# THE BELL SYSTEM TECHNICAL JOURNAL

# Step Stress Aging of Plated Wire Memories

### By I. DANYLCHUK, U. F. GIANOLA, and J. T. SIBILIA

(Manuscript received February 5, 1968)

*A rate of 0.3 failures per billion hours or less is desirable for memory components in large integrated arrays. This unusually stringent requirement complicates the determination of lifetime from accelerated aging studies. The value and limitations of step stress aging techniques are discussed in terms of experimental results obtained using plated wire memory arrays designed to withstand the high ambient temperatures required for accelerated aging. Step stress aging measurements alone are insufficient for confident lifetime prediction. Therefore, longer term measurements at lower temperatures must also be made to establish the validity of the lifetime extrapolations. It is essential to protect the plated wires against corrosion. Given proper protection a shelf life in the hundreds of years is forecast. The importance of duty cycle on lifetime in exercising the memory is discussed and the results of aging under extreme pulsed magnetic field stress conditions are reported. Criteria for wire selection, with long term stability in mind, are discussed.*

"And in short measures life may perfect be"—Ben Jonson

I. INTRODUCTION

In spite of their advanced state of development, there is still uncertainty concerning the long term stability of magnetic film memories. As early as 1958 E. N. Mitchell showed that the anisotropy of

permalloy thin films could be modified by a magnetic anneal at moderate temperatures.[1] Chang, Gianola, and Sagal subsequently pointed out that such a phenomenon could be detrimental to the lifetime of a magnetic film memory element, and that in the case of plated wire intolerable changes in magnetic anisotropy and coercivity occur in freshly plated films.[2] These changes result in an increased minimum digit current for reliable writing along with a serious decrease in the digit disturb level. The result is a monotonic reduction in operating margins that, in time, can lead to a complete loss of range.

It was, however, also shown that the magnitude of the rate of change of magnetic properties could be substantially reduced by following electrodeposition with a stabilization anneal in an easy direction field.[2, 3] From a practical standpoint, a stabilization anneal has been found essential and is now in general use in plated wire fabrication. It is worth noticing that a post-deposition anneal is automatically provided in vacuum deposited magnetic films, since the latter are deposited on a hot substrate which is allowed to cool in the vacuum system.

From a simple physical viewpoint the effectiveness of the stabilization anneal should increase with annealing temperature and annealing time. However, Chang, Von Neida, and Calbick have shown that detrimental changes can also occur in electro-deposited films even when annealed in an easy direction field, thereby setting effective limits on the maximum annealing temperature for a given annealing time.[3] These limits were shown to correlate with discontinuous grain growth in the film. Grain growth, however, may have been complementary rather than causative to the increased dispersion in magnetic anisotropy noted. Copper diffusion from the substrate wire into the permalloy film has been shown not to be a primary contributor to the aging mechanism.[4]

These early studies showed that the post-deposition anneal extended the mean lifetime of the plated wire memory element to years or decades. Subsequent studies, however, have shown that a clear prediction of lifetime is complicated by a distribution of aging rates between individual bit elements in a large capacity memory, reflecting both nonuniformity in initial element properties as well as the distribution expected of a stochastic process. Since in a highly integrated large memory system the time to occurrence of the first few failures is of greater interest than the mean time before failure that is commonly used to define the lifetime of discrete components, the

distribution of failures is of prime importance. The aging phenomenon in magnetic films is also field dependent, being accelerated by hard direction fields (word fields) and retarded by easy direction fields (digit fields).[2] Therefore, since hard and easy direction fields are applied intermittently in an operating memory system, an added complication to useful lifetime prediction is the need to establish the average field environment that a memory element experiences over years or decades of use.

Earlier work has primarily been concerned with changes in dispersion and skew produced by a magnetic anneal, thereby delineating the magnitude of the aging phenomenon, but falling short in that a clear relation between skew and dispersion and the functional memory parameters of interest was not simultaneously established. Rabinovici and Renton subsequently examined the effect of aging on the functional parameters directly,[5] but, while a step in the right direction, the sample population used was insufficiently large to permit lifetime predictions at low failure levels.

To estimate the problem, consider some of the current thinking about high speed memories for which the plated wire is well suited. To achieve an economical system, a compact construction with a minimum number of interconnections, is desirable. We estimate that a suitable module size is about $4 \times 10^5$ bits; for example, 4096 words $\times$ 100 bits. Such a store operating as part of the central processor in an electronic switching office would be unacceptable if it were necessary to rework a single spare word or digit line once a year. On this basis then, if each bit in the memory is considered an independent device (not a fully justified assumption since failures are often grouped), an accelerated aging technique capable of predicting with some accuracy one failure in greater than $3.5 \times 10^9$ device hours (0.3 FIT) is needed.*

Clearly, it is impractical to build, age, and test large stores for the extended periods needed to derive statistically significant aging data under normal operating conditions. To overcome this difficulty, we have examined the possibility of adapting step stress accelerated aging techniques to magnetic memories in order to estimate lifetime at normal operating temperature. Such techniques are commonly used, particularly for determining the reliability of semiconductor components,[6] but have not been applied previously to memory arrays. This paper attempts to define the value and limitations of the tech-

---

* 1 FIT = one failure in $10^9$ device hours.

nique based on exploratory step stress aging studies of typical plated wire memory arrays.

## II. STEP STRESS AGING

It is well established that the aging process in thin magnetic films is thermally activated, and a reasonable starting assumption is that the aging process can be described by a characteristic time constant $\tau$, which is governed by the Arrhenius rate equation. More specifically, it is assumed that the average behavior of a device parameter of interest $P(t)$ can be adequately represented by $P(t) = P_0 + F(t/\tau)$, where $P_0$ is the average initial value. For the present purposes $F(t/\tau)$ need not be defined, the only assumption necessary is that it be a continuous monotonic function, albeit a complicated one. In addition, although a number of different rate mechanisms enter into the aging process we postulate a single time constant only. That condition, in fact, must be satisfied if the lifetime extrapolation is to be valid as discussed later.

The Arrhenius rate equation relates the aging time constant to the aging temperature, as follows:

$$\tau = \tau_0 \exp q/kT.$$

Where $\tau_0$ is a characteristic time constant, $q$ = activation energy, $k$ = Boltzmann constant, and $T$ = absolute temperature. This equation describes many of the mechanisms responsible for device degradation; for example, interatomic diffusion, chemical reactions, and crystallite growth.

The procedure followed in step stress aging is first to define a pass-fail criterion, defined by limits on one selected device parameter. In testing memory arrays a convenient device parameter is the output signal, which may be required to exceed a set discrimination level for a given set of operational write, disturb, and read current levels. All bits which give outputs exceeding the discrimination threshold pass, all others fail.

Such a pass-failure criterion is equivalent to setting $P(t_f)$ equal to a predetermined constant $P_f$ at the time of failure $t_f$. Consequently, the value of $F(t_f/\tau) = P_f - P_0$ and therefore $t_f/\tau$ are in turn predetermined constants and the Arrhenius rate equation can be used to relate the time of failure to the temperature at which aging is being carried out as follows:

$$ln\ t_f = \frac{q}{kT} + ln\ \tau_0 + ln\ F^{-1}(P_f - P_0).$$

In other words, $ln\ t_f$ is inversely proportional to $T$.

The first step in a step stress aging measurement is to stress (that is, anneal) the sample population at an elevated temperature $T_1$ for a prescribed time $t_1$, and then retesting at ambient temperature to determine the number of failures produced, if any. The sample is then once more stressed for the same period $(t_1)$ as before but at a higher temperature $(T_1 + \Delta)$. The temperature increment $\Delta$ is such that any failures produced are large compared to those accumulated at the lower stress temperature.

Experiment has shown the aging process to have an activation energy of about one electron volt. Therefore, a 20°C increment is appropriate. After aging, the sample is retested to determine the cumulative number of failures. This procedure is continued at successively higher temperatures using approximately the same temperature increment each time until the entire population has failed according to the original test specifications. This series of measurements provides a distribution of failures as a function of temperature for a given exposure time. Figure 1 shows a hypothetical distribution plotted as a function of inverse stress temperature. For the particular exposure
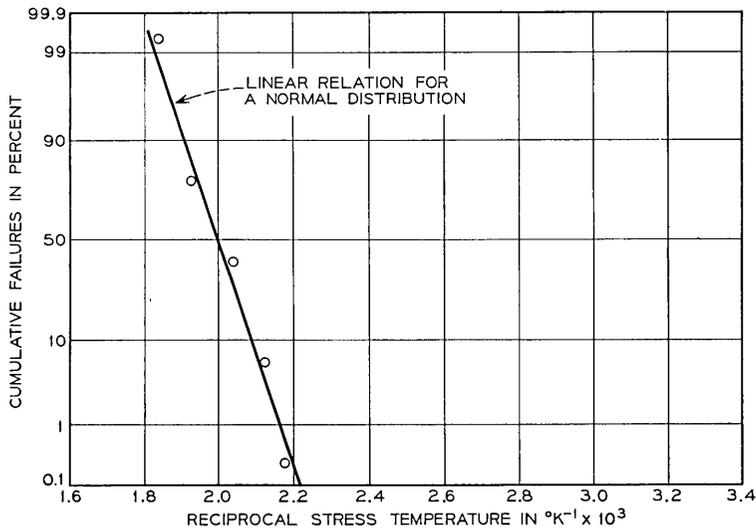


Fig. 1 — Hypothetical step stress aging data.

time chosen, no failures are observed to temperatures up to 200°C, and 1, 10, 70, and 100 percent cumulative* failures are obtained after stressing at 200, 220, 240, and 260°C, respectively.

The distribution of failures in temperature will depend upon both the initial distribution of $P_0$'s and the statistical variations expected of a thermally activated process. If these were random, a gaussian distribution of failures after aging might be expected as suggested in Fig. 1. However, such a distribution is virtually precluded by an initial selection criterion that truncates the distribution of $P_0$'s. If that were not the case, the initial sample population would contain that percentage of bad bits expected of an initially random distribution. This complicates the extrapolation of aging measurements based upon a small sample population. To reduce the uncertainty, the predictions of a step stress aging experiment must ultimately be confirmed on larger sample populations.

The distribution of failures in temperature may be converted to a distribution of failures in time. To do this, a second set of step stress measurements are obtained using a second sample population as identical to the first as possible, but in this case stressed at an exposure time $t_2$ that is approximately an order of magnitude larger than $t_1$. The failure distribution in this case will be centered at a lower temperature as illustrated in Fig. 2, which plots the logarithm of time to a given percentile failure versus the inverse stress temperature. A straight line extrapolation through points of equal failure in Fig. 2 provides an estimate of the time required to reach that level of failure at normal operating temperatures, for example, $\leq 50$°C. The slope of the linear relationship is determined by the activation energy $(q)$.

Figure 3 illustrates the difficulties encountered if aging is not the result of a simple thermally activated process. If it is assumed, for example, that two distinct and independent processes exist, then, depending upon the relationship between the $\tau_0$'s and $q$'s, a lifetime extrapolation from an accelerated aging measurement may or may not be valid. If the $\tau_0$ for the high $q$ process is larger than that of the low $q$ at all temperatures then the extrapolation will be valid. If, however, the high $q$ process has the shorter $\tau_0$, it may dominate at the temperatures used in step stress aging, thereby leading to an invalid life prediction because at the operating temperature the low $q$

---

* Averaging of the number of failures would provide an improved fit at the lower and upper ends of the distribution, but is not justified by the small statistics in the measurements reported here.
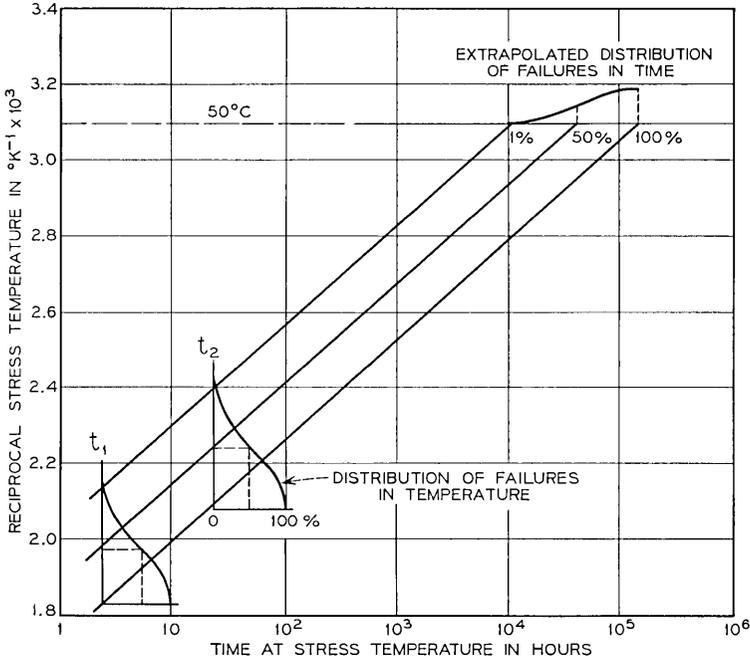
Fig. 2 — Translation of distribution of failures in temperature to distribution in times at lower temperatures.

process will be dominant and will produce earlier failures. It follows that there will always be some uncertainty regarding the validity of such extrapolations unless it can be positively established that a single aging mechanism is dominant up to and including the temperatures used in accelerated aging. In order to reduce this uncertainty as far as practicable, the results of the step stress aging measurements should be confirmed by longer term aging experiments at lower temperatures.

### III. AGING PLATED WIRE MEMORY ARRAYS

The essential structure of the conventional plated wire memory consists of plated wire pairs used as digit lines intersected by orthogonal word solenoids to form a regular memory array.[7] The sample populations used for the step stress aging experiments described in this paper each consisted of a 32 word × 31 bit array (992 bits). The test planes used had word solenoids on 50 mil centers with plated wires on 25 mil centers. A 2 mil thick permalloy overlay was used to
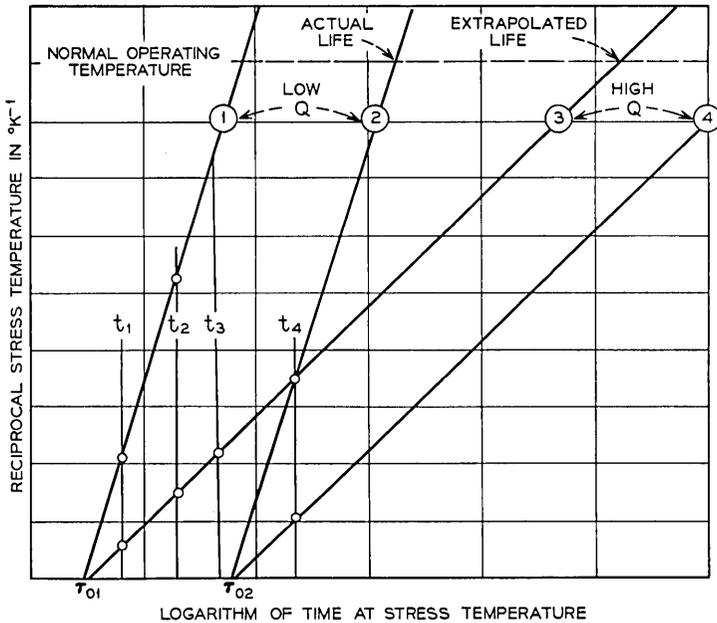
Fig. 3 — The effects of several aging processes on step stress aging extrapolations. If both low $Q$-high $\tau_o$ and high $Q$-low $\tau_o$ processes (2) and (3) are present, step stress measurements made with exposure times less than $t_4$, (for example, $t_1$, $t_2$ and $t_3$) will lead to longer extrapolated lifetimes than will actually be obtained. If $\tau_o$ for the low $Q$ process is lower than for the high $Q$ process—for example, (1) relative to (3) or (4)—the extrapolation will be valid or pessimistic.

provide shielding, enhance the word field, and limit word to word interaction.

In order to withstand the high temperatures used in the aging experiments the memory substrate was made of a slotted ceramic block and the word solenoids were Teflon insulated as illustrated in Fig. 4. Normal substrate materials and insulators are unsatisfactory at the higher aging temperatures. As shown, the plated wire is used in the shape of a hairpin. This construction has two important advantages; first, any uniform skew existing in the plated wire is nullified, second, since one end of the digit structure is free to move, the plated wires are not stressed by differential thermal expansion between the wire and memory plane. For most of the experiments the plated wire consisted of a 3500 angstrom thick, nominally nonmagnetostrictive, permalloy film on a 5 mil diameter conducting wire substrate preplated with a micron of copper. The wires had passed a functional on-line,
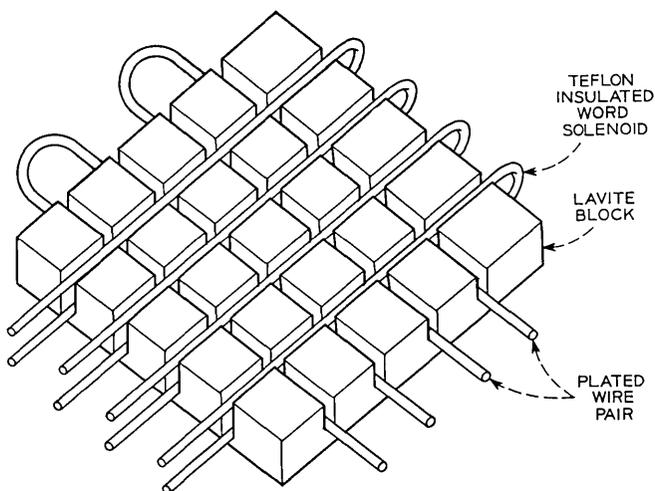
Fig. 4 — Structure of memory planes used for accelerated aging measurements at high temperatures.

destructive readout memory test and had received a stabilization anneal at 350°C for approximately one minute. During the course of these experiments wires from other sources were examined also; but, apart from differences reflecting different operating points and uniformity resulting from different processing details, no substantial difference in aging characteristics was found.

A functional test of normal destructive readout operation with adjacent word interaction using the following program was used to evaluate test planes. (*i*) Write zero in the test location using nominal word current and maximum digit current. Repeat 250 times. (*ii*) Write one in the test location using nominal word current and minimum digit current. (*iii*) Write zero in one adjacent address using nominal word current and maximum digit current. Repeat $10^5$ times. (*iv*) Write zero in the other adjacent address as in step *iii*. Repeat $10^5$ times. (*v*) Read the test bit and determine whether it has passed or failed. This program provides a worst case memory history that biases the test location towards minimum outputs.

A ±20 percent operating range on digit current was used to ensure a worst case test. It is assumed that a ±10 percent range will be more representative in actual system operation. The center values of currents used varied with the source of available wire. However, the following was typical for the wire used in most of the measurements: word cur-

rent equal to 800 ma with 40 ns rise time and 200 ns duration; digit current equal to 25 ma with 10 ns rise time and 200 ns duration. The word and digit pulses were overlapped by approximately 100 ns. The threshold level for pass was set at 2.5 mV corresponding to approximately one half the nominal output.

Test results were recorded using an xy plotter to show the failure locations in the array. Figure 5 is a typical map for a plane that has been aged to a 22 percent failure level. It clearly illustrates the tendency for failures to cluster in particular locations along the digit line, and demonstrates a marked variability in number of failures from wire to wire. By mapping failures for both digit senses (that is, 0's and 1's) the cause of failure can often be diagnosed. A negative correlation between failures for the two senses at corresponding locations indicates skew induced failures, while a positive correlation indicates either failure to write adequately or a low disturb threshold. Further diagnoses can be obtained by noting correlations produced under modified test programs.

Aging measurements were performed using several different ambiences: (*i*) air atmosphere, zero applied field; (*ii*) hydrogen reducing atmosphere, zero applied field; and (*iii*) hydrogen reducing atmosphere, pulsed hard direction field.
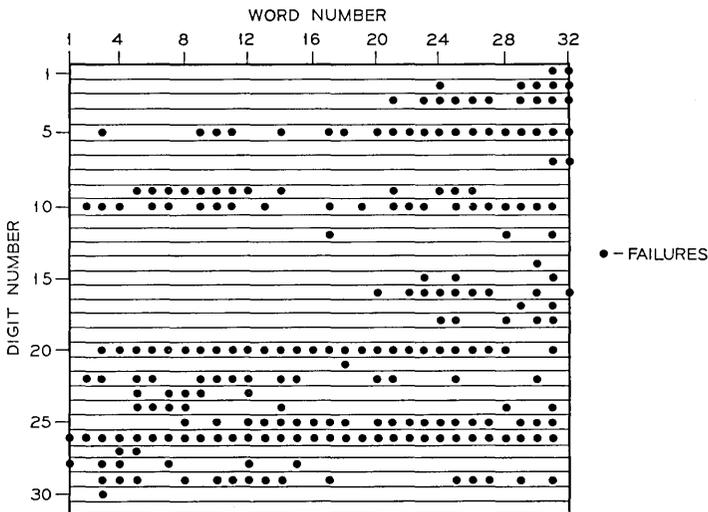


Fig. 5 — Typical distribution of failures in an aged memory plane.

IV. AGING UNPROTECTED WIRES IN AIR

Figures 6 and 7 show the result of step stress aging of unprotected plated wires in air for exposure time of 2, 20, and 200 hours at temperatures up to 320°C. In this experiment the onset of failures was so abrupt that the earliest failures were not observed in using a 20° stress temperature increment. Instead, Fig. 6 shows the last temperature at which no failures were obtained as 0.1 percent failure points for each of the exposure times. The 0.1 percentile represents the first measurable point in the test population of 1,000 bits. Notice that the percentage cumulative failures are plotted on a logarithmic scale. As pointed out previously there is no reason to expect a normal distribution of failures in temperatures, nor is one obtained.

Equal percentage failure points interpolated from the data of Fig. 6 are plotted in Fig. 7 on a $1/T$ vs $\log_{10} t$ graph. A good approximation to the linear relationship called for by the simple thermal activation model is obtained, and leads to extrapolated lifetimes for 0.1 percent failure at 50 and 25°C of 2 and 20 years, respectively. This extrapolation is additionally supported by a constant temperature aging experiment run at 80°C for several thousand hours using another similar 1,000 bit plane. The results of that experiment are also plotted in Fig. 7 and show excellent agreement with the step stress

Fig. 6 — Distribution of failures in temperature for unprotected wires aged in air.
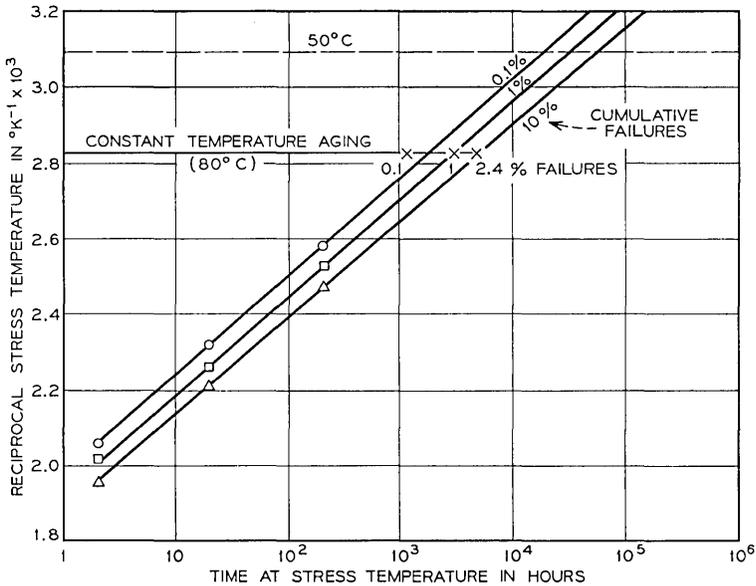
Fig. 7 — Loci of constant percentile cumulative failure in time and temperature for unprotected wires aged in air.

prediction for the 1 percent failure point. The agreement with the 0.1 percent failure point is not unreasonable in view of the large experimental inaccuracy in determining the first failure in a population of 1,000 bits. These results lead to extrapolated lifetimes at normal operating temperatures that are marginal for large memory systems.

Physical examination of failed wires in these studies showed that corrosion of the substrate wire with a resulting eruption of the magnetic film at pinholes appeared to be a primary cause of failure. Indeed, the activation energy of 0.84eV derived from the slope of the $1/T$ vs log $t$ relation is consistent with the value for the oxidation of copper (0.87 eV).[8] These observations, together with the results of the following experiment, lead to the conclusion that chemical passivation or protective encapsulation of the plated wire is essential if adequate lifetimes are to be obtained.

## V. AGING PROTECTED WIRES

A dramatic reduction in failure rates is obtained if the wires are protected either by chemical passivation or through the use of an

inert or reducing atmosphere. Figures 8 and 9 show the result of step stress aging protected wires. The distribution of cumulative failures versus inverse absolute temperature is seen in Fig. 8 to be better behaved than when corrosion occurs. The 0.1 and 0.2 percent data points in this experiment were obtained experimentally, so that extrapolation to 0.01 percent failure is reasonably justified. The extrapolation of equal percentile failure points on a $1/T$ vs $\log_{10} t$ relationship in this case yields lifetimes to early failures in the many thousands of years at normal operating temperatures.

The validity of the step stress extrapolations is well confirmed by a constant temperature aging experiment carried out at 140°C that has been in progress for 5600 hours with no failures observed in a population of 1,000 bits as indicated in Fig. 9. The slope of the $1/T$ vs $\log_{10} t$ relationship in this case yields an activation energy of 1.3 electron volts in reasonable agreement with the value 1.25 electron volts that we estimate from the grain growth data of Chang, Von Neida, and Calbick.[3] Therefore, it appears that crystallite growth, or a common causative phenomenon, is responsible for the aging observed in protected wires in the absence of applied magnetic fields. Complementary measurements indicated that the prime functional cause of failure was a reduction in the digit disturb threshold as



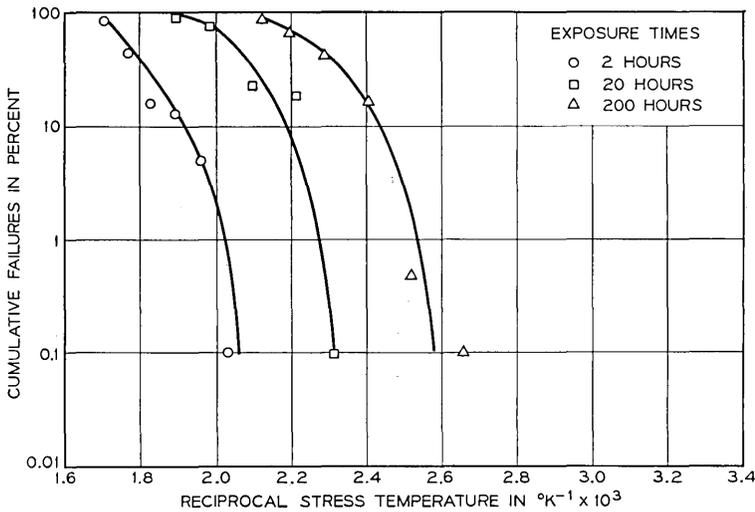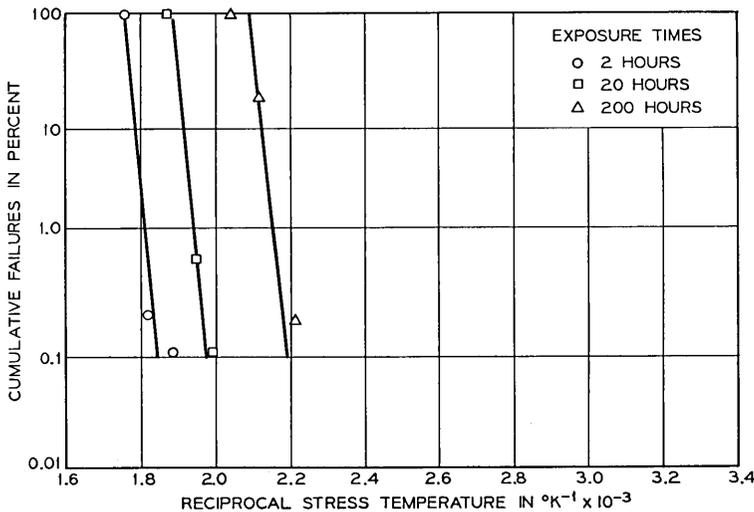Fig. 8 — Failure distribution in temperature for protected wires.
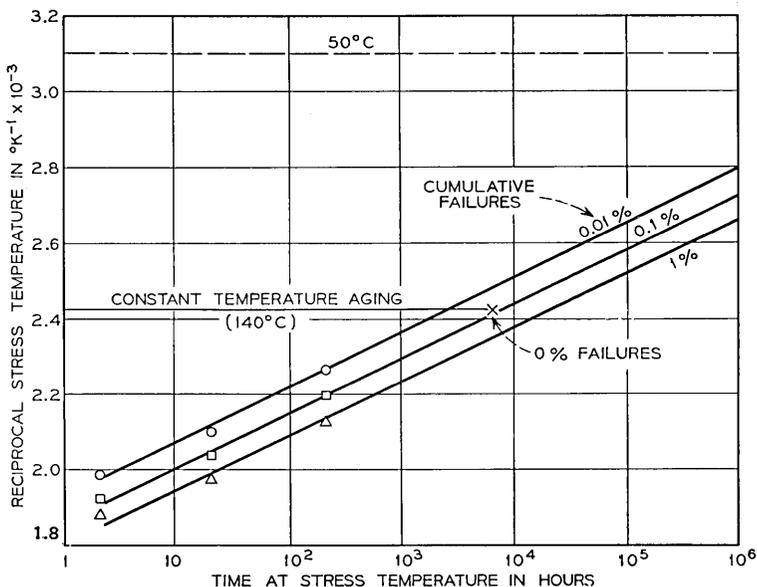
Fig. 9 — Loci of constant percentile cumulative failure in time and temperature for protected wires.

would be expected from the monotonic reduction in coercivity that occurs in permalloy as a result of a strain relief anneal.

The results of this experiment show that properly stabilized wires protected from corrosion should have an adequate shelf life.

## VI. EFFECTS OF MAGNETIC FIELDS ON AGING

Hard direction fields produce a rotation of the uniaxial anisotropy in a magnetic film memory element. This result is a natural consequence of the well known magnetic annealing properties of the permalloys.[9] The sensitivity to hard direction fields is substantially reduced by the stabilization anneal discussed previously. In a memory environment the individual magnetic film memory elements are subject to aperiodic pulsed magnetic fields so that measurements of the effect of dc hard direction fields are not simply applicable.

To further complicate matters, the magnetic anisotropy is not uniformly affected by an applied field, but in first approximation it appears to have a relatively stable component in addition to an easily rotatable component. The magnitude of the latter is considerably reduced by the stabilization anneal. This situation can be described

in terms of a simple model that assumes that the rotatable component of anisotropy experiences a torque proportional to a function of the angle between the anisotropy and the direction of magnetization, and that the rotatable component relaxes towards the direction of magnetization under a characteristic time constant. Under these circumstances, the rotatable anisotropy component relaxes back towards the easy axis direction established by the stable anisotropy component in the interval between word field pulses.

An analysis based upon this simple model is given in the Appendix. It shows that the worst case effect of hard direction field pulses applied at *low* duty cycles is to induce a skew that is approximately proportional to both the duty cycle and the ratio of the rotatable to stable anisotropy components. Such a growth in skew with a dependence on duty cycle has been verified experimentally, although there is as yet insufficient data to provide the exact form of the dependence over a wide range of duty cycles.

One problem, then, in designing a meaningful field aging experiment is to decide upon a representative duty cycle. In a 4,000 word memory with a ratio of cycle time to word pulse duration of 5, no one memory element would be subject to hard direction fields for more than 0.005 percent of the time if the memory were exercised in a completely random fashion. On the other hand, it would not be out of the question for any one memory word to be interrogated once every 10 instructions over substantial periods. In that case some memory words will be subject to hard direction fields 2 percent of the time. On the other hand those same memory words would be unlikely to be exercised continuously over many years of operation. In the absence of any definitive data, it is suggested that a representative worst case duty cycle will be about 0.1 percent.

Figures 10 and 11 give the results of an exploratory step stress aging measurement under pulsed field stress. In this experiment *every* bit in the test population was subject to hard direction fields under a 1 percent duty cycle to determine the effect of extreme aging conditions. It will be clear from the previous arguments that this is an unrealistically severe test which, however, serves to illustrate the problem of making extrapolations from aging measurements under pulsed magnetic field stress.

The test program used was the same as described previously, except that a ±14 percent digit current range was used. Unsatisfactory extrapolated lifetimes were obtained for the ±20 percent margin used in the previous experiments. This was not surprising since a ±20 percent
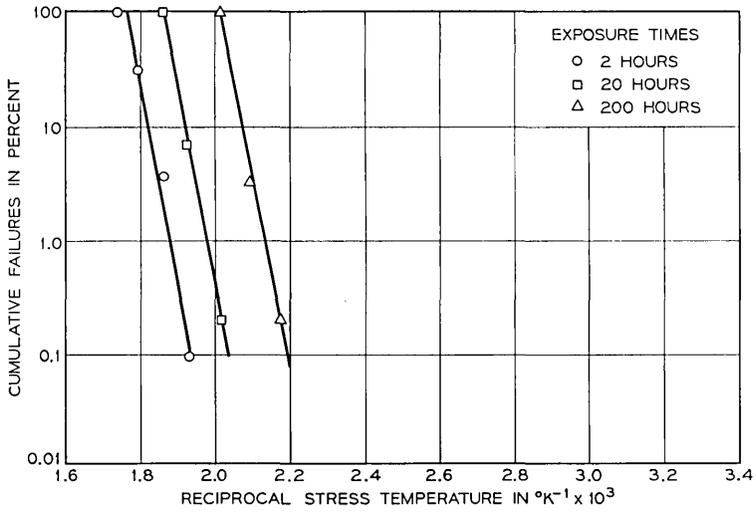
Fig. 10 — Distribution of failures in temperature for protected wires subjected to pulsed field stress during aging.
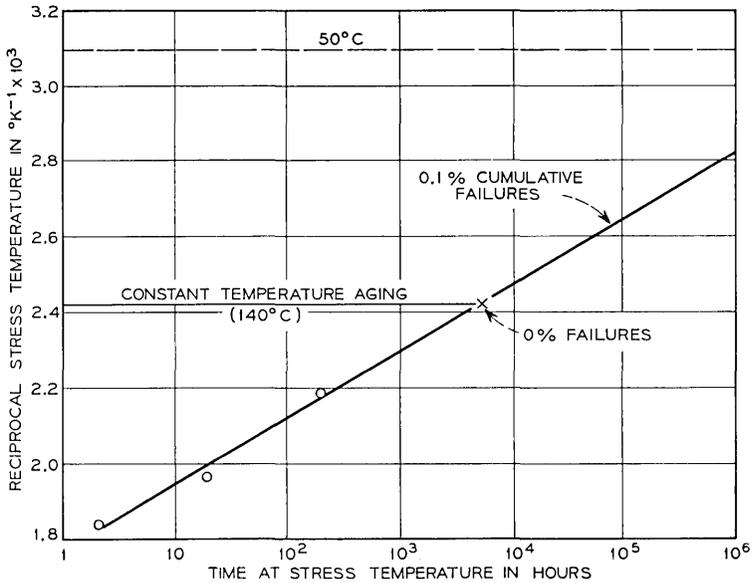


Fig. 11 — Loci of constant percentile cumulative failure in time and temperature for protected wires subjected to pulsed field stress during aging.

criterion was used for initial selection of wires. During step stress aging, all word solenoids and the test planes were connected in series and pulsed with 400 mA amplitude word current pulses at a 1 percent duty cycle. Thus, all bits in the test population were subjected to field stress during aging.

The 400 mA word current level was chosen for nondestructive readout during aging. This is a necessary condition since each bit must remain in an approximated single domain state during accelerated aging if anisotropy "recovery" during the interpulse interval is to take place. This procedure provides a not unreasonable simulation of the actual state of the memory sites in operation and avoids the complications that would be introduced, both in instrumentation and interpretation, if the memory were exercised in the destructive readout mode during aging.

In contrast to the situation for "no-field" aging, failures were observed to result primarily from increased skew as evidenced by an increase of the minimum digit write current and an asymmetry in bit failures for opposite polarities of digit write currents. In addition, the sense of skew correlated with the direction of the hard direction field pulses applied during aging.

The results of these experiments are given in Figs. 10 and 11. In Fig. 10 the 0.1 percent failure points represent the last stress temperature at which no failures were observed. An approximately logarithmic failure distribution is again found, in Fig. 10, although this relationship is not as well obeyed as the case for no-field aging, suggesting that the plated wires used were not homogeneous in field aging property. Once again, early failures tended to cluster on particular digit lines, but no obvious correlations to initial physical properties have as yet been established.

As already mentioned, no simple distribution could be obtained for a $\pm 20$ percent range of digit current indicating the severity of the additional aging induced under field stress. In addition, Fig. 11 shows that the extrapolated lifetime for 0.1 percent failures and a $\pm 14$ percent digit current range is substantially less than for "no-field" aging (Fig. 9), with a $\pm 20$ percent margin. Furthermore, the step stress data points do not as satisfactorily fit a linear $1/T$ vs log $t$ relationship, suggesting the possibility of even lower lifetimes as discussed in connection with Fig. 3.

On the other hand, the extrapolation shown in Fig. 11 is consistent with the results of a constant temperature aging experiment, which has been under way for 6500 hours at 140°C with no failures using the

same ±14 percent digit margin criterion. However, a different batch of wires was used for the constant temperature aging experiment. Thus there is a residual uncertainty regarding the validity of the lifetime extrapolation, and more comprehensive measurements will be needed before lifetime predictions can be made with reasonable confidence. Assuming that the data of Fig. 11 is representative it can be concluded that the plated wire memory has an adequate lifetime for many applications even under pulsed magnetic field stress, provided that proper attention is paid to duty cycle and choice of operating range.

VII. DISCUSSION

As stated at the beginning, this paper's prime purpose is to describe the procedures that have been developed for accelerated aging of memory arrays using functional test criteria, and to illustrate the pitfalls that must be taken into account. It is not intended to provide definitive answers to the plated wire lifetime question. The limited data presented is encouraging and suggests that, given proper selection and use, reasonable lifetimes can be ensured. We caution, though, that a much more comprehensive study is needed before confident predictions can be made. Measurements using larger sample populations will reduce the statistical uncertainty in extrapolating failure distributions, and longer term measurements at lower temperatures will reduce the uncertainty in extrapolation to long periods.

Because a memory is a large integrated entity, and because in present day usage only small numbers of failures can be tolerated over years or decades of operation for economical reasons, the problem of determining reliability is difficult. It has not arisen in the case of ferrite core memories since no short term degradation in properties has been reported. The time scale of degradation in the properties of anisotropic magnetic alloy films, however, is such that the possibility of a lifetime limitation needs to be considered seriously. The physical mechanisms responsible for aging deserve as much attention as the origins of the induced anisotropy. The step stress aging technique used in this study followed those in common use for determining the reliability of semiconductor devices. With the increasingly large number of functional cells being integrated into semiconductor circuits, similar attention will need to be given to the limitations of the technique. This will be especially true should large semiconductor memories be realized.

It has been established that the aging mechanisms in magnetic films

are thermally activated. Indeed, physical reasoning leads to that expectation. There is no evidence for a temperature threshold, which might only occur should a cooperative mechanism be responsible for aging. The activation energy is such that at normal temperatures, a 20°C temperature increment produces approximately one order of magnitude increment in predicted lifetime. Thus, if the long term reliability proves to be marginal for more extreme applications, improved heat sinking should be used. Magnetic film memory elements are unique in the small energy dissipated in the cell itself. Normally, the main cause of temperature rise above ambient is dissipation in peripheral circuits. It is also worth noticing that magnetic film elements are relatively unaffected by reduced temperatures. P. I. Bonyhard, in unpublished work, has shown that the plated wire can be operated to at least −70°C with no significant changes in operating margins.

Since we first drew attention to the potential severity of aging in plated wire memories, industry wide practice has followed the post-deposition stabilization anneal that we recommended. This practice has reduced aging from a first order to a less significant problem and, accordingly, has made reliable long term lifetime predictions more difficult. The measurement problem is also increased by the wide distribution of failure rates found. The step stress aging technique described in this paper reduces the measurement problem to experiments of reasonably short duration. Furthermore, the complementary longer term aging experiments at lower temperatures have provided reasonable confirmation of the extrapolations from the short term step stress measurements, although further corroborative studies are desirable.

The necessity for corrosion protection has been established by this study. In turn, it has been established that protected wires have adequate shelf life. Furthermore, this study also focusses attention on the importance of word pulse duty cycle in an operational memory. The desirability for adequate magnetic shielding against static environmental magnetic field follows implicitly.

The prime functional result of aging has been shown to be an erosion of both upper and lower limits of digit current. It is axiomatic, therefore, that for high reliability plated wire should be selected to have wider digit current margins than required operationally. Since an increase of the lower limit on digit current is likely more severe than the decrease of the upper limit, the center of the selection margin should be offset with respect to the nominal operational value. It also

follows that correlations of aging with manufacturing process should best be performed, at least initially, with attention to the rate of change of the lower limit on digit current as determined by short term step-stress aging.

APPENDIX

*Analysis Based on Magnetic Model*

Let

$K_1$ = "stable" component of anisotropy
$K_2$ = "rotatable" component of anisotropy.

Assume that the rotation of $K_2$ is reversible. Hard direction field pulses (applied normal to $K_1$) rotate $K_2$ through an angle $\beta$ with respect to $K_1$. Relaxation of $K_2$ towards $K_1$ takes place during the zero field interval between hard direction field pulses.

Of interest is the steady state situation under repetitive pulsing. Steady state is achieved when $\beta = \beta_1$ such that the incremental rotation $\Delta\beta$ produced by each successive field pulse of duration $t_1$ equals the relaxation $-\Delta\beta$ during each interpulse interval $t_2$.

We postulate that the rotation of $K_1$ results from a torque exerted by the magnetization $M$ upon $K_1$. We further postulate that this torque has the same $\sin 2\theta$ dependence as the torque exerted by $K$ upon $M$ in a uniaxial material, where $\theta$ is the angle between $K$ and $M$. These are physically reasonable assumptions, but are presented without experimental confirmation.

The dynamics of the rotation is assumed to be governed by some characteristic relaxation time constant $\tau$. Its exact form is unimportant for the present discussion, but we do assume $\tau \gg t_1$ or $t_2$ in which case $\Delta\beta \ll \beta_1$. In that case for the steady state solution it is sufficient to consider the conditions required for a time averaged balance of the torques applicable during intervals $t_1$ and $t_2$. This condition can be expressed:

$$t_1 \sin 2(\theta_h - \beta_1) = -t_2 \sin 2(\theta_0 - \beta_1) \qquad (1)$$

where $\theta_h$ is the angle between $M$ and $K_1$ during interval $t_1$ when the hard direction field pulse is applied, and $\theta_0$ the angle between $M$ and $K_1$ during the interpulse interval $t_2$. $\theta_0$ is related to $K_1$ and $K_2$ through minimization of the energy relation for zero applied field and negligible magnetostatic field. In this case:

$$0 = K_1 \sin 2\theta_0 + K_2 \sin 2(\theta_0 - \beta_1). \qquad (2)$$

Thus, substituting for $(\theta_0 - \beta_1)$ from equation (2) in equation (1)

$$\sin 2\theta_0 \approx [K_2 t_1 \sin 2(\theta_h - \beta_1)]/K_1 t_2.$$

For cases of interest, (low duty cycle and stabilized material), we may assume small $\theta_0$ and $\theta_h \gg \beta_1$. Under these circumstances the steady state skew produced by repetitive field pulsing is

$$\theta_0 \approx [K_2 t_1 \sin 2\theta_h]/2K_1 t_2.$$

## VIII. ACKNOWLEDGMENTS

REFERENCES

1. Mitchell, E. N., "Effects of Heat Treatment of Thin Ferromagnetic Films at Intermediate Temperatures," J. Appl. Phys., 29, No. 8, part 2 (March 1958), pp. 286–287.
2. Chang, J. T., Gianola, U. F., and Sagal, M. W., "Aging and Stabilization of Electrodeposited Cylindrical Magnetic Thin Films," J. Appl. Phys., 35, No. 3, part 2 (March 1964), pp. 830–831.
3. Chang, J. T., Von Neida, A. F., and Calbick, C. J., "Structural and Magnetic Characteristics of Annealed Electrodeposited Permalloy Films," J. Appl. Phys., 37, No. 3, part 2 (March 1966), pp. 1472–1473.
4. Von Neida, A. R., and Hagedorn, F. B., "Cu Diffusion in Electrodeposited Permalloy Films," J. Appl. Phys., 38, No. 3, part 2 (March 1967), pp. 1436–1438.
5. Rabinovici, B., and Renton, C. A., "Some Aging Properties of Cylindrical Electrodeposited Permalloy Memory Wires," Proc. Intermag. Conf., Stuttgart, Germany, (April 1966), pp. 2–5.
6. Dodson, G. A., and Howard, B. T., "High Stress Aging to Failure of Semiconductor Devices," Proc. 7th Nat. Symp. Reliability and Quality Control in Elec. (January 1961), pp. 262–272.
7. Danylchuk, I., Gianola, U. F., Perneski, A. J., and Sagal, M. W., "Plated Wire Magnetic Film Memories," Proc. Intermag. Conf., Washington, D. C., (April 1964), pp. 4–5.
8. Kubaschewski, O., and Hopkins, B. E., Oxidation in Metals and Alloys, London: Butterworth Scientific Publications, 1958, p. 163.
9. Chikazumi, S., Physics of Magnetism, New York: John Wiley and Sons, 1964, p. 359.

