	10100100	−1.3	8	3
13	11110000	+6.8	8	1
14	11101000	+0.7	8	2
15	11100100	−1.8	8	2
16	11100010	+0.9	8	2
17	11011000	+3.1	8	2
18	11010100	−0.4	8	3
19	11010010	−1.5	8	3
20	11001100	+0.9	4	1
21	11001010	−0.9	8	3
22	10101010	0 (reference)	2	1
23	11111000	+2.8	8	1
24	11110100	+1.5	8	2
25	11110010	+1.8	8	2
26	11101100	−1.1	8	2
27	11101010	+2.4	8	3
28	11100110	−0.2	8	2
29	11011010	+0.9	8	3
30	11111100	+5.9	8	1
31	11111010	+3.7	8	2
32	11110110	+3.9	8	2
33	11101110	+2.4	4	1
34	11111110	+5.0	8	1
35	11111111	+11.2	1	0
	10RRRRRR	+0.2 (test random)	sum 255	sum 64

Note: p_j = weight/sum of the weights.

An example of a sudden transition from pattern i to pattern j is shown in Fig. 5. The jitter θ which is injected into each regenerator suddenly jumps from $\bar{\theta}_i$ to $\bar{\theta}_j$. The jitter at the end of a chain of N repeaters can be found by applying (7)

$$\theta_N(t) \approx B \int_{t-(N/B)}^{t} \theta(t) \, dt \qquad (14)$$

where
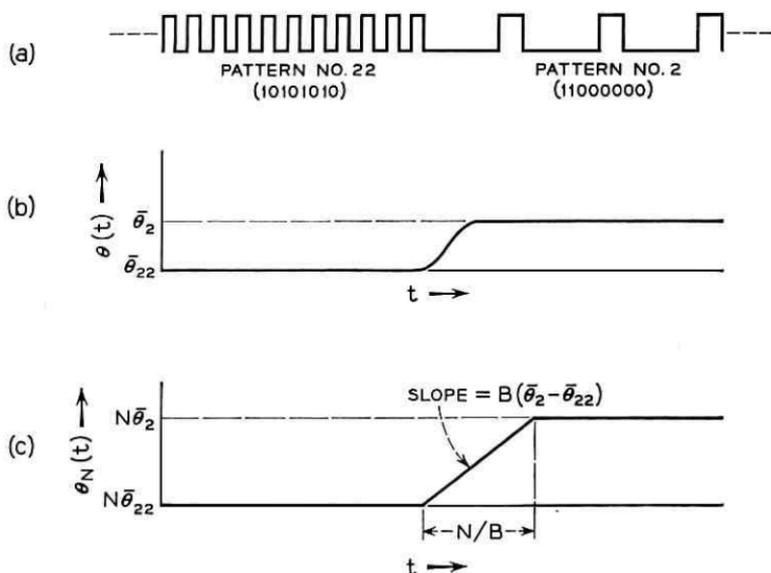
Fig. 5 — Sudden pattern transition: (a) bit pattern, (b) injected jitter, and (c) resultant jitter.

$$\theta(t) = \bar{\theta}_i, \qquad t < 0$$
$$\theta(t) = \bar{\theta}_j, \qquad t > 0.$$

Carrying out the integration, we have

$$\theta_N(t) \approx \begin{cases} N\bar{\theta}_i, & t < 0 \\ N\bar{\theta}_i + B(\bar{\theta}_j - \bar{\theta}_i)t, & 0 < t < \dfrac{N}{B} \\ N\bar{\theta}_j, & t > \dfrac{N}{B}. \end{cases} \qquad (15)$$

Equation (15) is plotted in Fig. 5(c).

The approximation used in deriving (15) is valid for low frequencies, and therefore (15) will be a good approximation when N is large.

Note that the total change of phase at the end of N regenerators will be $N(\bar{\theta}_j - \bar{\theta}_i)$. This change is unbounded as the length of the chain increases.

The rate of change of the phase during the transition is

$$\frac{d}{dt} \theta_N(t) = B(\bar{\theta}_j - \bar{\theta}_i). \qquad (16)$$

This rate of change of phase, which acts like a frequency change, is limited and independent of N. Further, since $(\bar{\theta}_j - \bar{\theta}_i)$ will be a fraction of a radian, the frequency shift will be well within the bandwidth of the timing filter (B is the half bandwidth). Note that (16) can be derived directly from (3) using the initial-value theorem.

The worst case of jitter, for both phase and frequency deviations, is characterized by the maximum value of $(\bar{\theta}_i - \bar{\theta}_j)$ for all possible pairs of repetitive patterns. This value can therefore be used as a figure of merit for the regenerator. The smaller this figure is, the less will be the maximum phase excursion for a given length of chain, and the less will be the maximum frequency deviation.

It may seem unlikely that such a transition will take place on a real communication channel because simple repetitive patterns are rare. However, they can occur in some systems, particularly when a line is idle. Another example is a PCM system being used to transmit slowly changing analog data.

V. RANDOM PATTERNS

In normal operation a digital channel can be expected to have an essentially random bit pattern. We can find the jitter injected by such a signal by breaking it up into successive 8-bit blocks. The patterns in the blocks are assumed to be independently selected from the set of 8-bit patterns listed in Table I. The patterns do not occur with equal probability. For example, the pattern 10000000 can also occur as 01000000, 00100000, etc., a total of eight distinct forms. However, the pattern 10101010 can occur in only two forms, the other being 01010101. The probability p_j of each pattern is shown in Table I as a relative weight.

From this model of the jitter $\theta(t)$ injected by a random pattern, we can derive the jitter power density Φ. Since the model is valid only for low frequencies, we will ignore the high-frequency structure of Φ.

The spectrum of the model is derived in Appendix B. It is shown that the low-frequency power density is flat (independent of frequency) and that the value of the density is

$$\Phi = 2T_B\sigma^2 \tag{17}$$

where T_B is the time duration of each block.

σ is the standard deviation of the block phase shifts

$$\sigma^2 = \frac{1}{M} \sum_j p_j \left(\bar{\theta}_j - \frac{1}{M} \sum_j p_j\bar{\theta}_j \right)^2 \tag{18}$$

where M is the total number of distinct patterns and p_j is the probability of pattern j.

The value of Φ from (17) can be substituted into (8) to give the jitter at the end of a chain of N repeaters

$$\Phi_N = \left(\frac{B}{\omega}\right)^2 \left| 1 - \left(\frac{1}{1 + j(\omega/B)}\right)^N \right|^2 2T_B\sigma^2. \tag{19}$$

Because Φ is independent of frequency, the plot of jitter in Fig. 3 can also be taken as a plot of the jitter spectrum. The 0 db line represents Φ, the jitter injected in each regenerator.

A number of conclusions can be drawn from the spectra plotted in Fig. 3. The jitter at a given frequency is bounded for all N but the bound grows without limit as the frequency approaches zero. The low-frequency asymptote grows with an amplitude proportional to N.

For very long chains of regenerators, the jitter spectrum approaches

$$\lim_{N\to\infty} \Phi_N \to \left(\frac{B}{\omega}\right)^2 \Phi. \tag{20}$$

The mean-square jitter can be found by integrating the jitter spectrum over all frequencies (see Appendix C). The result is

$$\overline{\theta_N^2} = \Phi B P(N) \tag{21}$$

where

$$\Phi = 2T_B\sigma^2$$

B is the half bandwidth of the timing tank, and
P_N is given in Table II for values of N from 1 to 100.

For chains of more than 100 repeaters,

$$\overline{\theta_N^2} \approx \tfrac{1}{2}\Phi BN. \tag{22}$$

Therefore, we see that the mean square jitter in long chain increases as N and the root mean square jitter increases as the square root of N. Rowe[4] and DeLange[2] have shown that root mean square jitter due to nonsystematic sources increases only as the fourth root of N. Therefore, systematic jitter must dominate in sufficiently long chains.

It is interesting that the total rms jitter is unbounded, even though the jitter components near any one frequency are bounded. This behavior has been confirmed by Kinariwala[6] for a very general model of a regenerator. The explanation of the paradox lies in the fact that the bound for the very low-frequency components becomes very large as the chain grows longer.

TABLE II — $\overline{\theta_N^2} = \Phi B P(N)$

N	$P(N)$
1	0.250
2	0.625
3	1.031
4	1.453
5	1.887
6	2.308
7	2.753
8	3.200
9	3.652
10	4.105
20	8.74
30	13.27
40	18.22
50	23.01
60	27.82
70	32.64
80	37.48
90	42.32
100	47.18

For $N > 100$, $P(N) \approx N/2$.

Another interesting property of (21) is that the jitter power is proportional to the timing filter bandwidth. This confirms the intuitively obvious fact that high-Q tank circuits in the regenerators will reduce jitter.

The amplitude distribution of jitter is also interesting. In principle, it is truncated, since the amplitude is limited by the worst case bounds discussed in the previous section. However, the general shape of the distribution will be shown to be Gaussian, and for long chains of repeaters the truncation occurs far out on the tail of the distribution.

To study the amplitude distribution, we will use (7)

$$\theta_N(t) \approx B \int_{t-(N/B)}^{t} \theta(t) \, dt. \tag{23}$$

The approximation is valid for low frequency and therefore is good for long chains, where low frequencies dominate. When we consider $\theta(t)$ to be a sequence of rectangular blocks of duration T_B we can write

$$\theta_N(pT_B) \approx BT_B \sum_{p-M}^{p} \theta_i. \tag{24}$$

Where $\theta_N(pT_B)$ is the jitter at time pT_B

$$M \approx \frac{N}{T_B B} \approx \frac{QN}{\pi K} \text{ (nearest integer).}$$

K is the number of bits in each block of duration T_B.

Equation (24) shows that the jitter can be expressed as the sum of M independent samples from the set of block patterns. It is well known (central limit theorem) that such a sum has an amplitude distribution which approaches a Gaussian form as M becomes large.

To see how well a distribution of this nature approaches the Gaussian shape, a sample problem was computed. The probability distribution of θ_j was assumed to consist of five equally spaced impulses of equal probability. Actually, the distribution would probably be better spread than this. The value of M was taken as 50. This corresponds to a chain of 13 repeaters with a Q of 100 and the block length taken as 8 bits.

The cumulative probability deviation for this case is plotted on probability paper in Fig. 6. The straight line represents a true Gaussian distribution. Since the distribution is symmetrical, only half of it has been plotted. A normalized deviation of 1 corresponds to the standard deviation. Truncation of the example shown in Fig. 6 occurs at a normalized deviation of 10.
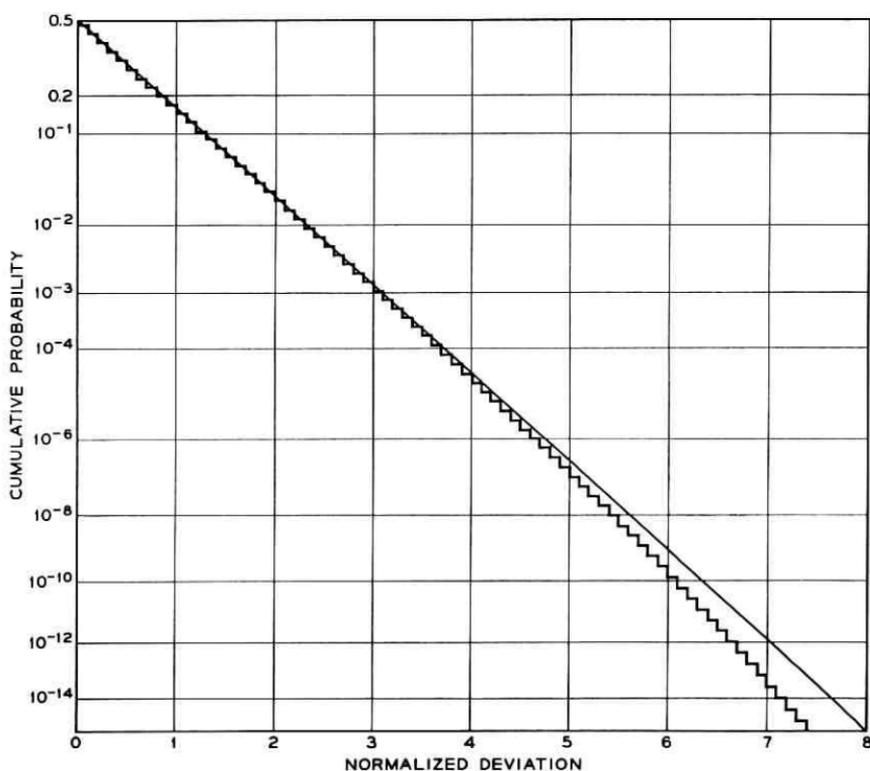

Fig. 6 — Amplitude distribution of jitter, true Gaussian and computed.

Fig. 6 shows that the distribution is Gaussian down to a cumulative probability of one part in 10^{10}. This seems to be sufficient justification for assuming a Gaussian distribution for long chains of regenerators.

In summary, a random pattern introduces jitter with a flat spectrum into each regenerator. The power density of this jitter can be predicted from measurements with a set of repetitive test signals.

The jitter at the end of a long chain of N repeaters is approximately Gaussian distributed, with an rms value which increases as the square root of N, without bound. The power density approaches a bound which is inversely proportional to the square of frequency.

VI. EXPERIMENTS

Measurements were made both in the laboratory and in a field installation which are in agreement with the theory mentioned in the previous sections.

In the laboratory, trains of one to ten repeaters were used, while trains of 14 to 84 repeaters were available at a field site located between Passaic and Newark, New Jersey. The laboratory repeaters were bipolar regenerators using ideal forward acting self-timing.[7] The field repeaters were quite similar to these generally, but differed in their physical shape, and in a minor respect in their clock circuitry. The interrepeater spacings were controllable in the laboratory only. In the field, all equipment was installed in one central office, the signals being sent on repeated lines to other offices and then returned on separate lines. The field situation was such that crosstalk and other interference was negligible. Since lines could not be terminated in the field conveniently at places other than central offices, the minimum number of repeaters that could be used was 14. All the different line lengths tested in the field were thus multiples of 14.

The repeated lines were driven by a word generator which could emit a train of bipolar pulses, and the necessary clock references for the phase measuring equipment. The generator could emit any of the fixed patterns possible with 8 digits, or a random pattern, or a combination of fixed and random. In addition, it could switch between two different fixed patterns at rates of several kc or less. A restriction on the patterns in this case was that all the pulses in the less dense pattern must also be present in the denser one.

The phase detector used for measuring the jitter operated in the following manner. The reference clock from the word generator was used to set a flip-flop which was reset by a clock derived from the last repeater in the train. Jitter in this clock would cause the flip-flop to gener-

ate a duration-modulated pulse. To extend the measuring range, both clocks were divided by two before being sent to the flip-flop. This operation did not reduce the amount of jitter presented to the instrument, however. The jitter information contained in the duration variations of the flip-flop output pulses was converted to corresponding amplitude variations by passing those pulses through suitable low-pass filters and amplifiers. This, of course, is a standard way of demodulating duration-modulated pulses.

The jitter information, now in the form of voltage variations, was passed to a detector. For measuring fixed pattern shifts, an oscilloscope and camera were used. For random patterns, a true rms voltmeter and high-quality tape recorder served as detectors.

A typical photograph of transitions between two fixed patterns is shown in Fig. 7. Note the close resemblance to the waveform derived from the model (Fig. 5c). It was shown in Section IV that the total change of phase at the end of N regenerators should be N times the phase difference between the two fixed patterns in question for one regenerator. Fig. 8 shows the total change of phase plotted against N for transitions between various fixed patterns. The relations are clearly linear. It is also apparent that pattern density is not the only factor influencing phase shift, as was mentioned in Section III. Note for example the transitions from 10000000 to 11000000 (1/8 to 2/8) and 10000000 to 10001000 (1/8 to 1/4). Both have equal density change but phase changes of different sign and amplitude. Examination of Fig. 8 would lead one to believe that the effect of intersymbol interference on phase
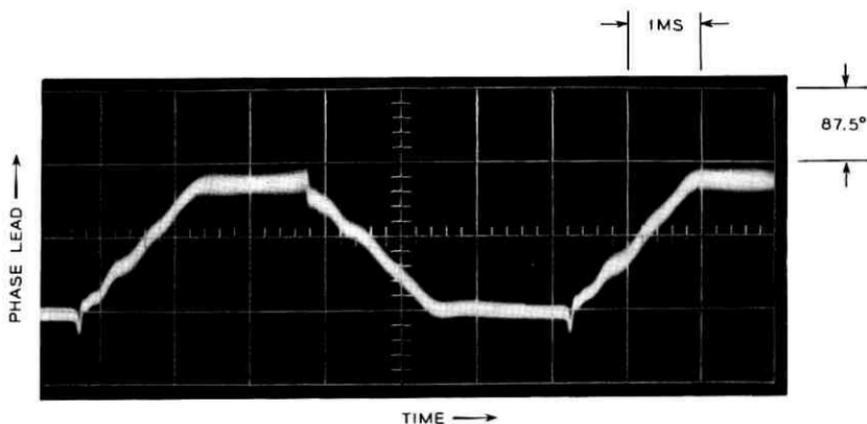

Fig. 7 — Photograph of phase shift due to a transition from a 1/8 pattern (top) to an 8/8 pattern (bottom), with 84 repeaters; phase scale, 87.5 degrees/cm; time scale, 1 msec/cm.
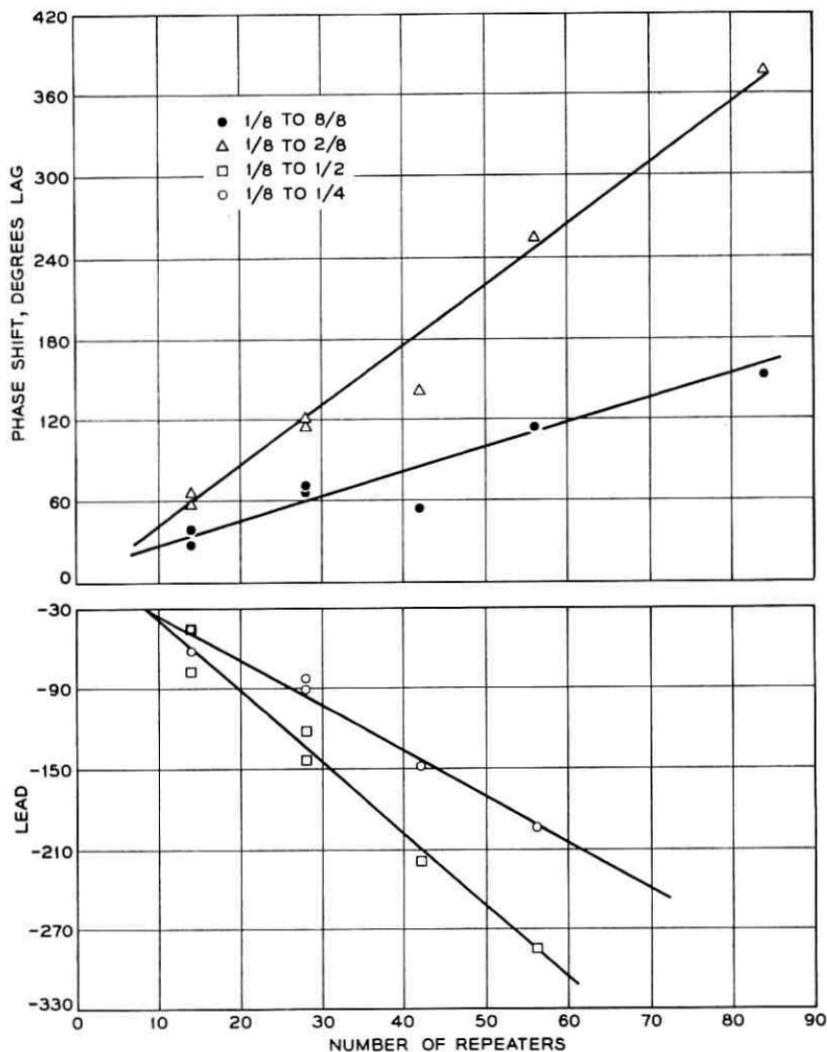
Fig. 8 — Steady-state phase shift between fixed patterns.

shift is comparable to the effect of density. The transitions from 1/8 to 1/4 and 1/8 to 1/2 (10101010) involve patterns with little or no inter-symbol interference. A doubling of density between 1/4 to 1/2 causes about a 50 per cent increase in phase shift. Changing the pattern geometry to 2/8, which has intersymbol interference, is sufficient to cause a change in the algebraic sign of the phase shift. In the case of the 1/8 to 8/8 (all pulse) transition, the effects of density and intersymbol inter-

ference partially cancel each other, but the resultant shift indicates that the intersymbol interference is the larger.

Equation (16) states that the rate of change of phase during the transition between two given fixed patterns is proportional to the half bandwidth of the timing filter and to the phase change between the two patterns. Restating this equation

$$\frac{d}{dt}\,\theta_N(t) \,=\, B(\bar{\theta}_j \,-\, \bar{\theta}_i) \,=\, \frac{\pi f_o}{Q}\,\Delta\theta \tag{25}$$

where $\Delta\theta$ is the phase change per repeater between the two patterns, f_o is the pulse repetition frequency (timing filter center frequency), and Q is the quality factor of the tank. Fig. 9 shows the measured rate of change of phase in degrees per millisecond plotted against the absolute value of the phase change per repeater. The slopes were obtained from photographs of the transients during the various transitions. Using the known pulse repetition frequency, (25), and the best line fit from Fig. 9, it is possible to arrive at a value of Q for the average repeater tank. This is found to be 72. This figure is quite compatible with the expected values of loaded Q for repeater timing tanks (about 80).

To check the effect of repeater spacing, the phase shift for one pattern transition (1/8 to 8/8) was measured in the laboratory with the repeaters first correctly spaced at 6000-foot intervals, and then spaced at 5000-foot intervals with no compensation. This is a more severe condition than would be encountered in actual repeater practice. The data of Fig. 10 show that shortening the span has little effect on phase shift. Photographs showed the transient behavior to be unaffected also.

The properties of jitter caused by a random pattern were investigated with a special quasi-random signal. The pattern was constrained to have one forced pulse followed by one forced space at the beginning of every block of eight slots, the remaining six spaces being filled with random pulses. Forcing the pulse was necessary to ensure that the clock of the self-timed repeaters did not die out. There is no particular reason for forcing the space except that it kept the over-all pulse density at one half.

In testing the field repeatered lines with this quasi-random pattern, the output from the phase detector was recorded on tape and studied later with amplitude and frequency distribution analyzers and a true rms voltmeter. The results of these measurements are plotted in Figs. 11–15, and will be used to corroborate the theoretical predictions of Section V.

The measured spectral distributions of the output data are shown in
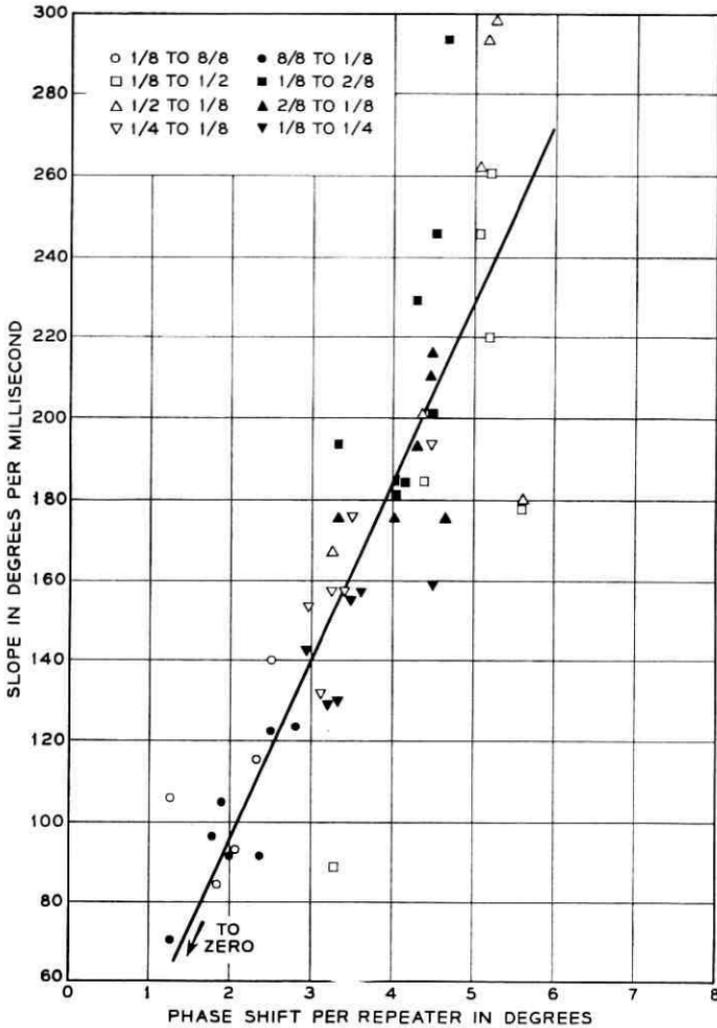
Fig. 9 — Measured rate of change of phase versus the absolute value of the phase change per repeater.

Fig. 11. We have shown in Section V that the spectrum of random jitter injected at each repeater is flat; therefore, the spectrum at the end of a chain of repeaters should have the shape of the corresponding jitter transfer function, plotted in Fig. 3. A comparison of Figs. 3 and 11 shows that they are in excellent agreement, even to the details of the nulls. A direct comparison of the measured and theoretical spectra for a chain of 84 repeaters is shown in Fig. 12. The theoretical curve is given by (8).
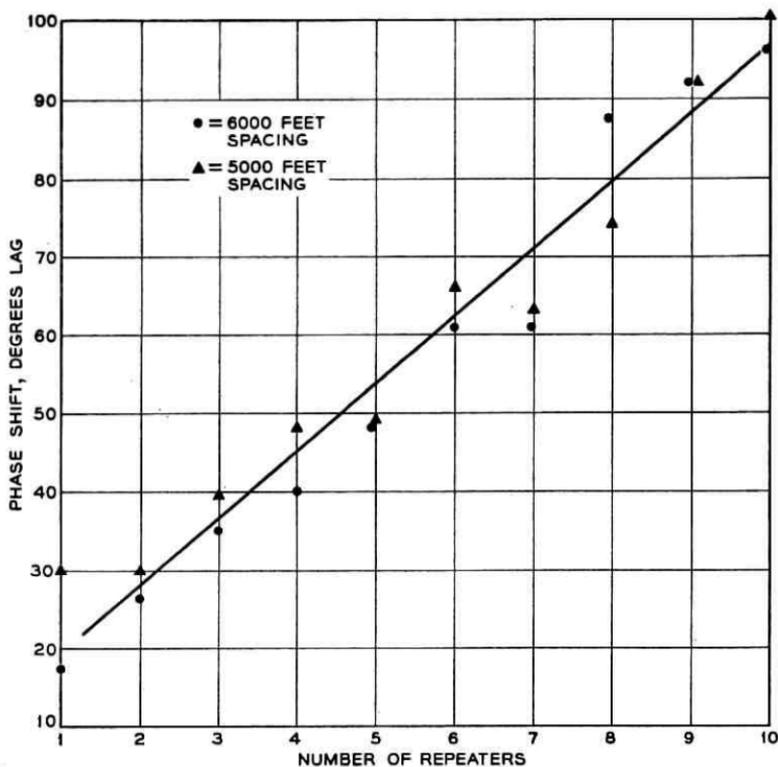
Fig. 10 — Measurement of the effect of repeater spacing.
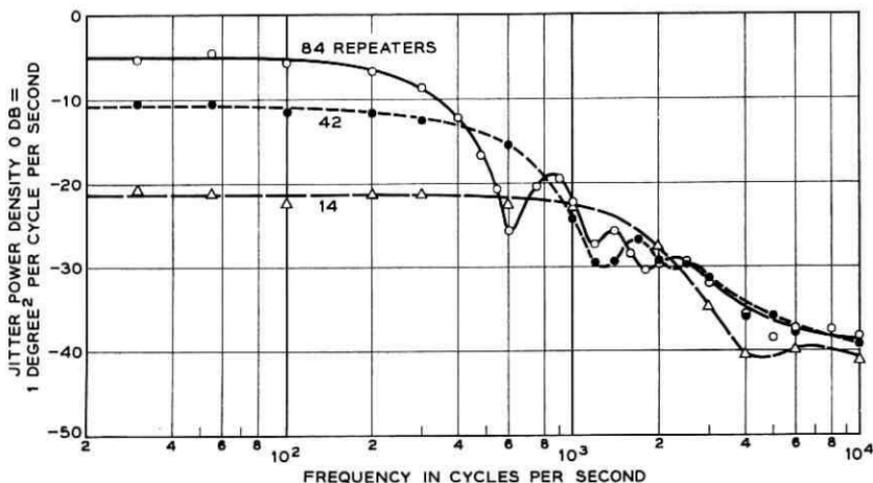
Fig. 11 — Measured spectral distributions for a random pattern (10 - - - - - -).
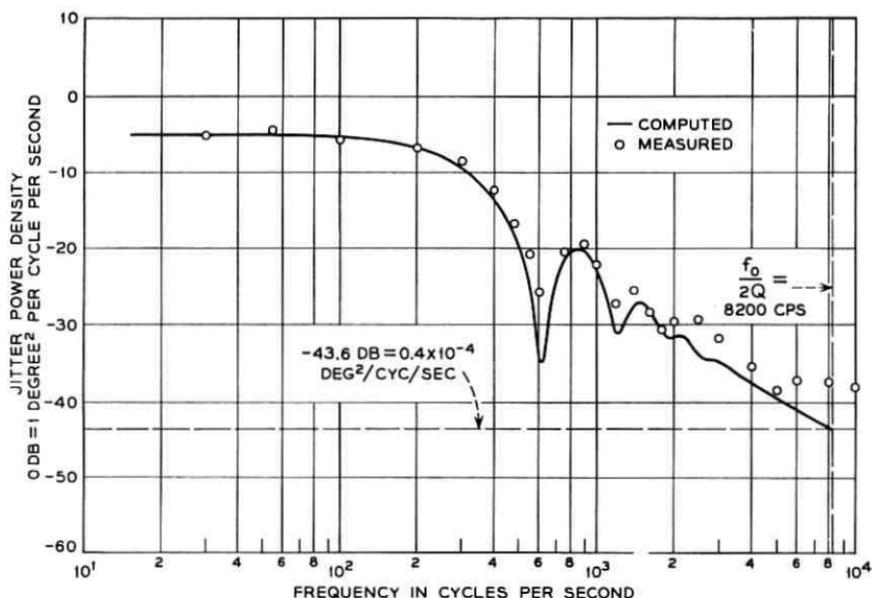
Fig. 12 — Comparison of measured and theoretical spectra for a chain of 84 repeaters.

The two parameters Φ and B were adjusted to give the best fit. The corresponding jitter injection density was 0.44×10^{-4} degrees squared per cycle per second and the tank Q was 94 (compare a value of 72 from pattern transition slopes).

Inspection of (8) shows that the jitter amplitude density at very low frequencies should be directly proportional to the number of repeaters. This is a consequence of assuming systematic jitter sources. The measured low-frequency jitter density is plotted against the number of repeaters in Fig. 13. The curve is indeed linear. The slope of this line can be used to find the power density of jitter injected at each repeater (0.44×10^{-4} degrees squared per cycle per second). The rms value of the jitter was measured for each chain of repeaters. The result is shown in Fig. 14. The solid curve is plotted from (21), with the product (ΦB) adjusted for a best fit. Using a value of 0.44×10^{-4} for Φ (from Fig. 13) the value of Q is found to be 69, in good agreement with the value 72 found from slope data. The dotted line in Fig. 14 is the asymptote for long chains, given by (22).

The amplitude distribution of the random jitter was also measured. All of the curves were Gaussian in form, with no indication of truncation within the limits of the amplitude analyzer. A typical distribution (54
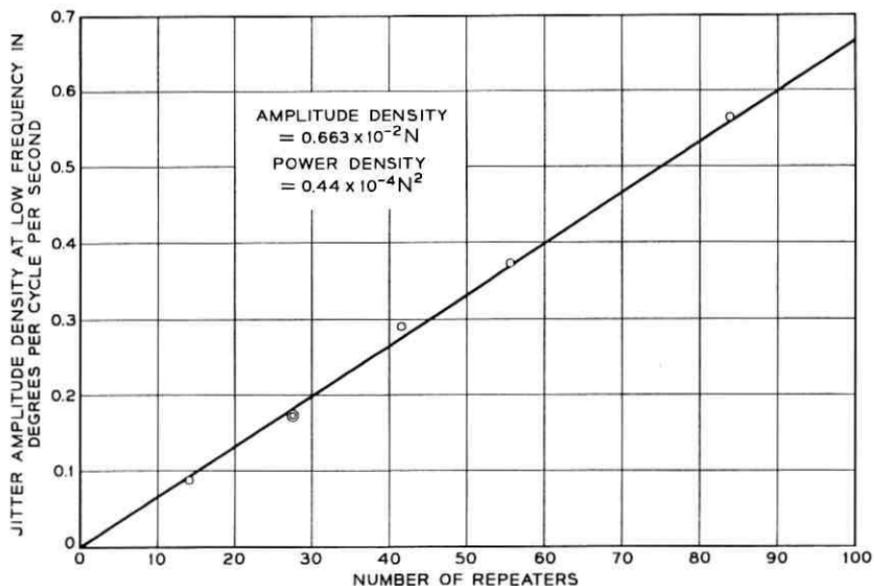
Fig. 13 — Measured low-frequency jitter amplitude density versus the number of repeaters in the chain.
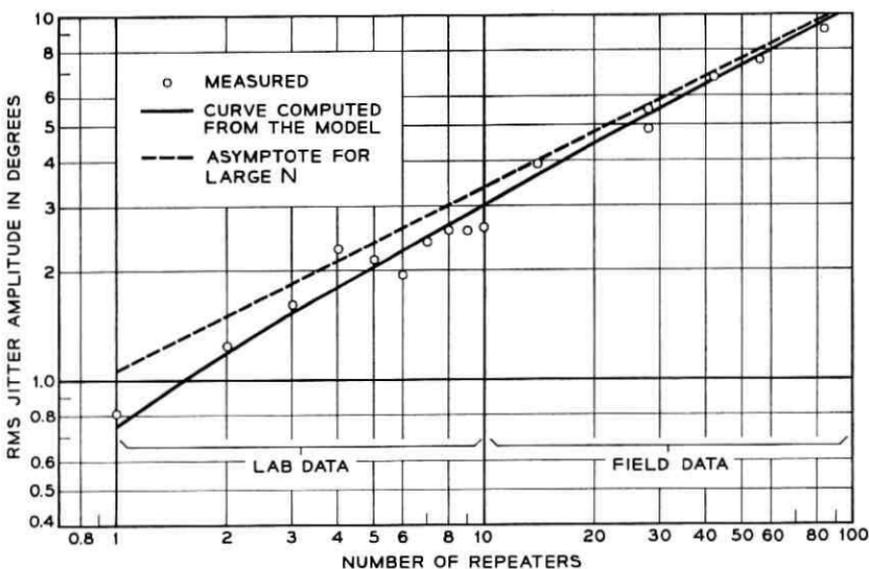

Fig. 14 — Root-mean-square jitter due to a random pattern (10 - - - - - -) versus the number of repeaters in the chain, measured and calculated.

repeaters) is plotted on probability paper in Fig. 15. Note that the curve extends to amplitudes which have a probability of only 2×10^{-4} of being exceeded, with no sign of truncation.

The data described above establish all the predicted properties of the model. However, the important parameter Φ, the jitter power density injected at each regenerator, has been measured using random patterns on long chains of repeaters. Equation (17) relates Φ to measurements with simple repetitive patterns on a few repeaters. This allows the performance of long chains of repeaters to be predicted from short chains, measured in the lab. To test the method, a chain of 10 repeaters was set up in the lab, as described above. The average phase shift of these repeaters was measured for every possible 8-bit repetitive pattern, as shown in Table I. Pattern 22 was used as a reference.

The standard deviation of the phase shifts was found to be $2.225°$, calculated according to (18). The weights used were those for the test random pattern. Using this value, (17) was used to find the power density Φ, yielding a value of 0.513×10^{-4} deg²/cps. Since the measured
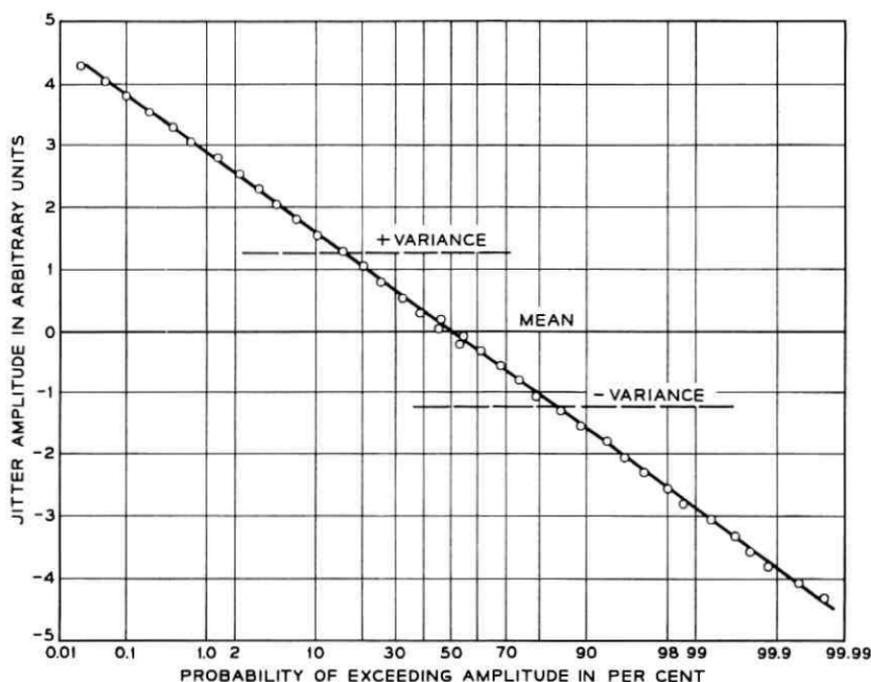

Fig. 15 — Typical amplitude distribution of jitter due to a random pattern (10 - - - - - -) (54 repeaters).

value on long chains in the field was 0.44×10^{-4} deg^2/cps, this result validates the assumptions made in deriving the model and shows that simple measurements in the lab can adequately predict the behavior of long chains in the field.

VII. ALIGNMENT JITTER

We have shown that jitter is unbounded; however, we will now show that the nature of jitter is such that it will not contribute strongly to digital errors in the regenerator chain.

Alignment jitter has been defined by Rowe[4] as deviations in alignment between the input signal pulse and the corresponding timing pulse. In terms of our model, the output jitter of one repeater is the input jitter for the next, and the timing jitter of a repeater becomes its output jitter. Therefore, we will redefine alignment jitter in a particular repeater as the difference between its output jitter and the output jitter of the preceding regenerator:

$$\theta_{aN} = \theta_N - \theta_{N-1} \tag{26}$$

Substituting from (3)

$$\Theta_{aN}(s) = \Theta(s) \left(\frac{1}{1 + (s/B)} \right)^N \tag{27}$$

Since the jitter injected at each repeater is bounded, we can show by a straightforward analysis that the alignment jitter is bounded, and further, the bound does not increase as the length of the chain of repeaters increases. (See Appendix D.) These properties are easily seen if we consider simple pattern transition. In that case the alignment jitter increases from zero to a value equal to the phase shift due to the final pattern. (See Fig. 16.) An important assumption in our discussion of
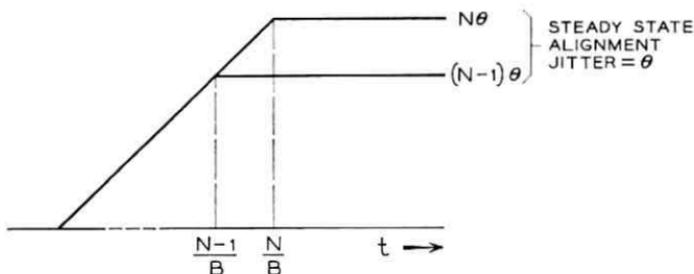

Fig. 16 — Alignment jitter for a fixed pattern.

alignment jitter is that the frequencies and Q's of the timing filters are identical.

The error margin for random noise interference has been measured at the first and last repeater in a long chain of repeaters in the field. No significant difference was found. In addition, analog measurements on voice channels in T1 carrier terminals showed no noticeable change as the length of the repeatered line coupling these terminals was increased from about 11 to about 130 miles (14 repeaters to 154 repeaters).

It is clear that even though jitter becomes quite large in long chains, the error margin of the repeaters does not decrease.

VIII. SUMMARY

We have measured the jitter in a field installation of the T1 24-channel PCM carrier system. The maximum number of repeaters in tandem was 84. The jitter was produced by pattern variations in the signal.

The largest jitter observed is produced by a sudden transition from one repetitive pattern to another. This causes a steady-state phase shift which increases linearly with the number of repeaters. The amount of the shift depends on the patterns used and is of the order of two degrees per repeater. The rate of change of the phase shift also depends on the patterns used and is of the order of 2000 radians per second. This phase shift slope becomes independent of the number of repeaters. These measurements are consistent with the work of R. C. Chapman.

Jitter produced by random signals has the character of noise. The root mean square phase deviation after 84 repeaters is about 10°. For long chains of repeaters the rms jitter increases in proportion to the square root of the number of repeaters. At a given frequency, the jitter power density approaches a limit as the number of repeaters grows large. However, at low frequencies this limit is very large and is not approached for a reasonable number of repeaters. This low-frequency jitter power density is directly proportional to the number of repeaters. Therefore, for long chains of repeaters, the jitter is concentrated at low frequencies. For 84 repeaters the jitter power is mostly below 400 cycles per second. It is interesting to note that this is much less than the half bandwidth of the timing filters, which is about 10,000 cycles per second. The amplitude distribution of jitter is Gaussian.

We have derived a model which is in agreement with our measurements. This model assumes that the causes of jitter are systematic: that is, the same at each repeater. While nonsystematic effects are undoubtedly present, their jitter does not accumulate as rapidly as that

due to systematic sources and is not significant in long chains. We further assume that only the low-frequency components of jitter are important, since high-frequency components are strongly attenuated by the timing filters. The model quantitatively predicts the measurements of random jitter from measurements of the steady-state phase shift of 8-bit repetitive patterns.

We have shown that the model is valid for many sources of jitter, including amplitude-to-phase conversion, intersymbol interference, and finite width effects. It is also valid for the effects of mistuning in very long chains of repeaters, although mistuning was not an important source of jitter in the repeaters we measured.

IX. ACKNOWLEDGMENTS

APPENDIX A

Amplitude and Jitter Transfer Functions of a Single-Tuned Tank Circuit

The following argument is based on work by S. O. Rice on the analysis of FM systems. The timing wave is extracted from the signal by a high-Q single-tuned filter. Therefore, for the input to this filter, we need only consider that component of the signal which is near the bit rate. This component is amplitude modulated by the pulse density and phase modulated by the jitter.

Therefore, let the signal incident on the filter be given by the real part of

$$v_i(t) = \exp\left[j\omega_c t + j\theta(t) + a(t)\right] \tag{28}$$

where ω_c is the bit frequency, and $\theta(t)$ and $a(t)$ are arbitrary functions of time with the restriction that their time derivatives are small.

We will write the output of the filter as

$$v_o(t) = \exp\left[j\omega_c t + j\theta_o(t) + a_o(t)\right]. \tag{29}$$

Further, if we express the impulse response of the filter as

$$g(t) = \int_{-\infty}^{\infty} Y(j\omega)e^{j\omega t}\, df \tag{30}$$

we have

$$v_o(t) = \int_{-\infty}^{\infty} \exp\left[j\omega_c(t - x) + j\theta(t - x) + a(t - x)\right]g(x)\,dx. \quad (31)$$

Equating (29) and (31), we have

$$j\theta_o(t) + a_o(t)$$
$$= \ln \int_{-\infty}^{\infty} \exp\left[j\theta(t - x) + a(t - x) - j\omega_c x\right]g(x)\,dx. \quad (32)$$

We will assume that the transfer function of the filter is zero for frequencies above $(f_c + r)$ and below $(f_c - r)$, for some r. When we impose this condition on (30) and make the change of variable

$$f = f_c + f' \quad (33)$$

we have, to a high degree of approximation

$$j\theta_o(t) + a_o(t) = \ln \int_{-\infty}^{\infty} \exp\left[j\theta(t - x) + a(t - x)\right]h(x)\,dx \quad (34)$$

where

$$h(x) = \frac{1}{2\pi} \int_{-r}^{r} Y(j\omega_c + j\omega')\, e^{j\omega' x}\, d\omega'. \quad (35)$$

For a tuned tank

$$Y(j\omega_c + j\omega) \approx \frac{1}{1 + j(\omega/B)} \quad (36)$$

where B is the half bandwidth.

For a high-Q filter we may let r in (35) go to infinity, whereupon

$$h(x) = \begin{cases} 0, & x < 0 \\ Be^{-Bx}, & x > 0. \end{cases} \quad (37)$$

Now we return to the evaluation of (34), which is our problem. We will approximate

$$e^{j\theta(t-x)+a(t-x)} = e^{a(t)+j\theta(t)} \sum_{0}^{\infty} \frac{(-x)^n}{n!} F_n(t) \quad (38)$$

where

$$F_n(t) = e^{-j\theta(t)-a(t)} \frac{d^n}{dt^n} e^{a(t)+j\theta(t)}. \quad (39)$$

We now define

$$m_n = \int_{-\infty}^{\infty} x^n h(x)\, dx \tag{40}$$

or

$$m_n = n!B^{-n}. \tag{41}$$

Then from (38) and (40) we have

$$\int_{-\infty}^{\infty} \exp\,[j\theta(t-x) + a(t-x)]h(x)\, dx$$
$$= e^{j\theta(t)+a(t)} \left\{ 1 + \sum_{n=1}^{\infty} \frac{(-1)^n m_n}{n!}\, F_n(t) \right\}. \tag{42}$$

Now in general

$$\ln\left(1 + \sum_{1}^{n} \frac{\alpha_n t^n}{n!}\right) = \frac{t}{1!}\,\alpha_1 + \frac{t^2}{2!}\,(\alpha_2 - \alpha_1{}^2)$$
$$+ \frac{t^3}{3!}\,(\alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1{}^3) + \cdots. \tag{43}$$

So if we take $t = -1$, $\alpha_n = m_n F_n(t)$, we have

$$\ln \int_{-\infty}^{\infty} \exp\,[j\theta(t-x) + a(t-x)]h(x)\, dx$$
$$= j\theta_o(t) + a_o(t)$$
$$= a(t) + j\theta(t) + \left[-\frac{m_1 F_1}{1!} + \frac{1}{2!}\,(m_2 F_2 - m_1{}^2 F_1{}^2) \right. \tag{44}$$
$$\left. - \frac{1}{3!}\,(m_3 F_3 - 3m_2 m_1 F_2 F_1 + 2m_1{}^3 F_1{}^3) + -\cdots \right].$$

Now we replace the F_n by their values from (39), and after some manipulation we have

$$a_o(t) + j\theta_o(t) = \int_{-\infty}^{\infty} [j\theta(t-x) + a(t-x)]h(x)\, dx \tag{45}$$

+ (terms in powers of the derivatives of $\theta(t)$ and $a(t)$ normalized to the half bandwidth)

+ (cross products of the derivatives of $\theta(t)$ and $a(t)$ normalized to the half bandwidth)

where we have made use of the fact that

$$[a(t) + j\theta(t)] - m_1[a'(t) + j\theta'(t)] + \frac{m_2}{2}[a''(t) + j\theta''(t)] + \cdots$$

$$= \int_{-\infty}^{\infty} [a(t - x) + j\theta(t - x)]h(x)\ dx. \tag{46}$$

Now, if we neglect powers and cross products of all the derivatives of $\theta(t)$ and $a(t)$, and use the value of $h(x)$ from (37), we have

$$a_o(t) + j\theta_o(t) = \int_0^t Be^{-Bx}a(t - x) + j\theta(t - x)\ dx. \tag{47}$$

Equating real and imaginary parts and recognizing the convolution integral, we have

$$\Theta_o(s) = \frac{1}{1 + (s/B)}\ \Theta(s) \tag{48}$$

and

$$A_o(s) = \frac{1}{1 + (s/B)}\ A(s). \tag{49}$$

The relationship given in (48) is one of the results we are seeking. The amplitude modulating signal, however, is $e^{a(t)}$, not $a(t)$. If now we restrict $a(t)$ to be small, then

$$e^{a(t)} \approx 1 + a(t).$$

From (47) it is clear that if $a(t)$ is small, then $a_o(t)$ is small, and then the output amplitude modulation

$$e^{a_o(t)} \approx 1 + a_o(t).$$

Then we see that (49) gives the relationship in the transform domain between the varying term of the input AM and the varying term of the output AM.

APPENDIX B

Power Density of a Jitter Produced by a Random Pattern

The jitter $\theta(t)$ produced by a series of 8-bit blocks may be represented by

$$\theta(t) = \sum_{-\infty}^{\infty} \bar{\theta}_n A(t - n8T_o) \tag{50}$$

where

T_o is the bit interval,

$\bar{\theta}_n$ is the phase shift that would be produced by the pattern of block n, in a repetitive pattern, and

$$A(t) = 1, \quad -4T_o < t < 4T_o$$
$$\quad\quad 0, \quad\quad |t| > 4T_o.$$

The patterns in successive blocks are statistically independent, and $\bar{\theta}_n$ can be taken to have zero mean. When we add an arbitrary starting time we make the process whose sample functions are given by (50) stationary. Then the autocorrelation function can be shown to be

$$\psi(\tau) = \sigma^2 \Delta(\tau) \tag{51}$$

where

$$\Delta(\tau) = \left(1 + \frac{\tau}{8T_o}\right), \quad -8T_o < \tau < 0$$

$$= \left(1 - \frac{\tau}{8T_o}\right), \quad 0 \le \tau < 8T_o$$

$$= 0, \quad |\tau| \ge 8T_o$$

where τ is the time difference, and σ is the standard deviation of $\bar{\theta}_n$.

The one-sided power density is given by

$$\Phi(\omega) = 2 \int_{-\infty}^{\infty} \psi(\tau) e^{-j\omega\tau} \, d\tau. \tag{52}$$

Substituting (51) in (52) gives

$$\Phi(\omega) = \frac{4\sigma^2}{\omega} \frac{\sin^2 \omega 4T_o}{\omega 4T_o}. \tag{53}$$

Applying L'Hospital's rule we obtain

$$\Phi(0) = 2(8T_o)\sigma^2. \tag{54}$$

Furthermore, it is clear from (53) that at zero frequency the slope of the power density is zero, and because T_o is small for the system under consideration we may assume to a very high degree of accuracy that over the band of frequencies which are of interest, the power density is flat.

The values of $\bar{\theta}_n$ have a discrete probability distribution, indicated in Table I. Therefore,

$$\sigma^2 = \frac{1}{M} \sum_j p_j \left(\bar{\theta}_j - \frac{1}{M} \sum_j p_j \bar{\theta}_j\right)^2 \tag{55}$$

where p_j is found by dividing the weights by the sum of the weights. M is the total number of distinct patterns.

APPENDIX C

Integration of the Jitter Spectrum

We are indebted to A. J. Goldstein for the following integration.

The spectrum of jitter at the end of a chain of N repeaters was given in the section on the model in (8).

$$\Phi_N(\omega) = \Phi \frac{B^2}{\omega^2} \left[1 - \left(\frac{1}{1 + j(\omega/B)} \right)^N \right]^2. \tag{56}$$

The mean-square jitter is given by

$$\overline{\theta_N^2} = \frac{1}{2\pi} \int_0^\infty \Phi_N(\omega) \, d\omega$$

$$= \frac{\Phi}{2\pi} \int_0^\infty \frac{B^2}{\omega^2} \left| 1 - \left(\frac{1}{1 + j(\omega/B)} \right)^N \right|^2 d\omega \tag{57}$$

$$= \frac{\Phi B}{2\pi} \int_0^\infty \frac{1}{\omega^2} \left| 1 - \left(\frac{1}{1 + j\omega} \right)^N \right|^2 d\omega$$

$$\overline{\theta_N^2} = \Phi B P(N) \tag{58}$$

where

$$P(N) = \frac{1}{2\pi} \int_0^\infty \frac{1}{\omega^2} \left| 1 - \left(\frac{1}{1 + j\omega} \right)^N \right|^2 d\omega$$

$$P(N) = \frac{1}{2\pi} \int_0^\infty \frac{1}{\omega^2} \left[1 - \left(\frac{1}{1 + j\omega} \right)^N \right] \left[1 - \left(\frac{1}{1 - j\omega} \right)^N \right] d\omega$$

$$= \frac{1}{2\pi} \int_0^\infty \frac{1}{\omega^2} \left[1 - \left(\frac{1}{1 + j\omega} \right)^N - \left(\frac{1}{1 - j\omega} \right)^N \right.$$

$$\left. + \left(\frac{1}{1 + \omega^2} \right)^N \right] d\omega. \tag{59}$$

Let us introduce the dummy variable a and define

$$\Sigma(a) = \sum_{N=0}^\infty P(N) a^N$$

$$= \frac{1}{2\pi} \int_0^\infty \frac{1}{\omega^2} \sum_{N=0}^\infty \left[a^N - \left(\frac{a}{1 + j\omega} \right)^N - \left(\frac{a}{1 - j\omega} \right)^N \right.$$

$$\left. + \left(\frac{a}{1 + \omega^2} \right)^N \right] d\omega. \tag{60}$$

We recognize that each of the terms in the summation of (60) is a geometric series whose sum is given by

$$\sum_0^\infty r^n = \frac{1}{1 - r}. \tag{61}$$

Using this, we can write

$$\Sigma(a) = \frac{1}{2\pi} \int_0^\infty \frac{1}{\omega^2} \left[\frac{1}{1-a} - \frac{1}{1 - \dfrac{a}{1+j\omega}} - \frac{1}{1 - \dfrac{a}{1-j\omega}} \right.$$

$$\left. + \frac{1}{1 - \dfrac{a}{1+\omega^2}} \right] d\omega \tag{62}$$

$$= \frac{1}{2\pi} \int_0^\infty \left[\frac{a}{1-a} \frac{1}{\omega^2} - \frac{2a(1-a)}{\omega^2[\omega^2 + (1-a)^2]} \right.$$

$$\left. + \frac{a}{\omega^2[\omega^2 + (1-a)]} \right] d\omega .$$

All of the terms above can be integrated by using standard tables. The result is

$$\Sigma(a) = \sum_{N=0}^\infty P(N)a^N = \frac{1}{2} \left[\frac{a}{(1-a)^2} - \frac{1}{2} \frac{a}{(1-a)^{3/2}} \right]. \tag{63}$$

If we expand the right-hand term of (63) in a power series, the coefficients of the a^N will be the $P(N)$

$$P(N) = \frac{1}{2} \left[N - \frac{1}{2} \frac{(2N-1)!}{4^{(N-1)}[(N-1)!]^2} \right]. \tag{64}$$

For values of N greater than 5 we can approximate the factorials by Stirling's formula

$$n! \approx \sqrt{2n\pi} \, (n/e)^n. \tag{65}$$

The result is

$$P(N) \approx \frac{1}{2} \left(N - \frac{1}{\sqrt{\pi}} \frac{N - \frac{1}{2}}{\sqrt{N-1}} \right). \tag{66}$$

For $N > 10$ we can make the further approximation

$$P(N) \approx \frac{1}{2} \left(N - \frac{1}{\sqrt{\pi}} \sqrt{N} \right). \tag{67}$$

And for $N > 100$

$$P(N) \approx N/2. \tag{68}$$

Values of $P(N)$ from 1 to 100 are listed in Table II.

The mean-square jitter at the end of a long chain of repeaters is approximately [using (58) and (68)]

$$\overline{\theta_N{}^2} \approx \frac{\Phi B N}{2} \tag{69}$$

This last approximation can also be derived from direct integration of (9).

APPENDIX D

Growth of Alignment Jitter with Number of Repeaters in Chain

The alignment jitter at the nth repeater is given by (27)

$$\Theta_{aN}(s) = \Theta(s) \frac{1}{(1 + (s/B))^N}, \tag{70}$$

where $\Theta(s)$ is the jitter injected at each repeater.

Equation (70) can be written in the iterative form

$$\Theta_{aN}(s) = \Theta_{a(N-1)}(s) \frac{1}{1 + (s/B)} . \tag{71}$$

In the time domain equation (71) becomes:

$$\theta_{aN}(t) = \int_0^t \theta_{a(N-1)} (t - \tau) B e^{-B\tau} \, d\tau. \tag{72}$$

Then,

$$| \theta_{aN}(t) | \leqq M_{N-1}(1 - e^{-Bt})$$

or

$$| \theta_{aN}(t) | \leqq M_{N-1}, \tag{73}$$

where M_{n-1} is the maximum value of $| \theta_{a(N-1)}(t) |$ on the interval.

Carrying the iteration back to the first repeater, we have

$$| \theta_{aN}(t) | \leqq M_o(1 - e^{-Bt})^N$$

or

$$| \theta_{aN}(t) | \leqq M_o \tag{74}$$

where

$$M_o \geqq |\theta(t)|$$

where $\theta(t)$ is the jitter injected at each repeater.

Equations (73) and (74) state that the maximum alignment jitter does not increase as the number of repeaters grows and is never greater than the maximum jitter injected into each repeater. Since the jitter injected at each repeater is limited by the circuitry to a fraction of a cycle, alignment jitter can never be greater than a fraction of a cycle.

REFERENCES

1. Sunde, E. D., Self-timing Regenerative Repeaters, B.S.T.J., **36**, July, 1957, pp. 891–937.
2. DeLange, O. E., The Timing of High-Speed Regenerative Repeaters, B.S.T.J., **37**, November, 1958, pp. 1455–1486.
3. Bennett, W. R., Statistics of Regenerative Digital Transmission, B.S.T.J., **37**, November, 1958, pp. 1501–1542.
4. Rowe, H. E., Timing in a Long Chain of Regenerative Binary Repeaters, B.S.T.J., **37**, November, 1958, pp. 1543–1598.
5. Aaron, M. R., PCM Transmission in the Exchange Plant, B.S.T.J., **41**, January, 1962, pp. 99–141.
6. Kinariwala, B. K., Timing Errors in a Chain of Regenerative Repeaters, B.S.T.J., **41**, November, 1962, pp. 1796–1797.
7. Mayo, J. S., A Bipolar Repeater for Pulse Code Modulation Signals, B.S.T.J., **41**, January, 1962, pp. 25–97.

Satellite System Interference Tests at Andover, Maine

By HAROLD E. CURTIS

Because of the shortage of suitable frequencies, it appears that it will be necessary for satellite communication systems and ground microwave systems to operate in the same frequency band. To do this without excessive interference between them requires that stringent criteria be met. This paper presents the results of radio interference tests from two test transmitter sites into the Andover, Maine, earth station receiver for experimental satellite communications. Although these sites were only 23.5 miles and 55 miles from Andover, an analysis of the results indicated that they could be used as locations for ground microwave systems with few limitations.

I. INTRODUCTION

In order that satellite communication systems and terrestrial microwave systems may share successfully the same frequency band, it is in general necessary that the earth station of the satellite system and microwave station be separated physically so that line-of-sight transmission between them is not possible. Under these conditions, either the scatter mode or diffraction mode of transmission predominates, depending on the path profile. While a considerable amount of data on such transmission is available with transmitting and receiving antennas pointed at one another and only elevated sufficiently to graze the horizon, very little data have been taken with one of the antennas, the earth station antenna in this case, randomly oriented in azimuth and elevation.

Predictions[1,2] have been made that it should be possible to operate point-to-point microwave systems in the same radio frequency band as the satellite system at distances of some 45 to 150 miles removed from the earth station, or even less, provided that (a) the earth station antenna is not operated at an elevation angle less than some minimum value, for which 7.5 degrees has been suggested, (b) suitable terrain blocking exists, and/or (c) a certain amount of angular discrimination is provided by the antenna of the radio relay terrestrial station.

Thus, one purpose of the tests to be described and evaluated was to examine critically the received interfering signal strength as a function of elevation and azimuth angle of the earth station antenna beam. A second purpose was to record for an extended length of time the received signal with the earth station antenna elevated in angle somewhat above the elevation value for maximum received interference to see if reflections from cloud masses, airplanes, birds, etc., could produce signal enhancements of sufficient magnitude and frequency of occurrence to degrade significantly the performance of the satellite communication system.

The American Telephone and Telegraph Company already had commercial operating sites at Cornish, Maine, and at West Paris, Maine — 55 miles and 23.5 miles, respectively, from the earth station at Andover, Maine — and these appeared to be suitable locations at which to place test transmitters for the interference tests. A contour map of the area is shown on Fig. 1. The interference path profiles (Figs. 2 and 3) indicated that these sites would be appropriate for evaluating the required mimimum antenna discrimination toward Andover of an operating TD-2 station[3] at these sites and/or the required limitation on minimum antenna elevation at Andover for compatible operation on a frequency sharing basis.

Although the results of these tests and the conclusions based thereon may not be valid for other sites and equipments, they do show that it is possible by careful choice of the earth station site to operate microwave radio stations in the same frequency band without undue physical separation between them.

Because of the considerable international interest in the question of frequency sharing, the criteria of permissible interference are those recommended by the Xth Plenary Assembly of the International Radio Consultative Committee (C.C.I.R.) held in 1963.

II. RECORDING SETUP

A sketch of the experimental setup for recording the received signal at Andover for the various test conditions is shown on Fig. 4. It consisted of the earth station antenna whose output was fed to a maser followed by a receiver-converter, pre-amplifier and main IF amplifier normally associated with the earth station receiver. At this point a 4170-mc RF input signal appeared as a signal whose nominal frequency was approximately 74.1 mc. The dynamic range of the entire receiver from antenna to strip chart was approximately 40 db when the gains of the various amplifiers were properly adjusted. This IF signal was then
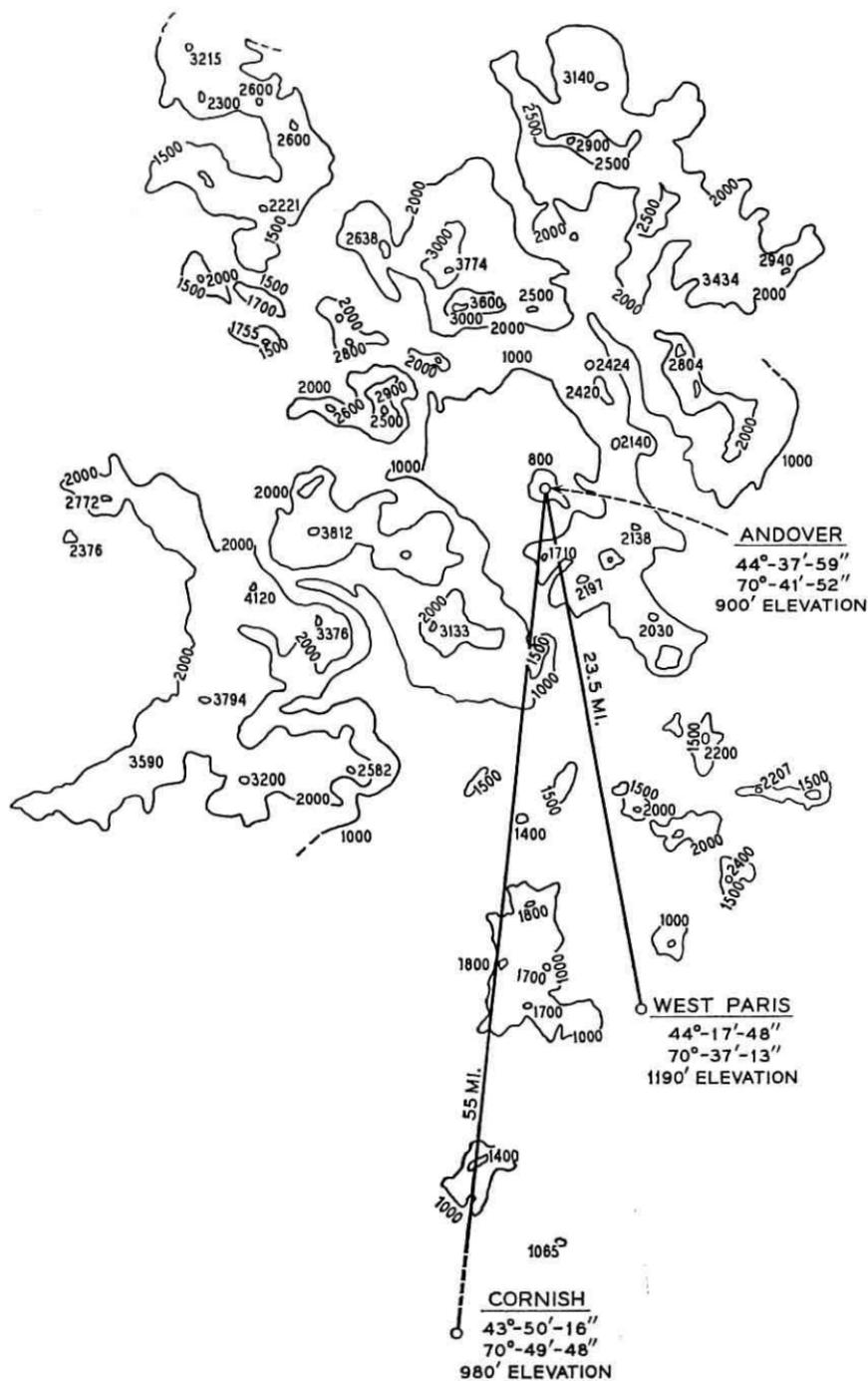
Fig. 1 — Topography surrounding site at Andover, Maine.

2717
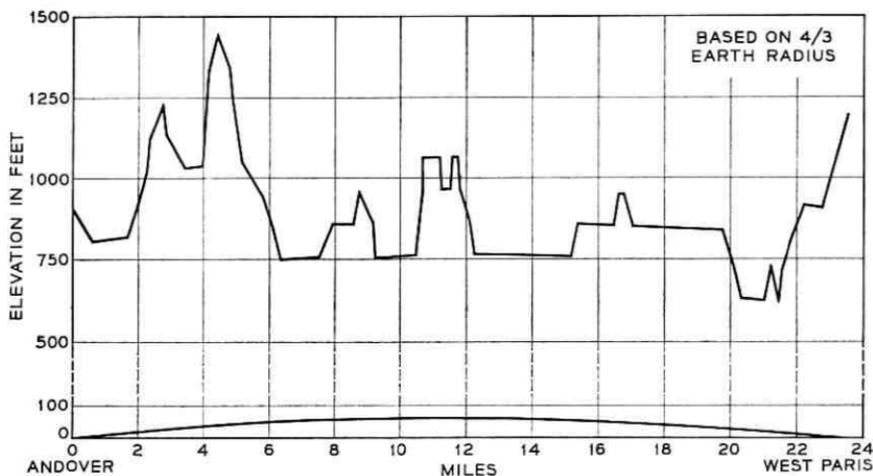
Fig. 2 — Profile of West Paris-Andover path.

detected by a variable frequency analyzer with a 10-kc bandwidth and the rectified output recorded on a strip chart after passing through a logarithmic amplifier. No automatic gain control was used, so that the amplitude of the 74.1-mc signal was directly proportional to the amplitude of the 4170-mc input signal, within the boundaries of noise and overload.

Provision was made, as shown on Fig. 4, for calibrating the recorder in terms of received RF power. The chart range was approximately 40 db, but the over-all sensitivity of the setup could be changed, if required,
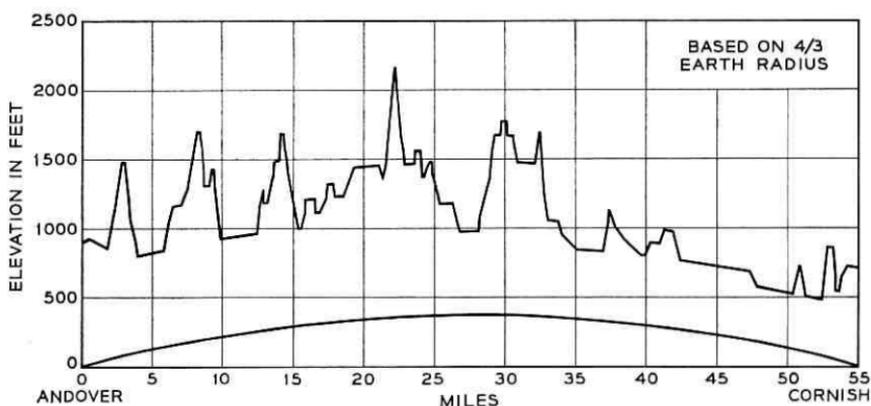
Fig. 3 — Profile of Cornish-Andover path.
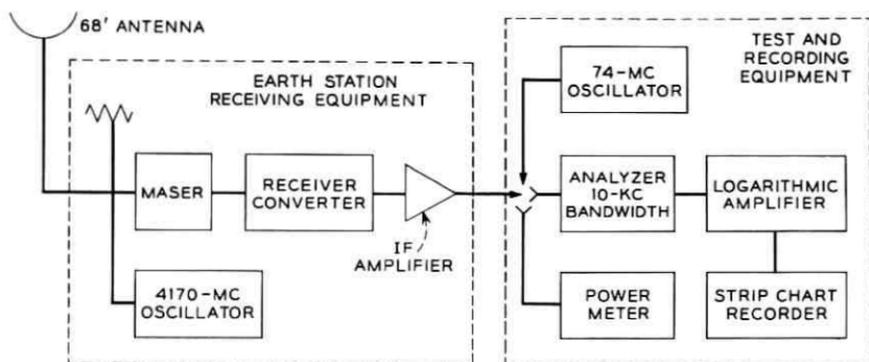
Fig. 4 — Test setup.

by the insertion of an IF pad of suitable value. The total range covered was from about −142 dbm (noise only) to about −70 dbm. The frequency response of the receiving system was limited by the strip chart recorder and was in the order of 60 cycles per second.

It will be noted that the gain of the RF-IF portion and the sensitivity of the analyzer-recorder portion of the system were calibrated separately. This was necessary because the RF calibrating oscillator was not stable enough in frequency to stay within the 10-kc band of the analyzer. However, the frequencies of the test transmitters at Cornish and West Paris were adequately stable, and only occasional checks of the tuning of the analyzer were required during an extended run.

All of the values of received signal power given herein are, strictly speaking, referred to the input point of the maser amplifier. In view of the fact that the loss of the waveguide and associated circuitry connecting the maser to the receiving antenna is only about 0.1 db, they are also essentially the same as would have been observed at the antenna output. Table I presents information pertinent to these tests.

III. PATH PROFILES

The profiles of the West Paris–Andover and Cornish–Andover paths are shown on Figs. 2 and 3, respectively. An analysis of the first path indicates that, because of the mountain 4.4 miles from Andover, the diffraction mode would be expected to predominate over the scatter mode. The opposite is true for the Cornish–Andover path.

The Andover antenna, as will be noted on Fig. 1, is located in a relatively flat area surrounded by low mountains. Elevation scans of noise only were available at 5-degree intervals in azimuth using the Andover

TABLE I — PERTINENT INFORMATION

Station	Andover, Maine	West Paris, Maine	Cornish, Maine
Status	Earth Station	Test transmitter	Test transmitter
Latitude	44° 37′ 59″	44° 17′ 48″	43° 50′ 16″
Longitude	70° 41′ 52″	70° 37′ 13″	70° 49′ 48″
Path length	—	23.5 miles	55.0 miles
Frequency source	—	Crystal oscillator	Crystal oscillator
Frequency, RF	—	4170 mc	4170 mc
Antenna diameter (actual)	68 feet	10 feet	16 feet
Effective radiated power	—	+64 dbm	+78 dbm
Antenna gain (free-space)	54 db*	40 db	43 db
Orientation			
Azimuth	Variable	Toward Andover	Toward Andover
Elevation	Variable	Toward horizon	Toward horizon
Bandwidth of test setup	10 kc	—	—
Noise (at zenith)	−142 dbm	—	—
Elevation†	—	1.57°	2.43°
Azimuth†	—	170.69°	184.92°

* The Andover antenna is arranged normally to receive a circularly polarized wave, and for this condition the maximum response is 57 db referred to an isotropic antenna. The response of the antenna to a linearly polarized incident wave under the same condition is 3 db less.

† These are the elevation and azimuth angles of the Andover, Maine, antenna for maximum received signal as determined experimentally. The elevation angle is referred to the horizontal in all cases unless specifically stated otherwise.

antenna, and the elevation angle of the antenna at which the noise abruptly changed was noted as the "electrical" horizon. A 360-degree azimuth profile thus obtained is shown on Fig. 5.

IV. ANTENNA DIRECTIVITY PATTERNS

Figs. 6 and 7 show vertical and horizontal directivity patterns about the main beam of the 68-foot horn-reflector antenna at Andover, as measured by J. N. Hines of Bell Telephone Laboratories. A circularly polarized signal was transmitted from a reference point at Black Mountain about 4.9 miles away, and the vertical component of the electrical field out of the antenna was recorded as the beam of the antenna at Andover was swung in azimuth or raised in elevation about the bearing of the reference point. The azimuth and elevation angles of the reference point relative to the Andover antenna were 142.097 degrees and 3.977 degrees, respectively, and the path was line-of-sight.

The maximum response of the antenna arranged to receive in this fashion is about 54 db referred to an isotropic antenna. It will be noted
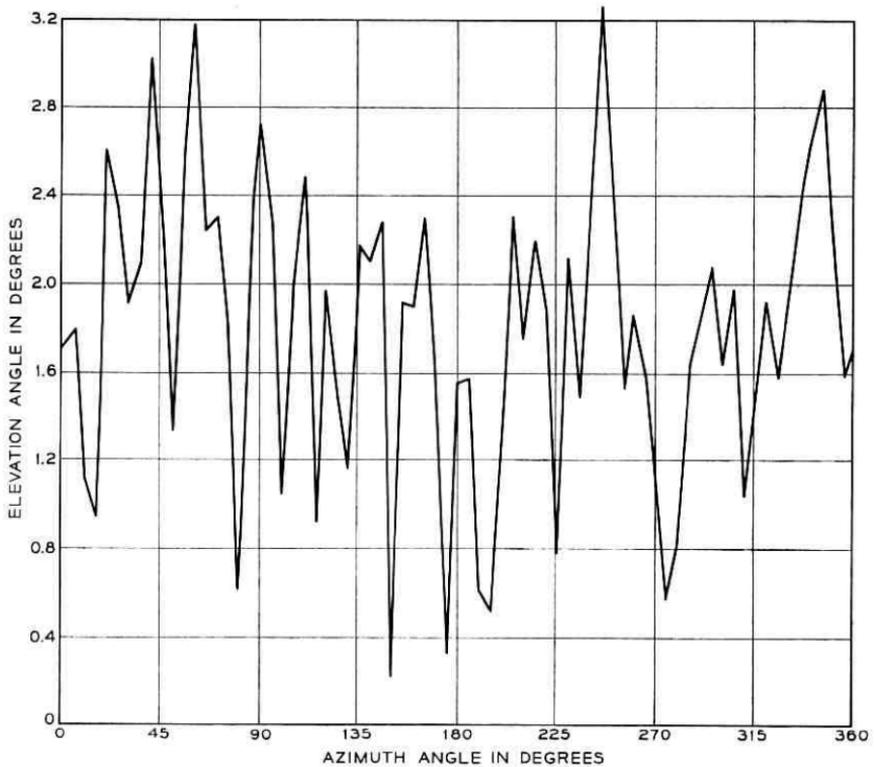
Fig. 5 — Azimuth profile based on noise measurements.

that the antenna beam had to be elevated about 10 degrees relative to the reference direction, in order to reduce the gain to that of an isotropic antenna.

Since the noise power in the measuring equipment was also about 54 db below the maximum received signal, the maximum value of antenna directivity discrimination that could be measured was also limited to this same value.