# The Transmission Performance of Bell System Intertoll Trunks

By I. NÅSELL, C. R. ELLISON, Jr., and R. HOLMSTROM

(Manuscript received February 28, 1968)

*A systemwide survey of the transmission performance of Bell System intertoll trunks was undertaken in 1964. The sample design used for the survey is described briefly. The main purpose of the paper is to present survey results. Thus, the physical composition and some physical attributes of the trunks are given. The transmission measurement procedures are summarized, and measurement results are presented in distributional form for 1000-Hz loss, frequency response, background noise, impulse noise, and relative envelope delay. Among the results noted are an increase in average noise level and a decrease in noise level standard deviation as the trunk length is increased. The frequency response of long trunks is superior to that of short trunks. Measurement results are presented separately for major transmission facilities.*

## I. INTRODUCTION

The Bell System toll network consists of a hierarchy of toll offices interconnected by transmission paths called intertoll trunks. A toll call between two subscribers is built up of a tandem connection of several transmission paths which are joined by switching.

At each end of such a connection is a loop that connects the subscriber's telephone set with a local telephone office. The local telephone offices connect with toll offices by toll connecting trunks. These toll offices are connected by either a single intertoll runk or by several intertoll trunks through intermediate toll offices. The transmission performance of a toll connection between subscribers is thus influenced by the performance of each trunk and loop in the connection.

Systemwide improvement in transmission performance by a category of trunks is directly measured by the corresponding improvement in built-up connections. Systems engineering on transmission objectives, therefore, requires information about the relation between

transmission performance on trunks and the corresponding transmission performance that results on built-up connections. Such information is meaningful only if it is based on accurate information about the transmission performance of the various entities of importance; that is, the intertoll trunks separated into mileage categories, the major transmission facilities used in the toll plant, the toll connecting trunks, the loops, and the built-up connections. This need for information constitutes the basic reason for undertaking the intertoll trunk survey discussed in this article.

The introduction of data transmission over the switched telephone network has brought with it the need for information about transmission parameters that are of only minor importance for the transmission of spoken messages. The discussions of impulse noise and envelope delay in Sections VII and VIII both deal with this category of parameters.

Each intertoll trunk consists of trunk facility and office equipment in tandem. The trunk facility supplies a transmission path between the two toll offices connected by the trunk. The office equipment contains signaling devices and attenuators, and sometimes echo suppressors and hybrid transformers. There are several different transmission facilities in the Bell System toll network. The most important are voice-frequency facilities, compandored short-haul cable carrier, coaxial cable carrier, and microwave radio.

Survey results are presented both for various trunk lengths and for selected transmission facilities. The results for trunks take a slightly different form than for facilities. The frequency response for facilities gives the difference between the loss at a certain frequency and at 1000 Hz, while the frequency response for trunks gives the actual switch-to-switch loss at each frequency. Thus the trunk results depend on facility mixture, loss design, and loss maintenance of trunks, while none of these factors influence the facility results. Background noise levels for facilities are referred to a standard zero transmission level point, while those for trunks are referred to the receive switch of each trunk. Facility mixture, loss design, and loss maintenance of trunks affect the trunk results but not the facility results.

There is not such a separation of results for impulse noise; both facility results and trunk results are referred to the receive switch. The reason for this is the expectation that the switching equipment contributes to the impulse noise level on an intertoll trunk. Relative envelope delay is given separately for facilities and trunks; varying

facility mixture and various types of office equipment influence the trunk results in different mileage categories.

The facility composition of intertoll trunks considerably influences their transmission performance. Facility composition is continuously changing; the most important change in recent years is the introduction of new short-haul carrier systems with improved transmission characteristics.

## II. SAMPLING CONSIDERATIONS

### 2.1 Definition of Population

A sampling plan tailored to the structure of the Bell System intertoll network was established and followed carefully. The objective was to design the survey so that the sample data could be used to make estimates of characteristics associated with the entire population. An important preliminary step is to give a precise definition of the population so that the extent and limitation of the survey results are known.

An intertoll trunk is defined as a trunk between two toll offices, that is, between two separate toll switching units that both have one of the long distance switching plan classifications: regional center (class 1), sectional center (class 2), primary center (class 3), or toll center or toll point (class 4). In case one toll office building (or complex of buildings) contains only one toll switching machine, it is counted as one toll office. This means that a manual switchboard in the same building as a switching machine is not counted as a separate toll office, and trunks between such a switchboard and the switching machine do not qualify as intertoll trunks. If a building contains two or more switching machines that are separately identified and that have access to different groups of trunks and where the trunks between the machines are designed as intertoll trunks, then these switching machines are considered as separate toll offices. So-called tandem offices that can connect with the long distance network are classified as toll offices. Trunks between tandem offices are regarded as intertoll trunks if they can carry traffic to or from the long distance network.

Every intertoll trunk allows transmission in two directions, but measurements in the survey and hence characterization of performance are made only in the receive direction. Therefore, it is clear that the trunks themselves do not constitute the population elements.

Rather, a population element is identified with each direction of transmission of an intertoll trunk. One requirement for inclusion in the population is that the trunk have both of its terminations within the USA (excluding Alaska and Hawaii) or Canada. A further requirement is dictated by the administrative necessity of confining the measurements to Bell System toll offices. This further requirement states that the receive termination of a particular direction of transmission of an intertoll trunk should be located in a Bell System toll office. This implies that those intertoll trunks that have both endpoints in Bell System toll offices give rise to two population elements, while those that have one termination in a Bell System toll office and the other in an independent office are counted just once.

## 2.2 Sampling Plan

The sampling plan used in the survey is a two-stage plan with substratification which is self-weighting within each substratum and where the first-stage sample is selected with probabilities proportional to measures of size. (See Hansen, Hurwitz and Madow[1] for a general discussion of this type of sampling plan.) The primary units were Bell System toll offices as defined in the preceding section. A substratification is a stratification of the population elements in each sampled primary unit. The substrata were defined in the same way for all primary units; they are identified with length-categories of intertoll trunks as shown in Table I.

The first step in the sampling plan was to establish a frame for the first-stage selection. The frame listed all Bell System toll offices and gave for each the number of intertoll trunks terminating in the office. The number of trunks per office was used to form probabilities for the first-stage selection of primary units. Such a probability was computed for each office as the quotient of the number of intertoll

TABLE I — DEFINITION OF SUBSTRATA

| Substratum number | Trunk length $l$ | |
|---|---|---|
| | Miles | km (approx.) |
| 1 | $0 < l \leq 62.5$ | $0 < l \leq 100$ |
| 2 | $62.5 < l \leq 125$ | $100 < l \leq 200$ |
| 3 | $125 < l \leq 250$ | $200 < l \leq 400$ |
| 4 | $250 < l \leq 500$ | $400 < l \leq 800$ |
| 5 | $500 < l \leq 1000$ | $800 < l \leq 1600$ |
| 6 | $1000 < l \leq 2000$ | $1600 < l \leq 3200$ |
| 7 | $2000 < l \leq 4000$ | $3200 < l \leq 6400$ |

trunks terminating in that office to the total number of trunk termina-
tions found in all Bell System toll offices.

After the sample size had been determined, as discussed in the
next section, this frame was used for selecting a first-stage sample of
48 primary units. Randomness was assured by using lists of random
numbers. The selection was made with replacement. As a result, three
toll offices were selected twice. The first-stage sample therefore con-
tains 45 different toll offices representing 48 primary units. The sam-
pling was done with replacement in order to correspond with the
specific assumptions made in the derivation of the estimation form-
ulas.[1] More efficient sampling without replacement is being considered
for future surveys.

The next step of the sampling plan was to acquire detailed in-
formation about the selected offices. This consisted of lists of all inter-
toll trunks terminating in the selected offices, giving for each the trunk
number, the distant termination, and the actual trunk length in miles.
These lists were used to establish frames for the second-stage sampling
in each substratum of each toll office.

The final step in the sampling plan was to select sample elements
from these frames. In this selection, all population elements in a given
substratum of a sampled toll office were given the same probabilities
of inclusion. The selection was made with tables of random numbers,
without replacement. It resulted in lists of specific trunks to be meas-
ured in the survey. To these lists were appended lists of "spare"
trunks which were resorted to only when a trunk in the original list
was not available for testing when the measurements were taking
place.

2.3 *Determination of Sample Size*

The size of a survey sample ideally should be determined to give
maximum precision for fixed cost, or to minimize the cost while
achieving a required precision. Many transmission parameters were
measured for each trunk in the intertoll trunk survey. An ideal sam-
ple size would, therefore, recognize precision requirements for each
of these parameters; but for most of them it was far from obvious
how the precision requirements should be stated. The sample size
therefore was determined to maximize the precision of estimates of
background noise levels (measured with the 3A noise level meter[2]),
combined with some basic cost constraints. This parameter was chosen
as the crucial one in determining sample size because meaningful
precision requirements could easily be stated.

To estimate the precision expected from a survey requires estimation of variance components. There had been no systemwide survey of Bell System intertoll trunk transmission performance before 1964, so direct variance estimates based on previous survey results were not available. Variance estimates were therefore derived, based partly on a small pilot survey and partly on an indirect approach that used data available for some selected transmission facilities.

Experience has indicated that many transmission parameters show a dependence on trunk length. It is therefore of interest to present survey results by mileage categories. This in turn carries with it the desirability of making precision estimates within the same mileage categories.

Table II lists the widths of the 90 percent confidence intervals for the mean 3A noise level that were expected in each of the seven mileage categories of Table I. These expectations are based on the above-mentioned variance estimates combined with a sample size of 151 trunks in each mileage category. This sample size was acceptable from a cost standpoint, and the expected confidence interval widths were in line with precision requirements. The increased precision with longer trunks results from the smaller noise variance for long trunks, discussed in Section VI. Since the longest trunks have the highest average noise level, it was deemed desirable to require greater precision for these trunks. The final sample contained a total of 1069 trunks.

The last column in Table II gives the precision that was actually achieved in the survey. A higher variability than expected was found in the first two substrata; hence the achieved precision was somewhat poorer than expected. Otherwise the precision achieved was uniformly better than expected. This is a reflection of using unnecessarily pessimistic variance estimates in substrata 3 to 7.

TABLE II — WIDTH OF 90 PER CENT CONFIDENCE INTERVALS
FOR MEAN OF DISTRIBUTION OF 3A NOISE LEVELS

| Substratum | 90 percent confidence interval (dB) | |
|:---:|:---:|:---:|
| | Expected | Achieved |
| 1 | ±1.3 | ±1.7 |
| 2 | ±1.3 | ±1.4 |
| 3 | ±1.2 | ±1.0 |
| 4 | ±1.1 | ±0.6 |
| 5 | ±1.0 | ±0.5 |
| 6 | ±0.8 | ±0.4 |
| 7 | ±0.9 | ±0.4 |

2.4 *Data Analysis*

All estimation formulas associated with the sample design described above are based on so-called ratio estimators.[1] Such estimators have the undesirable feature that in general they are biased. However, the bias decreases with sample size and can be ignored for large enough samples. Furthermore, ratio estimators have the desirable property that their sampling variance is small and that they allow a large amount of flexibility in the data analysis. The most important aspect of this flexibility is related to the analysis of subclasses of the population. Examples of this usage of the ratio estimator appear in most of the sections that follow where results are presented for specific transmission facilities.

All of the results presented here refer to the population defined above. In many cases where means, standard deviations, and proportions are discussed, these are estimates of population parameters based on the sample data. Because of the structuring of the sample, the estimate of the population mean, for example, is often weighted and therefore not identical with the unweighted sample mean.

The amount of data treated here is large and the data analysis formulas are complicated. Therefore, digital computer programming has been used extensively in all of the data analysis work.

Much effort was put into "data cleaning," that is, in scrutinizing the data collected in the field for errors, omissions, and inconsistencies. Several errors were unveiled and corrected. Most important among these were erroneous readings of the measurement instrument, errors in the facility classification, and errors in transcribing the data onto IBM cards.

III. PHYSICAL CHARACTERISTICS OF INTERTOLL TRUNKS

The results reported here are based on various record data distinct from the transmission measurement results discussed in Sections IV through VIII.

3.1 *Trunk Lengths and Airline Distances*

Trunk length information was included for each trunk listed in the frame from which the second stage sample was selected. For each of these trunks it was also possible to compute the airline distance between toll offices. This information was used to estimate the distribution of trunk lengths and the distribution of the ratio of trunk length to airline distance. The sampling plan used to achieve this was a one-

stage cluster sampling plan where the selection of clusters coincided with the selection of primary units discussed in Section II.

Table III gives the distributions of both trunk length and trunk mileage. It shows that although only 3 percent of the intertoll trunks are longer than 2000 miles, no less than 25 percent of the total intertoll trunk mileage is accounted for by trunks longer than 2000 miles.

The distribution of the ratio of trunk length to airline distance shows some small variations with the airline distance between toll offices. Thus, ratios up to five were found on a small proportion of the trunks that connect toll offices separated by less than 180 airline miles. On the other hand, the distribution was confined to ratio-values less than two when the airline distance exceeded 1450 miles. The average ratio over all trunk lengths was found to be 1.40.

The distribution of the ratio is quite naturally truncated by one at the lower end, and it has a strong positive skewness. The transformed variable $y = \log_{10}(r - 1)$, where $r$ is the ratio of trunk length to airline distance, is, however, close to normal in its distribution. Computed over all trunk lengths, $y$ has a mean of $-0.53$ and a standard deviation of 0.64. The mean of $y$ corresponds to a trunk length to airline distance ratio of 1.29, which is close to the median of the ratio. The fact that the median of $r$ is lower than the mean reflects the positive skewness of the distribution of the ratio.

### 3.2 Toll Office Characteristics

At the time the first-stage sample was selected in early 1964 there were 1544 Bell System toll offices in the USA and Canada that qualified under the toll office definition in Section II. It was estimated that the same geographical area contained more than 600 additional toll offices that were independently owned and therefore excluded from selection as primary units in the sample.

TABLE III — DISTRIBUTIONS OF TRUNK LENGTH

| Trunk length (miles) | Percentage of trunks | Percentage of trunk mileage |
|---|---|---|
| 0– 62.5 | 35 | 3 |
| 62.5– 125 | 21 | 6 |
| 125 – 250 | 14 | 8 |
| 250 – 500 | 11 | 12 |
| 500 –1000 | 9 | 19 |
| 1000 –2000 | 7 | 27 |
| 2000 –4000 | 3 | 25 |
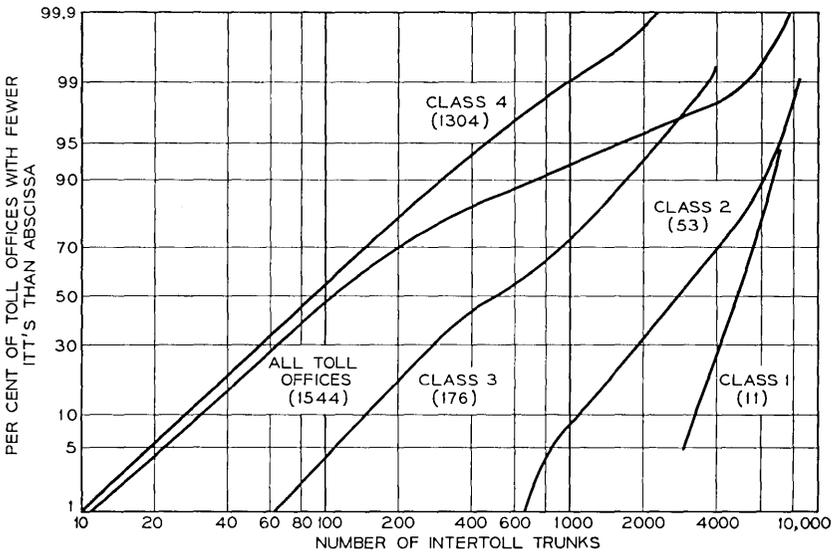
Fig. 1 — Distributions of sizes of toll offices.

The sizes of Bell System toll offices are represented by distribution curves in Fig. 1. Office size is measured by the number of intertoll trunks terminated in the toll switching machine. The figure demonstrates that office size increases with office class, and that the office size distribution is approximately log-normal within each of the four classes of offices. The number of Bell System toll offices belonging to each class is also given in the figure.

Table IV is another way to demonstrate the high concentration of intertoll trunks in large toll offices. This table gives the estimated percentage of intertoll trunks within each of seven mileage categories and over-all trunk lengths that interconnect toll offices of indicated classes. Notice that an estimated 50 percent of all intertoll trunks interconnect toll offices of class 1, 2, or 3. From Fig. 1 we find that toll offices of these classes constitute only 16 percent of all Bell System toll offices. Table IV also shows that 94 percent of the intertoll trunks have at least one of their end-points in a toll office of class 1, 2, or 3. The table further indicates that the concentration of trunks to high level toll offices is even more pronounced if attention is restricted to mileage categories that contain trunks longer than 125 miles.

The combined percentage estimates for all intertoll trunks terminat-

TABLE IV—PERCENT OF INTERTOLL TRUNKS INTERCONNECTING
TOLL OFFICES

| Office class | | Per cent of intertoll trunks within mileage category | | | | | | | Percent of all intertoll trunks |
|---|---|---|---|---|---|---|---|---|---|
| From | To | 0– 62.5 | 62.5– 125 | 125– 250 | 250– 500 | 500– 1000 | 1000– 2000 | 2000– 4000 | |
| 1, 2 | 1, 2 | 8 | 6 | 20 | 38 | 66 | 66 | 89 | 24 |
| 1, 2, 3 | 1, 2, 3 | 23 | 35 | 58 | 79 | 96 | 97 | 100 | 50 |
| 1, 2, 3 | 4 | 63 | 60 | 40 | 21 | 4 | 3 | — | 44 |
| 4 | 4 | 14 | 5 | 2 | — | — | — | — | 6 |
| Percent of all intertoll trunks | | 35 | 21 | 14 | 11 | 9 | 7 | 3 | 100 |

ing on at least one end in each of the four classes of toll offices are:

| | |
|---|---|
| Regional centers | 21 percent |
| Sectional centers | 58 percent |
| Primary centers | 47 percent |
| Toll centers or toll points | 50 percent |

Closely associated with office rank is the type of switching machine used in the office. From the standpoint of transmission performance the major distinction in toll switching equipment arises from the use of 4-wire versus 2-wire switching, since the latter will ordinarily require additional equipment (a hybrid transformer and impedance matching network) to convert a 4-wire transmission path to a 2-wire path for switching.

Table V lists the estimated percentages of intertoll trunks within each mileage category that interconnect two 4-wire machines, two

TABLE V—PERCENT OF INTERTOLL TRUNKS INTERCONNECTING
SWITCHING MACHINES

| Switch type | | Percent of intertoll trunks within mileage category | | | | | | | Percent of all intertoll trunks |
|---|---|---|---|---|---|---|---|---|---|
| | | 0– 62.5 | 62.5– 125 | 125– 250 | 250– 500 | 500– 1000 | 1000– 2000 | 2000– 4000 | |
| From | To | | | | | | | | |
| 4-wire | 4-wire | 9 | 6 | 25 | 44 | 70 | 77 | 91 | 26 |
| 2-wire | 4-wire | 41 | 54 | 60 | 47 | 29 | 23 | 9 | 44 |
| 2-wire | 2-wire | 50 | 40 | 15 | 9 | 1 | — | — | 30 |
| Percent of all intertoll trunks | | 35 | 21 | 14 | 11 | 9 | 7 | 3 | 100 |

2-wire machines and a 2-wire with a 4-wire machine. In 1964 there were 73 4-wire toll switching machines (types 4A or 4M crossbar) in service in the Bell System; the table shows that 26 percent of all intertoll trunks interconnect these toll offices and that 70 percent touch at least one of them.

The combined percentage estimates for all intertoll trunks terminating on at least one end in one of the four major types of Bell System toll switching machines are:

|  |  |  |
|---|---|---|
| 4A or 4M crossbar | (4-wire) | 70 percent |
| Crossbar tandem | (2-wire) | 40 percent |
| Number 5 crossbar | (2-wire) | 19 percent |
| Step-by-step | (2-wire) | 29 percent |

### 3.3 *Facility Composition*

More important than toll offices in terms of transmission performance is the facility makeup of intertoll trunks. Percentage estimates of intertoll trunks within each mileage category and over-all trunk lengths are listed according to line facility makeup in Table VI.

TABLE VI—FACILITY COMPOSITION OF INTERTOLL TRUNKS

| Facility | Percent of intertoll trunks within mileage category | | | | | | | Percent of all intertoll trunks |
|---|---|---|---|---|---|---|---|---|
|  | 0–62.5 | 62.5–125 | 125–250 | 250–500 | 500–1000 | 1000–2000 | 2000–4000 |  |
| Voice frequency | 29 | 3 | 1 |  |  |  |  | 11 |
| N1-carrier | 42 | 11 | 9 |  |  |  |  | 19 |
| ON-carrier | 22 | 47 | 17 | 1 |  |  |  | 20 |
| C or J carrier |  | 6 | 3 | 1 |  |  |  | 2 |
| K-carrier | 1 | 5 | 8 | 5 | 2 | 1 |  | 3 |
| L-carrier | 2 | 5 | 14 | 15 | 15 | 4 | 4 | 7 |
| Microwave radio | 3 | 16 | 25 | 51 | 42 | 33 | 55 | 21 |
| L-carrier and radio |  | 1 | 6 | 9 | 28 | 53 | 37 | 9 |
| N1 and ON carrier | 1 | 5 | 2 |  |  |  |  | 2 |
| Noncompandored carrier combinations |  |  | 4 | 4 | 8 | 7 | 4 | 3 |
| Compandored and noncompandored carrier combinations |  | 1 | 11 | 14 | 5 | 2 |  | 3 |
| All intertoll trunks | 35 | 21 | 14 | 11 | 9 | 7 | 3 | 100 |

The table partitions all intertoll trunks into eleven facility categories. (For detailed descriptions of the commonly found telephone carrier systems see Refs. 3 through 7.) The microwave radio category includes all the commercial telephone carrier systems using line-of-sight radio as the transmission medium. A majority of the trunks in this category used the TD-2 radio system. Four specific single facility categories include the two short haul cable carrier systems, N1 and ON, the long-haul K-carrier system, and coaxial cable carrier, designated L-carrier. (Most intertoll trunks in this category used L3-carrier.) Trunks made up entirely of the older C or J open wire carrier systems were combined into one category because they represent such a minor contribution to the toll network. The most widely used combination of the two long-haul line facilities, coaxial cable carrier and microwave radio, and the two short-haul facilities, N1-carrier and ON-carrier, are listed separately for emphasis. Intertoll trunks made up of all other combinations of carrier facilities are divided into two categories on the basis of whether any compandored carrier (N1-, ON- or O-carrier) was used in their makeup.

The voice frequency category in Table VI includes only those intertoll trunks made up entirely of voice frequency facilities. Eleven percent of the intertoll trunks fall into this category but 80 percent

TABLE VII — CHANNEL BANK MAKEUP OF INTERTOLL TRUNKS

| Channel banks | | Percent of intertoll trunks within mileage category | | | | | | | Percent of all intertoll trunks |
|---|---|---|---|---|---|---|---|---|---|
| Number | Type | 0-62.5 | 62.5-125 | 125-250 | 250-500 | 500-1000 | 1000-2000 | 2000-4000 | |
| 0 | Voice Frequency | 29 | 3 | 1 | | | | | 11 |
| 2 | N1 | 41 | 10 | 9 | | | | | 18 |
| 4 | N1 | 1 | 1 | | | | | | 1 |
| 2 | O | 22 | 47 | 13 | 1 | | | | 19 |
| 4 | O | | | 4 | | | | | 1 |
| 2 | C | | 5 | 1 | 1 | | | | 1 |
| 2 | A | 6 | 27 | 56 | 74 | 76 | 66 | 52 | 37 |
| 4 | A | | 1 | 3 | 9 | 17 | 26 | 37 | 6 |
| 6 | A | | | | 1 | 2 | 6 | 10 | 1 |
| 8 | A | | | | | | | 1 | 0 |
| ≥4 | A and N1 | | | 4 | 3 | 2 | 1 | | 1 |
| ≥4 | A and O | | 1 | 7 | 11 | 3 | 1 | | 2 |
| ≥4 | N and O | 1 | 5 | 2 | | | | | 2 |
| All intertoll trunks | | 35 | 21 | 14 | 11 | 9 | 7 | 3 | 100 |

of these are shorter than 15 miles and 93 percent shorter than 62.5 miles. Trunk records revealed that about 1 percent of all intertoll trunks contain a section of voice frequency facilities in tandem with carrier facilities. These trunks were classified by their carrier facility makeup for Table VI. Using this rule for facility classification, 83 percent of all intertoll trunks have a homogeneous line facility composition and more than half of the remainder are made up of the long-haul combination of coaxial cable carrier and microwave radio.

The estimated percentages of all intertoll trunks containing any of the eight major carrier facility types in their facility makeup, in descending order of facility occurrence, are:

| *Facility* | *Percent* |
|---|---|
| Microwave radio | 33 |
| ON-carrier | 24 |
| N1-carrier | 22 |
| L-carrier | 18 |
| K-carrier | 6 |
| C-, J-, or O-carrier | 4 |

The estimates given in Table VI do not differentiate between carrier facility combinations connected at voiceband frequencies and those connected at group, supergroup, or mastergroup frequencies, nor do they distinguish tandem combinations of the same facility type connected at voice frequency from single facility trunks.

Information about the number of voice frequency modulators and channel bank filters used in intertoll trunks is, however, given by Table VII. This table lists percentage estimates of intertoll trunks within each mileage category and over-all trunk lengths classified by the types and number of voiceband channel banks in their makeup. The 13 categories listed identify all carrier trunks with five types of channel bank, but not with a specific equipment configuration, for several generations of a particular type of channel bank are found in service. A-type channel banks[8] are used on the long haul carrier facilities, J, K, L, and microwave radio; N1-type is used on N1-carrier; O-type on O- and ON-carrier; and C-type on C-carrier. Note that when intertoll trunks made up entirely of voice frequency facilities are included, 86 percent of all intertoll trunks encounter no more than one pair of channel banks and only 5 percent are equipped with combinations of dissimilar channel banks.

The degree of interconnection of various facilities to form a trunk

is only party exposed by Tables VI and VII. It is common to combine facilities that use A-type channel banks without demodulating to voice frequency. This is done by the use of group, supergroup, and mastergroup connectors.

Table VIII shows the use of such high frequency connectors in Bell System intertoll trunks. Notice that this table is only concerned with those trunks that contain at least one pair of A-type channel banks. Furthermore, any portion of such a trunk that contains channel banks different from the A-type is disregarded. The table indicates a strong trend toward a larger number of connectors for the longer trunks. Thus, the average number of group connectors increases monotonically from 0 for short trunks to 1.8 for trunks longer than 2000 miles. Likewise, the average number of supergroup connectors increases from 0 for short trunks to 2.3 for trunks in the 2000-4000 mile category.

The average number of high frequency connectors, that is, of connectors at group, supergroup, or mastergroup frequencies on trunks with A-type channel banks can be found in each mileage category by addition of the corresponding averages for each of the three categories of connectors. Thus, the average number of high frequency connectors on trunks with A-type channel banks in the 2000-4000 mile category equals 4.2. No trunks in the sample shorter than 1000 miles contained more than 5 high frequency connectors. Among the trunks in the 1000-2000 mile category, 2 percent contain 8 or more high frequency connectors, while 3 percent of the trunks longer than 2000 miles contain 8 or more high frequency connectors.

3.4 *The Number of FM Terminals per Radio Facility*

The intertoll trunks in the sample that used microwave radio as a transmission facility were analyzed to determine the number of FM terminals per radio facility. Frequency modulation and demodula-

TABLE VIII—AVERAGE NUMBER OF PAIRS OF A-TYPE CHANNEL
BANKS AND HIGH FREQUENCY CONNECTORS ON INTERTOLL
TRUNKS WITH A-TYPE CHANNEL BANKS

| Miles | 0–62.5 | 62.5–125 | 125–250 | 250–500 | 500–1000 | 1000–2000 | 2000–4000 | All Trunks |
|---|---|---|---|---|---|---|---|---|
| Pairs of A-banks | 1.0 | 1.0 | 1.0 | 1.1 | 1.2 | 1.4 | 1.6 | 1.2 |
| Group connectors | 0.0 | 0.1 | 0.2 | 0.5 | 1.1 | 1.7 | 1.8 | 0.7 |
| Supergroup connectors | 0.0 | 0.1 | 0.2 | 0.5 | 0.8 | 1.4 | 2.3 | 0.7 |
| Mastergroup connectors | 0.0 | 0.0 | 0.0 | 0.1 | 0.1 | 0.3 | 0.1 | 0.1 |

tion is resorted to when groups, supergroups, or mastergroups of voice frequency channels are dropped from a radio system at an inter-mediate point. The number of pairs of FM terminals in a radio facility is important from a design standpoint because of its effect on the resulting noise level.

Table IX lists the average radio facility length, the average num-ber of pairs of FM terminals per radio facility, and the average length between FM terminals for seven different mileage categories. The length of a radio facility is the sum of the lengths of all sections of radio facility used on a given trunk. This rule of computation is used even when the radio facility sections of a trunk are not all adjacent to each other. The most important result indicated by Table IX is that the length between FM terminals is strongly correlated with the radio facility length.

The average length between FM terminals computed over all radio facilities is 240 miles, while the average length between FM terminals for radio facilities longer than 2000 miles is 495 miles. The average length between FM terminals is approximately proportional to the square root of the radio facility length. For a radio facility length of 4000 miles, the average length between FM terminals is estimated at 620 miles. Correspondingly, the average number of pairs of FM terminals for a 4000 mile long radio facility is approximately 6.5. The distribution of the number of pairs of FM terminals for radio facilities longer than 2000 miles shows that only one percent contain 9 or more pairs of FM terminals in tandem.

IV. 1000-HZ LOSS

The switch-to-switch loss of intertoll trunks at 1000 Hz have been analyzed. The term switch-to-switch loss is used to refer to the total loss between outgoing switch appearances at the originating and ter-minating ends of a trunk. It equals the loss inserted into a connection by switching the trunk into an operating condition.

The loss at 1000 Hz can be studied from three different viewpoints: design, performance, and maintenance. The switch-to-switch loss at which each trunk in the sample was designed to operate was ex-tracted from trunk records in each of the toll offices visited. The actual switch-to-switch loss was found by measurement. The main-tenance effect is measured by the difference between measured loss and design loss, if it is assumed that all trunks met the design value exactly when they were first put into service.

TABLE IX — RADIO SYSTEM COMPOSITION VERSUS RADIO
SYSTEM LENGTH

| Radio facility length (miles) | Average radio facility length (miles) | Average number of pairs of FM terminals per radio facility | Average length between FM terminals (miles) |
|---|---|---|---|
| 0–62.5 | 39 | 1.0 | 39 |
| 62.5–125 | 80 | 1.3 | 64 |
| 125–250 | 173 | 1.5 | 120 |
| 250–500 | 341 | 2.1 | 168 |
| 500–1000 | 708 | 2.8 | 255 |
| 1000–2000 | 1264 | 3.8 | 337 |
| 2000–4000 | 2565 | 5.2 | 495 |

## 4.1 Design Loss

The scatter diagram of Fig. 2 shows how the design losses of inter-toll trunks vary with trunk length. Intertoll trunks are designed according to the via net loss concept discussed by H. R. Huntley.[9] According to this concept, the design loss for carrier trunks increases linearly with trunk length from 0.5 dB to 2.6 dB, from 165 to 1565 miles; this is shown as an exponential trend in Fig. 2 because of the logarithmic mileage scale. Trunks longer than 1565 miles and trunks between regional centers are designed to have a loss of 0 dB and to be equipped with echo suppressors. Fig. 2 shows that in 1964, such trunks had generally a design loss of 0.5 dB. The deviations from these rules
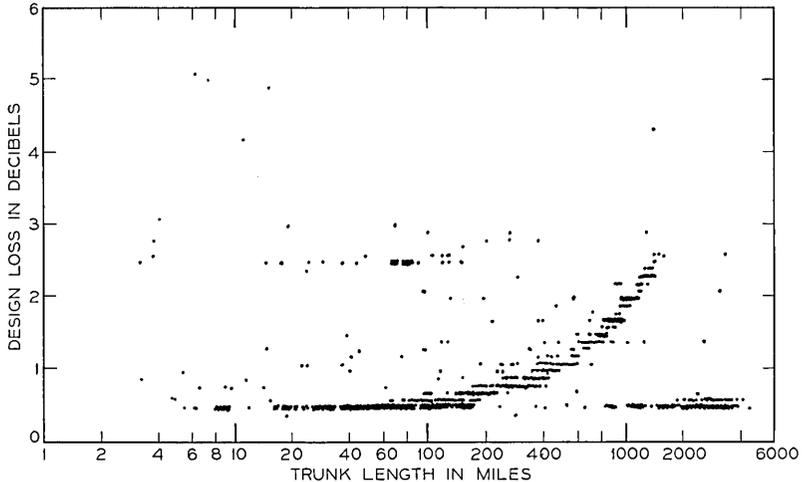


Fig. 2 — Scatter diagram of switch-to-switch design loss versus trunk length.

are given by trunks on noncarrier facilities and trunks in "unbalanced" toll offices, that is, 2-wire toll offices of class 1, 2, or 3, that do not meet certain objectives for uniformity in office cabling impedance. The scatter diagram indicates that the design loss distributions are non-normal; definite modes exist at loss values given by the via net loss computation. The scatter diagram also shows that the adherence to a uniform loss design improves as the trunk length increases.

A summary of the results of the data analysis for design loss is given in columns 2 and 3 of Table X. As in most tables in Sections IV to VIII, estimates are given of the mean and the standard deviation of the population distribution, and the mean is equipped with its estimated 90 percent confidence interval. The first seven mileage categories in this table correspond to the seven substrata defined earlier. A further breakdown has been made of the sixth substratum, which contains trunks between 1000 and 2000 miles long. The reason for this additional breakdown is the design rule already mentioned that prescribes the use of echo suppressors and a switch-to-switch loss of 0 dB for all trunks longer than 1565 miles as well as on trunks between regional centers.

The break in the sixth mileage category is at 1465 miles rather than at 1565 miles because all trunks in the sample between 1465 and 1565 miles long were equipped with echo suppressors and had a design loss of 0.5 dB. None of these trunks interconnected two regional centers, and according to present design practices only 17 percent of them

TABLE X — INTERTOLL TRUNK SWITCH-TO-SWITCH LOSSES
AT 1000 Hz

| Trunk length (miles) | Design loss (dB) | | Measured loss (dB) | | Measured loss minus design loss (dB) | |
|---|---|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 0 – 62.5 | 0.9 ± 0.2 | 1.0 | 1.2 ± 0.2 | 1.4 | 0.3 ± 0.2 | 1.0 |
| 62.5– 125 | 1.0 ± 0.3 | 0.9 | 1.2 ± 0.3 | 1.3 | 0.1 ± 0.2 | 1.1 |
| 125 – 250 | 0.8 ± 0.1 | 0.4 | 1.1 ± 0.2 | 1.2 | 0.3 ± 0.2 | 1.2 |
| 250 – 500 | 1.0 ± 0.1 | 0.3 | 1.4 ± 0.2 | 1.1 | 0.4 ± 0.2 | 1.1 |
| 500 –1000 | 1.5 ± 0.1 | 0.3 | 1.9 ± 0.2 | 1.1 | 0.4 ± 0.2 | 1.0 |
| 1000 –2000 | 1.6 ± 0.2 | 0.9 | 2.1 ± 0.2 | 1.3 | 0.4 ± 0.2 | 1.1 |
| 2000 –4000 | 0.6 ± 0.1 | 0.2 | 0.9 ± 0.1 | 1.2 | 0.3 ± 0.1 | 1.2 |
| 1000 –1465 | 1.9 ± 0.2 | 0.7 | 2.3 ± 0.3 | 1.2 | 0.3 ± 0.2 | 1.1 |
| 1465 –2000 | 0.6 ± 0.1 | 0.4 | 1.3 ± 0.3 | 1.2 | 0.8 ± 0.3 | 1.2 |

would be equipped with echo suppressors and have a design loss of 0 dB, while the remaining 83 percent would be without echo suppressors and have a design loss of 2.6 dB. This difference does not mean that the 1964 trunks were incorrectly designed; rather, it reflects a small change in design rules that has been introduced after 1964.

Table X shows the same trend in mean design loss observed from the scatter diagram. The improved adherence to a uniform loss design with increasing trunk length is seen from the decrease in standard deviation with trunk length. The only notable exception is found in the 1000 to 1465-mile category. The higher standard deviation here is caused by the fact that 14 percent of the trunks in this category interconnect regional centers. These trunks would, by present design practices, be equipped with echo suppressors and have a design loss of 0 dB, while in 1964 their design loss was 0.5 dB. An additional one percent of the trunks in this mileage band are estimated to have been equipped with echo suppressors and accordingly have a design loss of 0.5 dB. According to present design rules, these trunks would not contain echo suppressors, and their design loss would be 2.6 dB.

## 4.2 *Measured Loss*

All survey measurements of intertoll trunks were made so as to describe as closely as possible the transmission characteristics from switch through switch as they appear when the trunk is being used in a built-up connection between subscribers. The maintenance and testing facilities provided in each toll office, such as toll testboards and code test lines, are geared toward the same objectives and were therefore used extensively in the survey.

The loss measurements were made with the far end of the trunk connected to a 1000 Hz one-milliwatt testing power source. This connection was either supplied by a dialable test termination or it was made manually at the far-end test board. The received level was then measured at the near-end toll office.

Figure 3 is a scatter diagram of measured loss versus trunk length. The fourth and fifth columns of Table X list the corresponding means with 90 percent confidence intervals and standard deviations in each of the mileage categories already discussed. The distributions are essentially normal with a slight positive skewness in some length categories. It is quite remarkable that no trace remains in the distributions of measured loss of the very pronounced non-normality of the
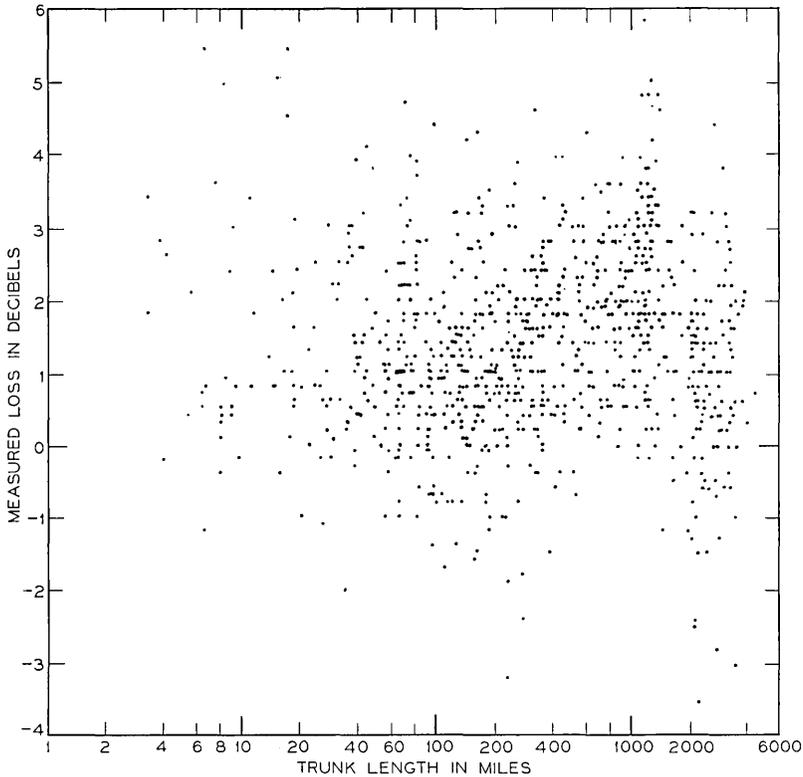
Fig. 3 — Scatter diagram of switch-to-switch measured loss versus trunk length.

distributions of design loss. The reason for this is that the variability resulting from maintenance overshadows the design variability.

Table X shows that the average measured loss follows the same pattern as the average design loss as a function of trunk length. The standard deviations of measured loss decrease as trunk length increases over the first four mileage categories, similar to what was observed for design loss in the previous section. After this, however, the standard deviation remains generally constant and it can be noted that it is substantially larger than the standard deviation of the corresponding design loss distributions.

### 4.3 Loss Maintenance

The loss maintenance of intertoll trunks can be studied by considering the difference between measured loss and design loss. The last two

columns of Table X summarize the results. The distributions of these loss differences within mileage categories are all close to normal.

The table shows that the measured losses are on the average somewhat larger than the design losses. Comparison of the third and seventh columns shows that the standard deviation of the loss difference is larger than the standard deviation of the design loss, as stated in the previous section.

Finally, we notice that the standard deviation shows a small but statistically significant increase with trunk length over the mileage categories that contain short-haul trunks (up to 250 miles), and that it also increases (again significantly) with trunk length over the mileage categories to which long-haul trunks belong (longer than 500 miles). This indicates that within each of these two broad classes of trunks, the longer trunks are somewhat more difficult to maintain at the loss value at which they have been designed to operate. This is to be expected since the longer trunks contain more sources for loss variation and more points where loss adjustment can be applied.

## V. FREQUENCY RESPONSE

### 5.1 *Measurement Procedure*

Frequency response was measured by noting the loss at each of 9 frequencies throughout the voiceband. Tones were sent from a variable frequency oscillator located at the far end toll testboard and the received level was noted at the near end toll testboard. The frequency of each tone was verified by a frequency counter at the receive end and received levels were measured using the meter available at the toll testboard. The frequencies of measurement were 200, 300, 400, 1000, 1700, 2300, 3000, 3200, and 3400 Hz. The frequency response of the office meter was also noted and measured loss values were arithmetically corrected for any roll-off in the meter.

Since the range of the office meters was generally limited to switch-to-switch losses less than 31 dB, it was not always possible to measure loss at all nine frequencies. This was the case for trunks with compandored carrier systems where the loss at 3400 Hz was almost always beyond the range of the meter. In this situation a value of 31 dB was arbitrarily assigned as the switch-to-switch loss. Such a procedure is necessary to provide a realistic, albeit conservative, estimate of the mean loss at 3400 Hz for groups of trunks with varying facility composition.

5.2 *Frequency Response of Facilities and of Office Equipment*

As we mentioned, all frequency response measurements in the survey were made from test board to test board, that is, they included the sum of the effects of the facility and the office equipment at each end. In order to isolate these separate contributions, the assumption was made that the frequency response of office equipment and facility are independent random variables which add to produce the overall switch-to-switch frequency response of a trunk.

Each major facility category was divided into three different subclasses: those in which the trunk has 2-wire switching at both ends, those with 2-wire at one end and 4-wire at the other, and those with 4-wire at both ends. For the largest categories, that is, those containing one and two pair of A-channel banks, the largest proportion of the sample was found in the category with 4-wire switching at each end. The two subclasses consisting of trunks with one pair of A-channel banks and 4-wire switching at each end and trunks with two pair of A-channel banks and 4-wire switching at each end were used to estimate the frequency response of A-channel banks and of 4-wire office equipment.

The difference between the estimated frequency response characteristics for these two subclasses of trunks was taken as the frequency response of facilities with one pair of A-channel banks. An estimate of twice the frequency response of office equipment for 4-wire offices was obtained by subtracting the above estimate for facilities with one pair of A-channel banks from the frequency response for trunks with one pair of A-channel banks and 4-wire switching at each end.

The frequency response of two sets of 2-wire office equipment was estimated in an analogous manner, using results for trunks with one pair of A-channel banks and 2-wire switching at each end in conjunction with the facility estimates. The separate estimates for the frequency response of the facility and the two types of office equipment were combined and found in excellent agreement with observed frequency response for trunks with 2-wire switching at one end and 4-wire at the other.

The estimates for office equipment thus obtained were then used to obtain estimates of frequency response for the other types of facilities. Table XI shows the estimates of mean, standard deviation, and 90 percent confidence interval for the mean of facility loss differences relative to 1000-Hz loss. The loss estimates omitted from the table correspond to switch-to-switch losses in excess of 31 dB.

TABLE XI — FACILITY LOSS DIFFERENCES IN dB RELATIVE TO 1000 Hz LOSS

| Frequency (Hz) | A channel banks | | O channel filters | | N1 channel filters | | VF cable | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 200 | $1.2 \pm 0.6$ | 1.2 | $13.1 \pm 1.5$ | 5.2 | $4.4 \pm 1.4$ | 2.7 | $2.0 \pm 1.5$ | 2.7 |
| 300 | $0.1 \pm 0.3$ | 0.5 | $2.9 \pm 0.7$ | 1.8 | $0.7 \pm 0.5$ | 0 | $0.1 \pm 0.7$ | 1.4 |
| 400 | $0.3 \pm 0.2$ | 0.5 | $0.4 \pm 0.3$ | 1.0 | $0.3 \pm 0.3$ | 0.7 | $0 \pm 0.4$ | 1.1 |
| 1700 | $-0.2 \pm 0.2$ | 0.4 | $0.3 \pm 0.3$ | 0.8 | $0.3 \pm 0.2$ | 0.6 | $-0.1 \pm 0.2$ | 0.3 |
| 2300 | $-0.3 \pm 0.2$ | 0.2 | $0.2 \pm 0.4$ | 1.6 | $0.4 \pm 0.4$ | 0.8 | $-0.1 \pm 0.3$ | 0.2 |
| 3000 | $0.3 \pm 0.3$ | 0.7 | $2.9 \pm 0.5$ | 2.0 | $4.5 \pm 0.6$ | 1.6 | $2.0 \pm 0.7$ | 1.4 |
| 3200 | $0.7 \pm 0.4$ | 0.8 | $8.9 \pm 0.6$ | 2.1 | $17.4 \pm 1.2$ | 4.3 | $4.2 \pm 1.0$ | 2.3 |
| 3400 | $2.3 \pm 0.4$ | 0.7 | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $8.9 \pm 2.0$ | 4.6 |

Notice that the bandwidth of a facility with A-channel banks is superior to that of the short haul facilities with O-carrier or N1-carrier channel filters. In fact, a facility with three pair of A-channel banks in tandem will have greater bandwidth than either of the short-haul carrier facilities. It may also be noticed that between the short-haul carrier facilities, N1-carrier has the superior frequency response characteristic at low frequencies while facilities with O-carrier terminals are superior at the high frequencies.

The loss differences for office equipment are given in Table XII. Comparison with Table XI shows that office equipment may contribute more to switch-to-switch loss at low frequencies than a pair of A-channel banks. This is especially true when both offices use 2-wire switching, in which case the mean loss difference of office equipment at low frequencies exceeds even that of two pair of A-channel bank facilities in tandem.

### 5.3 Frequency Response of Trunks

The loss data were also analyzed within the mileage categories defined previously. The results of this analysis are presented in Table XIII. A general trend toward smaller values of mean loss and standard deviation with increasing trunk length is evident at both the lower and higher frequencies of the voice band. Hence the longer trunks are seen to have a frequency response characteristic that is superior to that of the shorter ones. This reflects the transition from compandored carrier on short-haul trunks to A-channel banks for long-haul trunks and is consistent with the superior performance of long-haul carrier noted previously. The high percentage of voice frequency cable and N1-carrier in the shortest category accounts for

TABLE XII — LOSS DIFFERENCES IN dB RELATIVE TO 1000 Hz
LOSS FOR OFFICE TRUNKING EQUIPMENT

| Frequency (Hz) | 4-Wire office | | 2-Wire office | |
|---|---|---|---|---|
| | Mean | Standard deviation | Mean | Standard deviation |
| 200 | 0.7 ± 0.8 | 0.4 | 1.4 ± 0.9 | 1.3 |
| 300 | 0.3 ± 0.5 | 0.4 | 0.7 ± 0.5 | 0.8 |
| 400 | 0.3 ± 0.4 | 0.3 | 0.3 ± 0.2 | 0.1 |
| 1700 | 0.1 ± 0.3 | 0.2 | 0.2 ± 0.2 | 0.2 |
| 2300 | 0.2 ± 0.3 | 0.3 | 0.4 ± 0.3 | 0.5 |
| 3000 | 0.3 ± 0.5 | 0.1 | 0.4 ± 0.4 | 0.4 |
| 3200 | 0.3 ± 0.5 | 0 | 0.4 ± 0.4 | 0.4 |
| 3400 | 0.6 ± 0.7 | 0.5 | 0.5 ± 0.6 | 0.9 |

TABLE XIII — SWITCH-TO-SWITCH LOSS IN dB ON INTERTOLL TRUNKS

| Frequency (Hz) | Miles | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0–62.5 | | 62.5–125 | | 125–250 | | 250–500 | |
| | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| 200 | 8.9 ± 1.5 | 6.0 | 12.6 ± 2.0 | 7.9 | 8.8 ± 1.3 | 7.5 | 7.0 ± 1.1 | 6.1 |
| 300 | 3.4 ± 0.5 | 2.4 | 4.1 ± 0.7 | 2.4 | 3.5 ± 0.4 | 3.0 | 3.2 ± 0.4 | 2.6 |
| 400 | 2.1 ± 0.3 | 1.8 | 2.2 ± 0.3 | 1.5 | 2.2 ± 0.2 | 1.7 | 2.5 ± 0.3 | 1.7 |
| 1000 | 1.2 ± 0.2 | 1.4 | 1.2 ± 0.3 | 1.2 | 1.1 ± 0.2 | 1.2 | 1.4 ± 0.2 | 1.1 |
| 1700 | 1.6 ± 0.3 | 1.6 | 1.6 ± 0.3 | 1.5 | 1.5 ± 0.2 | 1.6 | 1.5 ± 0.2 | 1.3 |
| 2300 | 2.1 ± 0.3 | 1.9 | 1.8 ± 0.3 | 1.8 | 1.7 ± 0.2 | 1.8 | 1.8 ± 0.2 | 1.6 |
| 3000 | 5.1 ± 0.6 | 3.2 | 5.3 ± 1.3 | 6.2 | 4.2 ± 0.7 | 4.2 | 3.5 ± 0.5 | 3.2 |
| 3200 | 12.4 ± 1.6 | 8.2 | 10.0 ± 1.6 | 7.3 | 8.2 ± 1.9 | 7.8 | 5.1 ± 0.9 | **5.7** |
| 3400 | 22.8 ± 2.6 | 11.4 | 22.5 ± 2.8 | 12.4 | 16.4 ± 3.3 | 12.8 | 9.6 ± 1.6 | 9.7 |

| Frequency (Hz) | Miles | | | | | |
|---|---|---|---|---|---|---|
| | 500–1000 | | 1000–2000 | | 2000–4000 | |
| | Mean | Standard Deviation | Mean | Standard Deviation | Mean | Standard Deviation |
| 200 | 5.7 ± 0.4 | 3.4 | 5.3 ± 0.3 | 2.6 | 4.2 ± 0.4 | 2.4 |
| 300 | 2.9 ± 0.2 | 1.4 | 3.0 ± 0.2 | 1.5 | 1.9 ± 0.3 | 1.5 |
| 400 | 2.8 ± 0.2 | 1.4 | 3.0 ± 0.3 | 1.5 | 1.9 ± 0.2 | 1.4 |
| 1000 | 1.9 ± 0.2 | 1.1 | 2.1 ± 0.2 | 1.3 | 0.9 ± 0.1 | 1.2 |
| 1700 | 1.8 ± 0.2 | 1.3 | 1.9 ± 0.2 | 1.4 | 0.9 ± 0.2 | 1.4 |
| 2300 | 2.0 ± 0.2 | 1.3 | 2.1 ± 0.2 | 1.5 | 1.1 ± 0.2 | 1.4 |
| 3000 | 3.2 ± 0.4 | 2.9 | 3.2 ± 0.3 | 1.7 | 2.3 ± 0.2 | 1.6 |
| 3200 | 4.2 ± 0.5 | 3.9 | 3.8 ± 0.3 | 2.2 | 2.9 ± 0.2 | 1.7 |
| 3400 | 7.2 ± 0.8 | 6.1 | 7.3 ± 0.6 | 4.5 | 6.1 ± 0.5 | 2.7 |

the superior low frequency performance of trunks in that category compared with trunks from 62.5 to 125 miles long where O-carrier terminals predominate (see Table VII).

In the first two mileage categories notice that the estimated mean loss at 3400 Hz is above 20 dB. Included in these estimates are a high percentage of assigned values of 31 dB for loss too high to be measured. Hence these estimates are best interpreted as indicating a mean loss in excess of 20 dB rather than as true estimates of mean loss. A similar statement holds for the estimated mean loss at 3400 Hz in the 125 to 250 miles category; that is, the estimate of 16.4 dB indicates mean loss exceeding 16 dB. With these exceptions the effect of assigned values is to provide more realistic estimates rather than to alter the interpretation of the estimates.

A comparison between frequency responses for short and long intertoll trunks is made in Fig. 4. It contains plots of the median switch-to-switch losses as a function of frequency for trunks in the first and last of the seven mileage categories. Notice that these are not median curves in the sense that 50 percent of all frequency response curves lie on or below these curves; rather, they connect such points at each frequency that 50 percent of the trunks have a lower
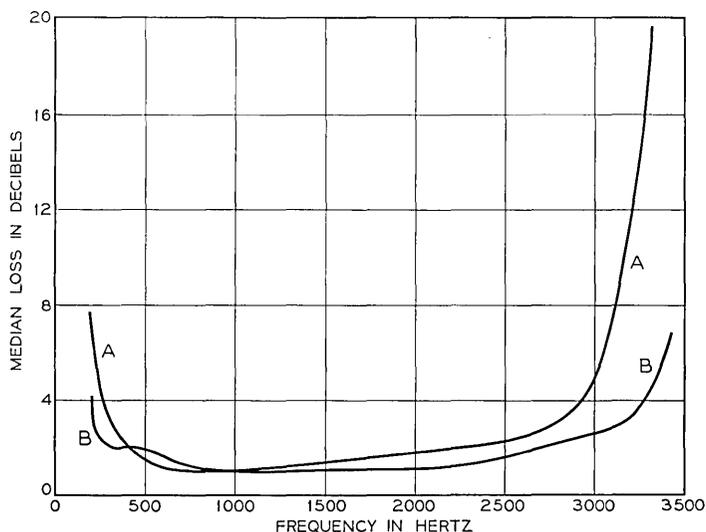


Fig. 4 — Median switch-to-switch loss for (a) short (0 to 62.5 miles) and (b) long (2000 to 4000 miles) intertoll trunks.

loss at that frequency. Fig. 4 shows clearly the larger bandwidth for the longer trunks.

## VI. BACKGROUND NOISE

The noise on a telephone communication channel affects the transmission of both spoken messages and data signals. From a practical standpoint, the noise is important only when it is interfering with or disturbing to the transmission of a signal. Hence the noise evaluation of a telephone communication channel seeks to quantify the interfering or disturbing effect of the noise in relation to particular types of information bearing signals. The character of this evaluation takes different forms for different types of signals. The disturbing or annoying effect of noise for spoken messages is related to the time-average noise power, while the most interfering effect of the noise on a data signal is related to the peaks of the noise voltage. Thus, although we are dealing with the same noise in either case, different aspects of that noise are important for different types of signals. In presenting survey results, we refer to "background noise" when dealing with those noise aspects that basically affect the transmission of spoken messages, while we use "impulse noise" to describe those aspects that most seriously affect data transmission.

### 6.1 *Measurement Procedure*

All measurements of background noise in the survey were made with the far end of the trunk connected to a quiet termination, supplied either by a dialed test termination or by a manually established connection. Noise levels were then measured at the near end with the 3A noise measuring set used both with C-message weighting and with 3-kHz flat weighting. All noise measurements were made during the busy period of an ordinary business day.

### 6.2 *Results of 3A Noise Level Measurements*

A scatter diagram of 3A noise levels with C-message weighting versus trunk length is shown in Fig. 5. All noise levels here are given "as measured," that is, referred to the receive switch of each measured intertoll trunk. The scatter diagram exhibits clearly the important facts that the mean noise level increases with the length of the trunk, while the variability decreases.

The regression line in Fig. 5 gives an estimate of the mean noise level under the assumption that the mean noise level is linearly
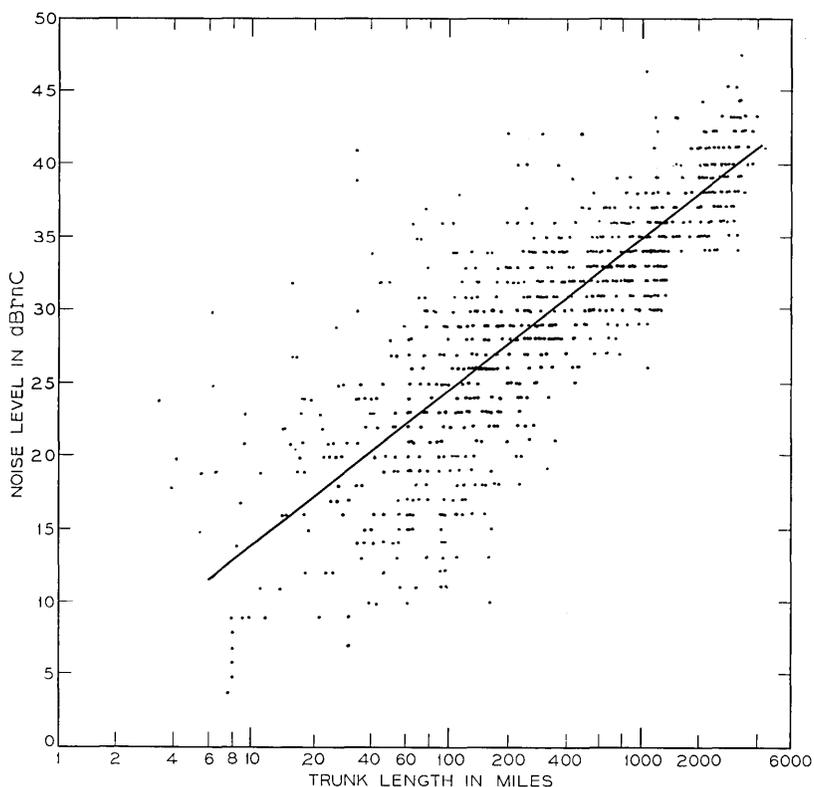
Fig. 5 — Scatter diagram of 3A noise level at receive switch versus trunk length.

related to the logarithm of the trunk length. The equation for the regression line is

$$N = 3.7 + 3.1 \log_2 l$$

where $l$ is the trunk length in miles and $N$ is the average 3A noise level in dBrnC. This equation shows that the average noise level increases by 3.1 dB for each doubling of length of trunk and that the average noise level at 4000 miles is 40.8 dBrnC.

The fact that the variance is not constant with the trunk length affected this regression analysis; the least squares fit to the noise levels as a function of the logarithm of the trunk length was weighted in inverse proportion to the variance about the regression line. This means that higher weight was given to those observations that show

a small spread about their mean. The variance about the regression line was in each case computed as the variance in a mileage category minus the contribution to that variance that occurred because the mean noise level varies with the trunk length.

The means and standard deviations of the noise distributions in each of the mileage categories discussed earlier are listed in the second and third columns of Table XIV. As before, the mean is given with its 90 percent confidence interval. The strong dependence of both mean and standard deviation on the trunk length is again clearly exhibited. The discontinuity in loss design rules around 1500 miles is seen to result in a noticeable extra increase in the average noise level, since the rise of 3.1 dB from 34.4 to 37.5 dBrnC occurs for an increase in trunk length that is approximately half of what corresponds to a double length. The noise distributions in each of the mileage categories are all close to normal with only a small tendency toward positive skewness in some of the categories.

Compandored carrier facilities are used extensively on short-haul trunks as shown in Section III. These facilities have the property that the noise level during quiet intervals is lower than the noise level during periods when speech or some other signal is transmitted over the facility. Tests have shown that the subjective reaction to this noise behavior is approximately accounted for by adding 5 dB to the noise level measured in a quiet interval. The resulting noise level is commonly referred to as an "effective" noise level. The distributions

TABLE XIV — 3A NOISE LEVEL AT RECEIVE SWITCH

| Trunk length (miles) | Noise level | | | |
|---|---|---|---|---|
| | Measured | | Effective* | |
| | Mean (dBrnC) | Standard deviation (dB) | Mean (dBrnC) | Standard deviation (dB) |
| 0  –  62.5 | 18.1 ± 1.7 | 7.5 | 21.6 | 8.6 |
| 62.5– 125 | 22.0 ± 1.3 | 6.2 | 25.2 | 5.8 |
| 125  – 250 | 26.4 ± 1.0 | 5.3 | 28.0 | 4.8 |
| 250  – 500 | 30.6 ± 0.6 | 4.4 | 30.9 | 4.2 |
| 500  –1000 | 33.1 ± 0.5 | 2.9 | | |
| 1000  –2000 | 35.1 ± 0.4 | 3.0 | | |
| 2000  –4000 | 39.4 ± 0.4 | 2.6 | | |
| 1000  –1465 | 34.4 ± 0.4 | 2.8 | | |
| 1465  –2000 | 37.5 ± 0.7 | 2.5 | | |

* Including subjective compandor penalty.

of effective noise levels for various length categories of trunks are obviously influenced by the proportion of trunks in the category that contain compandored carrier.

The fourth and fifth columns of Table XIV give means and standard deviations of distributions of effective noise levels in those mileage categories where there is a noticeable difference between the average effective noise level and the average measured noise level. These results have been derived by adding 5 dB to the measured noise level on each trunk composed entirely of compandored facilities, and adding a correspondingly lower value to the measured noise level on the trunks made up of a tandem connection of compandored carrier facilities and noncompandored facilities. Comparison of the second and fourth columns shows that the average difference between effective and measured noise levels ranges from 3.5 dB for trunks shorter than 62.5 miles to 0.3 dB for trunks from 250 to 500 miles long.

The noise performance of the most important transmission facilities used in the intertoll trunk plant was estimated from the survey data by subclass analysis. The results are summarized in Table XV. Notice that noise distribution estimates are not given for all mileage categories of each transmission facility. The short-haul carrier facilities N1 and ON are restricted to the first three mileage categories by their capabilities (compare with Table VI). Results for these as well as for other facilities are presented only in those cases where the sample

TABLE XV — MEASURED 3A NOISE LEVEL REFERRED TO 0 TLP

| Trunk length (miles) | Voice frequency facility | | N1 carrier | | ON-carrier | |
|---|---|---|---|---|---|---|
| | Mean (dBrnCO) | Standard deviation (dB) | Mean (dBrnCO) | Standard deviation (dB) | Mean (dBrnCO) | Standard deviation (dB) |
| 0–62.5 | 16.6 ± 3.1 | 8.6 | 25.1 ± 2.0 | 5.7 | 18.9 ± 1.5 | 4.6 |
| 62.5–125 | · · · | · · · | 29.0 ± 1.9 | 4.3 | 21.8 ± 2.0 | 5.5 |
| 125–250 | · · · | · · · | 29.3 ± 1.5 | 3.0 | 23.0 ± 1.5 | 4.0 |

| Trunk length (miles) | Coaxial cable carrier | | Microwave radio carrier | |
|---|---|---|---|---|
| | Mean (dBrnCO) | Standard deviation (dB) | Mean (dBrnCO) | Standard deviation (dB) |
| 62.5–125 | · · · | · · · | 29.8 ± 1.9 | 4.1 |
| 125–250 | 31.1 ± 1.9 | 3.0 | 31.2 ± 2.1 | 4.7 |
| 250–500 | 34.5 ± 1.4 | 4.0 | 34.0 ± 0.9 | 3.9 |
| 500–1000 | 36.0 ± 1.1 | 2.9 | 37.2 ± 0.6 | 2.5 |
| 1000–2000 | · · · | · · · | 39.1 ± 0.7 | 2.4 |
| 2000–4000 | · · · | · · · | 42.8 ± 0.6 | 2.3 |

contained at least 10 trunks. The noise levels have here been referred to a conventional reference point called the zero transmission level point (0 tlp). Intertoll trunks operate with the transmitting switch at a transmission level of −2 dB relative to 0 tlp, and with the receiving switch at a transmission level of − (2+design loss) dB. The mean noise level at 0 tlp was computed by adding 2 plus the mean measured loss to the mean measured noise level. The variance was calculated by subtracting the variance of measured losses from the variance of measured noise levels. In this way, it is seen that the noise level computed at 0 tlp does not include effects resulting from the loss variability of trunks. Thus the random variables representing noise level at 0 tlp and measured loss can be regarded as independent.

Table XV shows that voice frequency facilities have a lower average noise level than the compandored short-haul carrier systems N1 and ON, while the standard deviation is higher. The average noise level on N1-carrier, which is a double-sideband system, is more than 6 dB higher than the average noise level on the single-sideband ON-carrier system.

Microwave radio facilities are used for a wide range of trunk mileages. The trend with increasing mean and decreasing standard deviation as trunk length is increased is clearly visible from the table. This trend is in line with the results already mentioned for 3A noise levels on trunks. The trend can be explained theoretically by regarding a transmission facility as a tandem connection of a number of noise sources $n$, with $n$ being directly proportional to the facility

TABLE XVI — MEASURED FLAT-WEIGHTED 3A NOISE LEVEL AT RECEIVE SWITCH

| Trunk Length (Miles) | Noise level | |
|---|---|---|
| | Mean (dBrn flat) | Standard deviation (dB) |
| 0–62.5 | 29.4 ± 2.0 | 9.0 |
| 62.5–125 | 30.2 ± 1.4 | 7.3 |
| 125–250 | 31.9 ± 0.8 | 4.9 |
| 250–500 | 34.9 ± 0.7 | 4.8 |
| 500–1000 | 36.3 ± 0.5 | 3.3 |
| 1000–2000 | 37.7 ± 0.6 | 2.9 |
| 2000–4000 | 41.3 ± 0.5 | 2.4 |
| 1000–1465 | 37.2 ± 0.7 | 2.7 |
| 1465–2000 | 39.6 ± 0.9 | 2.9 |

length. The resulting noise level on the facility is then the power sum of $n$ components. Recent work by Marlow[10] and Nåsell[11] has shown that the mean of a power sum increases while the standard deviation decreases as the number of component variables $n$ is increased.

Table XVI summarizes the results of 3A noise readings with 3-kHz flat weighting. Trends similar to those for 3A noise levels with C-message weighting are indicated; the mean increases with trunk length while the standard deviation decreases. Noise level readings taken with a 3-kHz flat weighting network are mainly used to indicate the presence of low-frequency noise components on a measured channel. Such noise levels do not in general propagate along telephone trunks since their main frequencies fall below the lower cut-off frequency of most telephone channels.

VII. IMPULSE NOISE

One of the major sources of impairment to the successful transmission of digital data over telephone circuits is the appearance of short-term, high-level peaks of noise 12 to 50 dB above the background noise. These peaks are called "impulse noise" because they frequently resemble the impulse response of a bandpass filter when viewed on an oscilloscope.

Impulse noise and its interfering effects on various voice-band data signals have been studied extensively in recent years by Fennick and others.[12, 13, 14] It has been characterized in terms of peak amplitude distributions, burst durations, frequency spectra, distributions of the intervals between impulses and conditional probability of receiving a second impulse within "t" units of time from an initial impulse. Among these, the distribution of peak amplitudes above selected threshold levels provides an adequate and useful description of the impulse noise process with instrumentation of minimum complexity.[15] It is not the object of this section, therefore, to expand or improve upon the methods of characterizing impulse noise now used, but rather to use one of these methods to describe the impulse noise performance as measured during the 1964 intertoll trunk survey and to identify some factors that significantly influence its character.

7.1 *Impulse Noise Description*

The impulse noise on any one transmission channel can be described by the time-average of the number of noise bursts that exceed a threshold level as a function of threshold level. This functional rela-

tion is referred to as the peak amplitude distribution. Experience has shown that the relation between the logarithm of the average counting rate and the threshold level is in many cases approximately linear, that is,

$$L = L_1 - K \log_{10} C. \tag{1}$$

Here, $L$ is the threshold level in dBrn at which the average counting rate is $C$ counts per minute, as observed on a 6A impulse noise counter,[15] $L_1$ is the impulse noise level that corresponds to an average counting rate of one count per minute, and $K$ is the slope of the peak amplitude distribution. It expresses the dB decrease in threshold level that increases the average counting rate by a factor of ten.

## 7.2 *Measurement Procedure*

During the field measurement phase of the survey, a 15 minute magnetic tape recording was made of the trunk noise for each inter-toll trunk in the sample. Trunks containing no compandored carrier facilities in their makeup were terminated at the distant end either in a dialed quiet termination or, in the case of one-way incoming trunks, in a termination supplied manually at the transmitting end of the trunk.

If a trunk contained any compandored carrier, a low-frequency tone (325-350 Hz) was transmitted at −13 dBm from the distant toll testboard in order to operate the compandors with a simulated, fixed-power data signal. This tone was removed by filtering before the noise was recorded. To ensure that all impulse noise data were collected during a period of peak channel loading and switching activity at both ends of the trunk, it was required that noise be recorded only when the local time at both ends of the trunk fell between 9:15 and 11:45 a.m. or 1:15 and 4:15 p.m.

After the field measurement phase of the survey was completed, each tape recording was played back in the laboratory and monitored simultaneously by eight 6A impulse counters with threshold levels spaced 3 dB apart. Each 6A counter was equipped with a voiceband weighting network.[15] The playback gain was adjusted so that the 6A counts covered the range from 4 to 45 counts during the 15 minute observation period for each sample trunk. Using relation (1), the levels corresponding to 45, 15, and 4.5 counts, respectively, were determined by interpolation between the levels that in each case gave a count higher and lower than the count for which the corresponding
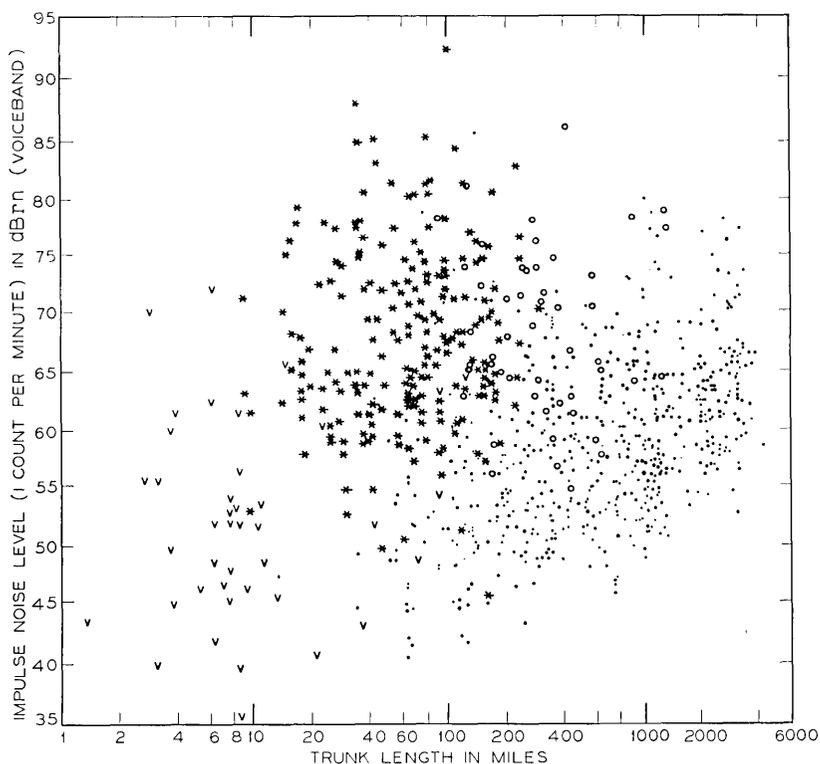
Fig. 6 — Scatter diagram of impulse noise level versus trunk length. v = voice-frequency trunks; * = trunks containing only compandored carrier; o = trunks containing both compandored and noncompandored carrier; · = trunks containing only noncompandored carrier.

level was sought. The slope of the peak amplitude distribution was computed for each trunk as the difference between the interpolated 4.5 and 45 count levels.

## 7.3 Results of Impulse Noise Measurements

The 1963 survey of impulse noise on Bell System carrier facilities revealed that mean threshold levels corresponding to a 6A counting rate of 3 counts per minute were significantly higher on the short haul compandored carrier facilities, N1 and ON, than on long haul carrier facilities.[16] In recognition of this, the scatter diagram of impulse noise levels versus trunk length, shown in Fig. 6, reflects the following partition of intertoll trunks:

(*i*) Trunks made up entirely of compandored carrier facilities (N1, ON, or O),

(*ii*) Trunks made up of any combination of compandored carrier with noncompandored carrier facilities (part N1, ON or O),

(*iii*) Trunks made up entirely of noncompandored carrier facilities (C, J, K, L and microwave radio), and

(*iv*) Trunks made up entirely of voice frequency facilities.

Trunks in the first three categories that include a short section of voice frequency facilities in tandem with carrier facilities were classified by their carrier facility makeup.

Fig. 6 suggests the following trends:

(*i*) Impulse noise levels measured on trunks containing *any* compandored carrier are distinctly higher than on noncompandored trunks.

(*ii*) Impulse noise levels measured on mixed facility trunks containing any compandored carrier are dominated by the compandored carrier impulse noise.

(*iii*) Impulse noise levels measured on noncompandored trunks are correlated with trunk length.

These trends are described quantitatively by the results shown in the figures and tables that follow. Tables XVII and XVIII summarize estimates of the mean, standard deviation, and 90 percent confidence interval for the mean for the impulse noise level (corresponding to an average of 1 count per minute) and the slope. All noise measurements were made on trunks and therefore are referred to the level of the receive switch.

In Table XVII estimates are presented for  all intertoll trunks and for nine subclasses defined by transmission facility. These subclasses include trunks made up entirely of the five major Bell System carrier facilities, voice frequency facilities, and the common long haul combination of coaxial cable carrier and microwave radio. The remaining categories partition all carrier intertoll trunks whether single facility or mixed, into those using only noncompandored carrier facilities and those using any compandored carrier facilities. Table XVIII summarizes estimates for the latter two trunk categories within seven mileage categories. Voice frequency trunks are eliminated from Table XVIII because 93 percent of them are shorter than 62.5 miles; impulse noise on these trunks is therefore adequately characterized by Table XVII.

TABLE XVII — SUMMARY OF IMPULSE NOISE RESULTS, AT RECEIVE SWITCH, OVER ALL TRUNK LENGTHS

| Facility type | Impulse noise level (1 count per minute) | | Slope | |
|---|---|---|---|---|
| | Mean (dBrn VB) | Standard deviation (dB) | Mean (dB per decade) | Standard deviation (dB) |
| Voice frequency | 51.4 ± 2.5 | 8.4 | 8.2 ± 0.8 | 3.4 |
| N1 carrier | 68.7 ± 1.8 | 7.9 | 6.9 ± 1.2 | 4.3 |
| ON-carrier | 66.7 ± 1.8 | 7.3 | 5.5 ± 0.8 | 3.2 |
| K-carrier | 58.3 ± 2.5 | 9.8 | 9.2 ± 1.3 | 6.3 |
| L-carrier | 54.9 ± 2.2 | 7.6 | 6.6 ± 0.7 | 3.5 |
| Microwave radio | 57.3 ± 1.4 | 7.0 | 7.9 ± 0.7 | 4.1 |
| L-carrier and microwave radio | 59.9 ± 1.7 | 6.3 | 7.3 ± 0.5 | 3.7 |
| Any compandored carrier | 67.8 ± 1.3 | 7.6 | 6.2 ± 0.7 | 3.9 |
| Noncompandored carrier | 57.8 ± 1.0 | 7.6 | 7.7 ± 0.4 | 4.1 |
| All intertoll trunks | 61.6 ± 1.2 | 9.7 | 7.1 ± 0.4 | 4.0 |

TABLE XVIII — SUMMARY OF IMPULSE NOISE RESULTS, AT RECEIVE SWITCH, FOR CARRIER INTERTOLL TRUNKS

| Mileage stratum | Compandored carrier | Impulse noise level (1 count per minute) | | Slope | |
|---|---|---|---|---|---|
| | | Mean (dBrn VB) | Standard deviation (dB) | Mean (dB per decade) | Standard deviation (dB) |
| 0–62.5 | ANY | 67.1 ± 1.8 | 8.0 | 6.3 ± 1.1 | 4.3 |
| | NONE | 52.5 ± 2.8 | 5.5 | 7.2 ± 1.8 | 3.1 |
| 62.5–125 | ANY | 69.4 ± 1.8 | 7.2 | 6.3 ± 0.9 | 3.4 |
| | NONE | 55.1 ± 3.3 | 9.2 | 9.2 ± 1.3 | 5.0 |
| 125–250 | ANY | 67.0 ± 1.5 | 6.3 | 6.0 ± 0.9 | 3.4 |
| | NONE | 56.7 ± 1.6 | 7.6 | 8.2 ± 0.8 | 4.6 |
| 250–500 | ANY | 67.8 ± 2.4 | 6.6 | 5.0 ± 0.5 | 1.9 |
| | NONE | 58.4 ± 1.6 | 7.3 | 7.7 ± 0.6 | 4.0 |
| 500–1000 | ANY | 66.4 ± 3.6 | 5.9 | 4.3 ± 1.9 | 2.6 |
| | NONE | 58.8 ± 1.5 | 6.4 | 6.9 ± 0.6 | 3.6 |
| 1000–2000 | NONE | 59.9 ± 1.1 | 6.1 | 7.2 ± 0.5 | 3.5 |
| 2000–4000 | NONE | 63.2 ± 1.3 | 5.0 | 7.1 ± 0.3 | 2.8 |

Table XVII supports the observation made from the scatter plot: the average impulse noise level is 10 dB higher on trunks containing any compandored carrier than on trunks using no compandored carrier. On the other hand, the average slope is 1.5 dB higher for noncompandored carrier trunks. The differences between mean impulse noise levels observed for the three facility categories L-carrier, microwave radio, and the combination of L-carrier and microwave radio is not a reflection of different impulse noise performance of these facilities. If a comparison between the three is made within mileage categories, one finds no significant differences. The differences observed in Table XVII depend on the varying length distribution of the three facilities combined with the fact that the average impulse noise level increases with trunk length for trunks on noncompandored carrier facilities.

Some of the results of Table XVII are depicted graphically in Fig. 7. The intertoll trunks are partitioned into three major facility categories, voice frequency trunks, trunks made up of noncompandored carrier facilities, and trunks containing any compandored carrier. For each facility category, a peak amplitude distribution is given by
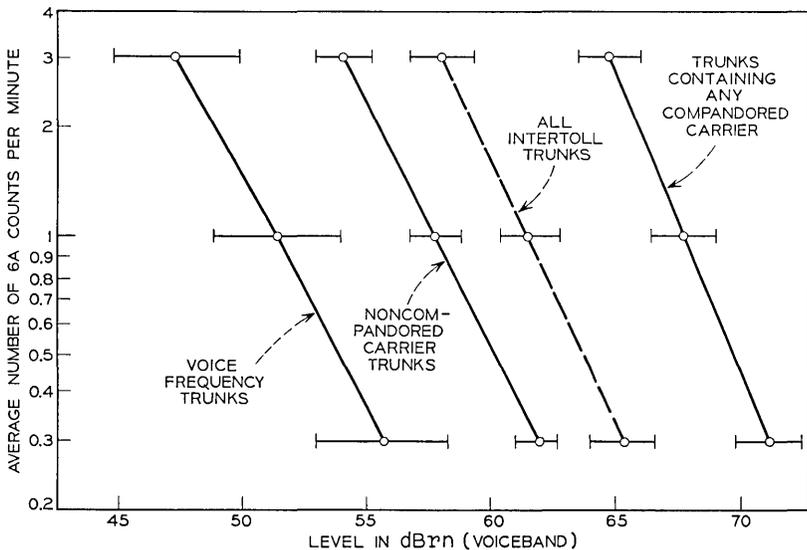
Fig. 7 — Average impulse noise levels at receive switch with 90 percent confidence intervals.
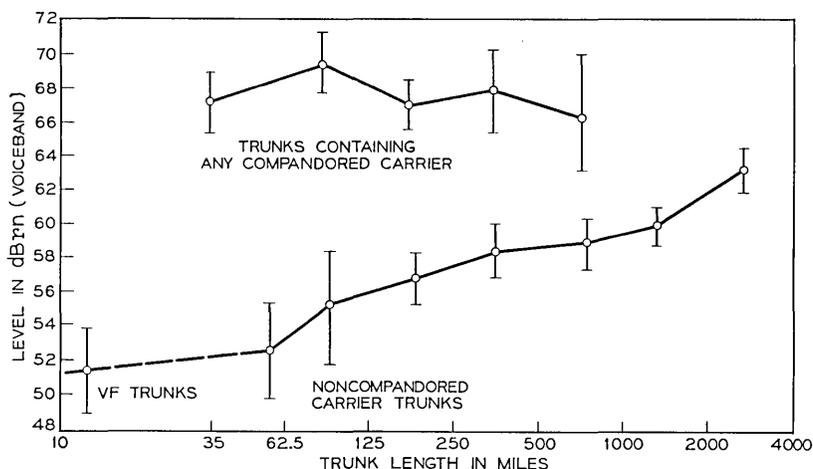
Fig. 8 — Average impulse noise level (1 count per minute) at receive switch with 90 percent confidence intervals.

the curve that connects the mean impulse noise levels at which an average of 3, 1, and 0.3 counts occur per minute.

Table XVIII demonstrates how the average impulse noise level on noncompandored carrier trunks increases with increasing trunk length, while the average impulse noise level on trunks containing any compandored carrier shows no significant change. The table also indicates a tendency for the slope to decrease with trunk length within each of these two facility categories. Furthermore, standard deviations of both impulse noise level and slope tend to decrease with trunk length for each of the two facility categories.

Fig. 8 clearly portrays the relationship between average impulse noise level and trunk length. Abscissa values for points plotted in this figure are the mean trunk lengths within each mileage-facility category.

For each of the categories given in Tables XVII and XVIII the distribution function was estimated for both the impulse noise level and the slope. Distributions of the impulse noise levels in all these categories are very nearly normal and therefore are adequately defined by the distribution parameters given for them. The distributions of the slope show a positive skewness, however, and so Fig. 9 is included to show the amount of this skewness for the two major trunk facility categories referred to throughout this section.

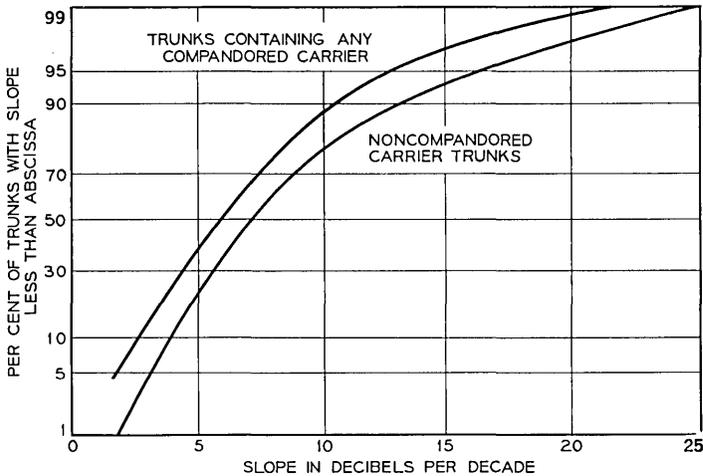Estimation of the slope of peak amplitude distributions is mean-

Fig. 9 — Distribution functions of slope of peak amplitude distributions.

ingful only to the extent that relation (1) is a good approximation to the relation between counting rate and threshold level. Fig. 7 indicates a high degree of linearity in the functional dependence between the logarithm of the average counting rate and the average threshold level in dBrn. However, this does not guarantee linearity of individual peak amplitude distributions. Some insight into the latter can be gained by studying the difference, $\Delta$, between the level at which 15 counts occur in 15 minutes and that level at which 15 counts in 15 minutes would be predicted on the basis of the levels at which 45 and 4.5 counts occur in 15 minutes and using the linearity assumption in (1).

The 90 percent confidence interval for $\Delta$ includes zero for all subclasses of trunks listed in Table XVII, except K carrier. Excluding this category, the mean $\Delta$ values range from 0.3 to 0.7 dB, and the standard deviations of $\Delta$ range from 0.8 dB for L carrier to 1.5 dB for microwave radio. The mean $\Delta$ for K carrier trunks is 1.2 dB with a 90 percent confidence interval of $\pm 0.5$ dB. This indicates a nonlinear peak amplitude distribution for trunks made up entirely of K carrier. The standard deviation of $\Delta$ is 2.7 dB for these trunks.

Fig. 8 demonstrates that the impulse noise level for noncompandored carrier trunks is correlated with length, but Fig. 5 shows that background noise level is also correlated with length. In an effort to assess the relative significance of these two factors on impulse noise, a weighted multiple linear regression analysis was performed on the

impulse noise level corresponding to 3 counts per minute for noncompandored trunks versus the logarithm of trunk length in miles and the background noise level as measured with the 3A noise measuring set. The least squares fit was weighted in inverse proportion to the variance of impulse noise levels about the regression plane.

This regression analysis gave the following result:

$$L_3 = 33.2 + 0.08 \log_2 l + 0.65 N, \tag{2}$$

where $L_3$ is the average impulse noise level in dBrn (VB) corresponding to an average 6A counting rate of 3 counts per minute, $l$ is the trunk length in miles, and $N$ is the background noise level in dBrnC. As with all results reported in this section, both $L_3$ and $N$ refer to noise levels measured at the receive switch of an intertoll trunk. Relation (2) shows that the correlation between impulse noise level and trunk length is very low if the background noise level measurement is included as an independent variable in the regression analysis.

A weighted linear regression analysis of the impulse noise levels as a function of trunk length alone gave the result:

$$L_3 = 37.9 + 1.9 \log_2 l. \tag{3}$$

Thus the average impulse noise level $L_3$ on noncompandored carrier trunks increases by 1.9 dB for each doubling of the trunk length. Such a strong correlation between these two variables was to be expected because of (2) and the dependence of average background noise level on trunk length discussed in Section VI.

## VIII. RELATIVE ENVELOPE DELAY

Relative envelope delay can be used to characterize nonlinearity of the phase characteristic. It is the delay of the envelope of an amplitude modulated carrier relative to the envelope delay at a reference carrier frequency. As such it provides an approximation to the derivative of the phase characteristic. Hence linear phase corresponds to constant delay, or zero relative envelope delay. The reference frequency chosen for all measurements in the survey was 1800 Hz. That is, the envelope delay of each trunk in the sample is given relative to the envelope delay of the same trunk at 1800 Hz.

### 8.1 *Measurement Procedure*

Envelope delay was measured with a loop-around technique. Whenever possible a trunk was looped back onto itself at a 4-wire point

in the distant office, thus ensuring the same facility composition in each direction of transmission. Delay was then measured from a 4-wire point in the near-end office. Since voice frequency patch bays are standard 4-wire points for most carrier systems, they were used both for point of measurement and point of loop back when available. For facilities not appearing at a voice frequency patch bay, an alternate point of measurement and loop-back was selected. In offices where compandored carrier systems terminated only at the circuit patch bay this point was used; and for voice frequency facilities, the repeater jacks were used. For 2-wire voice facilities the trunk to be measured was looped back onto its cable quad pair, thus assuring similarity of loading, repeater type, and spacing. The loop-around measurements were converted arithmetically to one-way values in a manner described later.