V. EXPECTED RECEIVED SIGNAL

5.1 *West Paris–Andover Path*

    1. Transmitter power   =   +27 dbm
    2. Waveguide and filter loss   =   3 db
    3. Antenna gain – vertical polarization   =   40 db
    4. Effective radiated power   =   +64 dbm
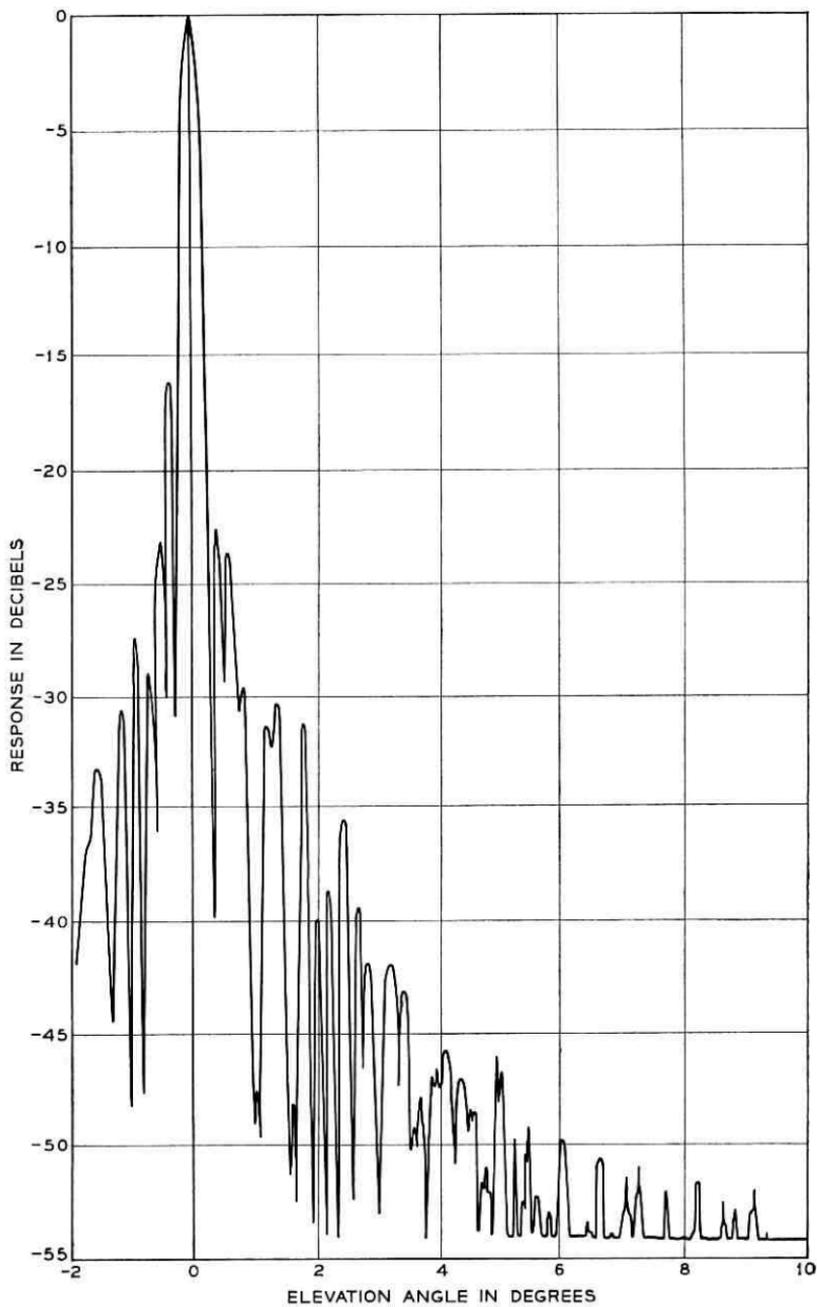
Fig. 6 — Antenna directivity in vertical plane.
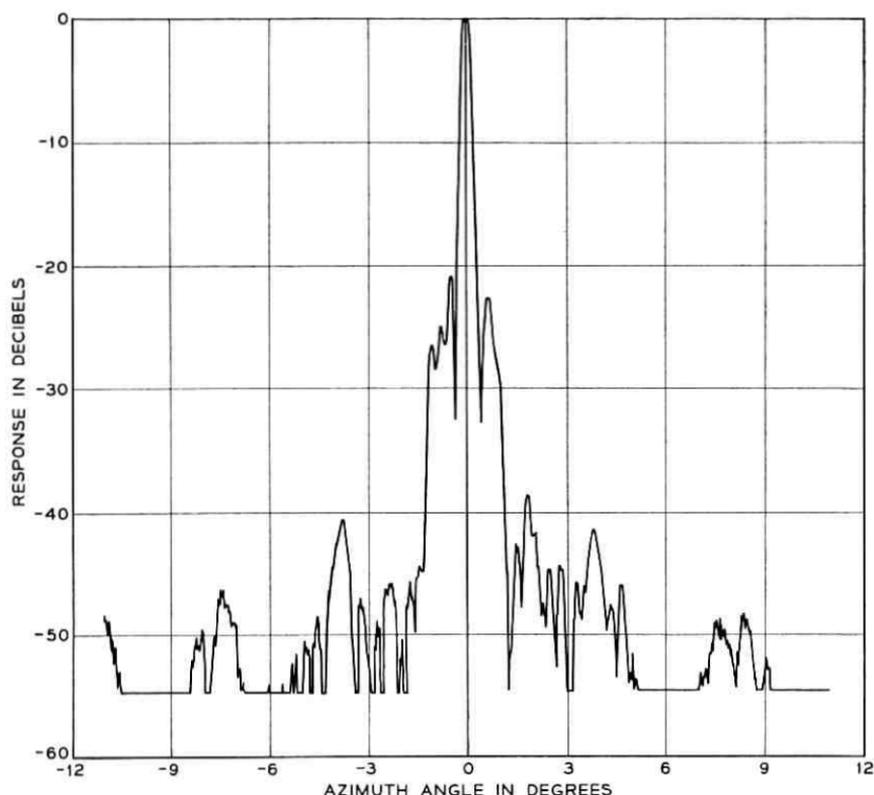
Fig. 7 — Antenna directivity in horizontal plane.

5. Estimated knife-edge diffraction basic transmission loss* (Median *scatter* loss estimated as 198 db and hence not controlling.)   =   168 db

6. Receiving antenna gain (see note on Table I)   =   +54 db

7. Estimated signal power at antenna output on basis of knife-edge diffraction   =   −50 dbm

5.2 *Cornish–Andover Path*

1. Transmitter power   =   +37 dbm

2. Waveguide and filter loss   =   2 db

---

* Defined in Ref. 4, "Basic Transmission Loss (L_b). The basic transmission loss (sometimes called path loss) of a radio circuit is the transmission loss expected between ideal, loss-free, isotropic, transmitting and receiving antennae at the same locations as the actual transmitting and receiving antennae."

3. Path antenna directivity gain.* Experience on
   similar scatter paths indicates that the effective
   antenna gains on such paths are substantially less
   than the free-space gains of 54 db and 43 db for
   the two antennas. This reduction, which is usually
   referred to as the antenna-to-medium coupling
   loss, is estimated to be 11 db for the larger antenna
   and 3 db for the smaller antenna. The path an-
   tenna directivity gain is therefore   =   83 db
4. Estimated median basic transmission loss (scatter
   propagation) during November   =   220 db
5. Estimated received signal power at antenna out-
   put on basis of scatter propagation   =   −102 dbm

The transmitter power at a standard TD-2 station[3] is +27 dbm, and
the waveguide and filter losses total about 3.5 db. A horn-reflector an-
tenna of 10-foot aperture and gain of approximately 40 db at 4200 mc
is also standard equipment. The transmitted wave is linearly polarized.
Thus the effective radiated power at Cornish was about 15 db greater
than that of a normal TD-2 transmitter.

VI. EXPERIMENTAL RESULTS

6.1 *West Paris–Andover*

As noted above, the predominant mode of transmission over this path
was expected to be by diffraction rather than by scatter, and hence a
comparatively steady signal was expected.

The principal tests over this path may be divided into two parts, (*i*)
"fine grain" elevation runs at the azimuth angle corresponding to maxi-
mum signal when "on-beam," and (*ii*) "fine grain" azimuth scans at vari-
ous fixed elevation angles. Certain runs, each of about one hour in dura-
tion, were made at various fixed elevations, all on the azimuth bearing
of maximum received signal, to establish the fact that the signal was
steady in each case.

The maximum received signal on this path ranged from about −75
dbm to −80 dbm during the various days of the tests. The expected
signal as computed above, based on knife-edge diffraction, was −50

---

* "Path Antenna Directivity Gain (G_p). The path antennae directivity gain is
equal to the increase in the transmission loss when lossless, isotropic antennae
are used at the same locations as the actual antennae." Transmission loss $L = L_b - G_p$. See Ref. 4.

dbm, indicating that the intervening hill was considerably more effective than an ideal knife edge in attenuating the signal. On the basis of the received signal levels the basic transmission loss, as defined above, was about 193 db for the West Paris–Andover path. Since this was fairly constant during the period of any single test, the variations in received signal power as the direction of the beam of the Andover antenna was changed reflected changes in the path antenna directivity gain.

In view of the fact that the basic purpose of this study is the evaluation of three specific sites from the standpoint of frequency sharing between satellite communication systems and ground microwave systems, the results are plotted in terms of actual received interfering power at the site in question.

Figs. 8 and 9 show fine structure plots of vertical scans during clear weather and during a heavy rain, respectively. Fig. 10 presents a smooth curve of the envelope of the peaks of the interference up to an elevation of 50 degrees.

Figs. 11 and 12 are fine grain azimuth scans at fixed elevations of zero and +2 degrees, relative to that of maximum signal, i.e., 1.573 degrees above the horizontal. Fig. 13 is a similar scan with the antenna elevated 8.573 degrees above the horizontal.

Finally, Figs. 14 and 15 present 360-degree azimuth scans at fixed
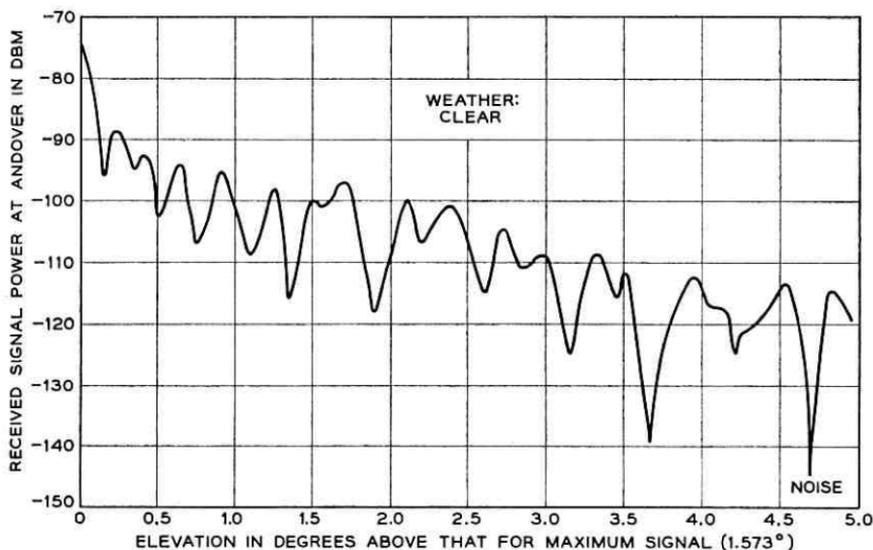
Fig. 8 — Elevation scan, West Paris transmitter activated (vertical scan showing fine structure). Signal is steady at any elevation of antenna.
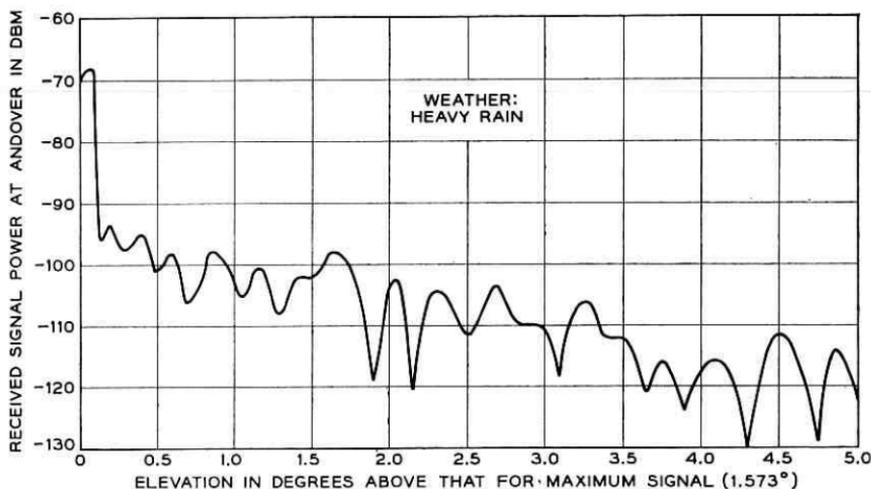
Fig. 9 — Elevation scan, West Paris transmitter activated (vertical scan showing fine structure). Signal is steady at any elevation of antenna.

elevations. In these figures the maximum signal values in sectors of various widths are plotted. It will be noted on Fig. 14 that the signal, when the antenna was directed at about 180 degrees from the direction of West Paris, was only about 10 db below the signal received when pointed directly at West Paris. A similar reflection down 14 db was noted when the Cornish transmitter was activated and the Andover antenna rotated through 360 degrees of azimuth at an elevation angle of 2.43 degrees. The source of the reflections is as yet unidentified.
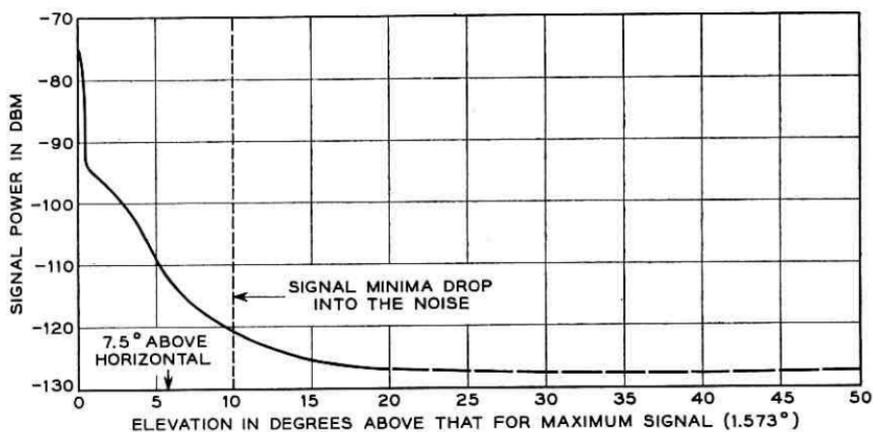
Fig. 10 — Elevation scan, West Paris transmitter activated (envelope of peaks of maximum signal).
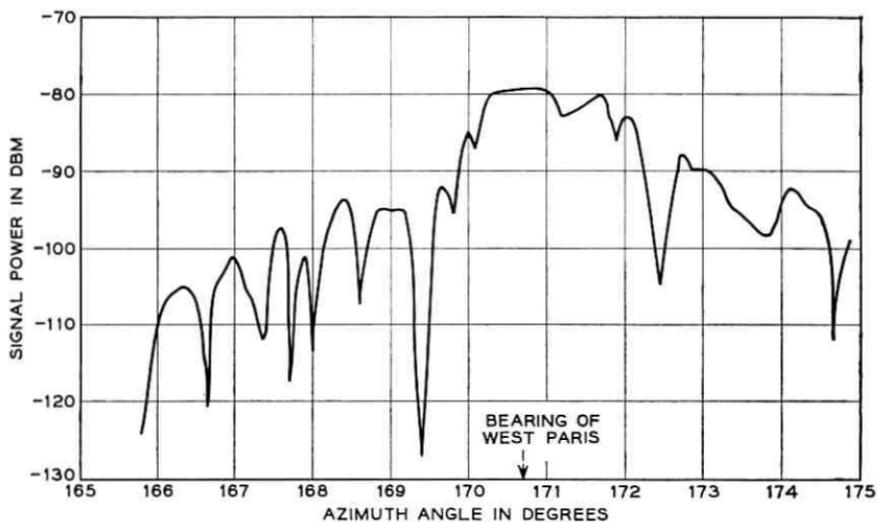
Fig. 11 — Azimuth scan, West Paris transmitter activated (receiving antenna elevated for maximum signal, i.e., 1.573° above the horizontal).
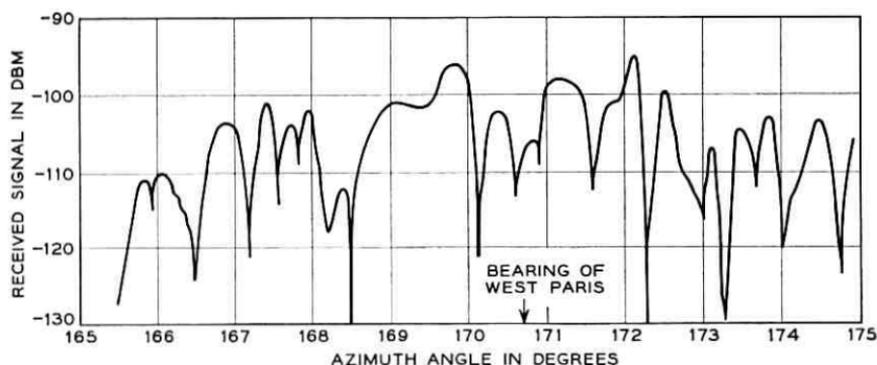
Fig. 12 — Azimuth scan, West Paris transmitter activated (receiving antenna elevated 2° above that for maximum signal, i.e., 3.573° above the horizontal).
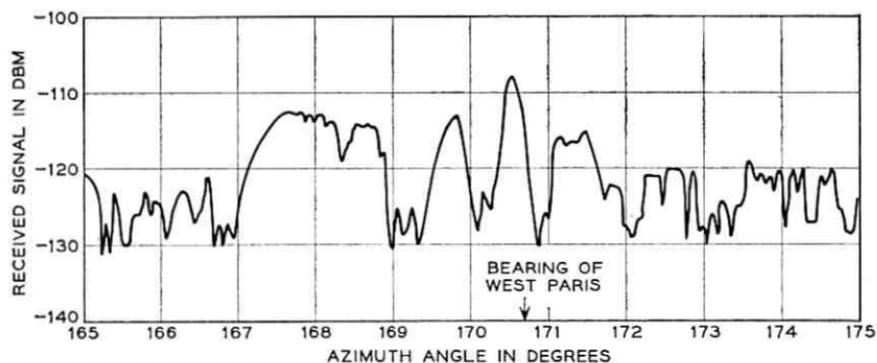
Fig. 13 — Azimuth scan, West Paris transmitter activated (receiving antenna elevated 7° above that for maximum signal, i.e., 8.573° above the horizontal).
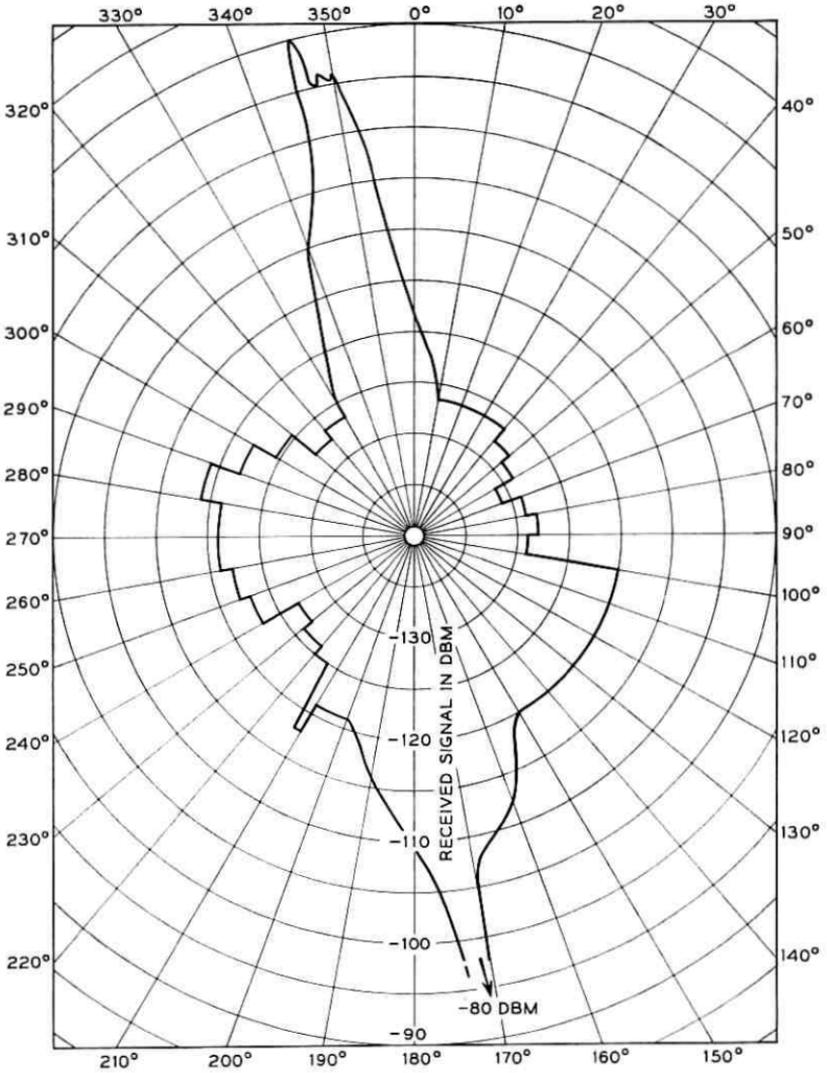
Fig. 14 — Azimuth interference profile, West Paris transmitter activated (receiving antenna elevated for maximum signal when "on-beam," i.e., 1.573° above the horizontal).

6.2 *Cornish–Andover, Maine*

As mentioned above, examination of the profile of the path indicates that propagation by scatter would be controlling, with an expected median basic transmission loss of 220 db between isotropic antennas for the time of year the measurements were made.

In addition to the elevation and azimuth scans such as were also made on the West Paris–Andover path, runs totaling 107 hours were made with the Andover antenna elevated in angle sufficiently to bring the median signal down close to the level of the noise in order that periods of possible signal enhancement could be accurately observed. The azimuth angle at all times was that for maximum signal when "on-beam," i.e., 184.92 degrees.

When on the coordinates of maximum interference, the signal was steady within a db, and measured about −108 dbm at the input of the maser, as compared with the expected value of −102 dbm. Thus the basic transmission loss of the path at this time was 226 db based on an estimated path antenna directivity gain of 83 db. Elevating the receiving antenna less than a degree altered the character of the signal so that it varied continually about ±10 db. Typical portions of the strip chart traces are shown for several antenna elevations on Fig. 16.

Fig. 17 shows the results of an elevation run giving median values of
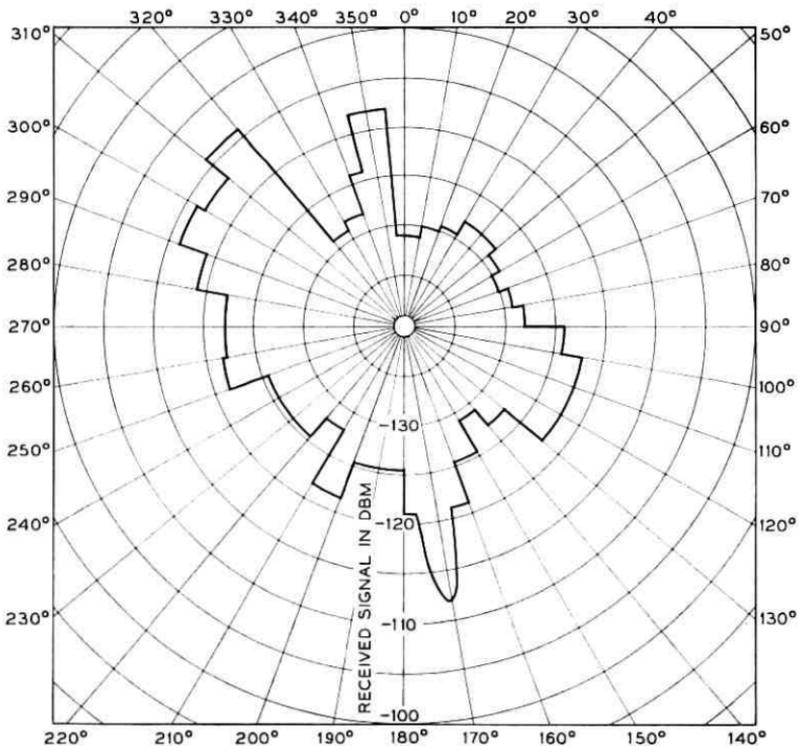
Fig. 15 — Azimuth interference profile, West Paris transmitter activated (receiving antenna elevated 7.5° above the horizontal).
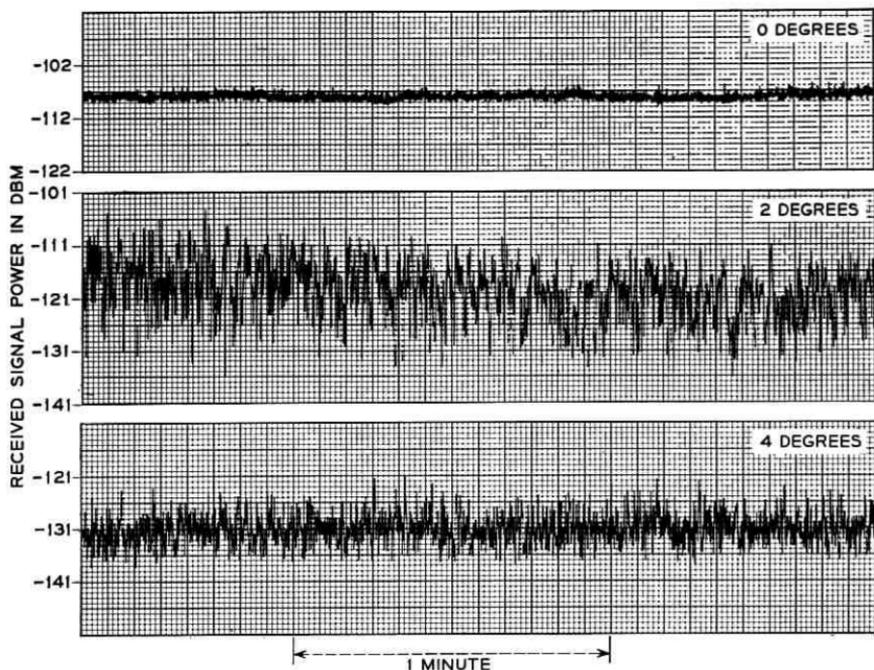
Fig. 16 — Received signal from Cornish for various antenna elevations at Andover (elevation angles are in degrees above that for maximum signal).

interference at intervals of 0.2 degree in elevation. The variation in signal strength was so great that it was impossible to reset the antenna elevation to a particular angle and obtain the same interference level to within the order of ±5 db. A reproduction of a portion of the strip chart from which the results on Fig. 17 were obtained is shown on Fig. 18. Here the rate of elevation scan is 0.043 degree per second. Also shown on Fig. 18 for comparison is an elevation scan made at Andover with the West Paris transmitter activated.

An azimuth scan with the antenna beam elevated 7.5 degrees above the horizontal with the Cornish transmitter activated showed that the interference power was below the noise, i.e., −141 dbm, for all angles except in the direction of Cornish. Here the interference power reached a maximum of −132 dbm.

Fig. 19 shows a running plot of hourly median received signal power at Andover extending over a period of nine days, the actual measuring time totaling 107 hours. Furthermore, since the main interest was in signal enhancement, several different elevation angles were used during

the period. Thus the changes in average hourly median from day to day reflect to a considerable extent changes in path antenna directivity gain.

Signal enhancements in which the median power rose 7 db or more for brief periods were individually examined, and the median value at the time of maximum rise relative to the normal median signal level at the time was tabulated. The duration of each enhancement at a level
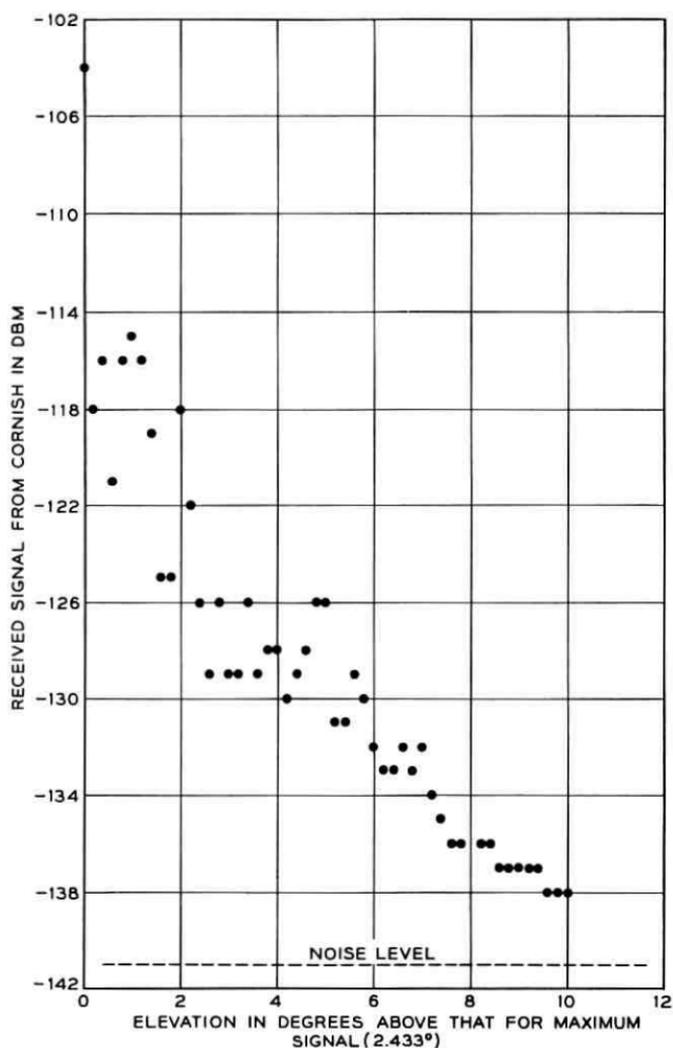
Fig. 17 — Elevation scan, Cornish transmitter activated (Andover antenna directed toward Cornish; data from Fig. 18).
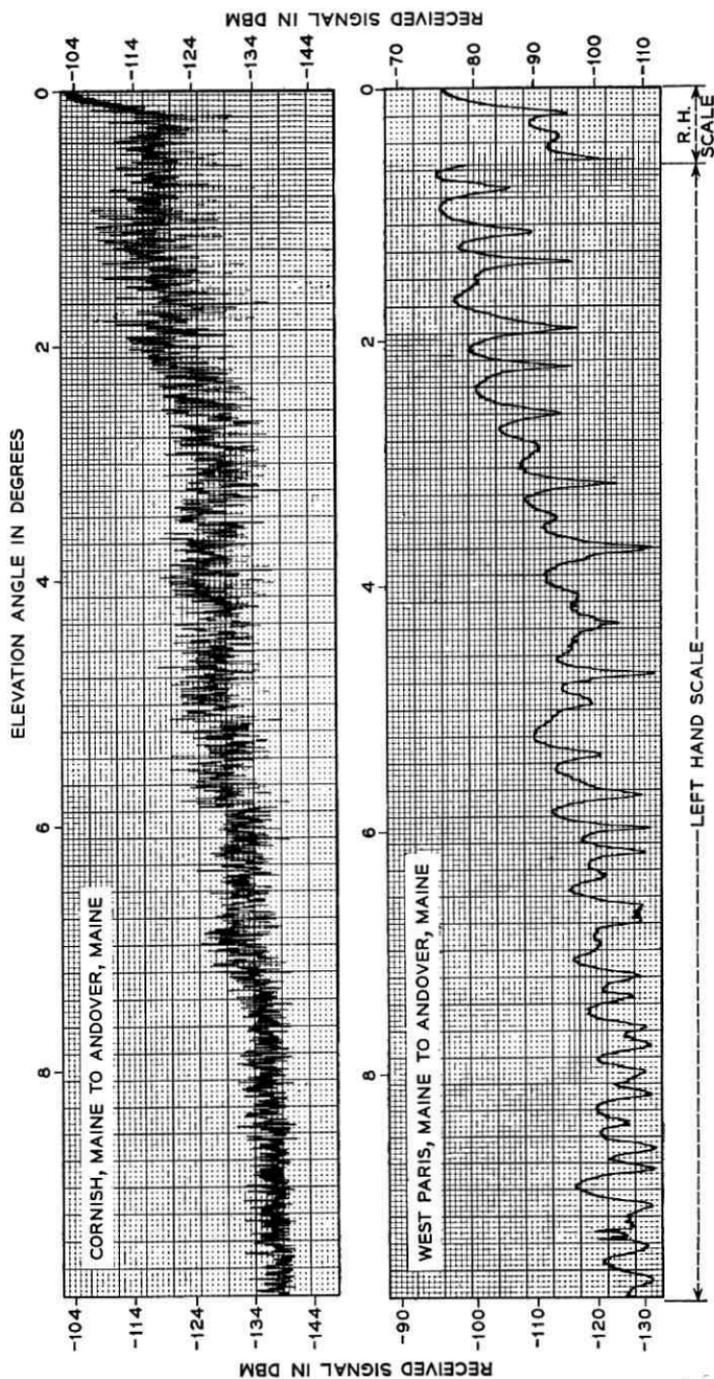
Fig. 18 — Elevation scans (elevation angles are in degrees above that for maximum signal in each case; azimuth angles are fixed at value giving maximum signal at reference elevation in each case; elevation scan rate is 0.043° per second).
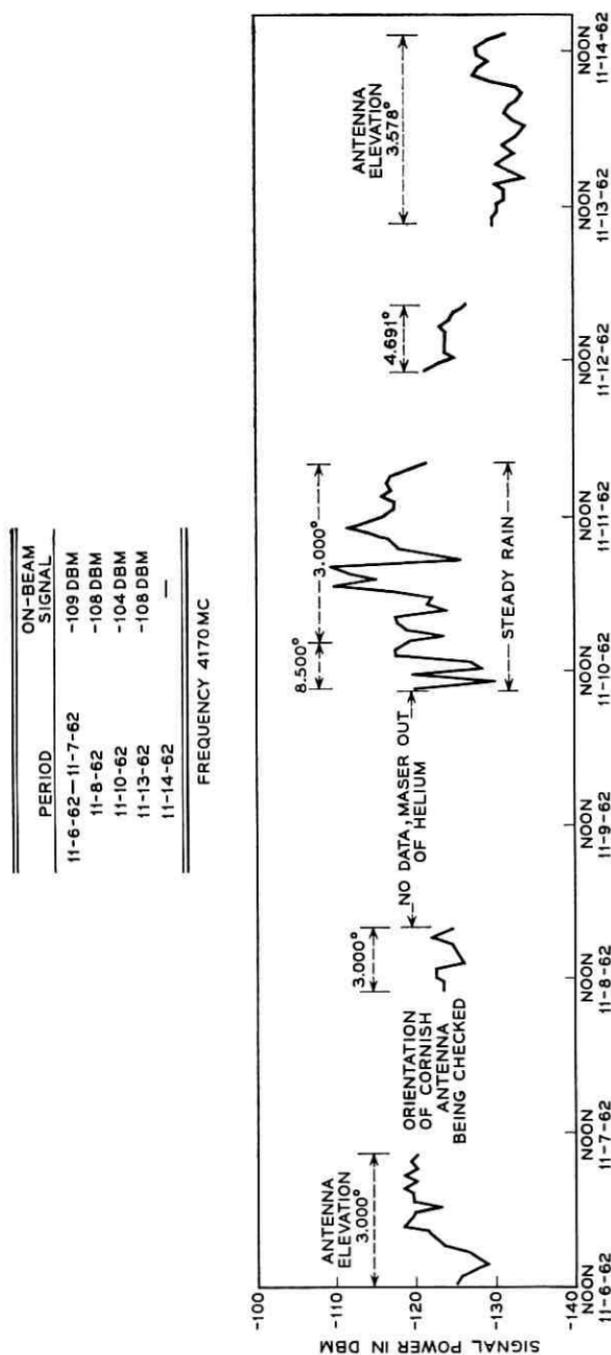
Fig. 19 — Hourly median received signal at Andover from Cornish (elevation angles are referred to that for maximum received signal, i.e., 2.433° above the horizontal).

halfway (in db) from the normal median to the peak median value was also recorded.

The short-term, e.g., five seconds, median varies normally about ±3 db in any brief period such as an hour, and such normally expected increases associated with these variations were not classified as enhancements. Three typical periods that were classed as enhancements are shown on the strip chart reproductions on Fig. 20. Here the chart speed is one small division per second. In the top example on this figure the average median is taken as −121 dbm, and the median at the peak as −106 dbm, and the enhancement is thus 15 db. The enhancement in the middle example is also 15 db, and for the bottom is 20 db.

A total of 56 enhancements thus defined occurred in the 107-hour period. Fig. 21 shows a plot of the cumulative percentage of the total
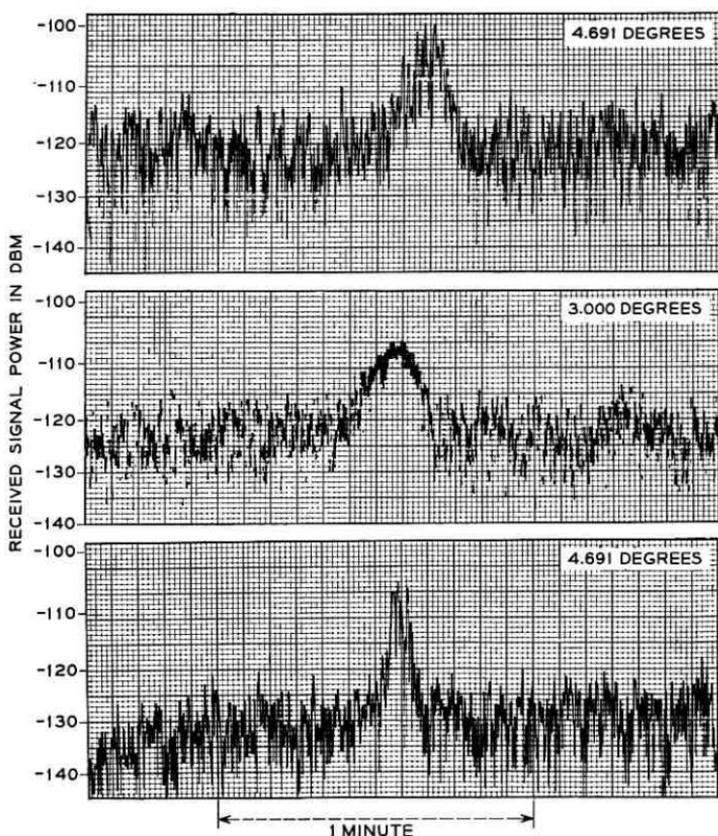
Fig. 20 — Typical periods of signal enhancements — Cornish-Andover path (elevation angles are in degrees above that for maximum signal).
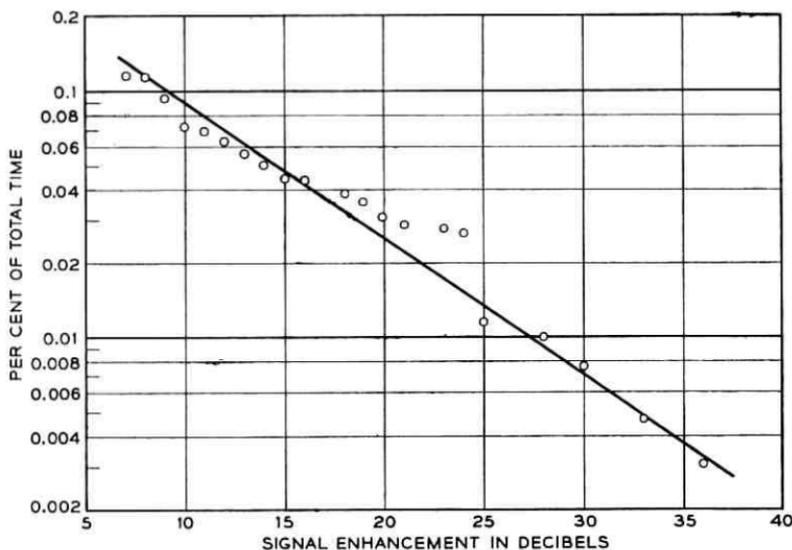
Fig. 21 — Probability of signal enhancements — Cornish-Andover path (cumulative percentage of time the signal strength was enhanced any given amount above the median, normal variations in median excluded; duration of observations 107 hours).

period of 107 hours the signal was enhanced, as defined above, any given amount or more above the median. Thus 0.01 per cent of the 107 hours the signal was enhanced 27 db or more above the median signal value.

Fig. 22 shows the distribution of the duration of signal enhancements.

VII. INTERPRETATION OF RESULTS AND CONCLUSIONS

7.1 *Cornish–Andover Path*

Recommendation No. 356* of the Xth Plenary Assembly of the C.C.I.R. held in 1963 provides that the mean value of interference from radio relay systems into an earth station receiver in any hour should not exceed 1000 picowatts, psophometrically weighted. On the basis of four exposures into the earth station receiver, the message channel interference per exposure is 250 picowatts or 18 dba at reference level.

The signal-to-interference ratio at baseband may be related to the ratio of the desired carrier to the interfering carrier by the Receiver Transfer Characteristic as defined in Recommendation No. 356. The

* This will appear in the forthcoming publications of the Xth Plenary Assembly, Geneva, 1963.
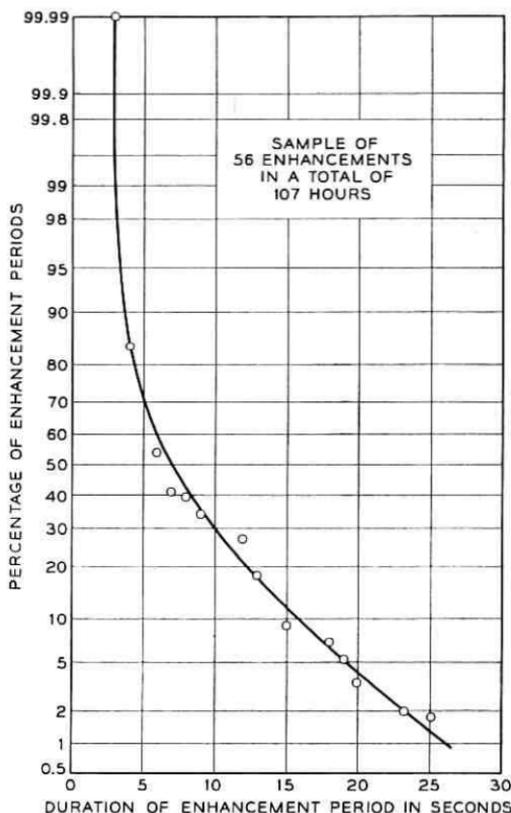
Fig. 22 — Duration of signal enhancements — Cornish-Andover path (percentage of enhancement periods having durations equal to or greater than values shown on abscissa; sample of 56 enhancements in a total of 107 hours).

magnitude of the Receiver Transfer Characteristic depends upon the frequency separation of the two carriers, the characteristics of the modulation, and the radio receivers themselves. Fig. 3 of Ref. 2 shows computed values of this relationship for a possible satellite system, a TD-2 system and a TH system.[5] Here the signal is Gaussian noise having the same power per channel as an average talker, and the interference is expressed in dba at a zero-db transmission level point, or reference level. Thus we find that the mean value of the ratio of the desired carrier to the interfering carrier should be 25 db or greater in order that the baseband noise not exceed 18 dba at reference level. A reasonable value for the received carrier power at the earth station receiver from the satellite is about −98 dbm. The interfering carrier should then be −123 dbm or less.

A signal power of about -128 dbm was observed on the average when the Andover antenna was directed at the Cornish transmitter and elevated 7.5 degrees above the horizontal. Since the power into the transmitting antenna was $+35$ dbm, the transmission loss is 163 db. Because of higher power and antenna gain, the effective radiated power of the Cornish test transmitter was about 15 db greater than that of a standard TD-2 transmitter. Therefore it may be concluded that a TD-2 transmitter could be operated at Cornish, Maine, with its antenna directed toward the earth station, and the mean interference during the period of measurement would have been 20 db below the permissible value recommended by the C.C.I.R.

Recommendation No. 356 also provides that the message channel interference into an earth station receiver should not exceed 80,000 picowatts, psophometrically weighted, or 43 dba at reference level, for more than 0.02 per cent of any month. On the basis of four interferences per satellite ground station, the probability of occurrence per exposure should not exceed 0.005 per cent.

Reference to Fig. 21 shows that enhancements of 33 db were observed during the test period with a probability of 0.005 per cent. This leads to an expected interference value of -110 dbm, or a ratio of desired to interfering carrier power of 12 db, at the antenna output of the satellite ground receiver. This corresponds to 31 dba, or 5000 picowatts, a value well within the above objective developed from the C.C.I.R. recommendation.

The earth station transmitter at Andover operates in the 6-kmc common carrier band, and therefore it could potentially interfere with a common carrier receiver such as the TH microwave system operating in this band. While no companion interference measurements were made in this frequency range, it is quite practical to estimate with adequate accuracy the interference from Andover that might fall into a TH receiver that might be located at West Paris or at Cornish.

Recommendation No. 357 of the Xth Plenary Assembly of the C.C.I.R. provides that the interference from an earth station transmitter into a radio relay system should not exceed in any hour 1000 picowatts, psophometrically weighted. Again allowing four interferences per relay route from such a source, the interference per station becomes 18 dba at reference level. Fig. 3 of Ref. 2 shows that for this value of interference from the earth station transmitter, the mean value of carrier-to-interference ratio at the converter of the TH receiver should not be less than 58 db, providing the earth station transmitter frequencies are interleaved between the TH frequencies.

Since the normal received carrier power of the TH receiver is -27 dbm, the interfering carrier must be -85 dbm or less. Although the measurements were made at 4170 mc, the estimated system loss is less than one db different at 6000 mc. Thus the transmission loss with the earth station antenna elevated 7.5° is about 163 db, as before. Therefore, the power into the earth station antenna at Andover could be as great as 48 dbw or 63,000 watts without exceeding the mean interference recommendation.

Recommendation No. 357 also provides that the interference into a relay system should not exceed 50,000 picowatts psophometrically weighted one minute mean power for more than 0.01 per cent of any month. On the basis of four exposures, this reduces to 50,000 picowatts for not more than 0.0025 per cent of any month.

The latter permissible power recommendation is 23 db greater than that for the per-exposure objective for a mean power of 250 picowatts as developed above. Fig. 21 shows that signal enhancements of 38 db were observed for 0.0025 per cent of the observational period. This implies that for these specific sites the 0.01 per cent recommendation is 15 db more restrictive on earth station power than the mean interference recommendation. This leads to a maximum transmitter power of 33 dbw or 2000 watts.

If the antenna at Cornish were not beamed directly at Andover, the maximum transmitter power at the latter site could be raised by the antenna directivity loss thus obtained. For the 10-foot horn-reflector antenna, this amounts to over 40 db for 10 degrees of angle.[6]

Furthermore, with a nonstationary satellite the probability of the earth station antenna being oriented at the interfered-with radio relay receiver, as assumed above, is quite low. This will reduce substantially the probability of the interference reaching 50,000 picowatts, and in turn will lead to a maximum transmitter power considerably greater than 2000 watts.

7.2 *West Paris–Andover Path*

The separation between the West Paris test transmitter and the Andover station was only 23.5 miles, and the median received signal with the antenna at the latter site elevated 7.5 degrees above the horizontal was about -112 dbm. With a standard TD-2 transmitting arrangement the power would have been about one db less or -113 dbm. This lacks 10 db of meeting the objective of -123 dbm developed above from the C.C.I.R. recommendation for mean interference in any hour.

This additional transmission loss could be provided by about 2 degrees of antenna discrimination[6] at West Paris.

A TH receiver operating at West Paris potentially could suffer interference from an earth station transmitter operating in the 6-kmc common carrier band. The basic transmission loss in this band is estimated to be within about one db, as in the 4-kmc band. Using the objective of −85 dbm developed above, and with the West Paris antenna 2 degrees off beam, the permissible power at Andover, as far as interference at West Paris is concerned, could be about 31 dbw.

No long-term data were obtained on this path from which one might predict the probability of occurrence of short-duration signal enhancements. However, it is quite possible that these might limit the maximum tolerable power at Andover to a lower value than that derived on the basis of tolerable mean interference.

REFERENCES

1. Response of the A. T. & T. Co. to the F.C.C. "In the Matter of an Inquiry into the Allocation of Frequency Bands for Space Communications." Docket No. 13522.
2. Curtis, H. E., Interference Between Satellite Communication Systems and Common Carrier Surface Systems, B.S.T.J., **41**, May, 1962, p. 921.
3. Roetken, A. A., Smith, K. D., and Friis, R. W., The TD-2 Microwave Radio Relay System, B.S.T.J., **30**, October, 1951, p. 1041.
4. Defined in Recommendation No. 241, IXth Plenary Assembly of the C.C.I.R., Los Angeles, 1959.
5. The TH Microwave Radio Relay System, B.S.T.J., **40**, November, 1961, entire issue.
6. Curtis, H. E., Radio Frequency Interference Considerations in the TD-2 Radio Relay System, B.S.T.J., **39**, March, 1960, p. 369.
7. For a detailed description of the Andover, Maine, earth station receiver with its antenna, the reader is referred to the Telstar issue of the B.S.T.J., **42**, July, 1963, which appeared after the present paper was submitted.

The Nature of and System Inferences of Delay Distortion Due to Mode Conversion in Multimode Transmission Systems

By S. E. MILLER

Quantitative estimates of delay distortion due to mode conversion in a multimode medium are made using an analysis based on modes coupled through power-transfer coefficients. This results in a simple translation from the spatial distribution of mode conversion to delay distortion without an intermediate step in the frequency domain. The expected value of the reconversion magnitude and its delay distribution relative to a driving impulse are found for (i) the case where the undesired mode loss is distributed (helix damped modes and higher-order circular electric modes) and (ii) the case where discrete mode filters are inserted (as in smooth-walled waveguide). Numerical estimates are given for TE_{01} in 2-inch I.D. guides at 55 kmc.

For both cases the power in the reconversion echo varies directly as the system length, and the shape of the echo is independent of length. For the case of distributed undesired mode loss the echo to impulse-excitation has an exponential shape in relative delay τ, varying as $e^{-\tau/\tau_0}$, and for the case of partially absorbing mode filters the echo is a line-segment approximation to an exponential in τ (Fig. 4). The characteristic delay constant τ_0 is about 0.035 nanosecond for helix damped modes in an all-helix line, and is about 0.106 microsecond for TE_{02} in either helix or smooth-walled guide. For solid-walled guide with mode filters every 300 feet the characteristic delay constant (similar to τ_0) is about 2 nanoseconds.

Estimates are made for signal interference effects from such echoes, taking account of the fact that the most limiting requirements on echoes in some system arrangements occur at $\tau \gg \tau_0$, where the reconversion power is small. For PCM in smooth-walled copper waveguide with mode filters every 300 or 150 feet, it is concluded that pulse rates of 200 or 400 megabits might be used, with up to 20 or 40 miles respectively between regenerators;

beat wavelength straightness variation mode conversion is controlling. For PCM in an all-helix waveguide, it is concluded that a pulse rate up to 5000 megabits and up to 746 miles between regenerators is permitted by mode conversion effects; diameter variations (TE_{02} conversion) are controlling.

For transmission of frequency division multiplex via a frequency-modulated carrier (FDM-FM), estimates based on the discrete-echo theory of Bennett, Curtis and Rice suggest that 4000-mile transmission of 2000 channel groups is possible in all-helix waveguide; diameter variations are controlling. An rms frequency deviation of about 15 mc (total band about 150 mc) would provide 40 db interchannel interference ratio at 900 miles, and larger deviations (σ) increase the allowed system length (z) according to $z \approx (\sigma)^{2.8}$. Even in solid-walled waveguide there is a good prospect for 4000-mile FDM-FM using guide tolerances already achievable.

Separate consideration is being given to delay distortion due to waveguide cutoff dispersion, which will be appreciable in some configurations described and will require equalization.

I. INTRODUCTION

This paper gives some quantitative estimates of the delay distortion due to mode conversion to be expected in multimode transmission lines such as the millimeter-wave circular electric waveguides. In principle, the results apply to any multimode system, including optical guided-wave systems; in the latter case, however, it is likely that the magnitude of delay distortion will be too small to be a limitation in practice. In millimeter waveguide systems, delay distortion effects can be important.

A very simple analytical approach is used, based on modes coupled through power-transfer coefficients. This permits a direct translation of the spatial distribution of mode conversion into the effect on the output waveform without recourse to the frequency domain as an intermediate step. The case treated herein is mode conversion that is an independent random function of distance, but the physical reasoning employed can be used to indicate the changes which alterations in the conversion distribution would cause. A perturbation method is employed, and limits are found for the distance for which this perturbation calculation is valid.

In a companion paper Dale T. Young[1] has shown how the power-transfer coefficients of the coupled-mode equations used herein are related to the more familiar coefficients between the amplitudes of coupled modes; this makes more quantitative the relation between the present work and the work of H. E. Rowe and W. D. Warters.[2] In another companion paper, L. H. Enloe[3] presents a precise technique for

analyzing the delay distortion effects due to mode conversion, starting with the work of Rowe and Warters[2] and D. T. Young;[4] additionally, Enloe has extended the work of Bennett, Curtis, and Rice[5] to facilitate the calculation of interchannel interference in FM systems in the presence of the exponential echo which is characteristic of mode conversion and reconversion effects.

Two different systems situations will be included in the numerical evaluations: (i) analog systems such as FM, wherein the important transmission line length is large, perhaps 4,000 miles, and (ii) regenerative systems, PCM, wherein the pertinent line length is the distance between regenerators, perhaps 15–50 miles.

Previous consideration has been given to FM on waveguide by Kazuhiro Miyauchi[6] and R. Hamer.[7] The present work seeks to add to our understanding by using impulse excitation and a simple physical model to clarify the nature of the distortion produced by mode conversion-reconversion effects and to make evident the controlling parameters. By using experimental results in combination with analysis, it turns out to be possible to define in a broad way the system capabilities of existing waveguides without specifying many of the exact waveguide tolerances. Also presented is a first treatment of the delay distortion effects in smooth-walled waveguides with partially absorbing mode filters inserted periodically.

II. PRELIMINARY FORMULATION OF THE PROBLEM

To evaluate delay distortion we assume the transmission of an impulse in the desired signal mode. Power will be converted to an undesired mode and reconverted back to the signal mode after some relative delay or advancement. Use will be made of the three-mode system previously employed[8] and sketched in Fig. 1. We designate the signal power as
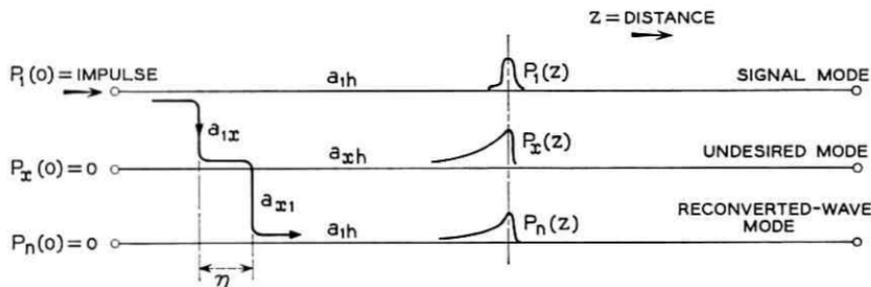
Fig. 1 — Diagram representing mode conversion and reconversion in multimode transmission systems.

$P_1(z)$, which decays exponentially as $e^{-(a_{1h}+a_{1x})z}$ due to heat loss a_{1h} and conversion loss a_{1x} coefficients. The undesired mode power $P_x(z)$ has a heat loss coefficient a_{xh} and a conversion loss coefficient a_{x1}. The third mode $P_n(z)$ contains the reconverted power, and has a heat loss coefficient a_{1h} and conversion loss coefficient a_{1x}. In Ref. 1 free conversion and reconversion is permitted between mode x and mode n, while conversion out of P_1 is permitted but no reconversion back from mode x to mode 1 takes place. Thus the signal mode total power is $(P_1 + P_n)$, but it is convenient to separate P_1 and P_n to identify the power which has undergone conversion at least once (P_n).

In the present pulse analysis we will employ only one mode conversion-reconversion sequence and identify the limits of applicability of this approximation.

We use power flow directly, as in the appendix of Ref. 8, and assume continuous random coupling between the modes. From this we get the expected value for the reconverted power under the conditions assumed; this is a useful first step, although more detailed knowledge would be desirable.

The key to this analysis is the fact that the converted pulse in the x mode suffers simultaneous delay and attenuation relative to the signal mode. No matter where in the line a component is converted and reconverted, a relative delay τ is always associated with a relative attenuation $e^{-(a_{xh}-a_{1h})\eta}$, where η is the distance the power travels in the x mode. Also,

$$\tau = \frac{\eta}{v_x} - \frac{\eta}{v_1} \tag{1}$$

where v_x and v_1 are the group velocities in mode x and mode 1 respectively.

The magnitude of the reconverted power density P_nD relative to the signal pulse P_1 is (for $z >$ a minimum distance to be specified)

$$P_nD(\eta) = K \exp\left[-(a_{xh} - a_{1h})\eta\right] \tag{2}$$

where K is to be determined. By using (1) we can replace η in (2) by τ, giving the reconverted power density versus τ (Fig. 2). We see immediately that the echo power due to conversion and reconversion varies as $e^{-\tau/\tau_0}$, where τ_0 depends only on the group velocities and the heat loss coefficients. Equation (2) can be integrated over all values of η, the distance travelled in the x mode, and equated to the average reconverted power P_n to evaluate K.

$$\int_0^\infty P_nD(\eta)\, d\eta = \text{average power in mode } n. \tag{3}$$

The next section establishes the right-hand side of (3).
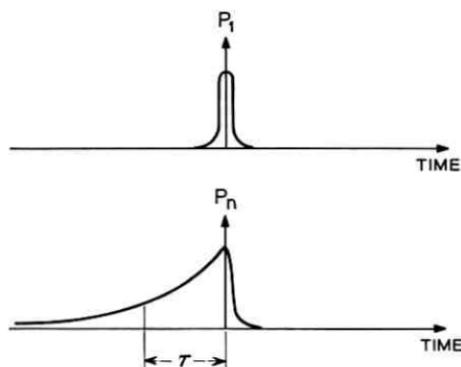
FIG. 2 — Signal impulse (P_1) and reconversion echo wave (P_n) versus time after transmission through a multimode line with mode conversion.

III. DETERMINATION OF THE AVERAGE RECONVERTED POWER

We use here the results of an earlier analysis.[8] We assume continuous excitation of the signal mode at the sending end, and write the following equations for power flow between modes

$$\frac{dP_1}{dz} = -(a_{1h} + a_{1x})P_1$$

$$\frac{dP_x}{dz} = -(a_{xh} + a_{x1})P_x + a_{1x}P_1 + a_{1x}P_n$$

$$\frac{dP_n}{dz} = -(a_{1h} + a_{1x})P_n + a_{x1}P_x.$$

In Ref. 8 these equations are solved for $P_1(z = 0) = 1$ and $P_x(z = 0) = P_n(z = 0) = 0$; dividing that work's equation (21) by equation (19) we obtain

$$\frac{P_n}{P_1} = -1 + \frac{1}{\sqrt{}}\left\{\frac{1}{2}[-(\alpha - \beta) + \sqrt{}]\exp\left[\frac{1}{2}[-(\beta - \alpha) + \sqrt{}]z\right]\right.$$
$$\left. - \frac{1}{2}[-(\alpha - \beta) - \sqrt{}]\exp\left[\frac{1}{2}[-(\beta - \alpha) - \sqrt{}]z\right]\right\} \qquad (4)$$

where

$$\alpha = a_{1h} + a_{1x}$$

$$\beta = a_{xh} + a_{x1}$$

$$\sqrt{} = \sqrt{(\alpha - \beta)^2 + 4a_{1x}a_{x1}}.$$

It can be verified that P_n/P_1 grows without limit for $z \to \infty$ because the exponent $[-(\beta - \alpha) + \sqrt{\,}]$ is positive. This is an expected result due to the fact that conversion and reconversion take place between mode x and mode n, whereas power converts one way from mode 1 to mode x.

We now seek to limit z in such a way that (4) is small compared to unity. We assume

$$a_{1x} = a_{x1} \tag{5}$$

$$a_{1x}^2 \ll (a_{xh} - a_{1h})^2. \tag{6}$$

Then (4) reduces to

$$\frac{P_n}{P_1} = -1 + \exp\left[a_{1x}^2 z/(a_{xh} - a_{1h})\right] + \frac{a_{1x}^2}{(a_{xh} - a_{1h})^2}$$

$$\exp\left[-(a_{xh} - a_{1h})z\right] - \frac{a_{1x}^2}{(a_{xh} - a_{1h})^2} \exp\left[a_{1x}^2 z/(a_{xh} - a_{1h})\right] \tag{7}$$

Now, by requiring

$$(a_{xh} - a_{1h})z \gg 1 \tag{8}$$

we drop the third term of (7), and by letting

$$\frac{a_{1x}^2 z}{(a_{xh} - a_{1h})} \ll 1 \tag{9}$$

we can expand the second term of (7) (noting the fourth term is negligible compared to the second), yielding

$$\frac{P_n}{P_1} = \frac{a_{1x}^2 z}{(a_{xh} - a_{1h})} . \tag{10}$$

Conditions (6) and (8) are not restrictive on the maximum length z of the system and are very typical. Condition (9) is restrictive on z, and by requiring it to be true, we discover P_n/P_1 has the same form. Hence, so long as P_n/P_1 calculated from (10) is small compared to unity, it is a valid calculation. Numerical values will be inserted at a later point.

IV. JUSTIFICATION OF FIRST-ORDER PERTURBATION SOLUTION

We now consider the validity of using only the first conversion and reconversion term. Using equations (19), (20) and (21) of the appendix to Ref. 8, it can be shown that when restriction (8) of this paper is met

it is also true that

$$\frac{P_1 + P_n}{P_x} = \frac{(a_{xh} - a_{1h})}{a_{1x}} \tag{11}$$

which is independent of z. We will see that this is typically a number of the order of 100 to 1000 or more. As stated above, when (9) is also true, $P_n \ll P_1$, and hence

$$\frac{P_1}{P_x} = \frac{a_{xh} - a_{1h}}{a_{1x}}. \tag{12}$$

This is believed to be important. P_1 decays simply as $\exp\left[-(a_{1h} + a_{1x})z\right]$ independently of all reconversion effects; P_1 therefore serves as a convenient reference, and the other powers P_x and P_n are normalized with respect to it. For any z [subject to (8) and (9)] P_1/P_x is a constant. With increasing z, power is being converted continuously from P_1 to P_x and from P_n to P_x, but P_1/P_x is constant despite the fact that P_n/P_1 grows through orders of magnitude. It must follow that negligible power in mode x comes from mode n in the range for which (9) is valid. Since negligible power in mode x was ever in mode n, the second- and higher-order processes of conversion and reconversion can be ignored.

V. RECONVERSION POWER DENSITY VERSUS RELATIVE DELAY

Equation (3) can now be written [using (10)]

$$\int_0^\infty K \exp\left[-(a_{xh} - a_{1h})\eta\right] d\eta = \frac{a_{1x}^2 z}{(a_{xh} - a_{1h})} P_1. \tag{13}$$

Performing the integration to evaluate K and substituting back into (2) gives

$$P_n D(\eta) = P_1 a_{1x}^2 z \exp\left[-(a_{xh} - a_{1h})\eta\right] \tag{14}$$

or, using (1) to eliminate η in (13)

$$P_n D(\tau) = \frac{P_1 a_{1x}^2 z}{\left(\frac{1}{v_x} - \frac{1}{v_1}\right)} \exp\left[-(a_{xh} - a_{1h})\tau / \left(\frac{1}{v_x} - \frac{1}{v_1}\right)\right]. \tag{15}$$

In either case we have assured that

$$\int_0^\infty P_n D(\tau) \, d\tau = \int_0^\infty P_n D(\eta) d\eta = \frac{P_1 a_{1x}^2 z}{(a_{xh} - a_{1h})}. \tag{16}$$

It is convenient to define a time constant τ_0 at which the reconversion

power density is down to $1/e$ times the $\tau = 0$ value

$$\tau_0 = \frac{\left(\dfrac{1}{v_x} - \dfrac{1}{v_i}\right)}{(a_{xh} - a_{1h})} \approx \frac{1}{2c}\frac{(\nu_x{}^2 - \nu_1{}^2)}{(a_{xh} - a_{1h})} \tag{17}$$

where

c = velocity of light in the dielectric in the waveguide

ν_y = $\lambda k_y / \pi d$

d = waveguide diameter

k_y = Bessel root appropriate to the cutoff of mode y, and

λ = intrinsic wavelength characteristic of the dielectric in the waveguide.

The part of (17) involving ν's instead of group velocities v assumes the waveguides are far enough from cutoff that

$$\sqrt{1 - \nu^2} \approx 1 - \frac{1}{2}\nu^2.$$

In some system applications it may be acceptable to ignore reconversion tails closer to the signal pulse than some time interval τ_r. It is of interest then to sum the reconverted powers for τ from τ_r to ∞, which is the shaded area in Fig. 3. From (15)

$$P_{nr} = \int_{\tau_r}^{\infty} P_n D(\tau)\, d\tau = \frac{P_1 a_{1x}{}^2 z}{(a_{xh} - a_{1h})} \exp\left[-\tau_r / \tau_0\right]. \tag{18}$$
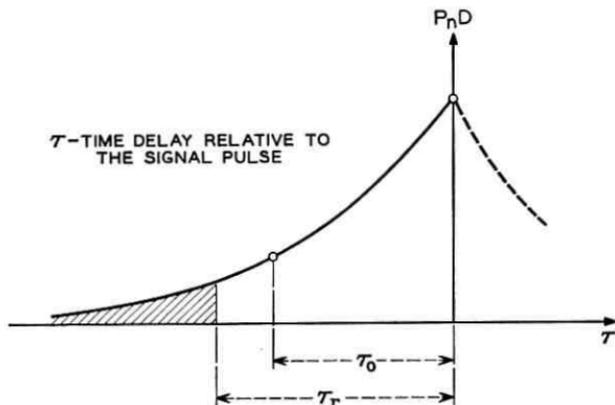
Fig. 3 — Density of reconverted wave power $(P_n D)$ versus time when there is coupling between modes with distributed differential loss.

VI. EFFECT OF MODE FILTERS ON DELAY DISTORTION

The preceding paragraphs cover the case of distributed differential loss. When the undesired mode loss is lumped, as is the case when mode filters are used, (1) still holds, but we need a new expression to replace (2) to represent the density of reconverted power as a function of the distance η that the power traveled in the undesired mode. The appropriate relation is

$$
\begin{aligned}
P_n D(\eta) &= A \left[1 - \frac{\eta}{s} + \frac{C_1 \eta}{s} \right] \text{ for } 0 < \eta < s \\
&= A \left[C_1 \left(2 - \frac{\eta}{s} \right) + C_1^2 \left(\frac{\eta}{s} - 1 \right) \right] s < \eta < 2s \quad (19) \\
&= A \left[C_1^2 \left(3 - \frac{\eta}{s} \right) + C_1^3 \left(\frac{\eta}{s} - 2 \right) \right] 2s < \eta < 3s
\end{aligned}
$$

in which s is the mode filter spacing, C_1 is the power transmission coefficient for the mode filters, and the number of mode filter sections n must be large compared to $(p + 1)$, where $ps < \eta < (p + 1)s$. The proportionality constant A is analogous to K of (2), and is to be evaluated by equating the integral of $P_n D(\eta)$ over all η to the average reconverted power P_n. To do so we evaluate P_n (assuming $a_{1x} = a_{x1}$)

$$
P_n = \int_0^{ns} a_{1x} P_x(y) \, dy = n \int_0^s a_{1x} P_x(y) \, dy + (n - 1)
$$

$$
\int_s^{2s} a_{1x} C_1 P_x(s) \, dy + (n - 2) \int_{2s}^{3s} a_{1x} C_1^2 P_x(s) \, dy + \cdots \quad (20)
$$

in which y is the distance in the direction of propagation, analogous to z of Fig. 1. We use a loose coupling approximation and assume random coupling effects as above. We can assume the loss in the driven mode (no. 1) is small in the length s between mode filters, so that we can write $P_x(y)$ in the interval $ps < y < (p + 1)s$, p any integer,

$$
P_x(y) = P_1(ps) \int_0^{(y - ps)} a_{1x} \, dy + C_1 P_x(ps). \quad (21)
$$

We evaluate (20) using the approximation that there is little difference in attenuation (other than the mode filter loss) between modes 1 and x in the length η which is ultimately important

$$
P_n = P_1 \frac{n a_{1x}^2 s^2}{2} \left[1 + \frac{2(n - 1)C_1}{n} + \frac{2(n - 2)C_1^2}{n} + \cdots \right]. \quad (22)
$$

For $n \gg k$, where C_1^k is the highest-order term of importance and $z = ns =$ total length of line

$$P_n \approx \frac{a_{1x}^2 sz}{2} \frac{(1 + C_1)}{(1 - C_1)} P_1. \tag{23}$$

This expression is analogous to (10) and P_1 is again the power at z in the driven mode. We now evaluate A of (19) by equating the two expressions for P_n

$$P_n = P_1 \frac{a_{1x}^2 sz}{2} \frac{(1 + C_1)}{(1 - C_1)} = \int_0^\infty P_n D(\eta) \, d\eta = \frac{s}{\tau_s} \int_0^\infty P_n D(\tau) \, d\tau \tag{24}$$

where

$$\frac{\eta}{s} = \frac{\tau}{\tau_s} \tag{25}$$

$$\tau_s = s \left(\frac{1}{v_x} - \frac{1}{v_1} \right) \tag{26}$$

$$\approx \frac{s}{c} \left(\frac{1}{2} v_x^2 - \frac{1}{2} v_1^2 \right).$$

Using (19) for $P_n D(\eta)$ we perform the integration of (24) and find, approximately

$$A = a_{1x}^2 z P_1. \tag{27}$$

We can now put (19) in final form

$$P_n D(\tau) = P_1 a_{1x}^2 z \left[1 - \frac{\tau}{\tau_s} + \frac{C_1 \tau}{\tau_s} \right] 0 \leq \tau \leq \tau_s$$

$$= P_1 a_{1x}^2 z \left[C_1 \left(2 - \frac{\tau}{\tau_s} \right) + C_1^2 \left(\frac{\tau}{\tau_s} - 1 \right) \right]$$

$$\tau_s \leq \tau \leq 2\tau_s \tag{28}$$

$$= P_1 a_{1x}^2 z \left[C_1^2 \left(3 - \frac{\tau}{\tau_s} \right) + C_1^3 \left(\frac{\tau}{\tau_s} - 2 \right) \right]$$

$$2\tau_s \leq \tau \leq 3\tau_s$$

or more generally

$$P_n D(\tau) = P_1 a_{1x}^2 z \left[C_1^k \left(k + 1 - \frac{\tau}{\tau_s} \right) + C_1^{k+1} \left(\frac{\tau}{\tau_s} - k \right) \right] \tag{28a}$$

$$k\tau_s \leq \tau \leq (k + 1) \tau_s$$

and $k = 0, 1, 2, 3, 4$, etc. This function is plotted in Fig. 4. The reconversion power density declines linearly from its $\tau = 0$ value to C_1 times its $\tau = 0$ value at $\tau = \tau_s$. For $\tau_s < \tau < 2\tau_s$ the reconversion power density again declines linearly from $C_1 a_{1x}^2 z P_1$ to $C_1^2 a_{1x}^2 z P_1$, and similarly for larger τ.

It is again of interest to evaluate the total reconversion power from $\tau = \tau_r$ to $\tau = \infty$ because echoes at very short time delays are not damaging. The ratio of reconversion power for $\tau > \tau_s$ to the total reconversion power may be shown to be

$$\int_s^\infty P_n D(\eta)\, d\eta = P_n - \int_0^s P_n D(\eta)\, d\eta = C_1 P_n \qquad (29)$$

and similarly for reconverted power at $\tau > 2\tau_s$

$$\int_{2s}^\infty P_n D(\eta)\, d\eta = \int_s^\infty P_n D(\eta)\, d\eta - \int_s^{2s} P_n D(\eta)\, d\eta = C_1^2 P_n. \qquad (30)$$

Thus for $\tau_r = m\tau_s$ the total reconverted power from $\tau = \tau_r$ to $\tau = \infty$ declines as C_1^m.

VII. DISTANCES FOR WHICH THE APPROXIMATIONS HOLD

Two approximations concerning z have been made, relations (8) and (9). The minimum z comes from (8). We calculate for three waveguide cases of interest
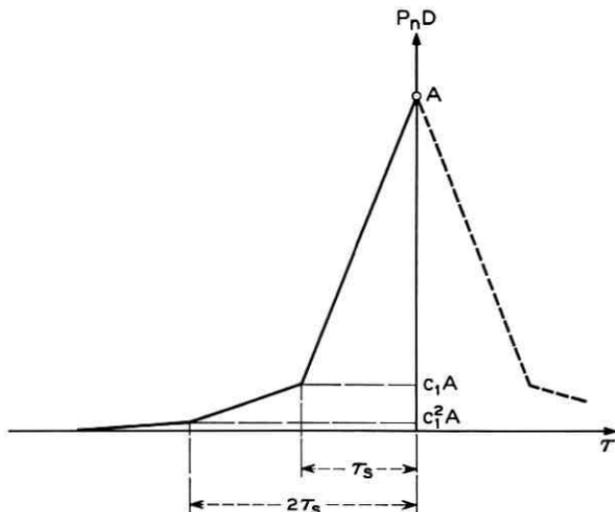

Fig. 4 — Density of reconverted wave power $(P_n D)$ versus time when there is coupling between modes with lumped differential loss.

(i) helix damped modes in a 2-inch I.D. all-helix waveguide,

(ii) modes (such as TE_{02}) which are attenuated only by the normal copper losses in the walls; this situation is the same in helix waveguide or in solid-walled copper waveguide, where again we use the 2-inch I.D. case, and

(iii) 2-inch I.D. solid-walled copper guide with a 15-foot helix mode filter every 300 feet.

Condition (8) appears because we want to make exp $[-(a_{xh} - a_{1h})z]$ negligible compared to unity; we will use as a criterion for minimum z (z_{min}) that $(a_{xh} - a_{1h}) z_{min} = 3$. The table below gives the typical values for $(a_{xh} - a_{1h})$ and the resulting z_{min}. For case (iii), $z_{min} \gg ms$, where m is chosen so that $C_1{}^m \ll 1$; this assures that the line-segment exponential reconversion trail (Fig. 4) is fully developed. Except possibly for case (ii) these restrictions are not very limiting.

The maximum $z(z_{max})$ for which these estimates are valid comes from (9) and (10) or (23); we take $P_n/P_1 = 0.1$ as the definition of z_{max}, although somewhat larger z's might give meaningful results. The values of $(a_{xh} - a_{1h})$ from Table I (known from measurements on physical waveguides) are again used. The values of conversion coefficient a_{1x} are known from measurements on experimental waveguides at the Holmdel, N. J., Bell Laboratories for cases (i) and (iii) but are not known from measurement for case (ii). For the latter the theory of H. E. Rowe and W. D. Warters[2] was used, in combination with known data on the diameter variations of our good waveguides; a 1-mil rms diameter variation was taken, and the Warters-Rowe calculation for the average loss for random discrete imperfections on 10-foot joint spacing used to get the case (ii) value of a_{1x} in Table II. These are conservative values when applied to the 55-kmc region; Holmdel waveguides have shown comparable conversion coefficients at higher frequencies.

For all-helix waveguide, it appears that these estimates should be good for a cross-country analog system; for copper tubing with mode filters the estimates are good for 100 miles or thereabouts.

TABLE I

Waveguide Case	Amplitude Decay Coefficient $(a_{xh} - a_{1h})/2$*	$(a_{xh} - a_{1h})$	z_{min}
(i) (ii) (iii)	1 db/foot 4.3 db/mile	0.23/ft. ≈1/mile	13 feet 3 miles 3000 feet

* With the necessary conversion factors to give the units shown.

TABLE II

Waveguide Case	Average Amplitude Conversion Loss $a_{1x}/2$*	$a_{1x}/(a_{xh} - a_{1h})$	a_{1x}	z_{max} miles
(i)	1 db/mile	1/5280	0.23/mile	2290
(ii)	0.005 db/mile	1/860	0.00115 per mile	74,600
(iii)	1 db/mile		0.23/mile	62

* With the necessary conversion factors to give the units shown.

VIII. NUMERICAL VALUES FOR RECONVERSION TIME CONSTANTS

In this section we will evaluate the relations previously derived for the time constants of the reconverted powers, using parameters typical of millimeter-wave circular electric waveguides.

It should be noted that the reconverted power will precede as well as lag the signal impulse. Significant mode conversion takes place between TE_{01} and TE_{11} due to straightness variations, and the group velocity for TE_{11} is greater than that for TE_{01}. This will not be labored at greater length here; a *typical* group velocity difference will be calculated, and part of the reconverted power for straightness variations will precede the signal and part will lag the signal.

It should also be noted that the time variation of the reconversion power as given by (15) and (28) is independent of distance z within the limits $z \leqq z_{max}$ calculated above. The magnitude of reconversion is of course dependent on z.

Equations (17) and (26) show that less delay occurs when the velocity difference is smaller, and this is in turn obtained when ν is smaller, i.e., the guide is farther from cutoff. From (17) or (26) the reconversion duration varies as $1/f^2 d^2$ where d = waveguide diameter and f = frequency. The estimates given here are for 2-inch I.D. guide at 55 kmc.

Consider the delay distribution of mode conversion due to random unintentional straightness variations, which causes coupling to the TM_{11}, TE_{11}, and TE_{12} modes. Because TE_{12} gives the longest echo trail, we use it as the mode for the computation of $(\nu_x^2 - \nu_1^2)/2$ and evaluate (17) and (26) as shown in Table III.

TABLE III

Waveguide Case	$(\nu_x^2 - \nu_1^2)/2$	Delay Evaluation
(i) (all-helix)	0.0079	$\tau_0 = 0.0348 \times 10^{-9}$ sec
(iii) (solid copper with mode filters)	0.0079	$\tau_s = 2.41 \times 10^{-9}$ sec

The other waveguide condition of interest, case (ii), either helix or solid-walled copper with circular electric mode conversion, gives an exponential reconversion shape; $(\nu_x^2 - \nu_1^2)/2 = 0.0197$ (for TE$_{01}$ — TE$_{02}$), $(a_{xh} - a_{1h})$ corresponds to an attenuation difference of 4.3 db/mile, and $\tau_0 = 105.8 \times 10^{-9}$ second from (17).

IX. SYSTEM EVALUATION OF MODE CONVERSION DELAY DISTORTION

Some comments can be made on the relation between expected system requirements and reconversion power magnitude and delay distribution, even though further analysis in the case of analog systems will be necessary to get precise numbers.

9.1 PCM on an All-Helix Waveguide Line

Consider first an all-helix waveguide line. Typical values of conversion coefficient a_{1x} and heat-loss difference $(a_{xh} - a_{1h})$ have already been given in connection with Tables I and II. The first row of Table I applies for random straightness deviations, and the second row applies for diameter variations. If we require a total P_n/P_1 ratio (summed over all delay times) of -30 db (as might be desirable in a PCM system of infinite pulse rate) we find from (18) with $\tau_r = 0$ that a distance z of 22.9 miles is permissible as far as straightness variations are concerned, and $z = 746$ miles is permissible from the point of view of diameter variations. The shorter distance controls, and we can conclude for helix waveguide there is no limit on the pulse width permitted by reconversion delay distortion in a PCM system with regenerators every 20 miles. We can go further, however. For reconversion delays very short compared to the pulse rate there will be no interference between time slots; the reconverted power shows as a variation in apparent loss in the line and provided that this loss shift is stable, the reconverted power can be relatively large. Suppose we let $z = 4000$ miles; then the total $P_n/P_1 = -7.6$ db with respect to straightness variations, and the total $P_n/P_1 = -22.6$ db with respect to diameter variations. Using (18), we can compute the τ_r for the total P_{nr}/P_1 in the range $\tau_r < \tau < \infty$ to be -30 db; we find $\tau_r/\tau_0 = 5.16$ or $\tau_r = 0.18 \times 10^{-9}$ second with respect to straightness variations, and $\tau_r/\tau_0 = 1.7$ or $\tau_r = 180 \times 10^{-9}$ second with respect to diameter variations. This implies that straightness variations (or other effect causing conversion to modes damped by the helix) do not limit and diameter variations do limit the length between regenerators in PCM systems with all-helix waveguide. By backing off from $z = 4000$ to $z = 746$ miles, the desired total $P_n/P_1 = -30$ db

including reconverted power at all time delays; thus satisfactory PCM operation with regeneration every 746 miles is permitted by reconversion delay effects in an all-helix waveguide line for pulse rates up to about $10^9/0.18$ or 5000 megabits/second.