In order to characterize each trunk as completely as possible, measurements were also made on the office trunking equipment between the toll test board and the voice frequency patch bay (or similar points where the facility measurements had been made). Measurements between the toll testboard and the voice frequency patch bay in 4-wire switching offices were made from the test board with a loop-back at the voice frequency patch bay. In offices with 2-wire switching, two separate measurements were made on the office equipment: one from the toll test board to the voice frequency patch bay and a second from the voice frequency patch bay to the toll test board.

## 8.2 *Loop-Around to One-Way Conversion*

As already mentioned, relative envelope delay on facilities was measured with a loop-back technique. The requirement to characterize the transmission performance by one-way data, therefore, makes it necessary to find a method for converting the loop-around data to one-way delay. This conversion presents a problem, since a direct division by two of each two-way delay reading would lead to an underestimation of the standard deviation of the distribution of one-way delay readings, even though the mean of the same distribution would be correctly estimated.

The classification of trunks into homogeneous facility categories is of importance here, since it allows us to view the delay in the two directions of transmission of the trunk as *independent,* identically distributed random variables. Within each such facility category, the

following conversion formula was used at each frequency:

$$d_1 = \frac{\bar{d}_2}{2} + \frac{d_2 - \bar{d}_2}{(2)^{\frac{1}{2}}} \tag{4}$$

where $d_1$ is the desired one-way delay for a given sample trunk, $d_2$ is the measured loop-around delay for the same trunk and $\bar{d}_2$ is the estimated mean loop-around delay for all trunks in the category.

The use of a ratio estimator in the sample survey estimation formulas means that $\bar{d}_2$ is not an unbiased estimator of the mean loop-around delay. However, this estimator is asymptotically unbiased. Therefore, if the sample size within the facility category is large enough, the bias of $\bar{d}_2$ can be neglected, and its variance can be neglected in comparison with the variance of $d_2$. The covariance of $d_2$ and $\bar{d}_2$ is likewise negligible. Under these conditions it is easily shown that:

$$E(d_1) = \tfrac{1}{2}E(d_2) \tag{5}$$

and

$$\text{Var } (d_1) = \tfrac{1}{2} \text{ Var } (d_2). \tag{6}$$

The conversion formula (4) is thus seen to have the desired properties for a large enough sample size in the corresponding trunk category. It has, however, been used for conversion from two-way to one-way delay in all of the facility categories regardless of size. The bias introduced in this way is certainly less serious than the errors that would occur through a simple division of each measured loop-around delay by two.

### 8.3 Delay on Facilities

The data were analyzed separately for each of the facility categories. The results include estimates of the mean, standard deviation, and 90 percent confidence intervals for the means. These results for facilities with one pair of A-type channel banks, one pair of O-carrier channel filters, one pair of N1-carrier channel filters are presented in Table XIX and in Fig. 10. The curves for facilities with A-channel banks and O-carrier channel filters are quite symmetric while the curve for N1-carrier channel filters is not. Measurements at high frequencies could not always be made on the compandored carrier systems because of excessive attenuation, as mentioned previously.

Delay characteristics are additive. That is, the relative envelope delay for two or more facilities in tandem may be obtained by adding the individual delay curves for each of the component facilities. For

TABLE XIX — RELATIVE ENVELOPE DELAY FOR MAJOR
FACILITY CATEGORIES

| Frequency (Hz) | Delay (µs) | | | | | |
|---|---|---|---|---|---|---|
| | A Channel banks | | O Channel Filters | | N1 Channel Filters | |
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 400 | 1096 ± 29 | 213 | 1545 ± 30 | 131 | 513 ± 11 | 50 |
| 600 | 538 ± 11 | 79 | 671 ± 9 | 61 | 143 ± 4 | 22 |
| 800 | 315 ± 6 | 48 | 357 ± 7 | 43 | 36 ± 5 | 18 |
| 1000 | 180 ± 3 | 27 | 212 ± 5 | 33 | −20 ± 4 | 14 |
| 1200 | 104 ± 3 | 25 | 118 ± 5 | 30 | −26 ± 4 | 14 |
| 1400 | 57 ± 2 | 19 | 54 ± 3 | 21 | −26 ± 3 | 11 |
| 1600 | 21 ± 1 | 12 | 15 ± 2 | 11 | −17 ± 2 | 9 |
| 1700 | 8 ± 1 | 9 | 5 ± 1 | 7 | −9 ± 2 | 6 |
| 1800 | 0 | | 0 | | 0 | |
| 2000 | 2 ± 2 | 15 | 5 ± 2 | 12 | 27 ± 2 | 8 |
| 2200 | 29 ± 4 | 29 | 31 ± 3 | 20 | 79 ± 2 | 10 |
| 2300 | 51 ± 4 | 33 | 53 ± 4 | 25 | 117 ± 4 | 14 |
| 2400 | 77 ± 5 | 37 | 81 ± 4 | 31 | 162 ± 4 | 17 |
| 2600 | 147 ± 6 | 44 | 161 ± 6 | 44 | 252 ± 7 | 23 |
| 2800 | 257 ± 7 | 51 | 293 ± 10 | 63 | 393 ± 9 | 31 |
| 3000 | 444 ± 8 | 61 | 537 ± 15 | 89 | 750 ± 23 | 84 |
| 3200 | 837 ± 10 | 79 | 936 ± 33 | 154 | ... | ... |
| 3300 | 1279 ± 14 | 109 | ... | ... | ... | ... |
| 3400 | 1837 ± 29 | 186 | ... | ... | ... | ... |
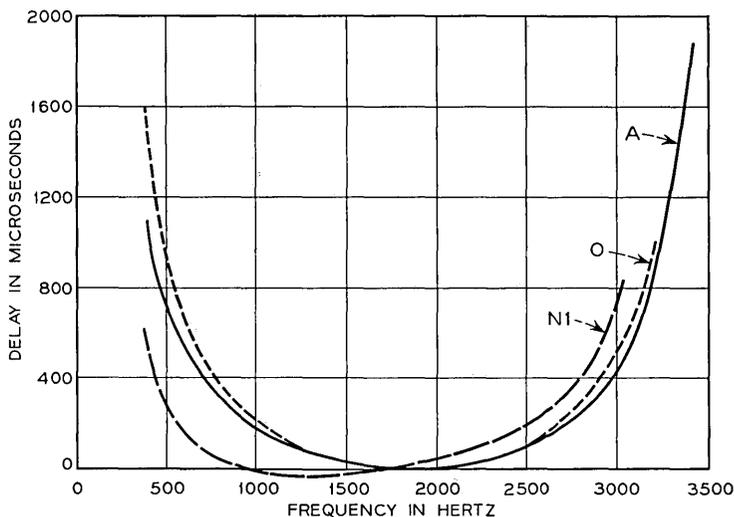


Fig. 10 — Mean relative envelope delay for facilities with A-type channel banks, O-carrier channel filters, and N1-carrier channel filters.

example, the delay characteristic of a trunk facility containing one pair of A-channel banks in tandem with an ON system may be obtained by adding the data for a facility with one pair of A-channel banks and that for a facility with one pair of O-carrier terminals. This follows from the fact that the voiceband channel filters are, with two notable exceptions dealt with later, the only significant contributors to envelope delay on carrier facilities.

Variance for a tandem facility category is similarly estimated by the sum of the variances. Precision in the form of 90 percent confidence intervals for the mean may also be estimated by "addition" of confidence intervals in the same manner as standard deviations (square root of sum of squares). Estimates of the cumulative distribution functions have indicated that the assumption of normality is justified within facility categories. Hence the percentage points of the distribution functions may be estimated from the means and standard deviations.

The property of additivity was directly used in the data analysis. That is, delay for facilities with two pair of A-channel banks was assumed to be the sum of that for two facilities, each with one pair of A-channel banks. Similarly, delay for facilities with three pair of A-channel banks was regarded as the sum of three separate A-channel bank facilities. The data for facilities with one, two, or three pairs of A-channel banks were then pooled to provide the estimates for a facility with one pair of A-channel banks given in Table XIX and Fig. 10. This procedure has the advantage of producing greater precision than use of only data for facilities with one pair of A-channel banks.

In addition to channel bank filters, other factors which may contribute to the over-all delay on a facility are group connectors and K-carrier modems. The most noticeable effects of these are found on edge channels of the basic group (channels 1 and 12). Estimates of the additional delay contributed by each of these are given in Table XX. The results for group connectors are also shown in Fig. 11 which presents the delay curves for A-type channel banks with and without the effects of group connectors on edge channels.

The delay contributed by group connectors reflects the specific combination of an older and a newer generation of such connectors that existed in the plant in 1964. The newer generation gives smaller additional delay than the older one. It is therefore expected that the additional delay resulting from group connectors will decrease as the proportion of new group connectors in the plant increases. Table XX shows that the low frequency effect of a K-carrier modem is similar

TABLE XX—ADDITIONAL MEAN RELATIVE ENVELOPE DELAY ON
EDGE CHANNELS FROM K-CARRIER MODEMS AND GROUP
CONNECTORS

| Frequency (Hz) | Delay (μs) | | | |
| --- | --- | --- | --- | --- |
| | K-Carrier modem | | Group connector | |
| | Channel 1 | Channel 12 | Channel 1 | Channel 12 |
| 400 | 517 | 842 | 455 | −138 |
| 600 | 334 | 253 | 314 | −121 |
| 800 | 231 | 114 | 233 | −109 |
| 1000 | 164 | 73 | 154 | −90 |
| 1200 | 117 | 32 | 103 | −67 |
| 1400 | 76 | 5 | 65 | −47 |
| 1600 | 37 | 0 | 32 | −27 |
| 1700 | 20 | −2 | 16 | −13 |
| 1800 | 0 | 0 | 0 | 0 |
| 2000 | −34 | −3 | −27 | 39 |
| 2200 | −63 | 0 | −50 | 90 |
| 2300 | −80 | −8 | −63 | 122 |
| 2400 | −93 | −17 | −68 | 158 |
| 2600 | −120 | −28 | −84 | 244 |
| 2800 | −147 | −27 | −99 | 360 |
| 3000 | −170 | −38 | −115 | 533 |
| 3200 | −238 | −121 | −127 | 851 |
| 3300 | −296 | −59 | −133 | 1058 |
| 3400 | −293 | 217 | −141 | 1349 |

to that of a group connector on channel 1. In fact, a K-carrier modem
contributes more excess delay at 400 Hz than one group connector.
This is seen to be true for both channels 1 and 12 associated with
the K-carrier modem. We may also notice that on channel 12, un-
like channel 1, a K-carrier modem will also add delay to the highest
frequencies of the voice band. In Fig. 11 we observe that for channel
12, the effect of a group connector is to produce a shift of the delay
curve toward the left and an attendant asymmetry. This is seen to
be opposite to the effect on channel 1 and also somewhat greater in
magnitude. For a K-carrier modem, the effect on both channels 1
and 12 will be similar to that of a group connector on channel 1;
that is, a shift of the entire delay curve to the right.

8.4 *Office Equipment Delay*

   The results for office equipment are presented in Tables XXI and
XXII for 4-wire and 2-wire offices, respectively. The left part of Ta-
ble XXI shows the results of loop-around measurements, with the

measurements being made at the toll test board (ttb) and the loop-back at the voice frequency patch bay (vfpb). The right part of this table indicates that the relative envelope delay is lower if the loop-back is made at the circuit patch bay (cpb) instead of at the vfpb. The difference between these two sets of readings is attributable to the single frequency signaling units located between the cpb and the vfpb. Voice frequency facilities and older types of compandored carrier systems generally do not use in-band single frequency signaling units, so the data on the right side of Table XXI apply specifically to the delay of office equipment associated with such facilities.

The results in Table XXII for office equipment in 2-wire offices show a small difference between the delays in the two directions of transmission. Also, slightly lower values of delay are recorded where the transmission path does not contain a signaling unit. The difference is, however, not as large as in the case of 4-wire offices. This could result from the fact that the delay in 2-wire offices shows greater variability than the delay in 4-wire offices, as indicated by the estimates of standard deviation. This greater variability and a smaller sample size combine to produce less precise estimates, thus masking the contribution of the signaling units.

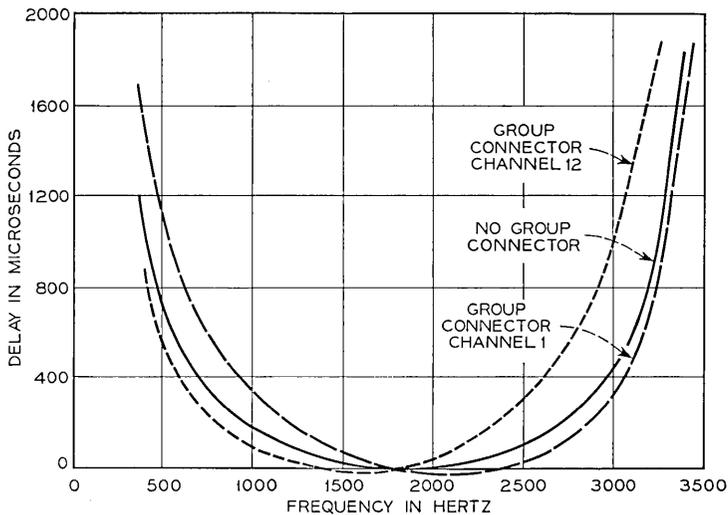The existence of 4-wire to 2-wire hybrid transformers in offices



Fig. 11 — Mean relative envelope delay on edge channels for facilities with one group connector and one pair of A-type channel banks.

TABLE XXI — RELATIVE ENVELOPE DELAY OF EQUIPMENT IN
4-WIRE OFFICES

| Frequency (Hz) | Delay (μs) | | | |
| | TTB-VFPB-TTB* | | TTB-CPB-TTB* | |
| | Mean | Standard deviation | Mean | Standard deviation |
|---|---|---|---|---|
| 400 | 307 ± 9 | 48 | 195 ± 6 | 15 |
| 600 | 122 ± 3 | 20 | 77 ± 2 | 6 |
| 800 | 62 ± 2 | 12 | 35 ± 1 | 4 |
| 1000 | 35 ± 1 | 8 | 19 ± 1 | 3 |
| 1200 | 20 ± 1 | 5 | 11 ± 1 | 4 |
| 1400 | 10 ± 1 | 5 | 6 ± 1 | 4 |
| 1600 | 5 | 4 | 2 ± 1 | 2 |
| 1700 | 2 | 2 | 1 | 1 |
| 1800 | 0 | | 0 | |
| 2000 | −4 | 2 | −2 ± 1 | 2 |
| 2200 | −7 ± 1 | 8 | −3 ± 1 | 3 |
| 2300 | −6 ± 1 | 5 | −4 ± 1 | 3 |
| 2400 | −5 ± 2 | 8 | −5 ± 1 | 3 |
| 2600 | −6 ± 1 | 6 | −6 ± 1 | 3 |
| 2800 | −6 ± 1 | 9 | −7 ± 1 | 3 |
| 3000 | −16 ± 1 | 7 | −8 ± 1 | 3 |
| 3200 | −14 ± 1 | 5 | −8 ± 1 | 3 |
| 3300 | −14 ± 1 | 5 | −8 ± 1 | 3 |
| 3400 | −15 ± 1 | 4 | −9 ± 1 | 3 |

* TTB, toll test board; VFPB, voice frequency patch bay; CPB, circuit patch bay.

with 2-wire switching is also noteworthy. Observe that the office equipment delay in 2-wire offices is approximately twice that in 4-wire offices. The use of different types of hybrid coils and wiring arrangements may also account for the higher variability encountered in 2-wire offices.

The office equipment contribution to the total relative envelope delay of intertoll trunks becomes appreciable compared with the facility contribution only at low frequencies. Tables XXI and XXII show that the average office equipment delay amounts to no more than 50 μs per office throughout the frequency range from 1000 to 3400 Hz.

8.5 *Delay on Trunks*

After conversion of the loop-around delay on facilities to one-way data the results were combined with the data on office equipment to obtain estimates of the total switch-to-switch delay on trunks. The

TABLE XXII — RELATIVE ENVELOPE DELAY OF EQUIPMENT IN 2-WIRE OFFICES

| Frequency (Hz) | Delay (µs) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | TTB-VFPB* | | VFPB-TTB | | TTB-CPB | | CPB-TTB | |
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 400 | 338 ± 24 | 67 | 294 ± 50 | 118 | 307 ± 28 | 70 | 272 ± 41 | 99 |
| 600 | 137 ± 10 | 28 | 130 ± 22 | 52 | 120 ± 10 | 25 | 108 ± 16 | 39 |
| 800 | 68 ± 6 | 16 | 67 ± 12 | 29 | 58 ± 4 | 12 | 53 ± 8 | 20 |
| 1000 | 38 ± 3 | 10 | 38 ± 8 | 18 | 33 ± 3 | 10 | 30 ± 4 | 12 |
| 1200 | 21 ± 2 | 7 | 22 ± 5 | 12 | 17 ± 1 | 4 | 16 ± 2 | 7 |
| 1400 | 12 ± 2 | 5 | 12 ± 2 | 7 | 10 ± 1 | 5 | 9 ± 1 | 4 |
| 1600 | 5 ± 1 | 3 | 5 ± 1 | 4 | 4 ± 1 | 3 | 4 ± 1 | 2 |
| 1700 | 2 ± 1 | 2 | 2 | 2 | 2 ± 1 | 2 | 1 | 2 |
| 1800 | 0 | | 0 | | 0 | | 0 | |
| 2000 | −3 | 2 | −4 ± 1 | 3 | −3 ± 1 | 2 | −3 | 2 |
| 2200 | −6 ± 1 | 3 | −6 ± 2 | 5 | −5 ± 1 | 3 | −5 ± 1 | 3 |
| 2300 | −7 ± 1 | 3 | −8 ± 3 | 6 | −6 ± 1 | 3 | −6 ± 1 | 3 |
| 2400 | −8 ± 1 | 3 | −9 ± 3 | 6 | −7 ± 1 | 3 | −7 ± 1 | 4 |
| 2600 | −10 ± 1 | 5 | −11 ± 3 | 8 | −8 ± 1 | 3 | −8 ± 1 | 4 |
| 2800 | −11 ± 1 | 4 | −12 ± 4 | 8 | −10 ± 1 | 4 | −10 ± 1 | 4 |
| 3000 | −12 ± 1 | 4 | −14 ± 5 | 9 | −11 ± 1 | 4 | −10 ± 2 | 5 |
| 3200 | −13 ± 1 | 4 | −15 ± 5 | 10 | −12 ± 1 | 4 | −11 ± 2 | 5 |
| 3300 | −14 ± 1 | 4 | −16 ± 5 | 11 | −13 ± 2 | 4 | −12 ± 2 | 6 |
| 3400 | −14 ± 1 | 5 | −16 ± 5 | 11 | −13 ± 2 | 4 | −12 ± 2 | 6 |

* TTB, toll test board; VFPB, voice frequency patch bay; CPB, circuit patch bay.

delay for the office equipment at the far end of a trunk was generally taken to be the mean delay characteristic for the type of office equipment involved. That is, for each trunk the far-end office equipment was identified by whether 2-wire or 4-wire switching was used and whether in-band single frequency signaling units were present. The appropriate mean delay curve was then taken to represent the delay of the far-end equipment. The exception to this rule occurred when both near-end and far-end offices had 4-wire switching and both used the same type of signaling (that is, both used in-band single frequency signaling or neither did). In this case, the delay as measured for both directions of transmission on the near-end office equipment of a given trunk was taken as representative of the total office equipment delay for that specific trunk. In all other cases only the receive direction of the near-end office equipment measurement was added (the mean delay curves used for far-end office equipment were for the transmit direction). When the near-end office equipment was measured on a loop-around basis and the far-end was not similar the loop-around value was divided by two to represent the receive direction of the near-end equipment. A loop-around to one-way conversion similar to that used on the facility data was not used here because of the relatively small variances encountered.

The resultant data for switch-to-switch delay were grouped into mileage categories. The mean, standard deviation, and a 90 percent confidence interval for the mean were then estimated for the trunks in each mileage category. The results of this analysis are presented in Table XXIII. The lower relative envelope delay in the shortest mileage category reflects its high percentage of N1-carrier and voice frequency cable facilities.

The data for delay at frequencies above 3200 Hz in this category refers mainly to the characteristics of voice frequency cable, since high loss generally precluded measurements on N1 and ON systems in this frequency range. A gradual increase in delay is also evident with increasing trunk length at the higher frequencies. This trend is caused by the increased use of tandem facilities and group connectors on longer trunks (compare with Tables VII and VIII). A similar increase in delay with trunk length is not observed at the lower frequencies since the transition from mainly 2-wire switching on short trunks to 4-wire switching on long trunks produces a decrease in the contribution of office equipment to overall delay (compare with Tables V, XXI, XXII).

Figure 12 compares delay characteristics for short and long inter-toll trunks. It contains plots of the median relative envelope delay curves for trunks in the first and last of the seven mileage categories. The asymmetry of the curve for short trunks reflects the previously noted facts that N1 carrier channel filters dominate in this length category and that the delay curves for that facility are not sym-metrical.

## IX. CONCLUDING REMARKS

The transmission performance of intertoll trunks has an important influence on the transmission performance of built-up connections between subscribers. Notice, however, that the latter cannot be derived in a simple manner from the former since the relation be-tween the two is influenced by a number of factors, such as customer calling habits, toll network routing patterns, and automatic alternate routing probabilities, as well as by the transmission performance of toll connecting trunks.

It should come as no surprise that important differences have been noticed between intertoll trunk performance and connection per-formance. One example illustrates this. The regression analysis re-ported in Section VI shows the mean 3A noise level with C-message weighting to increase by 3.1 dB per double length of intertoll trunk. In contrast to this, the 1966 connection survey[17] showed the average noise level on built-up connections to increase by only 2.0 dB for each doubling of the airline distance between end-offices.

Some earlier survey activities have been directed at establishing the transmission performance of selected transmission facilities.[16] The results of such surveys cannot be used to directly estimate the performance of trunks. On the other hand, the results discussed in this paper show that a survey of intertoll trunks can be used to ar-rive at performance estimates of both trunks and facilities. This is possible since any given facility constitutes a subclass of the popula-tion of trunks. The powerful technique of subclass analysis on sam-ple survey data can, therefore, be applied directly. It is seen from this that a systemwide trunk survey supplies more information than a survey of specific transmission facilities.

The dynamic growth of the toll plant should be taken into ac-count in applying the results given here. An important aspect of this growth is the introduction in recent years of the new short-haul carrier facilities N2, N3, and T1. Their transmission characteristics

TABLE XXIII—SWITCH-TO-SWITCH RELATIVE ENVELOPE DELAY OF INTERTOLL TRUNKS

| Frequency (Hz) | Delay (μs) | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | 0–62.5 miles | | 62.5–125 miles | | 125–250 miles | |
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 400 | 1177 ± 137 | 592 | 1869 ± 114 | 529 | 1877 ± 164 | 758 |
| 600 | 448 ± 65 | 274 | 815 ± 63 | 277 | 838 ± 75 | 360 |
| 800 | 199 ± 39 | 169 | 425 ± 33 | 165 | 446 ± 41 | 198 |
| 1000 | 89 ± 27 | 119 | 235 ± 18 | 110 | 253 ± 26 | 134 |
| 1200 | 37 ± 17 | 76 | 124 ± 11 | 77 | 140 ± 15 | 85 |
| 1400 | 7 ± 11 | 62 | 58 ± 7 | 46 | 71 ± 8 | 50 |
| 1600 | −4 ± 6 | 35 | 17 ± 4 | 23 | 25 ± 4 | 25 |
| 1700 | −5 ± 5 | 34 | 5 ± 2 | 12 | 8 ± 2 | 13 |
| 1800 | 0 | | 0 | | 0 | |
| 2000 | 14 ± 6 | 38 | 7 ± 5 | 28 | 7 ± 6 | 40 |
| 2200 | 51 ± 9 | 49 | 48 ± 12 | 68 | 49 ± 12 | 70 |
| 2300 | 76 ± 11 | 70 | 77 ± 16 | 89 | 85 ± 19 | 110 |
| 2400 | 110 ± 13 | 71 | 115 ± 23 | 120 | 131 ± 27 | 155 |
| 2600 | 186 ± 17 | 102 | 224 ± 47 | 240 | 224 ± 30 | 168 |
| 2800 | 305 ± 27 | 154 | 414 ± 100 | 501 | 380 ± 50 | 279 |
| 3000 | 559 ± 53 | 273 | 597 ± 51 | 362 | 624 ± 54 | 300 |
| 3200 | 712 ± 81 | 414 | 980 ± 80 | 548 | 1027 ± 76 | 461 |
| 3300 | 801 ± 151 | 648 | 1248 ± 135 | 714 | 1390 ± 116 | 640 |
| 3400 | 900 ± 230 | 861 | 1856 ± 195 | 789 | 1956 ± 182 | 961 |

| Frequency (Hz) | Delay (μs) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 250–500 miles | | 500–1000 miles | | 1000–2000 miles | | 2000–4000 miles | |
| | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation | Mean | Standard deviation |
| 400 | 2020 ± 146 | 854 | 1926 ± 108 | 783 | 1978 ± 143 | 889 | 2111 ± 156 | 958 |
| 600 | 929 ± 69 | 386 | 880 ± 50 | 359 | 938 ± 70 | 443 | 1007 ± 77 | 477 |
| 800 | 514 ± 38 | 214 | 497 ± 29 | 213 | 543 ± 44 | 285 | 579 ± 44 | 286 |
| 1000 | 292 ± 24 | 135 | 277 ± 17 | 132 | 308 ± 26 | 178 | 337 ± 27 | 177 |
| 1200 | 165 ± 15 | 87 | 154 ± 11 | 87 | 178 ± 17 | 116 | 196 ± 16 | 112 |
| 1400 | 85 ± 9 | 52 | 80 ± 7 | 55 | 97 ± 11 | 72 | 107 ± 9 | 69 |
| 1600 | 31 ± 4 | 26 | 29 ± 4 | 28 | 36 ± 4 | 34 | 37 ± 4 | 32 |
| 1700 | 10 ± 2 | 15 | 10 ± 2 | 17 | 14 ± 3 | 22 | 14 ± 2 | 17 |
| 1800 | 0 | | 0 | | 0 | | 0 | |
| 2000 | 2 ± 4 | 30 | 2 ± 3 | 33 | −2 ± 3 | 34 | 4 ± 5 | 36 |
| 2200 | 36 ± 10 | 71 | 36 ± 8 | 76 | 34 ± 7 | 75 | 53 ± 12 | 76 |
| 2300 | 70 ± 14 | 99 | 70 ± 10 | 97 | 67 ± 10 | 97 | 92 ± 16 | 102 |
| 2400 | 110 ± 19 | 136 | 110 ± 14 | 123 | 107 ± 12 | 124 | 138 ± 22 | 131 |
| 2600 | 210 ± 30 | 206 | 214 ± 24 | 201 | 212 ± 17 | 185 | 257 ± 34 | 200 |
| 2800 | 365 ± 45 | 316 | 378 ± 42 | 335 | 377 ± 26 | 271 | 445 ± 53 | 311 |
| 3000 | 619 ± 68 | 460 | 615 ± 53 | 403 | 656 ± 48 | 430 | 758 ± 83 | 463 |
| 3200 | 1066 ± 64 | 427 | 1129 ± 85 | 604 | 1246 ± 84 | 710 | 1406 ± 133 | 776 |
| 3300 | 1497 ± 94 | 531 | 1635 ± 103 | 732 | 1850 ± 111 | 926 | 2094 ± 178 | 1046 |
| 3400 | 2052 ± 119 | 632 | 2298 ± 136 | 965 | 2600 ± 169 | 1157 | 3033 ± 268 | 1458 |
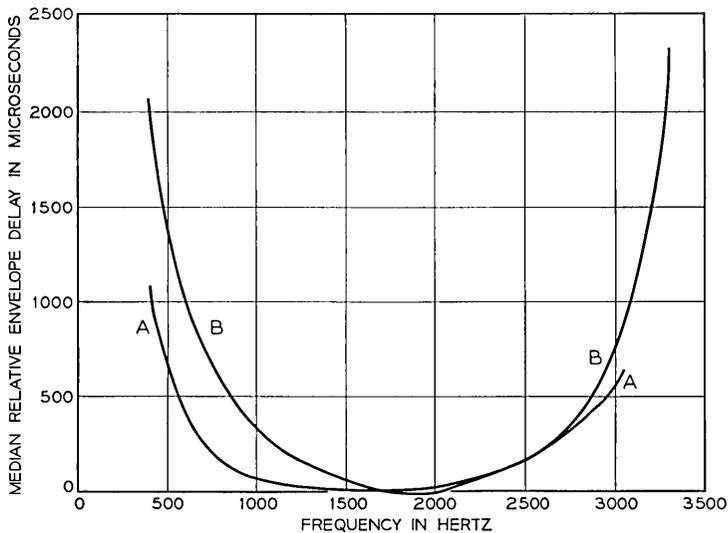
Fig. 12 — Median relative envelope delay for (a) short (0 to 62.5 miles) and (b) long (2000 to 4000 miles) intertoll trunks.

are, in practically all aspects, superior to those of the carrier facilities N1 and ON, which dominated the short-haul trunk plant in 1964. The trend in trunk performance because of this is expected to be toward improved performance. Specifically, the background noise levels on short-haul carrier trunks should decrease, and their bandwidth should increase to be more nearly comparable with the bandwidth of long-haul trunks.

## X. ACKNOWLEDGEMENTS

trunks for testing, while the measurements were made by 18 employees of Bell Laboratories. R. J. Christie, Miss M. L. Chubb, F. P. Duffy, J. H. Fennick, Mrs. C. A. Gadberry, J. E. Kessler, and J. T. Powers, Jr., all of Bell Laboratories, made important contributions in data analysis. M. Derzai (Bell Telephone Company of Canada) suggested the normalizing transformation of the ratio of trunk length to airline distance. R. C. Terreault (also Canada) estimated the effective background noise levels. The contributions of all of these people are gratefully acknowleged.

REFERENCES

1. Hansen, M. H., Hurwitz, W. N., and Madow, W. G., *Sample Survey Methods and Theory*, vols. I and II, New York: John Wiley and Sons, 1953.
2. Cochran, W. T., and Lewinski, D. A., "A New Measuring Set for Message Circuit Noise," B.S.T.J., *39*, No. 4 (July 1960), pp. 911–932.
3. Caruthers, R. S., "The Type N-1 Carrier Telephone System: Objectives and Transmission Features," B.S.T.J., *30*, No. 1 (January 1951), pp. 1–32.
4. Fracassi, R. D., and Kahl, H., "Type ON Carrier Telephone," A.I.E.E. Trans., *72*, Part I (January 1954), pp. 713–721.
5. Black, H. S., Brooks, F. A., Wier, A. J., and Wilson, I. G., "An Improved Cable Carrier System," A.I.E.E. Trans., *66* (1947), pp. 741–746.
6. Elmendorf, C. H., Ehrbar, R. D., Klie, R. H., and Grossman, A. J., "The L3 Coaxial System," A.I.E.E. Trans., *72*, Part I (September 1953), pp. 395–413.
7. Roetken, A. A., Smith, K. D., and Friis, R. W., "The TD-2 Microwave Radio Relay System," B.S.T.J., *30*, No. 4, Part II, (October 1951), pp. 1041–1077.
8. Blecher, F. H., and Hallenbeck, F. J., "The Transistorized A5 Channel Bank for Broadband Systems," B.S.T.J., *41*, No. 1 (January 1962), pp. 321–359.
9. Huntley, H. R., "Transmission Design of Intertoll Telephone Trunks," B.S.T.J., *32*, No. 5 (September 1953), pp. 1019–1036.
10. Marlow, N. A., "A Normal Limit Theorem for Power Sums of Independent Random Variables," B.S.T.J., *46*, No. 9 (November 1967), pp. 2081–2089.
11. Nåsell, I., "Some Properties of Power Sums of Truncated Random Variables," B.S.T.J., *46*, No. 9 (November 1967), pp. 2091–2110.
12. Fennick, J. H., "A Report on Some Characteristics of Impulse Noise in Telephone Communication Systems," I.E.E.E. Trans. Commun. and Elec., *83*, No. 75 (November 1964), pp. 700–705.
13. Fennick, J. H., "A Method for the Evaluation of Data Systems Subject to Large Noise Pulses," I.E.E.E. Int. Conv. Record, Part I (March 1965), pp. 106–110.
14. Kurland, M., and Molony, D. A., "Observations on the Effects of Pulse Noise in Digital Data Transmission Systems," I.E.E.E. Trans. Commun. Technology, *COM-15*, No. 4 (August 1967), pp. 552–556.
15. Favin, D. L. and Fennick, J. H., "The 6A Impulse Noise Counter and its Use," A.I.E.E. Fall Conv., October 1962.
16. Fennick, J. H., and Nåsell, I., "The 1963 Survey of Impulse Noise on Bell System Carrier Facilities," I.E.E.E. Trans. Commun. Technology, *COM-14*, No. 4 (August 1966), pp. 520–524.
17. Nåsell, I., "Some Transmission Characteristics of Bell System Toll Connections," B.S.T.J., *47*, No. 6 (July-August 1968), pp. 1001–1018.

# Effects Associated with the Thermal Response of the T1 Telephone Transmitter

## By C. A. FRITSCH

*The thermal response of the T1 transmitter, when excited by a bias current, is analytically obtained. The thermal expansions which produce a decrease in electrical resistance are described. Good agreement with experimentally measured temperature rises and displacements is demonstrated for the early time transients during which the resistance change occurs. The displacements which produce "thermal packing" are found to be a stronger function of the thermal expansion coefficient of the dome electrode than of any other part of the transmitter assembly.*

## I. INTRODUCTION

The design of an effective telephone transmitter requires that an acoustic signal (voice) be efficiently converted to an electrical output. One method of performing this function is to use a moving electrode (attached to a diaphragm) as one wall of a chamber containing granular carbon. Thus, if a dc bias current is impressed on the transmitter the resulting variations can be used to operate remote telephone apparatus.

An important factor in the design of such a transmitter is the control of thermal effects, not only caused by variations in ambient conditions but also arising from the heat generated each time the telephone set is connected to the line. These thermal effects result in dimensional changes in the transmitter body, resistance change resulting from temperature rise of the granular carbon itself, and the resistance change caused by thermal expansion of the carbon granules when they are heated by the biasing current. The total effect, which produces a loss of carbon transmitter efficiency, is referred to as "thermal packing."

To assay the relative importance of the various thermal effects it is first useful to ascertain whether the carbon granules should be

treated individually or as a continuum. It has been observed that the contact points resulting from the granule asperities cause a "bunching up" of the current.[1] Consequently, for ac heating in the kilohertz range, the high local rates of heat generation produce larger temperature changes at the points of contact with a proportionately large drop in electrical resistance.[1] However, it can be readily shown (see Appendix A) that the carbon particles are small enough and their thermal diffusivity is high enough so that any local temperature changes are virtually diffused in about one millisecond. Since the resistance changes in the Bell System T1 transmitter* have been observed to take place in about one second,[2] then the local heating caused by the asperities and the accompanying resistance change can be neglected and the granular carbon can be treated as a continuum.

To determine the effects of thermal expansion in both the microphone body and the granular carbon itself, the temperature distribution throughout the transmitter must be known. Thus, the work reported here consists of a first-approximation type of analysis to determine the temperature distribution in the carbon chamber of the T1 transmitter. The associated thermal effects are then considered with the hope of providing a better insight into what has been experimentally observed.

## II. ANALYSIS

### 2.1 Thermal Analysis of the Carbon Chamber

The telephone transmitter design of interest here is shown in Fig. 1. We wish to describe the transient thermal response of this transmitter as the above mentioned dc current is turned on. Geometrically, the carbon chamber consists of a dome (moving) electrode connected with a conical back electrode by a flexible nonconducting chamber closure. If we assume that the surfaces of constant voltage and constant temperature within the carbon aggregate are hemispherical then the walls of the carbon chamber can be considered as two concentric hemispheres (see Fig. 2).

If we also assume for the moment that the relatively heavy back electrode is held at the initial and ambient temperature $T_o$, then, the thermal response of carbon and its chamber can be conveniently

---

* Made by Western Electric Co., the manufacturing and supply unit of the Bell System, and available only to the Bell System.
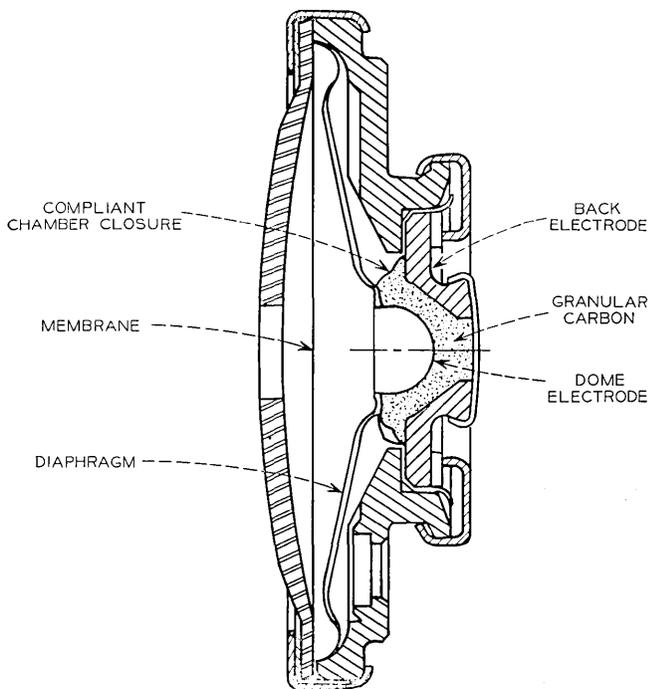
COMPLIANT
CHAMBER CLOSURE

BACK
ELECTRODE

GRANULAR
CARBON

MEMBRANE

DOME
ELECTRODE

DIAPHRAGM

Fig. 1 — Cross section of T1 transmitter.

written in terms of a temperature excess $\theta$ defined by

$$\theta = T - T_0. \tag{1}$$

If we further assume that the temperature excursions are small enough so that no significant variations in the physical properties take place then the energy-balance equation is given by

$$\rho c \frac{\partial \theta}{\partial t} = k\nabla^2\theta + q''', \tag{2}$$

where

  $k$ is the effective thermal conductivity of the carbon
  $\rho c$ is the heat capacity of the carbon
  $q'''$ is the rate of heat generation per unit volume.

Because of the angular symmetry of the boundary conditions and the heat generation, gradients in the directions of the angular coordinates can be neglected. Thus, on dividing through by the thermal
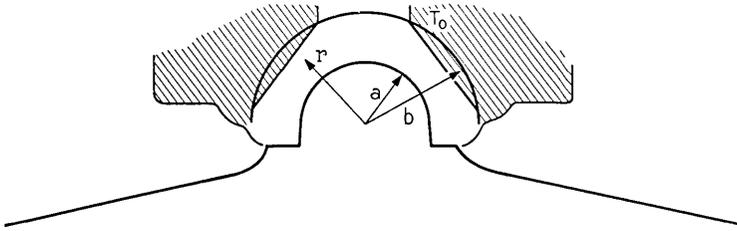
Fig. 2 — Coordinate system for carbon chamber analysis.

conductivity $k$, equation (2) in spherical coordinates becomes

$$\frac{1}{\kappa}\frac{\partial \theta}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \theta}{\partial r}\right) + \frac{q'''}{k}$$  (3)

where $\kappa$ is the thermal diffusivity.

In general, the heat generation per unit volume is given by the product of the square of the current flow per unit area times the electrical resistivity. Since here we have only radial flow of current $I$ then

$$q'''(r) = (I/2\pi r^2)^2\rho_e = i_a^2\rho_e .$$  (4)

It can be readily shown that the resistivity $\rho_e$ is related to a total measured resistance $R$ for a cavity formed by two hemispheres of inner radius, $a$, and outer radius, $b$, by

$$\rho_e = 2\pi[ab/(b - a)]R.$$  (5)

To specify the boundary conditions on our problem we recall that the initial and back electrode temperature are taken to be ambient. To approximate the heat lost by the dome electrode we specify a certain thermal resistance between the dome and some sink at ambient temperature and that resistance is represented by a coefficient $h$ in the so-called "radiation" boundary condition. Consequently, on using (4) and (5) in (3) the following boundary-value problem can be stated.

$$\frac{1}{\kappa}\frac{\partial \theta}{\partial t} = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial \theta}{\partial r}\right) + \frac{2}{r^4}\left(\frac{ab}{b - a}\right)\beta$$  (6)

$$\theta(r, 0) = 0$$  (7)

$$\theta(b, t) = 0$$  (8)

$$k\,\partial\theta/\partial r(a, t) = h\theta(a, t),$$  (9)

where

$$\beta = I^2 R / 4\pi k. \tag{10}$$

The above nonhomogeneous problem in spherical coordinates can be simplified by the following substitutions. If we define the dimensionless variables

$$\tau = \kappa t / (b - a)^2, \tag{11}$$

$$\eta = (b - r) / (b - a), \tag{12}$$

and

$$v(\eta, \tau) = \frac{r\theta(r, t)}{\beta} + \frac{ab}{b - a} (1/r), \tag{13}$$

then the problem specified by (6) through (10) becomes a homogeneous transient conduction problem in a slab with an initial temperature distribution:

$$\frac{\partial v}{\partial \tau} = \frac{\partial^2 v}{\partial \eta^2} \tag{14}$$

$$v(\eta, 0) = F/(1 - F)\left(\frac{1}{1 - (1 - F)\eta}\right) \tag{15}$$

$$v(0, \tau) = F/(1 - F) \tag{16}$$

$$\partial v / \partial \eta(1, \tau) + \frac{(1 - F)}{F} (N_{Bi} + 1)v(1, \tau) = (N_{Bi} + 2)/F, \tag{17}$$

where

$$F = a/b \tag{18a}$$

$$N_{Bi} = ha/k. \tag{18b}$$

$N_{Bi}$ is called the Biot number which characterizes the ratio of the rate at which heat is lost at the dome electrode to the rate at which heat is conducted to it through the granular carbon.

The analytical solution of the transformed problem is relatively straightforward if the dimensionless temperature $v(\eta, \tau)$ is first divided into two functions, one representing the steady-state temperature rise, the other corresponding to the transient response. Thus, we set

$$v(\eta, \tau) = \varphi(\eta) + \vartheta(\eta, \tau). \tag{19}$$

The steady-state portion, $\varphi(\eta)$, satisfies

$$d^2\varphi/d\eta^2 = 0 \tag{20}$$

$$\varphi(0) = F/(1 - F) \tag{21}$$

$$d\varphi/d\eta(1) + \frac{(1 - F)}{F} (N_{Bi} + 1)\varphi(1) = (N_{Bi} + 2)/F. \tag{22}$$

The steady-state solution is then

$$\varphi(\eta) = \left[\frac{2 - F + N_{Bi}(1 - F)}{1 + N_{Bi}(1 - F)}\right]\eta + F/(1 - F). \tag{23}$$

The transient portion of the solution results from letting $\vartheta(\eta,\tau)$ satisfy

$$\partial\vartheta/\partial\tau = \partial^2\vartheta/\partial\eta^2 \tag{24}$$

$$\vartheta(\eta, 0) = F/(1 - F)\left[\frac{1}{1 - (1 - F)\eta}\right] - \varphi(\eta) \tag{25}$$

$$\vartheta(0, \tau) = 0 \tag{26}$$

$$\partial\vartheta/\partial\eta(1, \tau) + \frac{(1 - F)}{F} (N_{Bi} + 1)\vartheta(1, \tau) = 0. \tag{27}$$

The solution to the problem in $\vartheta(\eta,\tau)$ is derived in Appendix B and can be stated as follows:

$$\vartheta(\eta, \tau) = \sum_{n=1}^{\infty} \frac{2(\lambda_n^2 + c_1^2)}{\lambda_n^2 + c_1^2 + c_1} \exp(-\lambda_n^2\tau) \sin \lambda_n\eta \int_0^1 \vartheta(\eta, 0) \sin \lambda_n\eta \, d\eta, \tag{28}$$

where $\lambda_n$ are the positive roots of

$$\lambda_n \cot \lambda_n = -c_1, \tag{29}$$

and the parameter $c_1$ is defined as

$$c_1 = \frac{(1 - F)}{F} (N_{Bi} + 1). \tag{30}$$

The series solution given above was found to converge too slowly for practical evaluation.* However, the problem specified by (14) through (17) can be readily solved through the use of finite differences. Once $v(\eta,\tau)$ was so determined for various values of $N_{Bi}$ and

---

* The difficulty resulted from the oscillating nature of the integral in (28). At small values of $\tau$ as many as 350 terms were inadequate for convergence.