9.2 FDM-FM on an All-Helix Waveguide

The work of Bennett, Curtis, and Rice[5] allows us to estimate the limits of transmitting a single-sideband group of telephone channels by FM (FDM-FM) as is done on the TD-2 and TH radio relay systems. Their work shows that the RF signal-to-echo ratio need not be as large as the interchannel interference requirement, and since the length of the permitted system is inversely proportional to the required RF signal-to-echo power ratio (Ref. 5, equation 18) this can greatly change the permissible system length. The work of Bennett et al.[5] (their Fig. 5.7) shows that when the ratio σ/f_b, (rms frequency deviation)/(highest baseband frequency), is unity or more the requirements on reconverted power are essentially independent of delay for $\tau > 0.2/f_b$ (f_b = highest baseband frequency) and becomes very rapidly more tolerant of reconverted power for smaller delays. At $\sigma/f_b = 1$ an FM advantage of about 7 db for large time delays is given by Bennett et al., and for larger σ/f_b the FM advantage increases at the rate of 8.4 db/octave.* It seems reasonable to assume an FM advantage of at least 10 db, and more probably could be realized. This means that for interchannel interference of -40 db, $P_{nr}/P_1 = -30$ db can be used, and since the requirement falls off rapidly below $\tau f_b = 0.2$, an assumed 10-mc SSB baseband signal (2000 channels) allows us to use $\tau_r = 20 \times 10^{-9}$ seconds in computing the permissible system length from (18). The resulting estimate is 900 miles, maximum system length for a 10-mc group of SSB-AM channels to have -40 db interchannel interference ratio when sent with an FM index σ/f_b of about $\sqrt{2}$ on an all-helix waveguide described by cases (i) and (ii) of Tables I and II. (The limitation is TE$_{02}$.) Mr. L. H. Enloe has made a much more rigorous analysis and found a similar result.[3]

Since the FDM-FM application is limited only by conversion to higher circular electric modes, there would be considerable advantage in adding mode filters for such modes.[6] This is certainly possible in principle but has not been reduced to practice. Estimates of requirements for circular electric mode filters can be deduced from (23), (26) and (28).

---

* This and the succeeding discussion mean that the FM advantage and the allowed system length vary as the 2.8 power of the frequency deviation when long-delay echoes are controlling.

9.3 *PCM on Solid-Walled Copper Waveguide*

Turning now to PCM transmission in solid-walled copper waveguide with (noncircular electric) mode filters, the circular electric mode conversion limitations are the same as in an all-helix waveguide line. Typical conversion loss and heat loss numbers are given in case (*iii*) of Tables I and II. For straightness variations (modes damped by the mode filters) and for $P_n/P_1 = -30$ db including all time delays, the permissible distance between regenerators is 0.62 mile for 300-foot mode filter spacing; P_1/P_n, and hence permissible system length, varies inversely as the mode filter spacings. This form of line will *not* handle indefinitely large PCM pulse rates for 15–20-mile repeater spacings. For 300-foot mode filter spacing the length of reconversion pulse (τ_s in Fig. 4) is $\tau_s = 2.5 \times 10^{-9}$ second (per Table III); a PCM system using a pulse rate of 200 megabits would probably be acceptable up to distances at which the "second" ($\tau_s < \tau < 2\tau_s$) reconversion pulse becomes limiting. For 15-db mode filters,* this would be approximately a distance z such that

$$10 \log_{10} (z/0.62) = 15 \text{ db}$$

or $z \approx 20$ miles between regenerators. For operation where the third ($2\tau_s < \tau < 3\tau_s$) reconversion interval is controlling — i.e., at PCM rates of about 100 megabits — the length z can be

$$10 \log (z/0.62) = 30 \text{ db}$$

or $z = 620$ miles† between regenerators. Alternatively, by using mode filters spaced 150 feet instead of 300 feet, the permissible pulse rate could be 400 megabits at 40 miles. We found above that diameter variations permit a length between regenerators ($P_1/P_n = -30$ db) of 746 miles, and hence this imperfection is not limiting for high-speed PCM in solid-walled guide with mode filters.

9.4 *FDM-FM on Smooth-Walled Copper Waveguide*

For the same guide with FDM-FM transmission where we may require [in (23)] $P_n/P_1 = -30$ db as above, the length z permitted by straightness variations is 0.62 mile (for all delay times considered interfering). If we ask what τ_r would make the permitted system length controlled by straightness variations equal to 900 miles, the figure found

---

* A 15-foot section of 2-inch I.D. Holmdel helix would have 15 db loss to TE_{12} and TE_{11} at 55 kmc and even more loss for TM_{11}, the other significant mode.

† Here we have run beyond the 100-mile limit of applicability of the perturbation theory and the result needs further justification.

permissible due to diameter variations, we find [as indicated in the discussion associated with (29) and (30)]

$$C_1^{\tau_r/\tau_s} = \frac{0.62}{900} .$$

Letting $C_1 = 0.0316$ (corresponding to a 15-db mode filter), we find $\tau_r/\tau_s \approx 2$ and $\tau_r \approx 5 \times 10^{-9}$ sec. Neglecting echoes with time delays less than 5×10^{-9} seconds is compatible with the assumptions made above for the FDM-FM transmission and appears to be a real possibility. However, to go to 900 miles system length is in violation of the $z_{max} = 62$ miles computed earlier, and indeed it is clear that the perturbation calculation is no longer valid. We might inquire what would be the practical effect of longer systems than that for which the perturbation calculation holds.* When all of the power in the driven mode has been reconverted once, the driving signal would no longer be an impulse but might roughly be a waveform of the character sketched in Fig. 4. Then further conversion and reconversion would broaden this waveform again. As a guess we might expect one perturbation-theory time-constant of delay distortion for every distance such that $P_n/P_1 \approx 1$. For the case being considered, $P_n/P_1 \approx 1$ at $z = 620$ miles, and in this vicinity the above calculation showed a τ_r of 5×10^{-9} seconds would be adequate to meet requirements; if we now allow $\tau_r = 20 \times 10^{-9}$ seconds as done above ($\tau f_b \approx 0.2$) we might expect to tolerate $(20/5) \approx 4$ times 600 miles system length. Thus, even in solid-walled guide with mode filters the reconversion power may be concentrated at such small delays as to permit FDM-FM for 900 miles and possible greater distances. This must be regarded as a tentative estimate, a major uncertainty being the nature of the delay distortion in a multimode system in the region where several successive conversions and reconversions must be accounted for (i.e., perturbation theory not valid).

---

* Another interesting question concerns the way echo power will be related to a possible "breaking" phenomenon at the FM demodulator. This is quite different from the breaking which occurs due to thermal noise. For echo delays which are short compared to the reciprocal of the modulating baseband frequency, the writer has concluded that no appreciable fluctuation in the envelope results, and the echo power has the principal effect of causing a static shift in carrier magnitude which would differ among members of an ensemble of waveguide lines but which would be approximately fixed (varying perhaps slowly due to temperature effects, drifts in carrier frequency, etc.) in any one line. For echo delays which are comparable to the reciprocal of the modulating baseband frequency, envelope fluctuations do result and limiters can be used to advantage. Since some waveguide modes produce reconversion echoes which occur predominantly at delays short compared to the modulating periods, the total tolerable reconversion power may be significantly greater than the tolerable thermal noise at the FM demodulator. Further study of specific situations is needed to answer the question.

In practice there is another factor which will make solid-walled waveguide less speculative for FDM-FM. As L. C. Tillotson has pointed out, one can demodulate the signal often enough so that the interference calculated from perturbation theory is valid. This is more than a ruse. By demodulating and filtering, the interchannel interference terms resulting from very short delay echoes are discarded. After remodulating, the short delay echoes are missing and hence they cannot go through successive steps of delay distortion to become longer delay echoes in the manner discussed above. Thus by demodulating at approximately the intervals at which the perturbation theory breaks down (100 miles approximately) one would filter out the short delay echoes and prevent their ever causing crosstalk. Since one usually drops and adds channels at intervals of 100–200 miles, it is possible that the demodulation procedure would not be a serious limitation in many cases.

X. CONCLUSIONS AND DISCUSSION

Using a simple power-flow analysis, expressions have been found for the magnitude and delay distribution for the expected value of the reconverted power in a multimode transmission system with mode conversion. Three cases are examined in detail:

(i) the helix-damped modes in an all-helix line,

(ii) circular electric mode conversions in either helix or smooth-walled guide, and

(iii) the mode filter damped modes in a line made up of smooth-walled copper plus mode filters.

For numerical examples, 2-inch I.D. guides at 55 kmc are assumed. The minimum and maximum lengths of line (between terminals in an analog system or between regenerators in a PCM system) for which the analysis is valid are given in Tables I and II respectively.

For impulse excitation, the reconversion echo as a function of delay τ relative to the signal pulse varies as $e^{-\tau/\tau_0}$ as given in Fig. 3 for cases (i) and (ii), and varies as a line segment approximation to an exponential illustrated in Fig. 4 for case (iii). For mode conversion due to straightness variations (TE_{12}), τ_0 in helix waveguide is about 0.035 nanosecond and τ_s (Fig. 4) in copper guide with mode filters spaced 300 ft. is about 2.4 nanoseconds. For diameter variations (TE_{02}) in either type of guide τ_0 is about 106 nanoseconds; this could be radically reduced by the addition of circular electric mode filters, but system estimates herein do not count on that, since such mode filters have not yet been developed.

Only the echo power at delays greater than some relative delay τ_r is subject to severe requirements in typical system layouts, and $\tau_r \gg \tau_0$ for certain imperfections in either helix waveguide or smooth-walled waveguide. The reconverted power P_n, relative to signal power P_1 at the same distance z, for $\tau_r < \tau < \infty$ is given in (18) for cases (i) and (ii)

$$P_{nr} = \int_{\tau_r}^{\infty} P_n D(\tau)\, d\tau = \frac{P_1 a_{1x}^2 z}{(a_{xh} - a_{1h})} \exp\left(-\tau_r/\tau_0\right) \qquad (18)$$

where a_{1x} is the average mode conversion coefficient and a_{xh} and a_{1h} are the heat loss coefficients in the undesired mode and signal modes respectively. Since reconverted power at delays less than some value τ_r will usually have negligible system effect, (18) evaluates the important reconversion power. For case (iii) similar results are given by (28), (29) and (30).

For the transmission of PCM on an all-helix line, the discussion of Section 9.1 indicates that diameter variations limit the regenerator spacing to about 746 miles and random straightness variations limit the pulse rate to 5000×10^6 per second. For analog transmission an FM advantage on the echo interference as computed for a single echo by Bennett, Curtis and Rice[5] is essential to get reasonable requirements on the RF signal-to-echo ratio. (See Section 9.2.) Using the theory of Bennett et al., a rough estimate indicates that 2000 channels could be sent a distance of 900 miles using an rms frequency deviation of $\sqrt{2}$ times the top baseband frequency, 10 mc; larger deviations would yield longer systems at a rate 6.9 times the system length for 2:1 increase in deviation. The length limit is again set by diameter variations. A more precise technique for computing the interchannel interference due to the exponential echo trails which are characteristic of multimode systems, Figs. 2 and 3, is given by L. H. Enloe in a companion paper.[3]

For the transmission of PCM on smooth-walled guide plus mode filters, the discussion of Section 9.3 indicates that pulse rates of 200 or 400 megabits with mode filters spaced 300 feet or 150 feet might be used with distances up to 20 or 40 miles respectively between regenerators. Longer distances between regenerators are possible if the bit rate or the mode filter spacing is reduced. In Section 9.4, analog transmission of the same FDM-FM signal noted above (2000 channels) may be possible in smooth-walled guide with mode filters for the same system lengths and deviation ratio noted above.

All of these estimates are very sensitive to the magnitude of the controlling imperfections in the medium; the permissible system length varies inversely as the fourth power of the geometric distortion magni-

tude. The echo duration estimates are also made for 2-inch guide at 55 kmc, and vary inversely as (frequency)^2 for fixed heat-loss coefficients. In practice, differential heat loss decreases with increasing frequency, tending to reduce the broad-range frequency variation in echo duration.

Separate consideration is being given to delay distortion due to waveguide cutoff dispersion, which will be appreciable in some system configurations described and will require equalization. Equalization of the delay distortion due to mode conversion is also being considered separately. It is interesting to note that each delay difference τ [of (28) or (15)] corresponds to a unique value of η, the distance the energy travelled in the undesired mode, and that the phase difference between all reconverted components and the unconverted (or "straight-through") signal component is also unique for a given τ. The random magnitudes (but not phase angles) of the conversion coefficients and the random location along the axis of propagation may make this surprising at first. There may be hope of equalizing out the expected value of the reconversion echo, but we must keep in mind that the variance of the echo magnitude will limit the degree to which this is possible. In this paper the system estimates are based on no equalization of the mode conversion effects.

REFERENCES

1. Young, D. T., Model for Relating Coupled Power Equations to Coupled Amplitude Equations, B.S.T.J., this issue, p. 2761.
2. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., **41,** May, 1962, pp. 1031–1170.
3. Enloe, L. H., Interchannel Interference in FM Systems Produced by Continuous Random Mode Conversion in Circular Electric Waveguide, B.S.T.J., this issue, p. 2765.
4. Young, D. T., Effect of Differential Loss on Approximate Solutions to the Coupled Line Equations, B.S.T.J., this issue, p. 2787.
5. Bennett, W. R., Curtis, H. E., and Rice, S. O., Interchannel Interference in FM and PM Systems under Noise Loading Conditions, B.S.T.J., **34,** May, 1955, p. 601.
6. Miyauchi, D., Interference Waves in Waveguide Transmission and Intermodulation Noise in FM Systems for Waveguide Communication, Review of the Electrical Communication Laboratory, **9,** May–June, 1961, pp. 361–393.
7. Hamer, R., Long Distance Waveguide Transmission, Electronic Engineering, **33,** 1961, pp. 218 and 279.
8. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33,** November, 1954, pp. 1209–1265; see especially the appendix.

Model for Relating Coupled Power Equations to Coupled Amplitude Equations

By D. T. YOUNG

Following a suggestion of H. E. Rowe, the first-order system of coupled differential equations with the independent variables representing amplitudes of coupled modes (the so-called coupled mode equations) has been rigorously transformed into similar equations with the independent variables representing the power in the coupled modes for the case where the amplitude coupling coefficient is a Gaussian random process with a white noise spectrum and zero mean value. Coupled mode equations in power have proven useful, and this work provides formal justification and identification of the variables and parameters.

The equations may allow one to approach coupled-mode problems directly for statistical results.

I. INTRODUCTION

Consider the coupled line equations:

$$I_0'(z) = -\Gamma_0 I_0(z) + jc(z)I_1(z) \tag{1}$$

$$I_1'(z) = jc(z)I_0(z) - \Gamma_1 I_1(z). \tag{2}$$

These equations are useful in describing effects of coupling between a signal mode, represented by a complex wave amplitude $I_0(z)$, and a single spurious mode, represented by $I_1(z)$, caused by geometric imperfections in a multimode transmission line. These equations may be derived in two ways from basic principles:[1,2,3,4] direct conversion of Maxwell's equations to generalized telegraphist's equations, or allowing discrete converters to become continuous. Exact solutions are known in only a few special cases, one of which is the case of constant $c(z)$. Making use of this solution and assuming $c(z)$ is a stationary Gaussian random function with a white noise spectrum S_0 and zero mean value, it is

2761

possible to derive coupled differential equations with the expected values of the power in $I_0(z)$ and $I_1(z)$ as independent parameters. That is

$$P_0'(z) = (-2\alpha_0 - S_0)P_0(z) + S_0 P_1(z) \tag{3}$$

$$P_1'(z) = S_0 P_0(z) + (-2\alpha_1 - S_0)P_1(z) \tag{4}$$

where

$$\underset{1}{P_0}(z) = \langle \underset{1}{I_0}(z) \underset{1}{I_0^*}(z) \rangle.$$

These equations are similar to equations used by S. E. Miller.[5]

II. DESCRIPTION OF MATHEMATICAL MODEL AND COMPUTATION

Consider (1) and (2) and assume the $\underset{1}{I_0}(z)$ are known and $c(z)$ has the constant value c in $(z, z + \Delta z)$; then $\underset{1}{I_0}(z + \Delta z)$ are given by[4]

$$\begin{bmatrix} I_0(z + \Delta z) \\ I_1(z + \Delta z) \end{bmatrix} = (e^{-\Gamma_0 \Delta z})\mathbf{G} \begin{bmatrix} I_0(z) \\ I_1(z) \end{bmatrix}$$

where $\mathbf{G}$ is a two-by-two matrix with elements

$$g_{11} = \left(\frac{a + j1}{2a}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 + ja)\Delta z\right]$$
$$+ \left(\frac{a - j1}{2a}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 - ja)\Delta z\right]$$

$$g_{22} = \left(\frac{a - j1}{2a}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 + ja)\Delta z\right]$$
$$+ \left(\frac{a + j1}{2a}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 - ja)\Delta z\right]$$

$$g_{12} = g_{21} = \left(\frac{c}{a\Delta\Gamma}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 + ja)\Delta z\right]$$
$$- \left(\frac{c}{a\Delta\Gamma}\right) \exp\left[\frac{\Delta\Gamma}{2}(1 - ja)\Delta z\right]$$

where

$$a = \sqrt{\left(\frac{2c}{\Delta\Gamma}\right)^2 - 1}$$

$$\Delta\Gamma = \Gamma_0 - \Gamma_1 = \Delta\alpha + j\Delta\beta.$$

Now compute $\langle I_0 I_0^* \rangle$ and $\langle I_1 I_1^* \rangle$ from

$$\begin{bmatrix} I_0 I_0^* \\ I_0 I_1^* \\ I_1 I_0^* \\ I_1 I_1^* \end{bmatrix}_{z+\Delta z} = (e^{-2\alpha_0 \Delta z})(\mathbf{G} \times \mathbf{G}^*) \begin{bmatrix} I_0 I_0^* \\ I_0 I_1^* \\ I_1 I_0^* \\ I_1 I_1^* \end{bmatrix}_z$$

where $\times$ denotes the Kronecker product (direct product) of the matrices $\mathbf{G}$ and $\mathbf{G}^*$.[6] If we compute ensemble averages, assuming the random function $c(z)$ is independent of previous values and $\langle c(z) \rangle = 0$, we have

$$P_0(z + \Delta z) = e^{-2\alpha_0 \Delta z}[AP_0(z) + BP_1(z)] \tag{5}$$

$$P_1(z + \Delta z) = e^{-2\alpha_0 \Delta z}[BP_0(z) + DP_1(z)] \tag{6}$$

where

$$\underset{1}{P_0}(z) = \langle \underset{1}{I_0}(z) \underset{1}{I_0^*}(z) \rangle$$

$$A = \langle g_{11} g_{11}^* \rangle$$

$$B = \langle g_{12} g_{12}^* \rangle$$

$$D = \langle g_{22} g_{22}^* \rangle$$

since $\langle g_{11} g_{12}^* \rangle = \langle g_{12} g_{11}^* \rangle = \langle g_{12} g_{22}^* \rangle = \langle g_{22} g_{12}^* \rangle = 0$, being odd functions of c. Expanding in powers $\langle c^2 \rangle$ and Δz and requiring

$$\lim_{\substack{<c^2> \to \infty \\ \Delta z \to 0}} \langle c^2 \rangle \Delta z = S_0$$

and after considerable algebraic manipulation, which we omit, we obtain (3) and (4)

$$\underset{1}{P_0}(z) = (-2\alpha_0 - S_0) \underset{1}{P_0}(z) + S_0 \underset{0}{P_1}(z).$$

These equations may be related to Miller's equations by replacing our P_0, P_1, $2\alpha_0$, $2\alpha_1$, and S_0 by his $P_1 + P_n$, P_z, $a_{1h} + a_{nh}$, a_{zh}, and a_{1z}.

III. ACKNOWLEDGMENT

REFERENCES

1. Miller, S. E., Coupled Wave Theory and Waveguide Applications, BSTJ, **33,** May, 1954, pp. 661–720.
2. Schelkunoff, S. A., Conversion of Maxwell's Equations into Generalized Telegraphist's Equations, BSTJ, **34,** September, 1955, pp. 995–1043.
3. Morgan, S. P., Theory of Curved Circular Waveguide Containing an Inhomogenous Dielectric, BSTJ, **36,** September, 1957, pp. 1209–1251.
4. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguides with Random Imperfections, BSTJ, **41,** May, 1962, pp. 1062–1068.
5. Miller, S. E., Waveguide as a Communication Medium, BSTJ, **33,** November, 1954, pp. 1209–1265.
6. Bellman, Richard, *Introduction to Matrix Analysis*, McGraw-Hill Book Co., Inc., New York. 1960.

Interchannel Interference in FM Systems Produced by Continuous Random Mode Conversion in Circular Electric Waveguide

By L. H. ENLOE

Long-distance, high-capacity transmission via the TE_{01} mode in circular waveguide is an attractive goal because the theoretical attenuation caused by heat loss decreases monotonically as the frequency of operation increases. However, continuous random mode conversion caused by the manufacturing or the laying of the waveguide causes severe random fluctuations with frequency in the transfer function of present-day waveguide, a fact which apparently eliminates all but the "toughest" modulation schemes, such as pulse code modulation and angle modulation.

In this paper, we present the derivation of a technique for analyzing the effects of continuous random mode conversion on an angle-modulated wave. The random coupling coefficient is assumed to have a Gaussian probability density. The modulating signal is assumed to be a single-sideband frequency division multiplex which may be simulated by a band of random Gaussian noise. The quantity of interest is the expected value of the interchannel interference noise which appears at baseband. A self-contained section on principal results and examples is presented. The examples use typical state-of-the-art data and are significant in that they demonstrate that FM appears to be an attractive modulation scheme. They are, however, not meant to be an exhaustive study of the problem. Copper waveguide (with and without mode filters) and helix waveguide are considered.

The Fourier series, multiple-echo representation of a stochastic transfer function presented in this paper should prove useful in handling other modulation schemes as well as other random channels.

I. INTRODUCTION

Long-distance, high-capacity transmission via the TE_{01} mode in the millimeter-wave circular-electric waveguide is an attractive goal be-

cause the theoretical attenuation caused by heat loss decreases monotonically as the frequency of operation increases. However, in order to obtain sufficiently small heat loss attenuation and delay distortion, the operating frequency must be well above the TE_{01} cutoff frequency and far into the multimode region. For example, a two-inch I.D. perfect copper circular guide will have a theoretical TE_{01} heat loss of 1.54 db per mile at 55 kmc and will propagate 223 additional (spurious) modes.[1] Real waveguides are never perfect right-circular cylinders but have continuous random imperfections arising from the manufacturing or the laying of the waveguide. These random imperfections cause conversion of the TE_{01} signal mode to spurious modes.[2] We call the attenuation produced by this energy loss "conversion attenuation." Usually more important, the imperfections also cause converted energy in the spurious modes to be reconverted into the TE_{01} signal mode. Since the group velocities of the various modes will be different, the reconverted energy from each mode will be delayed or advanced in time with respect to the energy which has remained in the TE_{01} signal mode. Thus we have a signal plus a continuum of echoes. The effect causes the amplitude and phase of the TE_{01} wave to fluctuate randomly with frequency. A transfer function which is a random variable can be defined to take this fluctuation into account in a statistical sense. We call it the "reconversion transfer function."

The fluctuations in the reconversion transfer function distort any signal which might be transmitted through the waveguide. A "tough" modulation scheme is required in order to overcome this effect. In this paper we consider the use of large-index frequency modulation (FM). The recent work of Rowe and Warters[2] provides us with a sound and tractable mathematical analysis of continuous random mode conversion. We shall use the results of their work as our starting point. The baseband signal used to frequency or phase modulate the carrier wave is assumed to consist of a single-sideband frequency division multiplex of telephone channels. For analysis purposes we can approximate this baseband signal by a band of random Gaussian noise.[3] When such an angle-modulated wave is passed through a network whose attenuation and phase are nonlinear functions of frequency, interchannel interference appears in the demodulated signal. In way of explanation, imagine that the energy in the baseband modulating signal corresponding to some particular telephone channel is removed. When such a signal is impressed on a frequency modulator and the resulting wave is transmitted through the network in question, detected, and finally demodulated, the received output in the originally clear channel represents interchannel interfer-

ence. Measurements of this type have been discussed by Albersheim and Schafer.[4]

To the author's knowledge, there is only one tractable technique for determining interchannel interference in angle-modulated systems which does not require that the fluctuation rate of the attenuation and phase functions be slow over the bandwidth occupied by the signal. This exception is the method of equivalent echoes.[5,6] Since the phase and attenuation of the reconversion transfer function fluctuates rapidly across the bandwidths of interest, we are necessarily confined to this technique or variations thereof. We see that some sort of multiple-echo approximation to the reconversion transfer function is required if any accuracy in our results is to be expected. How many echoes are required, and how should they be distributed and weighted? The answer is clear when we recall that a small echo produces or is equivalent to a sinusoidal variation with frequency in the transfer function, i.e., the Fourier transform of the echo. We simply expand the reconversion transfer function in a Fourier series over a frequency interval which we know to be sufficiently large to contain all of the significant energy in the FM wave. The Fourier transform of this series is the desired train of echoes and is an exact representation over the bandwidth under consideration. The echoes are regularly spaced, and their amplitudes are proportional to the magnitude of the coefficients of the Fourier series. The problem is now reduced to obtaining the Fourier series coefficients, and until recently this would have been a major stumbling block. Fortunately, however, Rowe, Warters, and Young have recently made important contributions[2,7,8] to the understanding of continuous random mode conversion. They have demonstrated that over the moderate percentage bandwidths in which we are interested, the effects of random mode conversion give rise to a transfer function which is a random function of frequency much as a noise wave is a random function of time. Using their results, we can evaluate the statistics of the Fourier series coefficients, i.e., echo amplitudes, which are, of course, random variables. Moreover, if the cross-correlation between coefficients is negligible, we can determine the expected value of the interchannel interference for each echo individually and then add the results to obtain the total expected value of the interchannel interference. The random coupling coefficient is assumed to have a Gaussian probability density. The effects of delay distortion caused by waveguide cutoff dispersion are not included in this paper. The system is assumed to be equalized.

Section II presents the principal results of this paper and contains examples demonstrating their use. The examples use typical state-of-

the-art data obtained from S. E. Miller[9] and are significant in that they demonstrate that FM appears to be an attractive modulation scheme for use on the millimeter-wave circular-electric waveguide as far as continuous random mode conversion is concerned. Both helix waveguide and copper waveguide with mode filters are considered. The helix waveguide has superior performance, of course, but copper waveguide with periodically placed mode filters may be useful in some situations.

Section III presents the mathematical development of the results.

II. PRINCIPAL RESULTS

The present section contains a statement and discussion of the principal results of this paper. We will restrict the discussion to the case of a single spurious mode and a single polarization. If other polarizations are present, they can be considered individually and the results added on a power basis.* The same is true for our purposes for different spurious modes. It is certainly true when the spurious modes arise because of different and uncorrelated physical imperfections in the waveguide. For instance, the TE_{02} spurious mode is caused primarily by random diameter variations. Its reconversion transfer function would be uncorrelated with that of the TE_{11} spurious mode, which is caused primarily by random straightness variations, if the diameter variations are uncorrelated with the straightness variations. Even when the spurious modes arise because of the same physical imperfections, it can be shown† that the cross-correlation in the same frequency interval is negligible.

There are two cases of interest. In the first, the difference between the heat loss attenuation constant of the TE_{01} and the spurious mode, i.e., the differential attenuation constant ΔA, is uniform. An example would be the TE_{0n} family on both the helix waveguide and copper waveguide with helix mode filters. The second case of interest is when the differential attenuation is effectively zero, except at periodically located points along the waveguide where a total differential attenuation of k is introduced. An example of this type of behavior would be the TE_{1n} family on a waveguide having periodically placed helix mode filters.

2.1 *Uniform Differential Attenuation Constant*

Consider a length L of waveguide having a differential attenuation constant of ΔA for the spurious mode of interest. Assume that the random coupling coefficient between the TE_{01} and spurious mode is Gauss-

---

* See Ref. 2, Section 4.3.2.
† See Ref. 2, Appendix G.

ian, pure real or pure imaginary, and has a white power spectrum of density S_c. If the differential attenuation of the total length of wave-guide is much larger than one, i.e., $|\Delta A| L \gg 1$, then the mean-square value of the equivalent echoes $\overline{r_n^2}$, normalized by the signal power, is given by the expression

$$\overline{r_n^2} = (T/\tau_0)(P_x/P_0)\epsilon^{-n(T/\tau_0)} \tag{1}$$

if $P_x/P_0 \ll 1$, where

$$P_x/P_0 = \text{reconversion echo-to-signal power ratio} \tag{2}$$

$$= S_c^2 L/2 |\Delta A| \tag{2}$$

$$\tau_0 = |\bar{v}_x^2 - \bar{v}_0^2|/4c |\Delta A| \tag{3}$$

and

n = positive integer designating the nth echo. (The echoes lag the signal if $\bar{v}_x^2 > \bar{v}_0^2$ and lead if $\bar{v}_0^2 > \bar{v}_x^2$. Because the interchannel interference is independent of this, we use the absolute value $|\bar{v}_x^2 - \bar{v}_0^2|$ in (3) and require n to be positive.)

P_x = reconversion echo power, i.e., the power in that portion of the signal which has been converted and reconverted once and only once.

P_0 = signal power which has suffered no conversion. This will hence-forth be considered *the* signal power.

$\bar{v}_0$, $\bar{v}_x$ = mode cutoff factors* of the TE_{01} and spurious modes at the band center frequency. This give rise to the differential phase shift.

c = velocity of light in free space.

$T \equiv 1/\Delta f$ = time delay between successive echoes.

Equation (2) has also been derived independently by Miller,[9] using a different approach. Equation (1) was developed from a Fourier series expansion of the reconversion transfer function over the frequency interval Δf. Hence, Δf must satisfy certain conditions if meaningful answers are to be obtained. First, all of the significant energy in the angle-modulated signal must be contained within the bandwidth Δf. Thus we require a bandwidth of

$$\Delta f \geqq 2(f_b + 4\sigma) \tag{4}$$

where f_b is the baseband bandwidth and σ is the root-mean-square fre-quency deviation of the angle-modulated wave. This is simply a state-ment of the familiar Carson bandwidth. Second, the mathematical

* See Ref. 2, Appendix A.

derivation of (1) requires the differential phase constant to be approximately proportional to frequency over the bandwidth used in the Fourier series expansion. This sets an upper limit on Δf. In order to satisfy this requirement we require that the percentage bandwidth and the square of the mode cutoff factors of the modes involved be small compared to unity, i.e.

$$\Delta f / 2 f_0 \ll 1, \quad \bar{v}_0^{\,2} \ll 1 \quad \text{and} \quad \bar{v}_x^{\,2} \ll 1. \tag{5}$$

This will usually be true for the bandwidth and modes of interest, but it should always be checked as a safeguard. Third, the Fourier series coefficients must be uncorrelated with one another. It is well known that Fourier series coefficients become uncorrelated as the expansion interval approaches infinity. In a practical case, "infinity" is a number which is much larger than the correlation interval of the random variable in question. In the present case the autocorrelation function of the reconversion transfer function varies as $1/[1 + (2\pi\tau_0 f')^2]$, where f' is measured about the band center frequency. Thus in order for the cross-correlation between coefficients to be negligible we require that

$$(\pi\tau_0\Delta f)^2 \gg 1. \tag{6}$$

Lastly, the expression (1) for $\overline{r_n^{\,2}}$ is limited by the accuracy of the perturbation theory from which it was developed. It should be good, however, as long as $P_0/P_x \gg 1$ as calculated from (2). A plot of the equivalent echo train is shown in Fig. 1. We see that τ_0 is the "time constant" associated with the envelope of the echo train, i.e., it is the value of time delay where the envelope of the echo train caused by reconversion has dropped to the relative value $1/\epsilon$.

If the cross-correlation between Fourier series coefficients, and hence echoes, is negligible, we may substitute the expression (1) directly
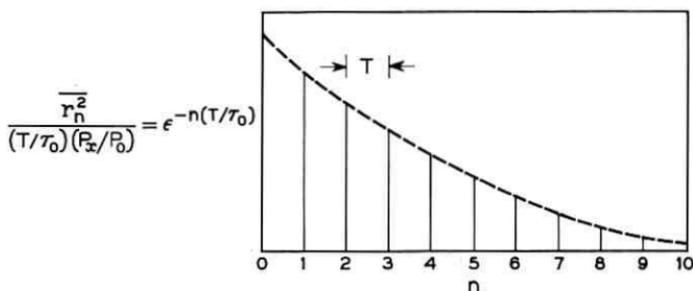
Fig. 1 — A plot of the reconversion echo train for uniform differential attenuation.

into the formulas given by Bennett et al.[5] and sum the results to determine the total baseband interchannel interference in the top channel.

As an illustration, for a two-inch I.D. helix waveguide at 55 kmc, the dominant spurious mode is the TE_{02}, as far as reconversion echoes are concerned. Typical data are[9]

$$| \Delta A | = 4.3 \text{ db/mile} = 0.495 \text{ neper/mile}$$
$$S_c = 0.01 \text{ db/mile} = 0.00115 \text{ neper/mile}$$
$$\bar{v}_0^2 = 0.0172 \ll 1$$
$$\bar{v}_x^2 = 0.058 \ll 1$$
$$\tau_0 = 105.8 \times 10^{-7} \text{ sec.}$$

For a waveguide length of 4000 miles, we have from (2)

$$P_0/P_x = 22.6 \text{ db} \gg 1.$$

Let us assume that a signal is applied to this waveguide which has been frequency modulated by a rectangular band of Gaussian noise. Let the baseband width be $f_b = 10$ mc, and the rms frequency deviation be $\sigma = 2f_b = 20$ mc. The significant energy in this wave would be contained in a bandwidth of $\Delta f = 10\sigma = 20f_b = 200$ mc. Thus, if we pick $\Delta f = 200$ mc we also satisfy (5) and (6), i.e.

$$\Delta f/2f_0 = 0.002 \ll 1, \qquad (\pi\tau_0\Delta f)^2 = 4,450 \gg 1.$$

From Fig. 5.7 of Bennett et al.[5] we can read off the factor $s(nf_bT)$, which multiplies the mean-square value of each echo to give the interchannel interference-to-signal power ratio for that particular echo. $s(nf_bT)$ takes into account the noise suppressing property of FM as well as the variation caused by the magnitude of the delay. A plot of $s(nf_bT)$ is shown in Fig. 2 for $\sigma/f_b = 2$ and 3. The total interchannel interference-to-signal power ratio is given by the sum

$$I/S = \sum_{n=1}^{\infty} s(nf_bT)\overline{r_n^2} \tag{7}$$

if $P_x/P_0 \ll 1$. We see that since $s(nf_bT)$ is essentially constant at -15 db for $nf_bT > 1/10$, i.e., $n > 1$, we have to a good approximation

$$I/S = s(nf_bT) \sum_{n=1}^{\infty} \overline{r_n^2}$$
$$= s(nf_bT)(P_x/P_0). \tag{8}$$

In decibels

$$I/S = -15 - 22.6 = -37.6 \text{ db.}$$
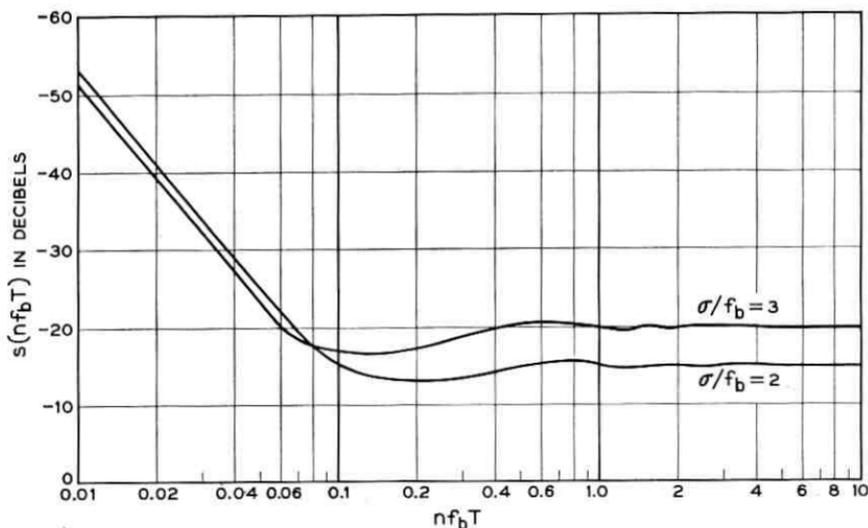
Fig. 2 — Weighting function $s(nf_bT)$ which multiplies individual echoes to give interchannel interference in top baseband channel for FM.

This corresponds to 27.4 dba at zero level. Thus, other things neglected, continuous random mode conversion does not preclude a 4000-mile two-inch I.D. helix system.

If this interchannel interference noise-to-signal ratio were too large, one could decrease it still further simply by increasing the modulation index. In other words, for the conditions of this example we are able to obtain FM improvement against reconversion echoes as well as additive thermal noise. However, from Fig. 2 we see that for $nf_bT \ll 1/10$ we would not be able to obtain this improvement. In fact, increasing the modulation index degrades the situation in that case.

2.2 *Mode Filters*

Mode filters may be placed at periodic intervals along the waveguide to provide large attenuation to the spurious modes. We assume here that each filter contributes k nepers of differential attenuation and that the waveguide between filters contributes essentially none. Let the spacing between mode filters be L_0 and the length of each mode filter be negligible in comparison. In the previous section, for uniform differential attenuation we found that the envelope of the equivalent echoes, when plotted versus delay, was an exponential. In the present case we find that the envelope consists of a piecewise approximation to an

exponential using straight line segments (see Fig. 3). The exponential is ϵ^{-2ik} and the straight line segments connect the points defined by letting $i = 0, 1, 2, \cdots$, etc. Let us number the trapezoids represented by the area under each straight line segment according to the value of i at the beginning of each trapezoid. The equation for the straight line in the ith trapezoid, i.e., between ϵ^{-2ik} and $\epsilon^{-2(i+1)k}$ on the ordinate, is

$$y_i(\tau/\tau_1) = \epsilon^{-2ik}[1 + [1 - \epsilon^{-2k}][i - (\tau/\tau_1)]] \tag{9}$$

where τ_1 is the length of the base of each trapezoid. The expression for the individual echoes is

$$\overline{r_{n_i}^2} = \frac{TS_c^2 LL_0}{\tau_1} \epsilon^{-2ik}[1 + [1 - \epsilon^{-2k}][i - (nT/\tau_1)]] \tag{10}$$

where i is chosen from the inequality $i \leqq (nT/\tau_1) \leqq i + 1$. Here L_0 is the length of waveguide between mode filters, L is the total length of waveguide and $\tau_1 \equiv |\bar{v}_x^2 - \bar{v}_0^2| L_0/2c$. All other quantities are as defined in the previous section. As in the previous section, we require that

$$\bar{v}_0^2 \ll 1, \qquad \bar{v}_x^2 \ll 1, \qquad \Delta f/2f_0 \ll 1 \quad \text{and} \quad \Delta f \geqq 2(f_b + 4\sigma). \tag{11}$$

The requirement on Δf to ensure negligible cross-correlation between the Fourier coefficients of the reconversion transfer function, i.e., between individual echoes, is changed, however. The autocorrelation function is given by (42). The series converges quite rapidly for large
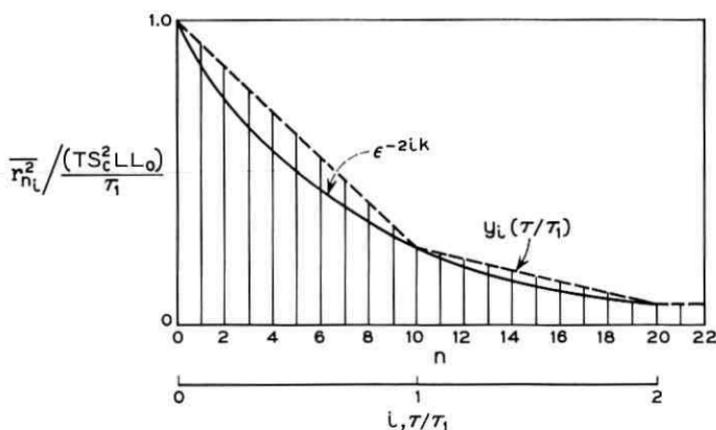
Fig. 3 — For periodically placed mode filters of k nepers differential attenuation each, the envelope of the equivalent echoes consists of a straight-line approximation $y_i (\tau/\tau_1)$ to the exponential ϵ^{-2ik}.

differential attenuation per mode filter k. In most cases of interest only the first term or two is required. In order for Δf to be much larger than the correlation interval we require

$$(\pi\tau_1\Delta f/2)^2 \gg 1 \tag{12}$$

since $|\zeta| L_0 = f\tau_1$.

Equation (10) requires the number of mode filters N to be large, i.e., $N \gg 1$, and that the square of the *total* differential loss of the N filters be large, i.e., $\epsilon^{-2kN} \ll 1$. Under these conditions, the reconversion echo-to-signal power ratio is given by

$$P_x/P_0 = (LL_0S_c^2/2) \coth k \tag{13}$$

which must be small compared to unity in order for (10) to be accurate. As a check, notice that if $k \ll 1$, we have

$$P_x/P_0 = LS_c^2/2 \, | \Delta A | \tag{14}$$

where $| \Delta A | \equiv k/L_0$. This is precisely the expression presented in the previous section for a constant differential attenuation.

It is frequently necessary to determine the ratio of the power in the ith trapezoid to the total. From simple geometric considerations this is given by

$$P_{x_i}/P_x = \epsilon^{-2ik} [1 - \epsilon^{-2k}] \tag{15}$$

Equations (13) and (15) have also been derived independently by S. E. Miller[9] using a different approach.

As an illustration, consider a two-inch I.D. copper waveguide system at 55 kmc having 15-db mode filters every 200 feet. The TE_{11}, TE_{12}, and TM_{11} spurious modes would all contribute significantly because of random straightness deviations. The energy in all of these modes will be lumped together and considered as if it were in the TE_{12} mode. This mode has the largest delay and consequently would produce the most interchannel interference. Thus we are assured that our calculations will be pessimistic. Numerical values were obtained from Miller[9]

$$k = 15 \text{ db} = 1.73 \text{ nepers}$$
$$L_0 = 200 \text{ feet} = 1/26.4 \text{ miles}$$
$$S_c = 2 \text{ db/mile} = 0.23 \text{ neper/mile}$$
$$\bar{v}_0^2 = 0.0172 \ll 1$$
$$\bar{v}_x^2 = 0.0334 \ll 1$$
$$\tau_1 = 16.08 \times 10^{-10} \text{ sec.}$$

Equation (10) requires that the signal-to-echo ratio be large compared to unity. If we arbitrarily select $P_0/P_x = 17$ db, we find from (13) that the permissible length of waveguide is $L = 18.7$ miles. If we now demodulate and pass the resultant signal through a baseband filter of bandwidth f_b to obtain the FM advantage, we obtain a new baseband signal not unlike the original. Thus we can remodulate and proceed for another 18.7 miles without analytical complications, as long as the baseband signal-to-interchannel interference ratio remains large. It should be emphasized that placing a demodulate-remodulate repeater every 18.7 miles is done to keep out of analytical difficulties. We have not shown and are not implying that performance requires this crutch.

As in the previous section, assume an FM system with $f_b = 10$ mc and $\sigma/f_b = 2$. We choose Δf to satisfy the inequalities (11) and (12). Picking $\Delta f = 2$ kmc yields $\Delta f/2f_0 = 0.018 \ll 1$, $\Delta f > 2 (f_b + 4\sigma) = 200$ mc and $(\pi\tau_1\Delta f/2)^2 = 25 \gg 1$. In order to calculate interchannel interference, we might again multiply each echo by the factor $s(nf_bT)$ given by Fig. 5.7 of Bennett et al.[5] and sum the result. However, that paper gives the analytical expression

$$s(\tau) = 2(\sigma/f_b)^2 (\pi f_b \tau)^4 \qquad (16)$$

which is valid for the small delays of this illustration, i.e., $2\pi f_b nT \ll 1$ and $(2\pi f_b nT)^2 (\sigma/f_b)^2 \ll 1$. We can integrate rather than sum. Thus in (10) we replace nT by τ and T by $d\tau$. For the interchannel interference-to-signal ratio contributed by the ith trapezoid, we have

$$\begin{aligned}
I_i/S &= \left[\frac{2LL_0(S_c\sigma f_b)^2\pi^4}{\tau_1}\right] \epsilon^{-2ik} \int_{i\tau_1}^{(i+1)\tau_1} \tau^4 \\
&\quad \{1 + (1 - \epsilon^{-2k})[i - (\tau/\tau_1)]\}\, d\tau \\
&= \left[\frac{2LL_0(S_c\sigma f_b)^2\pi^4\tau_1^4\epsilon^{-2ik}}{30}\right] [\{6 + (i - 5)(1 - \epsilon^{-2k})\} (1 + i)^5 \\
&\quad - \{6 + i (1 - \epsilon^{-2k})\} i^5].
\end{aligned}$$

It turns out, in this example, that the $i = 0$, 1 and 2 trapezoids all contribute significantly, with the $i = 1$ trapezoid producing the most interchannel interference. In physical terms, the value of i tells us how many mode filter spacings the converted energy traveled in the spurious mode before reconverting. The ith trapezoid represents energy which traveled a distance larger than i mode filters spacings but less than $i + 1$ spacings. In this example the $i = 1$ trapezoid has substantially less power in it than the $i = 0$ trapezoid, but its larger delay more than makes up

for this difference (see Fig. 2). The total signal-to-interchannel inter-ference ratio in the top channel is $S/I = 66.4$ db. For a total system length of 4000 miles, we would need $4000/18.7 = 214$ repeaters and would obtain a signal-to-interchannel interference ratio of $66.4 - 10 \log_{10} 214 = 43.1$ db if the only contribution were due to continuous random mode conversion. This corresponds to 22 dba at zero level.

This is only part of the story. The TE_{0n} spurious modes are not damped by helix mode filters. If it is conjectured[9] that the coupling coefficients, differential attenuation, etc., for these modes are the same for copper and helix waveguides, then the calculations and results of the previous section apply directly. The interchannel interference pro-duced by the TE_{02} spurious mode predominates over that from all other modes. Consequently, it looks as though the performance of this copper waveguide with mode filters is comparable with that of the helix wave-guide. The imposed requirement of demodulate-remodulate repeaters, however, is certainly a disadvantage. One must determine whether or not they are really necessary before any definite comparison can be made.

III. MATHEMATICAL DEVELOPMENT

Our approach is to represent the effects of continuous random mode conversion by an infinite number of uncorrelated echoes. These echoes follow directly from the expansion of the reconversion transfer function in a Fourier series over the frequency interval of interest. The coupled line equations serve as our starting point. They are

$$dI_0(z)/dz = -\Gamma_0 I_0(z) + jc(z)I_x(z)$$
$$dI_x(z)/dz = +jc(z)I_0(z) - \Gamma_x I_x(z)$$

where

$I_0(z), I_x(z)$ = complex wave amplitude of the TE_{01} mode and spuri-ous mode, respectively,

Γ_0, Γ_x = deterministic propagation constants of the TE_{01} and spuri-ous modes, respectively,

$c(z)$ = random coupling coefficient, assumed Gaussian and pure real or pure imaginary, and

z = distance along the waveguide.

If one imposes the initial conditions $I_0(0) = 1$ and $I_x(0) = 0$, then $I_0(z)$ is the transfer function of the waveguide. This may be factored into two components. One component, $\epsilon^{-\Gamma_0 z}$, is deterministic and repre-

sents the attenuation and phase which would be present in the absence of random mode conversion. Since in theory it is completely predictable and equalizable, we shall not consider this component further. The other component is a random variable and represents the attenuation and phase which is contributed by random mode conversion. This is by definition the reconversion transfer function and is given by

$$G_0(z) \equiv \epsilon^{+\Gamma_0 z} I_0(z). \tag{19}$$

Rowe[7] shows that it may be approximated directly by use of Picard's method of successive approximations, or with improved accuracy it may be obtained from the reconversion propagation function $\Lambda(z)$ defined by the relation $G_0(z) \equiv \epsilon^{-\Lambda(z)}$, where $\Lambda(z)$ is approximated by Picard's method. We use Rowe's second approximation throughout this paper

$$\Lambda(z) = \int_0^z c(s)\epsilon^{+\Delta\Gamma s} \, ds \int_0^s c(t)\epsilon^{-\Delta\Gamma t} \, dt.$$

The reconversion propagation function $\Lambda(z)$ may be written as $\Lambda(z, \xi, \Delta A)$, where $\xi \equiv \Delta\beta/2\pi \equiv (\beta_0 - \beta_x)/2\pi$, $\Delta A \equiv A_0 - A_x$ and $\Gamma_0 \equiv A_0 + j\beta_0$, $\Gamma_x \equiv A_x + j\beta_x$. Notice that ΔA will always be negative since A_x is positive and larger than A_0. Λ is nonstationary when considered as a function of z. Under certain conditions, however, Λ is approximately stationary, when considered as a function of ξ, over the small percentage bandwidths of interest.[2] The requirement is that the random coupling coefficient $c(z)$ have a white power spectrum; that is, the autocorrelation function $R_c(u) \equiv \overline{c(z)c(z + u)}$ be an impulse function or a close approximation thereof. The bar denotes an ensemble average. This seems to be true or at least approximately true for good waveguides over the moderate bandwidths required in our problem. Fortunately, it also is true that over the small percentage bandwidths in which we are interested, ξ is approximately proportional to frequency;* that is,

$$\xi \equiv \xi_0 + \xi' \doteq m \, (f_0 - f') \tag{20}$$

where ξ_0 and f_0 are measured at the band center, and $\xi \equiv \xi_0 + \xi'$ and $f \equiv f_0 + f'$. $m \equiv (\bar{v}_x^2 - \bar{v}_0^2)/2c$ where $\bar{v}_0$ and $\bar{v}_x$ are the mode cutoff factors of the TE_{01} and spurious modes, respectively, evaluated at the band center frequency f_0, and c is the velocity of light in free space.

---

* From Ref. 2, Appendix A, $\xi = (f/2c)[(1 - v_0^2)^{\frac{1}{2}} - (1 - v_x^2)^{\frac{1}{2}}] \doteq (f/4c)(v_x^2 - v_0^2) = \text{constant}/f$ if $v_0^2 \ll 1$ and $v_x^2 \ll 1$. Letting $f = f_0 + f'$ and $\xi = \xi_0 + \xi'$ we have $\xi' + \xi_0 \doteq (\text{constant}/f_0) (1 - f'/f_0) \equiv m \, (f_0 - f')$.

Thus we may determine the mean-square value of the coefficients of the Fourier series expansion of the reconversion propagation function Λ, when Λ is considered as a function of frequency, as a limiting case of the power spectrum of Λ, when Λ is considered as a function of ξ. The functional form of Λ depends upon whether ξ or f is the variable. For this reason we shall use a subscript 1 when ξ is the variable, and a subscript 2 when f is the variable.

Formulas for the power spectrum $S_1(u)$ of the random variable $\Lambda_1(\xi)$ are developed for the case of $\Delta A = 0$ by Rowe and Warters,[2] for $\Delta A = $ constant by Young,[8] and for the case of periodically placed mode filters later in this paper.

3.1 Transfer Function and Equivalent Echoes

The complex reconversion propagation function $\Lambda_2(f)$ may be expanded in a Fourier series over the interval Δf centered at f_0

$$\Lambda_{2s}(f) = \sum_{n=-\infty}^{+\infty} a_n \epsilon^{j2\pi n f T} \tag{21}$$

where $\Lambda_{2s}(f) = \Lambda_2(f)$ for $f_0 - \Delta f/2 \leq f \leq f_0 + \Delta f/2$. $T = 1/\Delta f$ and the subscript s reminds us that we have a periodic series approximation. If all of the significant energy in the FM wave, or any other wave for that matter, lies within Δf, then the periodic series (21) may be used for calculations. The $n \neq 0$ terms represent fluctuation components. If their sum is always small compared to unity with large probability, then the reconversion transfer function $G_0(f) \equiv \exp [-\Lambda_2(f)]$ can be approximated by the series

$$G_{0s}(f) = \exp \left[- \sum_{n=-\infty}^{+\infty} a_n \epsilon^{j2\pi f n T} \right] \doteq \epsilon^{-a_0} \left[1 - \sum_{n=-\infty}^{+\infty}{}' a_n \epsilon^{j2\pi f n T} \right] \tag{22}$$

where the prime on the summation indicates that the $n = 0$ term is omitted. The impulse response of this periodic transfer function, i.e., its inverse Fourier transform, is

$$h_{0s}(t) = \epsilon^{-a_0} \left[\delta(t) - \sum_{n=-\infty}^{+\infty}{}' a_n \delta(t + nT) \right]. \tag{23}$$

Thus we have an undistorted signal plus a series of small echoes.

The coefficients a_n are given by

$$a_n = (1/\Delta f) \int_{f_0-\Delta f/2}^{f_0+\Delta f/2} \Lambda_2(f) \epsilon^{-j2\pi n f/\Delta f} \, df. \tag{24}$$

Replacing f by $f' + f_0$ we have

$$a_n = (1/\Delta f)\epsilon^{-j2\pi n f_0/\Delta f} \int_{-\Delta f/2}^{+\Delta f/2} \Lambda_2(f' + f_0)\epsilon^{-j2\pi n f'/\Delta f} \, df'. \qquad (25)$$

Now, over the small percentage bandwidths in which we are interested, the differential phase constant is proportional to frequency: i.e., $\xi \equiv \xi' + \xi_0 \doteq m(f_0 - f')$. Since $\Lambda_2(f) = \Lambda_1(\xi)$, we may rewrite (25) as

$$a_n = \epsilon^{-j2\pi n f_0/\Delta f} \left\{ \frac{1}{m\Delta f} \int_{-m\Delta f/2}^{+m\Delta f/2} \Lambda_1(\xi' + \xi_0)\epsilon^{-j2\pi n\xi \, /(-m\Delta f)} \, d\xi' \right. \qquad (26)$$

$$= \epsilon^{-j2\pi n f_0 T} b_{(-\mathrm{sgn}\, m)n}$$

where b_n is the coefficient of the nth term of the Fourier series expansion of $\Lambda_1(\xi' + \xi_0)$ over the interval $-\Delta\xi/2 \leqq \xi' \leqq +\Delta\xi/2$. For $\Delta\xi$ much larger than the correlation interval of $\Lambda_1(\xi)$, we may approximate the expected value of $|b_n^2|$, and hence $|a_n^2|$, in terms of the power spectrum of $\Lambda_1(\xi)$. The result is

$$\overline{|a_n^2|} = (T/|m|)S_{\Lambda_1}(-nT/m). \qquad (27)$$

The imaginary part of the complex random variable $\Lambda_1(\xi)$ is the Hilbert transform[2,8] of the real part, so the power spectrum $S_{\Lambda_1}(u)$ vanishes for negative u. Thus, the coefficients a_n are nonzero only when the sign of n opposes that of m. Negative n indicates that the echoes lag the signal and vice versa. Since the interchannel interference is independent of whether the echoes lead or lag, we may drop the minus sign in (27), use the absolute value of m, and restrict n to positive values.

3.2 *Uniform Differential Attenuation*

The power spectrum of the real part of the reconversion propagation constant $\Lambda_1(\xi)$ is given by Rowe and Warters,[2] for the case of zero differential attenuation and a Gaussian random coupling coefficient, as

$$S(u) = \begin{cases} (\tfrac{1}{4})[\delta(u)L^2S_c^2 + S_c^2[L - |u|]], & |u| \leqq L \\ 0, \text{ otherwise.} \end{cases} \qquad (28)$$

Since the real and imaginary parts of $\Lambda_1(\xi)$ are Hilbert transforms, the power spectrum of $\Lambda_1(\xi)$ itself is given by*

$$S_{\Lambda_1}(u) = \begin{cases} (\tfrac{1}{4})L^2S_c^2\delta(u) + S_c^2[L - u], & (0-) \leqq u \leqq L \\ 0, \text{ otherwise.} \end{cases} \qquad (29)$$

---

* Discussions of the autocorrelation function and power spectrum of complex random variables, whose real and imaginary parts are Hilbert transforms, are given by Dugundji,[10] Zakai,[11] and Deutsch.[12]

Young[8] demonstrates that if the differential attenuation is nonzero but is a constant equal to ΔA, the real and imaginary parts of $\Lambda_1(\xi)$ are again Hilbert transforms and its power spectrum is found by multiplying that for $\Delta A = 0$ by $\epsilon^{-2|\Delta A u|}$. If $|\Delta A| L \gg 1$ we have

$$S_{\Lambda_1}(u) \doteq \begin{cases} (S_c^2 L^2/4)\delta(u) + S_c^2 L \epsilon^{-2|\Delta A|u}, & (0-) \leqq u < \infty \\ 0, \text{ otherwise.} \end{cases} \tag{30}$$

Equation (30) can also be obtained as a limiting case of the results of the next section.

It is often convenient to work in terms of the reconversion echo-to-signal ratio P_x/P_0, defined as the ratio of the power in the echoes in (23) to that in the signal. This is given by

$$\begin{aligned} P_x/P_0 &= \sum_{n=1}^{\infty} \overline{|a_n^2|} = \sum_{n=1}^{\infty} (T/|m|)S_{\Lambda_1}(nT/|m|) \\ &\rightarrow \int_{0+}^{\infty} (1/|m|)S_{\Lambda_1}(\tau/|m|)\, d\tau \\ &= S_c^2 L/2 |\Delta A|. \end{aligned} \tag{31}$$

3.3 *Mode Filters*

Consider a long section of waveguide of length L. Assume ideal mode filters are placed at $L_0, 2L_0, \cdots, NL_0 = L$. Let each mode filter be of zero length and produce k nepers of differential attenuation and a constant differential phase shift of θ radians, with no other side effects. If the differential attenuation of the length of waveguide L_0 between each mode filter is negligible in comparison with that produced by a mode filter, we may approximate the differential propagation function of the waveguide by the relation

$$\Delta\Gamma(z) = j2\pi\xi + (-k + j\theta)\sum_{n=1}^{N}\delta(z - nL_0) \tag{32}$$

where $2\pi\xi = \Delta\beta = $ differential phase constant of the waveguide alone.

In the coupled line equations (18) the propagation factors Γ_0 and Γ_z are constants. In the present case, however, they are functions of position along the waveguide, i.e., functions of z. Rowe has shown in unpublished work that in (18) we may replace

$$\Gamma_0 z \quad \text{by} \quad \gamma_0(z) = \int_0^z \Gamma_0(S)\, dS$$

and

$$\Gamma_{xz} \quad \text{by} \quad \gamma_x(z) = \int_0^z \Gamma_x(S) \, dS,$$

and use Picard's method of successive approximations to obtain the approximate solution

$$\Lambda(L) = \int_0^L c(t)\epsilon^{\Delta\gamma(t)} \, dt \int_0^t c(S)\epsilon^{-\Delta\gamma(S)} \, dS \qquad (33)$$

The derivation and result are exactly the same as Rowe's published derivation[7] for constant Γ's with the exception of the substitution mentioned. Substituting (32) into (33) we have

$$\Lambda(NL_0)$$

$$= \int_0^{NL_0} c(t) \exp\left[j2\pi\xi t - (k - j\theta) \sum_{n=1}^N u(t - nL_0) \right] dt \qquad (34)$$

$$\cdot \int_0^t c(S) \exp\left[-j2\pi\xi S + (k - j\theta) \sum_{l=1}^N u(S - lL_0) \right] dS,$$

where $u(x) = \begin{cases} 1, & 0 \leq x \leq \infty \\ 0 & \text{otherwise.} \end{cases}$

Simplifying and rearranging, we have

$$\Lambda(NL_0) = \sum_{n=0}^{N-1} \sum_{l=0}^{n-1} \epsilon^{(-n+l)(k-j\theta)} \int_{nL_0}^{(n+1)L_0} c(t)\epsilon^{j2\pi\xi t} \, dt$$

$$\cdot \int_{lL_0}^{(l+1)L_0} c(S)\epsilon^{-j2\pi\xi S} \, dS \qquad (35)$$

$$+ \sum_{n=0}^{N-1} \int_{nL_0}^{(n+1)L_0} c(t)\epsilon^{j2\pi\xi t} \, dt \int_{nL_0}^t c(S)\epsilon^{-j2\pi\xi S} \, dS$$

which after more rearranging and simplification reduces to

$$\Lambda(NL_0)$$

$$= \sum_{n=0}^{N-1} \sum_{l=0}^{n-1} \epsilon^{(-n+l)(k-j\theta-j2\pi\xi L_0)} \int_0^{L_0} c(t + nL_0)\epsilon^{j2\pi\xi t} \, dt$$

$$\cdot \int_0^{L_0} c(S + lL_0)\epsilon^{-j2\pi\xi S} \, dS \qquad (36)$$

$$+ \sum_{n=0}^{N-1} \int_0^{L_0} c(t + nL_0)\epsilon^{j2\pi\xi t} \, dt \int_0^t c(S + nL_0)\epsilon^{-j2\pi\xi S} \, dS.$$

If the random coupling coefficient $c(z)$ has a white power spectrum, the reconversion propagation function Λ is an approximately stationary random variable[2] when considered as a function of the differential phase variable ξ. We shall now calculate the autocorrelation function $R_{\Lambda_1}(\xi)$ of Λ. Since $c(z)$ has a white power spectrum of density S_c, it has an autocorrelation function which is an impulse, i.e.

$$R_c(\xi) = S_c\delta(\xi). \tag{37}$$

Thus, each term in the series (36) is uncorrelated with all others. The autocorrelation function of the sum is the sum of the autocorrelation functions of each individual term.

The autocorrelation function of a typical term

$$\int_0^{L_0} c(t + nL_0)\epsilon^{j2\pi\xi t}\, dt \tag{38}$$

is

$$R(\xi) = S_c L_0 \left(\frac{\sin \pi L_0 \xi}{\pi L_0 \xi}\right). \tag{39}$$

In view of the fact that $l < n$ in the double series of (36), the two integrals in each term are uncorrelated. The autocorrelation function of the double series is then

$$\sum_{n=0}^{N-1} \sum_{l=0}^{n-1} \epsilon^{2(l-n)(k-j\pi\xi L_0)}\, R^2(\xi) = \sum_{n=1}^{N-1} (N - n)\epsilon^{-2n(k-j\pi\xi L_0)}\, R^2(\xi) \tag{40}$$

since we have $N - 1$, $n - l = 1$ terms; $N - 2$, $n - l = 2$ terms; $\cdots$; 1, $n - l = N - 1$ terms.

The autocorrelation function of each term in the single series of (36) is the same as the autocorrelation function for a single piece of waveguide of length L_0. Thus, for $c(z)$ Gaussian, it is the inverse Fourier transform of the power spectrum given by (29), i.e., it is

$$\tfrac{1}{4} R^2(0) + \tfrac{1}{2} [R^2(\xi) + j\widehat{R^2}(\xi)] \tag{41}$$

where R is still defined by (39) and the circumflex stands for the Hilbert transform. Keeping in mind the fact that the dc terms, i.e., $R(0)$ terms, add in phase, we conclude

$$R_{\Lambda_1}(\xi) = (N^2/4)R^2(0) + (N/2)[R^2(\xi) + j\widehat{R^2}(\xi)] \\ + R^2(\xi) \sum_{n=1}^{N-1} (N - n)\epsilon^{-2n(k-j\pi\xi L_0)}. \tag{42}$$

In order to find the reconversion echo-to-signal ratio, we could integrate over the power spectrum of $\Lambda_1(\xi)$ as in (31). However, the alternative method of letting $\xi = 0$ in the autocorrelation function is easier. Note that the $0+$ limit on the integral in (31) corresponds to dropping the constant term, i.e., the $(\frac{1}{4})R^2(0)$ term in (42). Thus the reconversion echo-to-signal ratio becomes

$$P_x/P_0 = (N/2)R^2(0) + \sum_{n=1}^{N-1} (N - n)\epsilon^{-2nk} R^2(0).$$

If $N \gg 1$ and $\epsilon^{-2Nk} \ll 1$, then

$$P_x/P_0 \doteq \frac{LL_0S_c^2}{2} \coth k \tag{43}$$

where $L = NL_0$.

The power spectrum of $\Lambda_1(\xi)$ is the inverse Fourier transform of (42) and is given by

$$S_{\Lambda_1}(u) = (\tfrac{1}{4})N^2L_0^2S_c^2\delta(u)$$
$$+ S_c^2 \sum_{i=0}^{N-1} (N - i)\epsilon^{-2ik}((L_0 - |u - iL_0|)) \tag{44}$$

for $u \geqq (0-)$, where

$$((x)) = \begin{cases} x \text{ if } x > 0 \\ 0 \text{ if } x \leqq 0. \end{cases}$$

If $N \gg 1$ and $\epsilon^{-2Nk} \ll 1$, then the terms diminish quite rapidly. Thus, for $i \ll N$ we have the simpler expression

$$S_{\Lambda_1}(u) \doteq (\tfrac{1}{4})N^2L_0^2S_c^2\delta(u)$$
$$+ NL_0S_c^2\epsilon^{-2ik} [1 + (1 - \epsilon^{-2k})(i - (u/L_0))] \tag{45}$$

where $(-0) \leqq i \leqq u/L_0 \leqq i + 1$. A rough sketch of this function is shown in Fig. 3. The sketch consists of an impulse at the origin plus a straight-line approximation to an exponential.

It is often convenient to have an expression for the ratio of the reconversion echoes under the ith straight line, i.e., from $u = iL_0$ to $u = (i + 1)L_0$, to the total. This is given by

$$P_{x_i}/P_x = (P_{x_i}/P_0)/(P_x/P_0) = (1 - \epsilon^{-2k})\epsilon^{-2ik}. \tag{46}$$

If we allow the differential attenuation k per mode filter and the spacing L_0 between mode filters both to decrease without limit such that

their ratio remains finite, (43) reduces to (31), and (45) reduces to (30), where $|\Delta A| \equiv k/L_0$.

3.4 Interchannel Interference

Let the angle-modulated wave

$$e_i(t) = \epsilon^{j\omega_0 t + j\varphi(t)} \tag{47}$$

be applied to a filter having an impulse response given by (23), i.e., by

$$h_{0e}(t) = \epsilon^{-a_0}[\delta(t) - \sum_{n=-\infty}^{+\infty}{}' a_n \delta(t + nT)]. \tag{48}$$

$\varphi(t)$ is assumed to represent stationary random Gaussian noise. The output wave is found by convolution and is

$$e_0(t) = \epsilon^{-a_0}[\epsilon^{j\omega_0 t + j\varphi(t)} - \sum_{n=-\infty}^{\infty}{}' a_n \epsilon^{j\omega_0(t+nT) + j\varphi(t+nT)}]. \tag{49}$$

From (26) we have $a_n = \epsilon^{-jn\omega_0 T} b_{\pm n} \equiv \epsilon^{j\theta_n - jn\omega_0 T} r_n$, where r_n and θ_n are real. If there are a large number of sample points[2] contained in the small percentage bandwidths of interest, then to a very good approximation the θ_n's will be uniformly distributed in the interval $(0, 2\pi)$ and the b_n's will be uncorrelated. Substituting, (49) reduces to

$$e_0(t) = \epsilon^{-b_0}[\epsilon^{j\omega_0 t + j\varphi(t)} - \sum_{n=-\infty}^{+\infty}{}' r_n \epsilon^{j\omega_0 t + j\theta_n + j\varphi(t+nT)}]. \tag{50}$$

The output consists of the original signal plus an infinite number of small, uncorrelated echoes. If the total power contained in the echoes is much smaller than that contained in the signal, we may follow a procedure identical to that of Bennett et al.[5] to determine the baseband interchannel interference in the top channel as a function of time which results from each of the above echoes. We then determine the autocorrelation function of the interference and average over r_n and θ_n. Since the θ_n's are uniformly distributed, the interference caused by each echo adds on a power or mean-square basis. In their work, Bennett et al. carried along a deterministic term pT which would be replaced by θ_n in our case. This term appears in the final result as $\cos \theta_n$, which has an average value of zero. The curves of Fig. 5.7 of Ref. 5 are for $\cos pT = 0$ and so apply to our case directly.

IV. CONCLUSIONS

We have presented an analysis of the effects of continuous random mode conversion on an angle modulated wave. The modulating signal was assumed to be a band of random Gaussian noise which is used to simulate a single sideband frequency division multiplex. A technique was presented for determining the expected value of the interchannel interference noise which appears in the top channel at baseband. Examples using typical state-of-the-art data demonstrated that FM appears to be an attractive modulation scheme. Copper waveguide, with and without mode filters, and helix waveguide were considered.

It should be pointed out that the techniques and the results presented here should be applicable to other multimode and multipath situations virtually unchanged. The Fourier series characterization of the transmission channel should also be of use in the study of other modulation schemes.

Our work contains two key, but reasonable, assumptions which one must keep in mind:

(i) The perturbation technique upon which our work is based requires that the reconverted energy consist primarily of energy which has been converted and reconverted once and only once.[2,7]

(ii) The random coupling coefficient is assumed to be Gaussian and to have a power spectrum which is white, or uniform, over a bandwidth which is large compared to the bandwidth "occupied" by our signal.

V. ACKNOWLEDGMENT

REFERENCES

1. Miller, S. E., Waveguide as a Communication Medium, B.S.T.J., **33**, November, 1954, pp. 1209–1266.
2. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., **41**, May, 1962, pp. 1031–1170.
3. Holbrook, B. D., and Dixon, J. R., Load Rating Theory for Multichannel Amplifiers, B.S.T.J., **17**, October, 1939, pp. 624–645.
4. Albersheim, W. J., and Schafer, J. P., Echo Distortion in FM Transmission of Frequency-Division Multiplex, Proc. I.R.E., **40**, March, 1952, pp. 316–328.
5. Bennett, W. R., Curtis, H. E., and Rice, S. O., Interchannel Interference in FM and PM Systems Under Noise Loading Conditions, B.S.T.J., **34**, May, 1955, pp. 601–637.
6. Medhurst, R. G., and Small, G. F., Distortion in Frequency-Modulation Systems Due to Small Sinusoidal Variations of Transmission Characteristics, Proc. I.R.E., **44**, November, 1956, pp. 1608–1612.
7. Rowe, H. E., Approximate Solutions for the Coupled Line Equations, B.S.T.J., **41**, May, 1962, pp. 1011–1031.

8. Young, D. T., Effect of Differential Loss on Approximate Solutions to the Coupled Line Equations, B.S.T.J., this issue, p. 2787.
9. Miller, S. E., The Nature of and System Inferences of Delay Distortion Due to Mode Conversion in Multimode Transmission Systems, B.S.T.J., this issue, p. 2741.
10. Dugundji, J., Envelopes and Pre-Envelopes of Real Waveforms, Trans. I.R.E. PGIT, **IT-4,** March, 1958, pp. 53–57.
11. Zakai, M., Second-Order Properties of the Pre-Envelope and Envelope Processes, Trans. I.R.E. PGIT, **IT-6,** December, 1960, pp. 556–557.
12. Deutsch, R., *Nonlinear Transformations of Random Process*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1962.

Effect of Differential Loss on Approximate Solutions to the Coupled Line Equations

By D. T. YOUNG

(Manuscript received July 24, 1963)

The assumption of zero differential loss between coupling modes in a multimode transmission line reduces the complexity of theoretical analysis.

Here we show that in general the approximate solution including differential loss between modes may be computed by convolving the solution for the case of zero differential loss with the Fourier transform of $\exp(-|\Delta\alpha u|)$.

The rapidity of loss variation versus frequency is limited to $(2\Delta\alpha/\Delta\beta)f$ for transmission lines with high $\Delta\alpha$.

I. INTRODUCTION

Consider the coupled line equations

$$I_0'(z) = -\Gamma_0 I_0(z) + jc(z)I_1(z) \tag{1}$$

$$I_1'(z) = jc(z)I_0(z) - \Gamma_1 I_1(z). \tag{2}$$

These equations are useful in describing the effects of coupling between a signal mode, represented by a complex wave amplitude I_0, and a single spurious mode, represented by I_1, caused by geometric imperfections in a multimode transmission line. These equations may be derived in two ways from basic principles. The coupled line[1] or generalized telegrapher's equations[2,3] may be derived directly, or the geometric imperfections may be considered discrete; the case of continuous imperfections can then be considered as a limiting form of the discrete case.[4]

Exact solutions for these equations are known in only a few special cases, so considerable attention has been given to approximate solutions.[4,5] A second-order approximate solution is difficult to examine in general; however, Rowe and Warters[4,5] have given a very thorough investigation for the case of equal attenuation for the two modes or zero

differential loss ($\mathrm{Re}\,[\Gamma_0 - \Gamma_1] = 0$). Rowe[6] has shown in the case of a random coupling that the average loss for the TE_{01} in a circular guide mode may be calculated as the convolution of the Fourier transforms of three functions: an attenuation function, a triangular function, and the covariance function of the coupling. Here we show that in general the approximate solution for loss and phase of the TE_{01} mode may be calculated by convolving the solution for zero differential loss with the Fourier transform of the attenuation function.

II. PROOF

Approximate solutions to (1) and (2) are more conveniently described by normalizing the mode amplitudes in the following way

$$G_0(z) = I_0(z)e^{+\Gamma_0 z}$$
$$G_1(z) = I_1(z)e^{+\Gamma_1 z}.$$

The approximate solution[5] for $G_0(z)$ is

$$G_0(z) = 1 - \rho$$

where

$$\rho = \int_0^z e^{\Delta\Gamma u}\,du \int_0^{z-u} c(x)c(x + u)\,dx$$

$$\Delta\Gamma = \Delta\alpha + j\Delta\beta = \Gamma_0 - \Gamma_1$$

and the initial conditions are

$$I_0(0) = G_0(0) = 1$$
$$I_1(0) = G_1(0) = 0.$$

The normalized loss, $A = -\ln|G_0|$, may be further approximated by

$$A = \mathrm{Re}\,\rho$$

and the phase θ by

$$\theta = -\mathrm{Im}\,\rho.$$

2.1 Loss

Consider first the case of real coupling $c(z)$; then the normalized loss for a guide of length L is

$$A = \int_0^L e^{\Delta\alpha u}\cos\Delta\beta u\,du \int_0^{L-u} c(x)c(x + u)\,dx.$$

Let

$$\bar{c}(x) = p(x)c(x)$$

where

$$p(x) = 1 \qquad 0 \leqq x \leqq L$$
$$= 0 \qquad \text{otherwise.}$$

Then we write

$$R(u) = \int_0^{L-u} c(x)c(x+u)\, dx$$

as

$$R(u) = \int_{-\infty}^{\infty} \bar{c}(x)\bar{c}(x+u)\, dx.$$

We observe that $R(u)$ so defined is an even function of u and vanishes for $|u| > L$.

$$R(u) = R(-u)$$
$$R(u) = 0, \qquad |u| > L.$$

Assuming the signal mode is the lowest-loss mode of the transmission line, such as the TE_{01} mode in circular wave-guide, we will have $\Delta\alpha \leqq 0$ and may write the normalized loss as

$$A(\zeta) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-|\Delta\alpha u|} R(u)\, \cos 2\pi\zeta u\, du$$

or

$$A(\zeta) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-|\Delta\alpha u|} R(u)\, e^{-j2\pi\zeta u}\, du$$

where

$$2\pi\zeta = \Delta\beta.$$

Thus $A(\zeta)$ is the Fourier transform of the product of two functions and may be written as the convolution of their individual transforms. The transform of the first function is

$$B(\zeta) = \int_{-\infty}^{\infty} e^{-|\Delta\alpha u|}\, e^{-j2\pi\zeta u}\, du$$

$$= \frac{2}{|\Delta\alpha|} \frac{1}{\left[1 + \left(\dfrac{2\pi}{\Delta\alpha}\zeta\right)^2\right]}.$$

The transform of the second function is the solution for $\Delta\alpha = 0$,

$$A_0(\zeta) = \int_{-\infty}^{\infty} \frac{R(u)}{2} e^{-j2\pi\zeta u} \, du$$

$$= \tfrac{1}{2} \mid \bar{C}(\zeta) \mid^2$$

where

$$\bar{C}(\zeta) = \int_{-\infty}^{\infty} \bar{c}(u) \, e^{-j2\pi\zeta u} \, du.$$

Thus

$$A(\zeta) = B(\zeta) * A_0(\zeta)$$

where

$$B(\zeta) * A_0(\zeta) = \int_{-\infty}^{\infty} B(\zeta) A_0(\zeta - \xi) \, d\xi.$$

We observe that for $\Delta\alpha = 0$, $B(\zeta)$ becomes the unit impulse function which is the identity function for the convolution operator

$$A_0(\zeta) = B(\zeta) * A_0(\zeta), \qquad \Delta\alpha = 0.$$

2.2 *Phase*

The same analysis may be applied to the normalized phase as follows

$$\theta(\zeta) = -\int_0^L e^{\Delta\alpha u} \sin \Delta\beta u \, du \int_0^{L-u} c(x)c(x + u) \, dx$$

$$= -\int_0^{\infty} e^{\Delta\alpha u} R(u) \sin \Delta\beta u \, du$$

$$= -\frac{1}{2} \int_{-\infty}^{\infty} e^{-|\Delta\alpha u|} R(u) \, \operatorname{sgn} u \sin \Delta\beta u \, du$$

where

$$\operatorname{sgn} u = +1, \qquad u > 0$$

$$= -1, \qquad u < 0$$

$$= -\int_{-\infty}^{\infty} e^{-|\Delta\alpha u|} \frac{R(u)}{2} j \operatorname{sgn} u \, e^{-j2\pi\zeta u} \, du$$

then

$$\theta(\zeta) = -B(\zeta) * A_0(\zeta) * \frac{1}{\pi\zeta}$$

or

$$\theta(\zeta) = -A(\zeta) * \frac{1}{\pi\zeta}$$

which means the loss and phase functions are related by Hilbert transforms.[7]

The extension to complex coupling is straightforward and we obtain the following results:

Let

$$\tilde{c}(x) = [C_r + jC_i]c(x)$$

then

$$A(\zeta) = B(\zeta) * \left[(C_r^2 - C_i^2)A_0(\zeta) + (2C_rC_i)A_0(\zeta) * \frac{1}{\pi\zeta} \right]$$

$$\theta(\zeta) = -B(\zeta) * \left[(2C_rC_i)A_0(\zeta) + (C_r^2 - C_i^2)A_0(\zeta) * \frac{1}{\pi\zeta} \right]$$

and

$$\theta(\zeta) = -A(\zeta) * \frac{1}{\pi\zeta}$$

since

$$\frac{1}{\pi\zeta} * \frac{1}{\pi\zeta} = \delta(\zeta)$$

where $\delta(\zeta)$ is the unit impulse function.

III. DISCUSSION AND INTERPRETATION OF RESULTS

For simplicity we discuss only real coupling, but the results apply equally well to complex coupling. Consider the representation of $c(x)$ by a Fourier series over the length L with coefficients c_n. Rowe and Warters[4] have shown that for the case of zero differential loss, $A_0(\zeta)$ may be expressed as a double infinite summation

$$A_0(\zeta) = \frac{L^2}{2} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} c_m c_n *(-1)^{m-n} \frac{\sin \pi(\zeta L - m) \sin \pi(\zeta L - n)}{\pi(\zeta L - m)\pi(\zeta L - n)}$$

which is an expression for loss in terms of free-space wavelength, since at frequencies considerably above cutoff $2\pi\zeta = D\lambda_0(D$ constant).

$A_0(\zeta)$ is a band-limited function. Its sample points have relative frequency separation $\delta f/f = \delta\zeta/\zeta = 1/2\zeta L$, so the signal mode loss may vary more and more rapidly with frequency as the line length increases,

for $\Delta\alpha = 0$. Since $A_0(\zeta)$ is a band-limited function, $A(\zeta)$ is a band-limited function also. If we consider ζ as a time variable, then $A(\zeta)$ is the result of filtering $A_0(\zeta)$ with a low-pass filter whose impulse response is $B(\zeta)$. If $\Delta\alpha$ is very small $B(\zeta)$ has large amplitude at the origin and approaches zero rapidly away from the origin, and convolving $B(\zeta)$ with $A_0(\zeta)$ results in little change. The most noticeable change is a reduction in the peaks and an increase at the original minimums of the loss fluctuations. However, if $\Delta\alpha$ is increased until the spacing of the half-height points of $B(\zeta)$ is wider than the sample point spacing of $A_0(\zeta)$, the rapidity of the loss fluctuations with frequency will be controlled by $B(\zeta)$. In general for high $\Delta\alpha$ the frequency separation between half height points for $A(\zeta)$ is approximately the same as that of $B(\zeta)$ which is independent of length

$$\frac{\delta f}{f} = \left(\frac{2\Delta\alpha}{\Delta\beta}\right).$$

A numerical example pertaining to the waveguide problem is the following: Consider the previously mentioned loss function for a guide with a two-foot wiggle. The signal mode is TE_{01}. Let

$$L = 1000 \text{ ft.}$$

$$k = 500$$

$$\Delta\alpha = -0.184 \text{ neper/ft.}$$

Now the peak loss point is for $\zeta = 1/2$ which is near 50 kmc for TE_{12} in 2-inch diameter circular copper waveguide. The $\Delta\alpha$ value is typical of TE_{12} in lossy-jacketed helix waveguide. Now consider the half-height points for A_0. This bandwidth is approximately 0.084 kmc, which is very narrow compared to $B(\zeta)$, so that after convolution $A(\zeta) \approx KB(\zeta)$. The half-height points for $B(\zeta)$ are about 6.0 kmc. Thus the addition of the differential loss changes the TE_{01} loss from a very rapidly varying to a very slowly varying function of frequency. This effect is of great value for wideband transmission systems. It is also important in experimental measurements, since the number of measurements necessary for a guide with high loss to the spurious mode is greatly reduced.

Finally we recall that $\bar{c}(x) = p(x)c(x)$ so that

$$\bar{C}(\zeta) = \left(Le^{-j\pi\zeta L}\frac{\sin \pi L\zeta}{\pi L\zeta}\right) * C(\zeta)$$

which makes the solution for a Fourier series representation of $c(x)$ obvious, since $C(\zeta)$ would be a series of impulse functions, and the convolution operation is very easy.

$$A(\zeta) = \tfrac{1}{2}B(\zeta) * \left[\left(Le^{-j\pi\zeta L} \frac{\sin \pi L\zeta}{\pi L\zeta} * C(\zeta) \right) \left(Le^{+j2\pi\zeta L} \frac{\sin \pi L\zeta}{\pi L\zeta} * C^*(\zeta) \right) \right]$$

$$\Lambda(\zeta) = B(\zeta) * \left[\frac{L^2}{2} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} c_m c_n{}^*(-1)^{m-n} \right.$$
$$\left. \cdot \frac{\sin \pi(\zeta L - m) \sin \pi(\zeta L - n)}{\pi(\zeta L - m)\pi(\zeta L - n)} \right].$$

IV. ACKNOWLEDGMENT

REFERENCES

1. Miller, S. E., Coupled Wave Theory and Waveguide Applications, B.S.T.J., **33,** May, 1954, pp. 661–720.
2. Schelkunoff S. A., Conversion of Maxwell's Equations Into Generalized Telegraphist's Equations, B.S.T.J., **34,** September, 1955, pp. 995–1043.
3. Morgan, S. P., Theory of Curved Circular Waveguide Containing an Inhomogeneous Dielectric, B.S.T.J., **36,** September, 1957, pp. 1209–1251.
4. Rowe, H. E., and Warters, W. D., Transmission in Multimode Waveguide with Random Imperfections, B.S.T.J., **41,** May, 1962, pp. 1062–1067, 1087.
5. Rowe, H. E., Approximate Solutions to the Coupled Line Equations, B.S.T.J., **41,** May, 1962, pp. 1017–1018.
6. Unpublished work.
7. Guillemin, E. A., *Synthesis of Passive Networks,* John Wiley & Sons, Inc., N.Y., 1957.

Markov Processes Representing Traffic in Connecting Networks

By V. E. BENEŠ

A class of Markov stochastic processes x_t, suitable as models for random traffic in connecting networks with blocked calls cleared, is described and analyzed. These models take into account the structure of the connecting network, the set S of its permitted states, the random epochs at which new calls are attempted and calls in progress are ended, and the method used for routing calls.

The probability of blocking, or the fraction of blocked attempts, is defined in a rigorous way as the stochastic limit of a ratio of counter readings, and a formula for it is given in terms of the stationary probability vector p of x_t. This formula is

$$\frac{(p, \beta)}{(p, \alpha)}, \quad \text{or} \quad \frac{\sum_{x \epsilon S} p_x \beta_x}{\sum_{x \epsilon S} p_x \alpha_x},$$

where β_x is the number of blocked idle inlet-outlet pairs in state x, and α_x is the number of idle inlet-outlet pairs in state x. On the basis of this formula, it is shown that in some cases a simple algebraic relationship exists between the blocking probability b, the traffic parameter λ (the calling rate per idle inlet-outlet pair), the mean m of the load carried, and the variance σ^2 of the load carried. For a one-sided connecting network of T inlets ($=$ outlets), this relation is

$$1 - b = \frac{1}{\lambda} \frac{2m}{(T - 2m)^2 - (T - 2m) + 4\sigma^2};$$

for a two-sided network with N inlets on one side and M outlets on the other, it is

$$1 - b = \frac{m}{(N - m)(M - m) + \sigma^2}.$$

The problem of calculating the vector p of stationary state probabilities is fully resolved in principle by three explicit formulas for the components of p: a determinant formula, a sum of products along paths on S, and an expansion in a power series around any point $\lambda \geqq 0$. The formulas all indicate how these state probabilities depend on the structure of the connecting network, the traffic parameter λ, and the method of routing.

<div align="center">CONTENTS</div>

I. INTRODUCTION

A connecting system is a physical communication system consisting of (*i*) a set of terminals, (*ii*) a control unit which processes the information needed to set up calls, and (*iii*) a connecting network through which calls are switched between terminals.

Connecting systems have been described heuristically and at length in a previous paper.[1] Also, some of the algebraic and topological properties of connecting networks have been studied in another paper.[2] The models to be used here have been described (but not studied) in a third paper.[3] These papers are a source of background material for reading the present one; familiarity with them is desirable, but is not presupposed.

The principal problem treated here is the exact theoretical calculation of the grade of service (as measured by the probability of blocking) of a connecting network of given but arbitrary structure; the calculation is to be carried out in terms of a mathematical model for the operation of the network. The model used here is a Markov stochastic process x_t defined by some simple probabilistic and operational assumptions. The problem is first reduced to calculation of the stationary probability vector p of x_t from the "statistical equilibrium" equations. From the form of this reduction it follows that in many cases of practical interest

the probability of blocking is uniquely determined by the mean and variance of the carried load, a fact heretofore known only for very simple systems.

In the past, the application of A. K. Erlang's very natural method of statistical equilibrium has been visited by a curse of dimensionality, that is, by the extremely large number of equations comprised in the equilibrium condition. This difficulty has not only put explicit solutions apparently out of the question; it has even made it effectively impossible to reach a reliable qualitative idea of the dependence of the blocking probability on the structure of the network, the method of routing, etc.

Three explicit formulas for the solution p of the equilibrium equations will be given. One is based on purely algebraic considerations, and the others largely on combinatory and probabilistic notions. Because of the generality of the model with respect to network structure, these formulas are of necessity rather complex. Except in simple cases, they cannot be regarded as giving a final (or even a working) solution to the problem of calculating equilibrium probabilities. Still, they expose the mathematical character of the problem, and provide a badly needed starting point for well grounded approximations. For only after one has studied and understood this character can he seriously consider ignoring some of it in approximations.

II. PRELIMINARY REMARKS AND DEFINITIONS

Various combinatory, algebraic, and topological features of the connecting network play important roles in the analysis of stochastic models for network operation. Some of these features will now be described, and terminology and notations for them introduced.

Let S be the set of permitted (i.e., physically meaningful) states of the connecting network under study. It has been pointed out in earlier work[1,2] that these states are partially ordered by inclusion $\leq$, where

$$x \leq y$$

means that state x can be obtained from state y by removing zero or more calls. Also, these states can be arranged (in fact, partitioned) in an intuitive manner in a *state-diagram*, the Hasse figure for the partial ordering $\leq$. This figure is a graph constructed by partitioning the states in horizontal rows according to the number of calls in progress, the kth row consisting of all states with k calls in progress. The unique *zero*, or empty, state of the network, in which no calls are in progress, is placed at the bottom of the figure; above it comes the row consisting of states with exactly *one* call in progress, and so on. The figure is com-

pleted by drawing a graph with the states as nodes, and with adjacency matrix determined by the condition that states differing by exactly one call are adjacent. This means that in drawing the graph we place lines between states (in successive rows) that differ in point of one call. A *maximal* state is one that forms a summit of the state-diagram, i.e., has no states above it in the partial ordering.

For most systems the state-diagram has the following heuristic description: there is a unique "point" at the bottom corresponding to the zero state; there are usually many "points" at the top corresponding to the maximal states; and the diagram is very "fat" in the middle, because of the multitude of states with a moderate number of calls in progress.

We mention at this point that usually the number of states, i.e., the number of elements of S, is astronomically large. Indeed, this fact has been a principal obstacle to theoretical progress on problems of congestion in large connecting systems. For an illustration, in the network of No. 5 Crossbar type, illustrated in Fig. 1, made for 1000 lines out of square 10×10 switches, the number of *maximal states alone* is

$$(100 \times 10!)^4 = 1.734 \times 10^{34}.$$

The set of *inlets* of a connecting network is denoted by I, and the set
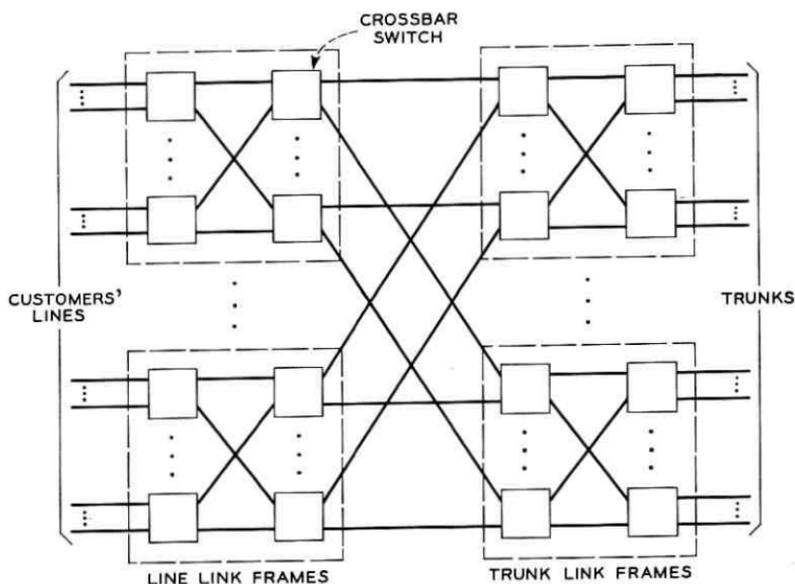
Fig. 1 — Structure of No. 5 Crossbar network.

of *outlets* by Ω. It is possible that $I \cap \Omega = \phi$, that $I \cap \Omega \neq \phi$, or even that $I = \Omega$, i.e., that all inlets are also outlets, depending on the "community of interest" aspects of the structure of the network. It is assumed that every call or connection is made only between an inlet and an outlet.

If x is a state, the notation $|x|$ (read "the norm of x") will denote the number of calls in progress in state x. If X is a set, then $|x|$ will denote the cardinality of X, i.e., the number of elements of X. We define the levels

$$L_k = \{x \in S: \quad |x| = k\}, \qquad k = 0, 1, \cdots, \max_{x \in S} |x|,$$

as the sets of states in which a specified number of calls is in progress. The $\{L_k\}$ form a partition of S,

$$\bigcup_k L_k = S$$

$$L_k \cap L_j = 0, \qquad k \neq j.$$

The "neighbors" of a state x are just those states which can be reached from x by adding or removing one call. These neighbors y of x can be divided into two sets according as $y > x$ or $y < x$; so we are led to define

A_x = set of neighbors *above* x
   = set of states accessible from x by *adding* one call
B_x = set of neighbors *below* x
   = set of states accessible from x by *removing* one call.

III. SUMMARY

The basic probabilistic assumptions that define the randomness in the traffic models to be studied are given precise statement in Section IV. They are, briefly, (*i*) the hang-up rate per call in progress is unity, and (*ii*) the calling rate per idle inlet-outlet pair is a constant $\lambda > 0$. Various operational aspects, such as the disposition of lost calls, and the method of routing, are specified and discussed in Section V. It is assumed that lost calls are refused without a change in state, and that routes for calls are chosen in a way that depends both on the call being set up or processed and on the current state of the system. In Section VI these probabilistic and operational assumptions are summarized in a transition rate matrix, Q. In Section VII, a Markov stochastic process x_t (the mathematical model for the operating system) is defined, and the statistical equilibrium condition $Qp = 0$ for the stationary probability vector p of x_t is formulated.

In Section VIII, the probability b of blocking is defined as the (probability one) limit of a ratio of counter readings, and a formula for b is given in terms of the stationary vector p. From this formula it is shown, in Section IX, that a simple algebraic relationship often exists between the blocking probability b, the traffic parameter λ, the mean load carried, and the variance of the load carried.

The remainder of the paper is devoted to the study and calculation of the vector p of stationary probabilities. Two explicit solutions, one algebraic and one combinatory in character, are given in Section X. In Section XI it is shown that the combinatory solution is a special case of a general formula for the stationary measure of an ergodic Markov process. The dependence of $p = p(\lambda)$ on the network structure and the method of routing is analyzed in an elementary way in Section XII. It is first shown that $p(\cdot)/p_0(\cdot)$ has components that are analytic in a neighborhood of the nonnegative real axis, and so are expressible in the form

$$\frac{p_x(\mu + \epsilon)}{p_0(\mu + \epsilon)} = \sum_{m=0}^{\infty} \epsilon^m c_m(x,\mu).$$

For $\mu = 0$ and $\epsilon = \lambda$ sufficiently small, this gives an expansion of p in powers of λ. It is then shown that with $|x|$ the number of calls in progress in state x, p_x is of order $\lambda^{|x|}$ as $\lambda \to 0$. This result renders possible a recursive calculation (Sections XII and XIV) of the coefficients $c_m(x,0)$ from the partial ordering $\leqq$ of S and a matrix used to specify the method of routing. Once p is developed as a power series in λ, a similar expansion is readily given (Section XIII) for the probability b of blocking.

In Section XV, finally, we completely solve the problem of calculating the coefficients $c_m(x,\lambda)$ for arbitrary values of $\lambda > 0$, giving each such coefficient both a combinatory interpretation, and an explicit formula, viz., a sum of products along paths through S which are trajectories for x_t permitted by the routing rule.

IV. PROBABILITY

To construct a Markov process for representing the random trajectory of the operating network through the set S of states, we shall make two simple probabilistic assumptions. The traffic models to be studied embody what has come to be known as a "finite-source effect," that is, a dependence of the instantaneous total calling-rate on the number of idle inlets, and on that of idle outlets.

In an attempt to describe this dependence in a simple rational way,

let us imagine a customer located at one of the inlets [outlets] of the connecting network, and seek to assign him a calling-rate, assuming that he is in an idle condition. We shall suppose that the traffic he offers is homogeneous in the sense that he calls every outlet [inlet] at the same rate, or with the same frequency. Indeed, we shall assume that all customers offer homogeneous traffic. Now on most occasions when he is making a call, a customer does not know whether the terminal he is calling is busy or idle. Thus, if he is on an inlet [outlet] it seems reasonable to suppose that there is a probability

$$\lambda h + o(h) \qquad \lambda > 0$$

that he attempts a call to a particular outlet [inlet] (distinct from his own) in the next interval of time of length h, as $h \to 0$, whether that outlet [inlet] is busy or not. The qualifying phrase "distinct from his own" is inserted to cover the case in which some inlets are also outlets, and in which it is reasonable to suppose that an idle terminal that is both an inlet and an outlet does not attempt to call itself.

We therefore make these two probabilistic assumptions:

(a) Holding-times of calls are mutually independent random variables, each with the negative exponential distribution of unit mean.

(b) If at time t the network is in a state x in which at least one member of the inlet-outlet pair $(u,v) \in I \times \Omega$ is idle (that is, one of u or v is not involved in a call in progress), the time elapsing from t until a call between u and v is attempted is a random variable having a negative exponential distribution with a mean $1/\lambda$, $\lambda > 0$. For different choices (u,v) and different occasions t, these times are all mutually independent and also independent of the call holding times.

These assumptions can be rendered in the informal terminology of "rates" as follows:

(i) The hang-up rate per call in progress is unity.

(ii) The calling-rate between an idle inlet u (outlet v) and an arbitrary outlet v (inlet u) with $u \neq v$ is $\lambda > 0$.

Assumptions (a) and (b) provide all the "randomness" needed to construct our models. The choice of a *unit* hang-up rate merely means that the mean holding-time is being used as the unit of time, so that only the one parameter λ need be specified.

V. OPERATION

To complete the description of the traffic models to be analyzed we must indicate how the network is operated. Since in the present work we are taking into account only the network configuration, and omitting consideration of the control unit, it suffices to describe how calls to busy

terminals are handled, how blocked calls are treated, and how routes or paths through the network are chosen.

It will be assumed that attempted calls to busy terminals are rejected, and have no effect on the state of the system; similarly, blocked attempts to call an idle terminal are refused, with no change in the state of the system. All successful attempts to place a call are completed instantly, with some choice of route.

To describe how routes are assigned to calls, we introduce a *routing matrix* $R = (r_{xy})$, with the following properties: for each x let Π_x be the partition of A_x induced by the equivalence relation of "having the same calls up," or satisfying the same "assignment" (of inlets to outlets); then for each $Y \epsilon \Pi_x$, r_{xy} for $y \epsilon Y$ is a probability distribution over Y; in all other cases $r_{xy} = 0$.

The interpretation of the routing matrix R is this: any $Y \epsilon \Pi_x$ represents all the ways in which a particular call c not blocked in x (between an inlet idle in x and an outlet idle in x) could be completed when the network is in state x; for $y \epsilon Y$, r_{xy} is the chance that if this call c is attempted, it will be routed through the network so as to take the system to state y. That is, we assume that if c is attempted in x, then a state y is drawn at random from Y with probability r_{xy}, independently each time c is attempted in x; the state y so chosen indicates the route c is assigned. The distribution or probability $\{r_{xy}, y \epsilon Y\}$ thus indicates how the calling-rate λ due to the call c is to be spread over the possible ways of putting up the call c. It is apparent that

$$\sum_{y \epsilon A_x} r_{xy} = \text{number of calls which can actually be put up in state } x$$

$$= s(x), \quad (\text{"successes" in } x),$$

the second equality defining $s(\cdot)$ on S. This account of the method of routing completes the description of the traffic models to be studied.

VI. TRANSITION RATES

For the purpose of defining a Markov stochastic process it is convenient and customary to collect the probabilistic and operational assumptions introduced above in a matrix $Q = (q_{xy})$ of *transition rates*, here given by

$$q_{xy} = \begin{cases} 1 & y \epsilon B_x \\ \lambda r_{xy} & y \epsilon A_x \\ -|x| - \lambda s(x) & y = x \\ 0 & \text{otherwise.} \end{cases}$$

The number q_{xy}, for $x \neq y$, has the usual interpretation that if the system is in state x, there is a chance

$$q_{xy}h + o(h)$$

that it will move to y in the next interval of time of length h, as $h \to 0$. Similarly

$$1 - q_{xx}h + o(h)$$

is the probability that the system will stay in x throughout the next interval of time of length h, as $h \to 0$

VII. MARKOV PROCESSES

In terms of the transition rate matrix Q it is possible to define a stationary Markov stochastic process $\{x_t, \ -\infty < t < +\infty\}$ taking values on the set S of states. The matrix $P(t)$ of transition probabilities

$$p_{xy}(t) = \Pr\{x_t = y \mid x_0 = x\}$$

of x_t satisfies the equations of Kolmogorov

$$\frac{d}{dt} P(t) = QP(t) = P(t)Q,$$

$$P(0) = I,$$

and is given formally by the formula

$$P(t) = \exp tQ.$$

Theorem 1: There exists a decomposition of the set S of states into a transient *set F and a single* ergodic *set $S - F$ containing the zero state; members of F have the property*

$$\lim_{t \to \infty} p_{xy}(t) = 0 \qquad y \ \epsilon \ F, \ x \ \epsilon \ S;$$

on $S - F$ there is a unique stationary (or equilibrium) *distribution* $\{p_x, \ x \ \epsilon \ S - F\}$ *such that*

$$\lim_{t \to \infty} p_{xy}(t) = p_y > 0 \qquad y \ \epsilon \ S - F, \ x \ \epsilon \ S$$

$$\sum_{x \epsilon S-F} p_x p_{xy}(t) = p_y \qquad y \ \epsilon \ S - F, \quad \text{all } t$$

$$\sum_{x \epsilon S-F} q_{xy} p_x = 0 \qquad y \ \epsilon \ S - F.$$

Proof: The existence of the unique ergodic set $S - F$ follows from the fact that the zero state is accessible from every other state by hang-

ups. The existence and character of the limit of $p_{xy}(t)$ as $t \to \infty$ is a consequence of exercise 19, p. 436 of Feller,[4] i.e., of the fact that the characteristic values r of Q satisfy $r = 0$ or $\mathrm{Re}(r) < 0$. (See also Bellman,[5] p. 294.)

To prove the uniqueness of p, suppose that q is a different probability vector on $S - F$ that also satisfies the "equilibrium" condition

$$\sum_{x \epsilon S-F} q_{xy} q_x = 0 \qquad y \epsilon S - F.$$

Then by Kolmogorov's equation

$$\frac{d}{dt} \sum_{x \epsilon S-F} g_x p_{xy}(t) = \sum_{x,z \epsilon S-F} q_x q_{xz} p_{zy}(t) = 0.$$

Integrating from 0 to t, and using $P(0) = I$, we find

$$\sum_{x \epsilon S-F} q_x p_{xy}(t) = q_y \qquad y \epsilon S - F.$$

Since $S - F$ is the only ergodic set, the left-hand side approaches p_y as $t \to \infty$. Hence $p = q$.

It is convenient to extend the dimension of p to $\mid S \mid$ by adding zero components for states in F, so that $p_{xy}(t) \to p_y \geqq 0$ for all $x,y \epsilon S$. The consideration of the transient set F is not just a mathematical fillip, since a "good" routing rule R may explicitly make certain "bad" states *unreachable* from the zero state, and thus place them in F to good purpose.

In the notation of Halmos,[6] p. 65, the stationary probability vector satisfies the equilibrium condition

$$Qp = 0.$$

This is the classical equation of state, or equation of statistical equilibrium, familiar in traffic theory. For our process x_t it takes on the rather simple form

$$[\mid x \mid + \lambda s(x)] p_x = \sum_{y \epsilon A_x} p_y + \lambda \sum_{y \epsilon B_x} p_y r_{yx}, \qquad x \epsilon S.$$

The left-hand side represents the average rate of exits from x, while the right-hand side is the average rate of entrances into x, in equilibrium. We define

$$p_k = \sum_{\mid x \mid = k} p_x = \Pr \{L_k\}.$$

Lemma 1: For $1 \leqq k \leqq w = \max_{x \epsilon S} \mid x \mid$

$$k p_k = \lambda \sum_{x \epsilon L_{k-1}} p_x s(x).$$

Proof: From the statistical equilibrium equation for $x = 0$ we obtain

$$\sum_{y \epsilon A_0} p_y = \lambda s(0) p_0 ,$$

which is Lemma 1 for $k = 1$. Assume that the lemma holds for a given $k \geqq 1$. Summing the statistical equilibrium equations over $x \epsilon L_k$ we find

$$k p_k + \lambda \sum_{x \epsilon L_k} s(x) p_x = \sum_{x \epsilon L_k} \sum_{y \epsilon A_x} p_y + \lambda \sum_{x \epsilon L_k} \sum_{y \epsilon B_x} p_y r_{yx} .$$

The second sum on the right is the same as

$$\lambda \sum_{y \epsilon L_{k-1}} p_y \sum_{x \epsilon A_y} r_{yx} ,$$

and by definition,

$$\sum_{x \epsilon A_y} r_{yx} = s(y).$$

Hence (induction hypothesis) the second sum equals $k p_k$. It is easy to see that in the first sum on the right each p_y is counted exactly $| y |$ times, i.e., $(k + 1)$ times, since for a given $y \epsilon L_{k+1}$ there are exactly $(k + 1)$ elements $x \epsilon L_k$ for which $y \epsilon A_x$. Thus the first sum is

$$(k + 1) \sum_{y \epsilon L_{k+1}} p_y = (k + 1) p_{k+1}$$

and the Lemma follows by induction. This result could also be obtained from the general observation that the statistical equilibrium equations are equivalent to the principle that for any set X of states the average rate of exits from X equals the average rate of entrances into X. (See Morris and Wolman.[7])

VIII. PROBABILITY OF BLOCKING

The fraction of calls that are refused because they are blocked, or the probability of blocking, is a quantity of particular interest to traffic engineers; they use it to assess the grade of service provided by an operating connecting network. The rigorous theoretical calculation of blocking probabilities has long been an outstanding problem of traffic theory. This problem is outstanding in both senses of the word: it is *conspicuous*, and it is *unsolved*. In fact, not even the definition (let alone the calculation) of the probability of blocking has received adequate treatment; for example, the otherwise monumental treatise of R. Syski[8] does not give a general account of blocking probability.

Since it is desirable to have a close connection between theoretical quantities and their physical meanings in terms of measurements, we

shall approach the study of blocking probabilities by asking how these probabilities might be measured "in the field." The most natural method of measuring the fraction of blocked attempts seems to be this: to the control unit of the connecting system under consideration we attach two counters; the first will count up one unit every time an attempted call is blocked, and the second will register one unit every time a call is attempted; the ratio of the reading of the first counter to that of the second should, after a long time during which the system's parameters remain constant, be an approximate measure of the fraction of blocked attempts. For mathematical convenience, one can then *define* the probability of blocking to be the limit (as time increases without end) of this ratio of the counter readings. This mathematical definition was first proposed by S. P. Lloyd, although, of course, the ratio has been the practical definition for 50 years, being the "peg count and overflow ratio."

A precise mathematical version of this measurement procedure can be given as follows: on the same sample space as that of the process x_t that describes the operating network, we define two additional stochastic processes $\{b(t),\ t \geqq 0\}$ and $\{a(t),\ t \geqq 0\}$ by the (respective) conditions

$$b(t) = \text{number of blocked attempted calls in } (0,t],$$

$$a(t) = \text{number of attempted calls in } (0,t].$$

These stochastic processes are the mathematical analogs of the counter readings. It is reasonable to use the limit

$$\lim_{t \to \infty} \frac{b(t)}{a(t)}$$

of the ratio of $b(\cdot)$ to $a(\cdot)$ as a mathematical definition of the probability of blocking, provided that the limit exists in a suitable sense. We show that this limit exists and is constant with probability one, and we give a formula for it.

Theorem 2: The probability of blocking b, defined by

$$b = \lim_{t \to \infty} \frac{b(t)}{a(t)},$$

exists and is constant with probability one; its almost sure value is

$$b = \frac{(p, \beta)}{(p, \alpha)} = \frac{\sum_{x \in S} p_x \beta_x}{\sum_{x \in S} p_x \alpha_x},$$

where p is the vector of stationary probabilities, and

β_x = *number of idle inlet-outlet pairs that are blocked in state x,*

α_x = *number of idle inlet-outlet pairs in state x.*

Proof: It can be seen that $a(t)$ and $b(t)$ can be written as sums over S,

$$a(t) = \sum_{x \epsilon S} a_x(t)$$

$$b(t) = \sum_{x \epsilon S} b_x(t)$$

where

$a_x(t)$ = number of attempted calls made in $(0,t]$ with the system in state x,

$b_x(t)$ = number of blocked attempts made in $(0,t]$ with the system in state x.

Now a blocked attempt occuring at an epoch u such that $x_u = x$ does not change the state of the system. Such an epoch u is a regeneration point of the process x_t. A successful attempt occurring at an epoch u at which $x_u = x$ *does* change the state of the system. The time interval from u back to the last previous epoch v at which a successful attempt occurred in state x, however, is independent of the behavior of x_t for $t > u$; it depends only on the fact that the system left x by adding a new call, not on *what* new call it was, nor on where into A_x x_t went as this new call was completed. This can be seen as follows: we have

$$u - v = \tau - v + u - \tau$$

where τ is the epoch at which x was last entered prior to u. Now $\tau - v$ is independent of x_t for $t > \tau$ if $x_{\tau+0}$ is known to be x, because x_t is a Markov process.

Let U be an event measurable on $\{x_t, t > u\}$. Then

Pr $\{U$ and $u - \tau \leqq \mu \mid x_{\tau+0} = x\}$

$$= \text{Pr}\{u - \tau \leqq \mu \mid x_{\tau+0} = x\} \sum_{y \epsilon A_x} \frac{r_{xy}}{s(x)} \text{Pr } \{U \mid x_{u+0} = y\},$$

where

$$\text{Pr}\{u - \tau \leqq \mu \mid x_{\tau+0} = x\} = \sum_{n=1}^{\infty} \left(\frac{\mid x \mid}{\mid x \mid + \lambda s(x)} \right)^{n-1} \frac{\lambda s(x)}{\mid x \mid + \lambda s(x)}$$
$$\int_0^\mu \frac{[\mid x \mid + \lambda s(x)]^n \, t^{n-1} \, e^{-t[\mid x \mid + \lambda s(x)]}}{(n - 1)!} \, dt.$$

Thus the time intervals β_1, β_2, $\cdots$ elapsing between successive blocked attempts in state x, and those α_1, α_2, $\cdots$ elapsing between successive attempts in state x, both form sequences of mutually independent, and except possibly for the first elements β_1 and α_1, identically distributed random variables. That is, the elements of each sequence are mutually independent, but the sequences are not independent since one consists of partial sums over blocks of the other.

Both these sequences can be studied, then, in terms of a sequence x_1, x_2, $\cdots$ of mutually independent random variables, all (except possibly x_1) identically distributed. We define for $t \geqq 0$ and $k \geqq 0$

$$S_0 = 0$$

$$S_n = \sum_{i=1}^{n} x_i, \qquad x_i = \alpha_i \quad \text{or} \quad \beta_i$$

$$n(t) = k \quad \text{if and only if} \quad S_k \leqq t < S_{k+1},$$

$$n(t) = a_x(t) \quad \text{or} \quad b_x(t).$$

It is now straightforward to show that $t^{-1}n(t)$ approaches a limit with probability one, and to find the limit. Let us put, for $t > S_1$

$$\frac{n(t)}{t} = \frac{n(t)}{S_{n(t)}} \cdot \frac{S_{n(t)}}{t}.$$

The first factor converges to $E^{-1}\{x_2\}$ with probability one, by the law of large numbers. The local suprema of

$$\frac{t - S_{n(t)}}{t}$$

for $t > S_1$ occur at the points

$$t = S_k \qquad k = 1, 2, \cdots,$$

and have the values

$$\frac{x_k}{S_k} = \frac{k}{S_k} \cdot \frac{x_k}{k} \qquad k = 1, 2, \cdots.$$

Again, the first factor converges to $E^{-1}\{x_2\}$ with probability one by the law of large numbers. Since $E\{x_2\} < \infty$, and $\{x_k, k \geqq 2\}$ are identically distributed,

$$\sum_{k=2}^{\infty} \Pr\{x_k > \epsilon k\} = \sum_{k=2}^{\infty} \Pr\{x_2 > \epsilon k\} < \infty$$

and it follows from the Borel-Cantelli lemma that for any $\epsilon > 0$

$$\Pr\{x_k > \epsilon k \text{ for infinitely many values of } k \geq 2\} = 0.$$

Hence $x_k = o(k)$ as $k \to \infty$, with probability one, and with the same probability,

$$\frac{S_{n(t)}}{t} \to 1 \quad \text{as} \quad t \to \infty.$$

It follows that with probability one,

$$\lim_{t \to \infty} t^{-1} a_x(t) = E^{-1}\{\text{time interval between successive attempted calls}$$
$$\text{in } x\},$$

$$\lim_{t \to \infty} t^{-1} b_x(t) = E^{-1}\{\text{time interval between successive blocked attempted calls in } x\}.$$

Furthermore (cf. Ref. 11, p. 247, equation (1.2) and p. 249)

$$\frac{d}{dt} E\{b_x(t) \mid x_0 = y\} = \lambda p_{yx}(t)\beta_x \to \lambda p_x \beta_x \quad \text{as} \quad t \to \infty.$$

However, by Feller's renewal theorem (cf. Ref. 11, p. 246), we know that

$$\lim_{t \to \infty} t^{-1} E\{b_x(t) \mid x_0 = y\} = E^{-1}\{\text{time interval between successive blocked attempted calls in } x\}.$$

Hence, with probability one,

$$t^{-1} b_x(t) \to \lambda p_x \beta_x \quad \text{as} \quad t \to \infty.$$

A similar argument shows that with probability one

$$t^{-1} a_x(t) \to \lambda p_x \alpha_x \quad \text{as} \quad t \to \infty$$

and completes the proof of Theorem 2.

IX. A BASIC FORMULA

Engineers have recognized (at least) four quantities as significant for the study and design of connecting networks carrying random traffic. These are the calling rate, the average load carried, the variance of the load carried, and the probability of blocking. In our model these quantities are given respectively by

λ = calling rate per idle inlet-outlet pair

$m = \sum\limits_{x \epsilon S} |x| \, p_x$ = average number of calls in progress

$\sigma^2 = \sum\limits_{x \epsilon S} (|x| - m)^2 p_x$

$b = (p,\beta)/(p,\alpha)$.

It is natural to ask whether there exist any systematic relationships between these quantities, or between these and (possibly) other simple parameters of the network under study. Such relationships would be particularly useful and significant if they were largely independent of the structure or configuration of the connecting network, and were valid either for all networks or for large classes of them. We shall show that there often exists a simple *algebraic* relation among λ, m, σ^2, and b. Its exact form depends on which inlets are also outlets. First we prove

Theorem 3: The probability b of blocking can be written as

$$b = 1 - \frac{m}{\lambda \sum\limits_{x \epsilon S} p_x \alpha_x} \qquad (1)$$

or, in words, as

$$b = 1 - \frac{average \ load \ carried}{(calling \ rate \ per \ idle) \times (average \ number \ of \ idle \ pairs)}.$$

Proof: In equilibrium, the average rate of successful attempts must equal the average rate of hangups. Hence, intuitively,

$$\lambda \sum\limits_{x \epsilon S} p_x s(x) = \sum\limits_{x \epsilon S} |x| p_x = m. \qquad (2)$$

Since $\beta_x = \alpha_x - s(x)$, the result follows from Theorem 1. The actual validity of the identity (2) can be inferred from Lemma 1, by summation on k.

Formula (1), rewritten in the form

$$1 - b = \frac{average \ load \ carried}{average \ rate \ of \ attempts},$$

should be viewed as a direct generalization of Erlang's classical loss formula for c trunks, blocked calls cleared, and calls arising in a Poisson process of intensity $a > 0$. In that case the probability of loss is

$$E_1(c,a) = \frac{\dfrac{a^c}{c!}}{\sum\limits_{j=0}^{c} \dfrac{a^j}{j!}},$$

and it can be seen that

$$1 - E_1(c,a) = \frac{\displaystyle\sum_{j=0}^{c-1} \frac{a^j}{j!}}{\displaystyle\sum_{j=0}^{c} \frac{a^j}{j!}}$$

$$= \frac{\displaystyle\sum_{j=1}^{c} j \cdot \frac{a^j}{j!}}{a \displaystyle\sum_{j=0}^{c} \frac{a^j}{j!}}$$

$$= \frac{\text{average number of busy trunks}}{\text{total calling rate}}.$$

To exhibit useful special cases of the general formula (1) of Theorem 3, we introduce a partial classification of connecting networks. A network is called *one-sided* if $I = \Omega$, i.e., if all inlets are also outlets; a network is *two-sided* if $I \cap \Omega = \phi$, i.e., if no inlet is an outlet.

Corollary 1: For a one-sided network of T terminals

$$b = 1 - \frac{1}{\lambda} \frac{2m}{(T - 2m)^2 - (T - 2m) + 4\sigma^2}.$$

Proof: For the one-sided network in question, we have $I = \Omega$, $|I| = |\Omega| = T$, and so

$$\alpha_x = \binom{T - 2|x|}{2}$$

$$\sum_{x \epsilon S} p_x \alpha_x = \tfrac{1}{2}\{T^2 - (2T - 1)2m - T + 4m^2 + 4\sigma^2\}.$$

Corollary 2: For a two-sided network with M terminals on one side and N on the other

$$b = 1 - \frac{1}{\lambda} \frac{m}{(M - m)(N - m) + \sigma^2}.$$

Proof: It is clear that in this case

$$\alpha_x = (M - |x|)(N - |x|)$$

so that

$$\sum_{x \epsilon S} p_x \alpha_x = (M - m)(N - m) + \sigma^2.$$

Each of the foregoing corollaries exhibits an explicit algebraic rela-

tionship between λ, m, σ, and b, based only on the one- or two-sidedness of the network.

The preceding corollaries can be used to show that in a large system, the numerical value of the constant λ will be small—indeed, of the order of the reciprocal of the number of inlets and outlets. This can be seen by the following heuristic argument, carried out for a one-sided network with T terminals: suppose that each terminal carries q ($0 \leq q < 1$) erlangs and that the blocking probability b is so small that we can ignore it and set

$$b = 1 - \frac{1}{\lambda} \frac{2m}{(T - 2m)^2 - (T - 2m) + 4\sigma^2} = 0.$$

Since the network is one-sided, any load carried by one terminal is also carried by some other terminal, and so

$$qT = 2m$$

whence

$$\lambda = \frac{1}{T} \frac{q}{(1 - q)^2 - \frac{(1 - q)}{T} + \frac{4\sigma^2}{T^2}}.$$

Because

$$\sigma^2 = \sum_{x \in S} p_x (|x| - m)^2$$

and

$$(|x| - m)^2 \leq \frac{T^2}{4},$$

we have $0 \leq 4\sigma^2/T^2 \leq 1$, and so

$$\lambda \approx \frac{\text{const.}}{T}$$

with

$$\frac{q}{1 + (1 - q)^2} \leq \text{const.} \leq \frac{q}{(1 - q)^2 - T^{-1}(1 - q)}.$$

X. SOLUTION OF THE EQUATIONS OF STATISTICAL EQUILIBRIUM

So far, we have shown that the theoretical determination of the blocking probability b reduces to that of the stationary vector p or, in many cases, to that of the mean m and variance σ^2 of the carried

load. In either case, some knowledge of p is required. Most of the rest of this paper, therefore, is devoted to the calculation of p and to the study of its properties.

In the past, the application of A. K. Erlang's very natural method of "statistical equilibrium" to congestion in connecting networks has been visited by the curse of dimensionality, that is, by the extremely large number $| S |$ of equations comprised in the stationarity condition $Qp = 0$. This difficulty has not only put explicit solutions apparently out of the question; it has even made it effectively impossible to reach a reliable qualitative idea of the *dependence* of the state probabilities $\{ p_x , x \in S \}$ on the structure of the network and on the method of routing.

To be sure, it has always been possible in principle to solve $Qp = 0$ by successive elimination of unknowns; however, when the dimension of p is of order 10^{40} or so, this remark is hardly helpful. Since successive elimination can be used to solve $Qp = 0$ for any "ergodic" transition rate matrix Q, it neither elucidates nor uses any of the special features of the matrices Q that arise in problems of congestion in networks. Thus, even were it algebraically feasible, the method of successive elimination treats our matrices Q as indistinguishable from other matrices possessing a zero characteristic value.

We shall give several explicit solutions of the equilibrium equations. One is based on purely algebraic considerations, and the others largely on combinatory and probabilistic notions. Because of the generality of our model with respect to network structure, the formulas appearing in the solutions are necessarily rather complex. Except in simple cases, they cannot be regarded as giving a final (or even a working) solution to the problem of calculating equilibrium probabilities. Nevertheless, they expose the mathematical structure of the problem and provide a badly needed starting point for well grounded approximations. For only after one has studied and understood this structure can he seriously think about throwing some of it away in approximations.

To describe the solutions in full detail, we need various preliminary definitions and conventions.

It will be shown in a later paper[9] that the minimum value of the blocking probability b is achieved by a routing matrix R consisting entirely of zeros and ones, i.e., by a deterministic rule. So it is assumed henceforth that R has only zeros or ones for entries.

A *path* on S of length $l \geqq 0$ is any ordered sequence $x_0 , x_1 , \cdots , x_l$ of $(l + 1)$ elements of S. A lower case pi, π, will be used as a symbol for a generic path on S, and we write

$$\pi = \{ x_0 , x_1 , \cdots , x_l \}, \qquad l = l(\pi)$$

to indicate that π is a path of length $l(\pi)$ consisting of x_0, x_1, $\cdots$, x_l in that order. Note that paths of length zero are countenanced. A path π is a *loop* if

$$x_0 = x_l,$$

and also

$$x_i \neq x_j$$

whenever $0 < i < j \leq l(\pi)$. A loop of length zero is a path of length zero. If $\pi = \{x_0, x_1, \cdots, x_0\}$ is a loop, each element x_0, x_1, $\cdots$ etc. will be spoken of as being *on* π.

The elements x and y of S are called *adjacent* in the graph of $(S, \leq)$, i.e., in the state diagram, if one of the following equivalent conditions holds:

  (*i*) x covers y or y covers x
  (*ii*) $y \in A_x$ or $x \in A_y$
  (*iii*) x and y differ by exactly one call in progress.

A path on S is called *continuous* if successive elements of the path are adjacent.

In order that x_t have positive probability of following a path π, it is not enough that π be continuous. For evidently the action of the routing matrix R (assumed to consist solely of zeros and ones) is to prohibit certain paths on S as (parts of) possible realizations of the process x_t. Here "possible" of course means "having positive probability." There exists then a class of those paths that are permitted by R, definable in several ways. One such way is as follows: A path $\pi = \{x_0, x_1, \cdots x_l\}$ on S is *permitted by* R if for each i in the range $1 \leq i \leq l$,

$$x_i \in B_{x_{i-1}} \qquad \text{or} \qquad r_{x_{i-1}x_i} = 1.$$

The set of paths permitted by R is denoted by P.

With X a subset of S,

$$\text{perm}(X)$$

will denote the set of all *permutations* of X, i.e., one-to-one maps of X onto itself. We let

$$y_1, y_2, \cdots y_{|S|}$$

be an arbitrary simple ordering of S, and we define the *ordinal number* $\omega(x)$ of a state $x \in S$ by the condition

$$\omega(x) = n \quad \text{if and only if} \quad x = y_n, \qquad n = 1, 2, \cdots, |S|.$$

For each m,n in the region $1 \leqq m,n \leqq |S|$, we define a function $\zeta_{mn}(\cdot)$ on the domain $1 \leqq i \leqq |S|$ by the condition

$$\zeta_{mn}(i) = \begin{cases} i & 1 \leqq i < \min(m,n) \\ 0 & i = m \text{ or } i = n \\ i - 1 & \min(m,n) < i < \max(m,n) \\ i - 2 & \max(m,n) < i \leqq |S|. \end{cases}$$

We observe that $\zeta_{mn}(\cdot)$ has an inverse for each m and n. Now let $\varphi(\cdot)$ be a permutation of the set of states with the mth and the nth removed; then

$$i \rightarrow \zeta_{mn}\{\omega[\varphi(y_{\zeta_{mn}}^{-1}(i))]\} \qquad i = 1, 2, \cdots, |S| - 2$$

defines the permutation $a_{mn}(\varphi)$ *associated with* φ. Also, sgn $a_{mn}(\varphi)$ is $+1$ or -1 according as the permutation $a_{mn}(\varphi)$ is even or odd.

The "hang-up" matrix $H = (h_{xy})$ is defined by the condition

$$h_{xy} = \begin{cases} 1 & \text{if } y \in B_x \\ 0 & \text{otherwise.} \end{cases}$$

Let x and z be states, and suppose that $\pi = \{x_0, \cdots, x_l\}$ is a path in P beginning at z and ending at x, so that $x_0 = z$ and $x_l = x$. Suppose also that the trajectory represented by π contains m new calls, i.e., there are exactly m values of i in the range $1 \leqq i \leqq l$ such that

$$x_i \in A_{x_{i-1}}.$$

Since π starts at z, in which $|z|$ calls are in progress, and ends up at x, in which there are $|x|$ calls in progress, it is evident that

$$l(\pi) = 2m + |z| - |x|.$$

The set of paths which start at a state z, never return to z, and end up at $x \neq z$, is denoted by

$$K_{zx}.$$

Thus π belongs to K_{zx} if and only if $x_0 = z$, $x_i \neq z$ for $0 < i \leqq l(\pi)$, and $x_l = x$.

Let $\pi = \{x_0, x_1, \cdots, x_l\}$ be a path on S, and let $f(\pi, \cdot)$ be a function defined for $x_0, x_1, \cdots, x_l$. In terms of $f(\pi, \cdot)$ and π, we define a product along the path π by the expression

$$\prod_{\substack{i=1 \\ \pi=\{x_0, x_1, \cdots, x_l\}}}^{l(\pi)} f(\pi, x_i).$$

It is convenient to abbreviate this product by the readily understandable expression

$$\prod_{x \epsilon \pi} f(\pi, x).$$

In the special case that $f(\pi, \cdot)$ has the form

$$f(\pi, x_i) = h(x_{i-1}, x_i), \qquad i = 1, \cdots, l$$

we abbreviate the product by

$$\prod_{x \epsilon \pi} h(px, x).$$

The notation $p \cdot$ is supposed to represent the *predecessor* of an element in π.

The first and simplest solution of the equilibrium equation $Qp = 0$ to be given is based on an observation made by I. W. Sandberg, namely, that $\det(Q)$, and hence $\det(Q')$, are zero, so that $Q' \mathrm{adj}(Q') = 0$, and thus columns of the matrix of cofactors of Q should give solutions of $Qp = 0$. The author has not succeeded in elucidating the probabilistic significance of these simple algebraic facts. It will be seen later that the other solutions to be given are, on the other hand, natural, plausible, or even obvious from a probabilistic viewpoint, but are algebraically involved.

Theorem 4: Let m be an integer in the range $1 \leqq m \leqq |S|$. An unnormalized nonnegative solution p of $Qp = 0$ is given by

$$p_{y_n} = (-1)^{|S|-1+m+n} \sum_{\varphi \epsilon \text{ perm } (S-[y_m, y_n])} \mathrm{sgn}\ a_{mn}(\varphi)$$

$$\prod_{\varphi(z)=z} (-|z| - \lambda s(z)) \prod_{\varphi(z) \neq z} (h_{z\varphi(z)} + \lambda r_{z\varphi(z)}).$$

Proof: Since $\det(Q) = 0$, it follows that $\det(Q') = 0$, the prime indicating the transposed matrix. Hence (Birkhoff and MacLane,[10] p. 290) no matter what ordering of S is used.

$$Q' \mathrm{adj}(Q') = 0,$$

where 'adj' denotes the adjoint matrix, i.e., the transposed matrix of cofactors. Let $C = (c_{xy})$ be the matrix of cofactors of Q corresponding to the ordering $y_1, y_2, \cdots, y_{|S|}$ of S, and suppose that the entries of Q are also arranged according to this ordering. Then

$$C = \mathrm{adj}(Q')$$

and we find that

$$\sum_{z \epsilon S} q_{zx} c_{zy} = 0.$$

Thus any column of the matrix C of cofactors of Q gives a solution of the equilibrium equations. It follows from a result of W. Ledermann (Bellman,[5] p. 294, exercise 10) that

$$(-1)^{|S|-1}c_{xy} \geqq 0.$$

We see that all the cofactors c_{xy} have the same sign, and each column of the matrix C yields a nonnegative solution of $Qp = 0$. Hence all columns are proportional, because there is only one nonnegative solution, up to normalization. The theorem follows from the standard formula for a cofactor as a determinant.

   Theorem 5: If $z \in S$ is any state, then an unnormalized solution of the statistical equilibrium equations $Qp = 0$ is given by

$$p_z = 1$$

and for $x \neq z$,

$$p_x = \sum_{\pi \in P \cap K_{zx}} \lambda^{\frac{l(\pi)+|x|-|z|}{2}} \prod_{y \in \pi} \frac{1}{|y| + \lambda s(y)}.$$

Proof: The formula given can be verified by direct substitution in the equations

$$[\lambda s(x) + |x|]p_x = \sum_{y \in A_x} p_y + \lambda \sum_{y \in B_x} p_y r_{yx}, \qquad x \in S.$$

Convergence of the infinite sum will follow from our Theorem 7 and exercise 19, p. 378 of Feller.[4]

XI. STATIONARY PROBABILITY MEASURES FOR ERGODIC MARKOV PROCESSES

   In order to shed light on Theorem 5 (and also to prove it by a probabilistic argument) we shall consider in this section the general problem of calculating the stationary probability measure of an ergodic continuous parameter Markov process on a finite number of states. Our object is to give an explicit formula for the measure in terms of the transition rate matrix. Again, it is needless to mention that a formula of such generality must be fairly complex. Applied to familiar Markov processes whose stationary measures are well known, the formula to be given yields some unexpected combinatory identities, not pursued here.

   We shall now use the notations x_t, S, Q, and $P(\cdot)$ to describe an *arbitrary* Markov stochastic process x_t in continuous time, taking values in a finite set S of states with transition rate matrix $Q = (q_{xy})$ and transition probability matrices $P(t) = (p_{xy}(t))$, t real. It is assumed that there is a single ergodic class of states. Such a general interpretation of notations already introduced (for specific processes describing traffic in

connecting networks) is made to avoid defining new terminology; it is made in this section only, and should cause no confusion.

If z is a state, a *return to z* is defined to be an epoch of time at which x_t reaches z, i.e., u is a return to z if for some $\epsilon > 0$, $x_t \neq z$ for $u - \epsilon < t < u$ and $x_t = z$ for $u < t < u + \epsilon$. A *departure from z* is an epoch of time at which x_t leaves z, i.e., u is a departure from z if for some $\epsilon > 0$, $x_t = z$ for $u - \epsilon < t < u$ and $x_t \neq z$ for $u < t < u + \epsilon$. A *return time* to z is a period of time elapsing between a departure from z and the next return to z. We set, for $t \geq 0$,

$$H_z(t) = E\{\text{number of returns to } z \text{ in } (0,t] \mid x_0 = z\}$$

$$\mu_z = E\{\text{return time to } z\}$$

$$q_z = -q_{zz} = E^{-1}\{\text{length of a stay in } z\}$$

The notation $H_z(\cdot)$ has been chosen because the defined quantity has an obvious resemblance to the classical renewal function. (See Smith.[11])

There is a simple relationship between the equilibrium probability of a state x, and the quantities μ_x and q_x; this is expressed in the next theorem which, though probably familiar, is included for completeness.

Theorem 6: For $x \in S$, $p_x = [1 + q_x \mu_x]^{-1}$.

Proof: The transition probability $p_{xx}(t)$ approaches p_x as $t \to \infty$, and is expressible as

$$p_{xx}(t) = e^{-q_x t} + \int_0^t e^{-q_x(t-u)} \, dH_x(u).$$

Since stays in x and returns to x are all mutually independent, the stays being identically distributed, and the returns also, the renewal theorem [Smith,[11] p. 247, formula (1.3)] implies that the right side approaches

$$\frac{\int_0^\infty e^{-q_x t} \, dt}{E\left\{\begin{matrix}\text{interval between successive} \\ \text{returns to } x\end{matrix}\right\}} = \frac{1}{1 + q_x \mu_x}.$$

Thus p_x can be calculated from μ_x where

$$\mu_x = \int_0^t u \, d\mathrm{Pr}\{\text{return time to } x \leq u\}$$

$$= \int_0^\infty \mathrm{Pr}\{\text{return time to } x > u\} \, du.$$

For our purposes it is convenient to approach the calculation of p_x in a slightly different way. Let z be any state, and let x be a state dis-

tinct from z, $z \neq x$. Define $q_{zx}(t)$ in $t \geq 0$ to be the probability that if the stochastic process start at z at time zero, it be at x at time t *without having returned to z*. Thus

$$q_{zx}(t) = \Pr\{x_t = x \quad \text{and} \quad (\text{epoch of first return to } z) > t \mid x_0 = z\}.$$

For convenience, we set $q_{zz}(t) \equiv 0$ in $t \geq 0$.

Lemma 2: For $z \neq x$, $t \geq 0$,

$$p_{zx}(t) = \int_0^t q_{zx}(t - u) \, dH_z(u).$$

Proof: Let t_i, $i = 1, 2, \ldots$ be the epoch of the ith return to z in $t > 0$, and let $A_i(t)$ be the event

$$\{x_t = x \quad \text{and} \quad t_i \leq t < t_{i+1}\}.$$

Then

$$\Pr\{A_i(t) \mid x_0 = z\} = \int_0^t q_{zx}(t - u) \, d\Pr\{t_i \leq u \mid x_0 = z\}.$$

However (cf. Ref. 11, p. 251, formula (1.7)),

$$H_z(t) = \sum_{i=1}^{\infty} \Pr\{t_i \leq t \mid x_0 = z\}$$

and

$$p_{zx}(t) = \sum_{i=1}^{\infty} \Pr\{A_i(t) \mid x_0 = z\}.$$

The integration and the summation can be interchanged by the monotone convergence theorem, and the lemma follows.

Lemma 3: For $z \neq x$,

$$p_x = \frac{q_z}{1 + q_z \mu_z} \int_0^{\infty} q_{zx}(u) \, du.$$

Proof: The integral on the right exists, since

$$\int_0^{\infty} q_{zx}(u) \, du = E\{\text{time spent in } x \text{ between successive returns to } z\} \leq \mu_z.$$

The lemma follows from Lemma 2 and the renewal theorem.

The matrix A is defined by the condition $A = (a_{xy})$ with

$$a_{xy} = \begin{cases} \dfrac{q_{xy}}{q_x} & x \neq y \\[2mm] 0 & x = y. \end{cases}$$

It can be verified that A is a stochastic matrix, indeed, the one-step transition probability matrix of a Markov stochastic process $\{x_n , n$ an integer$\}$ taking values on S; x_n is a discrete-time analog of x_t obtained by ignoring the lengths of time spent in a state.

Lemma 4: For $z \neq x$,

$$\int_0^\infty q_{zx}(u) \ du = \frac{1}{q_x} \ E \left\{ \begin{array}{l} \text{number of arrivals at } x \text{ between} \\ \text{successive returns to } z \end{array} \right\}.$$

Proof: The integral is the expected time spent in x between successive returns to z. Each stay in x has mean length $1/q_x$, and the stays are independent of the rest of the trajectory followed.

Lemma 5: For $z \neq x$,

$$E \left\{ \begin{array}{l} \text{number of arrivals at } x \text{ between} \\ \text{successive returns to } z \end{array} \right\}$$

$$= \sum_{n=1}^\infty \Pr\{x_n = x \quad \text{and} \quad x_j \neq z \quad \text{for} \quad 1 \leqq j \leqq n \mid x_0 = z\}.$$

Proof: We remark that the expectation on the left is the same for both x_t and x_n . The lemma is then a special case of the theorem that if $\{A_i , i = 1, 2, \ldots\}$ are any events, then the expected number of A_i that occur is

$$\sum_{i=1}^\infty \Pr\{A_i\}.$$

Lemma 6:

$$\mu_z = \sum_x \int_0^\infty q_{zx}(u) \ du.$$

Proof: This is an immediate consequence of $q_{zz}(\cdot) \equiv 0$ and

$$\int_0^\infty q_{zx}(u) = E\{\text{time spent in } x \text{ between successive returns to } z\}.$$

for $z \neq x$.

Lemma 7: Let $x \neq z$. Then

$$\Pr\{x_n = x \quad \text{and} \quad x_j \neq z \quad \text{for} \quad 1 \leqq j \leqq n \mid x_0 = z\}$$

$$= \frac{q_x}{q_z} \sum_{\substack{\pi \in K_{zx} \\ l(\pi)=n}} \prod_{y \in \pi} \frac{q_{py,\iota}}{q_y}.$$

Proof: The event in question can occur in as many ways as there are

paths of length n in K_{zx}. The probability that $\{x_j , 0 \leq j \leq n\}$ follow a path π from z to x is

$$\frac{q_x}{q_z} \prod_{y \in \pi} \frac{q_{py,y}}{q_y} ,$$

the quotient in front correcting for the end points.

Combining Theorem 6 with Lemmas 4, 5, 6, and 7, we obtain the following explicit formula for the stationary probabilities:

Theorem 7: If $x \neq z$, then

$$p_x = \frac{1}{1 + q_z \mu_z} \sum_{\pi \in K_{zx}} \prod_{y \in \pi} \frac{q_{py,y}}{q_y}$$

with

$$q_z \mu_z = \sum_{x \neq z} \sum_{\pi \in K_{zx}} \prod_{y \in \pi} \frac{q_{py,y}}{q_y} ,$$

and

$$p_z = \frac{1}{1 + q_z \mu_z} .$$

We remark that Theorem 5 follows from the above if we choose $z = 0 =$ zero state, omit normalization, and observe that only products along permitted paths ($\pi \in P$) are nonzero. Theorem 7 is an analog for continuous parameter processes of a theorem of Derman[12] for Markov chains.

XII. EXPANSION OF THE STATIONARY VECTOR p IN POWERS OF λ

We now turn to examining, in an elementary way, the analytical dependence of the state probabilities $\{p_x , x \in S\}$ on the calling rate λ, on the structure of the network, and on the routing matrix R. It will be shown that the partial ordering $\leq$ of the set S of states can be used to calculate the elements of p by expanding the ratios

$$\frac{p_x}{p_0} \qquad x > 0$$

in powers of the traffic parameter λ in a neighborhood of $\lambda = 0$, and then determining the coefficients of this expansion from the structure of the network and the routing matrix by a recursive procedure. The solution so obtained is later (Section XV) extended to arbitrary real positive values of λ by analytic continuation, and the coefficients are calculated.

Our approach to studying the stationary probability vector p will be guided by these intuitive remarks: it is known that in various simple models (of connecting systems carrying random traffic with blocked calls refused) the probability that k calls be in existence is proportional to the kth power of a constant associated with the calling rate divided by k factorial. For example, in Erlang's model for c trunks with blocked calls cleared, the chance that k calls are in progress is proportional to

$$\frac{a^k}{k!}, \qquad 0 \leqq k \leqq c$$

where a is the calling rate. Note that the exponent of a is the number of calls in progress, i.e., the current difference between the cumulative number of new calls and that of hang-ups, assuming that the system started in the zero state. The factorial in the denominator is the number of orders in which the k calls in progress could all hang up, or alternatively, could all have arisen.

The situation in our model is very similar. Each call still in progress required an event occurring at the rate λ to put it in existence; for each state x, there are exactly $|x|!$ orders in which the $|x|$ calls in progress in x could arise, or terminate. These circumstances suggest that for $x > 0$, p_x might be of order $\lambda^{|x|}$ as $\lambda \to 0$, and that the coefficient of $\lambda^{|x|}$ in p_x might involve $|x|!$ in the denominator. These conjectures are true, and are the first step in the systematic calculation of p_x by expansion in powers of λ, to be carried out in this section.

We first record some analytical properties of p as preliminary results. Some of these results could be obtained as consequences of the basic solutions given in Section X. Most of the proofs to be given, however, are independent of Section X, and proceed by simple arguments from the equilibrium equation.

When we need to view p_x as a function of the parameter λ, we write $p_x = p_x(\lambda)$, $x \in S$, or in vector form, $p = p(\lambda)$.

Lemma 8: $\lim_{\lambda \to 0} p_x(\lambda) = \delta_{x0}$.

Proof: Let x be a maximal state in the partial ordering $\leqq$ of the set S of all states. Then $s(x) = 0$, and

$$|x| p_x(\lambda) = \lambda \sum_{y \in B_x} p_y(\lambda) r_{yx}.$$

Since $0 \leqq p_y(\lambda) \leqq 1$ for all $\lambda > 0$ and all $y \in S$, the lemma is true for maximal states. Assume, as a hypothesis of induction, that the lemma is true for all y with $|y| \geqq k + 1$. Then for $x \in L_k$, $k > 0$,

$$[|x| + \lambda s(x)]p_x(\lambda) = \sum_{y \in A_x} p_y(\lambda) + \lambda \sum_{y \in B_x} p_y(\lambda) r_{yx},$$

and so $p_x(\lambda) \to 0$ as $\lambda \to 0$. The proof is completed by observing that for each $\lambda > 0$,

$$p_0(\lambda) = 1 - \sum_{x>0} p_x(\lambda).$$

Lemma 9: For each $x \, \epsilon \, S$, p_x is the restriction to real positive argument of a rational function $p_x(\cdot)$ of a complex variable μ. The function $p_x(\cdot)$ has no poles in a neighborhood of the half-line $\mathrm{Re}(\mu) \geq 0$, $\mathrm{Im}(\mu) = 0$, and an expansion

$$\frac{p_x(\mu + \epsilon)}{p_0(\mu + \epsilon)} = \sum_{m=0}^{\infty} \epsilon^m c_m(x,\mu),$$

with real coefficients $c_m(x,\lambda)$ is valid for $\mathrm{Re}(\mu) \geq 0$, $\mathrm{Im}(\mu) = 0$, and $| \epsilon |$ small enough.

Proof: The equation $Qp = 0$ can be solved for a normalized (i.e., probability) vector $p(\lambda)$ by successive elimination or by use of Theorem 4. Either procedure gives rise to an algebraic expression for $p_x(\lambda)$, $x \, \epsilon \, S$. Let $p_x(\mu)$ be that rational function of a complex variable μ defined by substituting μ for λ in this algebraic expression. Since $0 \leq p_x(\lambda) \leq 1$, $p_x(\mu)$ has no poles in a neighborhood of the nonnegative real axis. To justify the expansion we show that $p_x(\cdot)/p_0(\cdot)$ is also analytic in that neighborhood. But this is immediate because by Lemma 8,

$$p_x(\lambda) \to \delta_{0x} \quad \text{as} \quad \lambda \to 0,$$

and by Theorem 1, $p_0(\lambda) > 0$ for $\lambda \geq 0$ because the zero state belongs to the ergodic class $S - F$.

Setting $\mu = 0$ and $\epsilon = \lambda$ in Lemma 9, we obtain an expansion of p_x/p_0 in powers of the traffic parameter λ,

$$\frac{p_x(\lambda)}{p_0(\lambda)} = \sum_{m=0}^{\infty} \lambda^m c_m(x,0),$$

valid for λ small enough.

Theorem 8: For $k \geq 0$ and $x \, \epsilon \, L_k$,

$$p_k = \sum_{x \epsilon L_k} p_x = O(\lambda^k) \quad \text{as} \quad \lambda \to 0,$$

and

$$p_x = O(\lambda^k) \quad \text{as} \quad \lambda \to 0.$$

Proof: We prove both results simultaneously by induction. By Lemma 8, the result is true for $k = 0$. Assume that it is true up through $k - 1 \geq 0$. From Lemma 1 and the induction hypothesis, we find

$$p_k = \frac{\lambda}{k} \sum_{y \epsilon L_{k-1}} p_y s(y)$$

$$= O(\lambda^k) \quad \text{as} \quad \lambda \to 0.$$

By Lemma 9, for λ small enough p_x can be written as a power series around $\lambda = 0$

$$p_x = \sum_{m=0}^{\infty} \lambda^m c_m(x,0).$$

Thus

$$\sum_{m=0}^{\infty} \lambda^m \sum_{x \epsilon L_k} c_m(x,0) = p_k = O(\lambda^k) \quad \text{as} \quad \lambda \to 0.$$

The first nonvanishing coefficient in the expansion on the left must be positive, else $p_k < 0$ for $\lambda > 0$ small enough, which is impossible since $p_x(\lambda) \geqq 0$ for $\lambda > 0$. Hence

$$\sum_{x \epsilon L_k} c_0(x,0) \geqq 0.$$

However, the first nonvanishing coefficient in the expansion of p_x must also be positive, for the same reason as above, namely, that $p_x(\lambda) \geqq 0$ for $\lambda \geqq 0$. Thus $c_0(x,0) \geqq 0$. Hence $p_k = O(\lambda^k)$ as $\lambda \to 0$ implies $c_0(x,0) = 0$. We apply the same argument successively to show that for $x \epsilon L_k$, the coefficients $c_1(x,0), \ldots, c_{k-1}(x,0)$ are all zero, and the theorem is proven.

Theorem 9: For $x > 0$

$$p_x = p_0 \frac{\lambda^{|x|}}{|x|!} r_x + o(\lambda^{|x|}) \quad \text{as} \quad \lambda \to 0$$

where
$r_x = (R^{|x|})_{0x}$
 $= \text{the } 0,x \text{ entry of the } |x| \text{th power of the routing matrix } R$
 $= \text{the number of permitted strictly ascending paths from } 0 \text{ to } x.$

Proof: The equation of statistical equilibrium that defines p is

$$[|x| + \lambda s(x)]p_x = \sum_{y \epsilon A_x} p_y + \lambda \sum_{y \epsilon B_x} p_y r_{yx}, \quad x \epsilon S.$$

For convenience, suppose that $|x| = k$. We divide the equation by p_0, use Lemma 9 to expand the components of p/p_0 in powers of λ, and equate the coefficients of λ^k on each side of the equation. This gives

$$k c_k(x,0) + s(x)c_{k-1}(x,0) = \sum_{y \epsilon A_x} c_k(y,0) + \sum_{y \epsilon B_x} c_{k-1}(y,0)r_{yx}.$$

By Theorem 8, $c_{k-1}(x,0) = 0$ and $c_k(y,0) = 0$ for $y \epsilon A_x$. Therefore

$$kc_k(x,0) = \sum_{y \epsilon B_x} c_{k-1}(y,0)r_{yx},$$

or in general, with γ the vector with components $c_{|y|}(y,0)$,

$$c_{|x|}(x,0) = \frac{1}{|x|} \sum_{y \epsilon B_x} c_{|y|}(y,0)r_{yx}$$

$$= \frac{1}{|x|} c_{|y|}(y,0)r_{yx}$$

$$= \frac{1}{|x|} (R\gamma)_x.$$

Iterating this relation $|x|$ times, we find

$$c_{|x|}(x,0) = \frac{1}{|x|!} (R^{|x|} \gamma)_x.$$

Now it is easily seen that the y,x entry of R^k is zero unless $k = |x| - |y|$, and in particular, if $k = |x|$, this entry is zero unless $y = 0$. Thus

$$c_{|x|}(x,0) = \frac{1}{|x|!} (R^{|x|})_{0x}c_0(0,0),$$

and it is obvious from the definition of the $c_m(y,0)$ that $c_0(0,0) = 1$.

Theorem 10: Let the sequences $\{c_m(x,0), m \geqq 0, x \epsilon S\}$ be defined recursively by

$$c_m(0,0) = \delta_{m0}$$

$$c_m(x,0) = 0 \quad for \quad 0 \leqq m < |x|$$

$$and \quad x > 0$$

$$c_{|x|}(x,0) = \frac{r_x}{|x|!}$$

$$|x| c_m(x,0) + s(x)c_{m-1}(x,0) = \sum_{y \epsilon A_x} c_m(y,0) + \sum_{y \epsilon B_x} c_{m-1}(y,0)t_{yx}$$

$$for \quad m > |x| \quad and \quad x > 0.$$

If for $x > 0$

$$\lambda < (\limsup_{m \to \infty} |c_m(x,0)|^{1/m})^{-1}$$

then the component p_x of p is given by

$$p_x = p_0 \sum_{m=0}^{\infty} \lambda^m c_m(x,0).$$

If

$$\lambda < \min_{x>0} \; (\limsup_{m\to\infty} |\, c_m(x,0)\,|^{1/m})^{-1},$$

then the probability p_0 of the zero state is determined by the normalization condition $\sum_{x \epsilon S} p_x = 1$ as

$$p_0 = \frac{1}{1 + \sum_{x>0} \sum_{m=0}^{\infty} \lambda^m c_m(x,0)}.$$

Proof: This result follows immediately from Lemma 9 and Theorem 9, using the standard formula for the radius of convergence of a power series.

XIII. EXPANSION OF THE PROBABILITY OF BLOCKING IN POWERS OF λ

With a method of calculating equilibrium state probabilities for small λ at hand (in principle, at least) we now show how the probability b of blocking can be calculated, to any desired degree of accuracy, by an expansion in powers of the traffic parameter λ, assumed sufficiently small. In most connecting networks of practical interest, none of the states near the bottom of the state-diagram has any blocked calls, so that it is necessary for a state x to have certain minimum number of calls in progress before it can have any blocked idle pairs. To take advantage of this situation in our calculation, we let

$n = $ least k such that some call is blocked in a state of L_k.

Theorem 11: The probability b of blocking can be expanded in a power series in λ in a neighborhood of $\lambda = 0$; only terms of order higher than or equal to λ^n appear.

Proof: From Theorem 2 we have, since $\beta_x = 0$ for $|x| < n$, and $c_k(x,\lambda) = 0$ for $k < |x|$,

$$b = \frac{\sum_{n \leq |x|} p_x \beta_x}{\sum_{x \epsilon S} p_x \alpha_x}$$

$$= \frac{\lambda^n \sum_{j=0}^{\infty} \lambda^j \sum_{n \leq |x| \leq n+j} c_{n+j}(x,0)\beta_x}{\sum_{j=0}^{\infty} \lambda^j \sum_{j \leq |x|} c_j(x,0)\alpha_x} = \frac{\lambda^n B(\lambda)}{A(\lambda)}.$$

(3)

Since the denominator is not zero in a neighborhood of $\lambda = 0$, $b = b(\lambda)$ is analytic there and can be expanded in powers of λ. Up to terms of order λ^{n+2} this expansion is

$$b = \lambda^n \frac{B(0)}{A(0)} + \lambda^{n+1} \frac{A(0)B'(0) - A'(0)B(0)}{[A(0)]^2}$$
$$+ \lambda^{n+2}\left(\frac{A'(0)B''(0) - A''(0)B(0)}{2[A(0)]^2}\right.$$
$$\left. - \frac{A'(0)A(0)B'(0) - A'^2(0)B(0)}{[A(0)]^3}\right)$$
$$+ o(\lambda^{n+2}).$$

The coefficients in the first two terms can be obtained by the following calculations:

$$A(0) = \sum_{|x| \geqq 0} c_0(x,0)\alpha_x$$
$$= c_0(0,0)\alpha_0$$
$$= \alpha_0 ,$$

$$B(0) = \sum_{|x| \geqq n} c_n(x,0)\beta_x$$
$$= \sum_{|x|=n} c_n(x,0)\beta_x$$
$$= \frac{1}{n!} \sum_{x \epsilon L_n} r_x \beta_x ,$$

$$A'(0) = \sum_{|x| \geqq 1} c_1(x,0)\alpha_x$$
$$= \sum_{|x|=1} c_1(x,0)\alpha_x$$
$$= \sum_{x \epsilon L_1} r_{0x}\alpha_x ,$$

$$B'(0) = \sum_{|x| \geqq n+1} c_{n+1}(x,0)\beta_x$$
$$= \sum_{x \epsilon L_{n+1}} c_{n+1}(x,0)\beta_x + \sum_{x \epsilon L_n} c_{n+1}(x,0)\beta_x$$
$$= \frac{1}{(n+1)!} \sum_{x \epsilon L_{n+1}} r_x \beta_x + \sum_{x \epsilon L_n} c_{n+1}(x,0)\beta_x .$$

The constants $\{c_{k+1}(x,0), |x| = k\}$ can be determined by the following recurrence, obtained from Theorem 10:

$$c_1(0,0) = 0,$$

$$c_{n+1}(x,0) = \frac{1}{n}\left\{\frac{1}{(n+1)!}\sum_{y\epsilon A_x} r_x + \sum_{y\epsilon B_x} c_n(y,0)r_{yx} - \frac{s(x)r_x}{n!}\right\}.$$

Our results can be put in a slightly more explicit form by expanding $\log b$ rather than b, and using the fact that $A(\cdot)$ and $B(\cdot)$, as defined by (3), are *generating functions*. We have

$$\log b = n\log\lambda + \log B(\lambda) - \log A(\lambda).$$

Except for the systematic absence of factorials, the coefficients in the expansion of $\log B(\lambda)$ are related to those in the expansion of $B(\lambda)$ as cumulants are to moments. Set

$$b_j = \sum_{n\leq|x|\leq n+j} c_{n+j}(x,0)\beta_x, \qquad j = 0, 1, \ldots,$$

$$a_j = \sum_{j\leq|x|} c_j(x,0)\alpha_x, \qquad j = 0, 1, \ldots,$$

so that

$$B(\lambda) = \sum_{j=0}^{\infty} \lambda^j b_j,$$

$$A(\lambda) = \sum_{j=0}^{\infty} \lambda^j a_j.$$

Then, by a standard formula (Riordan,[13] p. 37),

$$\log B(\lambda) = \sum_{j=0}^{\infty} \lambda^j z_j(b),$$

$$\log A(\lambda) = \sum_{j=0}^{\infty} \lambda^j k_j(a),$$

where for $u = a$ or b (sequences)

$$\kappa_n(u) = \sum (u_1)^{k_1} \cdots (u_n)^{k_n} \frac{(-1)^{k-1}(k-1)!}{(k_1)!\cdots(k_n)!},$$

with $k = k_1 + k_2 + \cdots + k_n$, and the sum over all partitions of n, i.e., all solutions in nonnegative integers of $k_1 + 2k_2 + \cdots + nk_n = n$.

XIV. COMBINATORY INTERPRETATION AND CALCULATION OF THE CONSTANTS $\{c_m(x,0), x\epsilon S, m \geq 0\}$

We shall now evaluate the coefficients in the power series expansion of p around $\lambda = 0$ explicitly as sums of products on paths in S. Additional combinatory notions that enter this calculation are discussed first.

A path is said to *contain a loop* if it returns (one or more times) to a place where it has been previously. Thus $\pi = \{x_0, \cdots, x_l\}$ has a loop if there are integers i and j, $0 \leq i < j \leq l$, such that

$$x_i = x_j.$$

A *circuit* is a path that ends where it begins. Thus the generic path $\pi = \{x_0, \ldots, x_l\}$ is a circuit if $x_0 = x_l$. A circuit is a *loop* if it contains no subcircuits, i.e., π is a loop if $x_0 = x_l$ and $i \neq j$ implies

$$x_i \neq x_j$$

for any $0 \leq i \leq l$ and $0 < j < l$.

A circuit of length two is of necessity a loop. Furthermore, the circuits of length two on S can be partitioned into two classes according to the direction in which they are traversed. Each such loop can be thought of as obtained either by first adding a new call and then removing that same call, or by removing a call and then replacing that same call. Thus if $\pi = \{x_0, x_1, x_2\}$ is a loop of length two on S, then

$$x_0 = x_2$$

and either

$$x_2 = x_0 \; \epsilon \; B_{x_1}$$

or

$$x_2 = x_0 \; \epsilon \; A_{x_1}$$

and, of course, not both. In the first instance we say that π is of the *first kind*. Thus a loop of the first kind is a path π of the form $\{x,y,x\}$ with $y \; \epsilon \; A_x$, i.e., it is a trajectory in S obtained by starting at a state x, adding a new call to go to a state y, and then removing that very same call to return to x.

For a path $\pi = \{x_0, \ldots, x_l\}$ and a state x we say that x is *on* π if x is one of $x_0, x_1, \ldots, x_l$. If x is on π, we say that a loop of the first kind on π ends at x if for some i in the range $2 \leq i \leq l$,

$$x_i = x_{i-2} = x$$

and

$$x_{i-1} \; \epsilon \; A_x,$$

that is, the subpath $\{x_{i-2}, x_{i-1}, x_i\}$ is a loop of the first kind, and x_i is x.

With π and x as above we define

$$g(\pi,x) = \begin{cases} |x|, & \text{if } x \text{ is on } \pi \text{ and a loop of the first} \\ & \text{kind on } \pi \text{ ends at } x \\ 1, & \text{otherwise,} \end{cases}$$

and we denote by $\nu(\pi)$ the total number of loops of the first kind contained in π.