$F$, then the temperature rise above ambient $\theta(r,t)$ could be found using (13).

## 2.2 *Temperature Distribution in the Conical Diaphragm*

In the analysis of the transient response of the carbon chamber, the boundary condition associated with the dome electrode has specified that the heat loss from every point on the interior surface of the dome electrode is proportional to the difference between the electrode temperature $T(a,t)$ and the ambient temperature $T_o$. The constant of proportionality has been designated as $h$, which is a measure of the impedance of the conical diaphragm to the flow of heat. Thus, the heat flux at the junction between the diaphragm and the dome electrode is given by*

$$q'' = h[T(a, t) - T_0] \cdot \gamma = -k \left. \frac{\partial T}{\partial s} \right|_{s=s_1} \qquad (31)$$

where the coordinate system is shown in Fig. 3. The factor $\gamma$ is defined
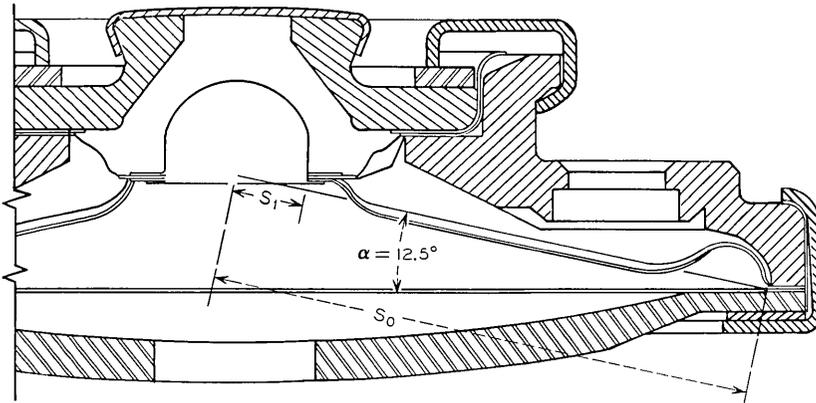


Fig. 3 — Coordinate system for heat flow along the conical diaphragm.

as the ratio of the surface area of the dome to the cross-sectional area of the diaphragm at $s = s_1$. If the thickness of the diaphragm is $\delta$ then $\gamma = a/\delta$.

In actuality, the coefficient $h$ as defined by (31) is not a constant since the $\partial T/\partial s|_{s=s_1}$ divided by $[T(a, t) - T_0]$ is still a function of

---

*It can easily be shown that the heat losses for the dome electrode to the air in contact with it are negligible.

time, somewhat larger at early times than at steady state. This means that the thermal impedance increases from some initial value in proportion to the loss of the capacitive effects as the transient dies out. Thus, the steady-state value of $h$ will be a minimum value.

A more exact approach to the problem would require the simultaneous determination of the transient response of the diaphragm linked to the carbon cavity by a matched heat flux boundary condition. However, if the capacitance of the diaphragm is small compared with the capacitance of the carbon cavity* then the determination of $h$ using the steady-state temperature gradient in (31) should be at least in the correct order of magnitude. Because of the crudeness of other approximations it was felt that the more exact approach was not warranted.

When we perform a heat balance for steady state axially symmetric heat conduction in the conical diaphragm, the following boundary value problem can be stated:

$$d/ds(s \, dT/ds) = 0 \tag{32}$$

$$T(s_1) = T(a, t) \tag{33}$$

$$T(s_0) = T_0 . \tag{34}$$

The solution is

$$\frac{T(s) - T(a, t)}{T_0 - T(a, t)} = \frac{\ln (s/s_1)}{\ln (s_0/s_1)}. \tag{35}$$

Differentiating (35) and using the results in (31), we then have for the steady-state approximation

$$h = \frac{\delta k_d \cos 12.5°}{a^2 \ln (s_0/s_1)} , \tag{36}$$

where

$$s_1 = a/\cos 12.5°.$$

## 2.3 Displacements Resulting from the Thermal Response

To calculate the displacements of the dome electrode we have chosen some rather simplified models for the geometrical configurations and constraints.

First, consider the dome electrode itself. If we assume that the

---

* The capacitance of the diaphragm is about one tenth the thermal capacitance of the carbon chamber for the Tl transmitter.

edge in contact with the diaphragm is essentially free it can be shown from both physical reasoning and the theory of elasticity that for an unconstrained hemisphere with a uniform temperature change the change in radius is given by

$$\Delta r = \alpha_t \theta(a, t) \cdot a. \tag{37}$$

The calculation of the displacements in the conical diaphragm are somewhat more involved because of the temperature gradient along the diaphragm. We first assume that the cone angle is small enough so that the conical diaphragm can be approximated by a disk. Thus, the displacement in the radial direction is given by[3]

$$u = (1 + \nu) \frac{\alpha_t}{s} \int_{s_1}^{s} [T(s) - T_0] s \, ds + C_1 s + C_2 / s, \tag{38}$$

where

$T(s)$ is the temperature along the diaphragm as given by (35)
$T_0$ is the (ambient) temperature in the stress free condition.

If we rewrite (35) in the following form:

$$T(s) - T_0 = [T(a, t) - T_0] \frac{\ln (s/s_0)}{\ln (s_1/s_0)} \tag{39}$$

where $T(a,t) - T_o = \theta(a,t)$, then (38) becomes

$$u = (1 + \nu) \frac{\alpha_t \theta(a, t)}{\ln (s_1/s_0)}$$

$$\cdot \left[ \frac{s}{2} \ln \left( \frac{s}{s_0} \right) - \frac{s_1^2}{2s} \ln \left( \frac{s_1}{s_0} \right) - \frac{s}{4} + \frac{s_1^2}{4s} \right] + C_1 s + C_2 / s. \tag{40}$$

The two constants can be determined from the boundary conditions on the diaphragm. We will assume that the outer edge of the diaphragm is fixed so that

$$u = 0, \qquad s = s_0. \tag{41}$$

If we also consider that the inner edge is free so that the radial stress at the location is zero then[3]

$$du/ds + \nu u/s - (1 + \nu)\alpha_t \theta(a, t) = 0, \qquad s = s_1. \tag{42}$$

Applying these conditions to (40) we find that

$$u = (1 + \nu) \frac{\alpha_t \theta(a, t)}{\ln (s_1/s_0)} \left[ \frac{s}{2} \ln \left( \frac{s}{s_0} \right) - \frac{s_1^2}{2s} \ln \left( \frac{s_1}{s_0} \right) - \frac{s}{4} + \frac{s_1^2}{4s} \right]$$

$$- \left[ s \left( \frac{1 - \nu}{1 + \nu} \right) + \frac{s_1^2}{s} \right] \left[ \frac{\frac{1}{4} - \frac{1}{4} \left( \frac{s_0}{s_1} \right)^2 + \frac{1}{2} \ln \left( \frac{s_0}{s_1} \right)}{1 + \left( \frac{s_0}{s_1} \right)^2 \left( \frac{1 - \nu}{1 + \nu} \right)} \right]. \qquad (43)$$

Since we are primarily interested in the displacement of the dome electrode which is fastened to the diaphragm at $s = s_1$, we first evaluate (43) for that location:

$$u(s_1) = \frac{2 s_1 \alpha_t \theta(a, t)}{\ln (s_0/s_1)} \left[ \frac{\frac{1}{4} - \frac{1}{4} \left( \frac{s_0}{s_1} \right)^2 + \frac{1}{2} \ln \left( \frac{s_0}{s_1} \right)}{1 + \left( \frac{s_0}{s_1} \right)^2 \left( \frac{1 - \nu}{1 + \nu} \right)} \right]. \qquad (44)$$

There are two methods by which the displacement along the slant height of the conical diaphragm can be exhibited as displacement of the dome electrode along the axis of symmetry of the T1 type transmitter. As shown in Fig. 4a, if the diaphragm is rigidly fastened to the dome then the dome displacement resulting from the thermal stresses in the diaphragm are given by

$$\Delta h_{\min} = -u \sin \alpha, \qquad (45)$$

so that the axial displacement, $\Delta h$, is somewhat less than the displacement along the slant height of the conical diaphragm. Considering the other extreme, as illustrated in Fig. 4b, if the fastening is loose but the dome electrode is rigid then the edge of the diaphragm moves as if up a rigid wall and the axial displacement is amplified
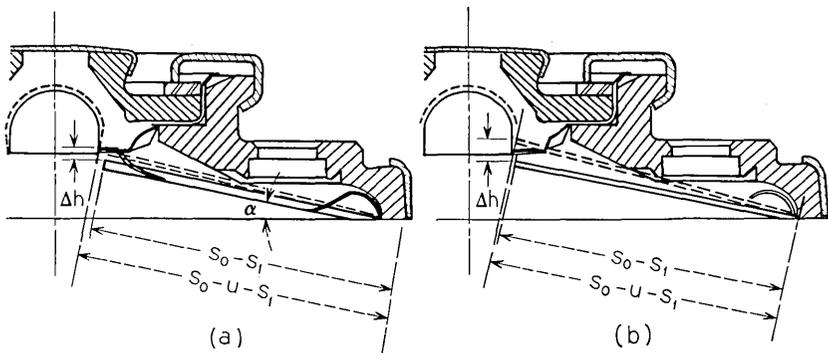


Fig. 4 — Displacement of dome electrode resulting from thermal stresses in the diaphragm.

by the sine of the angle. When the growth of the dome electrode, $\alpha_t \theta a$, is taken into account we have

$$\Delta h_{\max} = \frac{\alpha_t \theta a}{\tan \alpha} - \frac{u}{\sin \alpha}. \tag{46}$$

In the actual case the true axial displacement will be somewhere between these two extremes.

### III. ANALYTICAL RESULTS

The thermal response of the carbon chamber has been determined for various values of $N_{Bi}$ ranging from 0 to 100 and for two values of the shape parameter $F$. As noted earlier, the conical shaped back electrode has been approximated by a hemisphere of radius $b$. The choice of the value of $b$ (and hence $F$) was somewhat arbitrary. Thus, two values were taken for comparison. The character of the approximation is illustrated in Fig. 5. It was felt that the smaller value, $b = 0.44$ cm corresponding to $F = 0.636$, was the better choice.

The results in terms of the temperatures in the slab, $v(\eta, \tau)$, were transformed back to the temperatures in the carbon chamber using (13). Consequently,

$$\frac{a\theta(r, t)}{\beta} = \frac{F}{1 - (1 - F)\eta} [v(\eta, \tau) - v(\eta, 0)]. \tag{47}$$
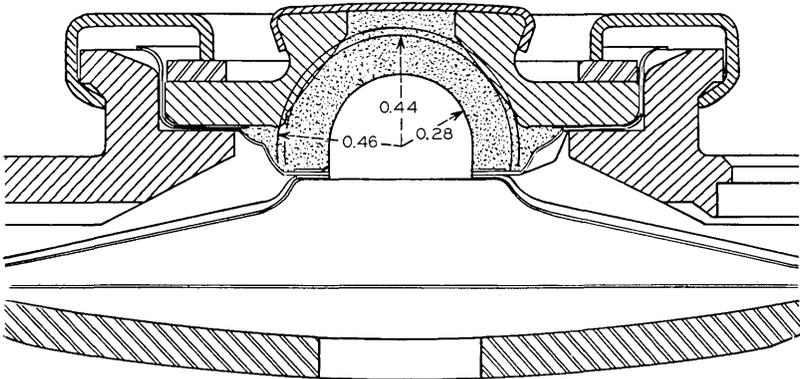


Fig. 5 — Approximation of carbon chamber by hemispherical cavity. Indicated radii are in cm.
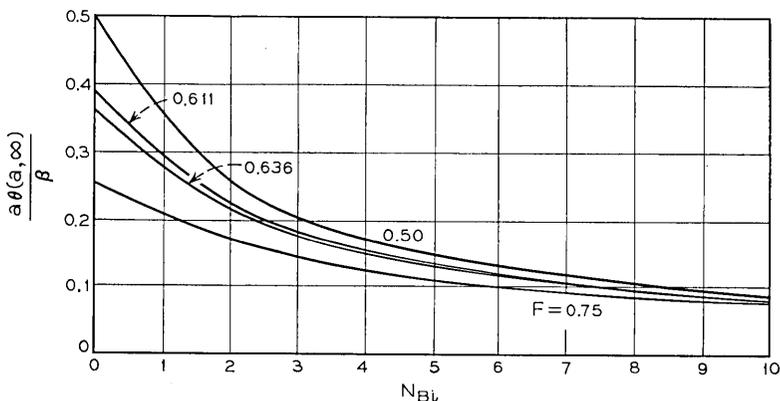
Fig. 6 — Dome electrode maximum temperature rise vs $N_{Bi}$. ($F$=0.636 for standard T1 transmitter.)

Evaluating (47) at $\eta = 1$, the temperature rise of the dome electrode can be expressed as:

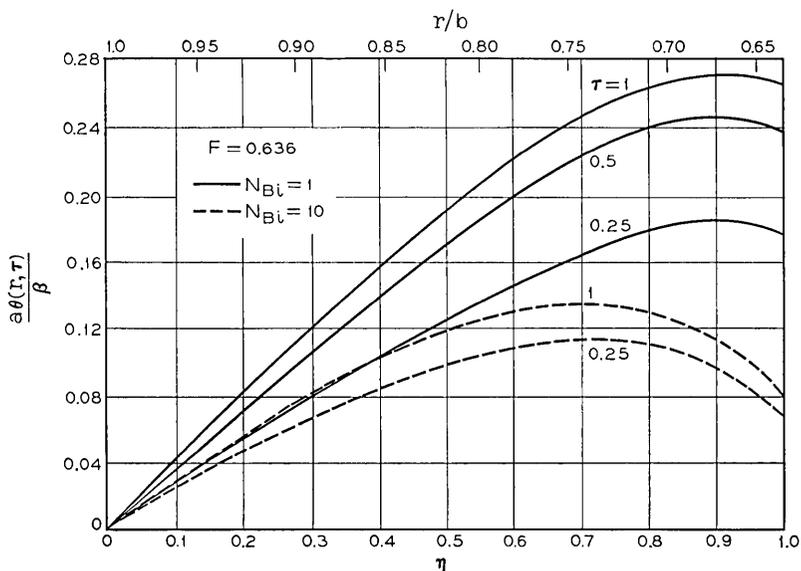$$\frac{a\theta(a, t)}{\beta} = [v(1, \tau) - 1/(1 - F)]. \qquad (48)$$



Fig. 7 — Temperature distributions at various times for two values of $N_{Bi}$.

From the steady-state solution (23) the maximum value of $\theta(a,t)$ becomes

$$\frac{a\theta(a, \infty)}{\beta} = \frac{(1 - F)}{1 + (1 - F)N_{Bi}}. \tag{49}$$

This function is plotted in Fig. 6 for various values of $F$.

The temperature distributions above ambient for $F = 0.636$ and two values of $N_{Bi}$ (1 and 10) are given in Fig. 7 for various times expressed in the nondimensional time variable $\tau$. The value of the abscissa denoted as $\eta = 1$ ($r/b = 0.636$) corresponds to the temperature rise of the dome electrode. As expected, the temperature rise of the dome electrode is substantially decreased as the Biot number is increased. This fact is also demonstrated in Fig. 6.

The "early time" transient response of the T1 transmitter to the "thermal packing" effect will be shown later to be closely linked to the transient temperature changes experienced by the dome electrode. This transient response is illustrated in Fig. 8 as the ratio of $\theta(a,t)/\theta(a, \infty)$ versus $\tau$. The effect of choosing a smaller value of $F$ is also shown for various values of $N_{Bi}$ by the broken lines. It is seen that virtually all of the transient temperature change has taken place by the time that $\tau = 1$. If the thermal time constant $\tau_o$ is defined as that value of $\tau$ where the transient has decayed to $1/e$ of its value, then the ratio $\theta(a,\tau)/\theta(a, \infty) = 0.632$ at $\tau = \tau_o$. Taking the values of $\tau_o$ from Fig. 8 then a curve of $\tau_o$ vs $N_{Bi}$ can be drawn for both values of $F$. These are indicated in Fig. 9 by the notation $\eta = 1$. The thermal time constant for a position which roughly characterizes the maximum temperature within the carbon chamber, that is, $\eta = 0.7$, is also shown.

## IV. EXPERIMENTAL OBSERVATIONS AND APPLICATION

Experimental observations have indicated that the electrical resistance of the T1 transmitter is significantly altered when a bias current is imposed. Fischer and Gaudet have attributed the major cause of this change to the thermal expansion of transmitter components resulting from the joulean heating in the carbon.[2] They tested both the standard T1 transmitter and a similar transmitter with a invar dome, diaphragm, and back electrode. The low expansion coefficient of invar reduced the thermal expansions and hence the invar transmitter packed less than the T1 transmitter. However, the reduction
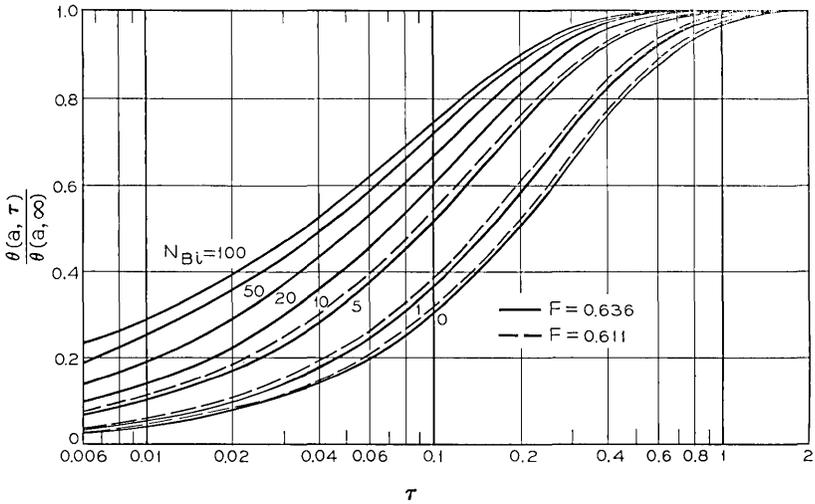
Fig. 8 — Transient temperature rise of the dome electrode for various $N_{Bi}$ and two values of $F$.

was not as significant as had been expected because the thermal resistance of the invar parts was greater and consequently higher temperatures were experienced on the dome electrode.
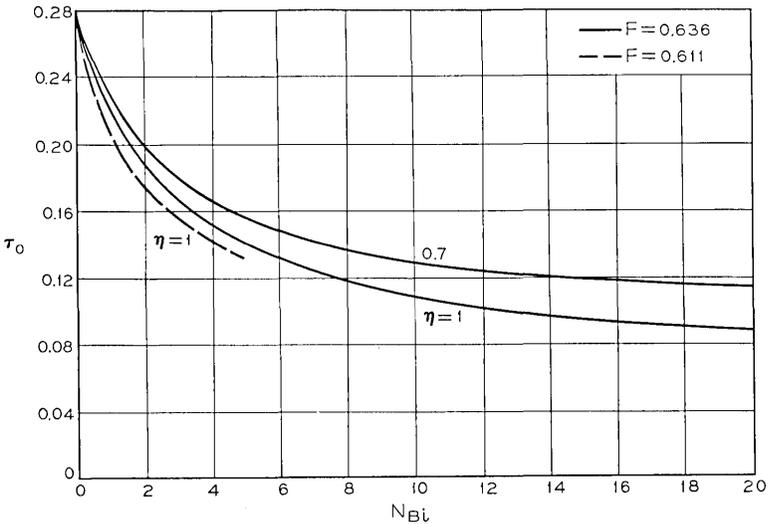


Fig. 9 — Thermal time constant at two positions as a function of $N_{Bi}$.

A detailed description of the thermal effects necessitated measurements of the dome electrode displacement and temperature. These were undertaken by C. Sandahl and P. E. Prettyman. The report has not been published but here is Mr. Sandahl's description of some of the instrumentation.

A flexible silicon semiconductor strain gage, 6 mils square by 50 mils long, was bonded to the diaphragm with epoxy in the region of maximum radial strain. A spherical bead thermistor 5 mils in diameter was attached to the underneath side of the dome electrode at the centerline with pressure sensitive tape.

The strain gage and thermistor were connected into Wheatstone bridge circuits. Bridge excitation was one volt, in order to minimize self-heating of the devices. The thermistor was calibrated in an oil bath using a NBS-calibrated thermometer graduated in $0.01°C$. The strain gage was calibrated in deflection of the dome by installing a small front-surface mirror (weight 0.204 gms) on a balsa wood plug in the dome. The transmitter was excited by the JRB circuit (a nearly constant current circuit) at various current levels and the steady-state deflection of the mirror was measured with an optical interferometer. The strain gage bridge output was measured for each increment of deflection.

During each test, thermistor bridge output, strain gage bridge output, JRB circuit voltage across the transmitter, and current through the transmitter were recorded simultaneously on a four-channel Sanborn recorder. Results for four typical tests on a particular transmitter are reproduced in Fig. 10. The test shown in 10d was a repeat of 10c changing only the scale factor on the original recording equipment.

In addition, detailed knowledge of the thermal resistance of granular carbon in a brass container was necessary for the application of the analysis presented here. This thermal resistance was found to consist of two parts: an effective thermal conductivity for a continuum representation of the granular carbon and a thermal contact resistance at the interface between the medium and the container walls. Experimental measurements of these two quantities have been recently completed and are reported in Ref. 4.

In Fig. 10, two types of response are apparent. There is an early time transient for which the back electrode temperature can be assumed constant. After a few seconds the back electrode and the whole transmitter begins to heat up. The subsequent over-all growth coupled with the thermal expansion of the granular carbon itself combines
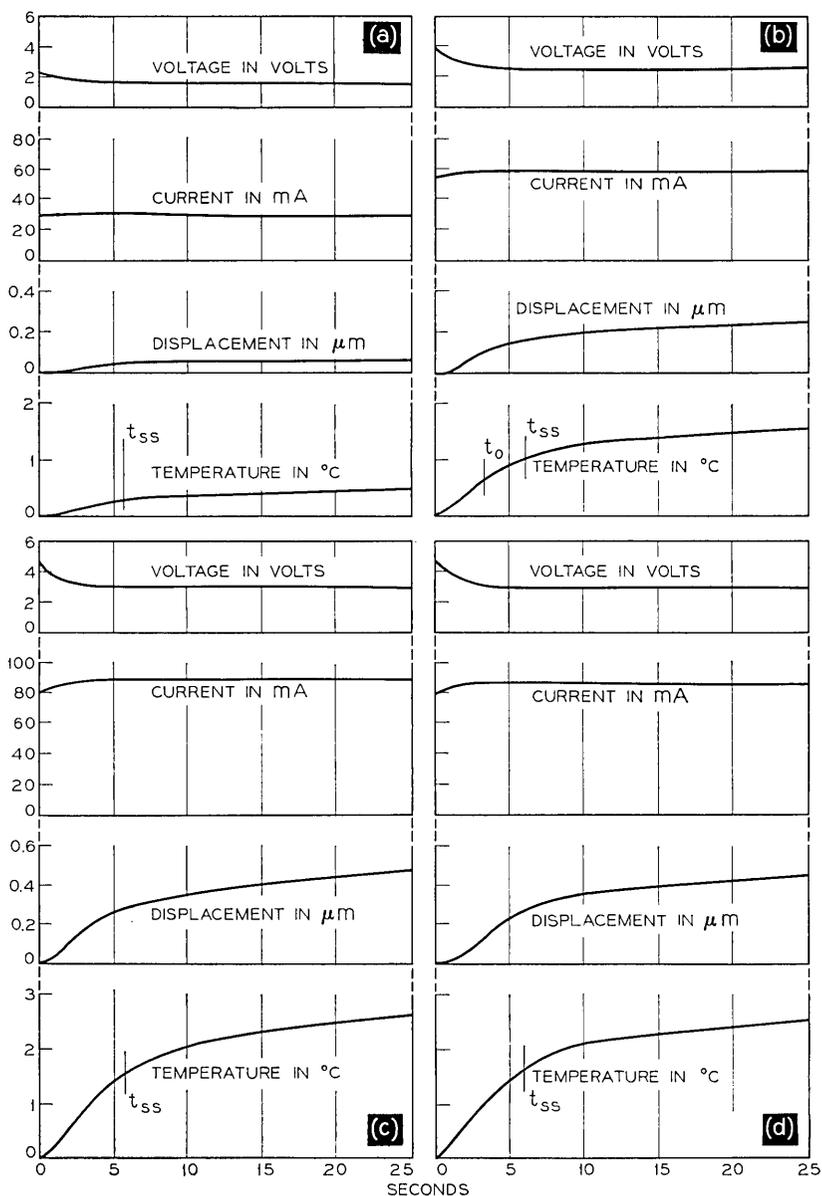
Figure 10. — Transient response of an excited T1 telephone transmitter. (a) Run 1, $\langle IV \rangle_{\text{av transient}} = 0.056W$; (b) run 2, $\langle IV \rangle_{\text{av transient}} = 0.156W$; (c) run 3, $\langle IV \rangle_{\text{av transient}} = 0.264W$; (d) run 4, $\langle IV \rangle_{\text{av transient}} = 0.264W$.

in such a manner that the transmitter resistance (voltage) reaches a constant value. No attempt will be made here to describe analytically this combination of effects at times later than about 6 seconds. However, the early time transient, which produces the resistance change, can be described.

We first calculate $N_{Bi}$ for the standard T1 transmitter. Calculating the thermal conductance of the diaphragm from (36)

$$h = \frac{\delta k_d \cos 12.5°}{a^2 \ln (s_0/s_1)} = 660 \text{ W}/(\text{m}^2 \cdot °\text{C})$$

when

$\delta$ = 0.0076 cm,
$k_d$ = 140 W/(m·°C),
$a$ = 0.28 cm,
$s_0$ = 2.1 cm,
$s_1$ = 0.286 cm.

From Ref. 4 we take the thermal conductivity of the carbon to be 0.24 W per (m·°C) and the thermal contact conductance as $h_c = 300$ W per (m²·°C). Consequently, the total conductance is the sum of the contact conductance in parallel with the conductance of the diaphragm:

$$1/h_{\text{tot}} = 1/660 + 1/300;$$

hence,

$$h_{\text{tot}} = 206 \text{ W}/\text{m}^2 \cdot °\text{C}).$$

Using $h_{\text{tot}}$ in $N_{Bi}$ we have

$$N_{Bi} = ha/k = 2.4.$$

Turning to Fig. 6 we see that $a\theta(a, \infty)/\beta = 0.19$ and Fig. 9 gives $\tau_o$ = 0.18 for this value of $N_{Bi}$. Now, the thermal diffusivity $\kappa = k/\rho c = 0.32 \times 10^{-6}$ m² per second for granular carbon where $\rho = 900$ kg per m³ and c = 840 J per (kg·°C) (for the loose state). Consequently,

$$t_0 = \frac{\tau_0(b - a)^2}{\kappa} = 1.4 \text{ seconds},$$

where $b = 0.44$ cm from Fig. 5. In the determination of $\beta$ we use the average values $\langle IV \rangle_{av}$ as obtained by a graphical integration of the transient portions of the appropriate curves of Fig. 10. Taking the

value from Fig. 10b as typical we have

$$\beta = \frac{\langle IV \rangle_{av}}{4\pi k} = 5.17 \text{ cm}°\text{C}.$$

Therefore, $\theta = 0.19\beta/a = 3.5°\text{C}$.

By comparing the time constant of 1.4 seconds to the measured values of $t_o$ in the range of 3 seconds, one could say that this is reasonable agreement especially since the capacitance effects of the diaphragm and the electrodes have been neglected in the analysis. The capacitance in the real situation would tend to give higher values of $t_o$ as observed.

The calculated temperature rise of 3.5°C compared with the measured "steady-state" value of 1.0°C needs some explanation. Of course, the capacitance and contact resistance of the thermistor would tend to give rise to the lower measured value. But more important is the fact that the predicted value is for the granular carbon continuum which is linked to the brass dome through a contact resistance. The temperature drop across that resistance can be calculated by multiplying the temperature rise above ambient ($\theta$) by the ratio of the contact resistance ($1/h_c$) to the total resistance ($1/h_{tot}$);

$$\Delta T_{\text{continuum to thermistor}} = \frac{\theta(1/h_c)}{1/h_{tot}} = 2.4°\text{C}.$$

Therefore, $\Delta T_{\text{thermistor above ambient}} = 3.5°\text{C} - 2.4°\text{C} = 1.1°\text{C}$; which is in remarkably good agreement with the measured value of 1.0°C. The other measured values, at the end of the early transient (namely, 0.33°C, 1.56°C, and 1.61°C), scale linearly with the input power; hence, the agreement with analysis remains good for all the results given in Fig. 10.

Consider briefly a transmitter made of invar. For invar the $k_d$ value is 18 times smaller than that of aluminum. Consequently,

$$1/h_{tot} = 1/36.6 + 1/300;$$

hence

$$h_{tot} = 32.6 \text{ W}/(\text{m}^2 \cdot °\text{C})$$

and $N_{Bi} = 0.4$. From Fig. 6 we see that the temperature rise would be 1.74 times that under corresponding conditions in the T1 transmitter. From Fig. 9 the value of $t_o$ would be 30 percent longer. This is in qualitative agreement with the observations of Fischer and Gaudet.[2]

Turning now to the displacement we first calculate the extension of the slant height of the aluminum diaphragm using (43). Taking $\nu = 1/3$ and $\alpha_t = 23.4 \times 10^{-6}°C^{-1}$ we obtain for a temperature rise of 1°C

$$u = -0.029 \ \mu m.$$

The minus sign indicates that the diaphragm grows toward the dome electrode. The expansion of the dome electrode is obtained from (37) using $\alpha_t = 16.2 \times 10^{-6}°C^{-1}$ for brass. For $\theta = 1°C$, $\Delta r = 0.04 \ \mu m$. The minimum displacement, given by (45), becomes 0.0064 $\mu m$, while the maximum displacement given by (46) becomes

$$\Delta h_{max} = \frac{\alpha_t \theta a}{\tan \alpha} - \frac{u}{\sin \alpha} = 0.180 + 0.134 = 0.32 \ \mu m.$$

The upper limit which assumes a perfectly rigid dome electrode compares reasonably well with the measured value of 0.18 $\mu m$ from Fig. 10b. Notice that the top of the dome electrode moves an additional 0.04 $\mu m$.

The most significant factor arising out of this rather crude model for displacements is that the thermal packing appears to be more sensitive to the thermal expansion coefficient for the dome electrode than any other part. Hence, a simple and inexpensive modification of the T1 transmitter would have been to change only the dome electrode to invar. Although the above analysis does not strictly apply to the modified transmitter design of Huffstutler,[1] the same conclusion with regards to an invar dome seems appropriate.

The thermal expansion of the carbon produces an additional packing effect. The change in volume caused by the heating could be calculated by using the expansion coefficient value[5] of $10.4 \times 10^{-6}°C^{-1}$ and the local temperature as given in Fig. 7. An estimate of such a calculation yields a dome displacement equivalent to 0.05 $\mu m$ in the direction opposite the displacements mentioned above for the conditions of Fig. 10b.

V. CONCLUSIONS

The resistance change of the T1 transmitter resulting from the joulean heating can be associated with the thermal expansion of the dome electrode and the aluminum diaphragm. The agreement of the model presented here with the measurements of Sandahl and Prettyman attests to its validity for the early time transients. The displacements which produce thermal packing are a stronger function of the

thermal expansion coefficient of the dome electrode than any other part. A simple modification would be to make this part of invar in the T1 or any similar transmitter. The model also demonstrates that the displacements associated with the temperature rise of the dome electrode are affected less by changing the thermal parameters of the system than by reducing the thermal expansion coefficient in going from aluminum and brass to invar. The thermal contact resistance between the granular carbon and the dome electrode reduces the temperature rise significantly.

## VI. ACKNOWLEDGMENT

The author wishes to acknowledge the stimulating discussions with R. J. Port and the suggestion by R. J. Grosh of the existence of a contact resistance.

## APPENDIX A

### Localized Heating in the Carbon Granules

Because of the asperities of carbon granules and of the "bunching up" of the current flow at the contact areas, the importance of localized heating should be checked. It can be assumed that the localized effects are virtually all diffused at the values of time for which the Fourier modulus $\kappa t_0/D^2$ is approximately unity. Then on setting the characteristic dimension of the granule, $D$ equal to 0.2 mm, and the thermal diffusivity, $\kappa$, of solid carbon equal to $0.2 \times 10^{-4}$ m$^2$ per second, the value of the time constant, $t_0$, becomes approximately

$$t_0 = D^2/\kappa = \frac{(0.2)^2 \times 10^{-6}}{0.28 \times 10^{-4}} = 1.4 \text{ msec.}$$

We see that the thermal time constant for the solid carbon granules is at least three orders of magnitude less than the thermal time constant of the thermal effects being observed. Consequently, any localized heating is rapidly diffused and the granular carbon can be treated as a continuum.

## APPENDIX B

### Transient Temperature Response of the Carbon Chamber

Consider the transient portion of the heat conduction problem specified by (24) through (27) and repeated below for convenience.

$$\partial\vartheta/\partial\tau = \partial^2\vartheta/\partial\eta^2 \tag{50}$$

$$\vartheta(\eta, 0) = F/(1 - F)\left[\frac{1}{1 - (1 - F)\eta}\right] - \varphi(\eta) \tag{51}$$

$$\vartheta(0, \tau) = 0 \tag{52}$$

$$\partial\vartheta/\partial\eta(1, \tau) + \frac{(1 - F)}{F}(N_{Bi} + 1)\vartheta(1, \tau) = 0. \tag{53}$$

Assuming a separable solution of the form

$$\vartheta(\eta, \tau) = X(\eta)T(\tau) \tag{54}$$

so that

$$X = A \sin \lambda\eta + B \cos \lambda\eta \tag{55}$$

$$T = Ce^{-\lambda^2\tau}. \tag{56}$$

But (52) requires that $B = 0$ and (53) specifies that

$$\lambda A \cos \lambda + c_1 A \sin \lambda = 0$$

or,

$$\lambda \cot \lambda = -c_1 \tag{57}$$

where

$$c_1 = \frac{(1 - F)}{F}(N_{Bi} + 1). \tag{58}$$

To satisfy the nonhomogeneous initial condition we set

$$\vartheta(\eta, 0) = \sum_{n=1}^{\infty} A_n X_n , \tag{59}$$

where $X_n = \sin \lambda_n\eta$ and $\lambda_n$ is the $n$th positive root of (57). Multiplying both sides of (59) by $X_m$, integrating, and using the orthogonality condition we have

$$A_n = \frac{\int_0^1 X_n(\eta)\vartheta(\eta, 0)\, d\eta}{\int_0^1 X_n^2\, d\eta}. \tag{60}$$

Evaluating the integral in the denominator where the boundary conditions have been used we have

$$\int_0^1 X_n^2\, d\eta = \frac{\lambda_n^2 + c_1^2 + c_1}{2(\lambda_n^2 + c_1^2)}. \tag{61}$$

The transient part of the solution can thus be written as

$$\vartheta(\eta, \tau) = \sum_{n=1}^{\infty} \frac{2(\lambda_n^2 + c_1^2)}{\lambda_n^2 + c_1^2 + c_1} \exp(-\lambda_n^2 \tau) \sin \lambda_n \eta \int_0^1 \vartheta(\eta, 0) \sin \lambda_n \eta \, d\eta.$$

(62)

To evaluate the integral in (62) we first notice that (51) can be written as

$$\vartheta(\eta, 0) = c_0 - c_2 \eta + \frac{c_3}{1 - c_4 \eta}$$

so that

$$\int_0^1 \vartheta(\eta, 0) \sin \lambda_n \eta \, d\eta = c_0/\lambda_n[1 - \cos \lambda_n] - c_2/\lambda_n^2[\sin \lambda_n - \lambda_n \cos \lambda_n]$$

$$+ c_3 \int_0^{\lambda_n} \frac{\sin \xi \, d\xi}{\lambda_n - c_4 \xi}.$$

(63)

The last term which can be expressed in various forms is the one which causes the series to converge very slowly.

REFERENCES

1. Huffstutler, M. C., Jr., and Kerns, B. T., unpublished work.
2. Fischer, C. A. and Gaudet, G. G., unpublished work.
3. Timoshenko, S. and Goodier, J. N., *Theory of Elasticity*, New York: Mc-Graw-Hill, 1951, p. 407.
4. Fritsch, C. A. and P. E. Prettyman, "Measurements of the Thermal Conductivity of Granular Carbon and the Thermal Contact Resistance at the Container Walls," Proc. 7th Conf. Thermal Conductivity, Nat. Bureau of Standards, Gaithersburg, Maryland, November 13–15, 1967.
5. Booth, T. C., unpublished work.

# Gain of Antennas with Random Surface Deviations

By H. ZUCKER

*On-axis gain of antennas with rough reflecting surfaces is computed as a function of rms surface deviation $\epsilon$, correlation distance c, antenna area A and wavelength $\lambda$. Gaussian stationary surface deviations, Gaussian correlation functions, and uniform illumination are assumed. Antennas with rectangular and circular apertures are considered. It is shown that a normalized gain can be defined which has the same functional form for both. A principal result of this work is quantitative calculation of the on-axis antenna gain when the normalized variance $(4\pi\epsilon/\lambda)^2$ of the rough surface is larger than 4. The off-axis gain is also considered, and it is shown that in the asymptotic limit (as $\lambda \rightarrow 0$), the gain reduces to that obtained by using geometrical optics.*

## I. INTRODUCTION

The gain of shallow paraboloid reflector antennas with random surface deviations has been derived by Ruze.[1,2] The deviation was based on scalar Kirchhoff approximation to the radiation from reflector antennas. The surface deviations were assumed to be gaussian stationary with gaussian correlation functions. On these bases, an approximate solution for the antenna gain was obtained in terms of an infinite series. The series has been evaluated for relatively small rms surface deviations, $\epsilon$, in comparison to the wavelength, $\lambda$, namely $(4\pi\epsilon/\lambda)^2 \leqq 4$. Asymptotic limits (as $\lambda \rightarrow 0$) for the gain were also given by Ruze[2] based on a similar analysis by Scheffler[3]. On-axis gain measurements of large reflector antennas as a function of frequency, exhibit the characteristics as predicted theoretically by Ruze.

The present work was motivated primarily to determine the gain in the intermediate region between very long and very short wavelengths and to establish a criterion for applicability of the asymptotic limit. Of primary interest was the near axis field distribution in the focal plane of a paraboloid reflector antenna illuminated by an inci-

dent plane wave. However, since both the far-field radiation pattern and the field distribution in the focal plane are Fourier transforms of the antenna aperture illumination, the derivations by Ruze are applicable for determining both the far-field and focal-plane distributions.

The series solution for the antenna gain obtained by Ruze does not seem to be suitable for numerical computations for large arguments. This is because some of the terms in the series will assume large values before the terms begin to decrease. However, the series for the on-axis gain is related to an exponential integral. The exponential integral also has an asymptotic series representation, which is particularly suitable for numerical computation for large arguments. On this basis, the on-axis gain has been computed as a function of the rms surface deviation to the wavelength ratio and for a range of correlation parameters. The asymptotic limit for the gain is evident from these computations.

The off-axis gain is also considered. Asymptotic representations of the series which may facilitate the off-axis gain computations are discussed. The limiting value ($\lambda \rightarrow 0$) for the off-axis gain is obtained, and it is shown that in this limit, the gain reduces to that obtained from geometrical optics.[4]

The gain of antennas with rectangular apertures and gaussian stationary surface deviations is presented by assuming uniform illumination. A generalization to include certain types of nonuniform illuminations is discussed. The on-axis gain for antennas with circular apertures also is given. It is shown that the on-axis gain for antennas with rectangular and circular apertures can be normalized, such that the normalized gain is the same for both. The off-axis gain is expressed in terms of series with known asymptotic expansions.

## II. ANTENNA GAIN

The far field gain, $G(\theta, \Phi)$, in the vicinity of the axis of a shallow paraboloid reflector antenna with surface deviations, $z(x,y)$ is, using the scalar Kirchhoff approximation[5]

$$G(\theta, \Phi) = \frac{4\pi}{\lambda^2}$$

$$\cdot \frac{\iint_s \iint_s E_a(x, y)E_a^*(x_1, y_1) \exp{(j\{\beta_x u + \beta_y v + 2k[z(x, y) - z(x_1, y_1)]\})}\, ds\, ds_1}{\iint_s E_a(x, y)E_a^*(x, y)\, ds} \tag{1}$$

where $E_a$ is the projected electric field on the antenna aperture and $s$ is the aperture area.

$$k = \frac{2\pi}{\lambda} = \text{free space propogation constant}$$

$$\lambda = \text{wavelength}$$

$$\beta_x = k \sin \theta \cos \Phi \tag{2}$$

$$\beta_y = k \sin \theta \sin \Phi \tag{3}$$

$\theta$ and $\Phi$ are the spherical coordinates indicated in Fig. 1.

$$u = x - x_1 \tag{4}$$

$$v = y - y_1 \tag{5}$$

The Kirchhoff approximation is based on the assumption that the surface is locally plane, and hence equation (1) is applicable to surfaces for which the curvatures are small.

Equation (1) can also be used to determine the power distribution in the focal plane of shallow paraboloid reflector antennas in the vicinity of the focal point, in which case (referring to Fig. 1)

$$\beta_x = kx_f/f \tag{6}$$

$$\beta_y = ky_f/f \tag{7}$$

where $x_f$ and $y_f$ are the coordinates in the focal plane and $f$ is the focal length.

If $z(x,y)$ is a Gaussian stationary random variable with zero mean it has been shown[1, 6] that by performing the statistical averaging, the expectation value for the gain, $\langle G(\theta, \Phi) \rangle_{av}$ is:

$$\langle G(\theta, \Phi) \rangle_{av} = \frac{4\pi}{\lambda^2} \exp(-\delta^2)$$

$$\cdot \frac{\iint_s \iint_s E_a(x, y)E_a^*(x_1, y_1) \exp[j(\beta_x u + \beta_y v)] \exp[\delta^2 r(u, v)]\, ds\, ds_1}{\iint E_a(x, y)E_a^*(x, y)\, ds} \tag{8}$$

where

$$\delta = \frac{4\pi}{\lambda} \epsilon \tag{9}$$
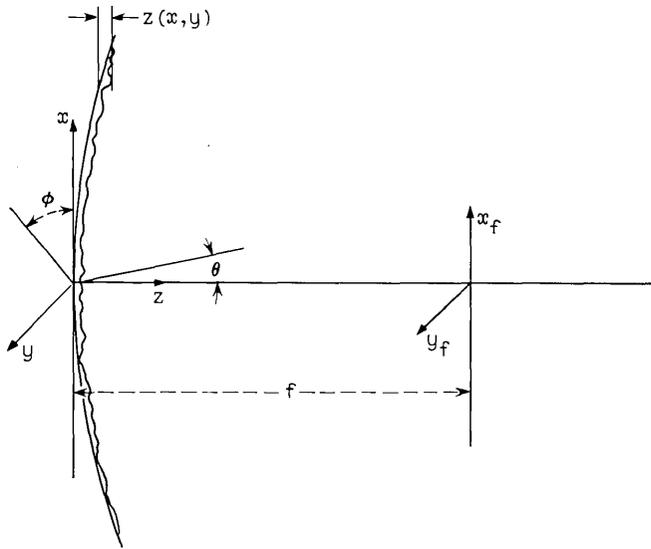
Fig. 1 — Antenna coordinates.

$\epsilon$ = rms surface deviation

$\delta^2 r(u, v)$ = correlation function.

To evaluate equation (8), four integrations have to be performed. It is shown in Appendix A that for antennas with rectangular apertures two integrations can be readily eliminated for certain types of illuminations, truncated cosine illuminations, for example.

In particular for uniform illuminations, $E_a(x, y) = 1$, and for a gaussian correlation function with

$$r(u, v) = \exp\left(-\frac{u^2 + v^2}{c^2}\right) \tag{10}$$

where $c$ is the correlation length, it is shown in Appendix A that the expectation value of the gain is:

$$\langle G(\theta, \Phi)\rangle_{\text{av}} = \exp(-\delta^2)G_0(\theta, \Phi)$$

$$+ \left(\frac{2\pi c}{\lambda}\right)^2 \exp(-\delta^2) \sum_{n=1}^{\infty} \frac{\delta^{2n}}{n!\, n} \left[\exp\left(-\frac{\beta^2 c^2}{4n}\right) - \Delta_n\right] \tag{11}$$

where $G_0(\theta, \Phi)$ is the antenna gain in the absence of surface deviations.

For an antenna with a rectangular aperture

$$G_0(\theta, \Phi) = \frac{4\pi A}{\lambda^2} \left( \frac{\sin \beta_x a}{\beta_x a} \frac{\sin \beta_y b}{\beta_y b} \right)^2 \tag{12}$$

where $A = 4ab$ is the aperture area.

$$\beta = \frac{2\pi}{\lambda} \sin \theta \tag{13}$$

and

$$\Delta_n < \frac{c}{2(\pi n)^{\frac{1}{2}}} \left[ \frac{1}{a} + \frac{1}{b} \right] \tag{14}$$

Equation (11) agrees with the gain derived by Ruze for antennas with circular apertures except for the term $\Delta_n$. This term is small, if the correlation distance $c$ is small compared with the linear dimensions of the antenna. This assumption is made in the subsequent computations.

For antennas with circular apertures the exact evaluation of equation (8) is in general more difficult. However, for uniform illumination, the on-axis gain $\langle G(0, 0) \rangle_{av}$ is evaluated exactly in Appendix B with the aid of $Q$ functions. The gain has the same functional form as equation (11) with $\beta = 0$. In particular for $n/2(D/c)^2 \gg 1$

$$\Delta_n \approx \frac{2c}{D(\pi n)^{\frac{1}{2}}} \tag{15}$$

where $D$ is the antenna aperture diameter.

### III. ON-AXIS GAIN

Equation (11) can be readily computed for small values of $\delta^2$. For large values of $\delta^2$ the terms $\delta^{2n}/n!n$ will become very large; therefore, the series is not suitable for direct computation if $\delta^2$ is large. Nevertheless, the gain on-axis can be readily computed by noticing that the series in equation (11) for $\beta = 0$ is related to an exponential integral, which also has an asymptotic representation.

The exponential integral, $E_i$, can be written[7]

$$E_i(x) = \gamma + \ln x + \sum_{n=1}^{\infty} \frac{x^n}{n!\,n} \tag{16}$$

where $\gamma$ is Euler's constant. The asymptotic series $(x \to \infty)$ for

$E_i(x)$ is

$$E_i(x) = \frac{\exp(x)}{x} \sum_{n=0}^{N-1} \left[ \frac{n!}{x^n} + O\left(\frac{1}{x^N}\right) \right]. \tag{17}$$

Though the asymptotic series diverges for all finite values of $x$ it can be used to evaluate $E_i(x)$ for large $x$ by using up to $N$ terms,[8] where $N$ is an integer nearest to the value of $x$.

In terms of the exponential integral, the on-axis gain for both rectangular and circular aperture antennas is

$$\langle G(0,0) \rangle_{av} = \left( \frac{D_0}{4\epsilon} \right)^2$$

$$\cdot \left\{ \delta^2 \exp(-\delta^2) + \left( \frac{2c}{D_0} \right)^2 \delta^2 \exp(-\delta^2)[E_i(\delta^2) - \ln \delta^2 - \gamma] \right\} \tag{18}$$

where $D_0$ is related to the antenna area, $A$, by

$$A = \frac{\pi D_0^2}{4}. \tag{19}$$

One parameter in (18) is readily eliminated by defining a normalized on-axis gain, $\langle g(0,0) \rangle_{av}$, by

$$\langle g(0,0) \rangle_{av} = \frac{\langle G(0,0) \rangle_{av}}{(D_0/4\epsilon)^2}$$

$$= \delta^2 \exp(-\delta^2) \left\{ 1 + \left[ \frac{2c}{D_0} \right]^2 [E_i(\delta^2) - \ln \delta^2 - \gamma] \right\}. \tag{20}$$

The normalized gain thus depends only on two parameters, $\delta^2$ and $(c/D_0)^2$.

Equation (20) has been computed by using a SHARE program for the computation of the exponential integral*. This program computes $E_i(x)$ with at least four-decimal accuracy.

Computations have been performed for $10^{-4} \leq \delta^2 \leq 80$ and for $10^{-3} \leq c/D_0 \leq 0.1$. The computed normalized gain is shown in Fig. 2.

The computations show the normalized antenna gain has three distinct regions which are characterized by the normalized rms surface deviation to wavelength ratio, $\delta$.

In the region $0 \leq \delta^2 \leq 1$ the normalized antenna gain is nearly independent of the correlation length $c$, and increases almost linearly with $\delta^2$. In the region $1 \leq \delta^2 \leq 20$ the gain is dependent on both $\delta^2$

---

* Contributed by D. S. Villars.
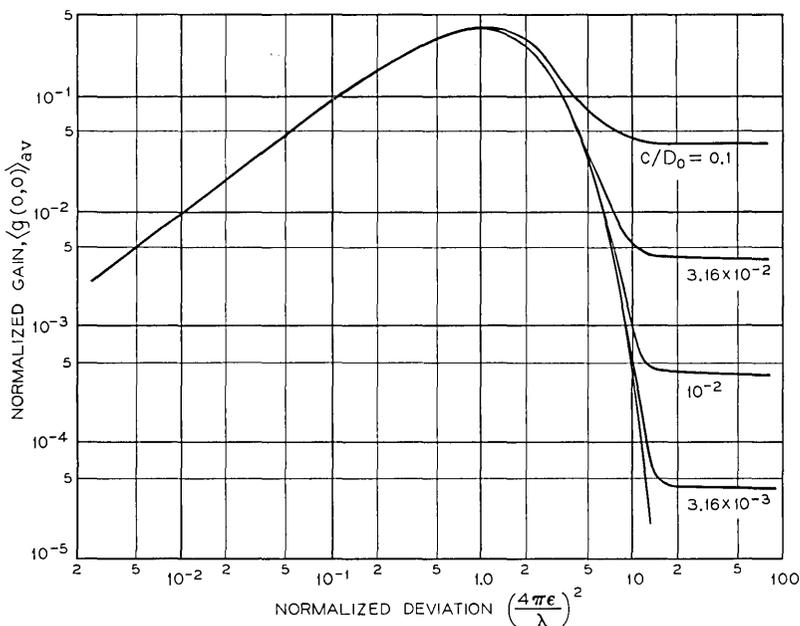
Fig. 2 — Normalized antenna gain.

and $c$. In the region $\delta^2 > 20$ the gain is almost independent of $\delta^2$ and is a function of $c/D_0$ only. This region is the asymptotic region. For the range of parameters used in the computation, the gain in the asymptotic region, for a given $c/D_0$, deviates by less than 5 percent from the asymptotic value.

The curves shown in Fig. 2 seem to confirm the general characteristics of the measured gain as a function of frequency of large reflector antennas presented by Ruze.[2] The presented measurements extend only slightly from the first into the second region but not sufficiently far to determine qualitative agreement between the theory and experiment in much of the intermediate and all of the asymptotic regions. A detailed comparison of the measured and computed gain can not be made since uniform illumination has been assumed in the computation.

## IV. ASYMPTOTIC VALUES FOR THE OFF-AXIS GAIN

Computation of the off-axis gain directly from equation (11) can only be readily performed for relatively small values of $\delta^2$. An

alternate representation for the off-axis gain for large values of $\delta$ is obtained by expanding the exponential in the second term of equation (11) in a power series. By using this expansion and neglecting $\Delta_n$, (11) can be rewritten as follows:

$$\langle G(\theta, \Phi)\rangle_{\text{av}} = \exp(-\delta^2)G_0(\theta, \Phi) + \left(\frac{c}{2\epsilon}\right)^2 \sum_{m=0}^{\infty} (-1)^m \frac{A_m(\delta^2)}{m!}\left(\frac{\beta c}{2\delta}\right)^{2m} \quad (21)$$

where

$$A_m(\delta^2) = \exp(-\delta^2) \sum_{n=1}^{\infty} \frac{(\delta^2)^{n+m+1}}{n! \, n^{m+1}}. \quad (22)$$

The series for $A_m(\delta^2)$ are special cases of functions considered by Barnes[9] who obtained their asymptotic expansions. These functions were also studied by Ford[10] who also presented a recurrence relation for the coefficients of the asymptotic series. The above functions designated by $G_\beta(x,\Theta)$ are:

$$G_\beta(x, \Theta) = \sum_{n=0}^{\infty} \frac{x^n}{n! \, (n + \Theta)^\beta}. \quad (23)$$

The functions $A_m(\delta^2)$ can be written as:

$$A_m(\delta^2) = (\delta^2)^{m+2} \exp(-\delta^2)G_{m+2}(\delta^2, 1). \quad (24)$$

Only the asymptotic limit for the off-axis gain is considered. For $x \to \infty$

$$G_\beta(x, \Theta) = \frac{\exp(x)}{x^\beta}\left[1 + O\left(\frac{1}{x}\right)\right]$$

hence for $\delta^2 \to \infty$

$$A_m = 1 + O\left(\frac{1}{\delta^2}\right). \quad (25)$$

The off-axis asymptotic gain will be designated by $G(\theta, \Phi)_\infty$ and is given by:

$$G(\theta, \Phi)_\infty = \left(\frac{c}{2\epsilon}\right)^2 \exp\left[-\left(\frac{c}{4\epsilon}\sin\theta\right)^2\right] \quad (26)$$

The corresponding normalized gain is found as

$$g(\theta, \Phi)_\infty = \left(\frac{2c}{D_0}\right)^2 \exp\left[-\left(\frac{c}{4\epsilon}\sin\theta\right)^2\right]. \quad (27)$$

The asymptotic value for the gain, equation (26), is in agreement with the gain obtained, based on an approximation to the gaussian correlation functions,[3] and also with results obtained by using geometrical optics.[4] Equation (24) is independent of frequency but strongly dependent on the ratio $\varepsilon/c$. This ratio has been interpreted as an average surface slope.[2] The range of $\delta$ over which the above off-axis approximation applies has not been determined precisely, however, it is reasonable to assume that this range will correspond to the asymptotic region for the on-axis gain.

## V. CONCLUSIONS

The on-axis gain of antennas with gaussian stationary random surface deviations and gaussian correlation functions has been determined for antennas with rectangular and circular apertures by assuming uniform illumination. For both types a normalized expression for the gain was derived which depends only on the normalized rms surface deviation $\varepsilon$, to wavelength $\lambda$ ratio, $\delta(=4\pi\varepsilon/\lambda)$, and the ratio of the correlation length $c$ to a defined linear antenna dimension $D_o$. For circular antennas, $D_o$ is the diameter.

The antenna gain as a function of $\delta$ exhibits three distinct regions: (i) $0 \leqq \delta^2 \leqq 1$, (ii) $1 \leqq \delta^2 \leqq 20$, and (iii) $\delta^2 > 20$. The last is called the asymptotic region. In this region the gain is nearly independent of wavelength.

The computed gain exhibits in general the characteristics of the measured gain as a function of frequency of large reflector antennas reported in the literature. These measurements extend only partially into the second region and have not been obtained in the third (asymptotic) region.

For large values of $\delta^2$ the off-axis gain can be expressed in terms of series with known asymptotic expansions. The limiting value for the off-axis gain has been obtained and reduces to that obtained by geometrical optics.

## VI. ACKNOWLEDGMENTS

APPENDIX A

*Gain of Antennas with Rectangular Apertures*

To evaluate equation (8) for antennas with rectangular apertures, consider the following integral

$$\langle I \rangle_{av} = \exp(-\delta^2) \int_{-a}^{a} \int_{-a}^{a} \int_{-b}^{b} \int_{-b}^{b} E_a(x, y) E_a(x_1, y_1)$$

$$\cdot \exp[\delta^2 r(u, v)] \exp[j(\beta_x u + \beta_y v)] \, dx \, dx_1 \, dy \, dy_1 \qquad (28)$$

with

$$u = x - x_1 \qquad (29)$$

$$v = y - y_1. \qquad (30)$$

Since equation (28) contains the correlation function in terms of $u$ and $v$, it is preferable to introduce the $u,v$ coordinate system.

In the $x,x_1$ coordinate system the integrations are over the square region shown in Fig. 3a. In the $y, y_1$ system the region is similar. With
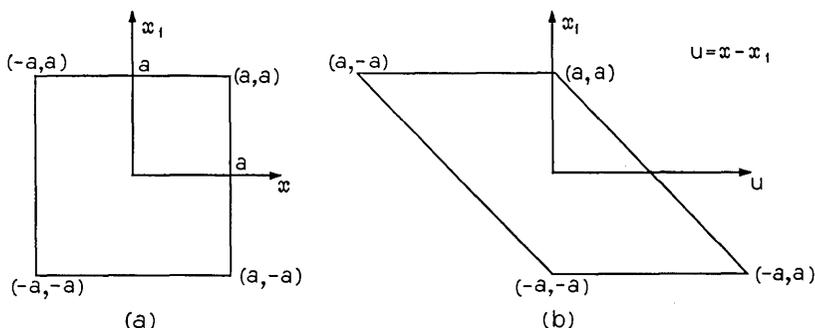


Fig. 3 — Coordinate transformation.

the coordinate transformation equation (29), the transformed region in the $x_1,u$ coordinate system is also shown in Fig. 3b.

In the $x_1,u$ plane the integration with respect to $x_1$ is readily performed for certain types of illumination functions.* In particular, let

$$E_a(x, y) = E_{ax}(x) E_{ay}(y) \qquad (31)$$

---

* A similar method has been used by Hoffman in his treatment of scattering of electromagnetic waves from a random surface.[6]

where.

$$E_{ax}(x_1 + u) = \sum_{n=1}^{N} g_n(x_1)f_n(u) \tag{32}$$

with a similar equation for $E_{ay}(y_1 + v)$. An example of such illumination are truncated cosine illuminations, where equation (32) will consist of two terms.

It is sufficient to consider the following integral

$$\langle I_1 \rangle_{\mathrm{av}} = \int_{-a}^{a} \int_{-a-x_1}^{a-x_1} g(x_1)f(u, v) \, du \, dx_1 . \tag{33}$$

Referring to Fig. 3, equation (33) can be written as

$$\langle I_1 \rangle_{\mathrm{av}} = \int_{0}^{2a} \int_{-a}^{a-u} g(x_1)f(u, v) \, dx_1 \, du$$

$$+ \int_{-2a}^{0} \int_{-a-u}^{a} g(x_1)f(u, v) \, dx_1 \, du \tag{34}$$

let

$$G(x_1) = \int g(x_1) \, dx_1$$

then

$$\langle I_1 \rangle_{\mathrm{av}} = \int_{0}^{2a} \{ [G(a - u) - G(-a)]f(u, v)$$

$$+ [G(a) - G(-a + u)]f(-u, v)\} \, du. \tag{35}$$

Using equation (35) and assuming uniform illumination, $E_a(x, y) = 1$, two integrations are readily eliminated and equation (28) reduces to

$$\langle I \rangle_{\mathrm{av}} = 4 \exp\left(-\delta^2\right) \int_{0}^{2a} \int_{0}^{2b} (2a - u)(2b - v)$$

$$\cdot \exp\left[\delta^2 r(u, v)\right] \cos \beta_x u \, \cos \beta_y v \, du \, dv. \tag{36}$$

By expanding the exponential function in a power series, equation (36) can be divided into two parts corresponding to the coherent and incoherent parts of gain, as follows

$$\langle I \rangle_{\mathrm{av}} = I_c + I_{inc} \tag{37}$$

where

$$I_c = A^2 \exp\left(-\delta^2\right)\left(\frac{\sin \beta_x a}{\beta_x a} \frac{\sin \beta_y b}{\beta_y b}\right)^2 \tag{38}$$

and

$$I_{inc} = 4A \sum_{n=1}^{\infty} \int_0^{2a} \int_0^{2b} \frac{[\delta^2 r(u, v)]^n}{n!} \cos \beta_x u \, \cos \beta_y v \, dv \, du - \Delta I_{inc} \quad (39)$$

where

$$\Delta I_{inc} = 4 \exp(-\delta^2) \sum_{n=1}^{\infty} \int_0^{2a} \int_0^{2b} [2(bu + av) - uv] \frac{[\delta^2 r(u, v)]^n}{n!}$$

$$\cdot \cos \beta_x u \, \cos \beta_y v \, du \, dv \quad (40)$$

and $A = 4ab$ is the aperture area.

The coherent part of the gain is the same as the antenna gain in the absence of surface deviations but multiplied by $\exp(-\delta^2)$. This follows from equation (28) by expanding the exponential function which contains the correlation function in a power series.

To obtain an estimate for the on-axis gain, equation (39) is evaluated for $\beta_x = \beta_y = 0$, and for a gaussian correlation function with

$$r(u, v) = \exp\left(- \frac{u^2 + v^2}{c^2}\right) \quad (41)$$

where $c$ is the correlation distance.

On-axis

$$I_{inc}(0, 0) = \pi A c^2 \sum_{n=1}^{\infty} \frac{(\delta^2)^n}{n! \, n} \left[ 1 - \frac{c}{2(\pi n)^{\frac{1}{2}}} \left(\frac{1}{a} + \frac{1}{b}\right) + \frac{c^2}{\pi A n} \right]. \quad (42)$$

In equation (42) terms of order $\exp[-n(2a/c)^2]$ and $\exp[-n(2b/c)^2]$ were neglected.

By extending the limits of integrations in equation (39) to $\infty$, the integration of the first part of this equation can be performed and gives equation (11).

APPENDIX B

*On-Axis Gain for Circular Aperture Antennas*

For circular aperture antennas the on-axis gain for uniform illumination and a gaussian correlation function is obtained from equation (8) by expanding the exponential function and performing the integrations for the $n = 0$ term, and the integration with respect to the azimuthal coordinates for the remaining terms, resulting in

$$\langle G(0, 0) \rangle_{\mathrm{av}} = \left(\frac{\pi D}{\lambda}\right)^2 \exp(-\delta^2) + \left(\frac{8\pi}{\lambda D}\right)^2 \sum_{n=1}^{\infty} \frac{(\delta^2)^n}{n!} I_{cn} \quad (43)$$

where $D$ is aperture diameter, and

$$I_{cn} = \int_0^{D/2} \int_0^{D/2} \exp\left[-\frac{n(\rho^2 + \rho_1)}{c^2}\right] I_0\left(\frac{2n\rho\rho_1}{c^2}\right)\rho \, d\rho\rho_1 \, d\rho_1 . \qquad (44)$$

$I_o$ = Modified Bessel function of order zero.

The two integrations in equation (44) can be performed either with the aid of the $Q(y, a_n)$ function defined by[11]

$$Q(y, a_n) = \int_{a_n}^\infty \exp\left(-\frac{x^2 + y^2}{2}\right) I_0(xy)x \, dx \qquad (45)$$

or by means of recently evaluated integrals of products of Bessel functions.[12] Let us use the former method. Let

$$x = (2n)^{\frac{1}{2}}\rho/c \qquad (46)$$

$$y = (2n)^{\frac{1}{2}}\rho_1/c \qquad (47)$$

$$a_n = D/c(n/2)^{\frac{1}{2}}. \qquad (48)$$

With equations (45) through (48), equation (44) can be written

$$I_{cn} = \left(\frac{c^2}{2n}\right)^2 \int_0^{a_n} [1 - Q(y, a_n)]y \, dy. \qquad (49)$$

Integrating by parts results in

$$I_{cn} = \left(\frac{c^2}{2n}\right)\left\{\frac{a_n^2}{2} [1 - Q(a_n, a_n)] + \int_0^{a_n} \frac{y^2}{2} \frac{\partial Q}{\partial y} \, dy\right\}. \qquad (50)$$

The derivative in equation (50) can be expressed as

$$\frac{\partial Q}{\partial y} = a_n \exp\left(-\frac{a_n^2 + y^2}{2}\right) I_1(a_n y). \qquad (51)$$

Equation (51) is readily derived from (45) and the following integral[13]

$$\int_0^\infty \exp(-t^2/2) J_v(xt) J_v(yt)t \, dt = \exp\left(-\frac{x^2 + y^2}{2}\right) I_v(xy) \qquad (52)$$

where $J_v$ is a Bessel function of order $v$. Substituting equation (51) into (50) and integrating by parts yields the result

$$I_{cn} = \left(\frac{c^2}{2n}\right)^2\left\{\frac{a_n^2}{2} [1 - Q(a_n, a_n) - \exp(-a_n^2)I_1(a_n^2)]\right.$$

$$\left. + \frac{a_n}{2} \int_0^{a_n} \exp[-\frac{1}{2}(a_n^2 + y^2)] \frac{d}{dy} [yI_1(a_n y)] \, dy\right\}. \qquad (53)$$

1650

Using the relation

$$\frac{d}{dy}\left[yI_1(a_ny)\right] = a_ny\,I_0(a_ny) \tag{54}$$

and the definition of the $Q$ function, equation (45), yields

$$I_{cn} = \left(\frac{c^2}{2n}\right)\left\{a_n^2[1 - Q(a_n,a_n)] - \frac{a_n^2}{2}\exp(-a_n^2)I_1(a_n^2)\right\}. \tag{55}$$

To evaluate $Q(a_n, a_n)$, the following relation, readily derived from (45) and (52), is used.

$$Q(\alpha, \beta) + Q(\beta, \alpha) = 2 + \int_0^\infty \exp(-\tfrac{1}{2}t^2)\left\{\frac{d}{dt}[J_0(\alpha t)J_0(\beta t)]\right\}dt. \tag{56}$$

Integrating (56) by parts and using (52), gives the known relation

$$Q(y, x) + Q(x, y) = 1 + \exp\left(-\frac{x^2 + y^2}{2}\right)I_0(xy). \tag{57}$$

With equations (55), (57), and (48), (44) is given by

$$I_{cn} = \left(\frac{cD}{4}\right)^2 \frac{1}{n}\left\{1 - \exp\left[-\frac{n}{2}\left(\frac{D}{c}\right)^2\right]\left[I_0\left(\frac{n}{2}\frac{D^2}{c^2}\right) + I_1\left(\frac{n}{2}\frac{D^2}{c^2}\right)\right]\right\}. \tag{58}$$

The gain on axis (43) can therefore be written by using equation (58) as:

$$G(0, 0) = \left(\frac{D}{4\epsilon}\right)^2\left[\delta^2 \exp(-\delta^2) + \left(\frac{2c}{D}\right)^2 \delta^2 \exp(-\delta^2) \sum_{n=1}^\infty \frac{\delta^2 n}{n!\,n}[1 - \Delta_n]\right] \tag{59}$$

with

$$\Delta_n = \exp\left[-\frac{n}{2}\left(\frac{D}{c}\right)^2\right]\left[I_0\left(\frac{n}{2}\frac{D^2}{c^2}\right) + I_1\left(\frac{n}{2}\frac{D^2}{c^2}\right)\right]. \tag{60}$$

For $n/2(D/c)^2 \gg 1$, when the modified Bessel functions can be approximated by the first terms of the asymptotic series, $\Delta_n$ is then given by equation (15).

REFERENCES

1. Ruze, J., "The Effect of Aperture Errors on the Antenna Radiation Pattern," Suppl. al Nuovo Cimento, 9, No. 3 (1952), pp. 364–380.
2. Ruze, J., "Antenna Tolerance Theory—A Review," Proc. IEEE, 54, No. 4 (April 1966), pp. 633–640.
3. Scheffler, H., "Über die Genauigkeitsforderungen bei der Herstellung Opti-

scher Flachen fur Astronomiche Teleskope," Z. Astrophys (German), *55,*
No. 1 (1962), pp. 1-20.

4. Chaevski, E. V., "Energy Characteristics of the Field Scattered by a Rough
Surface," Izvestiya VUZ. Radiofizika, *8,* No. 6 (1965), pp. 1128-1133.

5. Silver, S., *Microwave Antenna Theory and Design,* McGraw-Hill, New York,
1949.

6. Hoffman, W. C., "Scattering of Electromagnetic Waves from A Random
Surface," Quart. Appl. Math., *13,* No. 3 (1955), pp. 291-304.

7. A. Erdelyi et al, *Higher Transcendental Functions,* Vol. 2, McGraw-Hill,
1953, pp. 143-144.

8. Copson, E. T., *Asymptotic Expansions,* Cambridge: The University Press,
1965, pp. 19-21 and pp. 58-60.

9. Barnes, E. W., "On the Asymptotic Expansion of Integral Functions Defined
by Taylor Series," Phil. Trans. Royal Soc. London, Series A, *206* (1906),
pp. 249-297.

10. Ford, W. B., *Studies on Divergent Series and Asymptotic Developments,*
New York: Chelsea Publishing Co., 1960, pp. 269-272.

11. Luke, Y. L., *Integrals of Bessels Functions* New York: McGraw-Hill, 1962,
pp. 271-283.

12. Ford, F. A. J., "Some Infinite Integrals Involving Products of Bessel Func-
tions," J. London Math. Soc., *41,* No. 164 (October 1966), pp. 728-730.

13. Watson, G. N., *A Treatise on the Theory of Bessel Functions,* 2nd ed., Cam-
bridge: The University Press, 1962, pp. 393-395.

# Injection-Locked-Oscillator FM Receiver Analysis

By C. L. RUTHROFF

(Manuscript received April 18, 1968)

*The major components of an injection-locked-oscillator FM receiver are a linear mixer and a van der Pol type of negative resistance oscillator in a phase-locked configuration. In this paper the nonlinear differential equation describing the receiver output is solved and explicit expressions obtained for the output signal, noise, and controlling second and third order distortions. The input signal carrier is both amplitude and frequency modulated. Signal-to-distortion ratios have been computed and are presented for the case of a noise modulated FM input signal. The results indicate that excellent performance may be expected of such a receiver.*

## I. INTRODUCTION

It is well known that a conventional phase-locked loop can be used as a frequency modulation receiver.[1-3] It is perhaps less well known that the locking performance of the phase-locked loop and the injection-locked oscillator are described by the same differential equation.[4] These two facts suggest that an FM receiver using an injection-locked oscillator is possible. It is. And such a receiver is described here.

The principle of operation of the two receivers is the same but there are important practical differences. The baseband bandwidth of the phase-locked-loop receiver is limited by delay in the feedback loop to frequencies of about 1 MHz. The baseband bandwidth of the injection-locked-oscillator FM receiver can be as large as half the locking bandwidth of the injection-locked oscillator. With existing solid state oscillators such as the tunnel diode, locking has been achieved with bandwidths in excess of 200 MHz,[5] indicating that operation with basebands of about 100 MHz is possible.

This type of receiver is not used in present day systems and there has been little or no interest in it for about 20 years.[6-10] Woodyard[7]

is credited by Edson[10] with the first explicit receiver operating on the locking principle but the beginnings go back to Appleton in 1922.[6] Beers[8] and Bradley[9] reported excellent measured performance in configurations using vacuum tube oscillators; in the light of these results it is surprising that interest has flagged in recent years.

The existence of solid state oscillating devices such as the tunnel diode, the avalanche diode, and the Gunn diode is sufficient cause for renewed interest in injection-locked-oscillator FM receivers. They are especially attractive at the higher microwave and millimeter wave frequencies, where conventional receivers are difficult to build. This paper describes the principle of operation of a receiver configuration suitable for dominant mode transmission lines. Applications at optical frequencies are also of interest.

The distortion analysis which follows requires a mathematical description of the locking behavior of an oscillator. For this purpose the van der Pol sine wave oscillator model is used; there is abundant evidence in the literature that all oscillators which have nearly sinusoidal outputs are adequately described by the van der Pol model. The receiver output is derived in the form of a nonlinear differential equation. The solution of the equation gives the output signal and distortions explicitly in terms of the frequency and amplitude modulation on the receiver input carrier.

An example of receiver performance is computed in some detail for the case of a carrier modulated with a band of gaussian noise. Such a modulation signal is often used to simulate the output of a multichannel telephone multiplex terminal. The receiver input signal is corrupted by additive noise and the distortions resulting from the effects of envelope noise are computed in addition to those caused by signal-sensitive nonlinearities inherent in the demodulation process.

## II. DESCRIPTION AND ANALYSIS

The receiver is shown in block form in Fig. 1. This circuit configuration is convenient for description and is suitable for the microwave frequency range. Many other configurations are, of course, possible.

Let the input signal be a carrier, modulated in amplitude and frequency:

$$i_1(t) = I(t) \sin [pt + \theta(t)]. \tag{1}$$

The frequency modulation is $d\theta(t)/dt$; the envelope $I(t)$ is usually nearly constant with a small variable part representing noise or other
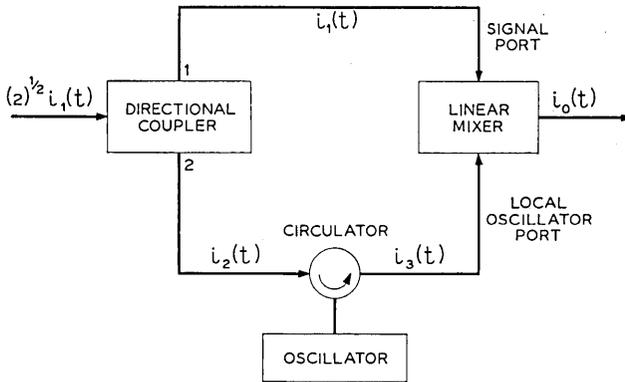
Fig. 1 — Injection-locked-oscillator FM receiver.

undesired amplitude modulation. Let

$$I(t) = I_0 + I_1(t), \qquad I_1(t) \ll I_0. \tag{2}$$

The function of the directional coupler in Fig. 1 is to divide the signal into two parts which are 90° out of phase. Letting the amplitudes be equal, the output of port 1 is given by (1), and the output of port 2 is

$$i_2(t) = I(t) \cos [pt + \theta(t)]. \tag{3}$$

The output of the injection-locked oscillator is

$$i_3(t) = I_3(t) \cos [pt + \theta(t) - \varphi(t)], \tag{4}$$

where $\varphi(t)$ is the phase tracking error. The receiver output is contained in $\varphi(t)$ and is discussed in detail later. The envelope variations in (3) are reduced in the passage of $i_2(t)$ through the oscillator. The output of the oscillator is used as the local oscillator for the linear mixer of Fig. 1, and since small envelope variations on the local oscillator port do not appear in the output, the envelope of (4) can be regarded constant, that is,

$$i_3(t) = I_3 \cos [pt + \theta(t) - \varphi(t)]. \tag{5}$$

The envelope variations on the input to the linear mixer signal port are not suppressed however, and the low frequency output of the mixer, from (1) and (5), is

$$i_0(t) = MI(t) \sin \varphi(t), \tag{6}$$

where $M$ includes the linear mixer conversion constant. The receiver FM output is contained in the term $\sin \varphi(t)$.

The differential equation describing the locking behavior of the van der Pol sine wave oscillator has been derived in many places.[11, 12] With the input signal of (3) it has the following form.

$$\varphi'(t) = \theta'(t) + (p - \omega_0) - \Delta(t) \sin \varphi(t), \tag{7}$$

where: the prime indicates differentiation with respect to the argument,

$$\Delta(t) = \frac{I(t)}{I_3 \Phi'(\omega_0)},$$

$\omega_0$ is the natural resonant frequency of the oscillator,

$| p - \omega_0 | \ll \omega_0$, and

$\Phi'(\omega_0)$ is the phase slope of the passive oscillator circuit at the resonant frequency.

Now, the injection-locked oscillator locking bandwidth is $2\Delta_0$ where $\Delta_0 = I_0/[I_3 \Phi'(\omega_0)]$, so the expression for $\Delta(t)$ becomes

$$\Delta(t) = \frac{\Delta_0}{I_0} I(t). \tag{8}$$

A solution of (7) has been found by an iteration procedure and substituted into (6) to obtain the receiver output.

$$
\begin{aligned}
i_0(t) = \frac{MI_0}{\Delta_0} \Bigg\{ &[\theta'(t) + p - \omega_0] - \frac{\theta''(t)}{\Delta(t)} \\
&+ \frac{[\theta'''(t) + \Delta'(t)(\theta'(t) + p - \omega_0)]}{\Delta(t)^2} \\
&- \frac{[\Delta''(t)(\theta'(t) + p - \omega_0) + \theta''(t)(\theta'(t) + p - \omega_0)^2/2]}{\Delta(t)^3} \\
&+ \frac{[\theta''''(t) + 3\theta''(t)\,\Delta'(t)]}{\Delta(t)^3} + \cdots \Bigg\}.
\end{aligned}
\tag{9}
$$

When the locking bandwidth $2\Delta_0$ is much larger than the baseband bandwidth, that is, when $\Delta_0 \gg \theta'(t)_{\max}$, the distortion will be small and the series (9) will converge rapidly. For this case the first few distortion terms will dominate. Expression (9) can be rearranged to identify these terms. Let $p = \omega_0$, and let

$$\Delta(t) = \Delta_0 + \Delta_1(t). \tag{10}$$

From (2) and (8) it follows that $\Delta_1(t) \ll \Delta_0$. The following relations will also be used.

$$1/\Delta(t) = \frac{1 - \Delta_1(t)/\Delta_0 + (\Delta_1(t)/\Delta_0)^2 + \cdots}{\Delta_0} \qquad (11)$$

$$\theta'(t - 1/\Delta_0) = \theta'(t) - \frac{\theta''(t)}{\Delta_0} + \frac{\theta'''(t)}{2! \, \Delta_0^2} - \frac{\theta''''(t)}{3! \, \Delta_0^3} + \cdots . \qquad (12)$$

Substitution of (10), (11), and (12) into (9) transforms the receiver output to

$$i_0(t) = MI_0/\Delta_0 \left\{ \theta'\left(t - \frac{1}{\Delta_0}\right) + \frac{\theta'''(t)}{2 \, \Delta_0^2} - \frac{5\theta''''(t)}{6 \, \Delta_0^3} + \cdots \right.$$
$$\left. + \frac{[\Delta_1(t)\theta'(t)]'}{\Delta_0^2} + \cdots - \frac{[\theta'^3(t)]'}{6 \, \Delta_0^3} + \cdots \right\}. \qquad (13)$$

The first term is the frequency modulation on the input signal, delayed by $1/\Delta_0$ seconds. The second and third terms are linear distortion and correspond to video roll-off. The fourth and fifth terms are second and third order nonlinear distortions, respectively. Notice that the second order distortion requires both amplitude and frequency modulation—if there is no AM there is no second order distortion. The third order distortion is independent of any AM on the input signal.

### III. EXAMPLE—NOISE MODULATION

Let the modulating signal, $\theta_s'(t)$, be a flat band of gaussian noise of bandwidth $O - W$ with a two-sided spectral density

$S_{\theta's}(f) = \sigma^2/2W$ watts per cycle of bandwidth.

$$-W \leqq f \leqq W. \qquad (14)$$

Assuming an FM transmitter sensitivity of 1 Hz per volt, the parameter $\sigma$ is also the rms frequency deviation. The notation $S_h(f)$ is used to indicate spectral density of $h(t)$ and is, of course, a function of frequency.

Normally the input signal to a receiver is contaminated by noise. Let this noise have a constant spectral density of $n_0$ watts per cycle of bandwidth. The input signal is $(2)^{\frac{1}{2}} i_1(t)$ where

$$i_1(t) = (2C)^{\frac{1}{2}} \sin [\omega_0 t + \theta_s(t)] + n(t), \qquad (15)$$

and $C$ is the input carrier power. The additive noise, $n(t)$, is narrow-

band gaussian noise and can be represented as follows.[13]

$$n(t) = x_c(t) \sin [\omega_0 t + \theta_s(t)] - x_s(t) \cos [\omega_0 t + \theta_s(t)] \qquad (16)$$

where $x_c(t)$ and $x_s(t)$ are gaussian random variables with zero mean and variances equal to the variance of $n(t)$. Using (16) and assuming a large carrier-to-noise ratio, that is, $(2C)^{\frac{1}{2}} \gg n(t)$, (15) becomes

$$i(t) \doteq [(2C)^{\frac{1}{2}} + x_c(t)] \sin [\omega_0\, t + \theta_s(t) - x_s(t)/(2C)^{\frac{1}{2}}]. \qquad (17)$$

From (1), (2), and (8) we have

$$\theta'(t) = \theta'_s(t) - x'_s(t)/(2C)^{\frac{1}{2}}, \qquad (18)$$

and

$$\Delta_1(t) = [\Delta_0/(2C)^{\frac{1}{2}}]x_c(t). \qquad (19)$$

Expression (18) contains the desired output signal $\theta'_s(t)$ and a noise term which represents the frequency modulation caused by the additive noise. The signal-to-noise relationships of FM receivers have been discussed widely in the literature;[14] since the noise is small relative to the signal, that is, $x'_s(t)/(2C)^{\frac{1}{2}} \ll \theta'_s(t)$, it will have a negligible effect on the distortion and need not be considered further.

The quantity of interest in broadband radio systems is the ratio of signal spectral density to distortion spectral density. From (13) the second and third order distortions are

$$D_2(t) = \frac{[\Delta_1(t)\theta'_s(t)]'}{\Delta_0^2} , \qquad (20)$$

and

$$D_3(t) = \frac{[\theta_s'^3(t)]'}{6\, \Delta_0^3}. \qquad (21)$$

The distortion spectra are

$$S_{D_2}(f) = \frac{1}{\Delta_0^4} S_{[\Delta_1\theta'_s]'}(f) = \frac{f^2}{2C\, \Delta_0^2} S_{x_c\theta'_s}(f), \qquad (22)$$

and

$$S_{D_3}(f) = \frac{1}{36\, \Delta_0^6} S_{[\theta'_s{}^3]'}(f) = \frac{f^2}{36\, \Delta_0^6} S_{\theta'_s{}^3}(f). \qquad (23)$$

To evaluate $S_{D_2}(f)$ notice that $x_c(t)$ and $\theta'_s(t)$ are statistically independent. The spectrum of the product can then be written

$$S_{x_c\theta'_s}(f) = S_{x_c}(f) * S_{\theta'_s}(f),$$

where the asterisk means convolution. Performing the convolution, and letting $B > 4W$, where $B$ is the RF bandwidth of the FM signal, the output distortion spectral density in the baseband is

$$S_{D_2}(f) = \frac{f^2}{2C \, \Delta_0^2} \, S_{x_e}(f) * S_{\theta'_s}(f)$$

$$= \frac{f^2 \sigma^2 n_0}{2 \, \Delta_0^2 C}, \qquad 0 \leqq f \leqq W, \qquad (24)$$

and the signal-to-distortion ratio is

$$\frac{S_{\theta'_s}(f)}{S_{D_2}(f)} = \left(\frac{C}{n_0 W}\right)\left(\frac{\Delta_0}{f}\right)^2, \qquad 0 \leqq f \leqq W. \qquad (25)$$

The evaluation of $S_{D_3}(f)$ requires an expression for $S_{\theta'_s}$ where $\theta'_s(t)$ is the input signal (14). From Hatch,[15]

$$S_{\theta'_s}(f) = \frac{9\sigma^6}{4W}\left[1 - \frac{1}{3}\left(\frac{f}{W}\right)^2\right], \qquad 0 \leqq f \leqq W, \qquad (26)$$

$$S_{D_3}(f) = \frac{f^2 \sigma^6}{16W \, \Delta_0^6}\left[1 - \frac{1}{3}\left(\frac{f}{W}\right)^2\right], \qquad 0 \leqq f \leqq W, \qquad (27)$$

and

$$\frac{S_{\theta'_s}(f)}{S_{D_3}(f)} = \frac{8 \, \Delta_0^6}{\sigma^4 f^2\left[1 - \frac{1}{3}\left(\frac{f}{W}\right)^2\right]}, \qquad 0 \leqq f \leqq W. \qquad (28)$$

For a locking bandwidth $2\Delta_o = 200$ MHz, a baseband $W = 5$ MHz, and an rms frequency deviation $\sigma = 15$ MHz, the signal-to-distortion ratios in the worst message channel, that is, at $f = W$ are:

$$\left.\frac{S_{\theta'_s}(f)}{S_{D_2}(f)}\right|_{f=W} = 52.2 \text{ Db} \qquad (29)$$

$$\left.\frac{S_{\theta'_s}(f)}{S_{D_3}(f)}\right|_{f=W} = 69.7 \text{ Db}. \qquad (30)$$

In computing (29), the carrier-to-noise ratio in the RF band $B$ was assumed to be near threshold, that is, $C/(n_o B) = 16$ (12 dB) where

$$B = 2W\left(1 + \frac{4\sigma}{W}\right) \qquad (31)$$

is the Carson bandwidth and $4\sigma$ is the peak frequency deviation. If

this receiver were used in a radio system with a fading margin of 30 dB the signal-to-distortion ratio for second order distortion would be 82.2 dB during periods of normal propagation and the performance would be limited by the third order distortion.

## IV. AUTOMATIC FREQUENCY CONTROL

If the natural resonant frequency of the oscillator, $\omega_0$, is not equal to the input signal carrier frequency $p$, the receiver output contains a direct current component. From (9),

$$i_{DC} = \frac{MI_0}{\Delta_0} (p - \omega_0). \tag{32}$$

This current gives the direction and magnitude of the frequency difference and, by using it to control a suitable oscillator parameter, the oscillator can be kept tuned automatically to the input carrier frequency.

## V. DISCUSSION

One may question whether $\Phi'(\omega)$ is really constant in the band of interest. If the oscillator circuit is a single resonance then the locking bandwidth is equal to the 3 dB bandwidth of the passive circuit divided by the current gain of the locked oscillator. For a 20 dB gain the circuit phase changes $\pm 4.5°$ over the locking band, therefore $\Phi'(\omega)$ is reasonably constant over bandwidths less than the locking bandwidth. Since $\Phi'(\omega)$ is the slope of a passive circuit, it can be made arbitrarily close to constant over the necessary frequency range by increasing the complexity of the circuit.

It would be nice to compare the theoretical performance of the injection-locked-oscillator receiver with the conventional FM receiver consisting of a limiter and a balanced discriminator. There is, however, no theoretical analysis of the conventional receiver comparable with the analysis presented here. Such an analysis can probably be done with the aid of recent work by Bedrosian and Rice,[16] although it would be difficult to describe analytically the amplitude-to-phase conversion in a real limiter and to take into account the nonlinearity of the envelope detectors which are used in the discriminator.

The situation with the injection-locked-oscillator receiver is different. The amplitude-to-phase conversion has been accounted for in the present analysis and practical linear mixers are certainly linear,

A most important question about the practical realization of the receiver is the extent to which real oscillators can be described by the van der Pol model. The evidence in the literature[5] answers the question favorably in that the locking equation (7) does indeed describe the locking behavior of tunnel diode and avalanche diode oscillators.

REFERENCES

1. Jaffee, R., and Rechtin, E., "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," I.R.E. Trans. Inform. Theory, *IT-1*, No. 1 (March 1955), pp. 66–76.
2. Gilchriest, C. E., "Application of a Phase-Locked Loop to Telemetry as a Discriminator or Tracking Filter," I.R.E. Trans. Telemetry and Remote Control, *TRC-4*, No. 1 (June 1958), pp. 20–35.
3. Moschytz, G. S., "Miniaturized RC Filters Using Phase-Locked Loop," B.S.T.J., *44*, No. 5, (May–June 1965), pp. 823–870.
4. Ruthroff, C. L., and Stover, H. L., unpublished work.
5. Shaw, R. C., and Stover, H. L., "Phase-Locked Avalanche Diode Oscillators," Proc. I.E.E.E., *54*, No. 4, (April 1966), pp. 710–711.
6. Appleton, E. V., "The Automatic Synchronization of Triode Oscillators," Proc. Cambridge Phil. Soc., *21*, Part 2 (1922–1923), pp. 231–248.
7. Woodyard, J. R., "Application of the Auto-Synchronized Oscillator to Frequency Demodulation," Proc. I.R.E., *25*, No. 5 (May 1937), pp. 610–619.
8. Beers, G. L., "A Frequency-Dividing Locked-In-Oscillator Frequency-Modulation Receiver," Proc. I.R.E., *32*, No. 12 (December 1944), pp. 730–737.
9. Bradley, W. E., "Single Stage FM Detector," Elec., *19*, No. 10 (October 1946), pp. 88–91.
10. Edson, W. A., *Vacuum-Tube Oscillators,* New York: John Wiley & Sons, 1953.
11. Khokhlov, R. V., "A Method of Analysis in the Theory of Sinusoidal Self-Oscillations," I.R.E. Trans. Circuit Theory, *CT-7*, No. 4, (December 1960), pp. 398–413.
12. Adler, R., "A Study of Locking Phenomena in Oscillators," Proc. I.R.E., *34*, No. 6 (June 1946), pp. 351–357.
13. Rice, S. O., "Mathematical Analysis of Random Noise," B.S.T.J., *23*, No. 3, (July 1944), pp. 282–332, and *24*, No. 1 (January 1945), pp. 46–156.
14. Rice, S. O., "Statistical Properties of a Sine-Wave Plus Random Noise," B.S.T.J., *27*, No. 1 (January 1948), pp. 109–157.
15. Hatch, R. W., unpublished work.
16. Bedrosian, E., and Rice, S. O., "Distortion and Crosstalk of Linearly Filtered, Angle-Modulated Signals," Proc. I.E.E.E., *56*, No. 1 (January 1968), pp. 2–13.