With these combinatory preliminaries behind us, we are ready to prove

Theorem 12: The coefficients $\{c_m(x,0), x \epsilon S, m \geq 0\}$ are given by

$$c_m(0,0) = \delta_{m0},$$

and for $x > 0$,

$$c_m(x,0) = \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-|x|}} (-1)^{\nu(\pi)} \prod_{y \epsilon \pi} \frac{g(\pi,y)}{|y|}. \tag{4}$$

Before proving the theorem it is probably helpful to state it in words, thus: To calculate $c_m(x,0)$, consider all the paths π that are permitted by R, start at 0, never return to zero, end up at x, and have length $2m - |x|$ (i.e., consist of m new calls and $m - |x|$ hangups); along each such path π take the product of the reciprocals of the numbers of calls in progress in the states traversed by π, omitting states at which a loop of the first kind ends; weight the product positive or negative according as π has an even or an odd number of loops of the first kind; add up all the weighted products.

Proof of Theorem 12: We already know that $c_m(x,0) = \delta_{m0}$ and that $c_m(x,0) = 0$ for $x > 0$ and $m < |x|$. The latter result is consistent with Theorem 12 because no path from 0 can reach x in fewer than $|x|$ steps. Consider then the case $m - |x| \geq 0$. Any path π from 0 to x of length $2m - |x| = m$ consists entirely of new calls, and for such a π

$$\prod_{y \epsilon \pi} \frac{g(\pi,y)}{|y|} = \frac{1}{|x|!}.$$

The number of such paths, summed over in (4), is easily seen to be r_x, the number of permitted strictly ascending paths from 0 to x. Thus for $m = |x|$, (4) states that

$$c_m(x,0) = \frac{r_x}{|x|!},$$

as was proven in Theorem 9.

The remainder of the proof is by upward induction on m and finite downward induction on x, using the recurrence formula

$$| x | c_m(x,0) + s(x)c_{m-1}(x,0) = \sum_{y \epsilon A_x} c_m(y,0) + \sum_{y \epsilon B_x} c_{m-1}(y,0)r_{yx} ,$$

given by Theorem 10. The starting step of the induction is the fact that formula (4) holds for $x \epsilon S$ and m such that

$$0 < | x | \leqq m \leqq 1.$$

This is a consequence of the fact, already proven, that (4) holds for $| x | = m$.

Let us assume, as a hypothesis of induction, that the theorem holds for all $x \epsilon S$ and m such that

$$0 < | x | \leqq m \leqq k. \tag{5}$$

We shall prove from this, by downward induction on x, that it also holds for $x \epsilon S$ and m such that

$$0 < | x | \leqq m = k + 1.$$

This last condition exactly describes the new cases covered in extending (5) to $k + 1$.

Where π is a path, we use the natural notation πx to denote the path obtained from π by adding x to π as a new ultimate element, assuming that x is adjacent to the last element of π. We now observe that if πx is a path that does not end in a loop of the first kind, then $g(\pi x,x) = 1$, and so

$$\prod_{z \epsilon \pi x} \frac{g(\pi x,z)}{| z |} = \frac{1}{| x |} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{| z |} , \tag{6}$$

$$\nu(\pi x) = \nu(\pi.)$$

A state x is *maximal* in the partial ordering $\leqq$ if no new calls can be put up in x, for whatever reason. If x is maximal, then A_z is empty and $s(x) = 0$. A state x is *maximal in (a set)* $X \subseteq S$ if $x \geqq y$ for every $y \epsilon X$, and $x \epsilon X$.

Let $0 \leqq | x | \leqq m = k + 1$ and suppose first that x is maximal. Then

$$c_m(x,0) = \frac{1}{| x |} \sum_{y \epsilon B_x} c_{m-1}(y,0)r_{yx} .$$

No path ending at a maximal state x can end in a loop of the first kind, and any such path must have a $y \epsilon B_x$ with $r_{yx} = 1$ as a penultimate

element if it is to be a permitted path. Then clearly

$$\{\pi \ \epsilon \ P \cap K_{0x} : l(\pi) = 2m - |x|\}$$
$$= \bigcup_{y \epsilon B_x} \{\pi x : \pi \ \epsilon \ P \cap K_{0y}$$
$$\text{and} \quad l(\pi) = 2m - |x| - 1 \quad \text{and} \quad r_{yx} = 1\}.$$

Let now $y \ \epsilon \ B_x$, $\pi \ \epsilon \ P \cap K_{0y}$, $l(\pi) = 2m - |x| - 1$, and $r_{yx} = 1$. Then πx satisfies $g(\pi x, x) = 1$ and also formula (6). Thus formula (4) holds for maximal x and $0 \leqq |x| \leqq m = k + 1$, by the hypothesis of induction.

Next, consider states x that are maximal in the set

$$\{y \ \epsilon \ S : 0 \leqq |x| \leqq k + 1\}.$$

These are just the elements of L_{k+1}, i.e., the states x with $|x| = k + 1$. Since we are assuming $m = k + 1$, the result (4) holds for these x by Theorem 10.

Finally, assume as a hypothesis of downward induction (on $|x|$) that the result is true for $y \ \epsilon \ S$ and m such that

$$1 < j + 1 \leqq |y| \leqq m = k + 1,$$

and suppose that $|x| = j$. Then

$$c_m(x,0) = \frac{1}{|x|} \sum_{y \epsilon A_x} [c_m(y,0) - r_{xy}c_{m-1}(x,0)] + \frac{1}{|x|} \sum_{y \epsilon B_x} c_{m-1}(y,0)r_{yx}.$$

If π is a path on S the notation $pe(\pi)$ denotes the *penultimate element* of π, i.e., $pe(\pi) = z_{l-1}$ for $\pi = \{z_0, z_1, \ldots, z_l\}$. The notation $ape(\pi)$ denotes the *antepenultimate element* of π, i.e., for $\pi = \{z_0, z_1, \ldots, z_l\}$, $ape(\pi) = z_{l-2}$.

A path of length $2m - |x|$ belonging to $P \cap K_{0x}$ reaches x either via A_x or via B_x. In the latter case the path cannot end in a loop of the first kind. By the hypothesis of induction,

$$\frac{1}{|x|} \sum_{y \epsilon B_x} c_{m-1}(y,0)r_{yx} = \sum_{y \epsilon B_x} r_{yx} \sum_{\substack{\pi \epsilon P \cap K_{0y} \\ l(\pi)=2(m-1)-|y|}} \frac{(-1)^{\nu(\pi)}}{|x|} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}$$

$$= \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-|x| \\ pe(\pi) \epsilon B_x}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}, \tag{7}$$

the second equality following from (6).

Now consider a path πx of length $2m - |x|$, belonging to $P \cap K_{0x}$,

reaching x via A_x, and not ending in a loop of the first kind, i.e., with $\mathrm{pe}(\pi) \neq x$. Using the hypothesis of downward induction, we find

$$\frac{1}{|x|} \sum_{\substack{y \epsilon A_x}} \sum_{\substack{\pi \epsilon P \cap K_{0y} \\ l(\pi)=2m-|x|-1 \\ \mathrm{pe}(\pi)\neq x}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}$$

$$= \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ \mathrm{pe}(\pi)\epsilon A_x \\ \mathrm{ape}(\pi)\neq x}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}, \tag{8}$$

the second equality again following from (6).

Finally, we consider those paths of length $2m - |x|$ belonging to $P \cap K_{0x}$ which *do* end in a loop of the first kind. Such a path is of the form

$$\pi y x$$

with $\pi \epsilon P \cap K_{0x}$, $l(\pi) = 2m - |x| - 2$, and $r_{xy} = 1$, for some $y \epsilon A_x$. We observe that in this case $|y| = |x| + 1$ and that

$$\frac{(-1)^{\nu(\pi)}}{|x|}\left\{\prod_{z \epsilon \pi y} \frac{g(\pi y,z)}{|z|} - \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}\right\} = \frac{(-1)^{\nu(\pi)}}{|x|}\left\{\frac{1}{|y|} - 1\right\}\prod_{z \epsilon \pi}\frac{g(\pi,z)}{|z|}$$

$$= -\frac{(-1)^{\nu(\pi)}}{|y|}\prod_{z \epsilon \pi}\frac{g(\pi,z)}{|z|}$$

$$= (-1)^{\nu(\pi)+1}\frac{g(\pi yx,y)}{|y|} \tag{9}$$

$$\cdot\frac{g(\pi yx,x)}{|x|}\cdot\prod_{z \epsilon \pi}\frac{g(\pi,z)}{|z|}$$

$$= (-1)^{\nu(\pi)+1}\prod_{z \epsilon \pi yx}\frac{g(\pi yz,z)}{|z|}.$$

We note that $\nu(\pi yx) = \nu(\pi) + 1$, and that $\nu(\pi y) = \nu(\pi)$. By summing formula (9) over paths of length $2m = |x| - 2$ belonging to $P \cap K_{0x}$ *and* over $y \epsilon A_x$ such that $r_{xy} = 1$, we obtain

$$\frac{1}{|x|}\sum_{\substack{y \epsilon A_x \\ r_{xy}=1}}\sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-|x|-2}}\left\{(-1)^{\nu(\pi t)}\prod_{z \epsilon \pi y}\frac{g(\pi y,z)}{|z|}\right.$$

$$\left. - (-1)^{\nu(\pi)}\prod_{z \epsilon \pi}\frac{g(\pi,z)}{|z|}\right\} = \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-|x| \\ \mathrm{pe}(\pi)\epsilon A_y \\ \mathrm{ape}(\pi)=x}} (-1)^{\nu(\pi)}\prod_{z \epsilon \pi}\frac{g(\pi,z)}{|z|}. \tag{10}$$

By the hypothesis of induction

$$\frac{1}{|x|} \sum_{\substack{y \epsilon A_x \\ r_{xy}=1}} \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-2-|x|}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|} = \frac{1}{|x|} \sum_{y \epsilon A_x} r_{xy} c_{m-1}(x,0). \quad (11)$$

Also, it can be seen that for $y \epsilon A_x$

$$\{\pi \epsilon P \cap K_{0y}: l(\pi) = 2m - |x| - 1 \quad \text{and} \quad \text{pe}(\pi) \neq x\}$$

$$\cup \{\pi y: \pi \epsilon P \cap K_{0x}, r_{xy} = 1, \quad \text{and} \quad l(\pi) = 2m - |x| - 2\} \quad (12)$$

$$= \{\pi \epsilon P \cap K_{0y}: l(\pi) = 2m - |y|\}.$$

Combining (8) and (10), using (11) and (12), and applying the hypothesis of downward induction, we find

$$\sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=2m-|x| \\ \text{pe}(\pi) \epsilon A_x}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|}$$

$$= \frac{1}{|x|} \sum_{y \epsilon A_x} \sum_{\substack{\pi \epsilon P \cap K_{0y} \\ l(\pi)=2m-|y|}} (-1)^{\nu(\pi)} \prod_{z \epsilon \pi} \frac{g(\pi,z)}{|z|} - \frac{1}{|x|} \sum_{y \epsilon A_x} r_{xy} c_{m-1}(x,0) \quad (13)$$

$$= \frac{1}{|x|} \sum_{y \epsilon A_x} \{c_m(y,0) - r_{xy} c_m(x,0)\}.$$

Together, (7) and (13) complete the inductive step.

XV. CALCULATION OF $c_m(x,\lambda)$

In order to give an explicit formula for $c_m(x,\lambda)$ for $\lambda > 0$ we suppose that, for each path π on S, all the upward (in $\leqq$) transitions on π are of two kinds, denoted by the symbols λ and ϵ. (The calculation we present is more easily understood if the new calls labeled ϵ are thought of as due to the increment ϵ in calling rate, while those labeled λ are due to the original calling rate λ.) In other words, we consider the set of all paths π on S as (partially) labeled by assigning either λ or ϵ to each upward transition. Formally, we define a *labeling* $\lambda(\cdot)$ of a path $\pi = \{x_0, x_1, \ldots, x_l\}$ to be any function defined for $x_1, x_2, \ldots, x_l$ with the property

$$\lambda(x_i) = \begin{cases} \epsilon \quad \text{or} \quad \lambda & \text{if} \quad x_i \text{ covers } x_{i-1}, \\ 0 & \text{otherwise.} \end{cases}$$

The set of *all possible labelings* of a path π is denoted by $\Lambda(\pi)$, and membership therein by the notation $\lambda(\cdot) \epsilon \Lambda(\pi)$. The *functions* $\lambda(\cdot)$

should not be confused with the *constant* λ. A path π together with a labeling $\lambda(\cdot)$ of π will be called a *labeled path* and denoted by $(\pi,\lambda(\cdot))$. The *index* $\epsilon(\pi,\lambda(\cdot))$ of a labeled path is the number of times $\lambda(x)$ assumes the value ϵ for $x \epsilon \pi$.

Let $\lambda(\cdot) \epsilon \Lambda(\pi)$ be a labeling of $\pi = \{x_0, x_1, \ldots, x_l\}$. The function $h(\pi,\lambda(\cdot),\cdot)$ is defined by the condition

$$h(\pi,\lambda(\cdot),x) = |x| + \lambda s(px)$$

if x is on π and a loop of the first kind on π ends at x, with the first (i.e., upward going) leg of the loop labeled ϵ by $\lambda(\cdot)$, i.e.,

$$\lambda(px) = \epsilon,$$

and by

$$h(\pi,\lambda(\cdot),x) = 1$$

in all other cases.

For a path π and a labeling $\lambda(\cdot) \epsilon \Lambda(\pi)$, the function $\zeta(\pi,\lambda(\cdot))$ is defined by

$$\zeta(\pi,\lambda(\cdot)) = \begin{cases} \text{the number of loops of the first kind on } \pi \\ \text{labeled } \epsilon \text{ on the upward leg by } \lambda(\cdot). \end{cases}$$

Theorem 13: For $\lambda > 0$ and $m \geqq 0$

$$c_0(0,\lambda) = \delta_{0m}$$

and for $x > 0$,

$$c_m(x,\lambda) = \lambda^{-m} \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ \lambda(\cdot)\epsilon\Lambda(\pi) \\ \epsilon(\pi,\lambda(\cdot))=m}} \lambda^{(l(\pi)+|x|)/2}(-1)^{\zeta(\pi,\lambda(\cdot))} \prod_{y\epsilon\pi} \frac{h(\pi,\lambda(\cdot),y)}{|y| + \lambda s(y)}. \quad (14)$$

Note that by Theorem 5, with $z = 0$, formula (14) for $m = 0$ reduces, as it should, to

$$c_0(x,\lambda) = \frac{p_x(\lambda)}{p_0(\lambda)}.$$

Also, formula (14) agrees with formula (4), Theorem 12, as λ is allowed to approach zero, since in this limit only $(\pi,\lambda(\cdot))$ which are full of ϵ's contribute, with $m = \frac{1}{2}(l(\pi) + |x|)$.

Proof of Theorem 13: The values of $c_0(x,\lambda)$ and $c_m(0,\lambda)$ are consequences of the definition

$$\sum_{m=0}^{\infty} \epsilon^m c_m(x,\lambda) = \frac{p_x(\lambda + \epsilon)}{p_0(\lambda + \epsilon)}. \quad (15)$$

The equilibrium condition $Qp = 0$ for traffic parameter $\lambda + \epsilon$ comprises exactly the equations

$$[\,|\,x\,| + (\lambda + \epsilon)s(x)]p_x(\lambda + \epsilon)$$
$$= \sum_{y \epsilon A_x} p_y(\lambda + \epsilon) + (\lambda + \epsilon) \sum_{y \epsilon B_x} p_y(\lambda + \epsilon)r_{yx}, \qquad x \epsilon S.$$

Dividing by $p_0(\lambda + \epsilon) > 0$, substituting the expansion (15), and collecting coefficients of like powers of ϵ, we find that $c_m(x,\lambda)$ for $x \epsilon S$ satisfies the equation

$$[\,|\,x\,| + \lambda s(x)]c_m(x,\lambda) = \sum_{y \epsilon A_x} [c_m(y,\lambda) - r_{xy}c_{m-1}(x,\lambda)]$$
$$+ \sum_{y \epsilon B_x} r_{yx}[c_{m-1}(y,\lambda) + \lambda c_m(y,\lambda)]. \tag{16}$$

It can be verified, using the fact that for any labeling $\lambda(\cdot)$ of a path π the number of times $\lambda(x)$ has the value λ for $x \epsilon \pi$ is

$$\frac{l(\pi) + |\,x\,|}{2} - \epsilon(\pi,\lambda(\cdot)),$$

that formula (14) gives a formal solution of equations (16).

To prove the theorem it suffices to show that the infinite sum over $\pi \epsilon P \cap K_{0x}$ in formula (14) is absolutely convergent, and that the left-hand side of formula (15), with the $c_m(x,\lambda)$ as given by the theorem is absolutely convergent for ϵ small enough.

We first observe that for each π and $\lambda(\cdot)$ summed over in formula the factors $h(\pi,\lambda(\cdot),y)$ in

$$\prod_{y \epsilon \pi} \frac{h(\pi,\lambda(\cdot),y)}{|\,y\,| + \lambda s(y)}$$

are uniformly bounded, and that at most $\min(m,\nu(\pi))$ of them are greater than unity. Also, the number of upward transitions (new calls) along a path $\pi \epsilon P \cap K_{0x}$ of length $l(\pi)$ is just

$$\frac{l(\pi) + |\,x\,|}{2}.$$

Thus there are exactly

$$\binom{\frac{1}{2}l(\pi) + \frac{1}{2}|\,x\,|}{m}$$

ways of labeling the upward transitions on a path $\pi \epsilon P \cap K_{0x}$ with length $l(\pi)$ and index m.

Hence for some constant $a > 0$

$$| c_m(x,\lambda) | \leq a^m \sum_{\pi \epsilon P \cap K_{0x}} \lambda^{(l(\pi)+|x|)/2} \left(\frac{\dfrac{l(\pi) + | x |}{2}}{m} \right) \prod_{y \epsilon \pi} \frac{1}{| y | + \lambda s(y)} \,.$$

By Lemma 7, with $\{x_i \,, i$ an integer$\}$ of the lemma defined in terms of the matrix Q appropriate to our congestion problem,

$$\sum_{x > 0} \frac{| x | + \lambda s(x)}{\lambda s(0)} \sum_{\substack{\pi \epsilon P \cap K_{0x} \\ l(\pi)=k}} \lambda^{(k+|x|)/2} \prod_{y \epsilon \pi} \frac{1}{| y | + \lambda s(y)}$$
$$= \Pr \{x_i \neq 0 \quad \text{for} \quad 1 \leq i \leq k \mid x_0 = 0\}.$$

By ex. 19, p. 378 of Feller,[4] there exists $0 < q < 1$ such that the probability on the right is at most q^k for all $k \geq | S |$. Hence

$$\sum_{x > 0} | c_m(x,\lambda) | \leq \text{const.} \ a^m \sum_{k=m} \binom{k}{m} q^k.$$

This proves that (14) converges absolutely, and that the left side of (15) converges absolutely for $| \epsilon |$ small enough.

REFERENCES

1. Beneš, V. E., Heuristic Remarks and Mathematical Problems Regarding the Theory of Connecting Systems, B.S.T.J., **41**, July, 1962, pp. 1201–1247.
2. Beneš, V. E., Algebraic and Topological Properties of Connecting Networks, B.S.T.J., **41**, July, 1962, pp. 1249–1274.
3. Beneš, V. E., A "Thermodynamic" Theory of Traffic in Connecting Networks, B.S.T.J., **42**, May, 1963, pp. 567–607.
4. Feller, W., *An Introduction to Probability Theory and its Applications*, **1**, 2nd Ed., John Wiley & Sons, New York, 1957.
5. Bellman, R., *Introduction to Matrix Analysis*, McGraw-Hill Book Co., Inc., New York, 1960.
6. Halmos, P. R., *Finite Dimensional Vector Spaces*, 2nd Ed., D. Van Nostrand Company, Inc., Princeton, 1958.
7. Morris, R., and Wolman, E., A Note on 'Statistical Equilibrium,' Operations Research, **9**, 1961, pp. 751–753.
8. Syski, R., *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, London, 1960.
9. Beneš, V. E., Theoretical Problems of Optimal Routing in Communications Networks, to appear.
10. Birkhoff, G., and MacLane, S., *A Survey of Modern Algebra*, The Macmillan Company, New York, 1950.
11. Smith, W. L., Renewal Theory and its Ramifications, J. Roy. Stat. Soc., Ser. B., **20**, 1958, pp. 243–302.
12. Derman, C., A Solution to a Set of Fundamental Equations in Markov Chains, Proc. Amer. Math. Soc., **5**, 1954, pp. 332–334.
13. Riordan, J., *An Introduction to Combinatorial Analysis*, John Wiley & Sons, New York, 1958.

Spectral Characteristics of Digit-Simulating Speech Sounds

By D. P. BORENSTEIN

(Manuscript received July 11, 1963)

A spectral analysis has been performed on a number of spoken vowel sounds, in particular those sounds causing digit registration in a TOUCH-TONE receiver. The analysis, implemented by computer methods, provides a definitive picture of the nature of digit simulation in TOUCH-TONE calling.

I. INTRODUCTION

A digit simulation in TOUCH-TONE calling (Ref. 1, pp. 9–12, 15–16) is, by practical definition, a speech segment capable of causing digit registration in a TOUCH-TONE signaling system. Spectral analyses have been performed on a number of speech segments, each of which was selected solely on the basis of having the above property. Briefly, a valid TOUCH-TONE signal requires the simultaneous presence of two code frequencies for a certain minimum length of time, and with some minimum signal-to-noise ratio. It was therefore theoretically anticipated (Ref. 1, pp. 10–12) that each of these speech segments would be linked by two other common characteristics: (1) a frequency spectrum having two sharply dominant peaks, and (2) a high degree of periodicity for some minimal length of time.

There is good reason to believe that speech segments of this general nature are likely to be troublesome in any signaling system based on the transmission of voiceband tones over speech channels.

Due to the inherent rarity and relatively brief duration of the voice-produced digit simulation, some special procedures were required both in obtaining and analyzing these speech segments. The remainder of this article comprises a description of these procedures, followed by a presentation and discussion of the resulting spectral analyses.
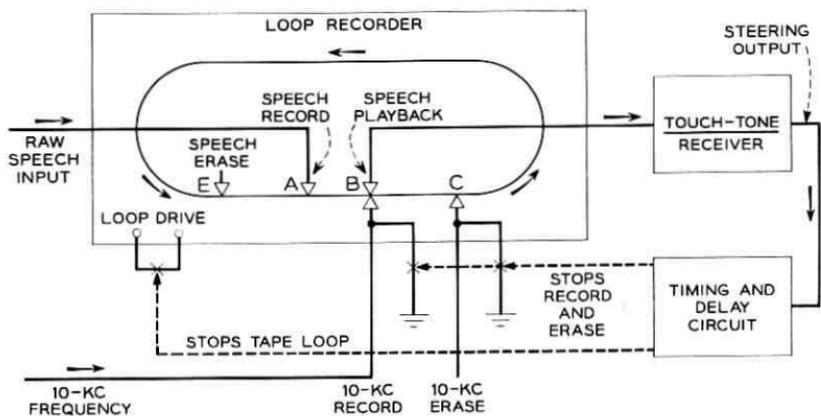
Fig. 1 — Apparatus for recording digit-simulating speech segments.

II. COLLECTING THE SPEECH SAMPLES

The digit-simulating speech segments were obtained by recording raw speech onto magnetic tape loops with the two-track recording arrangement shown in Fig. 1.

Using a 60-inch loop of tape at a speed of 15 in/sec, speech is continuously recorded at point A, played into a standard TOUCH-TONE receiver at point B, and, if there is no receiver output, erased at point E after traversal of the loop. Simultaneously, on a second track, a 10-kc pilot frequency is continuously recorded and erased at points B and C, respectively. If at any time there is a TOUCH-TONE receiver output, indicating the presence of a digit-simulating speech segment just past point B, the timing network is triggered. The timing network then performs two operations: (1) it disables the 10-kc record and erase after a delay of 35 ms, and (2) it stops the tape transport after a delay of 2 seconds (half the loop traversal time).

This process yields a 60-inch length of tape consisting of about 29 inches each of pre- and post-simulation speech plus a 1.5-inch (110 ms at 15 in/sec) segment which contains both the actual simulating speech sample and the 10-kc pilot frequency. In this manner, fourteen such samples were obtained, at the average rate of about one per ten hours of raw speech — an indication of the extreme rarity of simulation with the present TOUCH-TONE receiver.

III. ANALOG-TO-DIGITAL CONVERSION AND PRINTOUT

By means of encoding equipment developed by the Acoustics Research Department, the fourteen digit-simulating speech segments were con-

verted from analog form to an eleven-bit digital signal. The sampling rate of 10 kc was gated directly from the pilot track of the original analog tape, thus eliminating sources of error due to tape flutter during the original recording process. Once the digital tape was obtained, the conversion process was reversed to obtain an accurate X-Y recording of each of the fourteen speech waveforms. Visual inspection of these waveforms, two of which are shown in Fig. 2, confirms their periodic nature (the periodicity of the samples shown in Fig. 2 would be still more evident were it not for the fact that most speech fundamentals are considerably attenuated by telephone apparatus).

IV. SPECTRAL ANALYSIS

The fourteen speech samples, in eleven-bit digital format, were then subjected to a "pitch synchronous"[2] Fourier analysis on the IBM-7090 computer. The pitch synchronous analysis consisted essentially of a conventional Fourier analysis performed on each successive funda-mental pitch period in the speech sample. These pitch periods, in turn, were determined on the computer by counting the number of sampling intervals (each being 100 μsec) between successive maxima in the wave-form and then interpolating between samples for greater accuracy. This method of Fourier analysis is ideally suited to waveforms that maintain an almost-periodic structure over an appreciable length of time.

For each speech segment analyzed, the computer output consisted of a sequential set of bar graphs, one for each fundamental pitch period of the speech waveform. Each graph, in turn, is a plot of harmonic ampli-tude (the Euler coefficient) in db versus harmonic number. In addition, each graph gives the "instantaneous pitch" (i.e. the reciprocal of the period) of each fundamental period analyzed. Figs. 3 and 4 show the
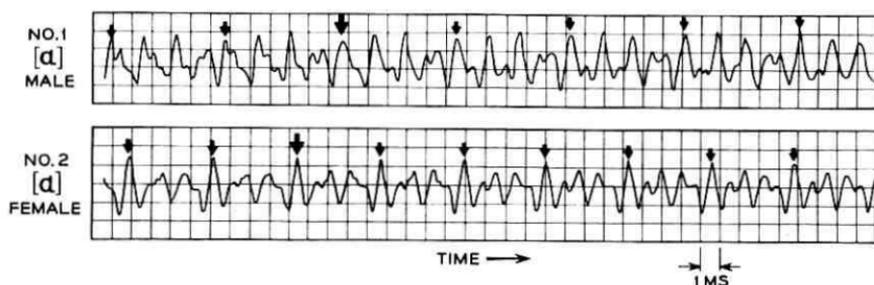
Fig. 2 — Analog waveforms of two digit-simulating speech segments. Also shown are the sex of each speaker and the particular phoneme causing the simulation. Fourier spectra of digit simulations 1 and 2 are shown in Figs. 3 and 4, respectively. Arrows indicate periodicity, with the large arrow showing the approximate start of the digital simulation.
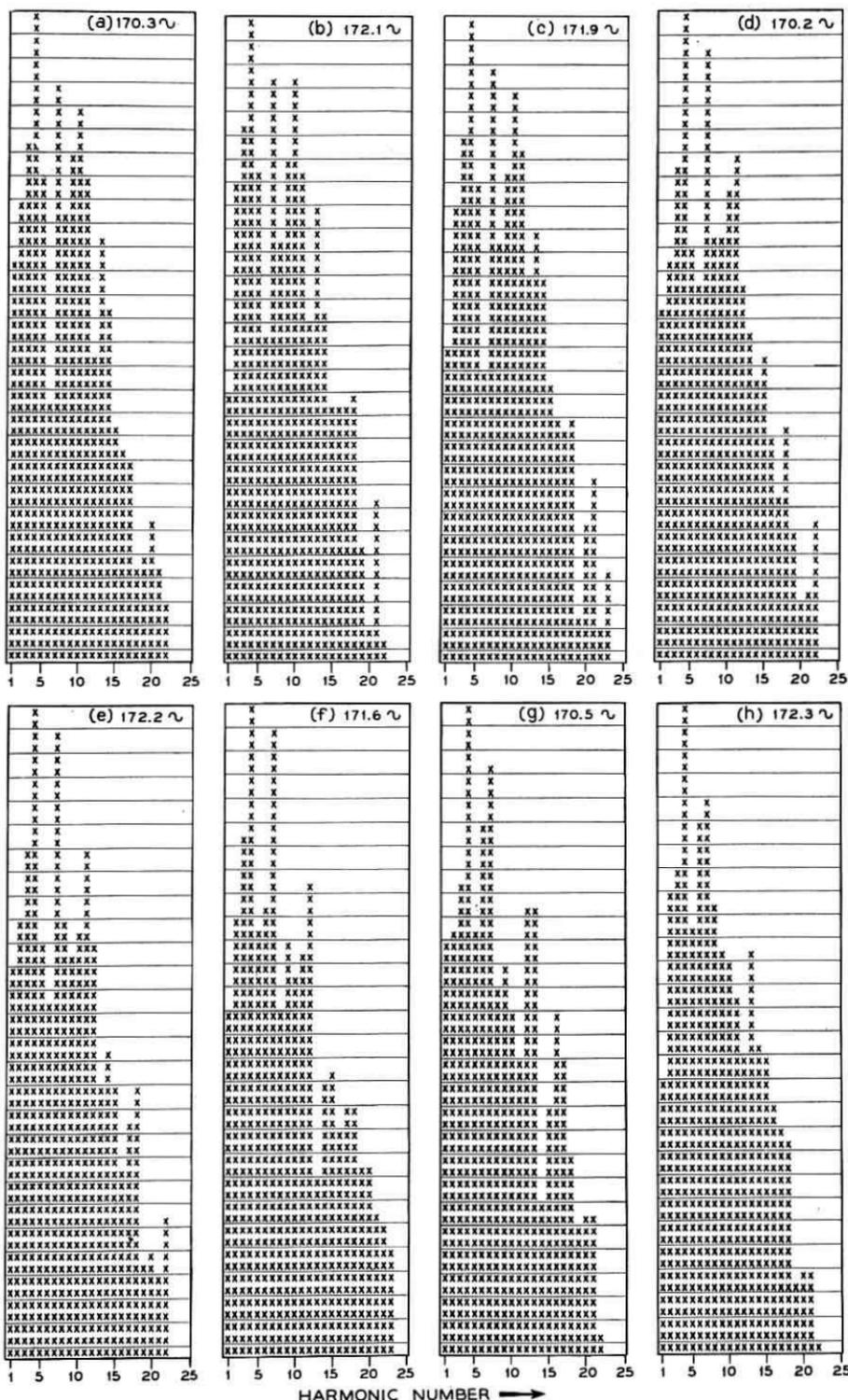
Fig. 3 — Set of Fourier spectra for digit simulation No. 1 (as shown in Fig. 2). Each X represents a 1-db relative amplitude increment. Spectra are in alphabetical order with respect to time.
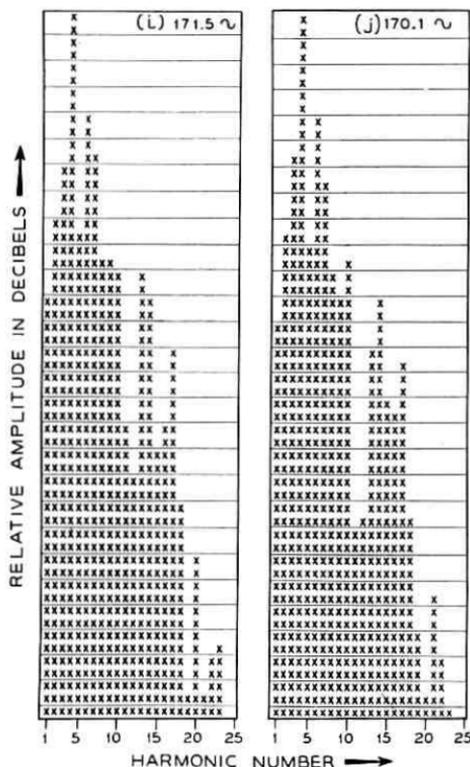
Fig. 3 — (continued)

two sets of spectra corresponding to the two speech segments whose time domain waveforms appear in Fig. 2.

V. DISCUSSION OF RESULTS

Several aspects of the spectra shown in Figs. 3 and 4 are worthy of note.

First of all, it is seen that these two speech segments (as well as the twelve others not shown here) do indeed satisfy the two properties anticipated in the introduction. The high degree of periodicity of these speech waveforms is spectrally confirmed by noting that in both sequences of spectra the harmonic structure remains extraordinarily uniform. (This result also confirms, by hindsight, the original validity of a period-by-period Fourier analysis.) By noting the fundamental pitch (thus the period) of each segment, it is seen that this highly stable harmonic structure is maintained for at least the 23 milliseconds which coincides with the duration requirements of the TOUCH-TONE receiver.
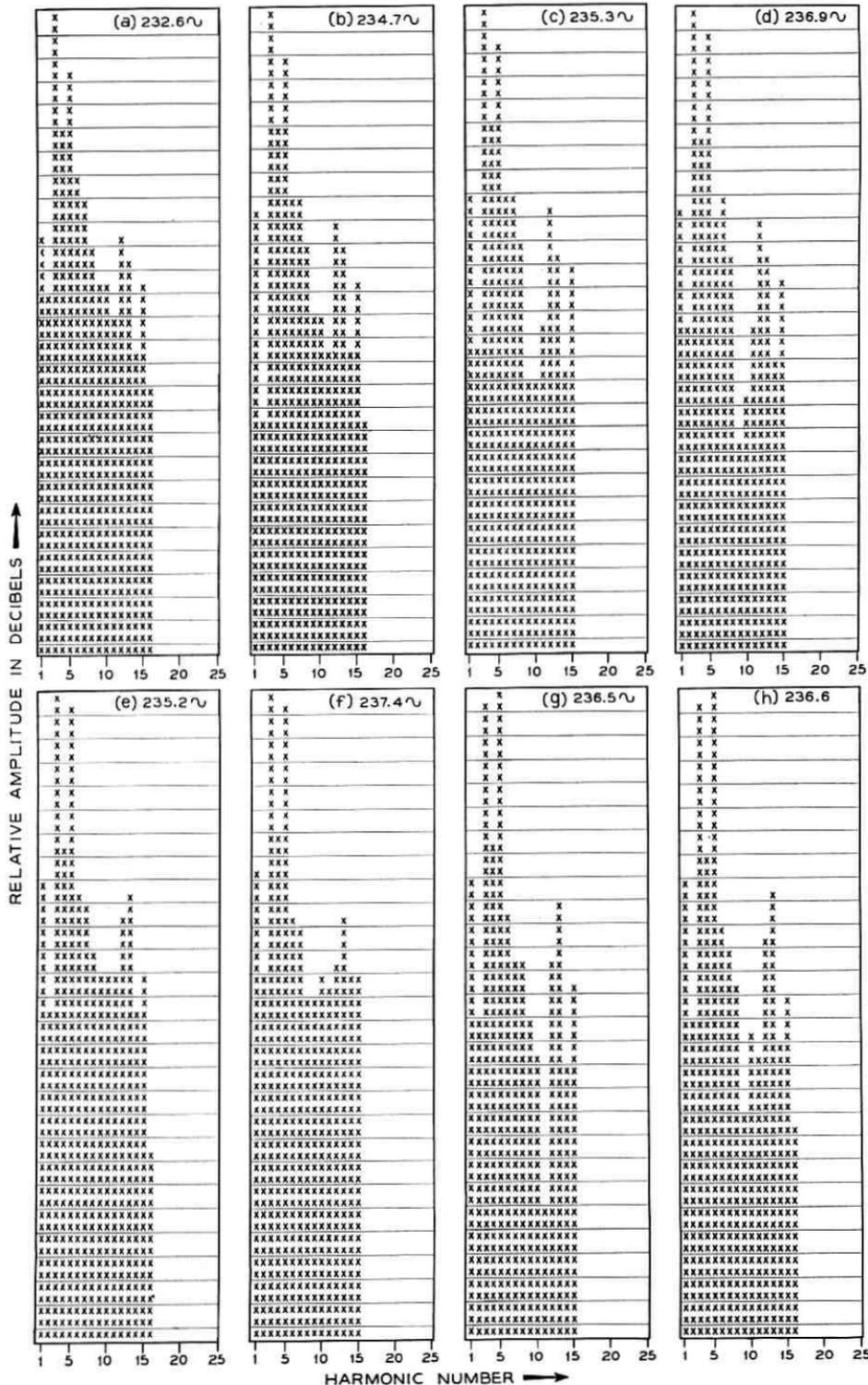
Fig. 4 — Set of Fourier spectra for digit simulation No. 2 (as shown in Fig. 2).

Fig. 4 — (continued)

2845
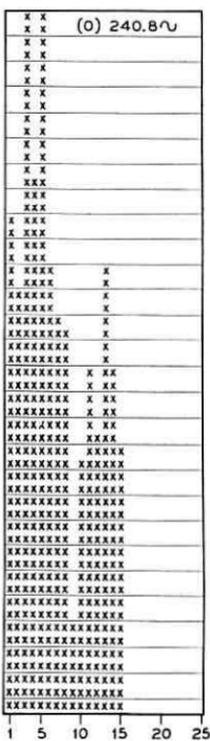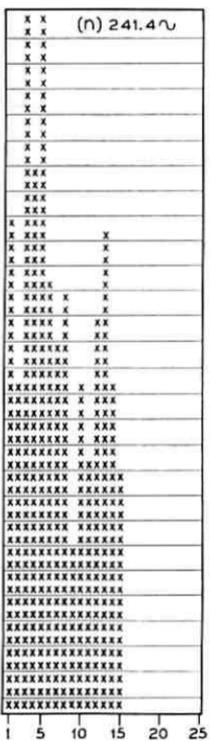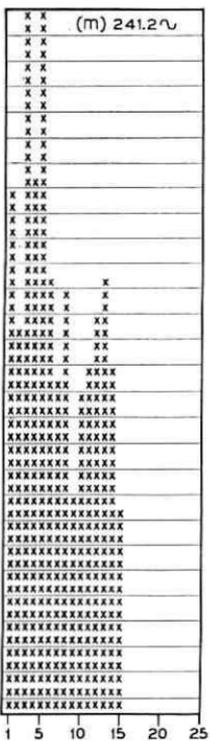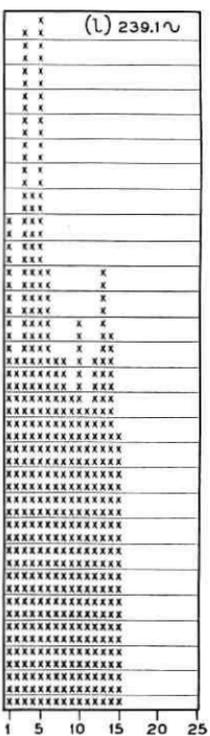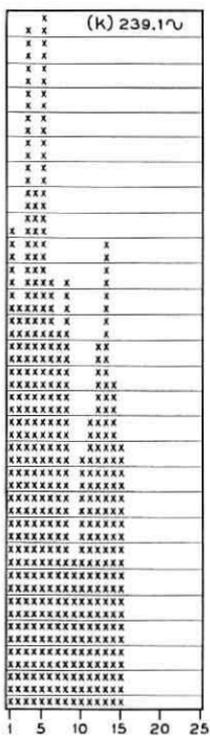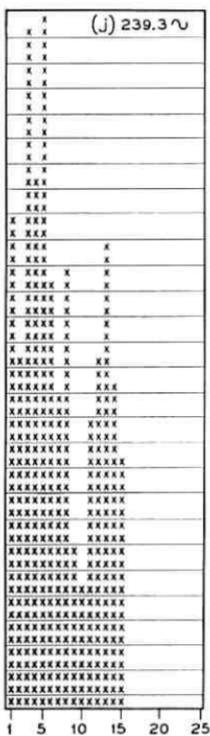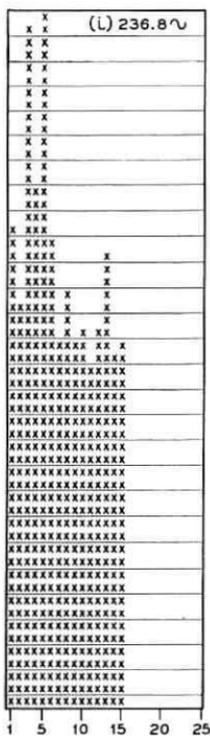
Secondly, one finds immediate justification for the fact that these speech segments caused digit simulation. By multiplying the fundamental pitch of any segment by the orders of its two dominant harmonics, a valid TOUCH-TONE calling signal* is derived. Thus a voice-produced digit simulation is spectrally analogous to a valid TOUCH-TONE signal accompanied by noise, with the sole exception that in the former case both the "noise" and signal components are integral multiples of a discrete fundamental frequency. Indeed, this sole distinction between a digit simulation and a valid signal might possibly be used to provide further simulation protection in future voice-frequency signaling applications. Specifically, a receiver might be designed to be sensitive to the presence of selected harmonics and/or sub-harmonics of valid signal frequencies, and thereby to reject many speech phonemes which would ordinarily cause simulation.

In the portions of Figs. 3 and 4 where the harmonic structure is noticeably changing with time (namely at the beginning and end of each series of spectra) pitch-synchronous Fourier analysis can be regarded as only an approximation of spectral density. For some applications, however, the approximation is still useful. In the first place, one can obtain a practical "feel" for the rate of change of pitch and harmonic structure in vowel-type speech sounds. Also, from the standpoint of digit simulation, by examining the spectra one can ascertain just how and when a speech segment becomes a digit simulation. For example, in the early spectra of Fig. 3, although pitch requirements for digit simulation are satisfied, the 10th harmonic competes with the 7th harmonic for limiter capture, and receiver recognition is prevented by limiter guard action (i.e., insufficient signal-to-noise ratio). (See Ref. 1, pp. 10–11, 13.)

On the other hand, although early spectra of Fig. 4 show an acceptable harmonic structure for digit simulation, the pitch is slightly too low for receiver recognition. In a similar manner, one can determine how and when a digit simulating wave-form starts to degenerate.

Admittedly, the speech segments chosen here are both rare and few in number. Thus, one cannot draw conclusions of statistical significance from this study. However, there is no reason to believe that any other group of frequencies of the same capacity in the voice band would not be simulated by the voice about as often as were these TOUCH-TONE calling frequencies. Therefore, such vowel-type speech segments may be

---

* A valid TOUCH-TONE calling signal consists of one frequency from each of two groups: a low group — 697, 770, 852 and 941 cps ±2.5 per cent — and a high group — 1209, 1336, 1477 and 1633 cps ±2.5 per cent.

looked upon as potential digit simulations in almost any proposed voice-frequency signaling application.

REFERENCES

1. Battista, R. N., Morrison, C. G., and Nash, D. H., Signaling System and Receiver for TOUCH-TONE Calling, Trans. IEEE, **65**, March, 1963.
2. Mathews, M. V., Miller, Joan E., and David, E. E., Pitch Synchronous Analysis of Voiced Sounds, J. Acoust. Soc. Am., **33**, Feb., 1961, pp. 179–186.

Reflections from an Exponential Atmosphere

By K. BULLINGTON

A quantitative explanation of tropospheric radio propagation is derived without the use of arbitrary numerical factors. It is based primarily on the average value and the standard deviation of the index of refraction at the earth's surface, both of which decrease exponentially with height.

This method seems to bridge the gap between the internal reflection and scatter hypotheses. The quantitative results are in good agreement with experimental data on distance dependence, frequency dependence, climatic and seasonal variations, and effective antenna gain. One consequence is a new concept of the decrease in effective gain of narrow-beam antennas, which accounts for some hitherto unexplained experimental results.

I. INTRODUCTION

Several conflicting hypotheses have been proposed over the years in an attempt to explain the transhorizon radio field intensities, which in some cases are hundreds of decibels stronger than predicted by smooth earth diffraction theory.[1-6] The rapid fading ordinarily associated with this type of transmission led naturally to the concept of many scattered components with random phase angles.

In order to account for the average field intensity, frequency dependence and other characteristics, the scatter hypotheses must make a number of more or less arbitrary assumptions regarding the number, size and distribution of the irregularities and whether the scatterers are spherical "blobs" or flat "pancakes." The scatter concept is a statistical framework that can be adjusted by arbitrary parameters to fit almost any consistent experimental data and, in fact, the values of the parameters have been changed significantly when the need to accommodate unexpected experimental results has arisen.

On the other hand, Snell's law of reflection indicates that reflections must occur from the gradual decrease in dielectric permittivity with height. This concept focuses attention on the average received power but,

of course, it is recognized that the fading must be associated with the motion of the irregularities in the atmosphere. The object of the internal reflection hypothesis is to relate the average received field to the rate of decrease in atmospheric density, and hence to eliminate some of the arbitrary parameters required by the scattering concepts. Many theorists have concluded that the internal reflections cancel, or at least that their effect is too small to be important.

If the problem of inhomogeneous atmosphere could be solved rigorously, there would be no difference of opinion, but a rigorous solution that can be readily evaluated has not yet been achieved without significant approximations, and the relative accuracy of the different approximations is in question.

The purpose of this paper is to present a new approach that to the present author seems to bridge the gap between the internal reflection concept and the scatter hypotheses. Although the method is not as rigorous as would be desired, it does provide quantitative predictions that are in excellent agreement with experimental data without the use of arbitrary numerical factors.

II. METHOD

The procedure is to find first the magnitude of the internal reflection from a smooth exponential atmosphere and then to superimpose random variations on the exponential decrease.

The difference in dielectric permittivity ϵ between two horizontal planes at heights y_1 and y_2 is assumed to be a random variable with a Gaussian distribution. The average value depends on the average index of refraction at the surface, together with the exponential decrease in atmospheric density with height; the variance at any height depends on the variance in the surface index of refraction together with the same exponential decrease with height. The assumed rate of exponential decrease or scale factor cuts in half the atmospheric density, mean value of permittivity and variance in permittivity for each increase in elevation of 7 or 8 kilometers. These quantities are based on meteorological data and are independent of transhorizon radio experiments.

It is further assumed that the reflection coefficient is proportional to the change in dielectric permittivity and hence is also Gaussian distributed. The average reflection coefficient is derived from electromagnetic theory as a function of the average exponential decrease in permittivity.

Finally, it is assumed that the variance of the reflection coefficient is related to the variance of the dielectric permittivity by the same function that relates the average reflection coefficient to the average permittivity.

The resulting reflection coefficient is Gaussian distributed and has both positive and negative values, which are almost equally probable when the standard derivation σ is large compared with the mean value μ. For comparison with experimental data it is necessary to find the mean of a Gaussian distribution taken without regard to sign, and this is given with sufficient accuracy by either the mean value μ or $\sqrt{2/\pi}\sigma$, whichever is the larger. The former is usually controlling below 100 mc and the latter is usually controlling at UHF and above.

This paper is not intended to be a complete description of all tropospheric phenomena. Some variations exist in the horizontal plane but they are small compared with the variations in the vertical plane and have been neglected in order to simplify the presentation. The resulting modified Gaussian distribution is similar to a Rayleigh distribution; consequently, this derivation has a fading component, but it does not give quantitative results on either the speed of fading or on the associated problem of useful bandwidth. There is a hint that the speed of fading depends on the number of Fresnel zones in the common volume and that the useful bandwidth varies inversely as this quantity, but further work is needed.

In addition, this paper is concerned only with the effects of the atmosphere that have been neglected in the smooth sphere diffraction theory. The latter will be controlling at short distances and low frequencies.

III. EFFECT OF SMOOTH EXPONENTIAL PROFILE

The permittivity of the atmosphere is assumed to vary only in the vertical direction, and to have a mean value and a variance, both of which decrease exponentially with height. Specifically, the mean value of dielectric permittivity ϵ at any height illustrated on Fig. 1 is assumed to be

$$\epsilon = \epsilon_s + (1 - \epsilon_s)[1 - e^{-b(y-y_s)}]. \tag{1}$$

It follows that the change in dielectric permittivity in the region between $y = y_1$ and $y = y_2$ is given by

$$\Delta\epsilon_{2,1} = \epsilon_2 - \epsilon_1 = (1 - \epsilon_s)e^{-b(y_1-y_s)}[1 - e^{-b(y_2-y_1)}] \tag{2}$$
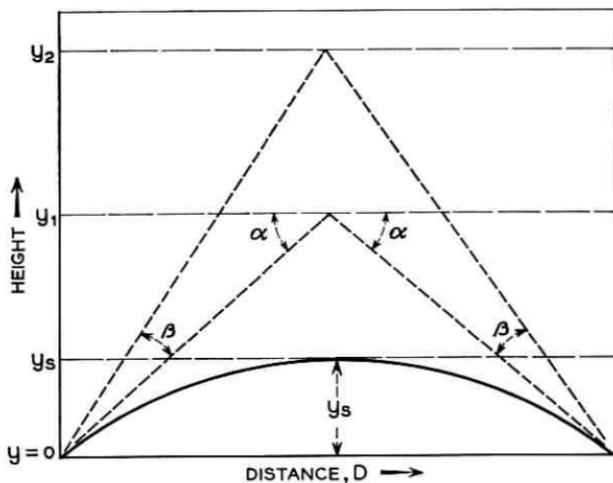
Fig. 1 — Geometry of transhorizon path.

where ϵ_s = dielectric permittivity at the surface $(y = y_s)$

b = rate of exponential decrease

≈ 0.14 per kilometer.

The parameter b is primarily a measure of the rate of decrease in atmospheric density, which is in turn determined by the laws of gas diffusion and the law of gravity. The average value of b varies from about $b = 0.12$ per km near the earth's surface to about $b = 0.16$ per km at elevations greater than 7 or 8 kilometers, but it can be considered a constant in this application.*

---

* The parameter b is related as follows to the widely used effective earth radius factor k (which is limited to regions where $1 - e^{-by} \approx by$)

$$k = \frac{1}{1 + \dfrac{a}{2}\dfrac{d\epsilon}{dy}} = \frac{1}{1 - 2b\left(\dfrac{N_s}{314}\right)}$$

where

$$N_s = \frac{(\epsilon_s - 1)}{2} \times 10^6 = (n_s - 1)\,10^6$$

n = index of refraction

a = true earth radius = 6370 km

For the typical case of $N_s = 314$ and $b = 0.125$/km, the effective earth radius factor is $k = 4/3$.

When a change in dielectric permittivity occurs at a sharp boundary as shown on Fig. 2(a), the reflection coefficient q for either polarization can be derived from Snell's law with the following result

$$q = \frac{\epsilon_2 - \epsilon_1}{4 \sin^2 \alpha} \tag{3}$$

or

$$4\alpha^2 q \approx \epsilon_2 - \epsilon_1$$

where α is the angle between the ray and the boundary. This result assumes $|\epsilon_2 - \epsilon_1| < 4 \sin^2 \alpha$ and $\sin \alpha \approx \alpha$. The corresponding reflection coefficient for an exponential decrease illustrated in Fig. 2(b) is derived in a later section with the following result

$$4\alpha^2 q = (\epsilon_2 - \epsilon_1) \frac{b}{b + i2A} F_1 \tag{4}$$

where

$$A = \frac{d\phi}{dy} = \frac{2\pi\alpha}{\lambda} \gg b$$

$$\phi = \frac{2\pi y \alpha}{\lambda}$$

$$\lambda = \text{wavelength}$$

$$F_1 = \left[\frac{1 - \exp\left[-(b + i2A)(y_2 - y_1)\right]}{1 - \exp\left[-b(y_2 - y_1)\right]} \right]$$

$$= \left[\frac{1 - \exp\left(\dfrac{-bD\beta}{2}\right) \exp\left(-i2M\pi\right)}{1 - \exp\left(\dfrac{-bD\beta}{2}\right)} \right]$$

D = path distance

β = antenna beamwidth $\approx \dfrac{2}{D}(y_2 - y_1)$

M = number of pairs of Fresnel zones in the common volume formed by the intersection of the two antenna beams

$$= \frac{2\alpha}{\lambda}(y_2 - y_1) = \alpha\beta \frac{D}{\lambda}.$$
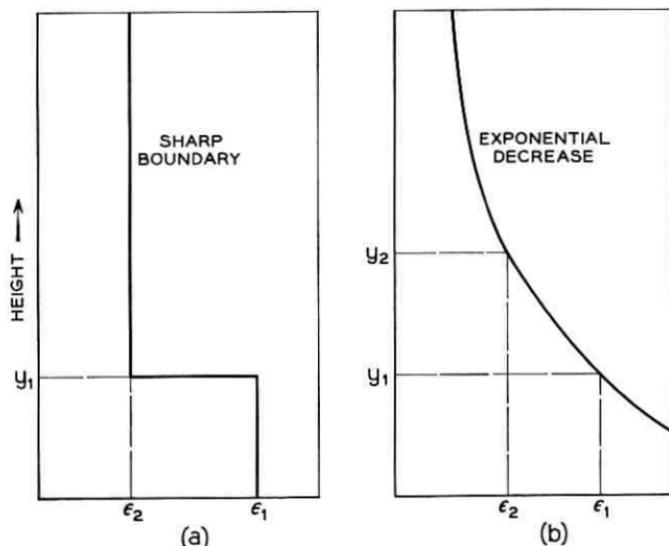
Fig. 2 — Idealized variation in dielectric permittivity.

The factor F_1 is unity whenever M is an integer; this is the minimum value and this is the case of interest.[*] The assumption of fractional Fresnel zones would lead to $| F_1 | > 1$, but realizable antenna patterns cannot cut off sharply enough to make this assumption meaningful. Nonexponential models of the atmosphere lead to factors analogous to F_1 which are either zero or indeterminate, and presumably this is the basis for the widespread assumption that internal reflections cancel.

Substitution of (2) into (4) leads to the following expression for the reflection coefficient of a smooth exponential atmosphere

$$
q = \left[\frac{(\epsilon_s - 1)}{4\alpha^2} e^{-bD\alpha/4} \right]
$$

$$
\cdot \left[\frac{b\lambda}{4\pi\alpha} \right] [1 - e^{-bD\beta/2}] F_1 \left[\frac{\exp\left[-b\left(\frac{y_1}{2} - y_s \right) \right]}{1 - \frac{ib\lambda}{4\pi\alpha}} \right]. \tag{5}
$$

---

[*] The use of pairs of Fresnel zones is an artifice introduced by Fresnel to solve related problems in optics. By this means, a difficult vector addition problem is changed into a much easier scalar one because the phase shift in each pair is exactly 360° by definition. A small percentage change in the number of pairs of zones cannot change the sum significantly, so the exact number of pairs is not critical. If, however, the summation is taken over a region that cuts off sharply at an odd or fractional zone, the sum depends critically on the exact nature of the cutoff. In this case, the magnitude of the vector sum oscillates over a relatively wide range and mathematically the summation is very difficult or indeterminate.

The term in the first brackets represents the reflection coefficient that would have been expected if all the decrease in dielectric permittivity had occurred at a sharp boundary at height $y_1 = D\alpha/2$. The term in the second brackets is the effect of an exponential decrease for wide-beam antennas, (as long as this factor is less than unity). The third brackets show the effect of narrow-beam antennas, which depends on distance as well as beamwidth. As discussed above, the factor F_1 can be set to unity. The term in the right-hand brackets can also be neglected for a relatively smooth earth path, because $y_1 \approx 2y_s$ for this case.

Typical values of the reflection coefficient q computed from (5) are shown in Table I for wide-beam antennas; that is, only the first two bracketed terms are used.

The estimated losses given in Table I are much closer to the measured results than are those obtained from smooth earth diffraction theory but they are still 10–30 db greater than the median measured data. In addition, (5) is incomplete because it is not quite adequate in other respects.

TABLE I

Distance beyond Line of Sight		α	20 log | q |, decibels		
km	miles		300 mc	1000 mc	3000 mc
100	62	0.45°	−49	−59	−69
200	124	0.9°	−68	−78	−88
500	310	2.25°	−97	−107	−117
1000	620	4.5°	−133	−143	−153

The experimental results on the variation of the received voltage with frequency seem, on the average, to be closer to $\sqrt{\lambda}$ rather than λ as shown in (5). Seasonal and climatic effects are known to be significant, but the only possible variable in (5) is ($\epsilon_s - 1$), and the seasonal changes in ($\epsilon_s - 1$) are too small to account for the observed results. Finally, the fixed exponential atmosphere assumed in (5) cannot hope to explain the fading phenomena unless the variations in the atmosphere are considered.

IV. EFFECT OF ATMOSPHERIC VARIATIONS

Random fluctuations in atmospheric pressure, temperature and humidity are superimposed on the average value. The resulting variations in the index of refraction ($n = \sqrt{\epsilon}$) at the earth's surface have been measured in many experiments, but perhaps the most complete and reliable results can be obtained from long-term weather data.[7-10]

An analysis of meteorological data for 8 years from 45 U.S. weather stations indicates that the standard deviation of the surface index of refraction has a long-term average in the vicinity of Washington, D.C. of 10–15 N units, which corresponds to a variation in ϵ_s of $\sigma(\epsilon_s) \approx 20$ to 30×10^{-6}. In the northern part of the U.S., $\sigma(\epsilon_s)$ may be as low as 4×10^{-6} (2 N units) and along the Gulf and in Florida it may be as high as 52×10^{-6} (26 N units) in certain hours and seasons. The variance at the earth's surface can be defined as

$$\text{variance } (\epsilon_s) = [\sigma(\epsilon_s)]^2 = [\delta(\epsilon_s - 1)]^2. \tag{6}$$

Since $(\epsilon_s - 1)$ is of the order of 600×10^{-6}, the average value of 10 to 15 N units corresponds to $\delta = 3$ to 5 per cent at the surface.

At very great heights the variance of ϵ must approach zero. If the dielectric permittivity followed a smooth exponential decrease and ϵ_s were the only variable, the standard deviation σ would decrease at the same rate as the mean value given in (2), but this does not agree with the experiment. Experimental data based on meteorological measurements indicate that the standard deviation σ decreases more slowly than the mean value.[10] A better fit with experimental data can be obtained by assuming that the variance (σ^2) and not the standard deviation σ decreases at the same rate as the mean value. Quantitatively this means that the standard deviation of the dielectric permittivity, which is measured to be 3 to 5 per cent of the mean value at the surface, is taken to be about 8 per cent of the mean value at an elevation of 14 kilometers, 16 per cent at 28 kilometers, 32 per cent at 42 kilometers, etc. At higher elevations this model breaks down because the standard deviation cannot be greater than the mean value in a nonionized medium. Heights greater than 50 kilometers are approaching the ionosphere and, hence, are outside the usual range of interest for tropospheric transmission. On this basis, the variance at y_1 over the interval $y_2 - y_1$ is given by

$$\text{variance } (\epsilon_2 - \epsilon_1) = [\delta(\epsilon_s - 1)]^2 e^{-b(y_1 - y_s)} [1 - e^{-b(y_2 - y_1)}] \tag{7}$$

or

$$\sigma(\epsilon_{2,1}) = \delta(\epsilon_s - 1) \exp\left(\frac{-b(y_1 - y_s)}{2}\right) \sqrt{1 - e^{-b(y_2 - y_1)}}.$$

The dielectric permittivity of the atmosphere, as used in this paper, has now been specified to be uniform in the horizontal plane and to have an exponential variation in the vertical coordinate with a mean value given by (2) and a variance given by (7). The mean value for the re-

flection coefficient corresponding to (2) has been given in (4) and (5). Inasmuch as the irregularities superimposed on the exponential decrease cause the variance of the dielectric permittivity to vary as its mean, it is assumed that the variance of the reflection coefficient is also proportional to its mean value. This critical assumption leads to

$$\text{variance } [4\alpha^2 q] = \text{variance } (\epsilon_2 - \epsilon_1) \frac{b}{b + i2A} F_1$$

which by means of (7) leads to

$$\sigma(q) = \left[\frac{\delta(\epsilon_s - 1)}{4\alpha^2} e^{-bD\alpha/8} \right] \sqrt{\frac{b\lambda}{4\pi\alpha}} \sqrt{1 - e^{-bD\beta/2}} \sqrt{F_1}$$

$$\sqrt{\frac{\exp\left[-b\left(\dfrac{y_1}{2} - y_s \right) \right]}{1 - \dfrac{ib\lambda}{4\pi\alpha}}}. \qquad (8)$$

The interpretation of (8) is essentially the same as for (5). The first two terms to the right of the equality sign are the important ones for wide-angle antennas, the third of the five factors is the correction for narrow-beam antennas and the two factors on the right are unity in most applications and hence can ordinarily be neglected.

When the standard deviation (8) is comparable to or larger than the mean value (5), the resulting distribution of reflection coefficients contains both positive and negative values. The significance of a change in sign is to introduce an uncertainty in phase delay of one half cycle, which is unmeasurable and unimportant in most radio applications and tests. In order to compare with experimental data it is necessary to "fold over" the negative part of the distribution, add it to the positive part and then to find the mean value of the resulting distribution.

The mean value of the folded distribution approaches the mean value μ of the Gaussian distribution when $\sigma \ll \mu$ and approaches $\sqrt{2/\pi}\sigma$ when $\sigma \gg \mu$. More precisely, the mean value of the folded distribution is given by

$$\mu'(\text{folded distribution}) = \sqrt{\frac{2}{\pi}} \sigma e^{-\mu^2/2\sigma^2} + \mu \operatorname{erf}\left(\frac{\mu}{\sigma\sqrt{2}} \right)$$

where μ and σ refer to the original Gaussian distribution.

The folded distribution indicates that fading is to be expected, at least on a diurnal and seasonal basis. The use of meteorological data taken at hourly intervals does not provide quantitative information on the fast

fading, but it does not preclude it. Moreover, the fading can be larger and at a faster rate than indicated by the variance in the dielectric permittivity taken over the entire vertical height of the common volume. The fading depends on the variations within each pair of Fresnel zones rather than on the standard deviation of the entire region. The standard deviation of the variation within each pair of zones is of the order of $\sqrt{M}$ times the standard deviation of their sum, where M is the number of pairs of Fresnel zones.

In addition to the reflection coefficient of the atmosphere, the transmission loss between two ground-based antennas relative to free space is also affected by ground reflections near each terminal. Ground reflections introduce image antennas which in effect double the amplitude of the received field intensity. This factor of 2 (or more precisely $1 + |r|$, where r is the reflection coefficient of the ground) appears in plane-earth, smooth-sphere and scatter theories, and it is unimportant numerically whether it is considered as twice the free-space field incident on the atmosphere with no effect near the receiving terminal or considered to be a power addition of the antenna and its image at both terminals.

Consequently, the mean value of transmission loss (relative to free space) to be compared with experimental data is the larger of $2q$ from (5) or $2\sqrt{2/\pi}\sigma$ from (8). At frequencies above 300 mc the latter is controlling, and typical values are shown in Table II for $\delta(\epsilon_s - 1) = 25 \times 10^{-6}$, which corresponds to a standard deviation at the earth's surface of 12.5 N units.

For comparison with the scatter hypothesis, the first two terms in (8) lead to the following ratio for the average received power relative to free space for wide-beam antennas

$$\frac{P}{P_0} = [2\sqrt{2/\pi}\sigma]^2 = \frac{b\lambda[\sigma(\epsilon_s)]^2 e^{-bD\alpha/4}}{8\pi^2\alpha^5}. \qquad (9)$$

V. INTERPRETATION OF RESULTS

The computed transmission losses shown in the above table for wide-beam antennas are compared on Fig. 3 with the average experimental

TABLE II

Distance		α	Transmission Loss Relative to Free Space, decibels		
km	miles		300 mc	1000 mc	3000 mc
100	62	0.45°	44	49	54
200	124	0.9°	60	65	70
500	310	2.25°	82	87	92
1000	620	4.5°	106	111	116
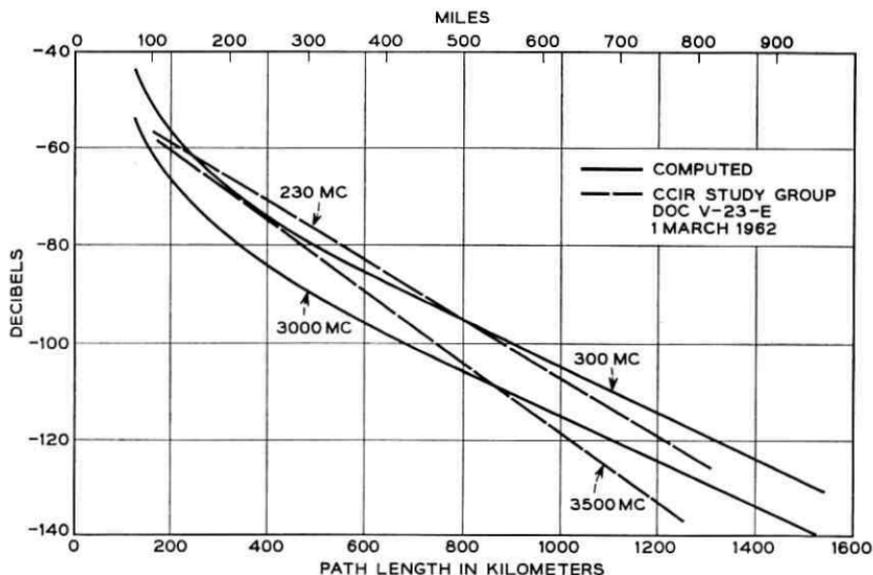
Fig. 3 — Transhorizon transmission loss for widebeam antenna.

data as given in a recent CCIR Study Group report. Since the computed values are based on the distance between the two horizons, while the experimental data are based on total path length, the computed values have been plotted on Fig. 3 at a distance of 30 km greater than given in the above table in order to allow for the line-of-sight distance between each terminal and its horizon. This assumption corresponds to an antenna height of about 50 feet over smooth earth. Any other reasonable assumption is not likely to change the comparison significantly.

The computed values are somewhat lower than the average experimental data as given in the CCIR Study Group Report, but at least part of the difference (in decibels) may be due to the assumed value of antenna to medium coupling loss used in the analysis of the original experimetal data. As shown in the following paragraph, the effect of narrow beam antennas derived herein is ordinarily less than obtained from the method used in the CCIR Study Group Report.[11,12]

The antenna degradation factor given in (8) is

$$\sqrt{1 - e^{-bD\beta/2}} = \sqrt{1 - e^{-b(y_2 - y_1)}},$$

and is evaluated in Table III.

The db loss shown in the second column of Table III depends on the height of the common volume ($y_2 - y_1$). When the path is symmetrical and the two antennas have equal beamwidth, the corresponding path length for a given loss and beamwidth is shown by the right-hand col-

TABLE III

$(y_2 - y_1)$	Decrease in Effective Antenna Gain (Both Antennas)	Distances for Various Beamwidths		
		$\beta = 2°$	$\beta = 1°$	$\beta = 0.5°$
300 meters	13.9 db	17 km	34 km	69 km
1000	8.9	58	115	230
3000	4.7	172	345	690
10,000	1.2	575	1150	2300
20,000	0.27	1150	—	—

umns. These results are compared on Fig. 4 with typical values of the antenna-to-medium coupling based on scatter concepts for a 1° antenna. Although most of the experimental data are near the crossover region, the opposing trends provide a good basis for a critical experiment.

A previously unexplained experimental result at 400 mc on a 1000-km path is that antennas up to 120 feet in diameter (beamwidths greater than $1\frac{1}{4}°$) showed less than 1 db loss in effective antenna gain when measured at one terminal only.[13] This result agrees with the present
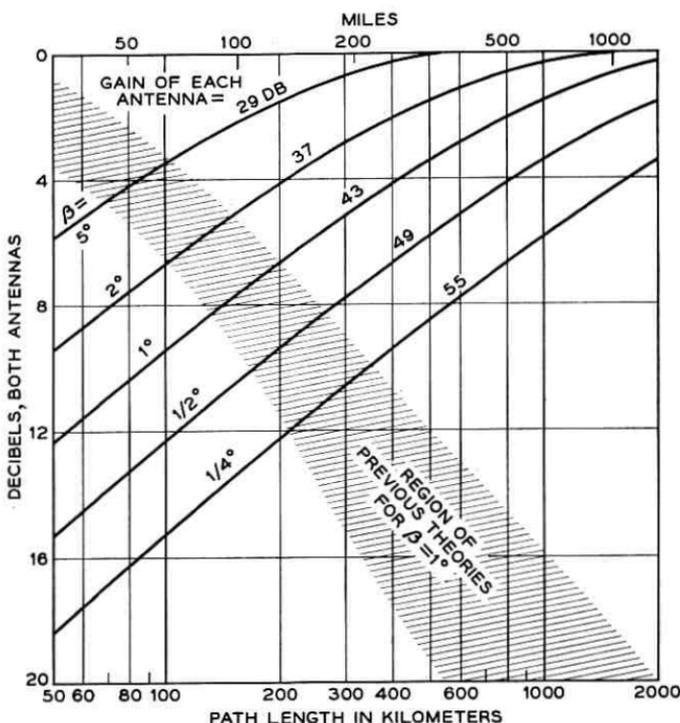
Fig. 4 — Loss in effective antenna gain.

method but is in sharp disagreement with earlier predictions. At the other extreme, measurements at 9000 mc on a 46-mile path with 4-foot and 1-foot antennas (about 1.6° and 6.6°) showed a loss of about 9 db when a change in antenna was made at both terminals.[14] The data on Fig. 4 would indicate no more than 6 db for a 4-to-1 change at short distances, but is closer to the measured result than the earlier methods which indicate 2 to 4 db.

Long-term comparisons between 8-foot and 60-foot antennas at 4000 mc (1.9° and 0.25°) on a 171-mile path indicated about 12 db loss in antenna effectiveness.[15] On the average the signal received on the 60-foot antenna was 5.7 db greater than on the 8-foot antenna, rather than 17.5 db, as would be expected in free space. The resulting loss of 11.8 db compares with an expected value of 10 log 6 = 7.8 db, since the path profile given in the reference indicates that the height of the common volume was reduced by a factor of 6.

The conclusion on the loss in effective antenna gain is that although the data on Fig. 4 seem a few db too low, the general shape of decreasing loss with increasing distance is to be preferred to the opposite trend derived from scatter theory.

It will be noted that the voltage ratio given in (8) varies as $\sqrt{\lambda}$, which is in general agreement with experiment, and that variations in $\delta(\epsilon_s - 1)$ can be 20 db or more with variations in climate, season and time of day. At UHF and above, the standard deviation (8) is large compared with the mean value (5), so considerable variation with time is to be expected; in the lower VHF and below, the standard deviation for the usual distances is less than the mean value, and this is consistent with the reduced amplitude of fading at the lower frequencies.

The remainder of this paper outlines the derivation of the reflection coefficient of a smooth exponential atmosphere, which was stated without proof in (4) and (5).

VI. DERIVATION OF REFLECTION FROM A SMOOTH EXPONENTIAL ATMOS-PHERE

Some reflections must occur whenever the dielectric permittivity changes. When the change occurs at a sharp boundary, the reflection coefficient is proportional to the difference in dielectric permittivity; when the dielectric permittivity is constant with height, the reflection is zero. The problem is to find the reflection coefficient for the intermediate case of a gradual change in permittivity.

A useful formulation of this problem was published by Schelkunoff in 1951.[16] The geometry and three possible variations are indicated on Fig
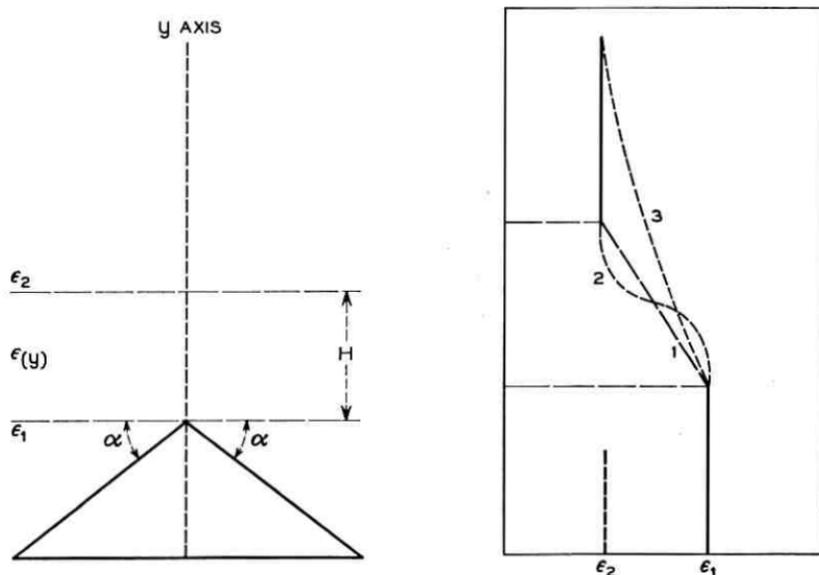
Fig. 5 — Idealized geometry for reflection from stratified media.

5. It is assumed that the region in which the dielectric permittivity is changing is sandwiched between two uniform media of dielectric permittivity ϵ_1 and ϵ_2, respectively. The reflection coefficient q of the stratified medium is derived from wave equations with the following result

$$q = \frac{1}{2} \int_0^\theta \frac{d \log K(\phi)}{d\phi} e^{-i2\phi} \, d\phi$$

$$= \frac{1}{2} \int_0^\theta \frac{1}{K} \frac{dK}{d\phi} e^{-i2\phi} \, d\phi$$

where

$$K = \sqrt{\frac{2\pi}{\lambda} (\epsilon - \epsilon_1 + \sin^2 \alpha)}$$

$$dK \approx \frac{K d\epsilon}{2 \sin^2 \alpha}$$

$$\phi = \frac{2\pi y \alpha}{\lambda} = Ay; \qquad \theta = \frac{2\pi H \alpha}{\lambda} = AH$$

λ = wavelength in same units as height y

$H = y_2 - y_1$.

It follows that for $(\epsilon_2 - \epsilon_1) \ll \sin^2 \alpha$

$$q \approx \frac{1}{4 \sin^2 \alpha} \int_0^\theta \frac{d\epsilon}{d\phi} e^{-i2\phi} \, d\phi. \tag{10}$$

For a linear variation corresponding to

$$\epsilon = \epsilon_1 + \frac{\phi}{\theta} (\epsilon_2 - \epsilon_1)$$

the reflection coefficient reduces to

$$q = R \frac{\sin \theta}{\theta} e^{-i\theta}, \qquad R = \frac{(\epsilon_2 - \epsilon_1)}{4 \sin^2 \alpha}.$$

The factor R is the reflection coefficient expected for a sharp boundary with the same total change in dielectric permittivity, and θ is the total phase change in the inhomogeneous media.

In a similar manner, the reflection coefficient for the antisymmetrical cubic illustrated by case 2 is given by

$$q = R \frac{3}{\theta^2} \left(\frac{\sin \theta}{\theta} - \cos \theta \right) e^{-i\theta}.$$

The above examples for linear and S shaped curves are given by Schelkunoff. The same general method has been used more recently by du Castel, Misme and Voge to obtain the reflection coefficient for the additional cases of parabolic, sinusoidal and hyperbolic variations.[17] Their results can be summarized as follows:

$$\text{linear } |q| = R \frac{\sin \theta}{\theta} \tag{11}$$

$$\text{sinusoidal } |q| = R \frac{\cos \theta}{1 - \theta} \tag{12}$$

$$\text{parabolic } |q| = R \left(\frac{\sin \theta}{\theta} \right)^2 \tag{13}$$

$$\text{hyperbolic } |q| = R \frac{\theta}{\sinh \theta}. \tag{14}$$

In each case, as θ approaches zero, the reflection coefficient approaches the value given by Snell's law for an abrupt change.

The effect of a gradual decrease in dielectric permittivity is to reduce the reflection coefficient below what would have been expected if all of the change occurred at a sharp boundary. The reduction in reflection

coefficient depends on the slope of transition and not on the boundary conditions alone. For the linear case it varies as $1/\theta$ and for the hyperbolic case it ultimately decreases exponentially. As θ approaches infinity the reflection coefficient goes to zero for these particular cases, but before reaching the conclusion that q always goes to zero as the upper limit goes to infinity, it is instructive to formulate the general case.

In general, the dielectric permittivity ϵ can be represented by

$$\epsilon = \epsilon_1 + m(\epsilon_2 - \epsilon_1)$$

where the parameter m can be any function that varies from $m = 0$ at the lower boundary ($\epsilon = \epsilon_1$) to $m = 1$ at the upper boundary ($\epsilon = \epsilon_2$). The corresponding reflection coefficient is

$$q = R \int_0^\theta \frac{dm}{d\phi} e^{-i2\phi} \, d\phi. \tag{15}$$

An exponential decrease with height provides a better match to the average atmospheric variation than is possible with any of the above models. Consequently, it is instructive to consider the case of $m = 1 - e^{-by}$, where b is the rate of decrease in atmospheric density. The magnitude of b is about 0.14 per km and is determined by the characteristics of gases and the law of gravity. For the case of a smooth exponential decrease, the reflection coefficient given by (10) or (15) reduces to

$$q = R \int_0^\theta b e^{-by} \frac{dy}{d\phi} e^{-i2\phi} \, d\phi \tag{16}$$

$$= R \int_0^H b e^{-(b+i2A)y} \, dy$$

$$= R \frac{b}{b + i2A} [1 - e^{-(b+i2A)H}]$$

$$= \frac{\epsilon_2 - \epsilon_1}{4 \sin^2 \alpha} \frac{b}{b + i2A} [1 - e^{-(b+i2A)H}] \tag{17}$$

where

$$\theta = AH$$

$$\phi = Ay.$$

It will be noted that the upper limit H appears only in the exponential factor and that the term in brackets approaches unity as $H \to \infty$. The reflection coefficient for a smooth exponential atmosphere is the reflec-

tion coefficient R that would result if all the change in dielectric permittivity occurred at a sharp boundary, multiplied by $b/(b + i2A)$. In the atmosphere, $b \ll A$ for VHF and above, as long as α is greater than 0.01 degree.

The bilinear model (uniform decrease with increasing height until $\epsilon = 1$) was used extensively in earlier studies, but its usefulness was questioned because of the discontinuity at the upper boundary $H = 1/b$. In the present nomenclature the reflection coefficient for the bilinear model with a slope b is

$$q = R \frac{b}{i2A} [1 - e^{-i2AH}]. \tag{18}$$

This is equivalent to the derivation of equation (2) in Ref. 2. The similarity between (17) and (18) shows that the use of the exponential decrease removes the effect of the upper discontinuity without any significant change in the magnitude of the reflection coefficient.

The foregoing analysis has been based on a rectangular coordinate system centered on the lower boundary to facilitate comparison with earlier papers. It is now desirable to shift to the geometry illustrated in Fig. 1. In this case

$$m = 1 - e^{-b(y-y_s)}$$

$$\epsilon = \epsilon_s + [1 - e^{-b(y-y_s)}][1 - \epsilon_s] \tag{19}$$

$$= 1 + (\epsilon_s - 1)e^{-b(y-y_s)}$$

$$d\epsilon = -(\epsilon_s - 1)be^{-b(y-y_s)}dy \tag{20}$$

where ϵ_s is the dielectric permittivity at the surface.

Substituting (20) into (10),

$$q = -\frac{(\epsilon_s - 1)}{4 \sin^2 \alpha} e^{by_s} b \int_{Ay_1}^{Ay_2} e^{-(b+i2A)y} \frac{dy}{d\phi} d\phi$$

$$= -\frac{(\epsilon_s - 1)}{4 \sin^2 \alpha} e^{by_s} b \int_{y_1}^{y_2} e^{-(b+i2A)y} dy$$

which reduces to

$$q = -\frac{(\epsilon_s - 1)}{4 \sin^2 \alpha} e^{-b(y_1-y_s)} e^{-i2Ay_1} \frac{b}{b + i2A} [1 - e^{-(b+i2A)(y_2-y_1)}]. \tag{21}$$

From the geometry in Fig. 1 it can be seen that for the symmetrical case the grazing angle $\alpha \approx 2y_1/D$ and the antenna beamwidth $\beta \approx$

$(2/D)(y_2 - y_1)$. In addition, each successive Fresnel zone corresponds to a change in phase of π radians, so $y_2 - y_1 = M\pi/A$, where M is the number of pairs of Fresnel zones. With these substitutions (21) can be rearranged as follows

$$
q = \frac{(\epsilon_s - 1)}{4 \sin^2 \alpha} e^{-bD\alpha/4} \frac{b}{2A} [1 - e^{-bD\beta/2}] \left[\frac{1 - e^{-bD\beta/2} e^{-i2M\pi}}{1 - e^{-bD\beta/2}} \right]
$$

$$
\left[\frac{\exp\left(-b\left(\dfrac{y_1}{2} - y_s\right)\right)}{1 - \dfrac{ib}{2A}} \right]. \quad (22)
$$

This equation has already been given by (5). The first bracket to the right of the equality sign is equivalent to equation (27) in Ref. 18 except that the latter is 2.24 times larger to account for the effect of the earth and for a gradual cutoff near the horizon.

This completes the derivation of (5), but a better understanding of the physical problem can be obtained by introducing the factor $(1 - e^{-b\pi/A})$ in both numerator and denominator as follows

$$
q = \left[\frac{(\epsilon_s - 1)}{4 \sin^2 \alpha} (1 - e^{-b\pi/A}) e^{-b(y_1 - y_s)} \right] \left[\frac{b}{b + i2A} \right]
$$

$$
\left[\frac{1 - e^{-b\pi M/A} e^{-i2M\pi}}{1 - e^{-b\pi/A}} \right]. \quad (23)
$$

The terms inside the first brackets represent the magnitude of the reflection coefficient that would result if the change in dielectric permittivity within one pair of Fresnel zones (phase change of 360°) were concentrated at a sharp boundary. The middle bracket corrects for a gradual change instead of a sharp boundary. The right-hand brackets represent the sum of M pairs of Fresnel zones and will be recognized as the sum of a geometric progression of M terms whose common ratio is $e^{-b\pi/A}e^{-i2\pi}$. The factor $(1 - e^{-b\pi/A})$ is the fraction of the reflected energy that remains in each pair of Fresnel zones after the almost complete cancellation caused by opposite phase in two adjacent Fresnel zones.

The concept of Fresnel zones is used to simplify the problem of adding a large number of components with a wide range of phases. By grouping into blocks of 2π radians, the contribution from each block adds in phase, and it is easy to see that the inclusion of weak components beyond nominal boundaries does not cause the answer to disappear. The cutoff merely neglects this contribution and the true answer is slightly greater than the computations indicate. In other words, the effect of the "discontinuity"

at the lower limit of integration, which has proved troublesome in some formulations of the problem, seems to have a negligible effect on the magnitude of the signal returned from the atmosphere.

Although the Fresnel zone concept provides a clearer understanding of the problem, it is easier for computation purposes to use (5).

REFERENCES

1. Gordon, W. E., Radio Scattering in the Troposphere, Proc. I.R.E., **43**, Jan. 1955, pp. 23–28.
2. Friis, H. T., Crawford, A. B., and Hogg, D. C., A Reflection Theory for Propagation Beyond the Horizon, B.S.T.J., **36**, May 1957, pp. 627–644.
3. Carroll, T. J., and Ring, R. M., Propagation of Short Radio Waves in a Normally Stratified Troposphere, Proc. I.R.E., **43**, Oct. 1955, pp. 1384–1390.
4. Villars, F., and Weisskopf, V. F., On the Scattering of Radio Waves by Turbulent Fluctuations of the Atmosphere, Proc. I.R.E., **43**, Oct. 1955, pp. 1232–1239.
5. Staras, H., Forward Scattering of Radio Waves by Anisotropic Turbulence, Proc. I.R.E., **43**, Oct. 1955, pp. 1374–1380.
6. Wheelon, A. D., Radio-Wave Scattering by Tropospheric Irregularities, J. Res. NBS, pt. D (Radio Propagation), **63D**, Sept.–Oct. 1959, pp. 205–233.
7. Crain, C. M., Survey of Airborne Microwave Refractometer Measurements, Proc. I.R.E., **43**, Oct. 1955, pp. 1405–1411.
8. Bean, B. R., and Riggs, L. P., Synoptic Variation of the Radio Refractive Index, J. Res. NBS, pt. D (Radio Propagation), **63D**, July-Aug. 1959, pp. 91–97.
9. Bean, B. R., Horn, J. D., and Ozanich, A. M. Jr., Climatic Charts and Data of the Radio Refractive Index for the United States and the World, NBS Monograph 22, Nov. 25, 1960.
10. Bean, B. R., and Thayer, G. D., Models of the Atmospheric Radio Refractive Index, Proc. I.R.E., **47**, May 1959, pp. 740–755.
11. Hartman, W. J., and Wilkerson, R. E., Path Antenna Gain in An Exponential Atmosphere, J. Res. NBS, pt. D (Radio Propagation), **63D**, Nov.–Dec. 1959, pp. 273–286.
12. Booker, H. G., and De Bettencourt, J. T., Theory of Radio Transmission by Tropospheric Scattering Using Very Narrow Beams, Proc. I.R.E., **43**, March 1955, pp. 281–290.
13. Chisholm, J. H., Morrow, W. E., Nichols, B. E., Roche, J. F., and Teachman, A. E., Properties of 400 mc Long Distance Trophospheric Circuits, Proc. I.R.E., **50**, Dec. 1962, pp. 2464–2482.
14. Trolese, L. G., Characteristics of Trophospheric Scattered Fields, Proc. I.R.E., **43**, Oct. 1955, pp. 1300–1305.
15. Crawford, A. B., Hogg, D. C., and Kummer, W. H., Studies in Tropospheric Propagation Beyond the Horizon, B.S.T.J., **38**, Sept. 1959, pp. 1067–1178.
16. Schelkunoff, S. A., Remarks Concerning Wave Propagation in Stratified Media, Communications in Pure and Applied Mathematics, June 4, 1951, p. 117–128.
17. Du Castel, F., Misme, P., and Voge, J., Sur le Rôle des Phénomènes de Réflexion dans la Propagation Lointaine des Ondes Ultracourtes, *Electromagnetic Wave Propagation*, New York, Academic Press, pp. 670–683.
18. Kalinir, A. I., Effect of Earth on UHF Tropospheric Scatter Propagation, Radio Engineering and Electronic Physics, No. 5, May, 1961.

The Effects of Time Delay and Echoes on Telephone Conversations

By J. W. EMLING and D. MITCHELL

A brief history of the problem of echoes and delay in telephone connections is first given. Actual delays involved in typical circuits are shown. This is followed by a discussion of the effects of delay only on typical conversations. The sources of echo in typical telephone connections are then discussed, together with measures which have been taken to reduce echo by improving return loss. Methods of controlling the effects of delayed echo are then summarized; the summary includes a brief introduction to echo suppressors and some of the new problems they introduce.

I. INTRODUCTION

About 30 to 40 years ago, the effects of time delay and echo became of great concern to engineers planning transcontinental telephone systems. As a result, much effort was devoted to understanding these effects and finding ways to control them.[1] Fortunately, the development of high-speed transmission systems made the problem less severe than originally anticipated, and satisfactory ways were found to handle the amounts of delay that were then involved. As a result, interest in the effects of delay declined and little work has been carried out in this field during the last twenty years. But, as happens so often, the historical cycle is beginning to repeat. With the growth of intercontinental telephony and particularly the proposal of satellite systems involving one-way path lengths of 50,000 miles or more, the problems of delay have again come to the fore.

Because of this renewed interest, Bell Laboratories has resumed its studies of echo and delay. The basic problems, first considered 30 or more years ago, have been reviewed to see if improved solutions might result from the technology that has evolved in the intervening period. This work has been addressed not only to technical problems but also to new testing techniques for the evaluation of echo, delay and the control of these factors. Throughout the recent program, emphasis has been

placed on the longer delays which are the unavoidable result of expanding the telephone network to provide a truly world-wide service.

This paper summarizes the earlier information on the effects of delay and echo, and outlines the fundamental problems they introduce. It thus serves as a guide to the literature of the past and also as an introduction to two companion papers dealing with the more recent studies in this area. The paper by Brady and Helder[2] traces the evolution of the echo suppressor, outlines the basic design problems involved and covers recent work aimed at applying modern technology to its design. The paper by Riesz and Klemmer[3] covers recent work on the subjective evaluation of delay and echo suppression in telephone communication.

II. HISTORICAL BACKGROUND

The problem of delay became acute during the 1920's and 1930's because voice-frequency circuits 500 miles and more in length were being set up for the first time in loaded cable. The propagation speed of these facilities was much slower than that of the open wire used previously, and the delay became sufficient to cause objectionable echoes. This initiated activity which led to the invention of the echo suppressor and also to intensive efforts to improve hybrid balance. Echo suppressors are capable of minimizing the most serious effects of echoes, but it was soon found that they also introduced other difficulties by interfering with the free two-way flow of speech. These impairments become more severe as the echo suppressor has to cope with longer delays. It was also found that delay alone introduces difficulties when it becomes large enough.

These facts naturally stimulated the development and use of the higher-speed carrier systems. Since the 1930's the use of these systems has grown rapidly, and at present practically all circuits of over 25 miles transmit at speeds of at least two-thirds that of light.

Table I shows the delays encountered on typical circuits of 30 years ago, on present-day circuits and on those which may be in service in the future through the use of satellites. This table clearly brings out the reasons for the intense interest in delay in the earlier period, the reduction of interest as high speed circuits were introduced, and the renewed current interest.

III. FUNDAMENTAL PROBLEMS

Problems due to echo and delay fall in three general categories. First are those due to the delay alone; second, those due to echo; and third,

TABLE I — AMOUNT OF DELAY IN LONG TELEPHONE CIRCUITS

Circuit	Approx. Length (miles)	Facility	Approx. Signal Velocity (mi/sec)	Approx. One-Way Delay* (msec)
(a) Actual Circuits in the 1930's				
New York to Chicago	900	VF on loaded cable (H-44)	20,000	13
New York to Dallas	1,850			93
(b) Considered in the 1930's but Not Established				
New York to San Francisco	3,200			160
(c) Present-Day Circuits				
New York to Hagerstown	260	VF on loaded cable (H-44)	20,000	13
New York to Chicago	900	K carrier	110,000	8
New York to San Francisco	3,200	TD-2	186,000	19
New York to London	4,100	submarine cable and others	120,000 for cable	35
Hawaii to London	10,000	submarine cable + 3000 mi of TD-2		80
(d) Circuits Involving Satellites				
New York to London	13,000† to 18,000	6,000-mile satellite + TD-2	186,000	70 to 97
	49,000	24-hour satellite + TD-2	186,000	265
Hawaii to London	29,000† to 38,000	two 6,000-mile satellites + 3,000 miles of TD-2	186,000	160 to 210
	98,000	two 24-hour satellites + 3,000 miles of TD-2	186,000	530

* Includes an allowance for terminals
† Length of path varies with satellite position.

those introduced by echo suppressors. This paper concentrates largely on the first two categories, leaving the last to the companion papers.

When two persons who are separated by a large distance wish to converse, there are two fundamental factors which must be taken into account. In their simplest form, these are illustrated in Fig. 1. This shows customer A at the left and customer B at the right connected by a four-wire circuit all the way from microphone to receiver (each line in the
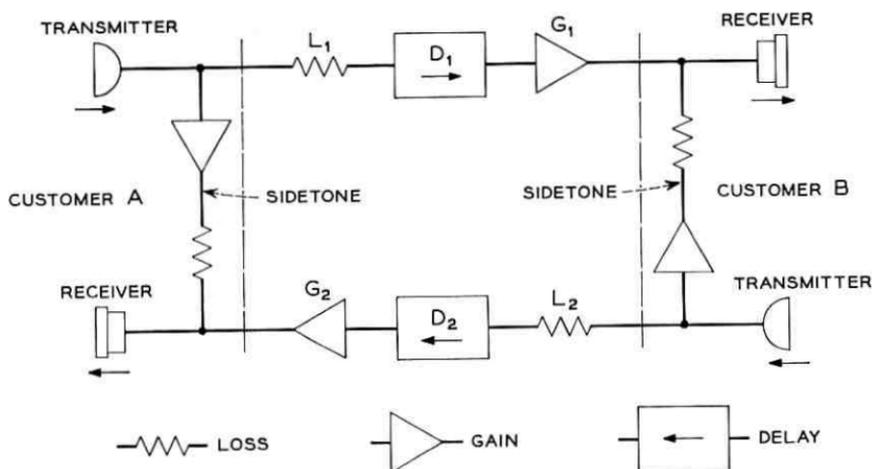
Fig. 1 — Four-wire toll circuit — no echo.

diagram actually represents two wires). There will be losses, indicated as L_1 and L_2, which must be compensated by gains G_1 and G_2 or the two will not be able to talk at all. In addition, there will be time delays between them, indicated by D_1 and D_2, for which no compensation is possible. In a practical case, the losses, gains, and delays are roughly equal in the two directions; they are not lumped as indicated in the figure, but are distributed over the entire circuit. For most long circuits, the net loss $(L - G)$ is made positive but close to zero.

Unlike losses that can be reduced as desired by amplification, delay cannot be reduced below an absolute minimum set by nature. This minimum equals the distance divided by the speed of light (186,000 mi/sec). As indicated previously, many of the older transmission facilities had speeds well below the speed of light, but current facilities achieve speeds closely approaching the theoretical maximum, and hence are minimum-delay facilities.

Delay changes the conversational process in a fundamental manner. We shall have more to say about this later, but for the moment it is only necessary to point out that when one talker stops talking he cannot hear the other reply until a time $D_1 + D_2$, or about $2D_1$, later.

For very long circuits, requiring the use of much gain, the separation of the two directions of transmission is economical, and it permits the use of high gains with a minimum of complication to avoid singing. On shorter circuits, however, it is far more economical to use "two-wire" circuits — that is, the same path for the two directions of transmission. Since by far the largest number of telephone circuits are short in length

(i.e., customers' loops and local trunks), the aggregate economic benefit of using two-wire transmission on these circuits is very large. In addition to the greater economy of the facilities themselves, it is obviously more economical from the standpoint of switching (or interconnecting) circuits to use two wires instead of four.

Thus, the usual long distance call today involves both two- and four-wire circuits, much as shown in Fig. 2. The long four-wire toll circuits are converted to two-wire for transmission over short trunks between long distance and local offices, and over the customer's loop. At the end of the loop, the telephone reconverts to four-wire for connection to the customer's transmitter and receiver.

The conversion from a four-wire path to a two-wire path is accomplished by a hybrid coil. The effectiveness of the hybrid coil depends on the degree of match (or balance) between the impedances of the line and the network N. This balance is never perfect and therefore results in some of the incoming energy being returned as echo E_1 or E_2.

It has been found that echo becomes increasingly objectionable as the delay is increased. This is because close-in echoes tend to be masked by sidetone speech, but the masking effect decreases rapidly after speech ceases. Echo can be made less objectionable by increasing net loss, but loss obviously must not be increased beyond the point where talking is
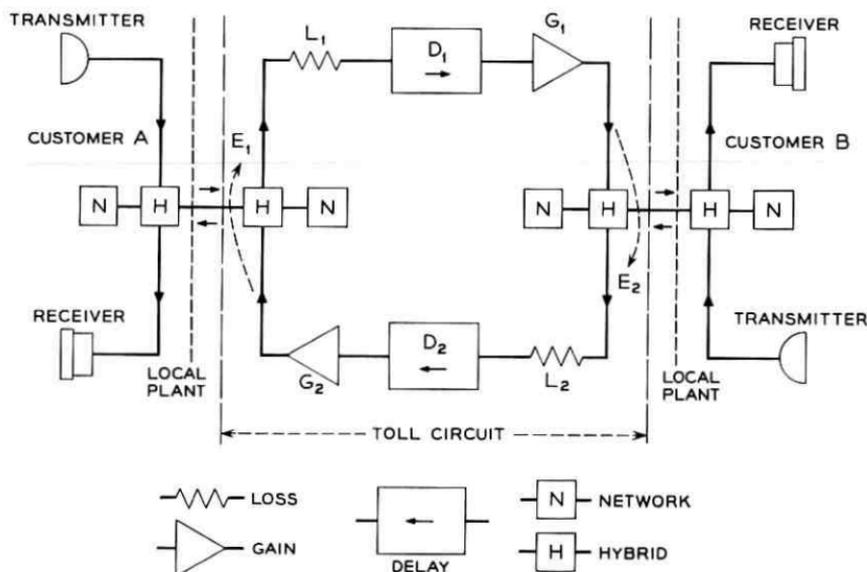
Fig. 2 — Four-wire toll circuit operated two-wire has echoes E_1 and E_2.

relatively comfortable. It was this situation which led to the invention of the echo suppressor. The fundamental point to be observed, however, is that increasing the delay results in two serious effects. The first is the increase in the response time and the second is increase in the severity of echo effects.

IV. EFFECTS OF DELAY ALONE

Delays of tenths of seconds are in the region of typical human reaction times and can be expected to have important effects on the structure of conversation even with four-wire circuits which are completely free from echo.

Fig. 3(a) illustrates a short portion of a typical conversation. The to-and-fro speech is shown for a case where there is zero delay in the telephone circuit. Line A shows the "talk spurts" TS from talker A at one end of the connection. A "talk spurt" is broadly defined as a portion of speech coming entirely from one talker.

In line B, the responses of talker B are shown. As is indicated, the talk spurts of each party vary considerably in length; and also the response times may vary. The "response time" RT, as shown, is defined as the interval from the time one party hears the other stop speaking until he himself starts speaking. RT may be negative if one party starts before the other stops.
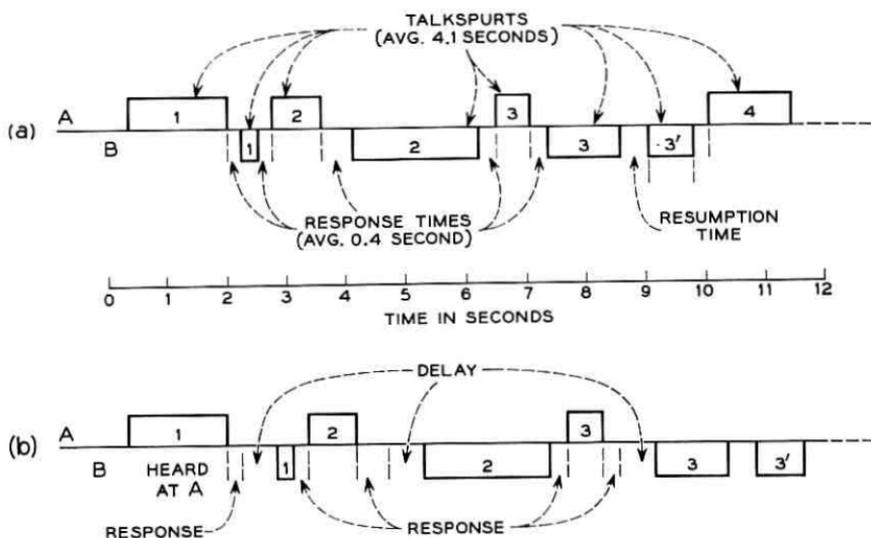
Fig. 3 — Typical time intervals in a two-way telephone conversation: (a) no circuit delay, (b) same conversation observed with added round-trip delay of 0.6 second.

Talk spurts, as defined, are not necessarily continuous. This is illustrated toward the right of line B by a break between B_3 and B_3' which provides an interval defined as "resumption time."*

In Table II some data are given showing typical TS and RT times as determined some years ago in observations made by A. C. Norwine and O. J. Murphy on actual telephone conversations.[4] These data were taken on slightly over 2800 telephone calls over a New York-to-Chicago private line used exclusively for Bell System business.

It is of interest that there are a rather large number of very short talk spurts. Thus the most frequent value (mode) is only 0.25 second. This probably represents the usual frequent monosyllabic replies. On the other hand, the median TS is two seconds and the average is a little over four. There are also occasional quite long talk spurts.

TABLE II—TALK SPURT (TS) AND RESPONSE TIME (RT)
INTERVALS (IN SECONDS)

	Min.	Mode	Median	Mean	Max.
TS	0.09	0.25	2.0	4.14	143.8
RT	−3.95	0.24	0.32	0.41	5.04

For response times, the differences between mode, median and mean are less but there are a few long pauses and some negative times which represent double talking.

The time diagram in Fig. 3(b) illustrates one way in which large round-trip delays can affect conversation. This is a somewhat oversimplified model in that it assumes that the delay does not affect conversational structure. The added response time of 0.6 second is actually almost three times the mode of the normal RT and about twice the median value. Thus, if people did not change their conversational habits, this increase in response time would substantially add to the amount of time required to carry on a given amount of two-way conversation. The percentage increase in total time depends also on the length of typical talk spurts. Using the figures given in the preceding table, it can be estimated that the average conversation might be lengthened by from 6 to 12 per cent if the talkers continued to talk in just the same manner as before.

Experience with long delays, which is covered in the Riesz-Klemmer paper,[3] however, indicates that the effect of delay is considerably more complex than this simple model. Actually, calls tend to be shorter instead of longer when long delay is present. Apparently there are subtle

---

* Some workers in this field define B_3 and B_3' as separate talk spurts.

factors of annoyance and discomfort which cause the talkers to change their conversational structure and which make them wish to terminate the conversation sooner.

The simplified illustration, however, shows that even if talkers learned to adapt themselves to delay and discipline themselves to wait patiently for replies, there still would be degradation. Thus it seems inescapable that the addition of delays of one-half second or more will substantially reduce the true value per unit time of a toll connection. The very essence of most telephone calls is the ability to conduct a rapid exchange of ideas. This may not apply to some other types of communication, however, which involve relatively large amounts of one-way information.

The illustration is perhaps even more useful in pointing out the possibility that delay modifies the conversational pattern. As noted previously, even with zero delay some response times are negative. That is, there is a tendency to start talking before the other user has completed his talk spurt. With zero delay, this interruption is noted promptly, and the period of "double talking" is small. With long delays, however, the person who breaks in may talk for some time before the other is aware of the attempted interruption. In fact, it may be associated with a later, and quite different, part of the talk spurt than that which occasioned the break-in. If a response is long delayed, the original talker may resume his conversation before it is received, in which case the response to one piece of conversational material may appear as an interruption to a later, and possibly different, idea.

Thus, delay not only increases the tendency to double talk, but also increases the potential for confusion. As will be shown elsewhere,[3] these characteristics greatly complicate the suppression problem.

V. ECHOES AND THEIR EFFECTS

5.1 Sources of Echo

Any impedance mismatch in a transmission system will reflect energy back toward the source and be a potential cause of echo. If there is a mismatch at each end of a transmission line, the energy will be repeatedly reflected back and forth until dissipated by the line. In a strictly four-wire system, these reflections are ordinarily of no moment to the telephone user, since the amplifiers are one-way transmission devices and confine the echoes to small sections of the system with short transmission time and sufficient loss to rapidly damp out the reflections.

Echoes are likely to become important, however, as soon as two-wire transmission is employed, since the echo can now be returned with not

much more attenuation than is encountered by the direct wave. Even though the major part of the system is four-wire, the effect of the two-wire portion needs to be considered since it usually controls the magnitude of the echoes.

As noted earlier, a commercial telephone connection always involves some two-wire transmission: a typical situation is illustrated in Fig. 2. Within the telephone station, the user's inherent four-wire transmission is converted to two-wire to fit into the exchange plant. At a toll office, it is reconverted to four-wire for transmission over long distances, and corresponding conversions are made at the far end. In practice, there may be other conversions, since four-wire transmission is sometimes used in the local plant, but is converted to two-wire for switching. Two-wire transmission is seldom employed in the toll plant, but some conversion from four-wire to two-wire for switching does occur. We need not consider these additional conversions for an understanding of the echo problem, but it should be appreciated that they complicate its solution.

Four-wire to two-wire conversions are made by means of a hybrid coil circuit. This is a form of bridge in which a network is provided to balance the line as shown in Fig. 4. If the impedance of the balancing network N is equal at all frequencies to the impedance of the two-wire line, the energy from the incoming four-wire branch is equally divided between the line and the network and the conversion is accomplished without echo. The "balance" in the hybrid circuit is then considered perfect. If there is any mismatch between the line and network impedances, the balance will be less than perfect and some of the energy will be transferred to the outgoing branch of the four-wire system.

Referring to Fig. 2, it will be noted that speech from customer A can
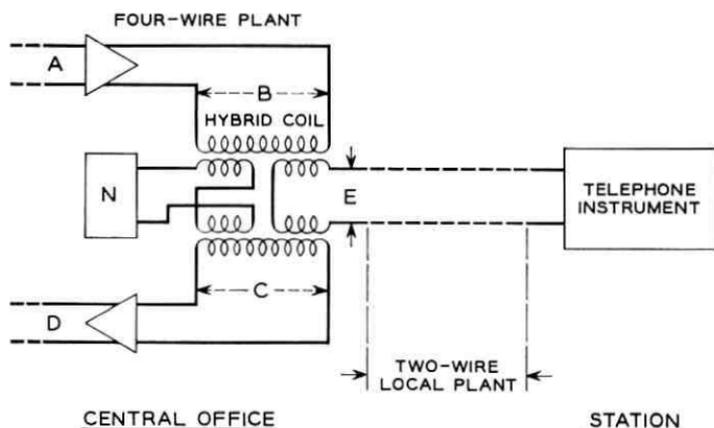
Fig. 4 — Occurrence of echo in two-wire local plant.

be transferred from one side of the four-wire circuit to the other at either his own telephone set, in which case it is called "sidetone," or at the distant, or B toll office. In this latter case, it is called "talker echo."

The principal difference between sidetone and talker echo is one of timing. Within the telephone station, transmission is essentially instantaneous and the sidetone appears concurrently with the talker's speech. The time for transmitting sounds to a distant toll office may, however, be many milliseconds. The talker echo may consequently be returned with a noticeable delay, in which case he will hear his own words reflected back much as if he were talking toward a cliff or other source of acoustical reflection.

Some of the talker echo can also be reflected at office A and back again toward the listener at the B office. This is referred to as "listener echo," since it is the distant listener who hears this some time following the arrival of direct speech. This process continues until the losses in the circulating path reduce the echo below audibility. This is illustrated in Fig. 5, reproduced from a paper[5] by G. M. Phillips.

It is customary to simplify discussions of echoes, as we have just done, by considering the transfer of energy as if it occurs at the hybrid coil in Fig. 4 (i.e., at the two to four-wire conversion point). In dealing with steady-state conditions, this is permissible. The magnitude of the steady-state echo is determined only by the relative impedances of the network and the line at the frequency of interest, and it can be computed if the steady-state impedance of the line is known at all frequencies. In problems involving transients, however, it is important to realize that the actual reflection occurs at any point or points in the two-wire line between the office and the station at which an impedance mismatch occurs. The time at which an echo occurs is therefore determined not only by the
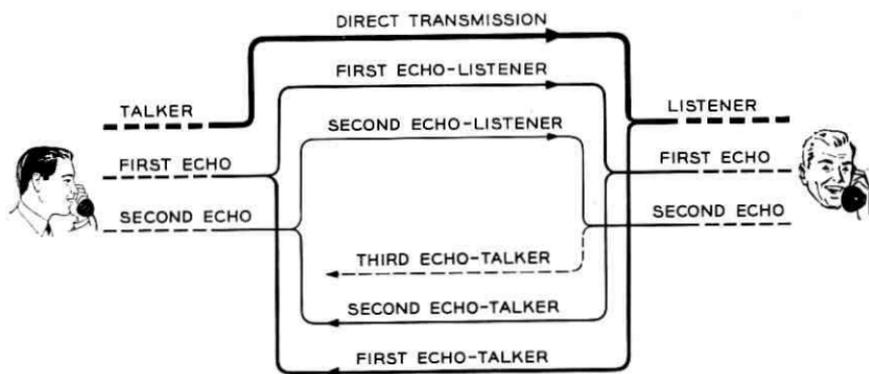
Fig. 5 — Telephone echo paths in a typical four-wire circuit.

transit time to the distant office but also by the transit time between the hybrid coil and the *point of reflection*.

Reflections do, in fact, occur at many places in a two-wire transmission line. Unless transformers have very large mutual inductance, they will usually cause some reflection at low frequencies. In voice-frequency circuits, variations in cable mutual capacitance and irregularities in the spacing of loading coils will introduce many small reflections. Thus in practice the echo tends to be a "smear" of many reflections.

However, by far the largest irregularity and associated reflection usually occur at the telephone station. Voice-frequency telephone lines will ordinarily have a characteristic impedance with a negative angle varying from a few degrees in the case of loaded lines to as much as 45 degrees with nonloaded lines. The magnitude of the impedance may be reasonably constant with frequency, for loaded lines, or decrease rapidly with increasing frequency for nonloaded. The telephone station impedance, on the other hand, is largely affected by the nature of the receiver. The angle is positive and the magnitude increases with frequency. It is evident that this is a very poor match with the line and a large source of echo.

Another important reflection may occur at the local central office. If the circuit (loop) from this office to the customer consists of loaded cable, it will provide a reasonably good match to the facility connecting the local office to the toll office. Where it is nonloaded cable, however, the match is poor because of the low magnitude and large negative angle of this facility.

The station and loop mismatch are problems of long standing, and there has been continuing effort devoted to their reduction. The problem is not so much technical as it is economic. The station and the loop connecting it to the central office are the most numerous elements in the plant (the Bell System has over 65,000,000 stations), and methods for solving the echo problem by improving their impedance characteristics would involve very large aggregate costs because of the large multiplying factor. Instead, it has proved more economical to reduce echo by measures applied to trunk circuits, since they are far less numerous.

The impedance of the two-wire local plant is a highly variable quantity. For any one loop, it varies greatly with frequency and also varies with the type of set and the length and type of facility. Thus, it is impractical to balance these impedances perfectly, and it is customary to select a network which provides the best compromise balance over the range of conditions commonly encountered. This network is 600 ohms in series with a capacitance of 2 mf in most toll offices, but 900 ohms with the same capacitance is used for local offices and some toll offices.

The most convenient way to express the amount of echo quantitatively is by using the concept of return loss, that is, the difference in db between the energy delivered to the circuit at a given point and the energy returned at this point as echo.* Obviously, the return loss of the local plant relative to the compromise network is a very complex matter, since it involves both frequency and facility. From the standpoint of echo, the frequencies between 500 and 2500 cps are of most importance, and it is customary to use a single value "average" over this band as a description. The effects of the various types and lengths of facility are expressed in terms of a statistical distribution of this "average" return loss or more particularly in terms of the mean and standard deviation of this distribution.

Typical values for return loss on telephone connections with compromise networks are a mean of 15 db with a standard deviation of 3 db at the toll office. At the user's local office, the corresponding figures are 11 and 3 db. The higher mean at the toll office is largely accounted for by the attenuation of the toll connecting trunks between the local and toll offices.

Return losses in the toll plant between one toll circuit and another are usually considerably higher than these values, and as noted earlier, return losses at the occasional points where toll circuits are switched on a two-wire basis are usually not controlling in the connection.

A long and continuing effort has been aimed at improving local return losses. So far, it has not proved economical to greatly reduce the unbalance at the telephone loop and station; however, impedance equalizers are used extensively on the toll office end of the exchange trunks giving access to the toll plant, and similar arrangements for other parts of the local plant are being developed. It appears that these and similar measures will achieve small but significant further improvements in local return loss over the next few years.

5.2 Effects of Echo

Not all echoes are harmful. For example, the telephone user likes some sidetone (which is talker echo without delay) because it gives him the impression that the circuit is "alive." Recent tests at Bell Laboratories[7] have indicated that the preferred sidetone volume is comparable to the volume he would like to receive from the distant talker. One reason for preferring this much sidetone is the high level at which the talker hears his own speech (largely via paths within the head) when he

---

* Appendix A discusses the return loss concept at somewhat greater length.

covers one ear with a telephone receiver. Obviously, his telephone side-tone must be noticeable compared to the head sidetone to obtain the desired impression of a live circuit.

When a telephone user is not talking, he does not have the masking due to his own voice, and he notices much weaker echoes. Delay which displaces talker echoes so that they can fall in silent intervals between speech sounds increases the chance that the talker will hear them. When the delays are short, the effects are small, since they are not much displaced from sidetone, and are manifested as sidetone having a hollow sound. With longer delays, the echoes clearly stand out, and, if the round trip delay is about one-quarter second and the returned echo is sufficiently loud (roughly 10 db above sidetone), it becomes so distracting that talking is virtually impossible. Even at very low levels, such long-delayed echoes are very annoying.

The annoyance of talker echo was recognized in the early days of long distance telephony, and tests to determine the tolerable echo were reported by Clark and Mathes[6] in 1925. In 1953, this work was repeated with current telephone circuits[5] with the results shown on Fig. 6. This shows the relation between round-trip delay and the minimum loss required to attenuate the echo sufficiently to provide a commercially tolerable condition in the judgment of an average listener. "Commercially tolerable" means that although an echo was discernible it was not loud enough to be objectionable. The curve represents the average tolerance to echo of all the listeners, but any individual may differ from the average by a considerable amount. About 68 per cent of customers have
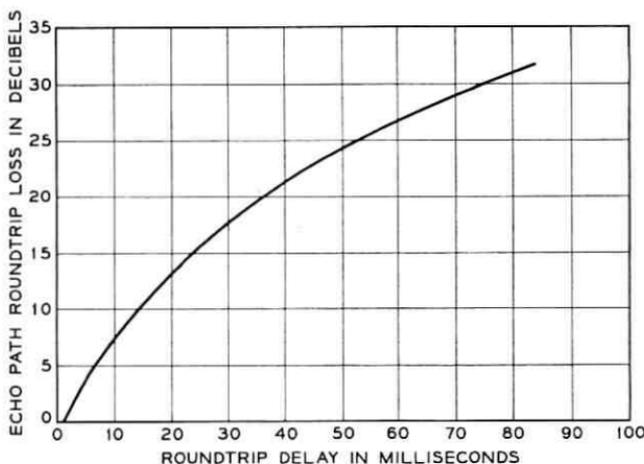

Fig. 6 — Average listener judgment of minimum echo attenuation necessary for just-tolerable talker echo conditions on modern telephone sets.

a tolerance to echo lying within ±5 db of this average curve. The remaining 32 per cent would be about evenly divided above and below but differing by more than 5 db from the average.

These tests were made with a typical loop connecting the talker to his local central office. The round-trip losses and delays shown are measured from the talker's central office to the point of reflection and return to the same central office. In a practical case, the loss in the echo path is made up of twice the net loss plus the far end return loss.

It might be thought that listener echo would be more troublesome than talker echo, since the user is then always in the "listening" condition, and hence does not have the masking effect of his own sidetone to reduce the effect of echo. In actuality, it has been found that where the round-trip delay is around one-quarter second or more, a given amount of talker echo is just about as annoying as the same amount of listener echo. Hence, talker echo is controlling on typical telephone circuits because the listener echo is always additionally attenuated by a second (near end) return loss.

5.3 *Controlling the Effects of Echo*

There are two fundamental ways in which echo can be controlled: (a) by reducing the delay and (b) by increasing the loss in the echo path, i.e., controlling echo magnitude. The use of carrier systems with their high propagation speeds has reduced the delay on a large part of the land line system to the point where echo is negligible. Since the speed is now close to that of light, little further reduction of delay is possible, and control of echo magnitude must be adopted where length is great enough to cause significant delay.

Control of echo magnitude has, over the years, taken many forms. Four-wire circuits and four-wire switching are used very largely in the long distance plant and have largely eliminated echoes except from the end links.* As noted earlier, it has so far not proved economical to reduce greatly the reflection at the telephone station.

Another way to reduce the effects of echo is to increase the loss between the talker and the point of major echo. It is to be noted that each db added will decrease the echo by 2 db. Attractive as this may be from an echo standpoint, it is obvious that it must not be pushed to the point where the direct transmission path is adversely affected.

---

* Even though the toll transmission paths are almost entirely four-wire, two-wire switching is still used extensively at smaller toll offices. By careful adjustment of the balancing networks and office wiring, the reflections from the four- to two-wire conversions are kept minor compared to the end link reflections.

The approaches just discussed are all simple, direct attacks on the problem. There is also a slightly more sophisticated approach, the echo suppressor, which is discussed in more detail in a companion paper.[2] This is a voice-operated switching device which reduces echo by introducing loss in the return transmission path of the four-wire system. Other sophisticated devices are conceivable. For example, since the return loss problem arises from the inability to obtain good return loss from a single balancing network facing a large variety of line impedances, it might be possible to make available a series of balancing networks from which the one giving the best balance for the particular circuit in use would be selected. This solution has been looked into quite extensively but, so far, the relatively small benefits which might result have not appeared to warrant the complications of implementation.

Another proposal frequently advanced has been the use of a self-balancing hybrid. Most of these schemes involve the measurement of current and voltage associated with the line impedance and the introduction of compensating currents or voltages in the balancing arm. F. B. Llewellyn has shown that this result is theoretically unattainable with any *linear networks*, i.e., with any arrangement for which the inserted compensating voltage or current is instantaneously, linearly related to the line value. This formerly unpublished document is reproduced in Appendix B with the author's permission.

In effect, this proof applies to a *single fixed* network which might of course be a complex device. However, it should be noted that this does not outlaw some more sophisticated approaches involving nonlinear networks or networks which are *adjusted* in accordance with some combination of the incoming signal and its echo. The latter class of device is sometimes called a "linear adaptive system." Although such arrangements may be technically achievable, they too have up to now appeared to be too complex and expensive to justify their use.

5.4 *Echo Control in the Bell System*

The use of four-wire circuits, four-wire switching and careful office balancing in two-wire offices has essentially eliminated echoes except from the end links. Also, some return loss improvement has been achieved through impedance correcting networks. Thus, for circuits with round-trip delays under about 45 milliseconds, it is possible to assure acceptable echo conditions on about 99 per cent of the calls by engineering the long distance circuits to have enough loss.[8]

When round-trip delay gets beyond about 45 milliseconds, however,

the losses required would be unacceptable. Such circuits are operated at a low loss with echo suppressors. In general, these devices are used on most circuits over 1000 to 1500 miles in length. Under these circumstances, acceptable echo conditions are obtained on most calls, and the received volume of direct speech approaches the preferred value.

In principle, the echo suppressor is very simple and can be illustrated by the central echo suppressor invented by Clark and Mathes[6] in the early twenties. This device, shown schematically in Fig. 7, consists basically of two voice-operated switches. Speech from A traversing the upper four-wire path operates switch X_2, which disables the lower return path for the time necessary to prevent the return of echoes generated at the B end. Similarly, speech from B operates the switch X_1 and blocks echoes generated at the A end.

It is apparent that such a device may well interfere with normal conversational patterns. When the device is fully effective against echoes, it modifies the transmission system so that it is no longer a full duplex link capable of carrying intelligence in both directions simultaneously. Instead, it approaches a half duplex system which can be used in either one direction or the other but not in both at once. This will interfere to some extent with the free two-way flow of conversation, especially in those cases noted in Section IV where talk spurts overlap, most often caused by one talker trying to interrupt the other.

The echo suppressors now commonly in use are considerably more sophisticated than indicated by the simple model described above. Since these devices are discussed in detail in a companion paper, it suffices here to point out that they have been used with considerable success over a period of 30 or more years with the round-trip delays of 50–100 milliseconds currently encountered. We should also add that an echo
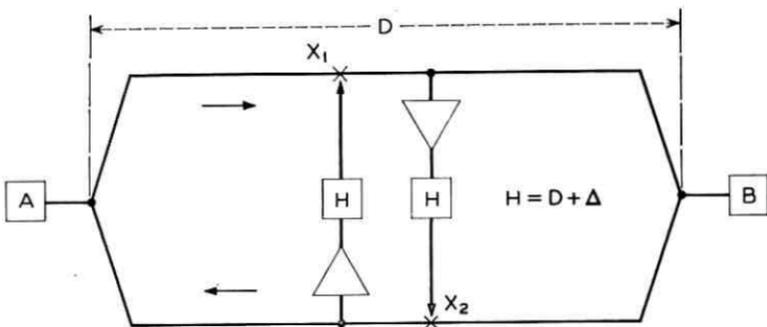
Fig. 7 — Central echo suppressor; H is slow release time of switch.

suppressor is essentially a compromise device. There is theoretically no
great difficulty in suppressing echoes. The problem arises in accomplish-
ing this while working with the wide variation in circuit characteristics,
noise, and speech volume of the commercial telephone system. At the
same time, there should be little interference with the rapid interchange
of ideas which gives telephony so much value. Many of the requirements
tend to be conflicting, and good design involves weighing the relative
effects of a considerable number of subtle psychological factors.

Additional problems arise when several echo suppressors are used in
tandem. Obviously the probability of interfering with normal conversa-
tional patterns increases with the number of suppressors used on a
circuit. A particularly serious interference can occur where two or more
complete suppressors are used with delay between them. In such a case
it is possible to completely block both directions of transmission. This
phenomenon is called "lock-out."

Lock-out can arise with the situation illustrated in Fig. 8. Let us as-
sume that A has momentarily stopped talking and the transmission path
is functioning in each direction. If B starts talking he will first operate
switch 3 and at time D later he would also operate switch 1. However,
if talker A resumes talking before time D his path is open to his sup-
pressor and he will operate switch 2. Now both directions of transmis-
sion will be blocked and remain so until one party stops talking. Since
neither one knows that the other is talking the "lock-out" may continue
for a long period.

It should be noted that the important factor in producing lock-out is
the delay D between X_2 and X_3. This delay provides a time storage which
allows each talker to independently capture control of the suppressor
nearest to him.
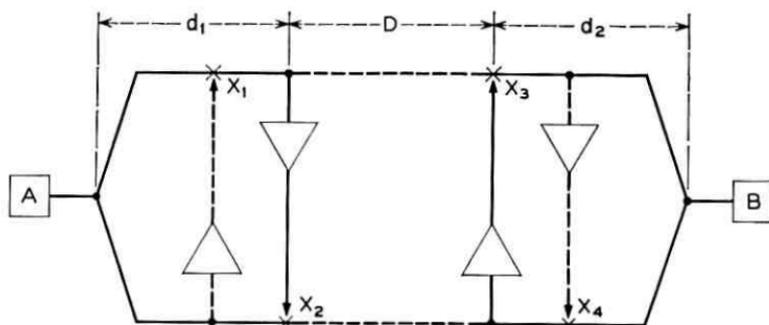
Fig. 8 — Multilink switched connection with two complete echo suppressors —
lock-out can occur.

Lock-out was studied at length by A. W. Horton in the 1930's.[9] He showed how the probability of its occurrence can be calculated from a knowledge of the circuit and speech parameters.

Some protection against lockout can in theory be provided through echo suppressor design, but in the long run it may be more desirable to avoid the use of tandem suppressors by suitable switching arrangements.

VI. CONCLUSION

In resuming the work on echo and delay, we found ourselves with a solid background of fundamental knowledge but were faced with a number of new problems. Perhaps first in importance was the determination and evaluation of the effect of using delays far longer than those experienced hitherto. Second, and closely related, was the question of designing suppressors to control these larger delays, and particularly the application of modern techniques in the expectation that they would provide more effective devices than those designed for the shorter delays. These are the main subjects of the papers which follow.

APPENDIX A

Further Notes on Return Loss

While return loss, in principle, can be measured at any point in a circuit, it is most conveniently specified and measured at the four-wire, two-wire junction. At this point, both the incident and the reflected wave can be determined from measurements on the transmitting side of the hybrid using a signal source on the receiving side. The relative strength of the wave delivered to the two-wire line is determined by opening the network: that is, producing a complete unbalance. The relative strength of the echo is then determined by connecting the network and measuring again. The difference between these two measurements, both expressed in db, is the return loss. If the network exactly balances the impedance of the line, there will be no reflected wave, and the return loss will, of course, be infinite.

It should be noted that it is the difference between the measurements made with the network leg open and terminated in a network that expresses the return loss; not the total loss from receiving to the transmitting side. This latter, or transhybrid, loss exceeds the return loss by the losses inherent in the hybrid coil. Referring to Fig. 4, the incident

energy at E will be seen to be lower than at B by the coil loss B to E (ideally 3 db and practically about 3.5 db). Similarly, the reflected energy at C is lower than that at E by the coil loss. Thus, the transhybrid loss B to C exceeds the return loss by twice the coil loss (about 7 db in practice). If amplifiers were introduced to just compensate for the coil losses, the difference in levels between A and D would be the return loss.

Return loss may also be expressed in terms of impedances. At the hybrid

$$\text{return loss} = 20 \log \left| \frac{Z_L + Z_N}{Z_L - Z_N} \right|$$

where Z_L and Z_N are the line and network impedance respectively. At any point in the two-wire circuit the return loss is also defined as above, where Z_L and Z_N are interpreted as the sink and source impedances.

Return loss is a function of frequency. From the standpoint of echo, the frequencies between 500 and 2500 cps are of most importance. Often a band of resistance noise covering this band is used as a source for measuring "average" return loss.

It will be appreciated that the reflections which give rise to echo can also cause oscillation or singing if the gain around the four-wire loop equals the loss at any frequency. Usually the high and low frequencies are controlling and return loss at the most critical frequency is referred to as a "singing" return loss to distinguish it from the loss at the frequencies important from an echo standpoint.

Return loss is often used to describe impedance characteristics and the uniformity of transmission lines. It is not practical to manufacture all cable pairs with precisely the same mutual capacitance; there will be some variation with length and from pair to pair. Similarly, it is not practical to introduce loading coils at precisely the theoretical spacing; some departures will be required to fit the geographical situation. It is customary to describe the effects of such variations in the impedance of the facility in terms of the "structural return loss." This is the return loss of an infinitely long (or a suitably terminated) practical cable compared to the impedance of an ideal cable with no irregularities. Typically, structural return losses run about 20–25 db for exchange plant circuits. Toll grade cable traversing areas which place few restrictions on the location of loading coils may have structural return losses of about 30–35 db. This is much higher than the 11–15 db typical of local return losses.

APPENDIX B

Proof That a Self-Balancing Hybrid Cannot Be Constructed with Linear Networks

By F. B. Llewellyn

In hybrid coils, the degree of balance that may be realized between the balancing network and the attached load impedance is one of the fundamental limitations in the design of telephone systems. From time to time, the proposal has been made that some means of producing an automatic self-balancing coil should be sought. Various schemes for accomplishing this have been suggested. Essentially most of them, in one way or another, involve the idea that it should be possible to insert vacuum tubes in such a way that they measure the current and voltage associated with the line impedance and introduce compensating values into the arm containing the balancing network. In this way, it was hoped to maintain the hybrid balance regardless of changes in the line impedance.

The following offers a proof that this result is theoretically unattainable with any linear network† whatever, regardless of whether it contains amplifiers, negative impedances, gyrators or any other linear elements, active or passive, that have been conceived in the past or may be thought of in the future. Of course, this does not rule out the possibility of obtaining operating improvements by certain *nonlinear* devices, and voice-operated echo suppressors provide a good example of such a nonlinear device.

As a starting point in the proof, Fig. B1 on the accompanying sketch shows the generalization of the configuration to be considered. The box contains the network elements of the hybrid itself while the four pairs of terminals represent, respectively, the attached line Z_L, the balancing network Z_N, the transmitting impedance Z_T, and the receiving impedance Z_R. Signal generators may exist in all of these except the balancing network Z_N. Consequently, the figure may be generalized even further, as shown in Fig. B2, by allowing Z_N to be included as part of the network within the box.

---

* This proof was prepared as a Bell Telephone Laboratories internal document dated October 13, 1950 and is published here with the author's permission.

† Note added by author in 1963: "With the exception of the line impedance Z_L all of the other impedances in the network are taken to be independent of time. Though implied in paragraph 1, this point has caused some confusion."
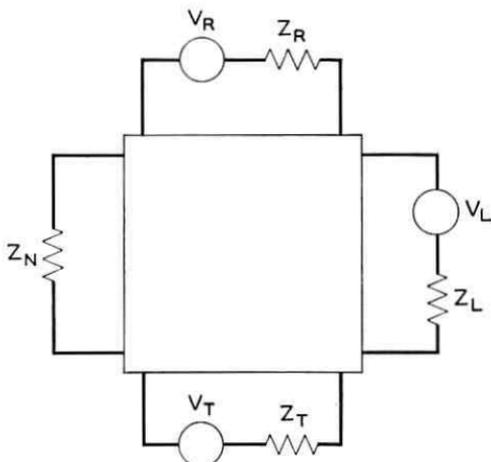
Fig. B1 — General prototype of hybrid terminations.

The conditions for an operative hybrid balance are three in number and may be set down as follows:

1. When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_T is zero.

2. When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_L is not zero.

3. When the system is driven from V_L (that is, when $V_R = V_T = 0$), the current I_T is not zero.

The requirement for self-balance is that these three balance conditions
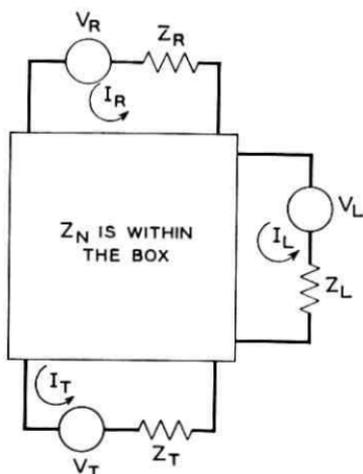
Fig. B2 — Equivalent of Fig. B1 for present analysis.

shall remain in force regardless of changes in the value of the line impedance Z_L.

Let us see what these conditions mean when applied to the general linear equations which describe the external behavior of the configuration in Fig. B2. These may be set down as follows:

I_R	I_L	I_T	
$Z_{11} + Z_R$	Z_{12}	Z_{13}	V_R
Z_{21}	$Z_{22} + Z_L$	Z_{23}	V_L
Z_{31}	Z_{32}	$Z_{33} + Z_T$	V_T

$$(1)$$

where the Z's with double subscripts are general impedance constants representative of the self-impedances and transfer impedances of the network within the box. Thus they are independent of the terminating impedances Z_R, Z_L, and Z_T.

When the system is driven from V_R (that is, when $V_L = V_T = 0$), the current I_T becomes, from (1)

$$I_{T_R} = \frac{V_R}{\Delta} [Z_{21}Z_{32} - Z_{31}(Z_{22} + Z_L)] \tag{2}$$

where Δ is the determinant of the matrix in (1). According to the first of the three balance conditions, this current must be zero, which requires:

$$Z_{21}Z_{38} - Z_{31}(Z_{22} + Z_L) = 0. \tag{3}$$

If self-balance is to be imposed then (3) must remain satisfied regardless of changes in Z_L. This can be accomplished only if *both* of the following are true:

$$Z_{31} = 0 \tag{4}$$

$$Z_{21}Z_{32} = 0. \tag{5}$$

To repeat this in words, we have found that the condition for self-balance requires that Z_{31} should be zero, and, moreover, that *either* Z_{21} or Z_{32} should be zero.

Let us see what these requirements mean in terms of the second and third of the three balance conditions. To do this, we need the equation for I_L when $V_L = V_T = 0$ and the equation for I_T when $V_R = V_T = 0$. From (1) these are, respectively:

$$I_{L_R} = \frac{V_R}{\Delta} [Z_{31}Z_{23} - Z_{21}(Z_{33} + Z_T)] \tag{6}$$

$$I_{T_L} = \frac{V_L}{\Delta} [Z_{31}Z_{12} - Z_{32}(Z_{11} + Z_R)]. \tag{7}$$

In the event that (4) is satisfied, these become

$$I_{L_R} = -\frac{V_R}{\Delta}[Z_{21}(Z_{33} + Z_T)] \tag{8}$$

$$I_{T_L} = -\frac{V_L}{\Delta}[Z_{32}(Z_{11} + Z_R)]. \tag{9}$$

Then when (5) is satisfied in addition, we see that I_{L_R} is zero in the event that $Z_{21} = 0$ and I_{T_L} is zero in the event that $Z_{32} = 0$.

It therefore must be concluded that the self-balance condition cannot be satisfied without violating either the second or the third of the three general hybrid balance conditions. Consequently, the proof is complete that a self-balancing hybrid cannot be constructed with linear networks.

REFERENCES

1. Blackwell, O. B., Time Factor in Telephone Transmission, Trans. A.I.E.E., **51**, 1932, pp. 141–147; reprinted as Bell System Monograph 619.
2. Brady, P. T., and Helder, G. K., Echo Suppressor Design in Telephone Communications, B.S.T.J., this issue, p. 2893.
3. Riesz, R. R., and Klemmer, E. T., Subjective Evaluation of Delay and Echo Suppression in Telephone Communications, B.S.T.J., this issue, p. 2919.
4. Norwine, A. C., and Murphy, O. J., Characteristic Time Intervals in Telephonic Conversation, B.S.T.J., **17**, April, 1938, pp. 281–291; reprinted as Bell System Monograph 1074.
5. Phillips, G. M., Echo and Its Effects on the Telephone User, Bell Laboratories Record, **32**, August, 1954, pp. 281–284.
6. Clark, A. B., and Mathes, R. C., Echo Suppressors for Long Telephone Circuits, Bell Telephone Laboratories Monograph 165, presented at A.I.E.E., June, 1925.
7. Noll, A. M., Subjective Effects of Sidetone During Telephone Conversation, to be published.
8. Huntley, H. R., Transmission Design of Intertoll Telephone Trunks, B.S.T.J., **32**, September, 1953, p. 1019.
9. Horton, Arthur W. Jr., The Occurrence and Effect of Lockout Occasioned by Two Echo Suppressors, B.S.T.J., **17**, April, 1938, pp. 258–280; reprinted as Bell System Monograph 1073.

Echo Suppressor Design in Telephone Communications

By P. T. BRADY and G. K. HELDER

An elementary echo suppressor is described which consists of a single voice-operated switch located at each end of a four-wire toll connection. It is shown that this circuit is unsuitable for telephone conversations if there is an appreciable delay between the speakers. Modifications and additions to the elementary echo suppressor are made in an effort to improve its performance. Action in the presence of break-in speech is discussed in some detail. The operation of four present-day echo suppressors is described.

I. INTRODUCTION

In a companion paper by J. W. Emling and D. Mitchell,[1] it was shown that the problems due to telephone echoes increase as the propagation time between speakers increases. As a result of the recent interest in satellite transmission, efforts have been made by several groups within and outside of Bell Telephone Laboratories to build echo suppressors particularly suited to the long delays encountered in such communications. In the first part of this paper, an elementary echo suppressor is described so that the reader may become familiar with some of the terminology used in this field and problems associated with the use of echo suppressors. Then, additions and modifications made to the simple echo suppressor to improve its performance are discussed. Finally, some of the more recent suppressor designs are illustrated. The echo suppressors to be described in the later sections of this paper are the same echo suppressors used in the subjective tests reported in a companion paper by Riesz and Klemmer.[2]

II. AN ELEMENTARY ECHO SUPPRESSOR

2.1 *Functional Description*

Assume a telephone circuit of the type shown on Fig. 1 is set up between two speakers A and B. The round-trip delay is appreciable, and is
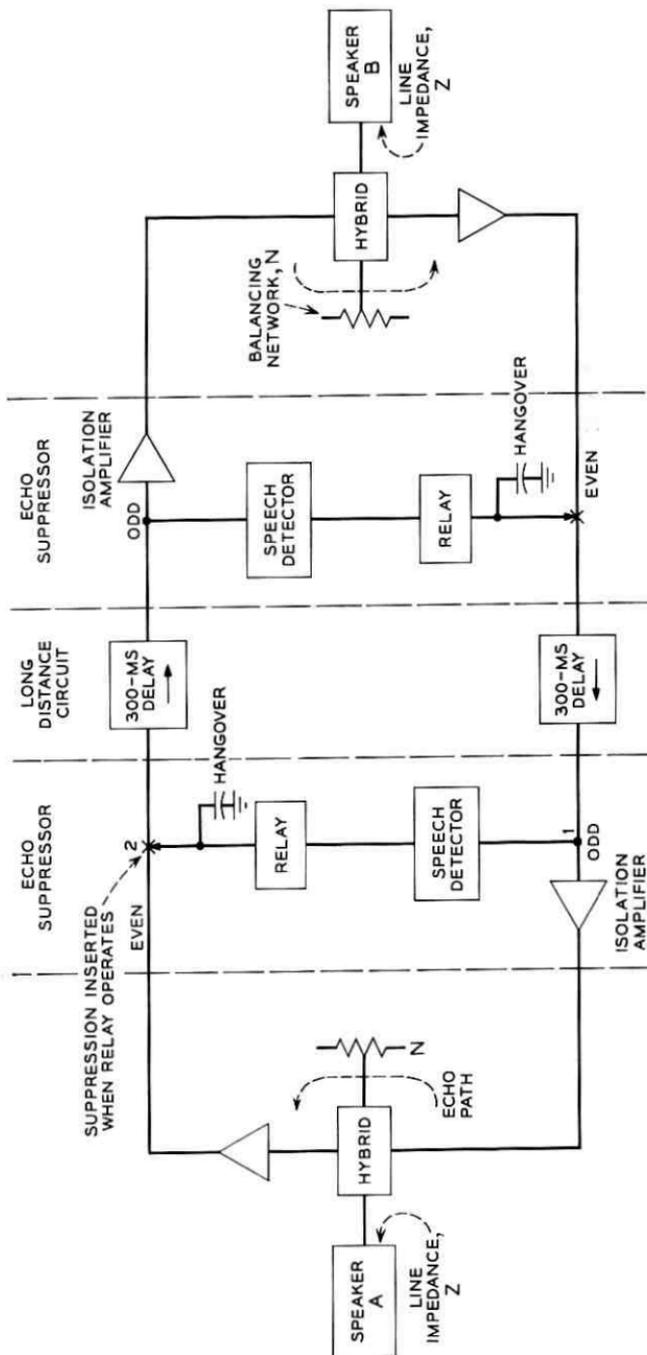
Fig. 1 — A simple echo suppressor.

shown for illustrative purposes as lumped and equal to 600 msec. Since the match of network N to the impedance Z of the 2-wire line at the hybrid is never perfect, echo is generated at the hybrids. If no devices were used to control the echoes on this circuit, it would be at best confusing and frustrating for the speakers if they attempted to converse. To block the echoes, a simple voice-operated relay or echo suppressor, as illustrated, can be installed at each end of the 4-wire portion of the circuit. Now, if speech from B appears at point 1 (generally called the "odd" side of the echo suppressor), the relay operates to open the line completely from A to B at point 2, the "even" side of the echo suppressor. These relays are effective in eliminating virtually all echoes, but they introduce other difficulties.

Assume that speaker B is talking and speaker A is silent. The transmission path from A to B will be blocked at point 2. If A now begins to talk at the same time B is talking (a condition called "double talking"), it is apparent that A cannot interrupt B's speech no matter how loud he talks. However, A has no way of knowing that his interrupting speech has failed to get through to B, and he may think that B is being impolite in not responding to his speech. This type of degradation would occur even if there were no delay between the speakers, and by itself could be a serious drawback of this echo suppressor design.

The delay coupled with the suppressor action causes an additional effect which is more pronounced and is very irritating. Assume B talking, A silent. At time t_0, B pauses momentarily, then at time $t_0 + 0.3$ sec, A hears this pause, and immediately interjects a short comment, such as "Uh huh," or "Yes, of course." Because there is no speech at point 1 (Fig. 1), A's comment gets through to B. Assume now that B had resumed talking at time $t_0 + 0.2$ sec. At $t_0 + 0.6$ sec, A's comment arrives at the echo suppressor near B and suppresses B's continuing speech for the length of A's comment. Then, at $t_0 + 0.9$ sec, A detects a short interruption of B's speech.

Two disturbing events have thus occurred. First, B heard A's comment at an inappropriate time, after he had resumed speaking, and second, A heard a break in B's speech. Although the second event is intuitively the more disturbing one, the first could well cause B to stop talking so he could hear the interrupting speech. Normal conversation would then come to a halt and be replaced by a frustrating attempt by each speaker to establish what the other person was trying to say. This and similar sequences could occur many times in a conversation and result not only in annoying the speakers but also in causing them to waste time repeating their remarks and commenting on the bad circuit.

The intermittent interruptions which the echo suppressors introduce are commonly referred to as "chopping." Chopping is most serious when the conversation alternates rapidly between speakers, and is nonexistent if only one person is talking. It becomes more objectionable as the delay between the speakers increases.

Methods to reduce this objectionable chopping have been devised and will be discussed later, after some characteristics of the simple echo suppressor are further described.

2.2 Characteristics of a Simple Suppressor

2.2.1 Amount of Suppression

It has been found that in circuits with long round-trip delays (say 200 msec and longer) and with echoes at their present-day levels, considerable attenuation must be supplied by the echo suppressors to reduce echoes so they are not annoying. Typically, about 50 db or more attenuation is provided. The present trend is to use some type of solid-state switching to introduce the loss, but a relay which short-circuits the line is very effective. In most recently designed echo suppressors, the attenuation introduced is so great that for all practical purposes the transmission path is completely blocked. Some suppressors, however, insert only a moderate amount of loss if the speech signal at point 1 of Fig. 1 is very weak, and insert larger amounts of loss for strong speech signals. They employ a "variolosser" which inserts a loss whose magnitude is a smooth function of the controlling signal. This would appear to be a good method of applying suppression since only the amount required is present at any time, and hence double talking should be easier. However, the variolosser provides satisfactory echo suppression only when it begins to insert loss on very weak speech signals. For signals of more typical magnitudes, at least 50 db suppression is required, and hence the variolosser is not much different from a relay. Further discussion of the operation of an echo suppressor will be with reference to relay operation, although an echo suppressor employing variolossers will be described later.

2.2.2 Sensitivity

The local sensitivity of the echo suppressor can be defined as the level of a 1000-cps tone, applied at the odd input, just sufficient to cause the speech detector to operate the relay. The signal level is specified as a power level in dbm working into a 600-ohm load. The local sensitivity includes within it an adjustment for the transmission level point at

which the suppressor is operated. Since it is much simpler to discuss its operation at the zero transmission level point (0TL),* all references to sensitivity hereafter will apply to the zero-level sensitivity, i.e., the sensitivity at 0TL.

If the speech detector shown in Fig. 1 has a flat frequency response, it will generally provide marginally acceptable suppression of long delayed echoes if its 1000-cps sensitivity is about −32 dbm. Greater (lower-level) 1000-cps sensitivities are required if the frequency response of the detector is weighted in such a way that it is less sensitive to frequencies below and above 1000 cps.

It would seem advisable to build a suppressor which has a very high sensitivity to assure complete suppression of echoes of all speech sounds. But if the sensitivity is too great, the suppressors will operate on noise as well as speech, which certainly is not desirable. Recent data on noise on telephone circuits indicate that less than 1 per cent of the longer toll calls have a noise level at 0TL greater than about 51 dbrnc.† This equates to −37 dbm if the noise is flat over the voice spectrum. Thus the sensitivity of an echo suppressor with flat bandwidth should not be greater than −37 dbm if operation on noise less than 1 per cent of the time is desired.

The sensitivity for adequate suppression and for minimal noise operation depends on the shaping of the dectector sensitivity characteristics. Previous studies have shown that a flat bandwidth between 500 to 3000 cps with reduced sensitivity is preferable for adequate suppression to a bandwidth with higher sensitivity at some of the frequencies in the band, if both suppressors are set for equal noise operation. The flat bandwidth helps to improve the sensitivity for initial consonants from many talkers.‡

2.2.3 Pickup Time

By definition, the pickup time is the time required for the suppressor to operate after the receipt at the odd input of a 1000-cps signal with a power level 3 db greater than the sensitivity. The pickup time should be short, since a long pickup time would cause the speaker to hear brief spurts of echo. These would sound like short "blips" and could be

---

* The zero transmission level point is a point to which all level points in a toll system can be referred. It is analogous to citing altitude by referring to height above sea level. The zero-level point is at the transmitting toll switchboard of the system under consideration.

† 51 dbrn as measured on a 3A noise meter with C message weighting.

‡ See Ref. 3, page 1458.

annoying even if they were not recognized as echo. A pickup time of about 5 msec is satisfactory, but this may produce operation on impulse noise, which again is not desirable. However, the bad effects of impulse noise operation can be minimized by proper control of the hangover, as discussed in Section 2.2.4.

2.2.4 *Hangover*

When the speech signal appearing at the odd input ceases, the suppression relay should remain operated briefly, for two reasons:

1. The speech signal may not terminate abruptly, but rather may contain low-level energy, such as that provided by a fricative. This speech may be below the sensitivity of the suppressor, but it should still be suppressed.

2. The echo may appear at the even side after speech appears at the odd side, due to the delay in the telephone circuit between the echo suppressor and the near telephone set (commonly called end delay).

To suppress these weak speech endings or delayed echoes, the echo suppressor is supplied with a slow release, or hangover time, which holds suppression after the speech level at the odd input has fallen below the sensitivity. The required hangover is dependent on at least three factors:

1. the end delay,
2. the echo suppressor sensitivity, and
3. the total circuit delay.

Generally, more hangover is required for long end and circuit delays and for lower sensitivities. No one hangover can be said to be acceptable in all cases. In the Bell System, where circuit delays of 30–40 msec and end delays of 20 msec are typical, the Western Electric 1A echo suppressors, having a detector with a shaped frequency response and a 1000-cps sensitivity of -31 dbm, have a suppression hangover of 50 msec.

If impulse noise operates the echo suppressor as described above in Section 2.2.3, the suppression hangover could chop appreciable portions of the speech if full hangover were applied for impulses. Most impulse noise is of very brief duration, and therefore the adverse effects of impulse noise operation can be minimized by providing a deferred or variable hangover which reaches full value only for speech sounds longer than about 50 msec.

2.3 *Full vs Split-Terminal Echo Suppressors*

In the suppressor just described the echo suppressor is split, with half located on either side of the major delay. The configuration shown

in Fig. 2 can also be used. In this case the echo suppressors for both speakers are located at one terminal. The echo suppressor indicated as 1 operates in an identical manner to the one described previously, but a very long hangover is required for echo suppressor 2 because the delay between speakers causes the echo of party A to appear at the suppression point long after the speech appears at the detection point. The combination of echo suppressors 1 and 2 is known as a full echo suppressor. A full echo suppressor is frequently used for circuits in which the round-trip delay is not very great (e.g., 50 msec or less). The hangovers required for full echo suppressors on circuits with long delay considerably increase the difficulty in conversing.*
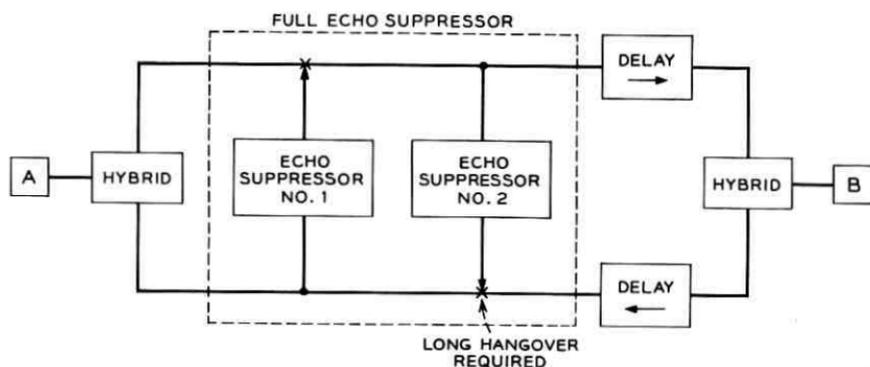
Fig. 2 — A full echo suppressor installed at one terminal.

The echo suppressors shown in Fig. 1 overcome this difficulty, and are called split-terminal echo suppressors because the functions of the full echo suppressor have been delegated to two units, located at the terminals of the intertoll trunk. Since the full echo suppressor requires only one installation, it has obvious economic advantages over the split-terminal echo suppressor, but because of its unsuitability for use in long-delay circuits, almost all recent design efforts have been devoted to developing an improved split-terminal echo suppressor.

III. DIFFERENTIAL ACTION

The conversational difficulties introduced by the echo suppressors of Fig. 1 arise from the failure to take into consideration the rapid inter-

---

* Years ago a full suppressor was sometimes installed midway between the speakers, and the long hangover previously required for one echo suppressor was divided between the echo suppressors. Such use today is in most cases impractical, since many circuits are multiplexed from end to end and revert to voice frequency only at the terminals.

changes of speech and the occurrence of simultaneous or double talking. The simple design which provides for silencing the even side at all times that speech appears at the odd side is not sufficient to provide a good communication path. A modification must be found which allows the echo suppressor to operate in a different mode if "break-in" speech is present at the even side at the same time that speech appears at the odd side, i.e., when double talking is taking place.

Ideally, during periods of double talking, the suppression should be removed so that free to and fro conversation can take place. Of course, when the suppression is removed, the echo is also transmitted with its degrading effects. The problems associated with double talking are discussed below.

3.1 Detecting Break-in Speech

Recognition of break-in speech in the presence of distant party speech is a most difficult problem in echo suppressor design, and no completely successful solution of this problem has been achieved. The nature of the problem is illustrated in Fig. 3, which is a simplified schematic of one end of a long distance telephone connection. Speech received from B appears at point 1, and speech from A appears at point 2 along with the echo of B's speech.

A simple speech detector installed at point 1 will easily recognize B's speech, and only B's speech, since the isolation amplifier prevents A's speech from reaching this point. But there is no point at which A's speech appears by itself, except of course at A's transmitter which is,
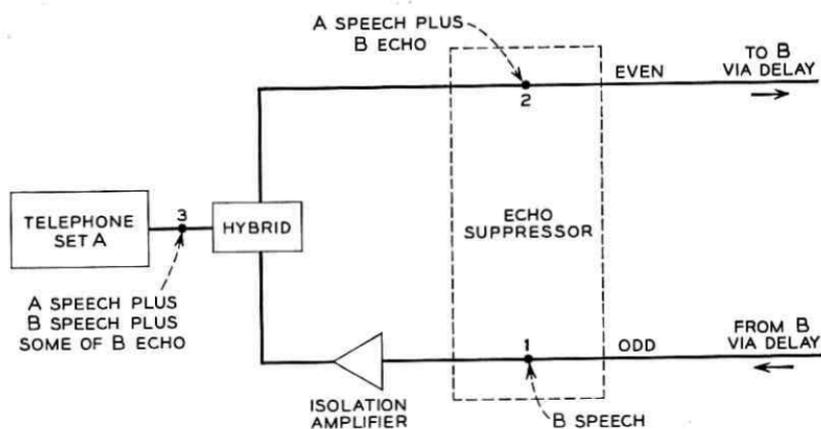

Fig. 3 — Speech signals at various points in a two-wire to four-wire connection.

unfortunately, inaccessible. A's speech is present at point 3, but B's speech is also present here, traveling not only toward A's telephone set but also reflected from the telephone set toward the hybrid. Furthermore, point 3 is generally inaccessible. It may be located miles away from the echo suppressor and is frequently not permanently associated with the same trunk on which the suppressor is installed. It is apparent that point 2 is the only point which can conveniently be examined for A's speech.

A simple speech detector, such as that used for suppression, installed at point 2 to detect break-in speech and remove the suppression, will operate on echo as well as local speech and is thus inappropriate as a local speech detector. However, many properties of the echo are known, because the echo is a distorted reflection of the signal at point 1. All existing break-in speech detectors compare a signal derived from point 2 with a signal derived from point 1. This "reference signal" derived from 1 is a measure of what the echo is expected to look like. If the signal at 2 possesses only those characteristics which the echo would contain, that is, if the signal derived from 2 is contained within the reference signal, then echo only is assumed at point 2. If, however, the signal at 2 is sufficiently different from the reference, speaker A is assumed to be talking and the break-in detector is activated. A good break-in detector should be as sensitive as possible to the break-in speech, but it should not operate when echo alone is present. A break-in detector triggered by echo alone is said to exhibit "false break-in."

Many different break-in detectors have had as their basis for design the method outlined above. Two of these devices are described in more detail.

3.2 The 1A Break-in Detector

Fig. 4 is a block diagram of the suppression and break-in scheme of the Western Electric 1A Echo Suppressor. (This suppressor is more fully described in Section IV.) The speech signals from points 1 and 2 are amplified, rectified, and applied directly to the differential device as illustrated. The return loss at the hybrid has an average value of 15 db with a standard deviation of 3 db.[1] Thus, the echo at 2 is certain to be at least 6 db below the signal at 1 if both points are at 0TL. The differential comparator is therefore adjusted so that if echo of this level or less is present at 2, the differential will yield a positive output. When this happens, relay 1 blocks the even path at point 4 with a hangover of 50 msec and also blocks the negative output of the differential at point 3.
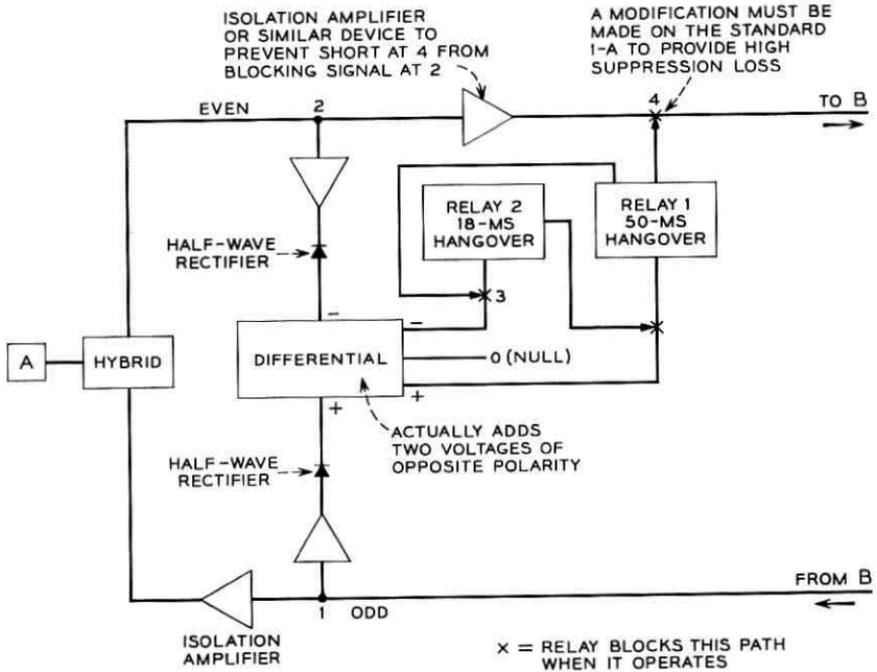
Fig. 4 — Block diagram of a 1A echo suppressor used in split-terminal fashion.

If speaker A begins to talk, the signal at 2 will occasionally override the differential and cause it to have an output on the negative side. If this occurs for 50 consecutive msec, relay 1 drops out and removes the block which it put in at points 3 and 4. Relay 2 now operates with an 18-msec hangover, assuring break-in for at least this hangover period. Break-in will continue until the differential becomes positive for at least 18 msec, in which case suppression will be restored. If no one is speaking, the differential has a null output and neither relay will be operated.

The "reference signal" referred to in Section 3.1 is in part the rectified signal derived from 1, but it also is contained in the hangover of the relays. For example, if B talks and then stops (A is silent), the echo may be delayed in reaching point 2. The echo could operate the differential for a brief period (up to about 20 msec) after speech ceases at 1, but the 50 msec hangover of relay 1 will prevent false break-in.

3.3 Break-In Detection by Means of Smoothed Rectified Speech

The majority of suppressors designed in the past few years employ a break-in scheme which uses a differential to compare two speech signals

which have been full-wave rectified and smoothed, as opposed to the 1A differential action which employs negligible smoothing. Fig. 5 shows a typical full-wave rectified speech signal which could appear at the odd side of the echo suppressor. The echo at the even side of the suppressor is certain to be at least 6 db (one-half amplitude) below the speech on the even side. It can also be delayed up to a maximum of about 20 msec. This "worst case" of echo is shown, full-wave rectified, as the dashed line of Fig. 5. Because of the delay, there are many times in which the echo signal exceeds the odd speech signal. But if the original signal is smoothed with an RC network so the voltage decays to one-half value in 20 msec, the echo will never exceed this smoothed signal. The smoothed signal is the reference signal and is indicated in Fig. 5. In practice, the echo signal is also smoothed, but with a smaller time constant (any value less than the time constant of the other). If the signal from the even side should ever exceed the reference, the local speaker must have been talking, and the break-in detector is activated. The break-in detector is completely independent of the suppressor relay operation, as opposed to the 1A action.