# Hybrid Digital Transmission Systems
## Part I: Joint Optimization of
## Analog and Digital Repeaters

### By ROBERT W. CHANG and S. L. FREENY

*A hybrid digital transmission system consists of analog repeaters placed between digital repeaters. Joint optimization of the analog and digital repeaters is considered in this paper, using minimum mean-square error between the transmitted and received symbols as the performance criterion. A general hybrid system is considered. The joint optimization problem is solved in closed form for deterministic sampling under two usually satisfied conditions. From the results the minimum mean-square error and the optimum repeater characteristics can be computed for given system parameters. Timing error is also considered. From a general result, it is concluded that in many practical systems it is not only economical, but also optimum, to use identical analog repeaters, and that hybrid systems can be used for either digital or voice transmission with no compromise in theoretical performance.*

I. INTRODUCTION

It is customary in long-haul digital transmission systems to regenerate the digital signal at each point that gain is introduced into the system. This is not necessary, however, and in fact there are circumstances in which it is advantageous to do otherwise. One such circumstance occurs when multilevel pulses are being transmitted and the associated digital repeater* is too complicated and costly to be placed at every gain point. In this case there is merit in interspersing a number of analog repeaters between digital repeaters, even though the digital device must usually be complicated further by the introduction of automatic equalization to compensate for the misalignment which accrues over several analog links in tandem.

---

* A digital repeater is also called a regenerative repeater,[1] a reconstructive repeater, or a regenerator.

Part I of this study addresses itself to the problem of jointly op-
timizing the various filters contained in a combination analog-digital
or "hybrid" multilevel transmission system. The criterion used is
minimizing the mean-square error between transmitted and received
symbols. The system studied is general in that: ($i$) the repeater spac-
ing may be nonuniform and the transfer functions of the transmission
media may be different, ($ii$) the noises introduced by the repeaters
may not be white and each may have a different spectral density
($iii$) the analog repeaters are not constrained to be identical, and
($iv$) the repeater output power levels are not constrained to be the
same.

The mathematical model is formed in Section II. Results are sum-
marized in the concluding section. Interesting characteristics and
potentialities of hybrid cable systems are explored in Part II.[2]

## II. MATHEMATICAL MODEL

Figure 1 illustrates a general hybrid digital transmission system.
Information symbols $\{a_k\}$ are transmitted from one digital repeater
to the next through $L$ analog repeaters. The output network of the
sending digital repeater is referred to as the transmitting filter with
transfer function $B_0(f)$. The input network of the receiving digital
repeater is referred to as the receiving filter with transfer function
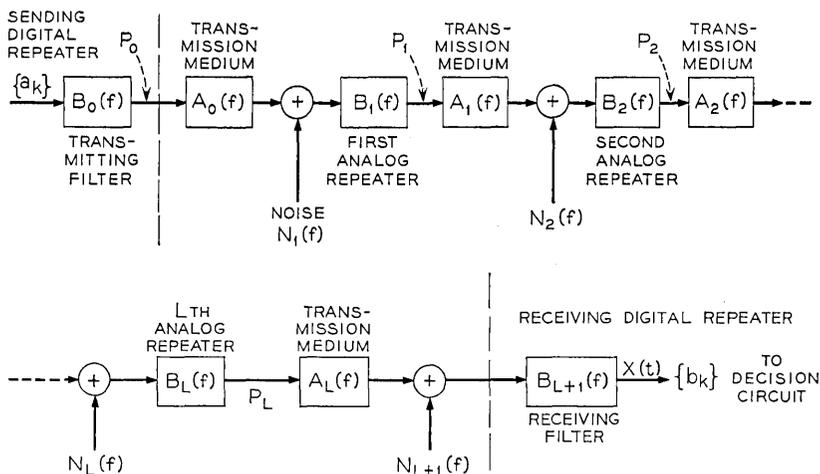


Fig. 1 — A general hybrid digital transmission system.

$B_{L+1}(f)$. The transfer function of the $i$th analog repeater is denoted $B_i(f)$, $i = 1, \ldots, L$. In this paper the analog repeaters are not constrained to be identical; hence, the ratio $B_i(f)/B_j(f)$ may be a function of frequency. The noise at the input of the $i$th analog repeater has a spectral density $N_i(f)$, $i = 1, \ldots, L$. The noise at the digital repeater input has a spectral density $N_{L+1}(f)$. Notice that $N_1(f), \ldots, N_{L+1}(f)$ may be all different and each may be a function of frequency.

As shown in Fig. 1, the transfer functions of the transmission media between the repeaters are denoted by $A_0(f)$, $A_1(f)$, ..., and $A_L(f)$. These transfer functions may be all different; hence, the repeater spacings may be nonuniform, and the transmission media may be different.

The average output signal power of the transmitting filter is constrained to be $P_0$. The average output signal power of the $i$th analog repeater is constrained to be $P_i$, $i = 1, \ldots, L$.

The information symbols $\{a_k\}$ are multilevel digits or real numbers. It is assumed that $\{a_k\}$ is stationary in the wide sense. The autocorrelation function is denoted by

$$m_k = E[a_l a_{l+k}], \qquad l, k = -\infty, \cdots, \infty.$$

Pulse amplitude modulation is considered. Let $1/T$ be the baud rate. The transmitting filter output is then

$$\sum_{k=-\infty}^{\infty} a_k s(t - kT)$$

where $s(t)$ is the impulse response of the transmitting filter.

As is well known, in linear PAM systems the receiving filter output, $X(t)$ in Fig. 1, is sampled sequentially at $T$-second intervals, and the $k$th time sample $b_k$ is used as an estimate of $a_k$. For analytical purposes, a constant time delay in the system may be neglected, and it may be assumed that $b_k$ is taken at

$$t = kT + \delta_k$$

where $\delta_k$ is the timing jitter.[3]

The system from the output of $B_0(f)$ to the input of $B_{L+1}(f)$ may be considered as a channel. For a given channel, Berger, Tufts, and Smith[4, 5] have considered methods for designing the transmitting and receiving filters for minimizing the mean-square error $E[(b_k - a_k)^2]$. By these methods the digital repeaters can be specified if the analog repeaters were given, and their output powers were not constrained.

In this paper, the analog repeaters are to be designed, and their output powers are constrained. We consider joint optimization of the analog and digital repeaters. For given $L$, $N_i(f)$, $i = 1, \ldots, L + 1$, $A_j(f)$, $j = 0, \ldots, L$, $\{m_k\}$, and $T$, we wish to design $B_k(t)$, $k = 0$, $1, \ldots, L + 1$, jointly to minimize the mean-square error

$$\mathcal{E} = E[(b_k - a_k)^2] \tag{1}$$

subject to the constraints of fixed repeater output signal power $P_i$, $i = 0, \cdots, L$.

The letter $E$ in (1) denotes the ensemble average taken over $\{a_k\}$, $\delta_k$, and the noise. The Fourier transform of the probability density function of $\delta_k$ is denoted by $P(f)$. The notation "*" denotes a complex conjugate and "$|\cdot|$" denotes a magnitude.

By a well known method,[4] the mean-square error in (1) can be expressed as

$$
\begin{aligned}
\mathcal{E} = m_0 &+ \int_{-\infty}^{\infty} M(f) \left[ \prod_{i=0}^{L+1} A_i^*(f) B_i^*(f) \right] \frac{1}{T} \\
&\cdot \sum_{k=-\infty}^{\infty} \left\{ P\left(\frac{k}{T}\right) \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right] \right\} df \\
&- \int_{-\infty}^{\infty} 2M(f) \left[ \prod_{i=0}^{L+1} A_i^*(f) B_i^*(f) \right] P(f)\, df \\
&+ \sum_{l=0}^{L+1} \int_{-\infty}^{\infty} N_l(f) \left| \prod_{i=l}^{L+1} A_i(f) B_i(f) \right|^2 df
\end{aligned} \tag{2}
$$

where

$$M(f) = m_0 + 2 \sum_{k=1}^{\infty} m_k \cos 2\pi f k T$$

is the spectral density of the stationary, random message sequence $\{a_k\}$.

By introducing the dummy variable $A_{-1}(f) = 1$ for all $f$, we can write the repeater average output signal powers all in the same form as

$$P_l = \frac{1}{T} \int_{-\infty}^{\infty} M(f) \left| \prod_{i=-1}^{l-1} A_i(f) \cdot \prod_{j=0}^{l} B_j(f) \right|^2 df, \quad l = 0, 1, \cdots, L. \tag{3}$$

III. NECESSARY AND SUFFICIENT CONDITIONS

Necessary and sufficient conditions for $B_n(f)$, $n = 0, 1, \cdots, L + 1$, to minimize the mean-square error $\mathcal{E}$ can be derived by using the

standard techniques of the calculus of variations. These conditions are rather lengthy. In order to conserve space and to facilitate the following manipulations, we write these conditions in the same form. To do so we use the dummy variables

$$A_{-1}(f) = 1 \quad \text{for all} \quad f$$

$$A_{L+1}(f) = 1 \quad \text{for all} \quad f$$

$$\lambda_{L+1} = 0.$$

Then it can be shown that the necessary and sufficient condition for $B_n(f)$, $n = 0, 1, \cdots, L + 1$, to minimize the mean-square error $\varepsilon$ subject to the power constraints in (3) is

$$M(f)\left[\prod_{i=0}^{L+1} A_i^*(f)\right]\left[\prod_{\substack{j=0 \\ j \neq n}}^{L+1} B_j^*(f)\right]\frac{1}{T}$$

$$\cdot \sum_{k=-\infty}^{\infty} \left\{P\left(\frac{k}{T}\right)\left[\prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right)B_i\left(f - \frac{k}{T}\right)\right]\right\}$$

$$- M(f)\left[\prod_{i=0}^{L+1} A_i^*(f)\right]\left[\prod_{\substack{j=0 \\ j \neq n}}^{L+1} B_j^*(f)\right]P(f)$$

$$+ \sum_{l=0}^{n}\left\{N_l(f)\left|\prod_{i=l}^{L+1} A_i(f) \cdot \prod_{\substack{j=l \\ j \neq n}}^{L+1} B_j(f)\right|^2 B_n(f)\right\}$$

$$+ \sum_{l=n}^{L+1}\left\{\lambda_l \frac{1}{T} M(f)\left|\prod_{i=-1}^{l-1} A_i(f) \cdot \prod_{\substack{j=0 \\ j \neq n}}^{l} B_j(f)\right|^2 B_n(f)\right\} = 0 \quad \text{for all } f \quad (4)$$

where $\lambda_l$, $l = 0, \ldots, L$, are Lagrange multipliers. The definition

$$\left|\prod_{\substack{j=k \\ j \neq k}}^{k} B_j(f)\right|^2 = 1$$

is used in (4).

In the following sections, we consider the problem of determining the optimum $B_n(f)$, $n = 0, 1, \ldots, L + 1$, from equations (2) to (4).

We can eliminate a trivial case first. In some correlation schemes (such as duobinary) $M(f)$ may be zero at some frequencies. It can be shown that $B_{L+1}(f)$ must be zero at the frequencies where $M(f) = 0$. Furthermore, $B_n(f)$, $n = 0, 1, \cdots, L$, can be arbitrarily chosen at these frequencies without affecting the mean-square error $\varepsilon$. In practice, they

may be chosen so that their amplitude and phases are continuous at these frequencies.

In the following sections $M(f) \neq 0$ is assumed. Furthermore, as is always the case in practice, $N_i(f)$, $A_i(f)$, $P_l$, and $T$ are assumed to be nonzero, finite quantities.

## IV. GENERAL RESULTS

By multiplying equation (4) by $B_n^*(f)$ one obtains

$$\sum_{l=0}^{n} \left\{ N_l(f) \left| \prod_{i=l}^{L+1} A_i(f) \right|^2 \left| \prod_{j=l}^{L+1} B_j(f) \right|^2 \right\}$$

$$+ \sum_{l=n}^{L+1} \left\{ \lambda_l \frac{1}{T} M(f) \left| \prod_{i=-1}^{l-1} A_i(f) \right|^2 \left| \prod_{j=0}^{l} B_j(f) \right|^2 \right\} = \zeta(f)$$

$$n = 0, 1, \cdots, L+1 \qquad (5)$$

where

$$\zeta(f) = M(f) \left[ \prod_{i=0}^{L+1} A_i^*(f) \right] \left[ \prod_{j=0}^{L+1} B_j^*(f) \right]$$

$$\cdot \left\{ P(f) - \frac{1}{T} \sum_{k=-\infty}^{\infty} P\left(\frac{k}{T}\right) \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right] \right\}$$

is not a function of $n$. It can be shown that we may use equation (5) instead of (4) without changing the solutions.

Letting $n = m$ and $m + 1$ by turns in equation (5), one obtains two equations. Subtracting the latter from the former gives

$$\lambda_m \frac{1}{T} M(f) \left| \prod_{i=-1}^{m-1} A_i(f) \right|^2 \left| \prod_{j=0}^{m} B_j(f) \right|^2$$

$$= N_{m+1}(f) \left| \prod_{i=m+1}^{L+1} A_i(f) \right|^2 \left| \prod_{j=m+1}^{L+1} B_j(f) \right|^2, \qquad m = 0, 1, \cdots, L. \quad (6)$$

Since the right-hand side of equation (6) cannot be zero for all $f$, one has $\lambda_m > 0$, $m = 0, \ldots, L$. From equation (6)

$$\frac{\lambda_{h+1} |A_h(f)|^2 |B_{h+1}(f)|^2}{\lambda_h} = \frac{N_{h+2}(f)}{N_{h+1}(f) |A_{h+1}(f)|^2 |B_{h+1}(f)|^2} \cdot$$

$$h = 0, 1, \cdots, L - 1. \qquad (7)$$

Equation (7) is equivalent to

$$| B_j(f) |^2 = \left[ \frac{\lambda_{j-1}}{\lambda_j} \cdot \frac{N_{j+1}(f)}{N_j(f)} \right]^{\frac{1}{2}} \frac{1}{| A_{j-1}(f) | | A_j(f) |}$$

$$j = 1, 2, \cdots, L. \qquad (8)$$

It can be shown that, regardless of the values of $m$ in equation (6), substituting equation (8) into equation (6) gives

$$\frac{1}{T} M(f) | A_0(f) | \left[ \frac{\lambda_0}{N_1(f)} \right]^{\frac{1}{2}} | B_0(f) |^2$$

$$= | A_L(f) | \left[ \frac{N_{L+1}(f)}{\lambda_L} \right]^{\frac{1}{2}} | B_{L+1}(f) |^2. \qquad (9)$$

From equations (8) and (9) one gets

$$\left| \prod_{i=0}^{L+1} A_i(f) B_i(f) \right| = \left[ \frac{\lambda_0 M(f)}{T N_1(f)} \right]^{\frac{1}{2}} | A_0(f) | | B_0(f) |^2. \qquad (10)$$

Let us define $\theta(f)$ to be the phase of $[ \prod_{i=0}^{L+1} A_i(f) B_i(f)]$, that is,

$$\prod_{i=0}^{L+1} A_i(f) B_i(f) = \left| \prod_{i=0}^{L+1} A_i(f) B_i(f) \right| e^{-i\theta(f)} \qquad (11)$$

Substituting equations (8), (10), and (11) into equation (5) and setting $n = 0$, one obtains after a few steps

$$\left[ \frac{\lambda_0 M(f)}{T N_1(f)} \right]^{\frac{1}{2}} | A_0(f) | | B_0(f) |^2 e^{i\theta(f)}$$

$$\cdot \left\{ \begin{aligned} \frac{1}{T} \sum_{k=-\infty}^{\infty} P\left(\frac{k}{T}\right) \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right] - P(f) \\ + \sum_{l=0}^{L} \left[ \frac{\lambda_l N_{l+1}(f)}{T M(f)} \right]^{\frac{1}{2}} \frac{1}{| A_l(f) |} e^{-i\theta(f)} \end{aligned} \right\} = 0. \qquad (12)$$

We have shown that the optimum digital and analog repeaters must satisfy the $L + 2$ equations in (8), (9), and (12). Some discussion is in order.

Let us refer to the frequencies at which $B_0(f) \neq 0$ as the transmission band. There is no signal transmitted outside this band. Clearly, the analog repeaters may have arbitrary amplitude characteristics outside the transmission band. Furthermore, the analog repeaters may have arbitrary phase characteristics at all frequencies.* Therefore, it is only

---

* It is seen from equation (2) that the mean-square error depends on the over-all phase characteristic of the system, but not on how the over-all phase is distributed among the repeaters. Thus, the analog repeaters may have arbitrary phases. The over-all phase can be adjusted at the digital repeaters.

necessary to specify their amplitude characteristics in the transmission band. Equation (8) shows that, in the transmission band, the $j$th analog repeater ($j = 1, \cdots, L$) should have an amplitude characteristic proportional to the function

$$\left[\frac{N_{j+1}(f)}{N_j(f)}\right]^{\frac{1}{4}}\left[\frac{1}{|A_{j-1}(f)A_j(f)|}\right]^{\frac{1}{2}}.$$

This simple specification holds regardless of the distribution of timing jitter and the spectral density $M(f)$ of the message sequence. Since the above function is in general well behaved, and since the phases can be arbitrary, the optimum analog repeaters can be closely realized.

For brevity, we say that several functions are similar when they differ only by multiplicative constants. In practice, the repeater noise spectral densities may be similar, and the transmission media may have similar transfer functions. In such cases, equation (8) shows that the amplitude characteristics of the analog repeaters are also similar. A rather important physical meaning of this is:

> The use of similar analog repeaters is not only an economical choice, but also an optimum one, for systems where the repeater noise spectral densities are similar and the transfer functions of the transmission media are similar.

It remains to determine the digital repeaters, the gain constants of the analog repeaters (the $L + 1$ LaGrange multipliers), and the transmission band. They must satisfy the $L + 2$ equations in (8), (9), and (12), and the $L + 1$ power constraints in (3). Furthermore, as will be shown, they must also satisfy some validity conditions because the repeater amplitude characteristics must be nonnegative. Since the solution depends on the distribution of timing jitter and since it is difficult to cover all cases in one paper, we shall consider only the important case of deterministic sampling (that is, the case in which timing jitter can be neglected) in the remainder of this paper.

V. DETERMINISTIC SAMPLING

From now on we consider deterministic sampling, that is, timing error $\delta_k = 0$, or

$$P(f) = 1, \quad \text{for all} \quad f. \tag{13}$$

Substituting (13) into (12) and noting that $\lambda_0$, $M(f)$, $T$, $N_1(f)$, and

$|A_0(f)|$ are nonzero, finite quantities (we consider $M(f) = 0$ in Section III), one can see that joint optimization requires either

$$| B_0(f) | = 0 \qquad (14)$$

or

$$\frac{1}{T} \sum_{k=-\infty}^{\infty} \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right] - 1$$

$$+ \sum_{l=0}^{L} \left[ \frac{\lambda_l N_{l+1}(f)}{TM(f)} \right]^{\frac{1}{2}} \frac{1}{| A_l(f) |} e^{-i\theta(f)} = 0. \qquad (15)$$

The first term in (15) is a periodic function in $f$ with period $1/T$, that is, it has the same value at the frequencies $f$ and $f - k/T$, where $k$ is any integer. Hence, a necessary condition for (15) to be satisfied at both $f$ and $f - k/T$ is that

$$\sum_{l=0}^{L} \left[ \frac{\lambda_l N_{l+1}(f)}{TM(f)} \right]^{\frac{1}{2}} \frac{1}{| A_l(f) |} e^{-i\theta(f)}$$

$$= \sum_{l=0}^{L} \left[ \frac{\lambda_l N_{l+1}\left(f - \frac{k}{T}\right)}{TM\left(f - \frac{k}{T}\right)} \right]^{\frac{1}{2}} \frac{1}{\left| A_l\left(f - \frac{k}{T}\right) \right|} e^{-i\theta(f-k/T)}. \qquad (16)$$

Since $M(f)$ is a periodic function in $f$ with period $1/T$, (16) is equivalent to the set of conditions

$$\alpha \cos \theta(f) = \beta \cos \theta\left(f - \frac{k}{T}\right) \qquad (17)$$

and

$$\alpha \sin \theta(f) = \beta \sin \theta\left(f - \frac{k}{T}\right), \qquad (18)$$

where

$$\alpha = \sum_{l=0}^{L} \frac{[\lambda_l N_{l+1}(f)]^{\frac{1}{2}}}{| A_l(f) |} \qquad (19)$$

$$\beta = \sum_{l=0}^{L} \frac{\left[\lambda_l N_{l+1}\left(f - \frac{k}{T}\right)\right]^{\frac{1}{2}}}{\left| A_l\left(f - \frac{k}{T}\right) \right|}. \qquad (20)$$

Noting that $\alpha$ and $\beta$ are positive, one can show that (17) and (18)

hold only if

$$\alpha = \beta \tag{21}$$

and

$$\theta\left(f - \frac{k}{T}\right) = \theta(f) + \nu\pi, \tag{22}$$

where

$$\nu = \text{any even integer, including zero.}$$

Thus, (15) can be satisfied at both $f$ and $f - k/T$ only if (21) and (22) are satisfied.

From (11), (10), and (22), one obtains

$$\sum_{k=-\infty}^{\infty} \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right]$$

$$= \left[ \frac{\lambda_0 M(f)}{T} \right]^{\frac{1}{2}} e^{-i\theta(f)} \sum_{k=-\infty}^{\infty} \frac{\left| A_0\left(f - \frac{k}{T}\right) \right|}{\left[ N_1\left(f - \frac{k}{T}\right) \right]^{\frac{1}{2}}} \left| B_0\left(f - \frac{k}{T}\right) \right|^2. \tag{23}$$

Substituting (23) into (15), one can show that

$$e^{-i\theta(f)} = 1. \tag{24}$$

Substituting (24) and (23) into (15) gives

$$M(f) \sum_{k=-\infty}^{\infty} \frac{\left| A_0\left(f - \frac{k}{T}\right) \right|}{\left[ N_1\left(f - \frac{k}{T}\right) \right]^{\frac{1}{2}}} \left| B_0\left(f - \frac{k}{T}\right) \right|^2$$

$$= T\left[ \frac{TM(f)}{\lambda_0} \right]^{\frac{1}{2}} - T \sum_{l=0}^{L} \left[ \frac{\lambda_l N_{l+1}(f)}{\lambda_0} \right]^{\frac{1}{2}} \frac{1}{|A_l(f)|} \tag{25a}$$

The optimum jitter-free system must satisfy either (14) or (25a). For convenience, we assume that for each $f$, (25a) is satisfied at $f - m/T$ for $m \, \varepsilon \, \mathfrak{R}_f$, where $\mathfrak{R}_f$ is a set of integers to be determined. The subscript $f$ indicates that $\mathfrak{R}_f$ may vary with $f$. Clearly (14) must be satisfied at $f - m/T$ for $m \notin \mathfrak{R}_f$, that is,

$$\left| B_0\left(f - \frac{m}{T}\right) \right| = 0, \quad m \notin \mathfrak{R}_f. \tag{26a}$$

If $\mathfrak{R}_f$ is an empty set, $|B_0(f - m/T)|$ must be zero for all $m$ (including $m = 0$).

Let us define a frequency set $\mathcal{I}$ as

$$\mathcal{I} = \left\{ f : -\frac{1}{2T} \leq f \leq \frac{1}{2T} \text{ and } \mathcal{R}_f \text{ is not an empty set} \right\}.$$

Clearly, (25a) is satisfied in the frequency set

$$\mathcal{F} = \{f : f = g - m/T, \quad g \, \varepsilon \, \mathcal{I} \quad \text{and} \quad m \, \varepsilon \, \mathcal{R}_g\},$$

and (14) must be satisfied for $f \notin \mathcal{F}$, that is,

$$| B_0(f) | = 0 \quad \text{for} \quad f \notin \mathcal{F}. \tag{26b}$$

Notice that $\mathcal{F}$ is the transmission band. Clearly, the transmission band can be determined from $\mathcal{I}$ and $\mathcal{R}_f$.

Substituting (13) into (5), letting $n = L + 1$, and integrating the resulting equation, one obtains

$$\int_{-\infty}^{\infty} M(f) \left[ \prod_{i=0}^{L+1} A_i^*(f) B_i^*(f) \right]$$

$$\cdot \left\{ \frac{1}{T} \sum_{k=-\infty}^{\infty} \left[ \prod_{i=0}^{L+1} A_i\left(f - \frac{k}{T}\right) B_i\left(f - \frac{k}{T}\right) \right] - 1 \right\} df$$

$$+ \sum_{l=0}^{L+1} \int_{-\infty}^{\infty} N_l(f) \left| \prod_{i=l}^{L+1} A_i(f) B_i(f) \right|^2 df = 0. \tag{27}$$

Combining (2), (13), and (27), one gets

$$\mathcal{E} = m_0 - \int_{-\infty}^{\infty} M(f) \left[ \prod_{i=0}^{L+1} A_i^*(f) B_i^*(f) \right] df. \tag{28}$$

Substituting (11), (10), and (24) into (28) yields

$$\mathcal{E} = m_0 - \int_{-\infty}^{\infty} M(f) \left[ \frac{\lambda_0 M(f)}{T N_1(f)} \right]^{\frac{1}{2}} | A_0(f) | | B_0(f) |^2 df. \tag{29a}$$

Using (26b) and the definitions of $\mathcal{F}$ and $\mathcal{I}$, and noting that $M(f)$ is a periodic function in $f$ with period $1/T$, one can cast (29a) in the form

$$\mathcal{E} = m_0 - \int_{\mathcal{I}} \left[ \frac{\lambda_0 M(f)}{T} \right]^{\frac{1}{2}}$$

$$\cdot \left[ M(f) \sum_{i \, \varepsilon \, \mathcal{R}_f} \frac{\left| A_0\left(f - \frac{i}{T}\right) \right|}{\left[ N_1\left(f - \frac{i}{T}\right) \right]^{\frac{1}{2}}} \left| B_0\left(f - \frac{i}{T}\right) \right|^2 \right] df. \tag{29b}$$

Since (25a) is satisfied at $f - m/T$ for $m \, \varepsilon \, \mathfrak{R}_f$, one has from (25a)

$$M\left(f - \frac{m}{T}\right) \sum_{k=-\infty}^{\infty} \frac{\left|A_0\left(f - \frac{m+k}{T}\right)\right|}{\left[N_1\left(f - \frac{m+k}{T}\right)\right]^{\frac{1}{2}}} \left|B_0\left(f - \frac{m+k}{T}\right)\right|^2$$

$$= T\left[\frac{TM\left(f - \frac{m}{T}\right)}{\lambda_0}\right] - T \sum_{l=0}^{L} \left[\frac{\lambda_l N_{l+1}\left(f - \frac{m}{T}\right)}{\lambda_0}\right]^{\frac{1}{2}}$$

$$\cdot \frac{1}{\left|A_l\left(f - \frac{m}{T}\right)\right|} , \qquad m \, \varepsilon \, \mathfrak{R}_f . \tag{25b}$$

By changing the index $m + k$ to $i$, using the periodicity of $M(f)$, and then using (26a), one can rewrite (25b) as

$$M(f) \sum_{i \varepsilon \mathfrak{R}_f} \frac{\left|A_0\left(f - \frac{i}{T}\right)\right|}{\left[N_1\left(f - \frac{i}{T}\right)\right]^{\frac{1}{2}}} \left|B_0\left(f - \frac{i}{T}\right)\right|^2 = T\left[\frac{TM(f)}{\lambda_0}\right]^{\frac{1}{2}}$$

$$- T \sum_{l=0}^{L} \left[\frac{\lambda_l N_{l+1}\left(f - \frac{m}{T}\right)}{\lambda_0}\right]^{\frac{1}{2}} \frac{1}{\left|A_l\left(f - \frac{m}{T}\right)\right|} , \qquad m \, \varepsilon \, \mathfrak{R}_f . \tag{25c}$$

Now we may substitute (25c) into (29b) to obtain

$$\mathcal{E} = m_0 - \int_{\mathcal{I}} TM(f) \, df$$

$$+ \int_{\mathcal{I}} [TM(f)]^{\frac{1}{2}} \sum_{l=0}^{L} [\lambda_l]^{\frac{1}{2}} \left[\frac{\left[N_{l+1}\left(f - \frac{m}{T}\right)\right]^{\frac{1}{2}}}{\left|A_l\left(f - \frac{m}{T}\right)\right|}\right] df, \qquad m \, \varepsilon \, \mathfrak{R}_f . \tag{30}$$

Equation (30) is the expression of $\mathcal{E}$ for the case of deterministic sampling.

VI. DETERMINATION OF THE TRANSMISSION BAND

Let us consider the determination of the frequency set $\mathcal{I}$ and the integer sets $\mathfrak{R}_f$. The ratios

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{|A_l(f)|}, \qquad l = 0, \cdots, L$$

have appeared in the previous equations, such as in (30). It is obvious from Fig. 1 that $[N_1(f)]^{\frac{1}{2}}/|A_0(f)|$ may be interpreted as the noise-to-signal ratio of the first analog link, $[N_2(f)]^{\frac{1}{2}}/|A_1(f)|$ as the noise-to-signal ratio of the second analog link, and so on. A similar noise-to-signal ratio has[4, 5] appeared in optimizing the transmitting and receiving filters for a given channel (see Section II). Since such noise-to-signal ratios are usually not periodic functions in $f$, it is customary to assume that for any $f$ and $k$ we have either

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{|A_l(f)|} > \frac{\left[N_{l+1}\left(f - \frac{k}{T}\right)\right]^{\frac{1}{2}}}{\left|A_l\left(f - \frac{k}{T}\right)\right|}, \tag{31}$$

or

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{|A_l(f)|} < \frac{\left[N_{l+1}\left(f - \frac{k}{T}\right)\right]^{\frac{1}{2}}}{\left|A_l\left(f - \frac{k}{T}\right)\right|}. \tag{32}$$

This assumption is valid for most practical cases.

In the following we assume that (31) or (32) holds simultaneously for $l = 0, 1, \ldots, L$. Physically, this means that the pass and attenuation bands (valleys and peaks of the noise-to-signal ratios) of the analog links coincide. Important applications where this assumption is valid are considered in Part II of this study.[2] It should be emphasized that this assumption is usually valid because carrier modulation can and should be used at the analog repeaters to shift the frequencies so that the pass and attenuation bands of the analog links coincide and the transmission media are best used.

From the above assumption, it is easily seen that

$$\sum_{l=0}^{L} [\lambda_l]^{\frac{1}{2}} \frac{[N_{l+1}(f)]^{\frac{1}{2}}}{|A_l(f)|} \neq \sum_{l=0}^{L} [\lambda_l]^{\frac{1}{2}} \frac{\left[N_{l+1}\left(f - \frac{k}{T}\right)\right]^{\frac{1}{2}}}{\left|A_l\left(f - \frac{k}{T}\right)\right|} \tag{33}$$

for any $f$ and $k \neq 0$, regardless of the values of the $\lambda_l$'s. Comparing (33) with (19) to (21), we see that (21) is not satisfied. Therefore,

from Section V, (15) and (25a) cannot be satisfied at both $f$ and $f -$ $k/T$ for any $f$ and $k \neq 0$. Consequently, the set of integers $\Re_f$ defined after (25a) cannot contain more than one element, and (26a) becomes

$$\left| B_0\left(f - \frac{m + k}{T}\right) \right| = 0, \quad \text{for} \quad m \varepsilon \Re_f, \quad k \neq 0. \tag{34}$$

Substituting (34) into (25b) gives

$$M(f) \frac{| A_0(f) |}{[N_1(f)]^{\frac{1}{2}}} | B_0(f) |^2$$

$$= T\left[\frac{TM(f)}{\lambda_0}\right]^{\frac{1}{2}} - T \sum_{l=0}^{L} \left[\frac{\lambda_l N_{l+1}(f)}{\lambda_0}\right]^{\frac{1}{2}} \frac{1}{| A_l(f) |}, \quad f \varepsilon \mathfrak{F}.$$

From this the transmitting filter is determined as

$$| B_0(f) |^2 = \frac{T[N_1(f)]^{\frac{1}{2}}}{M(f) | A_0(f) | \lambda_0^{\frac{1}{2}}}$$

$$\cdot \left\{[TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \frac{[N_{l+1}(f)]^{\frac{1}{2}}}{| A_l(f) |}\right\}, \quad f \varepsilon \mathfrak{F} \tag{35}$$

For the solution of the optimization problem to be valid, the solutions of $|B_l(f)|$ must satisfy the conditions

$$| B_l(f) | \geqq 0, \quad l = 0, 1, \cdots, L + 1.$$

These conditions can be used to determine the appropriate signs of $(\lambda_l)^{\frac{1}{2}}$, $l = 0, \cdots, L$. Consider first the possibility that $(\lambda_0)^{\frac{1}{2}} < 0$. It can be seen from (8) that $(\lambda_0)^{\frac{1}{2}} < 0$ and $| B_l(f) |^2 \geqq 0$, $l = 1, \cdots, L$, together require

$$(\lambda_l)^{\frac{1}{2}} < 0, \quad l = 1, \cdots, L.$$

But, from (35), the conditions

$$(\lambda_l)^{\frac{1}{2}} < 0, \quad l = 0, 1, \cdots, L,$$

would imply that

$$| B_0(f) |^2 \leqq 0,$$

which is not a valid solution. Therefore, $(\lambda_0)^{\frac{1}{2}}$ cannot be negative and must be positive. It can be shown from (35), (8), and (9) that $(\lambda_0)^{\frac{1}{2}} > 0$ and $| B_l(f) |^2 \geqq 0$, $l = 0, \cdots, L + 1$, together require that*

---

* $D(f)$ defined in (36) is an abbreviation used later.

$$D(f) = [TM(f)]^{\frac{1}{2}}$$

$$- \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \frac{\left[ N_{l+1}\left( f - \frac{m}{T} \right) \right]^{\frac{1}{2}}}{\left| A_l\left( f - \frac{m}{T} \right) \right|} \geqq 0, \qquad f \varepsilon \mathcal{I}; \qquad m \varepsilon \mathcal{R}_f \qquad (36)$$

and that $(\lambda_l)^{\frac{1}{2}} > 0$, $l = 1, \ldots, L$. From these and the fact that $(\lambda_0)^{\frac{1}{2}}$ must be positive, it is concluded that the solution is valid only when (36) is satisfied and

$$(\lambda_l)^{\frac{1}{2}} > 0, \qquad l = 0, \cdots, L. \qquad (37)$$

The necessary conditions in (36) and (37) are used later in determining $\mathcal{I}$ and $\mathcal{R}_f$.

By substituting (26b), (35), and (8) into (3), together with some algebraic manipulation, it can be shown that the power constraints in (3) can be expressed as:

$$\lambda_l^{\frac{1}{2}} P_l = B_l - \sum_{h=0}^{L} \lambda_h^{\frac{1}{2}} \alpha_{hl} , \qquad l = 0, 1, \cdots, L \qquad (38)$$

where

$$B_l = \int_{\mathcal{I}} [TM(f)]^{\frac{1}{2}} \frac{\left[ N_{l+1}\left( f - \frac{m}{T} \right) \right]^{\frac{1}{2}}}{\left| A_l\left( f - \frac{m}{T} \right) \right|} \, df, \qquad (39)$$

$$\alpha_{hl} = \alpha_{lh} = \int_{\mathcal{I}} \frac{\left[ N_{l+1}\left( f - \frac{m}{T} \right) \right]^{\frac{1}{2}}}{\left| A_l\left( f - \frac{m}{T} \right) \right|} \frac{\left[ N_{h+1}\left( f - \frac{m}{T} \right) \right]^{\frac{1}{2}}}{\left| A_h\left( f - \frac{m}{T} \right) \right|} \, df, \qquad m \varepsilon \mathcal{R}_f . \quad (40)$$

Clearly, if $\mathcal{I}$ and $\mathcal{R}_f$ are known, $\beta_l$, $\alpha_{hl}$, and $(\lambda_l)^{\frac{1}{2}}$ can be computed, the validity conditions in (36) and (37) can be checked, and the filter characteristics can be computed from (35), (8), and (9). Thus, the optimization problem is reduced to that of determining the $\mathcal{I}$ and $\mathcal{R}_f$ which minimize the mean-square error $\mathcal{E}$ in (30), subject to the power constraints in (38) to (40) and the validity conditions in (36) and (37).

6.1. *Mean-Square Error versus* $\mathcal{R}_f$

Before a design procedure can be proposed, it is necessary to understand the relationships among $\mathcal{E}$, $\mathcal{I}$, and $\mathcal{R}_f$. Such are the subjects of this section and the next, and Section 6.3 gives a simple design procedure based on the results.

From (30) and (39), the mean-square error can be written as

$$\mathcal{E} = m_0 - \int_\mathcal{J} TM(f) \, df + \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \beta_l . \tag{41}$$

From the definition of $M(f)$ in (2), it is easily shown that

$$\int_{-1/2T}^{1/2T} TM(f) \, df = m_0 . \tag{42}$$

From (41) and (42), we may decompose $\mathcal{E}$ into

$$\mathcal{E} = \mathcal{E}_1 + \mathcal{E}_2 \tag{43}$$

where

$$\mathcal{E}_1 = \int_{-1/2T}^{1/2T} TM(f) \, df - \int_\mathcal{J} TM(f) \, df \tag{44}$$

and

$$\mathcal{E}_2 = \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \beta_l . \tag{45}$$

Since $T$ and $M(f)$ are given, $\mathcal{E}_1$ depends on $\mathcal{J}$, but not on the integer in $\mathcal{R}_f$. Therefore, for any $\mathcal{J}$, the integer in $\mathcal{R}_f$ must be chosen to minimize $\mathcal{E}_2$ subject to the power constraints in (38) and (40) and the validity conditions in (36) and (37).

If we define

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_0^{\frac{1}{2}} \\ \lambda_1^{\frac{1}{2}} \\ \cdot \\ \cdot \\ \lambda_L^{\frac{1}{2}} \end{bmatrix}, \qquad \mathbf{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_L \end{bmatrix}, \tag{46}$$

$$\mathbf{Q} = \begin{bmatrix} P_0 + \alpha_{00} & \alpha_{10} & \cdots & \alpha_{L0} \\ \alpha_{01} & P_1 + \alpha_{11} & \cdots & \alpha_{L1} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{0L} & \alpha_{1L} & & P_L + \alpha_{LL} \end{bmatrix}, \tag{47}$$

then the power constraints in (38) take the compact form

$$\mathbf{Q\Lambda} = \mathbf{\beta} \tag{48}$$

and (45) becomes

$$\mathcal{E}_2 = \mathbf{\Lambda}'\mathbf{\beta} = \mathbf{\Lambda}'\mathbf{Q\Lambda}. \tag{49}$$

It can be shown that $\mathbf{Q}$ is positive definite; hence, $\mathbf{Q}^{-1}$ exists and we may combine (48) and (49) to obtain

$$\mathcal{E}_2 = \beta'\mathbf{\Lambda} = \beta'\mathbf{Q}^{-1}\beta. \qquad (50)$$

For each $f$, one may choose $m$ to minimize the ratios

$$[N_{l+1}(f - m/T)]^{\frac{1}{2}}/| A_l(f - m/T) | .$$

When this is done, $\beta_l$ and $\alpha_{hl}$ decrease—see (39) and (40)—and the elements of $\beta$ and $\mathbf{Q}$ decrease—see (46) and (47). However, it is difficult to see from (50) whether $\mathcal{E}_2$ would decrease or increase (the elements of $\beta$ and $\mathbf{Q}$ decrease, but the elements of $\mathbf{Q}^{-1}$ may increase).

We resolve this difficulty by first considering the increment in $\mathcal{E}_2$ resulting from arbitrary changes in the ratios

$$\frac{\left[N_{l+1}\left(f - \dfrac{m}{T}\right)\right]^{\frac{1}{2}}}{\left| A_l\left(f - \dfrac{m}{T}\right) \right|} , \qquad l = 0, \cdots , L.$$

For brevity, we use the abbreviations

$$C_l(f) = \frac{\left[N_{l+1}\left(f - \dfrac{m}{T}\right)\right]^{\frac{1}{2}}}{\left| A_l\left(f - \dfrac{m}{T}\right) \right|} , \qquad l = 0, \cdots , L; \qquad m \, \varepsilon \, \mathcal{R}_f . \qquad (51)$$

Let $\Gamma_l(f)$ denote the increment in $C_l(f)$, $l = 0, \cdots , L$. The resulting increments in $\lambda_l^{\frac{1}{2}}$, $\mathcal{E}_2$, $\mathbf{\Lambda}$, $\beta$, and $\mathbf{Q}$ are denoted, respectively, by $\Delta_l$, $d$, $\tilde{\mathbf{\Lambda}}$, $\tilde{\beta}$, and $\tilde{\mathbf{Q}}$.

Notice that the increments $\Gamma_l(f)$, $l = 0, \cdots , L$, are not necessarily small.

Replacing $\mathbf{Q}$, $\mathbf{\Lambda}$, $\beta$, and $\mathcal{E}_2$, by $(\mathbf{Q} + \tilde{\mathbf{Q}})$, $(\mathbf{\Lambda} + \tilde{\mathbf{\Lambda}})$, $(\beta + \tilde{\beta})$, and $\mathcal{E}_2 + d$, respectively, one has from (48) and (49)

$$(\mathbf{Q} + \tilde{\mathbf{Q}})(\mathbf{\Lambda} + \tilde{\mathbf{\Lambda}}) = \beta + \tilde{\beta} \qquad (52)$$

$$\mathcal{E}_2 + d = (\mathbf{\Lambda} + \tilde{\mathbf{\Lambda}})'(\beta + \tilde{\beta}). \qquad (53)$$

From (52) and (48)

$$\mathbf{Q}\tilde{\mathbf{\Lambda}} = \tilde{\beta} - \tilde{\mathbf{Q}}\mathbf{\Lambda} - \tilde{\mathbf{Q}}\tilde{\mathbf{\Lambda}}. \qquad (54)$$

From (53) and (49)

$$d = \mathbf{\Lambda}'\tilde{\beta} + \tilde{\mathbf{\Lambda}}'\beta + \tilde{\mathbf{\Lambda}}'\tilde{\beta}. \qquad (55)$$

Multiplying (48) by $\tilde{\Lambda}'$, transposing $\tilde{\Lambda}'Q\Lambda$ in the resulting equation, noting that $Q$ is symmetric, and using (54), one obtains

$$\tilde{\Lambda}'\mathfrak{B} = \Lambda'\tilde{\mathfrak{B}} - \Lambda'\tilde{Q}\Lambda - \Lambda'\tilde{Q}\tilde{\Lambda}. \tag{56}$$

Substituting (56) into (55) gives

$$d = \Lambda'\tilde{\mathfrak{B}} + \tilde{\mathfrak{B}}'(\Lambda + \tilde{\Lambda}) - \Lambda'\tilde{Q}(\Lambda + \tilde{\Lambda}). \tag{57}$$

By the definitions of $\beta_l$ and $\alpha_{hl}$ in (39) and (40), and by the definitions of the increments, one can show after some manipulation that

$$\Lambda'\tilde{Q}(\Lambda + \tilde{\Lambda}) = \sum_{h=0}^{L} [\lambda_h^{\frac{1}{2}} + \Delta_h] \int_{\mathscr{I}} \left[ \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f) \right] \Gamma_h(f) \, df$$

$$+ \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \int_{\mathscr{I}} \left\{ \sum_{h=0}^{L} [\lambda_h^{\frac{1}{2}} + \Delta_h][C_h(f) + \Gamma_h(f)] \right\} \Gamma_l(f) \, df \tag{58}$$

$$\tilde{B}'(\Lambda + \tilde{\Lambda}) = \sum_{l=0}^{L} [\lambda_l^{\frac{1}{2}} + \Delta_l] \int_{\mathscr{I}} [TM(f)]^{\frac{1}{2}} \Gamma_l(f) \, df \tag{59}$$

$$\Lambda'\tilde{\mathfrak{B}} = \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \int_{\mathscr{I}} [TM(f)]^{\frac{1}{2}} \Gamma_l(f) \, df. \tag{60}$$

We are looking for a condition to determine the sign of $d$. We must decompose or combine the terms in such a way that the condition, if it exists, can be detected. This is done by substituting (58) to (60) into (57) and casting the resulting equation in the following form:

$$d = \sum_{h=0}^{L} [\lambda_h^{\frac{1}{2}} + \Delta_h] \int_{\mathscr{I}} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f) \right\} \Gamma_h(f) \, df$$

$$+ \sum_{h=0}^{L} \lambda_h^{\frac{1}{2}} \int_{\mathscr{I}} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} [\lambda_l^{\frac{1}{2}} + \Delta_l][C_l(f) + \Gamma_l(f)] \right\} \Gamma_h(f) \, df. \tag{61}$$

From (61) we can prove a theorem about the selection of $\mathfrak{R}_f$ for any given $\mathscr{I}$.