A block diagram of the instrumentation is shown in Fig. 6. A negative signal of sufficient strength from the differential (a simple voltage adder) will trigger the threshold and cause break-in to be indicated. Once break-in speech has been detected, the suppression is removed and the echo suppressor is in the break-in mode. The manner in which the suppressor operates in the break-in mode varies with different suppressors, and illustrations of various break-in modes of operation will be given later in this paper.
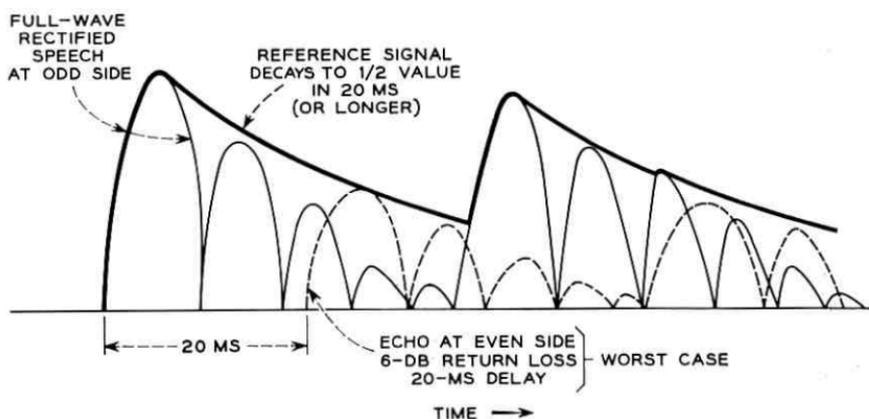

Fig. 5 — Typical waveforms appearing in rectified speech break-in detector.

3.4 *Sensitivity and Pickup of the Break-In Detector*

The parameters of the break-in detector will be discussed with reference to Fig. 6, with both the odd and even sides assumed at 0TL. These parameters are not, however, restricted to this circuit but apply to any break-in detector.

The break-in sensitivity is the 1000-cps signal power in dbm applied at the even input which will just cause the break-in detector to operate. This tone is applied with *no signal at the odd side*. A typical value is −32 dbm re 0TL. This sensitivity determines the ability to remove suppression during the suppression hangover after speech has disappeared from the odd side. It is also important in recognizing local speech just prior to the arrival of distant speech at the odd side. The break-in hangover (discussed later) will then prevent the local speaker from being immediately cut off.

The time required to operate the break-in detector or pickup time should be fairly fast, say 5 msec, but this is not so critical as suppressor pickup time. As shown before, long pickup times on the suppressor cause some echo to be returned. Long pickup times on the break-in detector cause initial fragments of speech sounds to be omitted or clipped if break-in is occurring when the suppressor is operated. Informal listening tests have shown that such front-end clipping is not very objectionable even when pickup times are as long as 20 or even 30 msec.

3.5 *Differential Sensitivity*

The differential sensitivity is a measure of the signal level at the even side required to operate the break-in detector in the presence of a signal at the odd side. A 1000-cps tone is applied to the odd terminal 3 db louder than the suppressor sensitivity. Then a 1000-cps tone is applied at the even side and is increased in level until break-in is detected. The difference in levels between the break-in signal and the odd signal is the differential sensitivity.

For example, assume the 0TL suppressor sensitivity is −32 dbm. A −29-dbm tone is applied at point 1 in Fig. 6. Break-in occurs when a −28-dbm tone is applied at the even terminal, point 2. The differential sensitivity is (−28) − (−29) db or +1 db.

The negative threshold detector requires a fixed voltage difference to operate. Its threshold is set to establish the break-in sensitivity by adjusting the amplifier on the even input, and the differential sensitivity is established by adjusting the indicated amplifier in the odd input. Because the threshold detector operates on a voltage which is the differ-
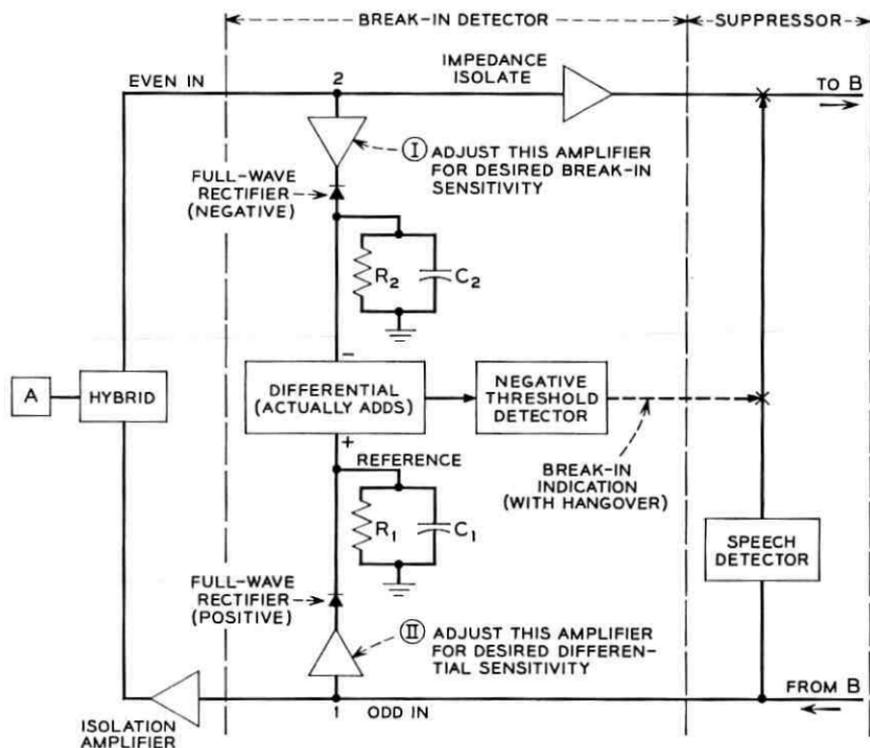
R₁ C₁ ADJUSTED TO DECAY TO HALF-VALUE IN
ABOUT 20 MS (MORE PRECISELY, $R_1C_1 > 28.9$ MS;
i.e. HALF-VALUE DECAY TIME IS AT LEAST 22.5 MS)

$R_2C_2 < R_1C_1$

Fig. 6 — Rectified speech break-in detector.

ence between the voltages at the even and odd inputs, if the differential sensitivity is 1 db for a -29-dbm signal at the odd input, the differential sensitivity will decrease for louder odd signals (but will still remain positive). For example, if the odd input is a tone of -15 dbm, the same voltage *difference* at the negative threshold detector will be required to overcome this signal, but the voltage *ratio* will be lower. This is why the definition of differential sensitivity must specify the magnitude of the test signal applied at the odd input.

3.6 *Break-In Hangover*

The break-in detector operates to remove suppression when the even speech is sufficiently louder than the odd speech. Because of this, it is apparent that loud speech at the odd input may prevent the break-in

detector from operating. In this case the echo suppressor reverts to the simple echo suppressor of Fig. 1, resulting in considerable speech mutilation. However, even when the odd speech volume is much greater than the even speech volume, because of the irregular nature of speech there are times when the even speech peaks will be sufficient to initiate break-in detector action and remove suppression. Shortly thereafter, the odd speech will once again be great enough to reinsert suppression. This alternate suppression and nonsuppression produces objectionable speech chopping.

To overcome this difficulty, a hangover is usually introduced on the break-in device. Once the local talker has broken in, he does not have to depend entirely on the differential, since the hangover will maintain break-in. Also, if he is talking and speech arrives from the distant talker, his speech will not be cut off by the suppressor for at least the break-in hangover period. (However, in the circuit of Fig. 6 if the distant speaker is talking and the local speaker breaks in, he does not have to wait for the suppression hangover to release before suppression is removed. The opposite is true in the 1A echo suppressor.)

If the break-in hangover is very long, practically no chopping will occur, but the break-in detector will remain activated when the local speaker has finished talking. During many of these times, distant speech will be present, and if transmission is allowed on the even side at these times, echo will be returned. This also can be very annoying on long-delay circuits.

The difficulties of achieving good break-in ability may now be summarized:

1. To avoid false break-in, the differential circuit must be made relatively insensitive to signals at the even side because the expected echo levels can be fairly high. This will cause chopping of local speech.

2. To avoid chopping, hangover is introduced on the break-in circuit. But this allows more echo to get through after the local speaker has stopped talking.

The break-in hangover is usually adjusted to effect a compromise between excessive chopping and echo. A typical value for recent designs is 150–200 msec. This is much longer than the 18-msec break-in hangover of the 1A suppressor. If a very good break-in detector could be built, the hangover would not need to be as long, and a better compromise between chopping and echo could be made. Also, if return losses could be improved (made greater), the existing break-in detectors could have even lower differential sensitivities, which would reduce the need for long hangovers. The accomplishment of either of these objec-

tives would be of considerable value in improving present-day echo suppressors.

3.7 *Asymmetry in Echo Suppressor Environment*

Little attention has yet been given in this paper to the speech levels at the suppressors. If each speaker is a "normal" talker and has relatively little loss between his subset and the suppressor, his speech will appear at 0TL at a long-term average level of roughly −15 dbm. There can be wide variations from this level. The speaker may be an especially loud or weak talker, or he may be prone (as many people are) to holding the transmitter under his chin, or there may be additional loss between him and the suppressor.

Fig. 7 is an illustration of a long-delay circuit in which two loud talkers with equal speech volumes are connected. The losses L_A and L_B account for all the speech level variations due to the factors described above. If L_A equals L_B, then each speaker has the same ability to break into the other's speech. Consider though, the case in which L_B equals zero and L_A equals 10 db. Speaker B will have more difficulty than A in conversing for two reasons:

1. The echo returning to A is 20 db less than that returning to B (assuming equal return losses at both hybrids). Therefore, B will hear louder echoes during break-in than will A.

2. A has 10 db more difficulty in breaking in than B; consequently B will hear more chopping.

If the losses L_A and L_B differ by 10 db because of the actual circuit configuration, as may happen, the circuit is said to have 10 db asym-
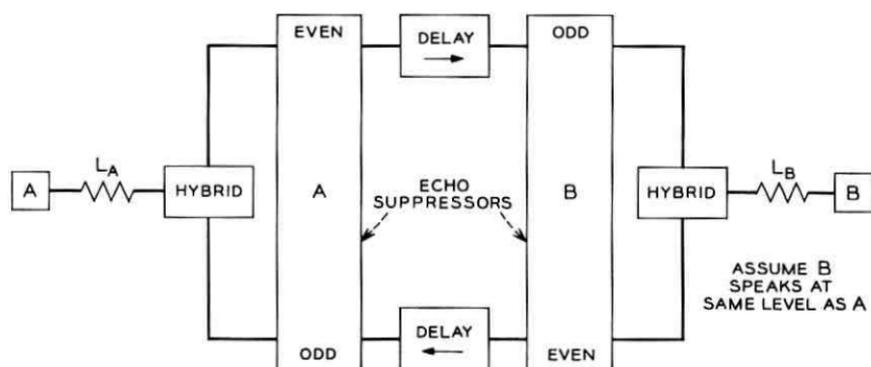

Fig. 7 — Asymmetry in an echo suppressor connection; the difference in losses L_A and L_B is a measure of the asymmetry.

metry. An asymmetry of only 10 db is sufficient to often cause one speaker to have considerable difficulty, whereas the other speaker may hardly notice anything wrong with the circuit. Differences in talker levels at the two ends, which may frequently occur, are also causes of asymmetry, much as if the physical losses differ.

IV. TYPES OF ECHO SUPPRESSORS

Four echo suppressors are described below to illustrate the various approaches taken to the problem of designing echo suppressors which permit double talking. All of the echo suppressors are similar in that they employ a suppression device and a differential circuit. However, each has some peculiarity which distinguishes it from the others.

4.1 *1A Echo Suppressor*

This echo suppressor has been standard in the Bell System since the late 1930's. About twenty thousand are presently in use on long distance continental and ocean cable circuits.

A simplified schematic of the 1A echo suppressor (used as a full echo suppressor) is shown on Fig. 8.* In the quiescent state with speech in neither the odd nor even path, the DA tube current flowing through the EM and OM relays is such that OM is operated and EM is not. If the tube current decreases, OM releases; if it increases, EM operates.

Consider speech in the odd path. Part of this speech enters the odd amplifier via the hybrid,† is amplified and half-wave rectified, and is then applied to the cathode of the DA tube. This makes the cathode more positive with respect to the grid, which decreases the tube current and releases relay OM. Release of OM removes ground from one winding of the OH relay and applies ground to the second OH winding, causing it to operate. Operation of OH places a ground return path in parallel with the EM relay, which prevents its operation and also removes ground from the hybrid balancing network N in the even path. Network N now balances the input impedance of the even amplifier and this provides a high (35 to 40 db) transhybrid loss across the hybrid in the transmission path which suppresses the echo. Suppression hangover is supplied by the RC network in one winding of the OH relay.

---

* The block diagram shown in Fig. 4 is for the split-terminal suppressor and is not a representation of the full suppressor of Fig. 8.

† The echo suppressor hybrids are branching devices for transferring energy from the speech path to the suppressor circuitry and should not be confused with the terminating hybrids that convert the 2-wire line to a 4-wire line.
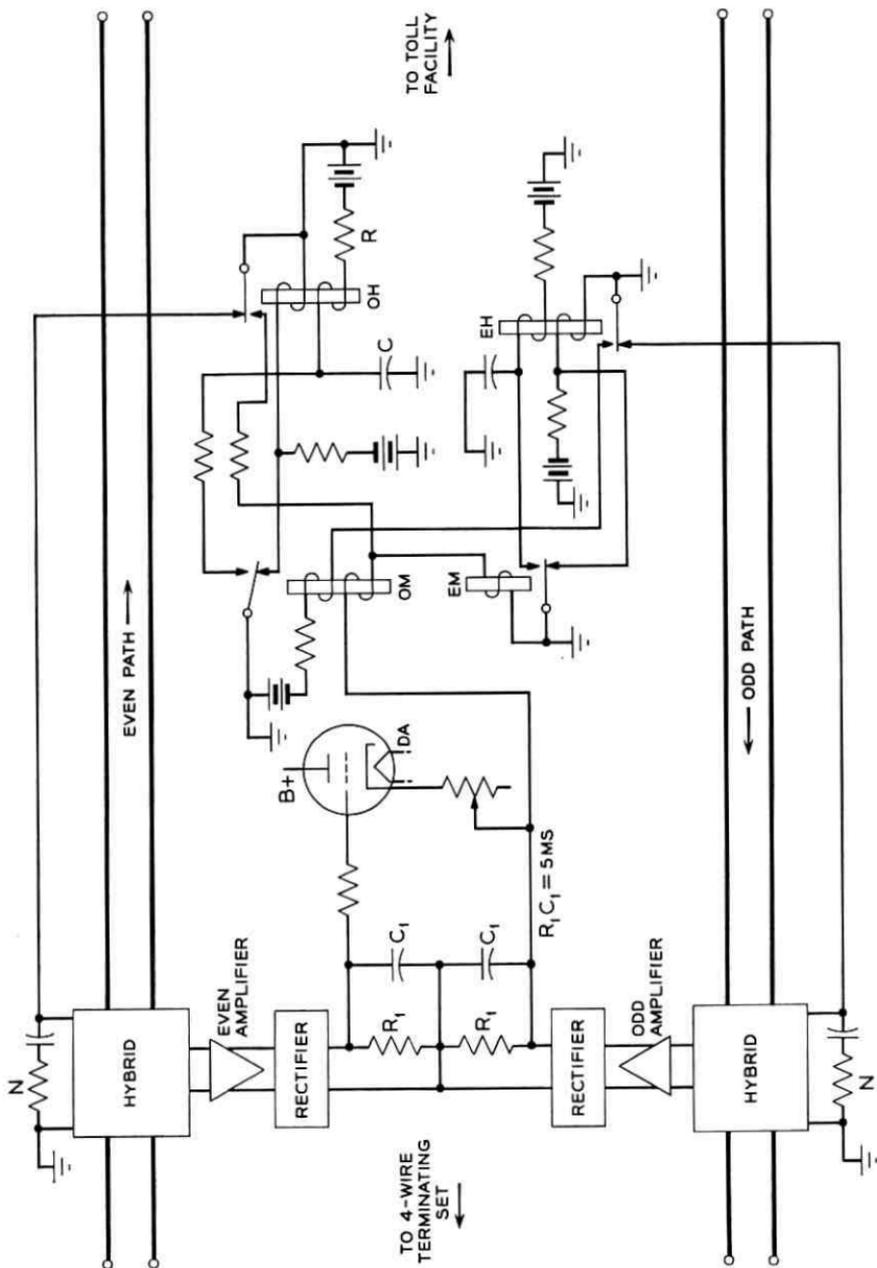
Fig. 8 — Simplified schematic of 1A echo suppressor.

If speech is present in the even path, the operation is similar except that the DA tube current is increased and the EM relay operates, causing the EH relay to operate and place loss in the odd path.

When speech is present in both paths during double talking, the echo suppressor operates as follows. Assume initially that speech in the odd path occurred first. This speech operates the OM and OH relays, suppressing the even path with a hangover of 50 msec. If the even speech level is less than the odd speech level, the suppression remains in. If the even speech level rises above the odd speech level for 50 msec, the OH relay releases and the EM and EH relays operate. This removes suppression in the even path but places suppression in the odd path with a hangover of 70 msec. For about equal-volume talkers, the suppression will alternate between the two paths as the syllabic or peak power points of the speech in both paths alternate. Almost all echo will be suppressed, but the speech during double talking will be chopped or mutilated.

Operation in the manner described above occurs with a full suppressor. Split echo suppressor operation is obtained by permanently grounding the balancing network in the odd path, with the hangover on the EH relay set to its minimum value of about 18 msec. Two echo suppressors with this modification are used on one circuit — one at each end. The operation during double talking is similar to that of a full echo suppressor.

The 35–40 db suppression supplied by the hybrid balance is sufficient for most circuits today, but the longer-delay cable circuits or future satellite circuits require more suppression than this. A modification (shown on Fig. 9) was made to the echo suppressor to increase the suppression for tests on long delay circuits. The OH relay, instead of balancing the hybrid, applies a short circuit across the even path to
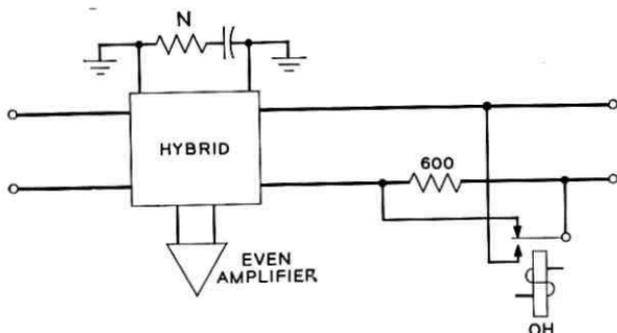
Fig. 9 — 1A echo suppressor modified for greater suppression.

suppress the echo. Two split suppressors modified in this fashion were subjectively tested on long-delay circuits at Bell Telephone Laboratories. The results of the subjective tests are given in a companion paper.[2]

4.2 BH Echo Suppressor

This echo suppressor, one of several experimental echo suppressors designed at Bell Telephone Laboratories, differs primarily from the 1A echo suppressor in that its suppression and break-in functions are independent of each other, while in the 1A these functions are combined into one differential circuit. A simplified schematic of the BH split echo suppressor is shown on Fig. 10.

If speech is present in the odd path, relay S operates and places a short across the even transmission path (provided relay B is operated) which suppresses the returning echo. Operation of suppression relay S is controlled by speech in the odd path only and is independent of speech in the even path. (This differs from the 1A where relay OM is controlled by speech in both paths as determined by the DA tube current.) Relay S provides a suppression hangover of about 50 msec.

Speech in the even path is amplified, rectified, and applied to the grid
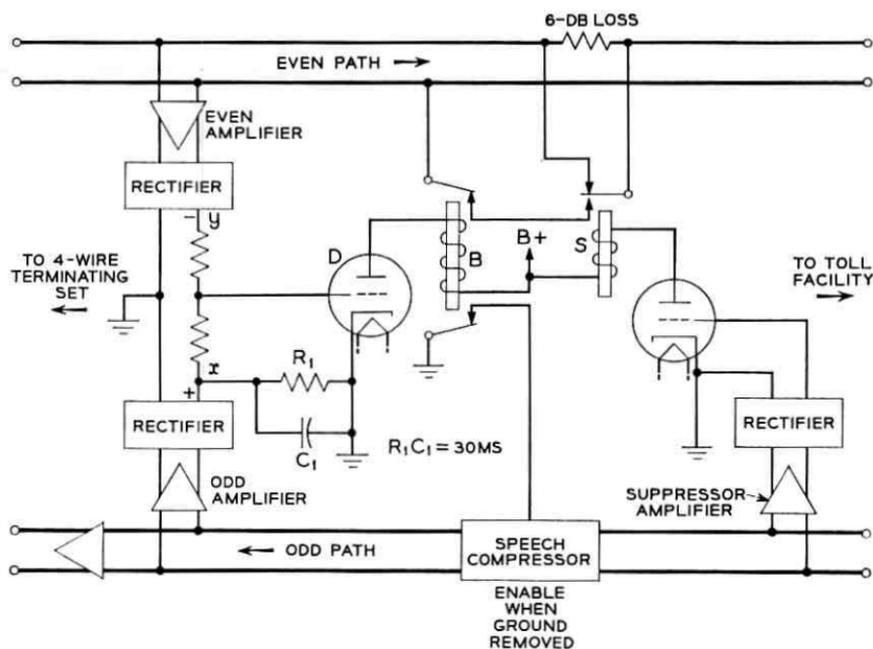
Fig. 10 — BH echo suppressor.

of tube D. The negative signal applied to the grid reduces the tube current which releases normally operated relay B. Release of B prevents or removes any suppression caused by operation of relay S and also removes ground to enable a speech compressor in the odd transmission path. (This is discussed later.) Once relay B has released, it remains released for a hangover time of 200 msec.

When speech is present in both paths, during double talking, the operation is as follows. Assume that speech in the odd path occurred first. This speech operates the S relay suppressing the even path with a hangover of 50 msec. If the even speech level is less than the odd speech level, relay B stays operated, and the suppression remains in. Since the even speech may be the echo of the odd speech, this speech must not operate relay B to remove suppression. Echo is prevented from operating relay B by proper choice of gains in the odd and even amplifiers and the R_1C_1 time constant. (This was discussed in Sections 3.1 and 3.3 above.) When true break-in speech of sufficient level is present in the even path, it will release relay B and remove the suppression for 200 msec. If relay S is operated when relay B releases, 6 db loss is inserted in the even path. Also, anytime relay B is operated, a speech compressor is enabled in the odd path. When the echo suppressor is used at 0TL, the speech compressor is adjusted to supply 0 db loss for a −50-dbm signal in the odd path and to smoothly increase this loss to 12 db for a 0-dbm signal. The 6-db loss in the even path and the speech compressor in the odd path are used to reduce the echo occurring during the double talking interval, when relay B is operated and suppression is removed.

If prolonged double talking is simultaneously present at the echo suppressors at both ends of the connection, all echoes would be attenuated by two compressors and two 6-db pads. This is not generally the case, however. Many echoes return at a time when the local talker is silent, and they are often reflections of speech generated when there was no distant speech. It is more accurate to say that all echoes are attenuated by at least one compressor and at least one 6-db pad.

4.3 *GN Echo Suppressor*

This split echo suppressor, also designed at Bell Telephone Laboratories, is unique in that it uses variolossers for suppression and thus applies a variable amount of suppression depending on the odd speech level. A simplified schematic is shown in Fig. 11.*

Speech present in the odd path is amplified, rectified, and applied to

---

* A more complete description of this echo suppressor is given in Ref. 4.
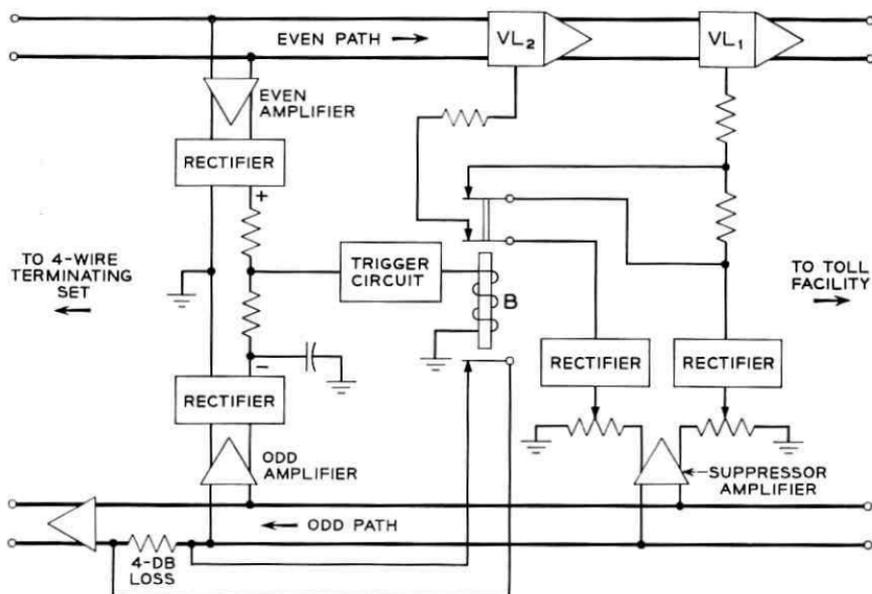
Fig. 11 — GN echo suppressor.

the variolossers VL_1 and VL_2. The combined suppression loss of VL_1 and VL_2 is shown on Fig. 12. It is seen that for very low-level speech the suppression slope is moderate, but for higher-level speech the suppression has a steep slope and rapidly reaches 80 db suppression. By proper choice of odd and even amplifier gains and the time constant on the odd input to the trigger circuit, echo is prevented from removing this suppression.

During periods of double talking, the operation is as follows. If odd speech is present first, suppression is applied. When break-in speech of sufficient amplitude then occurs, it is recognized as such at the input to the trigger circuit, which operates relay B. Operation of relay B removes the suppression of VL_2 to allow even speech to pass through the echo suppressor. It also reduces the suppression of VL_1 and places a 4-db loss in the odd path to reduce the echo. The suppression of VL_1 during double talking is shown on Fig. 12. Relay B provides a break-in hangover of about 400 msec.

4.4 *AM Echo Suppressor*

This echo suppressor, an experimental model by a United States manufacturer, differs primarily from other echo suppressors in its action
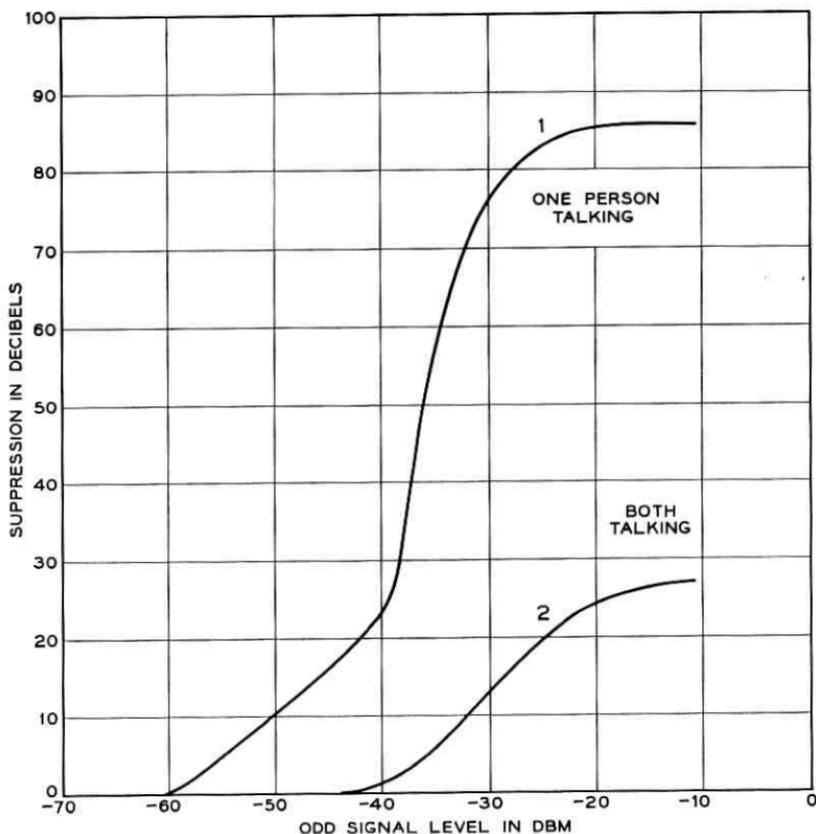
Fig. 12 — GN variolosser characteristics.

during double talking. A simplified block diagram of the AM echo suppressor is shown on Fig. 13.

If speech is present in the odd path, threshold detector TD_2 operates, producing a signal passing through gate 1 and enabling the transistor suppressor, which has about 60 db loss. Suppression hangover (about 100 msec) is supplied by hangover control HC_2.

Speech in the even path is amplified, rectified and applied to the differential amplifier. If the output of the differential amplifier operates threshold detector TD_1 as described in Section 3.3, gate 1 is inhibited, preventing suppression. The inhibit signal has a hangover time of about 200 msec.

When speech is present in both paths during double talking, the operation is as follows. Assume that speech in the odd path occurred first. This speech operates TD_2, suppressing the even path with a hangover
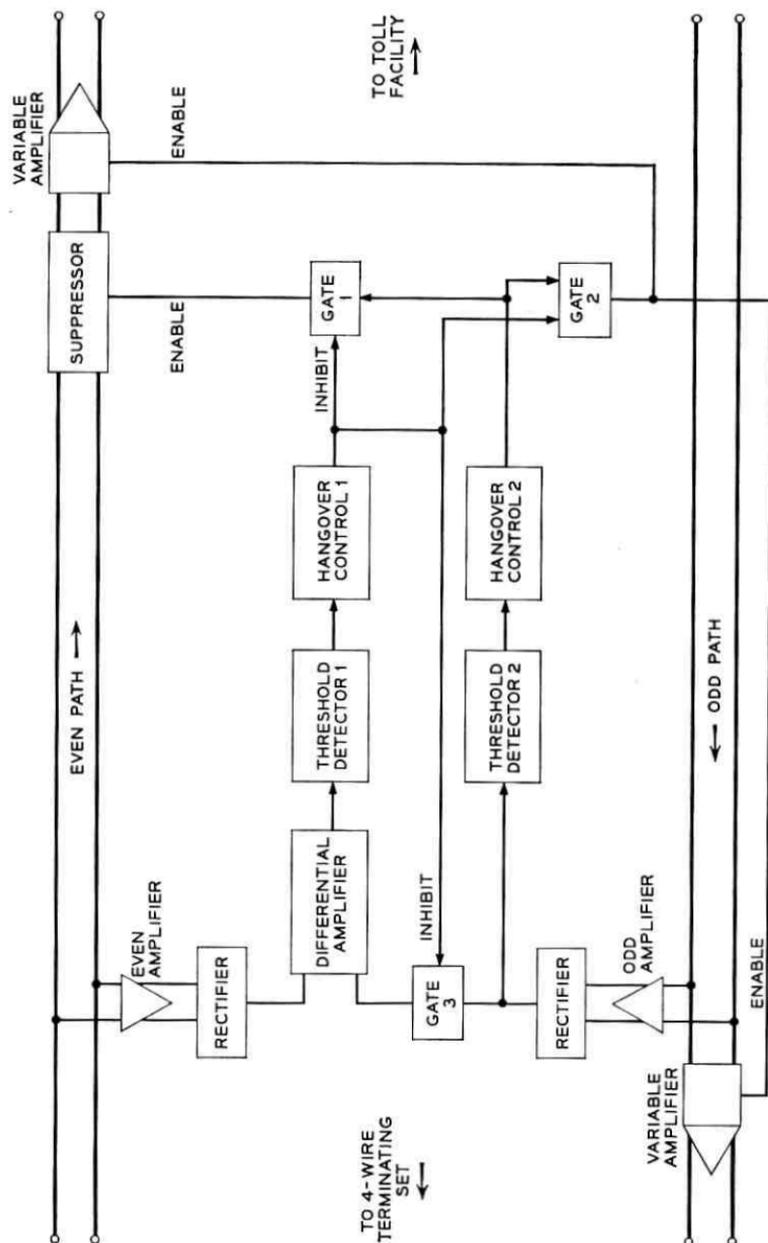
Fig. 13 — AM echo suppressor.

of 100 msec. Echo is prevented from operating TD_1 through use of proper time constants in the differential amplifier. When true double talking speech of sufficient volume is present in the even path, the differential amplifier will cause TD_1 to operate, removing suppression for at least 200 msec. TD_2 is still operated by speech in the odd path, and when TD_1 and TD_2 are operated, a 6-db loss is introduced in the amplifiers in the even and odd paths to reduce the echo. The operation of TD_1 also effectively inhibits a gate in the odd speech input to the differential amplifier. This inhibit signal prevents the release of TD_1 as long as the even speech is greater than the even sensitivity, irrespective of the speech present in the odd path. Thus, as long as the even speech level does not fall below the even sensitivity for over 200 msec (the hangover time of HC_1), gate 1 will always inhibit suppression no matter what the level of the odd speech. This action reduces the mutilation of even speech but also increases the unsuppressed echo, as discussed below.

Suppose the *odd* speech maintains an average level 10 db higher than the *even* sensitivity and the echo of the odd speech is reduced by an 8-db return loss such that the odd echo is 2 db higher than the even sensitivity. This echo will not by itself remove suppression because of the comparison action of the differential amplifier. However, if a short burst of break-in speech is present in the even path, TD_1 will operate to remove suppression and inhibit gate 3. Now the echo of the odd speech is sufficient to maintain TD_1 on, and gate 1 will be inhibited until the echo of the odd speech falls below the even side sensitivity for more than 200 msec. This action has been observed to produce considerable echo for some combinations of talkers and return loss.

V. CONCLUSIONS

In this paper we have discussed some of the design features of a particular type of echo suppressor. The philosophy behind the design of this type of echo suppressor is that a telephone channel should allow as much two-way conversation as possible consistent with proper echo control. Thus every effort is made to provide ease of break-in during double talking. Unfortunately, this generally results in a compromise between having too much echo or too much speech chopping during double talking. The echo and chopping are very disturbing to some people, and the disturbing effects may, in some cases, outweigh the beneficial aspects of attempting to provide a two-way circuit during simultaneous talking.

Ease of break-in is not the only principle which can be followed in designing echo suppressors. Other approaches have been suggested. For example, one echo suppressor has been proposed which is designed to train the conversants not to double talk by increasing the received volume of the distant speaker whenever the local speaker tries to interrupt, thereby "shouting down" the interrupter. Another proposal is aimed at preventing the chopping effect during double talking by allowing only one-way conversation at any time, the allowed direction being determined by examining which conversant spoke first.

All of these proposals include voice switching, and all produce transmission degradations. It is impossible to predict the subjective reaction to the degradation introduced by these various proposed echo suppressors merely by examining the design features of the suppressors. This reaction can be determined only through subjective tests, one type of which is described in a companion paper.

REFERENCES

1. Emling, J. W., and Mitchell, D., The Effects of Time Delay and Echoes on Telephone Conversations, B.S.T.J., this issue, p. 2869.
2. Riesz, R. R., and Klemmer, E. T., Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications, B.S.T.J., this issue, p. 2919.
3. Miedema, H., and Schachtman, M. G., TASI Quality — Effect of Speech Detectors and Interpolation, B.S.T.J., 41, July, 1962, pp. 1455–1473.
4. Gardner, Mark B., and Nelson, John R., Combating Echo in Speech Circuits with Long Delay, J. Acoust. Soc. Am., 35, Nov., 1963, p. 1762.

Subjective Evaluation of Delay and Echo Suppressors in Telephone Communications

By R. R. RIESZ and E. T. KLEMMER

(Manuscript received August 9, 1963)

The effect of transmission delay upon the quality of a telephone circuit was investigated using naturally occurring telephone conversations. Round-trip delays of 600 and 1200 msec went almost unnoticed at first when no echo sources or echo suppressors were present in the circuit. After exposure to pure delays of up to 2400 msec, considerable dissatisfaction, indicated by rejection of the circuit, developed with the 600- and 1200-msec delays. When echo sources and echo suppressors were added to the circuit, some dissatisfaction developed immediately for round-trip delays of 600 and 1200 msec. No adaptation to the effects of delay occurred; in fact dissatisfaction increased with experience under certain conditions.

I. INTRODUCTION

The preceding two papers describe the nature of the problems introduced into telephone circuits by delay and echo and the attempted solution of these problems through the design of echo suppressors. The present paper describes recent determinations of the degradation of transmission quality caused by pure delay and delay plus echo and echo suppressors. The studies reported were all done by the human factors research department of Bell Telephone Laboratories.

Other groups have studied the effect of delay upon voice transmission. These include the research department of Bell Telephone Laboratories, Stanford Research Institute, Italian Telecommunications Administration and the British Post Office. Unfortunately, none of this work has yet been published. In general, these studies have found little degradation with pure delay, even for round-trip delays over one second, although some objection did occur in one experiment with natural conversations with 1410-msec round-trip delay. These same subjects did

not find the delay objectionable during structured conversation, however.

All of the experiments described in the present paper were performed by introducing experimental circuits into naturally occurring telephone conversations. This method was chosen because of the previous finding, here and elsewhere, that the subtle conversational difficulties produced by delay and echo suppressor action do not often occur in structured conversations.

The general plan of the study was first to evaluate the effect of pure delays and then evaluate the effect of echoes, echo suppressors and delays for several different echo suppressors described in a companion paper.[1] When it became obvious in the first two experiments that continued exposure to long delays had a marked effect upon the user's reaction to shorter delays, a third experiment was conducted in which each subject was exposed for some weeks to only a single value of a given delay.

The primary measure of transmission quality in the present studies is the percentage of calls on which the circuit was rejected as unsatisfactory by the users for normal use. In addition, for some of the calls made under each condition, the users were called back and asked about the circuit.

II. SIMULATION APPARATUS: SIBYL

The simulator called SIBYL, which permits the insertion of experimental circuits into existing telephone lines, has been previously described by H. D. Irvin.[2] With this device it was possible to introduce controlled delay through the use of magnetic disc delay units (Echovox Sr.), echo suppressors and return loss to simulate field echo conditions. The simulator could be inserted into any call originated by any member of a panel of users without letting him know of its insertion.

III. EXPERIMENT 1: DEGRADATION IN TRANSMISSION QUALITY DUE TO ADDED PURE DELAY

3.1 *Apparatus*

For the pure delay tests a complete four-wire network was employed as shown in Fig. 1. Artificial sidetone was provided, and loss and noise were adjusted to values representative of the standard circuit. None of the calls were monitored; that is, complete privacy of conversation was maintained. For each call through the simulator, the originating number,
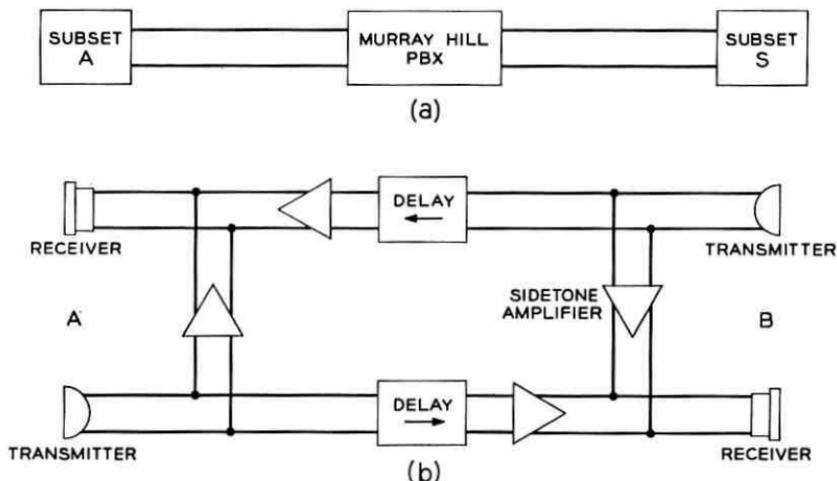
Fig. 1 — Circuit used to introduce delay in experiment 1: (a) standard circuit between PBX subsets, (b) experimental circuit between four-wire subsets used to introduce transmission delay.

called number and length of call were recorded as well as the time at which the call was rejected for the rejected calls.

3.2 *Subjects*

Eighteen members of the administrative staff at the Laboratories who called each other frequently were asked to serve as subjects. They were told that some of the calls that they originated would be routed over a simulated satellite circuit, but they did not know which calls would be affected or what the changes in the circuit were. They were instructed that if they found any circuit "unsatisfactory for normal telephoning" they could dial the digit "3" without hanging up or breaking the connection, and the standard circuit would be restored. Only the calling party was able to reject the circuit.

3.3 *Test Conditions*

Because of the necessity for a complete four-wire connection, the delay was inserted only on calls in which one of the subjects called another. The simulator was inserted on every such call unless it was in use by another pair of subjects. Since the simulator could handle only one call at a time, when one pair of subjects was routed through the simulator, all other subjects would receive the standard circuit even if they called each other.

The experiment continued for 12 weeks. The general plan of the experiment was to alternate between delays of 600 and 1200 msec each working day for 12 weeks. This schedule was changed in the fifth and sixth weeks, during which the 1200-msec delay was alternated each day with 2400-msec and the 600-msec delay was not used at all. The reason for inserting the 2400-msec delay in the fifth week was that there were almost no rejections during the first four weeks, which raised the question as to whether the subjects were not having any difficulty due to the delay or were not associating unusual conversational difficulties with the telephone circuit. It was correctly assumed that the 2400-msec delay would produce identifiable circuit-related difficulty. All but two of the 18 subjects made or received at least one call over the 2400 msec circuit during the two weeks it was used. The final six weeks at 600- and 1200-msec delay show the users reaction to these lesser delays after exposure to the longer delay. The trend during this six-week period is particularly important. The negative reaction of the users to the longer delay may be expected to generalize to the shorter-delay calls which are made in close time proximity. That is, after exposure to the 2400-msec delay, the users may reject any circuit on which they note any characteristics of a delay circuit. If this is the only effect, then the rejection rate should jump suddenly after exposure to 2400 msec and then return (perhaps slowly) to the former level when the 2400-msec condition is removed.

If, on the other hand, exposure to the longer delay teaches the users to identify conversational difficulties which are also present at shorter delays, then the rejection rate should not fall after removal of the longer delay. Indeed, it may well be expected to continue to rise as the learning continues.

3.4 *Results*

The percentage of calls rejected for each two-week period of the experiment is plotted in Fig. 2 for each delay separately. The grouping of weeks by twos provides an average of 29 calls per point with a minimum of 16 calls, whereas single weeks would have as few as six observations per point. None of the calls at 600 msec delay were rejected in the first four weeks and only 3 of 50 calls at 1200 msec were rejected by two different subjects. The 2400-msec delay was rejected more than half the time during weeks 5 and 6 when it was employed (18 of 34 calls). In the two weeks following exposure to 2400-msec delay there were still no rejected calls at 600 msec, but in the final four weeks 13 subjects
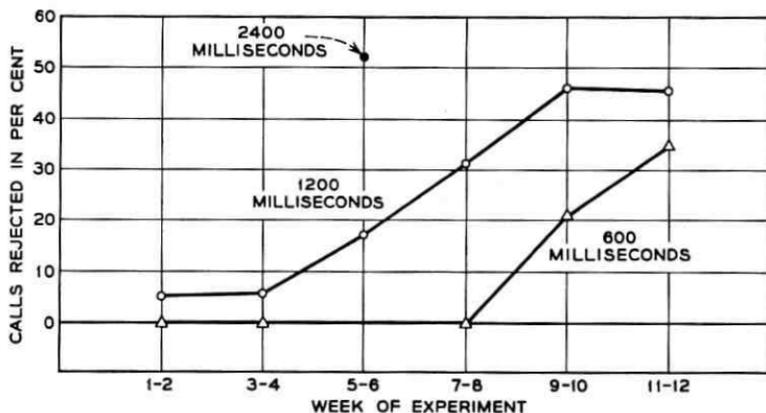
Fig. 2 — Percentage of calls rejected as a function of weeks for each pure delay condition separately. Combined data from 18 subjects, all of whom were exposed to all delays. Number of observations per point varies from 16 to 37. The 2400-msec delay was used during weeks 5 and 6 only. Rejections on standard circuit calls were less than 1 per cent.

made 61 calls at 600 msec and four of these subjects rejected a total of 17 calls.* The 1200-msec delay showed an increasing rejection rate after week 5, and during the last four weeks, 17 subjects made 66 calls at 1200 msec and eight of these subjects rejected a total of 31 calls.†

Because the circuit difficulties introduced by delay are not apparent until after some conversation has taken place, it is of interest to ask how long the call has been in progress before rejection occurs. The median time to dial out the simulator for all rejected calls was 22 seconds. The median length of rejected calls was 133 seconds, including talking time both before and after rejection. For comparison, the median length of nonrejected calls was only 94 seconds. Thus, rejected calls tend to be longer calls and they are rejected fairly early in the call. Lest this finding be interpreted as implying that users are more likely to reject calls early that they anticipate will last long, analysis of later data shows that such is not the case.

The rejection rates of the delayed calls may be compared to the probability of rejecting a call on the standard telephone circuit. In the pure delay experiment it was possible to record attempted rejections of standard circuits by the subjects. These numbered 36 in 6688 calls, or less than one per cent.

---

* These four subjects rejected 1 of 6, 2 of 12, 5 of 11, and 8 of 22 calls respectively.

† These eight subjects rejected 1 of 2, 1 of 5, 2 of 2, 2 of 3, 2 of 7, 3 of 4, and 18 of 21 calls respectively.

3.5 *Discussion*

The extent to which the 2400-msec exposure, actually experienced by 16 of the 18 subjects, influenced the over-all results has not been established. It cannot be assumed that the high rejection rate at 1200 msec would ever have been reached if 2400 msec had not been introduced. The rejection rate at 600 msec has undoubtedly been influenced by the exposure to 1200 and as well as 2400 msec. The importance of such interaction effects was not appreciated until after the more extensive tests at experiments 2 and 3. Regardless of the probable influence of the 2400-msec exposure and the 600–1200 mixture on the results, it is significant that the subjects show no sign of becoming accustomed to delay and adapting their conversations to its presence.

Individuals vary widely in their reaction to delay. One of the test subjects rejected 27 of the 36 calls he made at 1200 msec. Since the remaining 17 subjects rejected 17 of their 118 calls over the 10-week period at that value of delay, it can be seen that this one subject had an important effect on the total results. Of course, such strong objectors cannot be ignored in planning communications systems.

Because of the large differences in calling rate and rejection rate among the subjects and the small number of calls which could be sampled, the absolute levels of rejection rate shown in Fig. 2 cannot be taken as representative of larger populations of calls and users. The qualitative finding of increasing rejection rate with experience with the delays used is clear, however, for several of the subjects who must be assumed representative of a significant proportion of the total population.

3.6 *Summary*

Pure delays up to 1200 msec, round-trip, added to a telephone circuit did not result in much user dissatisfaction with the circuit for users with limited experience in using such circuits.

After further experience, which included limited exposure to 2400-msec delay, considerable dissatisfaction developed for delays of 600 msec and 1200 msec when these were intermixed.

IV. EXPERIMENT 2: DEGRADATION IN TRANSMISSION QUALITY DUE TO DELAY, ECHO AND ECHO SUPPRESSOR (INTERMIXED DELAYS)

4.1 *Apparatus*

The arrangement of the SIBYL simulator for experiment 2 is shown in Fig. 3. It differed from experiment 1 in that normal 2-wire connections
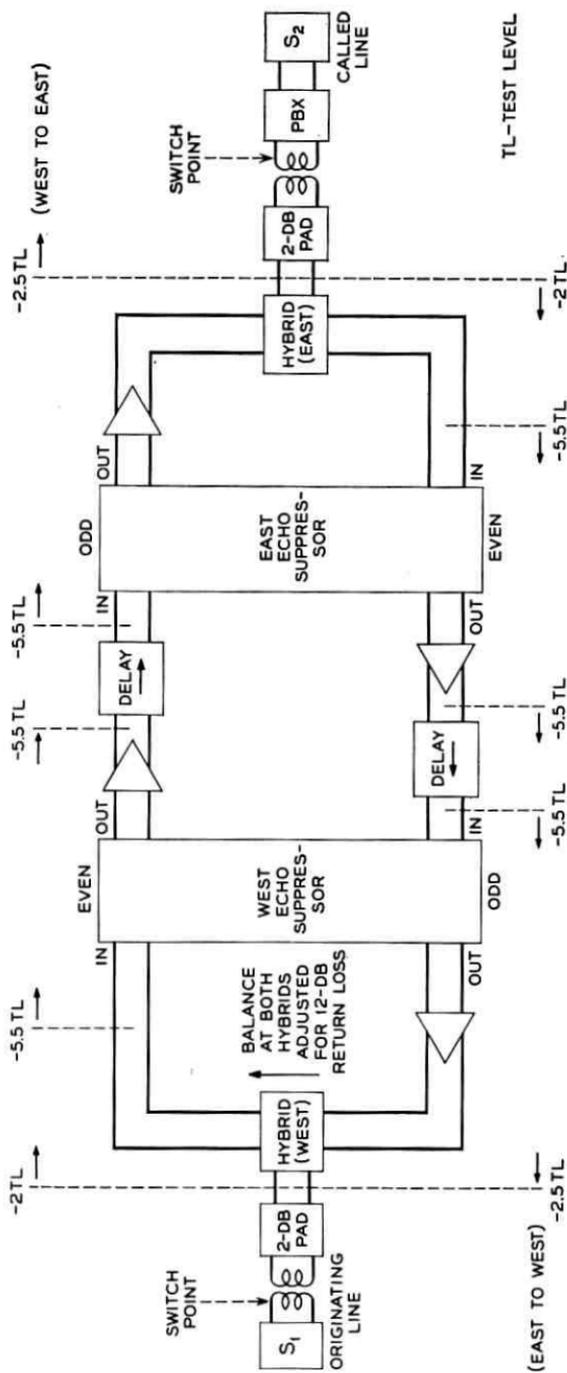
Fig. 3 — Satellite simulator circuit for experiments 2 and 3 showing method of introducing delay and echo suppressors. The level at the two-wire connection of the hybrid is arbitrarily designated as −2 test level. The levels at other points use the same reference. The overall loss between switch points is 4.5 db. For the modified 1A suppressor the adjustments were as follows. Zero test level sensitivity: even, −41 dbm; odd, −41 dbm. Actual sensitivity: even, −46.5 dbm; odd, −46.5 dbm. Hangover: even, 18 msec; odd, 50 msec.

were maintained between the telephone sets and hybrid transformers. Return loss[3] was set at 12 db; about 15 per cent of normal long-distance connections have a return loss and echo worse than this value.

The four echo suppressors employed are described in the preceding article by Brady and Helder.[1] They will be designated here, as in the previous article, by 1A, BH, GN and AM. Delay was produced by Echovox, Sr. magnetic disc units in the 4-wire portion of the circuit.

For each call through the simulator, the originating number, called number, length of call, and time of rejection (if rejected) were recorded.

4.2 *Subjects*

The subjects were 101 employees of the Murray Hill Laboratories representing a wide cross section of occupations and ages. None of the subjects were working on transmission quality or satellite projects. Because some of the subjects shared a telephone, there were only 94 telephone lines. Any call originating from these telephones could be routed through the simulator regardless of its destination. This is different from experiment 1, in which only calls between subjects could receive the experimental circuit.

4.3 *Instructions*

As in experiment 1, the subjects were not informed as to the nature of the degradation which would occur, nor were they cued as to which calls were routed through the simulator. They were instructed that some of the calls they originated would be routed over a simulated satellite circuit, and that if they found any such call unsatisfactory for normal telephoning they could restore the standard circuit merely by dialing a "3" without hanging up or breaking the connection.

4.4 *Test Conditions*

In addition to the four different echo suppressors and several delays, several experimental circuits were employed which involved adding loss or noise to an otherwise standard circuit. This was done to establish a relation between percentage of calls rejected and the more conventional types of circuit degradation.

On any one day of the experiment only one echo suppressor and four delays were employed; on other days the four loss conditions or the four noise conditions were employed. There were seven different daily conditions and four values of the appropriate parameter within each daily condition, as shown in Table I.

TABLE I — DESCRIPTION OF TEST CONDITIONS USED
IN EXPERIMENT 2

Suppressor 1A	50	200	600	1200	msec delay
Suppressor BH	50	200	600	1200	msec delay
Suppressor GN	50	200	600	1200	msec delay
Suppressor AM	50	200	600	1200	msec delay
Noise	26	32	38	44	dbrn noise*
Loss 1	6	9	12	15	db added loss
Loss 2	6	10	15	21	db added loss

* The noise was a recorded mixture of thermal noise and power hum selected to be representative of actual noise on telephone circuits. It was measured at the line terminals of the calling subject across his 500-type station set using a Western Electric 3A noisemeter with C-message weighting.

The main experiment lasted for eight weeks. On three days of each five-day week, three of the echo suppressors were used; on one of the remaining days the noise conditions were used, and on the other day the loss conditions were used. The order of suppressors, noise and loss was varied so that each of these occurred on different days each week. In addition, the sequence of values within days was varied so that each value occurred at a different time each day. The values within days were changed every twenty calls. Thus, the first 80 calls each day were assigned to the four experimental values given in Table I in some previously determined order. On most days 100 calls were made through the simulator, the final 20 calls being over a standard circuit to provide a basis of comparison. As an example of the schedule, suppressor 1A might be used Monday with the first 20 calls at 600 msec delay, with the second 20 calls at 200 msec, with the third 20 calls at 1200 msec, and with the fourth 20 calls at 50 msec; the final 20 calls might use the standard circuit. On Tuesday, the four noise conditions might be inserted.

Whether or not any particular subject received the simulator circuit when he placed a call was dependent upon whether or not the simulator was in use. It could handle only one call at a time. The other controlling factor was that no subject was given two calls on the simulator in immediate succession.

Echo suppressors BH and GN were used for the entire eight weeks of the experiment. Suppressor AM was not available until the fifth week, at which time it was substituted for suppressor 1A, which had been used along with BH and GN during the first four weeks.

Loss 1 (6–15 db range) was employed the first two weeks, during which time it became clear that the range was too small, whereupon

loss 2 (6–21 db) was substituted for the final six weeks. The noise conditions were used one day a week for all eight weeks.

Following the eight weeks of the main part of the experiment, an additional three weeks were run in which only one suppressor and one delay was used for the entire week. This was done in order to see if users would learn to adapt to the circuit changes with more practice under one condition. During the last of these additional weeks (week 11) the subjects were called back after each rejection and asked about the reason for rejection. After the eleventh week, the subjects were informed that the experiment was over.

Table II reviews the plan of the entire experiment by weeks.

TABLE II — PLAN OF EXPERIMENT 2 BY WEEKS

Week of Experiment										
1	2	3	4	5	6	7	8	9	10	11
1A suppressor			AM suppressor					BH at 600 msec only all week	GN at 600 msec only all week	BH at 600 msec only all week
BH suppressor										
GN suppressor										
Loss 1		Loss 2								
Noise										

(For variation within days during weeks 1–8, see Table I. Each horizontal row of Table II represents one weekday.)

4.5 Results

The percentage of calls rejected at each delay is shown in Table III for each suppressor separately. There is little difference among the four suppressors. Suppressor AM shows a slightly higher rejection rate than the other suppressors, but it was tested only during weeks 5–8 and the difference may well be due to an increased sensitivity after training on the delay circuits (suppressor BH showed exactly the same number of rejects as suppressor AM during weeks 5–8). A three-way analysis of variance confirms that there was no statistically significant difference in rejection rate among the four suppressors.* Therefore, the data from all suppressors are pooled in Fig. 4, which shows the rejection rate as the first eight weeks of the experiment.

---

* Differences between suppressors can be demonstrated by special techniques other than natural conversation, as has been shown by M. B. Gardner and J. R. Nelson.[4] Their acceptability in natural conversations would seem to be the ultimate criterion, however.

TABLE III — REJECTION RATE FOR EACH SUPPRESSOR
AND DELAY

	Round-Trip Delay in msec				
Suppressor	50	200	600	1200	All Delays
1A*	5	15	25	39	21
BH	7	11	24	41	21
GN	4	17	22	38	20
AM†	9	20	36	40	26
All suppressors	6	15	26	39	

Combined data for first eight weeks of experiment 2. Each table entry is based
on 80 calls for suppressors 1A and AM and 160 calls for suppressors BH and GN.
Table entries are percentage of calls rejected.
  * Tested during weeks 1–4 only
  † Tested during weeks 5–8 only

The standard condition is shown on Fig. 4 for comparison. Of 766
calls over the standard circuit, 10 (1.3 per cent) were rejected.

Fig. 5 shows the rejection rate for each value of added loss and the
standard circuit with no added loss. All subjects are combined over the
eight weeks during which the loss conditions were run.

Fig. 6 shows the rejection rate for each value of noise for all subjects
and weeks combined. The standard circuit for internal calls had a 6-dbrn
noise level and is plotted at that point.
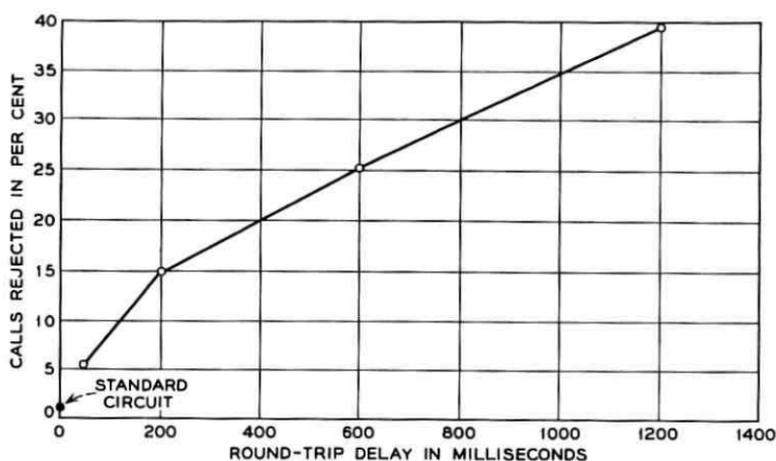
Fig. 4 — Rejection rate as a function of delay: combined data from 101 sub-
jects using four echo suppressors for a total of eight weeks. All subjects were
exposed to all delays and suppressors; each point represents 480 calls, except for
the standard circuit point, which represents 766 calls.
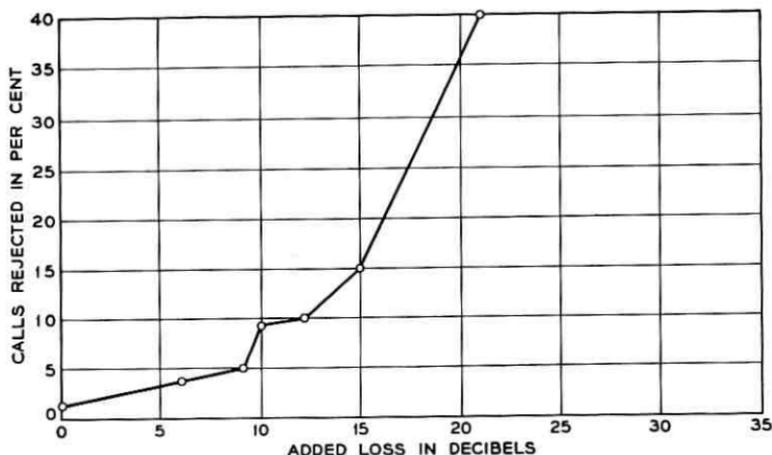
Fig. 5 — Rejection rate as a function of added loss: combined data from 101 subjects. The 9- and 12-db added-loss conditions were run during the first two weeks only, and represent 40 calls each. The other points represent 120 or 160 calls each, except for the zero added-loss point (standard circuit), which represents 766 calls.

4.6 *Effect of Experience*

The above figures summarize the data over the first eight weeks of the experiment and thus leave unanswered the question about change in rejection rate with experience. Do the circuits with delays, echoes and suppressors show an increased rejection rate with exposure, as shown in Fig. 2 of the pure delay tests? Evidence on this question is presented in Fig. 7, which shows rejection rate as a function of weeks of the experi-
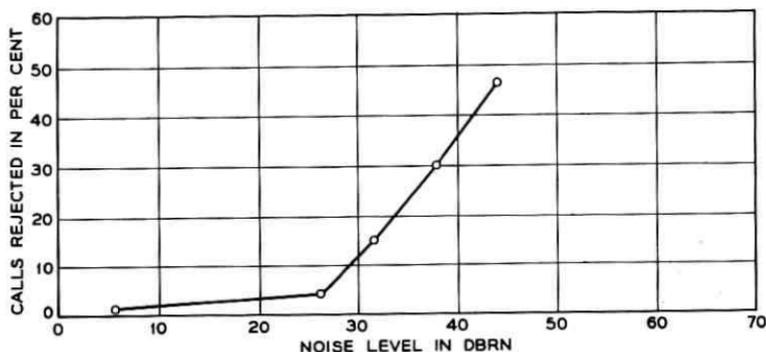
Fig. 6 — Rejection rate as a function of noise level: combined data from 101 subjects. Noise was a mixture of thermal noise and power line hum measured across the terminal set by a Western Electric 3A noise meter with C-message weighting. Each point represents 160 calls, except the 6 dbrn point, which represents 766 calls; the 6 dbrn point is the standard circuit condition.
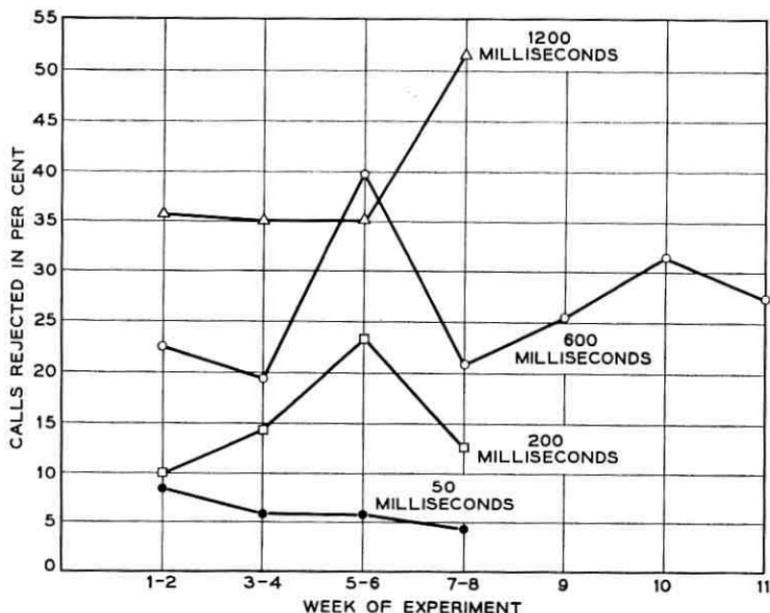
Fig. 7 — Rejection rate as a function of weeks of experiment: data from 101 subjects using all echo suppressors in use each week. Each point represents 120 calls, except points for weeks 9–11, which represent 268 or 480 calls each. Only the 600-msec delay condition was used on weeks 9, 10 and 11.

ment for each delay condition separately. The data from all suppressors are combined since there was no meaningful difference among suppressors. The changes in rejection rate over weeks shown in Fig. 7 are quite erratic, but for each delay above 50 msec the rejection rate was higher in the second half of the experiment than in the first half. The rejection rate at 50 msec actually decreased with weeks, but it should be pointed out that none of the changes with experience are statistically significant in this experiment.

4.7 *Interview Results*

The results of calling the subjects back after they had rejected a circuit is shown in Table IV. This procedure was followed only in week 11, and included 70 calls. Chopping and echo rank first among types of annoyance, which is to be expected from the action of the suppressor. Noise and loss comments do not refer to experimentally added noise and loss, since these conditions were not employed in the eleventh week. The echo suppressor does put in loss during double talking and its action does interrupt line noise originating at the other end of the circuit,

TABLE IV — RESPONSES OF SUBJECTS TO INTERROGATION CALL
IMMEDIATELY FOLLOWING THEIR REJECTION OF THE
CIRCUIT DURING WEEK 11 (BH SUPPRESSOR,
600 MSEC DELAY)

Why did you reject the circuit?	
Echo	47%
Chopping	36%
Noise	21%
Low volume	11%
Delay	13%
Would the circuit have been acceptable for a transatlantic call?	
No	51%
Marginal	28%
Yes	21%

A total of 70 interviews was made.

facts which may account for at least some of the loss and noise comments.

Table IV also shows the percentage of responses of several types to the question as to whether the circuit they had rejected would be acceptable for transatlantic calls. (They had been instructed to reject circuits unacceptable for normal telephoning.) Almost 80 per cent of the respondents said the quality would be unacceptable or marginal for transatlantic service.

4.8 *Length of Calls*

The relation between length of call and rejection rate was investigated by combining the data from all suppressors and delays of 200 msec or more. For 532 calls lasting 30 seconds or less the over-all rejection rate was 8 per cent. For 391 calls lasting 2.5 minutes or more the over-all rejection rate was 51 per cent. Another way of looking at the relation between length of call and rejection is to consider the length of rejected and nonrejected calls. Nonrejected calls had a median length of about one minute. Rejected calls have a median length of about 2.5 minutes and have a median time of rejection of about half a minute for the delay conditions. The probability of rejecting a call at any point was looked at as a function of the total length of the call. There was no relationship; that is, subjects do not show a tendency to quickly dial out calls which they know will last long.

Calls going outside the Murray Hill Laboratory* are longer on the

* The breakdown of calls by destination was: Murray Hill extension, 62 per cent; tie lines, 23 per cent; outside local, 14 per cent; DDD, 1 per cent.

average than calls to inside extensions and show a higher rate of rejection. There is no meaningful indication that the change in rejection rate for outside calls is either more or less than would be expected from their longer duration.

4.9 *Discussion*

In interpreting the results of experiment 2 it must be kept in mind that all subjects were potentially exposed to delays up to 1200 msec each week.* As in experiment 1, we must suspect that this exposure influenced the rejection rate at the lesser delays; experiment 3 provides data on this point. However, exposure of telephone customers to delays up to 1200 msec is not inconceivable in future satellite systems, and the present experiment — which gives the average user one such call every two weeks — is perhaps a reasonable approximation to one such system.

Two facts should be borne in mind in extrapolating from the laboratory experiments to field usage. One is that the return loss in the laboratory tests was set at a level somewhat worse than the average field expectation. Return losses in operating systems would be expected to be equal to or worse than the laboratory value in only about 15 per cent of all calls. The other factor is that the average laboratory call is much shorter than the average long-distance call; our results show that longer calls are rejected at a much higher rate. Thus, the laboratory rejection rates cannot be applied directly to all field calls, but do indicate potential difficulty in a significant proportion of long distance calls with intermixed delays such as those in experiment 2.

4.10 *Conclusions*

(*i*) Intermixed delays from 200 to 1200 msec produce substantial rejection rates under the conditions of the present experiment.

(*ii*) There was no meaningful difference among the four echo suppressors tested.

(*iii*) There is no evidence of increased tolerance of delay with experience nor on long distance calls.

(*iv*) Rejection rate increases greatly with length of call.

---

* Those people who made many calls did talk over the maximum delay each week. Many subjects who used their phone little did not talk over the maximum delay each week, but they do not influence the rejection rate as much, either.

V. EXPERIMENT 3: DEGRADATION IN TRANSMISSION QUALITY DUE TO
  DELAY ECHO AND ECHO SUPPRESSOR (FIXED DELAYS)

In the previous experiments, delays up to 1200 msec were intermixed each week so that all subjects were exposed to the entire range of delays. Experiment 3 modified the procedure of experiment 2 in that no subject ever talked over a longer delay than that assigned to his group, and separate groups of subjects were exposed to maximum delays of 50, 200, 400 and 600 msec.

5.1 *Apparatus*

The SIBYL simulator arrangement was the same as that used in experiment 2, except that only one echo suppressor was used (suppressor BH). Fig. 3 of experiment 2 shows a block diagram of the simulator which applies equally well to experiment 3.

5.2 *Subjects*

Eighty employees of the Murray Hill Laboratory served as subjects. None had served in any previous simulator experiment. During the course of the experiment 24 subjects changed rooms, left or otherwise became unavailable. Their data was removed from the analysis.

5.3 *Instructions*

The instructions to the subjects were the same as in experiment 2: that is, they were told to dial out any circuit which they found unsatisfactory, but they were not told the nature of the degradation and were not cued as to which calls were routed through the simulator.

5.4 *Test Conditions*

During the first week of the experiment all subjects were exposed to the 50-msec delay condition only. That is, on any call they initiated they would receive the simulator with a 50-msec delay unless another subject was already using the simulator. After the first week the 80 subjects were divided into four groups of 20 each, matched approximately on calling rate as determined from the first week. The sequence of delays to which each group was exposed is shown in Table V. Note that subjects in each group were never exposed to delays longer than that assigned from the fifth through fourteenth weeks.

Because the delay could not be changed instantaneously when any

TABLE V — SEQUENCE OF DELAYS FOR EACH GROUP OF
EXPERIMENT 3; SUPPRESSOR BH WAS USED THROUGHOUT

Group	Week of Experiment															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	50				→600											→
2	50	200												→		
3	50	200	400											→		
4	50	200	600											→		

Table entries are round-trip delay in msec.

particular subject picked up his telephone, it was necessary to divide
the 8-hour day into four 2-hour periods and make the simulator, with
appropriate delay, available to only those subjects who were assigned
the same delay. The sequence of conditions within each day was rotated
in such fashion that each group of users had access to the experimental
circuit at a different time each day. Of course, they were not informed
of this scheduling.

Group 1, which had the 50-msec delay for the first four weeks, was
changed to 600 msec on week five. Since so few calls at 50 msec were
rejected, it was not considered necessary to continue the 50-msec
condition beyond week four. Group 4 was changed to 600 msec in order
to increase the number of subjects at this delay, which is perhaps the
most important one for questions of satellite communications. In order
to combine groups 1 and 4 for an equivalent number of weeks at the
600-msec delay, data from two additional weeks (weeks 15 and 16)
were taken from group 1.

The noise and loss conditions of experiment 2 were not used in experi-
ment 3. Another difference from experiment 2 was that no control was
exercised over the number of calls made each day under each of the
experimental conditions. Rather, a time period was set aside each day
for each condition, as previously described.

As in the previous experiments, the data recorded consisted of calling
number, called number, length of call, and time of rejection if any.

During the eighth and fourteenth weeks the subjects were called back
immediately after rejecting a circuit and also were called on a sample of
simulator calls which were not rejected. They were asked what, if any-
thing, they had noticed about the circuit on the previous call.

Finally, after the above schedule had been completed, the subjects of
groups 1 and 4 were given one week of mixed delays of 200, 600 and 1200
msec in a fashion similar to experiment 2. This was done to check the

possibility that differences in results between experiments 2 and 3 might have been due to a chance difference in subjects.

5.5 *Results*

During the first week of experiment 3, when all subjects were on the 50-msec delay, 73 of them made at least one call over the simulator. A total of 475 calls were made, of which eight, or less than 2 per cent, were rejected. Group 1 continued on 50 msec through the fourth week. During those four weeks they made a total of 376 calls, of which two, or less than 1 per cent, were rejected. In view of the low rejection rate and the fact that no calls at all were rejected during the third and fourth weeks, group 1 was changed to 600-msec delay on the fifth week.

The 200-msec delay condition was given to all subjects except group 1 during the second week and to group 2 only from the third through fourteenth weeks. During the second week 13 of 265 calls, or less than 5 per cent, were rejected.

The main body of results from experiment 3 is shown in Fig. 8, which plots rejection rate as a function of delay. In the solid curve, different subjects are represented at each delay; that is, each subject appears only at the delay value which was the maximum to which his group was exposed, except for group 1, which appears in both the 50-msec and 600-msec conditions for the main part of experiment 3. This combined group was also given the added week of mixed delays, shown by the dashed line. Fig. 8 shows a low rejection rate at 200-msec delay (2 per cent), less than 8 per cent rejection at 400-msec delay and 14 per cent rejection at 600-msec delay. These figures may be contrasted with the rates from experiment 2 shown in Fig. 4 — which are roughly double these values — and the rejection rates for mixed delays during the one-week extension of experiment 3, which are only slightly below those of Fig. 4.

5.6 *Effect of Experience*

Experiment 1, with pure delays presented in an intermixed fashion, showed a large increase in rejection rate with continued experience. Experiment 2, with echoes and suppressors added, showed little increase. The results of experiment 3 are shown in Fig. 9, which plots average rejection rate as a function of weeks of experience at maximum delay. The data from the 200-, 400- and 600-msec delay groups are plotted in this figure. Data from all groups are combined in such fashion that the first point plotted represents the first two weeks each group was at its maximum delay (not including the added week of mixed delays to 1200 msec).
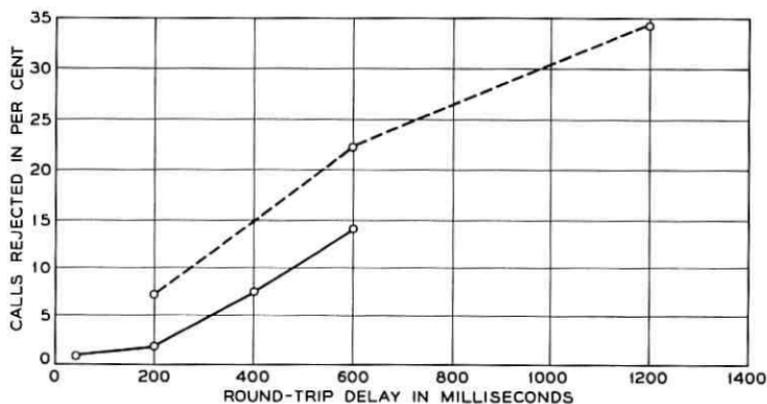
Fig. 8 — Rejection rate as a function of delay. Each point represents separate groups of 12 to 25 subjects, except that the 50-msec group (group 1) was later given 600-msec delay and is included with the subjects initially assigned to 600 msec (group 4). The light dashed line shows the result of exposing this combined group (1 and 4) to the mixed delays of experiment 2 for one week. Each point of the solid curve represents 376–774 calls. Each point of the dashed line represents 53–68 calls.

Fig. 9 shows a trend toward increasing rejection rate with experience on fixed delays, but as in experiment 2 the increase is small compared to the pure delay condition in experiment 1.

5.7 *Interview Results*

The results of calling the subjects back after they had made a call over the simulator are shown in Table VI. This table is in two parts: one for the case in which the simulator call had not been rejected and one for rejected calls. If the call had not been rejected, more than half the subjects noticed nothing different about the circuit; other comments were spread over many categories, echo being the largest. If the call had been rejected, half the subjects reported objectionable echo and all other categories of comments were higher than for the nonrejected calls. As in experiment 2, the suppressor will introduce loss during double talking which will subjectively make the circuit sound low in volume, and it will sometimes chop and mutilate echo, room noise and even speech in such a way as to sound like line noise. It is very difficult for the user to identify the exact nature of difficulties with those circuits.

5.8 *Destination and Length of Call*

The findings of experiment 2 on destination and length of call were substantiated in experiment 3. Long calls are rejected at a higher rate. Outside calls are longer and are rejected more than inside calls.
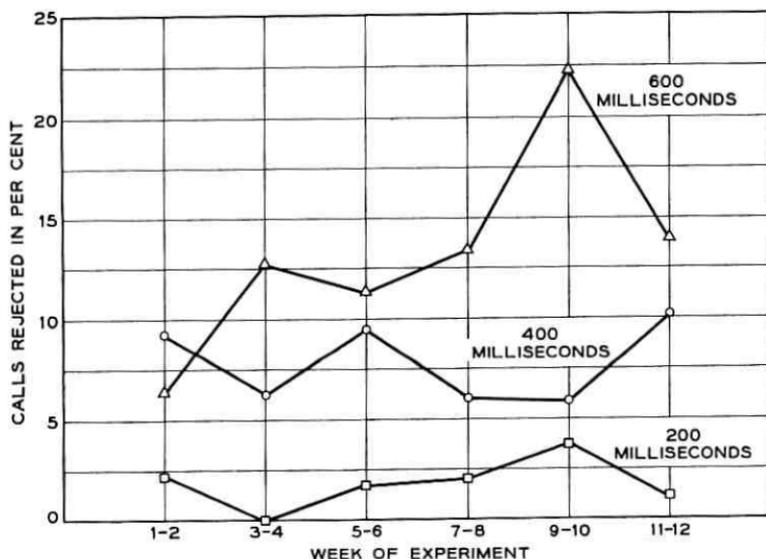
Fig. 9 — Rejection rate as a function of weeks at maximum delay. Data from all groups of experiment 3 combined as of the start of their respective weeks at the maximum delay to which each was exposed (200, 400, 600 msec). The added week of mixed delays to 1200 msec is not included. Each point represents 47 to 163 calls.

VI. GENERAL DISCUSSION AND SUMMARY

The effects of transmission delay upon the quality of a telephone circuit are not as obvious to the user as the more conventional degradations such as noise, loss and distortion. Indeed during most of a conversation over a delayed circuit there is no degradation. When degradation does occur it can often be misinterpreted by the user as being due to the other speaker. Slow responses, excessive interruptions and complete failures to respond (because the question was lost in the circuit) are examples of such difficulties.

It is not surprising, therefore, that in more than half the conversations involving pure delays of 1200 msec, or 600 msec plus echo and suppressors, the users are not aware that there is anything different about the circuit.*

In the pure delay condition the effect is particularly subtle since echo and speech mutilation are absent. Experiment 1 shows that it takes

---

* That there is an effect, however, is indicated by the fact that the users tend to talk for shorter times over these circuits. The median length of call over the 600-msec circuit with echo and suppressors is 24 per cent shorter than call length on the standard circuit.

TABLE VI — RESPONSES OF SUBJECTS TO INTERROGATION
CALL IMMEDIATELY FOLLOWING A CALL OVER
THE SIMULATOR

	Rejected Calls
Why did you reject the circuit?	
Echo	58%
Chopping	10%
Noise	22%
Low volume	18%
Delay	10%
	Nonrejected Calls
Did you notice anything different on your last call?	
Noticed nothing	52%
Echo	23%
Chopping	7%
Noise	0
Low volume	13%
Delay	11%

Only data from 600-msec delay (suppressor BH) are included, since rejection rates and therefore sample size at lesser delays were too small to be meaningful. Percentages are based on 40 rejected calls and 56 nonrejected calls.

considerable experience, and experience with long delays, before the effects of pure delay are recognized and rejected. Because of the limited scope of experiment 1 it is not possible to say exactly how much pure delay can be tolerated in telephone transmission, but it is clear that 2400 msec is completely unacceptable even on the first few calls and that intermixed 600 msec and 1200 msec are both rejected at fairly high rates by users who have been sensitized to delay circuits.

When echoes and echo suppressors are added to the circuit and delays up to 1200 msec are intermixed, even the 200-msec delay shows up with a significant rejection rate (15 per cent).

It might be thought that users would be more tolerant of long distance calls than internal calls and therefore that the rejection rates shown in all three of the present experiments are pessimistic. Analysis of the long distance calls made in experiments 2 and 3 does not support this view. In both cases the long distance calls were rejected at a significantly higher rate than internal calls. Although the difference in rejection rate can probably be attributed to length of call, long distance calls in operating systems average longer than those in the present study, and overseas calls are considerably longer on the average.

Experiment 3 showed that the use of fixed maximum delay conditions for each subject decreases the user dissatisfaction with circuits of shorter delays compared to the intermixed situation. When separate groups of

subjects were exposed to different values of maximum delay, the group exposed to 200 msec rejected only 2 per cent of the calls, and even at 600 msec the rejection rate was only 14 per cent, compared to 26 per cent in experiment 2, in which delays up to 1200 msec were intermixed.

Individual users' reactions to delay, echo and suppressor action differed widely. Among each group of subjects there was at least one who never rejected any of the simulator circuits. At the other extreme, the most critical subject had a rejection rate three to five times the group average. There was a slight tendency for the heavy users to be more critical, but this is likely due entirely to the increase in rejection rate with experience.

There was no meaningful difference in rejection rate among the four echo suppressors tested. None offered a measurable advantage over a modified version of the suppressor most commonly used today.

Finally, what do the results of these experiments mean for international communications? From the present experiments it would seem safe to conclude that noticeable degradation in transmission quality may occur under some circumstances with delays as low as 200 msec and with any of the currently available echo suppressors. The circumstances include low return loss and users who have been sensitized to the problems created by transmission delay.

Degradation increases with delay until at 1200-msec round-trip delay and any present echo suppressor, more than one-third of the calls were rejected as being unsatisfactory under the conditions of these experiments. An increase in circuit loss of approximately 20 db was required to produce the same rejection rate.

The degradation due to transmission delay in any telephone transmission system must be weighed against the cost and other degradations inherent in alternative transmission systems of less delay. It is clear, however, that the influence of delays of 200 msec or more should be considered in designing international voice communications systems to carry natural conversations.

nance of the echo suppressors. Mrs. N. W. Shrimpton wrote several computer programs for analysis of some of the data. Miss M. J. Billington interviewed the subjects for experiments 1 and 2. Mrs. V. A. Bull carried out call-back of the subjects and performed much of the data analysis for this article. J. E. Karlin consulted in all phases of the experiments.

REFERENCES

1. Brady, P. T., and Helder, G. K., Echo Suppressor Design in Telephone Communications, B.S.T.J., this issue, p. 2893.
2. Irvin, H. D., Studying Tomorrow's Communications . . . Today, Bell Laboratories Record, **36**, November, 1958, p. 399.
3. Emling, J. W., and Mitchell, D., The Effects of Time Delay and Echoes on Telephone Conversations, B.S.T.J., this issue, p. 2869.
4. Gardner, M. B., and Nelson, J. R., Combating Echo in Speech Circuits with Long Delay, J. Acoust. Soc. Am., **35**, Nov., 1963, p. 1762.

Calculation of the Spin-Axis Orientation of the *Telstar* Satellites from Optical Data

By D. W. HILL

The orientation of a satellite can be calculated if the aspect of two or more known reference vectors can be measured with respect to a body-fixed coordinate frame. On the Telstar satellites one such line may be determined by solar aspect cells, but the other is determined by ground observation of flashes of sunlight reflected from three mirrors mounted on the spinning satellite. The simple geometric considerations are discussed, as are the optimum placement of the mirrors, the amount of data to be used, and an error analysis. Spin-axis orientation data are used as the basis for steering the spin axis by ground command of the current in a coil of wire around the equator of the satellite.

I. INTRODUCTION

A novel method was devised to determine the orientation in space of the Telstar satellites. The method used grew out of limitations on weight, power, and telemetry facilities, and consists of three highly reflective mirrors and six solar aspect cells attached to the outside of the satellites. The mirrors reflect sunlight to a ground observation station,* and the solar aspect cells measure the direction of the sun with respect to reference coordinates in the satellite. The method has proved successful in determining the orientation of the Telstar satellites with sufficient accuracy both to deduce the residual magnetic moments of the satellites and to correct their precession by means of an equatorial torque coil. The remainder of the discussion concerns the details of the attitude determination scheme. See Ref. 1 for a description of the photoelectric equipment used. We will restrict the discussion to a spin-stabilized

---

* The use of mirrors for this purpose was first suggested by D. Gibble of Bell Laboratories.

satellite and further assume that the precessional motion occurs at a slow rate.

The orientation of a satellite can be calculated if the directions of two known lines in inertial space can be measured with respect to a body-fixed reference frame. One such reference line can be measured by solar aspect cells, and, in the case of a spin-stabilized satellite, another can be determined by observing flashes of sunlight reflected from a mirror attached to the satellite. Because of the indeterminancy of the angular displacement about each reference being observed from the satellite, we know only that any given body axis must lie on a cone about the reference line. The semivertex angle of the cone is the angle between the body axis and the reference line, and is the quantity determined by the attitude sensing scheme. Consequently, if two reference lines are known with respect to the satellite, the orientation of any desired body axis must lie along one of the two lines of intersection of the two cones which have been determined. Since the inertial coordinates of the reference lines are known, the orientation of the body axis can be calculated. Any ambiguity between the two possible orientations can usually be resolved by the general knowledge of the observations, and in fact, no difficulty was experienced with the Telstar satellites in resolving the ambiguity.

1.1 *Mirrors Attached to Satellite*

There are three mirrors attached to the satellite, one being a plane mirror approximately 4×6 inches in size mounted so the normal to the mirror makes an angle of 68° with the symmetry or spin axis of the satellite; see Fig. 1. The other two mirrors are each approximately half the size of the first and both are mounted so the normal is inclined 95° to the spin axis. The two 95° mirrors are 120° apart in aximuth around the satellite and provide a coding of the flash train to enable the ground station to determine which mirror is flashing.

If a train of flashes of reflected sunlight is observed at the earth and the mirror which is flashing can be identified, we know that the spin axis must lie on a cone of either 68° or 95° semiangle about the line normal to the mirror at the instant of a flash. This line can be calculated, since it lies in the plane of the sun, satellite, and ground station and bisects the angle between the incident and reflected rays.

The angles 68° and 95° were chosen because mirrors placed at those angles would afford nearly the maximum number of flash observations for the nominal orbit and attitude. Calculations showed that these
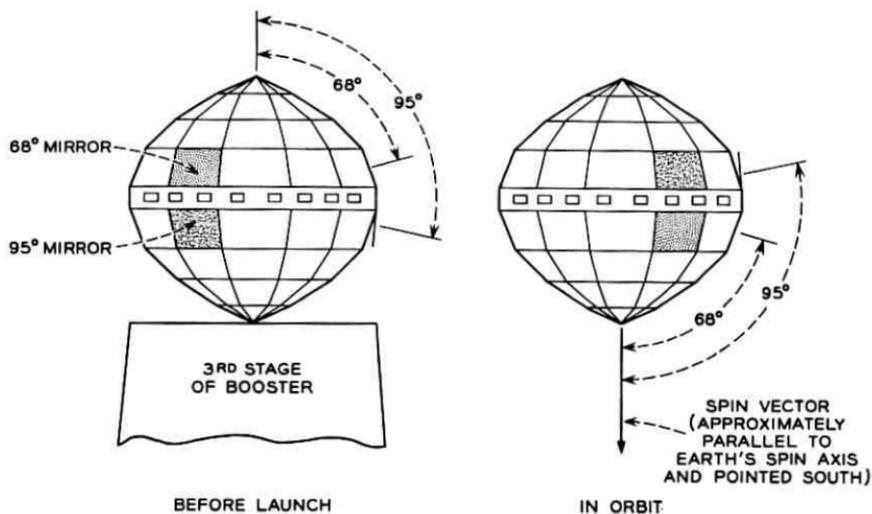
Fig. 1 — Placement of mirrors.

two mirrors would provide an average of more than one flash observation per night, weather permitting, for the first 55 nights, after which there would be no night passes for about one month due to precession of the satellite orbit and movement of the sun in the geocentric frame of reference.

1.2 *Solar Aspect Determination*

The determination of the sun's aspect is made on the satellite by six solar cells placed on the nearly spherical surface of the satellite. These solar cells are placed symmetrically and oppositely pointed on each of three orthogonal axes, Fig. 2. Each cell responds to sunlight over one hemisphere — i.e., 2π steradians — giving a current response approximately proportional to the cosine between the direction of the sun and the axis of the solar cell. Thus no more than three cells are illuminated at any one time, and the direction of the sun with respect to the orthogonal triad through the solar cells is found by relatively simple calculations using the currents generated by the three illuminated cells.

The orthogonal triad through the solar cells is located symmetrically with respect to the spin axis of the satellite, three axes piercing the upper hemisphere of the satellite and their opposite extensions piercing the lower hemisphere, the solar cells being attached at the intersections of the axes and the shell of the satellite. Thus the three axes in either
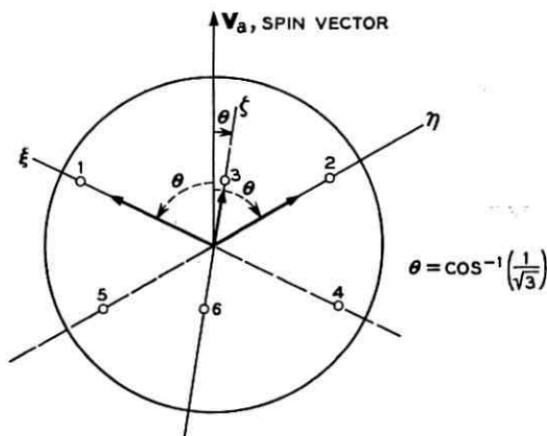
Fig. 2 — Solar aspect cell geometry.

hemisphere each have a direction cosine of $\sqrt{1/3}$ with the extension of the spin axis in that hemisphere.

Once the aspect of the sun is determined with respect to the orthogonal triad through the solar aspect cells, the angle between the satellite-sun line and the spin axis is thus easily computed.

In principle this solar aspect scheme is satisfactory, but it does suffer from an inherent defect. Fundamentally, the method is an analog method depending on absolute values of solar cell current for solar aspect determination. Thus any drift, either in the properties of the cells or in the data amplifying and transmitting system, will cause inaccuracies in the aspect determination. In order to minimize these effects, the cells have been preirradiated to lessen the effect of radiation in the satellite environment. Further, thermistors are used to measure the temperature of the cells and a temperature correction is applied to the solar cell current readings. Finally, two methods of computing the solar aspect are used to try to isolate any discrepancies that arise. However, there is no way to arrive at a correct aspect determination if the solar cells give erroneous readings. The best that can be achieved is to obtain some average result by distributing errors in some rational manner, such as distributing them in proportion to the currents measured in the three illuminated cells.

Consideration was given to correcting for angular rotation of the satellite during the time required to read the solar aspect cell currents. The six cell currents are measured a nominal 100 microseconds apart, so at 3 revolutions/second the satellite rotates 0.54° from first to last reading. However, calculations showed that the solar aspect cell data

would need an accuracy in excess of the capability of the solar aspect cells and data encoder to show these effects. Therefore, no spin correction was included in the analysis.

II. DETERMINATION OF SPIN-AXIS ORIENTATION

In this section we want to consider the problem of determining spin-axis orientation from observational data which give the directions with respect to the spin axis of two or more known reference lines. One of these reference lines will be the line normal to a mirror attached to the satellite at the instant a flash of reflected sunlight is observed at the ground station, and the other will be either a second such reference line or the line from the satellite to the sun. Since the change in the satellite-sun line is very slow, two solar aspect observations are not very useful for determining spin-axis orientation, as the two cones would nearly coincide and hence yield results of questionable accuracy. Clearly, any pair of observations could be used for spin-axis determination, but for useful information on the spin-axis precession the observations should be made as close together in time as possible. There would be no advantage in studying the use of three or more observations simultaneously, as the three or more cones would in general not have a common intersection because of observational errors. Thus the observed data should be used in pairs for determination of spin-axis orientation and any averaging should be done by operating on the resulting direction vectors.

2.1 Reference Line Obtained from Flash Observations

Let the inertial reference frame used be the x, y, z frame, and let the associated unit vectors be $\hat{\imath}$, $\hat{\jmath}$, and $\hat{k}$, respectively. At the instant a flash of sunlight reflected from the mirror attached to the satellite is seen at the ground station, the sun, satellite, and ground station coordinates in the x, y, z frame can be obtained by well-known methods.

Let the vector $\bar{R}_s$ from the earth mass center to the sun be (see Fig. 3)

$$\bar{R}_s = x_s \hat{\imath} + y_s \hat{\jmath} + z_s \hat{k}.$$

The vector $\bar{R}_g$ from the earth mass center to the ground station is

$$\bar{R}_g = x_g \hat{\imath} + y_g \hat{\imath} + z_g \hat{k}$$

and the vector $\bar{R}_b$ from the earth mass center to the satellite is

$$\bar{R}_b = x_b \hat{\imath} + y_b \hat{\jmath} + z_b \hat{k}.$$
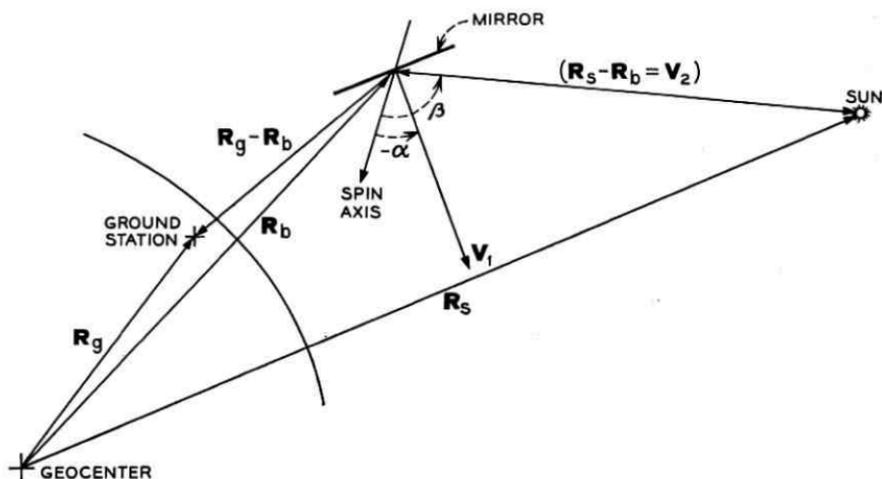
Fig. 3 — Reference frame for calculations from mirror flashes.

The vector $\bar{R}_i$ in the direction of the light ray incident to the mirror is

$$\bar{R}_i = \bar{R}_b - \bar{R}_s .$$

Since the solar parallax is only a few seconds of arc even from the satellite, we could approximate $\bar{R}_i$ by $-\bar{R}_s$, but this approximation will not be made, as very little simplification results.

The vector $\bar{R}_r$ in the direction of the reflected ray is

$$\bar{R}_r = \bar{R}_g - \bar{R}_b .$$

Let us adopt the symbol $\hat{R}$ to denote a unit vector, and the symbol R to denote the absolute value of a vector. The incident and reflected unit vectors are

$$\hat{R}_i = x_i \hat{\imath} + y_i \hat{\jmath} + z_i \hat{k}$$
$$\hat{R}_r = x_r \hat{\imath} + y_r \hat{\jmath} + z_r \hat{k}$$

where

$$x_i = \frac{x_b - x_s}{R_i}, \qquad y_i = \frac{y_b - y_s}{R_i}, \qquad z_i = \frac{z_b - z_s}{R_i}$$

$$x_r = \frac{x_g - x_b}{R_r}, \qquad y_r = \frac{y_g - y_b}{R_r}, \qquad z_r = \frac{z_g - z_b}{R_r}$$

$$\frac{1}{R_i} = [(x_b - x_s)^2 + (y_b - y_s)^2 + (z_b - z_s)^2]^{-\frac{1}{2}}$$

$$\approx \frac{1}{R_s} \left[1 + \frac{1}{R_s^2} (x_s x_b + y_s y_b + z_s z_b) \right]$$

$$R_r = [(x_g - x_b)^2 + (y_g - y_b)^2 + (z_g - z_b)^2]^{\frac{1}{2}}$$

$$= [R_g^2 + R_b^2 - 2(x_g x_b + y_g y_b + z_g z_b)]^{\frac{1}{2}}$$

$R_s = [x_s^2 + y_s^2 + z_s^2]^{\frac{1}{2}}$, distance from geocenter to sun

R_g = earth radius at ground station

$R_b = [x_b^2 + y_b^2 + z_b^2]^{\frac{1}{2}}$, distance from geocenter to satellite.

The vector $\bar{V}_1$ normal to the mirror is thus given by the expression

$$\bar{V}_1 = \hat{R}_r - \hat{R}_i$$

since $\hat{V}_1$ must bisect the angle between the incident and reflected rays and be normal to the mirror.

If we let x_1, y_1, z_1 be the direction cosines of the unit vector $\hat{V}_1$, we can write

$$\hat{V}_1 = x_1 \hat{i} + y_1 \hat{j} + z_1 \hat{k}$$

where

$$x_1 = \frac{x_r - x_i}{V_1}$$

$$y_1 = \frac{y_r - y_i}{V_1}$$

$$z_1 = \frac{z_r - z_i}{V_1}$$

$$V_1 = [(x_r - x_i)^2 + (y_r - y_i)^2 + (z_r - z_i)^2]^{\frac{1}{2}}$$

$$= [2 - 2(x_r x_i + y_r y_i + z_r z_i)]^{\frac{1}{2}}.$$

The vector $\hat{V}_1$ is our desired reference vector and can be computed explicitly by the above formulas for any time at which a flash is observed.

2.2 Second Reference Line

We will let our second reference line be represented by the unit vector $\hat{V}_2$

$$\hat{V}_2 = x_2 \hat{i} + y_2 \hat{j} + z_2 \hat{k}.$$

The vector $\hat{V}_2$ could represent the satellite-sun line in the case of a solar aspect determination, it could represent a second reference line determined from a flash observation, or it could be any other reference

line whose orientation is measured with respect to the satellite body axes as well as in the x, y, z system.

2.3 Spin-Axis Orientation

First we let $\hat{V}_3$ be the unit vector in the direction of the spin axis. If we observe a flash from the satellite and can identify the mirror which reflected the flash, then we know the angle between the normal to the mirror and the spin axis, say α. Similarly, if we determine the sun's aspect with respect to the spin axis or if we observe a flash from a second mirror, we know that the spin axis makes an angle, say β, with our second reference line. These two conditions give us sufficient information to determine the direction of the spin axis. If we know the spin axis makes an angle α with $\hat{V}_1$ and an angle β with $\hat{V}_2$, the spin axis must therefore lie along one of the two intersections of a cone of semiangle α about $\hat{V}_1$ and a cone of semi-angle β about $\hat{V}_2$. In general we do not know which intersection is the correct location of the spin axis, but we should be able to resolve the ambiguity by the general knowledge of the situation.

We can express the above statements analytically by the following set of equations

$$\hat{V}_1 \cdot \hat{V}_3 = \cos \alpha = x_1 x_3 + y_1 y_3 + z_1 z_3 \tag{1}$$

$$\hat{V}_2 \cdot \hat{V}_3 = \cos \beta = x_2 x_3 + y_2 y_3 + z_2 z_3 \tag{2}$$

$$\hat{V}_3 \cdot \hat{V}_3 = 1 = x_3^2 + y_3^2 + z_3^2. \tag{3}$$

The solution of (1)–(3) is found to be

$$x_3 = \frac{A + B y_3}{C} \tag{4}$$

$$y_3 = \frac{-(AB + A'B') \pm [(AB + A'B')^2 - (A^2 - C^2 + A'^2)(B^2 + C^2 + B'^2)]^{\frac{1}{2}}}{B^2 + C^2 + B'^2} \tag{5}$$

$$z_3 = \frac{-(A' + B' y_3)}{C} \tag{6}$$

where

$$A = z_2 \cos \alpha - z_1 \cos \beta$$

$$B = z_1 y_2 - z_2 y_1$$

$$A' = x_2 \cos \alpha - x_1 \cos \beta$$

$$B' = x_1 y_2 - x_2 y_1$$
$$C = z_2 x_1 - z_1 x_2 .$$

Given the direction cosines of the two reference vectors $\hat{V}_1$ and $\hat{V}_2$ plus α and β, we can use (4)–(6) to calculate the direction cosines of the spin axis. The plus and minus square root term in (5) is due to the two possible locations of the spin axis.

III. CONSIDERATION OF ERRORS

It is of interest to know the effect of observational errors on the calculated orientation of the spin axis. In the following the significant sources of error are discussed and the effects of these errors on the orientation of the spin axis are calculated.

Since we know the positions of the sun, satellite, and ground station to within a small fraction of a degree of arc, we can reasonably assume that the direction of V_2 is known with negligible error and that the uncertainty in V_1 is a fraction of a degree that does not exceed the errors in the observations themselves. Our main sources of observational error are in the determination of the aspect angles α and β, so we will consider the effect of small errors in α and β on the orientation of the spin axis. The principal errors we can expect in α and β are

1. Error in measuring the angle α between the spin axis and the normal to the mirrors. This is a mechanical error of approximately $\pm 1/2°$. There is also the possibility of some wobble in the spin motion, although the nutation damper should minimize wobbling.

2. Error in determining the direction cosines of the sun with respect to body axes. This gives an error in the other cone semi-angle, β. This error has been about $\pm 1/4°$ on most passes, but has been considerably worse on others, possibly due to interference from earth-shine. Since the sun subtends an angle of $1/2°$ at the satellite, not much increase in accuracy can be expected with the solar aspect system.

3. Error in determining the center of the sun's image on the ground, i.e., ascertaining the time of the center of the train of light flashes. This error has the effect of changing the direction of the reference vector $\bar{V}_1$ by $\pm \approx 1/4°$, since the semi-angle the sun subtends at the ground station is $1/2°$ to $3/4°$, depending on the flatness of the mirrors. We will simplify our error analysis by considering the effect of this error to be an additional error in α.

4. The angular errors of $\bar{V}_1$ resulting from the position ephemeris for the satellite would enter into the error analysis in a fashion similar to item 3.

3.1 *Analysis*

First let α_0 and β_0 be the values of the cone semiangles that determine the true orientation of the spin axis, and let $\hat{V}_0$ be the unit vector in the direction of the true spin axis. Similarly, let α and β be the observed values of the cone semi-angles that determine the orientation of an estimated unit vector $\hat{V}$. We will assume that α and β differ only slightly from α_0 and β_0, respectively, so that $\hat{V}$ and $\hat{V}_0$ differ only slightly. Once we have determined the reference vectors $\hat{V}_1$ and $\hat{V}_2$ and the cone semi-angles α and β, the orientation of the spin-axis unit vector is calculated from (4)–(6). We are assuming that $\hat{V}_1$ and $\hat{V}_2$ are known with negligible error and are considering only the effect on orientation of errors in α and β, so x_1, y_1, z_1 and x_2, y_2, z_2 are fixed quantities. The two unit vectors corresponding to the true and observed spin-axis orientation are thus

$$\hat{V}_0(\alpha_0, \beta_0) = x_0\hat{i} + y_0\hat{j} + z_0\hat{k}$$

$$\hat{V}(\alpha, \beta) = x\hat{i} + y\hat{j} + z\hat{k}.$$

We let $\delta\bar{V}$ be the small vector difference between $\hat{V}$ and $\hat{V}_0$, so we can write

$$\hat{V} = \hat{V}_0 + \delta\bar{V}.$$

For small changes of α and β we can approximate $\delta\bar{V}$ by the first- and second-order terms of a Taylor series.

$$\hat{V} = \hat{V}_0 + \delta\bar{V}$$

$$\approx \hat{V}_0 + \frac{\partial\hat{V}}{\partial\alpha}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\alpha + \frac{\partial\hat{V}}{\partial\beta}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\beta + \frac{1}{2}\frac{\partial^2\hat{V}}{\partial\alpha^2}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\alpha^2$$

$$+ \frac{1}{2}\frac{\partial^2\hat{V}}{\partial\beta^2}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\beta^2 + \frac{\partial^2\hat{V}}{\partial\alpha\,\partial\beta}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\alpha\,\delta\beta.$$

The small angle between $\hat{V}$ and $\hat{V}_0$, denoted ϵ, is found from

$$\hat{V}\cdot\hat{V}_0 = \cos\epsilon$$

so

$$1 + \hat{V}_0\cdot\left(\frac{\partial\hat{V}}{\partial\alpha}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\alpha + \frac{\partial\hat{V}}{\partial\beta}\bigg|_{\substack{\alpha=\alpha_0\\\beta=\beta_0}} \delta\beta\right)$$

$$+ \frac{1}{2} \hat{V}_0 \cdot \left(\frac{\partial^2 \hat{V}}{\partial \alpha^2} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\alpha^2 + \frac{\partial^2 \hat{V}}{\partial \beta^2} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\beta^2 + 2 \frac{\partial^2 \hat{V}}{\partial \alpha\,\partial\beta} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\alpha\,\delta\beta \right) \quad (7)$$

$$+ 0(\delta\alpha^3, \delta\beta^3)$$

$$= \cos \epsilon$$

$$= 1 - \frac{\epsilon^2}{2} + 0\ (\epsilon^4).$$

The effect of variations in $\delta\alpha$ and $\delta\beta$ on $\cos \epsilon$ is of second order, so the term containing first derivatives of $\hat{V}$ in (7) vanishes, as can be readily verified. Thus for ϵ we find

$$\epsilon^2 = - \hat{V}_0 \cdot \left(\frac{\partial^2 \hat{V}}{\partial \alpha^2} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\alpha^2 + \frac{\partial^2 \hat{V}}{\partial \beta^2} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\beta^2 + 2 \frac{\partial^2 \hat{V}}{\partial \alpha\,\partial\beta} \bigg|_{\substack{\alpha=\alpha_0 \\ \beta=\beta_0}} \delta\alpha\,\delta\beta. \quad (8)$$

The reference vectors $\hat{V}_1$ and $\hat{V}_2$ define a plane, so there is no loss of generality if we choose a ξ axis in the $\hat{V}_2$ direction and let $\hat{V}_1$ lie in the ξ, η plane where we transformed from the x, y, z axes to ξ, η, ζ axes. The details of the transformation are unnecessary, since we seek only magnitudes of error in orientation and not direction. If $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$ are the unit vectors in the ξ, η, ζ directions, respectively, then

$$\hat{V}_1 = \cos \gamma\ \hat{e}_1 + \sin \gamma\ \hat{e}_2$$

$$\hat{V}_2 = \hat{e}_1$$

where

$$\hat{V}_1 \cdot \hat{V}_2 = \cos \gamma$$

i.e., γ is the angle between the reference vectors $\bar{V}_1$ and $\bar{V}_2$. We infer from (1)–(3) that the unit vector $\hat{V}$ in the direction of the spin axis has the direction cosines

$$\xi = \cos \beta$$

$$\eta = \frac{\cos \alpha - \cos \gamma \cos \beta}{\sin \gamma}$$

$$\zeta = \pm \left[1 - \cos^2 \beta - \left(\frac{\cos \alpha - \cos \gamma \cos \beta}{\sin \gamma} \right)^2 \right]^{\frac{1}{2}}$$

where α and β are the angles between the spin axis and $\hat{V}_1$ and $\hat{V}_2$, respectively.

After rewriting (8) with the help of ξ, η, ζ we obtain for ϵ

$$\epsilon^2 = \frac{\delta\alpha^2 - 2\left(\dfrac{\cos\gamma - \cos\alpha_0 \cos\beta_0}{\sin\alpha_0 \sin\beta_0}\right)\delta\alpha\,\delta\beta + \delta\beta^2}{1 - \left(\dfrac{\cos\gamma - \cos\alpha_0 \cos\beta_0}{\sin\alpha_0 \sin\beta_0}\right)^2} \tag{9}$$

provided

$$\frac{\cos\gamma - \cos\alpha_0 \cos\beta_0}{\sin\alpha_0 \sin\beta_0} \neq 1. \tag{10}$$

The restriction (10) merely assures that the two cones have two intersections, i.e.,

$$\gamma > |\alpha_0 - \beta_0|$$

$$\gamma < \alpha_0 + \beta_0.$$

Vanishing of the denominator in (9) occurs for the singular instances when the two cones intersect in one line only or when they coincide. In these situations no solution exists if either α_0 or β_0 is varied, and so the expression for ϵ^2 is not defined.

There are four examples which are of interest, and these are examined next.

3.2 *Solar Aspect and Mirror Observation*

Usually the two reference vectors determined will be the satellite-sun line and the normal to the mirror, angles α and β being the angles between these reference lines and the spin axis.

Let $\hat{V}_2$ be directed toward the sun. The satellite is intended to be aligned nearly perpendicular to the ecliptic plane, so we will set $\beta_0 = 90°$. Equation (9) becomes

$$\epsilon^2 = \frac{\delta\alpha^2 - 2\dfrac{\cos\gamma}{\sin\alpha_0}\delta\alpha\,\delta\beta + \delta\beta^2}{1 - \left(\dfrac{\cos\gamma}{\sin\alpha_0}\right)^2}$$

and since there are two mirrors on the satellite having normals which are inclined 68° and 95° to the spin axis, α_0 can take on only the values 68° and 95°. The variation in ϵ with changes in γ is displayed in Fig. 4.
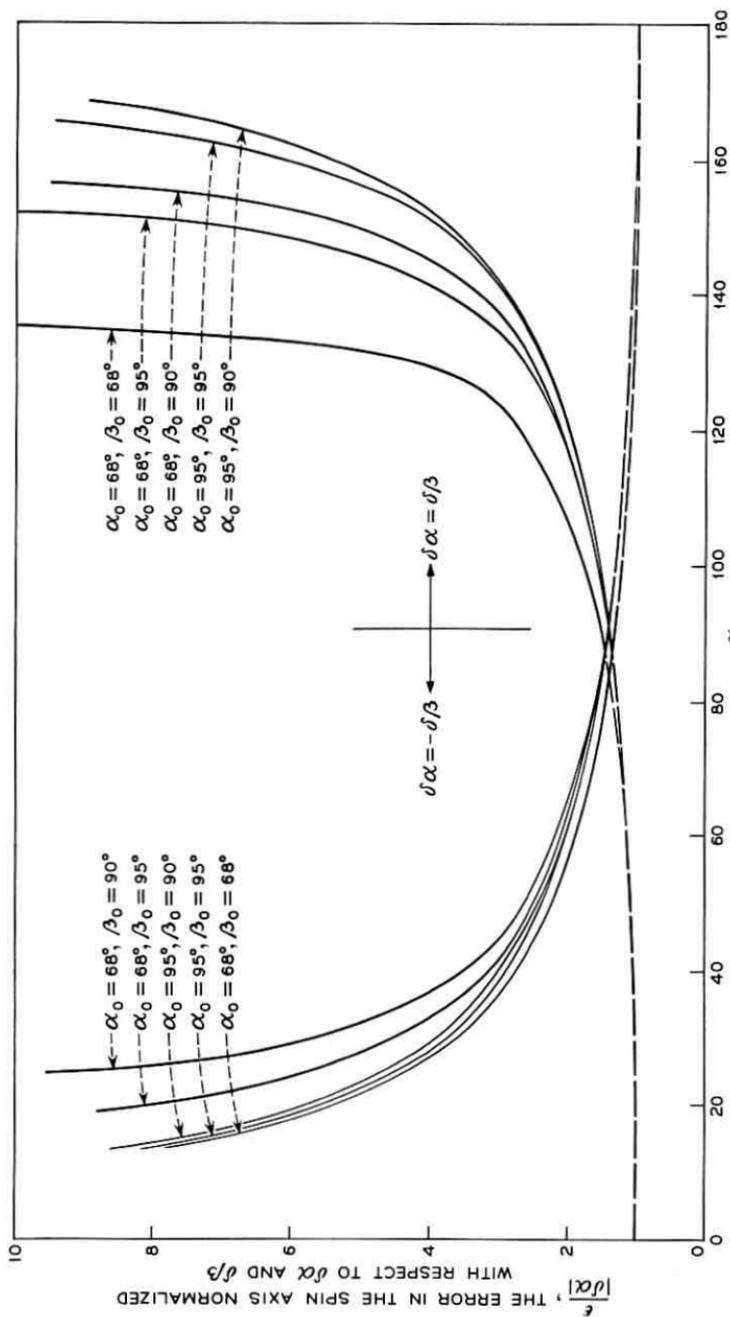
Fig. 4 — Variation of geometric errors as a function of γ.

3.3 *Two-Mirror Observations*

The three possible combinations of two mirror observations are

$$\alpha_0 = 68°, \qquad \beta_0 = 68°$$

$$\alpha_0 = 68°, \qquad \beta_0 = 95° \text{ (or vice versa)}$$

$$\alpha_0 = 95°, \qquad \beta_0 = 95°.$$

The results of varying γ for these three possible combinations, among others, are also shown in Fig. 4.

The curves of Fig. 4 show that if the angle between $\hat{V}_1$ and $\hat{V}_2$ is between 50° and 130° then

$$|\epsilon| < +2\sqrt{\delta\alpha^2 + \delta\beta^2}\,.$$

If $\delta\alpha = \delta\beta$ and $50° < \gamma < 130°$, then

$$\epsilon < 2.828\ \delta\alpha.$$

Thus if the observational errors in α and β are 1°, the error in determining spin-axis orientation would be less than 2.828° provided γ, the angle between $\bar{V}_1$ and $\bar{V}_2$, satisfies the above inequality.

IV. RESULTS

The motion of the spin axis of the Telstar satellites has been derived from observations of flashes of sunlight reflected from the mirrors attached to the satellite and from use of telemetry data from the solar aspect cells mounted on the satellite. Figs. 5 and 6 and Table I sum-
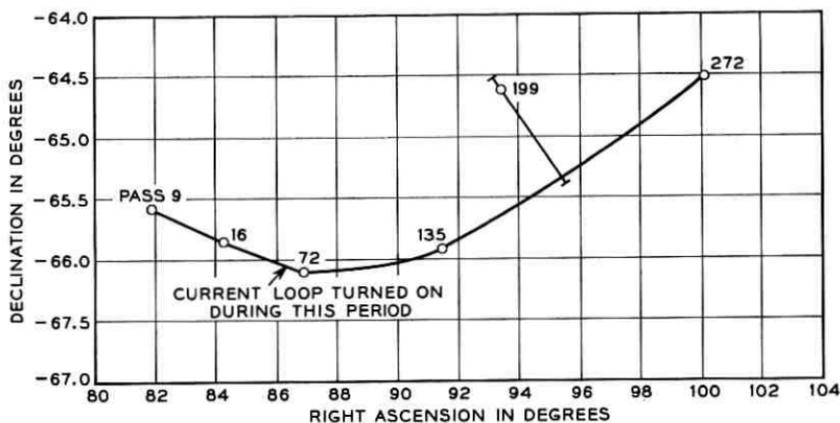

Fig. 5 — Motion of spin axis in geocentric coordinates.
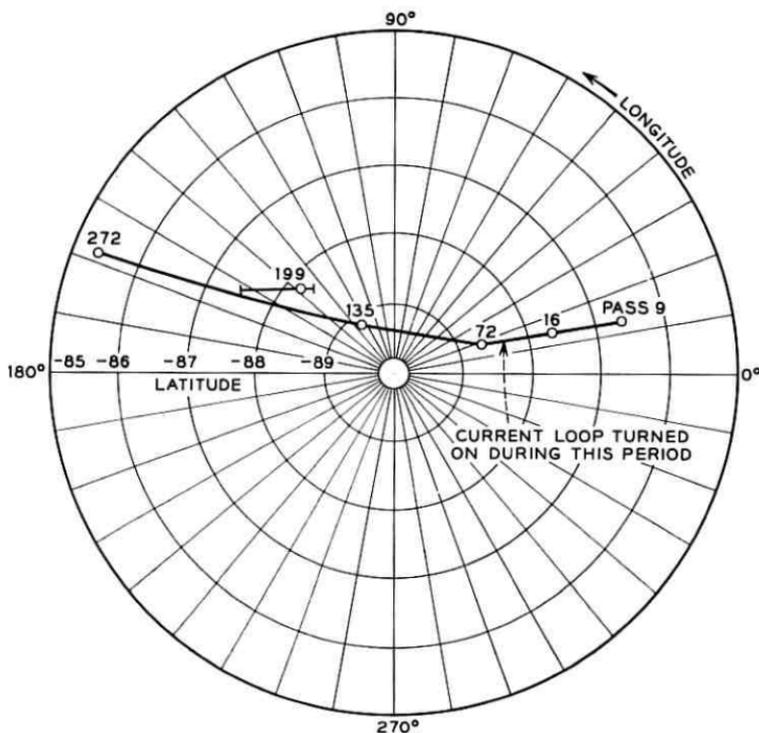
Fig. 6 — Motion of spin axis in polar coordinates referred to ecliptic plane.

marize the early precession of the Telstar I satellite.* The portion between passes 16 and 72 covers the period of time the torque coil on the satellite is known to have been activated (presumably from orbits 65 thru 70).

The accuracy obtained to date appears to be much higher than was originally anticipated. It is estimated that the angular error in determining the spin-axis orientation is less than $1/2°$ for all observations except pass 199. The orientation determined for pass 199 is questionable due both to poor geometric relations and to many inaccurate frames of telemetry data. Some more detailed information is given below for two of the fixes reported.

No high-frequency precession or "coning" has been detected so far, so apparently the nutation damper is operating satisfactorily.

---

* Right ascension is measured eastward from vernal equinox in the earth's equatorial plane; declination is measured plus or minus from earth's equatorial plane; latitude is measured plus or minus from ecliptic plane; longitude is measured eastward from vernal equinox in the ecliptic plane.

TABLE I — SPIN-AXIS ORIENTATION

Pass No.	Date	Universal Time	Right Ascension°	Declination°	Latitude°	Longitude°
7, 8, 9	7/10	0849	81.96	−65.57	−86.60	13.22
16	7/12	0214	84.39	−65.81	−87.62	15.52
72	7/18	0553	86.86	−66.10	−88.67	18.55
135	7/25	0328	91.22	−65.86	−89.19	125.95
199	8/1	0345	93.41	−64.58	−88.18	139.87
272	8/9	0341	100.08	−64.51	−85.36	158.43

Sensitivity checks have shown that timing mirror flashes to the nearest second is accurate enough.

4.1 *Attitude Calculations*

The times at which mirror flashes have been observed are the basis for attitude calculations; early observations for the Telstar I satellite are listed in Table II.

4.1.1 *Passes 7, 8, 9*

Two mirror flashes were observed on each of passes 7, 8, and 9, thus providing a fix for each of these passes. The accuracy of these three fixes was not sufficiently great to permit determining the motion between passes, so all three were averaged to obtain a single fix. The motion between successive passes is about 0.03°, probably below the precision of the present methods to resolve, although additional efforts were made on the data for passes 7, 8, and 9 to try to obtain such precision.

The results of various combinations of mirror flashes are given in Table III to show the scatter obtained. This exercise demonstrates that

TABLE II — MIRROR FLASH OBSERVATIONS

Observation No.	Pass No.	Date	Universal Time, h.m.s.	Mirror Observed
M1	7	7/11	0235 + 26	68°
M2	7	7/11	0250 + 47	68°
M3	8	7/11	0518 + 08	68°
M4	8	7/11	0535 + 36	68°
M5	9	7/11	0817 + 22	68°
M6	9	7/11	0821 + 30	95°
M7	16	7/12	0214 + 19	68°
M8	72	7/18	0553 + 26	68°
M9	135	7/25	0328 + 07	68°
M10	136	7/25	0609 + 31	68°
M11	199	8/1	0345 + 37	68°
M12	272	8/9	0340 + 58	68°

TABLE III — RESULTS FOR PASSES 7, 8, AND 9

Mirror Observations	Right Ascension°	Declination°
M1 & M2	82.09	−65.73
M3 & M4	82.59	−65.70
M5 & M6	81.00	−65.23
M1 & M5	82.97	−65.74
M2 & M5	81.27	−65.25
M3 & M6	81.77	−65.77

adequate attitude determination is possible from mirror data alone when solar aspect readings are unavailable or considered unreliable.

4.1.2 *Pass 135*

Mirror flashes were obtained for each of passes 135 and 136 and were used to obtain fix number 1 given in Table IV. Additionally, all the solar aspect data for pass 135 was combined with the mirror flash for pass 135, and 17 selected aspect readings averaged to obtain fix number 2. The average of these two was used for the pass 135 attitude fix.

4.1.3 *Remaining Observations*

All other attitude fixes reported here were obtained by combining one mirror observation with all the solar aspect telemetry data for the pass during which the mirror was observed.

4.2 *Residual Magnetic Moment*

The average residual magnetic moment has been determined approximately by comparing the observed spin axis precession against theoretical predictions. The value obtained is 0.5 ampere-turn-meter^2, which is of the same order of magnitude as the pre-launch estimates.

V. CONCLUSION

The conclusion drawn from these results is that it clearly is possible to track the motion of the spin axis with mirrors and solar aspect cells,

TABLE IV — RESULTS FOR PASS 135

No.	Right Ascension°	Declination°
1	91.10	−65.80
2	91.33	−65.97

and that controlling the spin-axis orientation slowly by means of a torque coil is feasible. Much more extensive tests of the procedure have been performed than described here, which cover the remaining life of the Telstar I satellite and the Telstar II satellite. The computational procedure has been modified so that no data from the solar aspect cells are required. In general, the resulting attitude angles fall well within the error estimates given in Section 3.3. A report on this work is being prepared by L. C. Thomas.

REFERENCE

1. Courtney-Pratt, J. S., Hett, J. H., and McLaughlin, J. W., Optical Measurements on *Telstar* to Determine the Orientation of the Spin Axis and the Spin Rate, J. Soc. Motion Picture and Television Engineers, 72, June, 1963, pp. 462–484; the abstract of this paper is reproduced here to briefly introduce the design and operation of the equipment used to detect the flashes of light from the mirrors:
    "One plane mirror and two faceted mirrors were fitted to Telstar. A photoelectric telescope was mounted on a radar antenna pedestal. Appropriate recording gear was built. The 12-in. aperture Cassegrain photoelectric telescope could be pointed at the satellite directly from prediction tapes, and the pointing accuracy could on occasion be improved by homing on the microwave beacon on the satellite or by applying corrections from visual observation through other telescopes with apertures of up to 6 in. It has been possible to detect the flashes of sunlight reflected from the mirrors on Telstar at slant ranges up to 3,700 miles. Directly from pairs of these observations, or by combining this information with telemetry data from the solar sensors on the satellite, it has been possible to determine the orientation of the spin axis of Telstar and to examine how this is moving in time. This normal precession of the spin axis is due mainly to the interaction of the residual magnetic moment of the satellite with the magnetic field of the earth. It was also possible to study how the spin axis moved when the torque coil was operated.
    Measurements have been made of the spin rate of the satellite and of its rate of decay. The spin rate will drop to $\frac{1}{10}$ of its initial value in about 750 days.
    Photographs of the glints have been taken, and a study is now being made of the feasibility of combining precise photoelectric observation of the time of the flashes with high-precision photographic observation of the instantaneous position of the satellite relative to the fixed star background."

For a more complete description, the reader will wish to consult the full article.

Contributors to This Issue

L. K. ANDERSON, B.Eng. (Engineering Physics), 1957, McGill University; Ph.D. (Electrical Engineering), 1962, Stanford University; Bell Telephone Laboratories, 1961—. He first engaged in work on problems of optical modulation and detection in microwave solid-state devices. Particular areas included modulation of light using ferrimagnetic resonance in garnets, development of high-speed semiconductor photodiode detectors and electro-optic light modulators. He presently supervises the microwave ferrite device group. Member, IEEE.

VACLAV E. BENEŠ, A.B., 1950, Harvard College; M.A. and Ph.D., 1953, Princeton University; Bell Telephone Laboratories, 1953—. Mr. Beneš has been engaged in mathematical research on stochastic processes, traffic theory, and servomechanisms. In 1959–60 he was visiting lecturer in mathematics at Dartmouth College. He is the author of *General Stochastic Processes in the Theory of Queues* (Addison-Wesley, 1963). Member, American Mathematical Society, Association for Symbolic Logic, Institute of Mathematical Statistics, Society for Industrial and Applied Mathematics, Mind Association and Phi Beta Kappa.

D. P. BORENSTEIN, B.E.E. and B.A., 1960, M.E.E., 1962, New York University; Bell Telephone Laboratories, 1960—. He was first engaged in the investigation of the statistical and spectral nature of digit simulation in TOUCH-TONE signaling systems. At present he is working on a development and testing program for the Videotelephone. Member, Eta Kappa Nu, Tau Beta Pi and Phi Beta Kappa.

PAUL T. BRADY, B.E.E., 1958, Rensselaer Polytechnic Institute; M.S.E.E., 1960, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1961—. His work in human factors engineering has been concerned with studies of speech and voice-operated devices, especially as applied to satellite communication circuits.

KENNETH BULLINGTON, B.S., 1936, University of New Mexico; M.S., 1937, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1937—. Mr. Bullington has worked on transmission

engineering problems on wire, radio and submarine cable systems. He is now Head, Transmission Projects Engineering Department. In 1956 he received the Morris Liebmann Memorial Prize of the Institute of Radio Engineers and the Franklin Institute's Stuart Ballantine Medal for contributions in tropospheric transmission and its application to practical communications systems. Fellow, IEEE; member, Phi Kappa Phi, Sigma Tau and Kappa Mu Epsilon.

CHARLES J. BYRNE, B.S.E.E., 1957, Rensselaer Polytechnic Institute; M.S., 1958, California Institute of Technology; Bell Telephone Laboratories, 1958–1962; Bellcomm, Inc., 1962—. At the Laboratories he investigated fast transistor logic, instrument noise in seismometers, and synchronization of digital systems. At Bellcomm he is engaged in studies of photographic reconnaissance systems and systems for Apollo site certification. Member, IEEE, Sigma Xi, Eta Kappa Nu and Tau Beta Pi.

HAROLD E. CURTIS, B.S. and M.S., 1929, Massachusetts Institute of Technology; American Telephone & Telegraph Co., 1929–34; Bell Telephone Laboratories, 1934—. He has specialized in work on transmission problems related to multichannel carrier telephony, including microwave radio relay, coaxial cable and waveguide systems.

SIDNEY DARLINGTON, B.S., 1928, Harvard College; B.S. in E.E., 1929, Massachusetts Institute of Technology; Ph.D., 1940, Columbia University; Bell Telephone Laboratories, 1929—. He has been engaged in research in applied mathematics with emphasis on network theory and military electronics. He holds more than 20 patents in these fields. Fellow, IEEE; member, AIAA.

M. DiDOMENICO, JR., B.S., 1958, M.S., 1959, Ph.D., 1963, Stanford University; Bell Telephone Laboratories, 1962—. He has been engaged in investigations on electro-optics and electro-optic light modulators. He has also worked on photoconductive detectors of microwave modulated light, and has been concerned with the physical processes involved in solid-state photodetectors. Member, American Physical Society, IEEE, Sigma Xi and Tau Beta Pi.

J. W. EMLING, B.S. in E.E., Univ. of Pennsylvania, 1925; Development and Research Department of American Telephone and Telegraph Co., 1925–34; Bell Telephone Laboratories, 1934—. While at A.T.&T.,

Mr. Emling was particularly concerned with transmission standards and with developing a system of effective transmission rating. He continued this work at Bell Laboratories. In World War II he was concerned with studies in the field of underwater acoustics. Subsequently he has been concerned with systems engineering studies in the fields of engineering economy, voice-frequency transmission, rural carrier, radio and television. He is currently Executive Director of the Transmission Systems Engineering Division, with responsibility for the systems engineering aspects of transmission over wire and radio. Member, Acoustical Society of America, IEEE, Eta Kappa Nu and Tau Beta Pi.

Louis H. Enloe, B.S., 1955, M.S., 1956, Ph.D., 1959, University of Arizona; Bell Telephone Laboratories, 1959—. As a member of the radio systems research department, he has worked in the field of modulation and noise theory with particular emphasis on problems associated with space communication. Member, IEEE, Tau Beta Pi, Sigma Xi, Pi Mu Epsilon, Sigma Pi Sigma and Phi Kappa Phi.

Gerald D. Haynie, B.S., 1956, Virginia Polytechnic Institute; M.E.E., 1961, New York University; Bell Telephone Laboratories, 1956—. Mr. Haynie has been engaged in the transmission measurements development area and at present is supervisor of a group concerned with the automation of laboratory measurement systems. Member, IEEE, Tau Beta Pi and Eta Kappa Nu.

George K. Helder, B.S. (Business), 1952, B.S.E.E., 1958, University of Colorado; M.E.E., 1960, New York University; Bell Telephone Laboratories, 1958—. He was first engaged in exchange area transmission, including methods of subscriber loop testing. More recently he has been concerned with the problems of echo control on telephone connections. At present, he supervises a group dealing with the transmission aspects of the toll telephone network. Member, Tau Beta Pi, Eta Kappa Nu and IEEE.

D. W. Hill, B.S.C.E., 1951, Duke University; M.S.C.E., 1956, California Institute of Technology; Ph.D., (Engineering Mechanics), 1960, Stanford University; Bell Telephone Laboratories, 1959–63; Assistant Professor of Engineering Mechanics, Duke University, 1963—. His first assignment was the study of submarine cable dynamics. This was followed by work on attitude control problems of satellites, in-

cluding the system of attitude determination and control used in the Telstar experiments. Member, Sigma Xi and Tau Beta Pi; associate member, ASME.

BARRY J. KARAFIN, B.S.E.E., 1961, Moore School of University of Pennsylvania; M.E.E., 1963, New York University; Bell Telephone Laboratories, 1961—. He was first engaged in the study of systematic jitter in a chain of digital regenerators. Since then he has worked on voltage-controlled oscillators, and he is currently working on computer languages and compilers.

EDMUND T. KLEMMER, B.S., 1944, Webb Institute of Naval Architecture; M.A., 1949, Ph.D., 1952, Columbia University; Bell Telephone Laboratories, 1962—. He has studied customer dialing behavior with particular emphasis on the frequency and types of errors made in dialing. He has also developed a preference scaling method based on the proximity scaling approach. Most recently Mr. Klemmer has been concerned with the influence of satellite transmission delay and echo suppressor action upon the quality of telephone transmission. Fellow, Society of Engineering Psychologists; member, American Psychological Association, Human Factors Society and Sigma Xi.

TINGYE LI, B.Sc., 1953, University of Witwatersrand (South Africa); M.S., 1955, and Ph.D., 1958, Northwestern University; Bell Telephone Laboratories, 1957—. He has been engaged in studies of microwave antennas and microwave propagation. Recently he has been primarily concerned with work on optical masers. Member, IEEE, Eta Kappa Nu and Sigma Xi.

JACK M. MANLEY, B.S. (Electrical Engineering), 1930, University of Missouri; Bell Telephone Laboratories, 1930—. He was first concerned with theoretical and experimental studies of nonlinear electric circuits. He later worked with new multiplex methods for communication systems, including early research work on PCM. Afterward, he was engaged in transmission line research, and at present he is working on noise problems in digital transmission systems. Fellow, IEEE; member, Sigma Xi, Tau Beta Pi and Eta Kappa Nu.

STEWART E. MILLER, B.S. and M.S., 1941, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1941—. He first worked on coaxial carrier repeaters and later shifted to microwave radar systems development. At the close of World War II he returned to coaxial

carrier repeater development until 1949, when he joined the radio research department. There his work has been in the fields of circular electric waveguide communication, microwave ferrite devices, and other components for microwave radio systems. As Director, Guided Wave Systems Research, he heads a group engaged in research on communication techniques for the millimeter wave and optical regions. Fellow, IEEE.

DOREN MITCHELL, B.S., 1925, Princeton University; American Telephone & Telegraph Co., 1925–34; Bell Telephone Laboratories, 1934—. Mr. Mitchell's early work was concerned with field studies of transmission on long telephone and radio circuits, including work on various types of voice-operated devices. During World War II he worked on military projects, including transmission systems and the problem of laying wire from airplanes. He also founded the Somerset Mechanics School to provide vocational training to residents of that county. Since the war, he has worked on radio systems and data transmission systems. Now working on satellite communication systems as Head, Satellite Systems Studies Department. Fellow, IEEE; licensed professional engineer; member, AFCEA and AAAS.

ROBERT R. RIESZ, A.B., 1924, Ripon College; M.A. (Physics), 1926, University of Wisconsin; Bell Telephone Laboratories, 1926–63. His earliest work was on vibratory mechanics, which led to research in the physics of speech and hearing, including such projects as the artificial larynx and visible speech. Prior to and during World War II, he was engaged in developing speech processing systems such as the Voder and Vocoder. More recently he engaged in the field of human factors as applied to telephone systems, including the development of simulation as a means of studying user reaction to new communications systems. Since his retirement from the Laboratories, he has been Associate Professor, Department of Physics, Union College, Barbourville, Kentucky. Fellow, American Physical Society; Senior Member, IEEE, and Chairman of Professional Group on Human Factors in Electronics; member, Human Factors Society.

DUDLEY B. ROBINSON, JR., B.S.E., 1950, M.S.E., 1951, Princeton University; Bell Telephone Laboratories, 1954—. He first engaged in exchange carrier transmission systems studies. He has since worked on design of test equipment used on early PCM field trials and later used for T-1 carrier trials. At present he is engaged in the study of a high-

speed PCM coder-decoder for use with FDM (frequency division multiplex) and TV signals, and field tests of the coder-decoder.

PETER E. ROSENFELD, Sc.B., 1957, Brown University; M.S., 1959, Harvard University; Bell Telephone Laboratories, 1959—. Mr. Rosenfeld's first assignment with the Laboratories was on the design of the 20–20,000-cps automated transmission measuring set. He is presently engaged in the design of an automated set which covers the frequency range from 20 cps to 250 mc and uses an on-line digital computer to perform the control functions. Member, Tau Beta Pi.

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. He is currently involved in a study of the signal-theoretic properties of nonlinear systems. Member, IEEE, Society for Industrial and Applied Mathematics, Eta Kappa Nu, Sigma Xi and Tau Beta Pi.

D. T. YOUNG, B.S., 1956, M.E.E., 1960, University of Oklahoma; Bell Telephone Laboratories, 1960—. He initially worked on mode conversion problems in multimode waveguides. At present he is working on a solid state repeater for a waveguide transmission system. Member, IEEE, Tau Beta Pi, Eta Kappa Nu and Sigma Xi.

B.S.T.J. BRIEFS

On the Spectrum of Optical Waves Propagated through the Atmosphere

By D. C. HOGG

It is well known from our day-to-day experience that distant objects appear to shimmer, especially on still, hot days when relatively large temperature and humidity gradients may exist. It is true, however, that refractive index gradients are in the air to some degree at all times; that is, they may be present at night and also when the air is disturbed by the winds. It is not surprising therefore that the power received at some distance from an optical source fluctuates randomly and possesses a characteristic low-frequency power spectrum.

The purpose of this note is to discuss typical low-frequency spectra resulting from propagation of 0.63-micron radiation over a 2.6-km path. A vertically polarized helium-neon maser[1] of power output 10 mw and a reflecting telescope of 9-cm diameter comprise the source. With all modes of the maser propagating, the beam spreads to an ill-defined and ever-changing disk of about 25-cm diameter.* The receiver is a refracting telescope of 5-cm diameter with associated filters, polarizers, and attenuators which feed a photomultiplier; it is located in the dense central region of the transmitted beam. The beamwidth of the receiver is large compared with that of the transmitter. For measurement of the power spectrum, the output of the photomultiplier is taken to a wave analyzer whose bandwidth is 4 cps.

A typical power spectrum is shown in Fig. 1, the measured points being indicated by open circles.† The abscissa, $F = f - f_o$, is sideband frequency, and the ordinate indicates relative received power. On this scale the direct-current level is 133 units; thus the power in the lower frequency components is about 13 db below the dc level. The curve has been extrapolated from 10 cps to 1 cps. These data were obtained with the full aperture of the source telescope illuminated.

At this point it is well to note that when the source is brought within

---

* Dependent upon weather conditions.

† Similar data, using an incoherent source, were reported by V. I. Tatarski in Ref. 2.
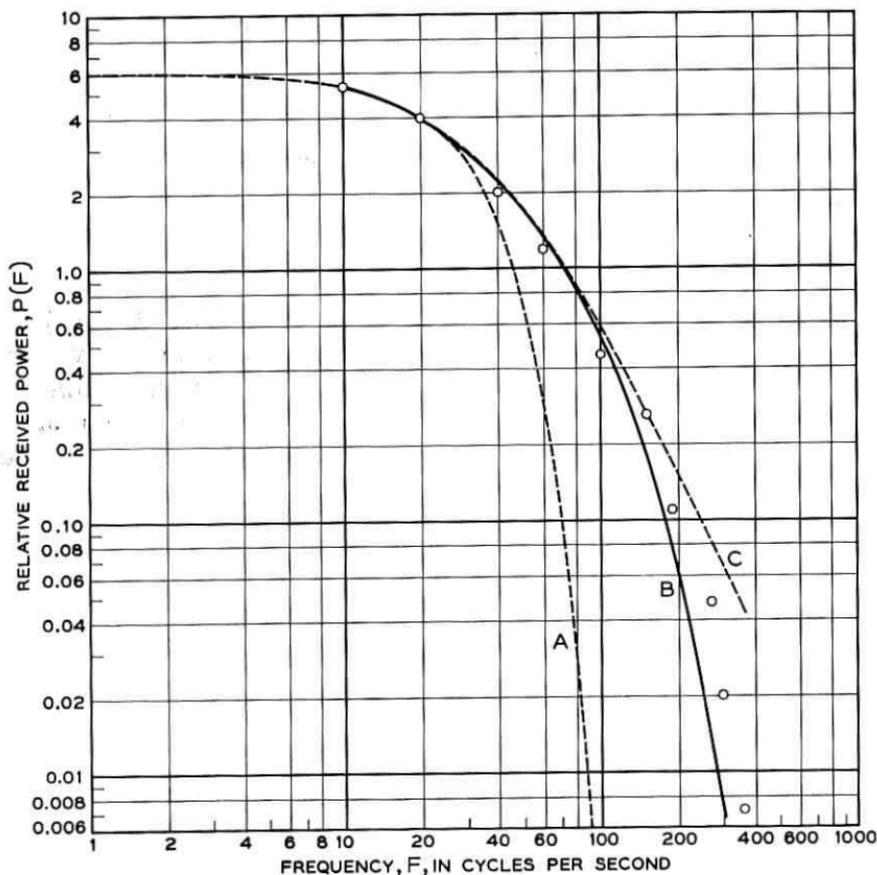
Fig. 1 — Low-frequency spectrum: $\lambda = 0.63\ \mu$ for a 2.6-km path length.

the laboratory housing the receiver and the same measurement procedure is followed, no such spectrum is observed.

Three theoretical distributions, $P(F)$, shown as curves in Fig. 1, accompany the experimental data. Curve A is a Gaussian, B an exponential; in C, $P(F)$ is of the form $(1 + KF^2)^{-1}$. In all cases, the theoretical curves have been fitted to the experimental data at $F = 30$ cps, where the spectrum has fallen to one-half its very-low-frequency value. The exponential appears to best represent the data.

Fig. 2 shows the effect of change in source beamwidth on the width (and shape) of the spectrum. For curve 2, the source beamwidth is about ten times that for curve 1. One notes that the half width increases from 40 to 120 cps with that increase in beamwidth. From this measure-
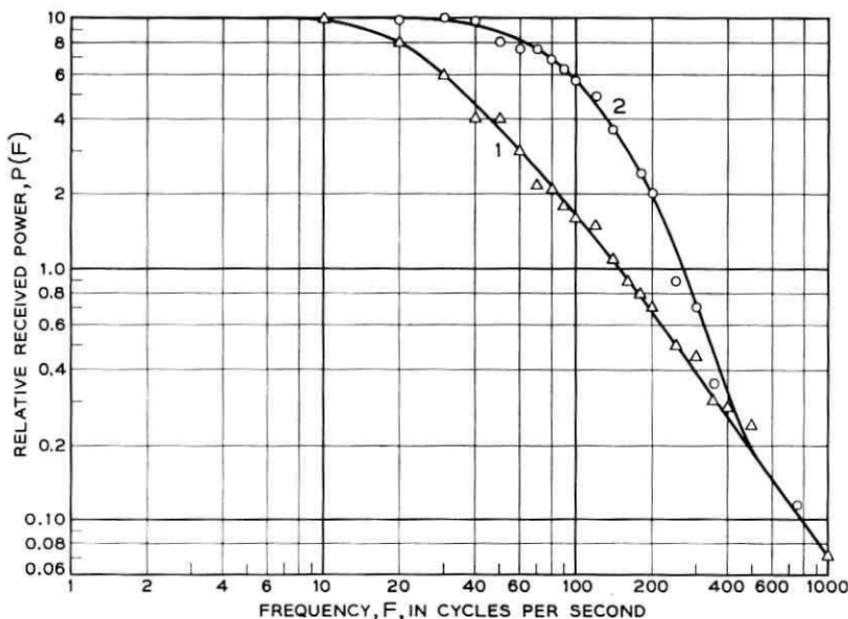
Fig. 2 — Low-frequency spectrum: $\lambda = 0.63\ \mu$ for a 2.6-km path length.

ment we deduce that the fluctuation rate is roughly proportional to the square root of the beamwidth. It is interesting that the two spectra appear to coalesce at the higher frequencies; however, in interpreting these data, one must recall that the near field of the source extends well toward the receiver in the case of curve 1.

Of course, when one listens to the detected envelope of the optical wave, it is heard as low-frequency audio noise.

REFERENCES

1. White, A. D., and Rigden, J. D., Continuous Gas Maser Operation in the Visible, Proc. I.R.E. Correspondence, **50**, July, 1962, p. 1697.
2. Tatarski, V. I., *Wave Propagation in a Turbulent Medium*, McGraw-Hill, New York, 1961.

A Relation Between the Basis Functions of Periodically Varying Nondissipative Circuits

By SIDNEY DARLINGTON

This note relates to circuits of linear, periodically time-varying, positive capacitors and inductors. More specifically, it notes a property of the basis functions, or natural modes which describe the transient behavior of the circuits. Some other properties of the basis functions are described in Ref. 1 in this issue.

In accordance with Ref. 1

$$I = (S + pCp)\Phi, \qquad E = p\Phi \tag{1}$$

where I and E are column matrices, or vectors representing the excitation currents into and the response voltages at the various nodes, and Φ is an auxiliary vector variable. The matrices C and S are the node matrices of the capacitances and of the reciprocals of the inductances. They are symmetric and at least positive semidefinite. We will assume that C is positive definite. If it is not so originally, a similar equation with a positive definite C can be derived from (1), for example by the transformations described in the Appendix of Ref. 1. The symbol p represents differentiation (*not* frequency) and operates on all quantities following it.

Setting $I = 0$ in (1) makes it a homogeneous, second-order, vector differential equation in, say n dimensions, with periodically varying coefficients. Thus the well-known Floquet-Poincaré theorem requires the solution to be as follows

$$\Phi = \sum_{\sigma=1}^{2n} k_\sigma \varphi_\sigma, \qquad \varphi_\sigma = H_\sigma(t)e^{s_\sigma t}. \tag{2}$$

Here the k_σ's are arbitrary scalar constants and the φ_σ's are the basis functions, or natural modes. The characteristic exponents s_σ are scalar constants. The coefficients H_σ are time-varying vectors. When the s_σ's are all different, the H_σ's vary periodically. Otherwise, they are at most polynomials in t with periodically varying vector coefficients.

The real parts of the s_σ's indicate the damping of the basis functions. When the inductors and capacitors are fixed, all basis functions are undamped. However, when the components vary periodically some of the basis functions may have nonzero damping (as is well known). Some bounds on the damping are derived in Ref. 1. It is also known that

the sum of all the characteristic exponents must be zero (when the circuit is nondissipative).

The purpose of this note is to point out the following: *Corresponding to any nondissipative circuit of periodically varying positive inductors and capacitors, any characteristic exponents which have nonzero real parts occur in equal positive and negative pairs.* Since complex exponents must occur in conjugate pairs, it follows that all the s_σ's must fit into pairs or quadruplets of the following sorts:

$$
\begin{array}{llll}
+i\omega_k & +\alpha_j & +\alpha_h + i\omega_h & \quad(3) \\
-i\omega_k\;, & -\alpha_j\;, & +\alpha_h - i\omega_h & \\
 & & -\alpha_h - i\omega_h & \\
 & & -\alpha_h + i\omega_h. &
\end{array}
$$

The theorem follows at once from certain general properties of adjointly related differential equations, which are stated below without derivation. The solution of the nonhomogeneous equation (1) may be expressed in terms of functions of the excitation time, τ, and the response time, t. The adjoint equation is the equation whose solution corresponds to interchanging the two times, τ and t. It can be shown that the characteristic exponents of the adjoint equation are the negatives of those of the original equation. It can also be shown that equation (1) is self-adjoint. (The equation is its own adjoint.) Thus the negatives of the characteristic exponents of the original equation are also the characteristic exponents of the original equation.

When the components are fixed, vectors I and E are related by a set of odd rational functions of a frequency variable. Then Φ, whose derivative is E, is related to I by even rational functions. It appears that, more generally, self-adjoint differential equations are useful counterparts, for time-varying circuits, of even rational functions in the theory of fixed circuits. (See also Ref. 2.)

Recall Foster's canonical fixed nondissipative one-ports, comprising series- or parallel-connected subcircuits of one or two components each. The grouping (3) suggests that there *may* be a time-varying counterpart, in which the most complicated subcircuits correspond to quadruplets of $\pm$ complex characteristic exponents. However, although the existence of the grouping (3) is necessary for such a configuration, it does not in itself prove, or even support a strong conjecture, that the configuration is, in fact, a canonical periodically varying nondissipative one-port.

REFERENCES

1. Darlington, S., Linear Time-Varying Circuits — Matrix Manipulations, Power Relations, and Some Bounds on Stability, B.S.T.J., this issue, p. 2575.
2. Darlington, S., Nonstationary Smoothing and Prediction Using Network Theory Concepts, IRE Trans on Circuit Theory, **CT-6**, Special Supplement, May, 1959.

A 5-Gigacycle Tunnel Diode Oscillator with 9-Milliwatt Output from a Single Diode

By M. V. SCHNEIDER

This brief paper describes a tunnel diode oscillator which has been built by using a single diode in a low-impedance strip transmission line.

Tunnel diodes were made from p-type zinc-doped gallium arsenide with a doping level of 7×10^{19} carriers per cubic centimeter. Peak currents in the range of 150–300 ma are obtained by alloying tin pellets with a diameter of approximately 1 mil to the gallium arsenide wafer.

A new type of mount has been used in order to reduce the package reactance as well as the contact resistance of the diode in the transmission


Fig. 1 — Tunnel diode mounted in balanced strip transmission line.

TABLE I—DIMENSIONS OF BALANCED STRIP TRANSMISSION LINE

Ground plane spacing	b = 6 mils
Width of center conductor	w = 100 mils
Thickness of center conductor	t = 1 mil
Distance diode to RF short	Δ = 100 mils
Impedance of strip transmission line	Z = 2.8 ohms

line. The mount is shown in Fig. 1. A hole with an 8-mil diameter is made in the 1-mil center conductor of the strip transmission line. The gallium arsenide wafer is alloyed to a 3-mil gold wire containing 3 per cent zinc. The diode is inserted into the hole, and the pellet is soldered


Fig. 2 — Complete tunnel diode oscillator.

to the tinned edge of the center conductor. A small drop of epoxy is used to lock the diode into the hole. The gold wire connects the upper and lower parts of the ground plane after the line is assembled.

It is advisable to have a radio-frequency short close to the diode in order to prevent the oscillator from multimoding. The distance between the diode and the RF short determines the frequency of oscillation. The short shown in Fig. 1 consists of a symmetrical step in the ground plane. The center conductor is insulated from the ground planes by an evaporated oxide film or a thin sheet of mica. This section serves at the same time as a bypass capacitor for the dc supply. The bias resistance consists of a thin nichrome film evaporated on Mylar.

The dimensions of a particular assembly are given in Table I.

A special line transformer is used for matching the low-impedance

TABLE II — ELECTRICAL PARAMETERS OF GALLIUM ARSENIDE
TUNNEL DIODE

Peak current	I_p = 210 ma
Valley current	I_v = 12 ma
Peak voltage	E_p = 225 mv
Valley voltage	E_v = 560 mv
Junction capacitance	C = 12 pf

structure into a 50-ohm coaxial line. The impedance transformation is obtained by increasing the ground plane spacing b from 6 mils to approximately 300 mils. This type of transformation leads to a structure with relatively low losses. A broadband launcher is used for the transistion from balanced strip transmission line to coaxial line.

The complete oscillator is shown in Fig. 2. An output power of 9.1 mw at 4.85 gc was obtained from a single diode. The electrical parameters of the diode are listed in Table II.

The junction capacitance has been measured in the valley by using the techniques developed by D. E. Thomas.[1] Several attempts were made in order to determine the series resistance of high-current diodes in a transmission line with the dimensions given in Table I. It was found that the DeLoach method[2] gives very satisfactory results for relatively large ground plane spacings (10 mils and above); however, measurements at smaller ground plane spacings are at the present time not possible because of increasing line losses and because of launching problems.

Several oscillators were built with diodes having peak currents in the range from 150 ma to 250 ma. The output power ranged from 6 mw to 10 mw and the frequency of oscillation from 3.5 gc to 5.5 gc. No degradation effects have been observed. One particular oscillator has been running continuously for two months with a power stability of better than ±2 per cent, measured with a temperature stabilized power meter.

The author expresses his thanks to R. Neeld, who has developed the necessary techniques for assembling and mounting devices with an over-all size of several mils.

REFERENCES

1. Thomas, D. E. Tunnel-Diode Junction Capacitance Measurement, IEEE Trans. Elec. Devices, **ED-10**, July, 1963, pp. 278–280.
2. DeLoach, B. C., A New Microwave Measurement Technique to Characterize Diodes, to be published in IEEE Trans. Microwave Theory and Tech., **MTT-12**, January, 1964.

A Relation Between the Basis Functions of Periodically Varying Nondissipative Circuits

By SIDNEY DARLINGTON

This note relates to circuits of linear, periodically time-varying, positive capacitors and inductors. More specifically, it notes a property of the basis functions, or natural modes which describe the transient behavior of the circuits. Some other properties of the basis functions are described in Ref. 1 in this issue.

In accordance with Ref. 1

$$I = (S + pCp)\Phi, \qquad E = p\Phi \tag{1}$$

where I and E are column matrices, or vectors representing the excitation currents into and the response voltages at the various nodes, and Φ is an auxiliary vector variable. The matrices C and S are the node matrices of the capacitances and of the reciprocals of the inductances. They are symmetric and at least positive semidefinite. We will assume that C is positive definite. If it is not so originally, a similar equation with a positive definite C can be derived from (1), for example by the transformations described in the Appendix of Ref. 1. The symbol p represents differentiation (not frequency) and operates on all quantities following it.

Setting $I = 0$ in (1) makes it a homogeneous, second-order, vector differential equation in, say n dimensions, with periodically varying coefficients. Thus the well-known Floquet-Poincaré theorem requires the solution to be as follows

$$\Phi = \sum_{e=1}^{2n} k_e \varphi_e, \qquad \varphi_e = H_e(t)e^{s_e t}. \tag{2}$$

Here the k_e's are arbitrary scalar constants and the φ_e's are the basis functions, or natural modes. The characteristic exponents s_e are scalar constants. The coefficients H_e are time-varying vectors. When the s_e's are all different, the H_e's vary periodically. Otherwise, they are at most polynomials in t with periodically varying vector coefficients.

The real parts of the s_e's indicate the damping of the basis functions. When the inductors and capacitors are fixed, all basis functions are undamped. However, when the components vary periodically some of the basis functions may have nonzero damping (as is well known). Some bounds on the damping are derived in Ref. 1. It is also known that

the sum of all the characteristic exponents must be zero (when the circuit is nondissipative).

The purpose of this note is to point out the following: *Corresponding to any nondissipative circuit of periodically varying positive inductors and capacitors, any characteristic exponents which have nonzero real parts occur in equal positive and negative pairs.* Since complex exponents must occur in conjugate pairs, it follows that all the s_e's must fit into pairs or quadruplets of the following sorts:

$$+i\omega_b \qquad +\alpha_j \qquad +\alpha_h + i\omega_h \qquad (3)$$
$$-i\omega_b \qquad -\alpha_j \qquad +\alpha_h - i\omega_h$$
$$\qquad\qquad\qquad\qquad -\alpha_h - i\omega_h$$
$$\qquad\qquad\qquad\qquad -\alpha_h + i\omega_h.$$

The theorem follows at once from certain general properties of adjointly related differential equations, which are stated below without derivation. The solution of the nonhomogeneous equation (1) may be expressed in terms of functions of the excitation time, τ, and the response time, t. The adjoint equation is the equation whose solution corresponds to interchanging the two times, τ and t. It can be shown that the characteristic exponents of the adjoint equation are the negatives of those of the original equation. It can also be shown that equation (1) is self-adjoint. (The equation is its own adjoint.) Thus the negatives of the characteristic exponents of the original equation are also the characteristic exponents of the original equation.

When the components are fixed, vectors I and E are related by a set of odd rational functions of a frequency variable. Then Φ, whose derivative is E, is related to I by even rational functions. It appears that, more generally, self-adjoint differential equations are useful counterparts, for time-varying circuits, of even rational functions in the theory of fixed circuits. (See also Ref. 2.)

Recall Foster's canonical fixed nondissipative one-ports, comprising series- or parallel-connected subcircuits of one or two components each. The grouping (3) suggests that there *may* be a time-varying counterpart, in which the most complicated subcircuits correspond to quadruplets of $\pm$ complex characteristic exponents. However, although the existence of the grouping (3) is necessary for such a configuration, it does not in itself prove, or even support a strong conjecture, that the configuration is, in fact, a canonical periodically varying nondissipative one-port.

REFERENCES

1. Darlington, S., Linear Time-Varying Circuits — Matrix Manipulations, Power Relations, and Some Bounds on Stability, B.S.T.J., this issue, p. 2575.
2. Darlington, S., Nonstationary Smoothing and Prediction Using Network Theory Concepts, IRE Trans on Circuit Theory, **CT-6,** Special Supplement, May, 1959.

A 5-Gigacycle Tunnel Diode Oscillator with 9-Milliwatt Output from a Single Diode

By M. V. SCHNEIDER

This brief paper describes a tunnel diode oscillator which has been built by using a single diode in a low-impedance strip transmission line.

Tunnel diodes were made from p-type zinc-doped gallium arsenide with a doping level of 7×10^{13} carriers per cubic centimeter. Peak currents in the range of 150–300 ma are obtained by alloying tin pellets with a diameter of approximately 1 mil to the gallium arsenide wafer.

A new type of mount has been used in order to reduce the package reactance as well as the contact resistance of the diode in the transmission


Fig. 1 — Tunnel diode mounted in balanced strip transmission line.

TABLE I—DIMENSIONS OF BALANCED STRIP TRANSMISSION LINE

Ground plane spacing	$b = 6$ mils
Width of center conductor	$w = 100$ mils
Thickness of center conductor	$t = 1$ mil
Distance diode to RF short	$\Delta = 100$ mils
Impedance of strip transmission line	$Z = 2.8$ ohms

line. The mount is shown in Fig. 1. A hole with an 8-mil diameter is made in the 1-mil center conductor of the strip transmission line. The gallium arsenide wafer is alloyed to a 3-mil gold wire containing 3 per cent zinc. The diode is inserted into the hole, and the pellet is soldered


Fig. 2 — Complete tunnel diode oscillator.

to the tinned edge of the center conductor. A small drop of epoxy is used to lock the diode into the hole. The gold wire connects the upper and lower parts of the ground plane after the line is assembled.

It is advisable to have a radio-frequency short close to the diode in order to prevent the oscillator from multimoding. The distance between the diode and the RF short determines the frequency of oscillation. The short shown in Fig. 1 consists of a symmetrical step in the ground plane. The center conductor is insulated from the ground planes by an evaporated oxide film or a thin sheet of mica. This section serves at the same time as a bypass capacitor for the dc supply. The bias resistance consists of a thin nichrome film evaporated on Mylar.

The dimensions of a particular assembly are given in Table I.

A special line transformer is used for matching the low-impedance

TABLE II — ELECTRICAL PARAMETERS OF GALLIUM ARSENIDE
TUNNEL DIODE

Peak current	I_p = 210 ma
Valley current	I_v = 12 ma
Peak voltage	E_p = 225 mv
Valley voltage	E_v = 560 mv
Junction capacitance	C = 12 pf

structure into a 50-ohm coaxial line. The impedance transformation is obtained by increasing the ground plane spacing b from 6 mils to approximately 300 mils. This type of transformation leads to a structure with relatively low losses. A broadband launcher is used for the transition from balanced strip transmission line to coaxial line.

The complete oscillator is shown in Fig. 2. An output power of 9.1 mw at 4.85 gc was obtained from a single diode. The electrical parameters of the diode are listed in Table II.

The junction capacitance has been measured in the valley by using the techniques developed by D. E. Thomas.[1] Several attempts were made in order to determine the series resistance of high-current diodes in a transmission line with the dimensions given in Table I. It was found that the DeLoach method[2] gives very satisfactory results for relatively large ground plane spacings (10 mils and above); however, measurements at smaller ground plane spacings are at the present time not possible because of increasing line losses and because of launching problems.

Several oscillators were built with diodes having peak currents in the range from 150 ma to 250 ma. The output power ranged from 6 mw to 10 mw and the frequency of oscillation from 3.5 gc to 5.5 gc. No degradation effects have been observed. One particular oscillator has been running continuously for two months with a power stability of better than ±2 per cent, measured with a temperature stabilized power meter.

The author expresses his thanks to R. Neeld, who has developed the necessary techniques for assembling and mounting devices with an over-all size of several mils.

REFERENCES

1. Thomas, D. E. Tunnel-Diode Junction Capacitance Measurement, IEEE Trans. Elec. Devices, ED-10, July, 1963, pp. 278–280.
2. DeLoach, B. C., A New Microwave Measurement Technique to Characterize Diodes, to be published in IEEE Trans. Microwave Theory and Tech., MTT-12, January, 1964.