It has been shown after (35) that the solution of the optimization problem is valid only when (36) and (37) are satisfied. From the definition of $C_l(f)$ in (51), (36) can be written as

$$D(f) = [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f) \geqq 0, \quad \text{for} \quad f \, \varepsilon \, \mathscr{I}. \tag{36}$$

For any given $\mathscr{I}$, let us define:

$$\{\mathfrak{R}_f\}_{\mathscr{I}} = \text{the set of all the choices of } \mathfrak{R}_f$$
$$\text{which, together with the given } \mathscr{I},$$
$$\text{satisfy (36) and (37).} \tag{62}$$

We have assumed that (31) or (32) holds for all $l$; therefore, there is a choice of $\mathfrak{R}_f$ in $\{\mathfrak{R}_f\}_\mathfrak{s}$ which simultaneously minimizes $C_i(f), l = 0, \cdots, L$. For later use, let us define:

$$\hat{\mathfrak{R}}_{f,\mathfrak{s}} = \text{the } \mathfrak{R}_f \text{ in } \{\mathfrak{R}_f\}_\mathfrak{s} \text{ which minimizes } C_i(f), \quad l = 0, \cdots, L. \quad (63)$$

If we change $\mathfrak{R}_f$ from $\hat{\mathfrak{R}}_{f,\mathfrak{s}}$ to $\tilde{\mathfrak{R}}_{f,\mathfrak{s}}$, which is also in $\{\mathfrak{R}_f\}_\mathfrak{s}$, $C_i(f)$ will be increased, say, from $C_i(f)$ to $C_i(f) + \Gamma_i(f)$, where

$$\Gamma_i(f) \geqq 0, \qquad l = 0, \cdots, L.$$

We have shown that if $C_i(f)$ is increased to $C_i(f) + \Gamma_i(f), l = 0, \cdots, L$, $\mathcal{E}_2$ is changed to $\mathcal{E}_2 + d$, where $d$ is given in (61). Since $\hat{\mathfrak{R}}_{f,\mathfrak{s}}$ is in $\{\mathfrak{R}_f\}_\mathfrak{s}$, (36) and (37) are satisfied. Since $\tilde{\mathfrak{R}}_{f,\mathfrak{s}}$ is also in $\{\mathfrak{R}_f\}_\mathfrak{s}$, (36) and (37) are again satisfied, but in the form

$$[TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} [\lambda_i^{\frac{1}{2}} + \Delta_i][C_i(f) + \Gamma_i(f)] \geqq 0, \quad \text{for} \quad f \, \varepsilon \, \mathfrak{s}$$

and

$$\lambda_i^{\frac{1}{2}} + \Delta_i > 0, \qquad l = 0, \cdots, L$$

because $C_i(f)$ is increased to $[C_i(f) + \Gamma_i(f)]$ and $(\lambda_i)^{\frac{1}{2}}$ is changed to $[(\lambda_i)^{\frac{1}{2}} + \Delta_i]$. Substituting (36), (37), and the two inequalities above into (61) shows that

$$d > 0.$$

Therefore, $\mathcal{E}_2$ and $\mathcal{E}$ increase when $\tilde{\mathfrak{R}}_{f,\mathfrak{s}}$ is chosen instead of $\hat{\mathfrak{R}}_{f,\mathfrak{s}}$ ($\mathcal{E}_1$ is fixed for a given $\mathfrak{s}$). This proves:

*Theorem 1: For any given $\mathfrak{s}$, the mean-square error $\mathcal{E}$ is minimized by selecting $\hat{\mathfrak{R}}_{f,\mathfrak{s}}$ in $\{\mathfrak{R}_f\}_\mathfrak{s}$.*

Clearly, $\hat{\mathfrak{R}}_{f,\mathfrak{s}}$ is the optimum $\mathfrak{R}_f$ for the given $\mathfrak{s}$ because it minimizes the mean-square $\mathcal{E}$, subject to the power constraints in (38) to (40) and the validity conditions in (36) and (37).

### 6.2 *Mean-Square Error versus $\mathfrak{s}$*

We now consider the variation in the mean-square error when a set of frequencies is deleted from a given $\mathfrak{s}$. Let us define a frequency set $\mathfrak{J}$ as

$$\mathfrak{J} = \left\{ f : -\frac{1}{2T} \leqq f \leqq \frac{1}{2T} \text{ and } \mathfrak{R}_f \text{ is an empty set} \right\}.$$

Clearly, $\mathfrak{s} \cap \mathfrak{J}$ is an empty set and $\mathfrak{s} \cup \mathfrak{J}$ is the frequency set $-1/2T \leqq$

$f \leqq 1/2T$. Equation (43) can be written as

$$\mathcal{E} = \int_{\mathcal{J}} TM(f) \, df + \Lambda'\beta. \tag{64}$$

Let $\Omega$ be a set of frequencies in $\mathcal{J}$. By deleting $\Omega$ from $\mathcal{J}$ we mean that $\mathcal{R}_f$ is changed to an empty set for $f \, \varepsilon \, \Omega$, but remains unchanged for $f \notin \Omega$. When $\Omega$ is deleted from $\mathcal{J}$ (that is, when $\mathcal{J}$ is changed to $\mathcal{J} - \Omega$), $\mathcal{J}$, $Q$, $\Lambda$, $\beta$, $(\lambda_l)^{\frac{1}{2}}$, and $\mathcal{E}$ are changed to $(\mathcal{J} + \Omega)$, $(Q - \tilde{Q})$, $(\Lambda - \tilde{\Lambda})$, $(\beta - \tilde{\beta})$, $[(\lambda_l)^{\frac{1}{2}} - \Delta_l]$, and $(\mathcal{E} + e)$, respectively. Equation (64) is changed to the form

$$\mathcal{E} + e = \int_{\mathcal{J}+\Omega} TM(f) \, df + (\Lambda - \tilde{\Lambda})'(\beta - \tilde{\beta}). \tag{65}$$

Since $\Omega \subset \mathcal{J}$ and $\mathcal{J} \cap \mathcal{J}$ is an empty set, $\Omega \cap \mathcal{J}$ is an empty set. From this we may subtract (64) from (65) and obtain

$$e = \int_{\Omega} TM(f) \, df - \Lambda'\tilde{\beta} - \tilde{\Lambda}'\beta + \tilde{\Lambda}'\tilde{\beta}. \tag{66}$$

It is seen from (48) that

$$\tilde{\Lambda}'\beta = \Lambda'Q\tilde{\Lambda}. \tag{67}$$

When $\mathcal{J}$ is changed to $\mathcal{J} - \Omega$, (48) is changed to

$$(Q - \tilde{Q})(\Lambda - \tilde{\Lambda}) = \beta - \tilde{\beta}. \tag{68}$$

Subtracting (48) from (68) and combining the resulting equation with (67) yields

$$\tilde{\Lambda}'\beta = \Lambda'\tilde{\beta} - \Lambda'\tilde{Q}\Lambda + \Lambda'\tilde{Q}\tilde{\Lambda}. \tag{69}$$

Substituting (69) into (66) gives

$$e = \int_{\Omega} TM(f) \, df - \Lambda'\tilde{\beta} - (\Lambda - \tilde{\Lambda})'\tilde{\beta} + \Lambda'\tilde{Q}(\Lambda - \tilde{\Lambda}). \tag{70}$$

The $i$th element of the vector $\tilde{\beta}$ is

$$\int_{\Omega} [TM(f)]^{\frac{1}{2}} C_{i-1}(f) \, df.$$

The $i$th element of the vector $\tilde{\Lambda}$ is $\Delta_{i-1}$. The element in the $i$th row and the $j$th column of $\tilde{Q}$ is

$$\int_{\Omega} C_{i-1}(f)C_{i-1}(f) \, df.$$

Using these element values, it can be shown from (70) that

$$e = \int_{\Omega} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} (\lambda_l^{\frac{1}{2}} - \Delta_l) C_l(f) \right\} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f) \right\} df.$$

$$(71)$$

We have shown in Theorem 1 that $C_l(f)$ should be minimized. Thus, $\mathfrak{R}_f$ should be selected to avoid those frequencies where $C_l(f) = \infty$ (for instance, where discrete tone interferences exist). Furthermore, the validity conditions in (36) and (37) cannot be satisfied at such frequencies. Therefore, we may assume, without loss of generality, that $C_l(f) \neq \infty$ for the given $\mathfrak{R}_f$. The variation in $(\lambda_l)^{\frac{1}{2}}$, $\Delta_l$, then approaches zero when the variation in $\mathcal{I}$, $\Omega$, approaches zero. Therefore, if $\Omega$ is replaced by an infinitesimal frequency set $\Theta$, $\Delta_l$ becomes negligible and (71) becomes

$$e = \int_{\Theta} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f) \right\}^2 df.$$

$$(72)$$

We have defined in (36) the abbreviation

$$D(f) = [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} C_l(f).$$

$$(73)$$

It is seen from (71) that if

$$D(f) = 0, \quad \text{for all} \quad f \, \varepsilon \, \Omega,$$

$$(74)$$

then $e$ is zero and $\mathcal{E}$ is unchanged when $\Omega$ is deleted from $\mathcal{I}$.

If $D(f) \neq 0$ for some $f \, \varepsilon \, \Omega$, there is, in $\Omega$, an infinitesimal frequency set $\Theta$ in which $D(f) \neq 0$. If $\Theta$ is deleted from $\mathcal{I}$, $e$ is given by (72). The integrand of (72) is $[D(f)]^2$ and is positive; therefore, when $\Theta$ is deleted from $\mathcal{I}$, $e > 0$, and $\mathcal{E}$ increases. Repeating the deleting process we see that $\mathcal{E}$ can only increase when any frequency set $\Omega$ is deleted from $\mathcal{I}$ and $D(f) \neq 0$ for some $f \, \varepsilon \, \Omega$.

The above proves the theorem:

*Theorem 2: For any given $\mathcal{I}$ and $\mathfrak{R}_f$ which may or may not satisfy the validity conditions in (36) and (37), and for any $\Omega \subset \mathcal{I}$, deleting $\Omega$ from $\mathcal{I}$ will not change the mean-square error $\mathcal{E}$ if $D(f) = 0$ for all $f \, \varepsilon \, \Omega$, and will increase $\mathcal{E}$ if $D(f) \neq 0$ for some $f \, \varepsilon \, \Omega$.*

### 6.3 *A Design Procedure*

The ambiguity in (50) is resolved in Section 6.1. It is proven in Theorem 1 that, for any given $\mathcal{I}$, the $\mathfrak{R}_f$ which minimizes $C_l(f)$, $l =$

$0, \cdots, L$, is the optimum choice among all the solutions of $\mathfrak{R}_f$ which satisfy the validity conditions and power constraints. Theorem 2 shows that deleting a frequency set $\Omega$ from a given $\mathscr{S}$ will increase or not change the mean-square error (never decrease). It is clear from these results that, in searching for the optimum $\mathscr{S}$ and $\mathfrak{R}_f$, one should begin with the largest possible $\mathscr{S}$ and with the $\mathfrak{R}_f$ which minimizes $C_l(f)$, $l = 0$, $\cdots, L$. From the definition of $\mathscr{S}$ in (27), the largest possible $\mathscr{S}$ is seen to be

$$\mathscr{S}_{\max} = \left\{ f : -\frac{1}{2T} \leqq f \leqq \frac{1}{2T} \right\}. \tag{75}$$

Thus, we can propose the following simple design procedure:

Choose $\mathscr{S} = \mathscr{S}_{\max}$. For each $f$ in $\mathscr{S}_{\max}$, choose the $\mathfrak{R}_f$ which minimizes $C_l(f)$, $l = 0, \cdots, L$. Compute $\beta_l$, $\alpha_{hl}$, and $(\lambda_l)^{\frac{1}{2}}$ from (39), (40), and (48), respectively. If the resulting values of $(\lambda_l)^{\frac{1}{2}}$ satisfy the validity conditions in (36) and (37), the above choice of $\mathscr{S}$ and $\mathfrak{R}_f$ is optimum. The power constraints are satisfied by computing $(\lambda_l)^{\frac{1}{2}}$ from (48). The mean-square error $\mathcal{E}$ is minimized.

Increasing $C_l(f)$ or deleting some frequencies from $\mathscr{S}_{\max}$ will increase $\mathcal{E}$. See Theorems 1 and 2.

The optimum filter amplitude characteristics are given by (35), (8), and (9). The over-all phase of the system, $\theta(f)$, is given by (24) (the system may have an additional time delay). As discussed previously, $\theta(f)$ may be distributed arbitrarily among the repeaters. The minimum mean-square error is given by

$$\hat{\mathcal{E}} = \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \beta_l .$$

Thus closed form results are obtained if the choices of $\mathscr{S}$ and $\mathfrak{R}_f$ in the above design procedure satisfies the validity conditions in (36) and (37). As illustrated by the applications in Part II, such validity conditions are usually satisfied under normal operating conditions.[2]

VII. CONCLUSION

The joint optimization problem is solved in closed form for deterministic sampling under two conditions:

(i) The pass and attenuation bands of the transmission media must coincide: (31) or (32) holds for all $l$. This is usually the case, because similar transmission media are usually used. Moreover, carrier modulation can and should be used at the repeaters to shift the frequencies

so that the pass and attenuation bands coincide and the transmission media are best used.

(*ii*) The validity conditions in (36) and (37) must be satisfied (see Section 6.3). As illustrated in applications in Part II, such conditions are usually satisfied.[2]

The closed-form expressions for the optimum repeater characteristics and the minimum mean-square error can be computed using the procedure in Section 6.3.

Two theorems are proven in Section VI for resolving the ambiguity in the selection of the transmission band. These theorems hold regardless of the second condition above.

Timing error is also considered. It is shown that in the transmission band, the $j$th analog repeater $(j = 1, \ldots, L)$ should have an amplitude characteristic proportional to the given function (see Fig. 1).

$$\left[\frac{N_{j+1}(f)}{N_j(f)}\right]^{\frac{1}{4}}\left[\frac{1}{|\ A_{j-1}(f)A_j(f)\ |}\right]^{\frac{1}{2}}.$$

This simple specification holds regardless of the timing jitter distribution, the message sequence spectral density, and the two conditions above. These conclusions are deduced from this result:

(*i*) Since the above given function is in general well-behaved, and since the analog repeaters may have arbitrary phases, the optimum analog repeaters can be closely realized.

(*ii*) It is not only economical, but also optimum, to use identical analog repeaters (which may have different gain factors) in many systems where the repeater noise spectral densities differ only by multiplicative constants (but are not necessarily flat with frequency), and the amplitude characteristics of the transmission media differ only by gain constants.

(*iii*) If the repeater noise spectral densities differ only by multiplicative constants (and are not necessarily white), each analog repeater will be required to provide amplitude equalization for its adjacent transmission media (with arbitrary phase equalization). This specification for digital transmission is the same as the requirement for analog repeaters in a voice system.[1] Thus, by installing a digital as well as an analog repeater at the $(L+1)$th repeater location, a hybrid system can be used for either digital or voice transmission without changing the $L$ analog repeaters between.

REFERENCES

1. *Transmission Systems for Communications,* by Members of the Technical
     Staff, Bell Telephone Laboratories, Incorporated, 3rd ed., 1964.
2. Freeny, S. L. and Chang, R. W., "Hybrid Digital Transmission Systems—
     Part II: Information Rate of Hybrid Coaxial Cable Systems," B.S.T.J.,
     this issue, pp. 1687–1711.
3. Tufts, D. W. and Berger, T., "Optimum Pulse Amplitude Modulation, Part
     II: Inclusion of Timing Jitter," IEEE Trans. Information Theory, *IT-13,*
     No. 2 (April 1967), pp. 209–216.
4. Berger, T. and Tufts, D. W., "Optimum Pulse Amplitude Modulation, Part
     I: Transmitter-Receiver Design and Bounds from Information Theory,
     IEEE Trans. Information Theory, *IT-13,* No. 2 (April 1967), pp. 196–208.
5. Smith, J. W., The Joint Optimization of Transmitted Signal and Receiving
     Filter for Data Transmission Systems, B.S.T.J., *44,* No. 10 (December
     1965), pp. 2363–2391.

# Hybrid Digital Transmission Systems
## Part II: Information Rate of
## Hybrid Coaxial Cable Systems

By S. L. FREENY and ROBERT W. CHANG

*The information rate of a hybrid coaxial cable transmission system using multilevel pulse amplitude modulation is studied, assuming that the additive repeater noise has a flat spectral density and that statistically independent message symbols are transmitted. Questions considered theoretically are: (i) Reduction in information rate when some repeaters in an "all digital repeater" system are replaced by analog repeaters, (ii) Number of digital repeaters required for converting an analog system to digital service, (iii) Information rate versus number of added analog repeaters in a fixed digital repeater section, (iv) System sensitivity to repeater output power and noise spectral density variations, and (v) Bit rate versus baud rate and achieving the greatest bit rate. Curves and tables answer these questions.*

*It is economical and theoretically optimum to use identical analog repeaters and uniform repeater spacing for the coaxial cable systems considered. The optimum gain-frequency characteristic for the analog repeaters is the same for both analog and digital transmission. Analog cable systems can be adapted directly to hybrid digital service with no compromise in theoretical performance.*

I. INTRODUCTION

In Part I,[1] the general problem of optimizing the parameters in a hybrid (combination digital and analog) transmission system was considered. Closed form expressions were obtained for the transmitting, receiving, and analog repeater filters which would minimize the total mean square error at each digital regenerator. In this part these formulas are applied to the important special case where the transmission medium is coaxial cable, under the assumption that the additive repeater noise has a flat spectral density and that statistically in-

dependent message symbols are transmitted. No attempt was made to include practical details of circuit and filter design.

## II. A COAXIAL CABLE HYBRID DIGITAL TRANSMISSION SYSTEM

A hybrid digital transmission system is illustrated in Fig. 1. Information symbols $\{a_k\}$ are transmitted from one digital repeater to the next through $L$ analog repeaters. Multilevel pulse amplitude modulation is considered. Each symbol $a_k$ can assume any one of $\nu$ equally spaced levels with probability $1/\nu$. The spacing between two adjacent levels is denoted by $d$. The levels are assumed symmetrically spaced about zero; hence,

$$E[a_k] = 0$$

$$E[a_k^2] = \frac{d^2(\nu^2 - 1)}{12}. \tag{1}$$

Notice that $\nu$ can be an odd as well as an even integer. As usual, the $a_k$'s are assumed to be statistically independent.

The $a_k$'s are transmitted sequentially at $T$ second intervals. The baud rate of the system is $1/T$, and the bit rate is

$$R = (1/T) \log_2 \nu \text{ bps}. \tag{2}$$

It is assumed that the input amplifiers of the analog and digital
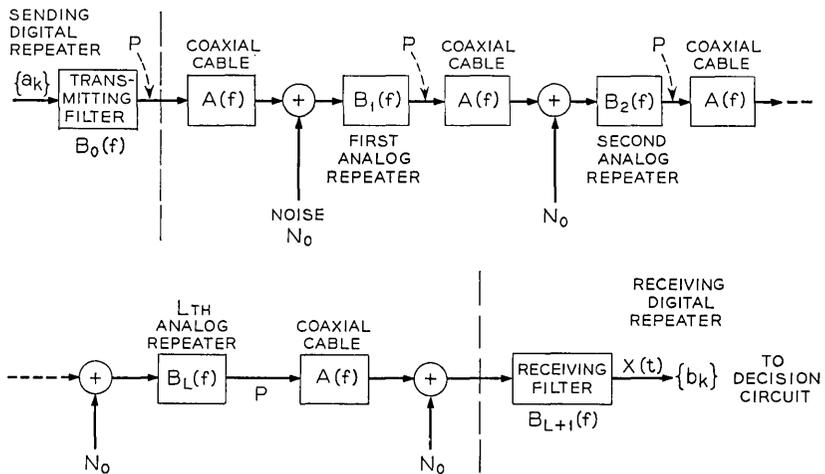


Fig. 1 — A hybrid coaxial cable digital transmission system.

repeaters introduce zero mean thermal noise of constant spectral density $N_0$ watts per hertz over the frequency band of interest. The average signal powers at the analog and digital repeater outputs are constrained to be $P$.

As in all linear pulse amplitude modulation systems, the signal at the decision circuit input—$X(t)$ in Fig. 1—is sampled sequentially at $T$ second intervals, and the $k$th time sample $b_k$ is used as an estimate of $a_k$. The system's performance is measured by the familiar criterion of mean square error between $b_k$ and $a_k$, that is, by the quantity

$$\mathcal{E} = E[(b_k - a_k)^2]. \tag{3}$$

The system is said to be optimum if $\mathcal{E}$ is minimized by jointly designing the analog and digital repeater characteristics.

It is shown in Appendix A that for the coaxial cable systems considered the mean square error $\mathcal{E}$ is further minimized if the analog repeaters are equally spaced along the cable. Therefore, uniform repeater spacing is henceforth considered. The transfer function of the coaxial cable between each two repeaters is denoted by $A(f)$ (see Fig. 1). Over most of the useful frequency range one may assume[2]

$$| A(f) | = e^{-S(|f|/f_0)^{\frac{1}{2}}} \tag{4}$$

where

$S$ = cable length in miles

$f_0$ = frequency at which one mile of cable has attenuation of one neper (a cable constant).

The analog and digital repeater characteristics that minimize $\mathcal{E}$ can be determined using the general results in Part I (see Appendix B). The main purpose of this part is to explore the interesting characteristics and potentialities of hybrid cable systems. Let us define

$N = (L + 1)N_0$

$\hat{\mathcal{E}}$ = The minimum value of $\mathcal{E}$ that can be attained by jointly designing the analog and digital repeater characteristics.

$$\mu = \frac{S}{(2Tf_0)^{\frac{1}{2}}}. \tag{5}$$

Notice that $\mu$ is the cable attenuation in nepers measured at a frequency equal to one half the symbol rate. Using results in Part I, it

is shown in Appendix C that if

$$\frac{P}{N_0} \geqq (L + 1) \frac{1}{2T} \frac{1}{\mu^2} [2\mu e^{2\mu} - 3e^{2\mu} + 4e^{\mu} - 1] \tag{6}$$

then the bit rate of the optimum system can be related to all other system parameters by

$$R = \frac{1}{2T} \log_2 \left\{ 1 + \frac{\dfrac{S^2 P}{f_0 N} + (2\mu - 1)e^{2\mu} + 1}{\dfrac{2}{3\mu^2} [(\mu - 1)e^{\mu} + 1]^2} \cdot \frac{\hat{\mathcal{E}}}{d^2} \right\}. \tag{7}$$

The condition in equation (6) requires that the signal-to-noise ratio of the system be larger than a certain value. This condition is satisfied under normal operating conditions for the following reason. It can be shown that the quantity

$$\frac{1}{\mu^2} [2\mu e^{2\mu} - 3e^{2\mu} + 4e^{\mu} - 1]$$

in equation (6) is zero when $\mu$ is zero, and increases with $\mu$. Therefore, the right side of equation (6) increases with the number of analog repeaters, the symbol rate $1/T$, the repeater spacing $S$, and the attenuation in the cable (that is, $1/f_0$). It can be shown that if $L$, $1/T$, $S$, and $1/f_0$ are made so large that equation (6) is not satisfied, the optimum system will be forced to use a bandwidth less than the Nyquist bandwidth $1/2T$ to reduce thermal noise. A system should not be designed to operate under such an extreme condition since intersymbol interference increases rapidly as the bandwidth is reduced to less than the Nyquist bandwidth $1/2T$.

For equation (7) to be useful one must assume something about the probability distribution of the total interference (intersymbol plus noise). This allows one to relate the ratio $\hat{\mathcal{E}}/d^2$ to the average probability of error. In the remainder of this paper the natural and useful assumption will be made that the total interference is normally distributed. Evidence to date indicates that this assumption is actually conservative and that average error probabilities even less than those stated would actually be obtained in most cases.[3]

III. SELECTION OF SYMBOL RATE

To facilitate comparing bit rates of systems which have different values, we ignore the variation of error probability with number of

levels, which can be at most a factor of two (for details see Ref. 4, pp. 114–118), and assume that it depends only on $\hat{\varepsilon}/d^2$. Specifically, a value of $d^2/\hat{\varepsilon} = 126$ gives an error rate of $10^{-8}$ for binary transmission. Unless stated otherwise, this value for the ratio will be used throughout.

Consider now the system parameters. The cable constant $f_0$ is usually a given parameter. Values which the thermal noise spectral density $N_0$ and the power constraint $P$ may assume are restricted in most cases. As already discussed, if the error rate is specified, the ratio $d^2/\hat{\varepsilon}$ is also approximately fixed. Thus, the factors over which the system designer may exercise the most control are the symbol rate $1/T$, the number of levels $\nu$, the number of analog repeaters $L$, and the repeater spacing $S$.

Let us consider the selection of $1/T$.

$$R = (1/T) \log_2 \nu \text{ bps} \qquad (2)$$

where $\log_2 \nu$ is the number of bits per symbol. It is proven in Appendix D that under the normal operating condition represented by equation (6), $\log_2 \nu$ decreases when $1/T$ increases. Thus, usually there exists a symbol rate which maximizes the bit rate $R$. To illustrate this and the significance of selecting the symbol rate, we consider a typical system in the following.

Consider a system using standard ⅜-inch coaxial cable which has

$$f_0 = 5 \times 10^6 \text{ hertz.} \qquad (8)$$

The analog or digital repeater output power is constrained to be

$$P = 0.1 \text{ watt.} \qquad (9)$$

The thermal noise spectral density depends on the noise figure of the amplifiers. A reasonable assumption[*] is that

$$N_0 = 1 \times 10^{-19} \text{ watts/hertz} \qquad (10)$$

corresponding to a noise figure of 13.8 dB.

Let us assume a repeater spacing $S$ of 1.25 miles. $\qquad$ (11)

Consider the case $L = 9$, that is, nine analog repeaters are used between each two digital repeaters. The ratio $d^2/\hat{\varepsilon}$ is fixed to 126, corresponding to an error rate of approximately $10^{-8}$ per 12.5 miles.

If we vary the baud rate $1/T$, we obtain the results in Fig. 2. When the baud rate $1/T$ increases from zero, the bit rate $R$ first increases and then decreases. There is a peak of $R$ at $1/T \cong 2.8 \times 10^8$. Also

---

[*] A thermal noise spectral density of $1.67 \times 10^{-19}$ watts per hertz was used in Ref. 2 based on a noise figure of about 16.2 dB.
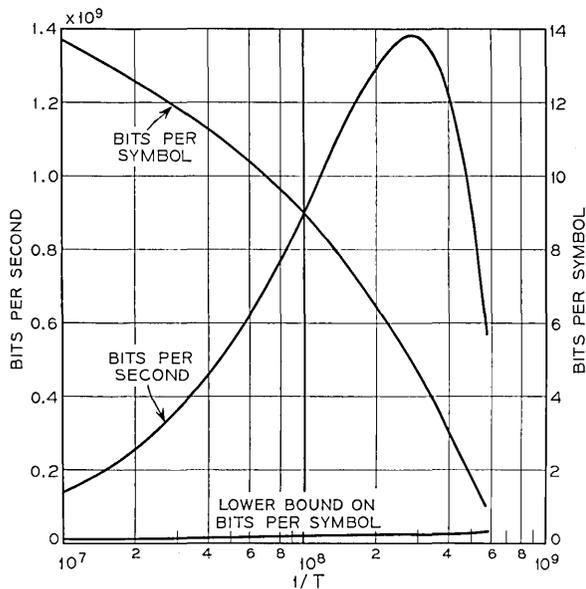
Fig. 2 — Bit rate and number of levels vs symbol rate $1/T$.

shown in Fig. 2 is the quantity "bits per symbol" (that is, $\log_2 \nu$) which, as proven in theory, decreases when the baud rate increases. Notice that the results are meaningful only when $\nu$ is an integer. Therefore, one should consider only those points where $\nu$ is an integer, or commonly used integers such as 2, 3, 4, 8, and 16.

Two observations are made from Fig. 2:

(*i*) If a low symbol rate such as $1/T \cong 10^7$ is selected, not only is the resulting bit rate too low (about $1/10$ of the maximum $R$), the number of levels must also be extremely large (approximately $2^{14}$ levels) in order to attain this very low bit rate.

(*ii*) At the maximum bit rate $\nu$ is approximately 32, an impractically large number. However, reducing $\nu$ to 16, 8, or 4 levels only reduces $R$ from $1.38 \times 10^9$ to $1.35 \times 10^9$, $1.21 \times 10^9$, or $0.96 \times 10^9$, respectively.

These observations clearly show the significance of selecting the baud rate and how a baud rate can be chosen for best use of a given system.

Notice that the above results are computed from equations (7) and (2) and that equation (7) is valid if equation (6) holds. By rearranging the terms of equation (6) together with some algebraic

manipulation, it can be shown that equation (7) is valid if the $R$ computed from equation (7) satisfies the inequality

$$R \geqq \frac{1}{2T} \log_2 \left[ 1 + \frac{6\mu^2}{\mu - 1 + e^{-\mu}} \frac{\hat{\varepsilon}}{d^2} \right]$$ (6a)

of if $\nu$ computed from equations (7) and (2) satisfies the inequality

$$\log_2 \nu \geqq \tfrac{1}{2} \log_2 \left[ 1 + \frac{6\mu^2}{\mu - 1 + e^{-\mu}} \frac{\hat{\varepsilon}}{d^2} \right].$$ (6b)

Notice that equations (6a) and (6b) represent lower bounds on bit rate and bits per baud, respectively. The lower bound on bits per symbol is plotted in Fig. 2. It is seen that, as discussed in Section II, equation (6) is easily satisfied in practice.

## IV. REPLACEMENT OF DIGITAL REPEATERS

What happens if some of the repeaters in an all-digital repeater system are replaced by analog repeaters? Because in multilevel transmission analog repeaters might be less expensive than digital repeaters, this may reduce the cost of the system.

Since analog repeaters introduce thermal noise, replacing digital repeaters with analog repeaters decreases the bit rate of the system (assuming a fixed error rate), but the reduction might not be much.

Let us consider the same system specifications (8), (9), (10), and (11), as in Section III, and let us consider three cases:

(*i*) The repeaters in the system are all digital.

(*ii*) 90 percent of the digital repeaters are replaced by analog repeaters (10 percent digital).

(*iii*) 99 percent of the digital repeaters are replaced by analog repeaters (1 percent digital).

The ratio $d^2/\hat{\varepsilon}$ is set to 144, 126, and 108, for cases *i*, *ii*, and *iii*, respectively. As discussed in Appendix E, this gives an error rate of approximately $10^{-7}$ over a distance of 125 miles for all three cases.

Under the above conditions, the bit rates of the three cases are computed using equations (7) and (2). The results are compared in Table I and plotted in Fig. 3.

In comparing cases *i* and *ii*, we see that the bit rate decreases only 18 to 28 percent when 90 percent of the digital repeaters are replaced by analog ones. From cases *i* and *iii*, we see that the bit rate de-

TABLE I— COMPARISON OF BIT RATES

| Type of Transmission | Bits per second ($\times 10^9$) | | |
| --- | --- | --- | --- |
| | Case *i* | Case *ii* | Case *iii* |
| Binary | 0.71 | 0.58 | 0.46 |
| Ternary | 1.02 | 0.82 | 0.63 |
| 4-level | 1.21 | 0.96 | 0.74 |
| 8-level | 1.56 | 1.21 | 0.91 |
| 16-level | 1.77 | 1.35 | 0.98 |
| 32-level | 1.86 | 1.38 | 0.98 |
| 64-level | 1.85 | 1.33 | 0.91 |

creases 35 to 50 percent when 99 percent of the digital repeaters are replaced by analog ones. The reductions in bit rate are moderate compared with the amount of replacement.

It is important to observe that the bit rate of case *i* is best at 32-level transmission, but it is difficult, if not impossible, to realize a 32-level transmission system. Therefore, one is forced to consider a reduced bit rate. If one uses only digital repeaters, there is only one choice, reducing bits per symbol. However, hybrid systems give another degree of freedom: one may consider various combinations of transmission levels and numbers of analog repeaters.
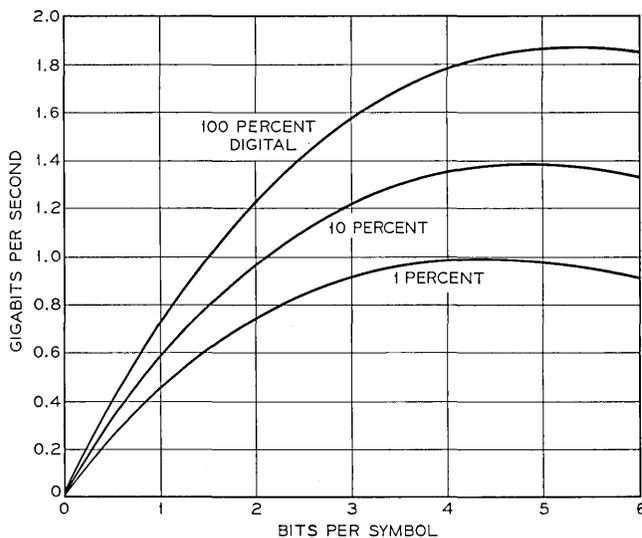


Fig. 3 — Comparison of bit rates of three arrangements.

For example, consider the two choices from Table I:

(*i*) Repeaters all digital, ternary transmission used, bit rate reduced to $1.02 \times 10^9$ bits per second.

(*ii*) 90 percent of the repeaters are analog, 4-level transmission, bit rate reduced to $0.96 \times 10^9$ bits per second.
Since the bit rates are very close, the selection depends largely on the cost of the repeaters, installation, maintenance, and so on.

## V. HYBRID SYSTEM FLEXIBILITIES

We have considered a digital system and computed the reductions in bit rate when digital repeaters are replaced by analog repeaters. But in other applications, the system may be originally built for voice communication with all analog repeaters. As is well known, the analog repeater gain-frequency characteristic for voice communication is shaped to match the loss-frequency characteristic of the coaxial cable.[5] Appendix B shows that such shaping is also optimum for multi-level digital transmission. Thus, an analog system is basically suited for digital transmission provided that digital repeaters are inserted; the question is how many.

Tables such as Table I can be helpful in making such decisions. Case *ii* in Table I corresponds to inserting one digital repeater after every nine analog repeaters, and case *iii* corresponds to inserting a a digital repeater after every ninety-nine analog repeaters. Bit rate can be easily computed for other values of $L$, $v$, $S$, $f_0$, $P$, $N_0$, and error rate using equations (7) and (2). The results reveal the capacities of various systems.

A hybrid system can be used for either digital or voice communication by installing a digital as well as an analog repeater at the $(L + 1)$st repeater location. The $L$ analog repeaters between can be used for both services without sacrificing the system performance because of common gain-frequency shaping requirements.

## VI. INSERTION OF ANALOG REPEATERS

Replacing digital repeaters with analog ones or inserting digital repeaters into an analog system amount to changing the parameter $L$ of a hybrid system. The repeater spacing $S$ is unchanged.

In certain cases, one might wish to fix the distance between two digital repeaters, and vary the number of analog repeaters between. In these cases $S$ varies with $L$.

Let the distance between two digital repeaters be $\eta$ miles. Then $S = \eta/L + 1$. Thermal noise spectral density at the input of the receiving digital repeater is $N = (L + 1)N_0$. When $L$ increases, thermal noise increases, but $S$ and cable loss decrease. Hence, the bit rate may increase or decrease depending on thermal noise and cable loss. For instance, if $N_0$ and $f_0$ are both very small, increasing $L$ will increase the bit rate; if $N_0$ and $f_0$ are large, increasing $L$ will decrease bit rate. The following shows that for typical values of $N_0$ and $f_0$, increasing $L$ increases bit rate.

Let $f_0 = 5 \times 10^6$, $P = 0.1$ and $N_0 = 10^{-19}$ as considered earlier, and assume a distance $\eta$ of 100 miles. Table II shows $L$ for each of the following:

$(1/T)_m$ = The baud rate that maximizes the bit rate
$R_{max}$ = Maximum bit rate in bits per second at $(1/T)_m$
$(\log_2 \nu)_m$ = bits per symbol baud at $(1/T)_m$.

For all $L$, $d^2/\hat{\varepsilon}$ is set to 126 (error rate $\cong 10^{-8}$ for each digital repeater section). Notice that the table contains values of $L$ which are both impractically small and impractically large, which are included only for completeness.

From Table II, we see that when $L$ increases from 1 to 10, 100, and 1000, $R_{max}$ increases 17, 550, and 12,700 times, respectively. Similarly rapid increases in bit rate are also obtained for $\eta$ as small as 10 miles or as large as 200 miles. It is concluded that, for the typical values of $N_0$ and $f_0$ considered, the insertion of analog repeaters increases the theoretical bit rate rapidly. The number of analog repeaters, how-

TABLE II — BIT RATE VERSUS $L$ FOR $\eta = 100$ MILES

| $L$ | Bits per second | | |
| --- | --- | --- | --- |
| | $(1/T)_m$ | $R_{max}$ | $(\log_2 \nu)_m$ |
| 1 | $3.68 \times 10^5$ | $2.58 \times 10^6$ | 7.00 |
| 2 | $7.60 \times 10^5$ | $5.10 \times 10^6$ | 6.71 |
| 5 | $2.60 \times 10^6$ | $1.62 \times 10^7$ | 6.22 |
| 10 | $7.53 \times 10^6$ | $4.37 \times 10^7$ | 5.80 |
| 20 | $2.32 \times 10^7$ | $1.24 \times 10^8$ | 5.35 |
| 50 | $1.06 \times 10^8$ | $4.99 \times 10^8$ | 4.73 |
| 100 | $3.31 \times 10^8$ | $1.41 \times 10^9$ | 4.25 |
| 200 | $1.02 \times 10^9$ | $3.85 \times 10^9$ | 3.77 |
| 500 | $4.30 \times 10^9$ | $1.36 \times 10^{10}$ | 3.16 |
| 1000 | $1.22 \times 10^{10}$ | $3.27 \times 10^{10}$ | 2.68 |

ever, is limited by such practical considerations as misalignment, equalization, and economy.

## VII. BIT RATE, POWER, AND NOISE

We have assumed a repeater output power, $P$, of 0.1 watt and a thermal noise spectral density, $N_0$, of $10^{-19}$ watts per hertz. Although these are conservative figures, it is nevertheless interesting to ask how sensitive the system is to variations in $P$ and $N_0$.

Let us consider the coaxial cable system specified by (8), (11), and $d^2/\hat{\mathcal{E}} = 126$. Notice from (7) that we need to consider only the ratio $P/N_0$, not $P$ and $N_0$ separately.

$$\frac{P}{N_0} = \frac{0.1}{10^{-19}} = 10^{18}.$$

It is very unlikely that $P/N_0$ will vary by a factor of $10^7$, but let us consider such a range.

Since $N = (L + 1) N_0$ and since $L$ appears only in the ratio $P/N$ in (7), we may vary $P/N$ instead of $P/N_0$ so that the results can be used for all $L$. For instance, when $L = 9$, varying $P/N$ from $10^{18}$ to $10^{11}$ corresponds to varying $P/N_0$ from $10^{19}$ to $10^{12}$.

In Fig. 4, $P/N$ is varied from $10^{18}$ to $10^{11}$. For each $P/N$, bit rate is shown versus bits per symbol (that is, versus the logarithm of the number of levels of transmission). We see that the reduction in bit rate is moderate compared with variation in $P/N$. For instance, when $P/N$ reduces from $10^{18}$ to $10^{16}$ (by a factor of $10^2$), bit rate reduces only 37 percent at binary transmission, or 40 percent at 4-level transmission. Thus, the system can tolerate a reasonable amount of variation in $P/N$. However, as one should expect, an extremely severe reduction in $P/N$ is not tolerable. For example, if $P/N$ reduces from $10^{18}$ to $10^{11}$, the maximum bit rate would be reduced to $6 \times 10^7$.

Figure 4 shows that binary transmission is the least sensitive to variation in $P/N$. As $P/N$ decreases, the peak of the curve shifts to the left, reducing the theoretical advantage of multilevel transmission over binary.

## VIII. CONCLUSIONS

The information rate of a hybrid coaxial cable digital transmission system has been evaluated theoretically. Because of the assumptions made, the various curves involving information rate are to be inter-
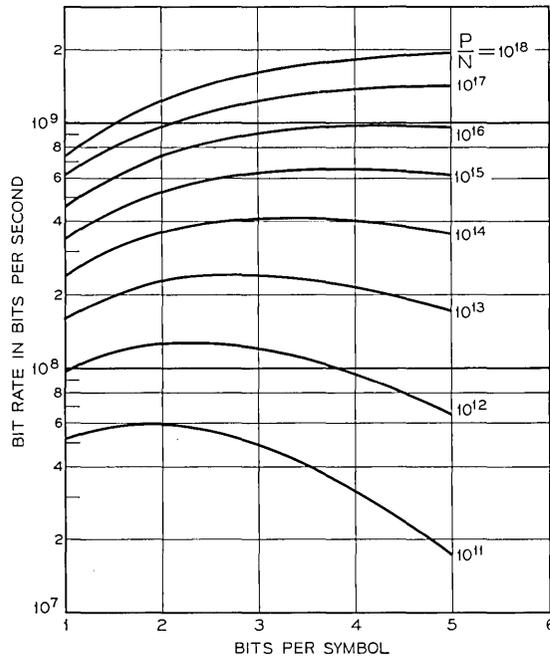
Fig. 4 — Bit rate vs bits per symbol for $P/N$ from $10^{18}$ to $10^{11}$.

preted more in the nature of upper bounds than actual performance curves to be attained in practice. Taken as such, the curves nevertheless illustrate the interesting characteristics and potentialities of hybrid cable systems. Among the more important results of the study are:

(i) In general it is not only economical but also optimum to use uniform repeater spacing and identical analog repeaters. Moreover, the optimum gain-frequency characteristic for the analog repeaters is the same for both analog and digital transmission. Therefore, an analog system can be adapted directly to hybrid digital service with no compromise in theoretical performance over the frequency band that the analog repeaters were originally designed for.

(ii) In general, hybrid systems give system designers an additional degree of freedom. For example, the curves of Fig. 3 show that, for the particular system illustrated, the sacrifice in theoretical information capacity for binary transmission between a system using all digital repeaters and one in which only one in ten repeaters is digital is about 20 percent. In order to remove low frequency energy from

the transmitted pulse spectrum, it is customary in present binary PCM systems to actually use some form of 3-level transmission, thus incurring a sacrifice of 37 percent of the information capacity. Therefore, another method of solving this dc problem (for example, some form of amplitude modulation) which is too expensive for use in each repeater might profitably be used here where it would appear in only every tenth repeater.

(iii) System parameters which have a first order effect on information capacity are the symbol rate, repeater spacing, and cable diameter. On the other hand, the hybrid cable system is relatively insensitive to variations in repeater output power, repeater noise figure and average probability of error.

APPENDIX A

## Best Uniform Repeater Spacing

$\hat{\varepsilon}$ denotes the minimum value of $\varepsilon$ attained by jointly designing the analog and digital repeater characteristics as pointed out in Part I.[1] Some terms in $\hat{\varepsilon}$ are extremely small for the coaxial cable systems considered. With such terms neglected, $\hat{\varepsilon}$ can be further minimized by using uniform repeater spacing.

Minimum notations are used in the text for clarity, but it is necessary to add a rather large number in the appendices.

Part I showed that under two conditions we have

$$\hat{\varepsilon} = \beta' Q^{-1} \beta \tag{12}$$

where

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_L \end{bmatrix}, \qquad Q = \begin{bmatrix} P_0 + \alpha_{00} & \alpha_{10} & \cdots & \alpha_{L0} \\ \alpha_{01} & P_1 + \alpha_{11} & \cdots & \alpha_{L1} \\ \vdots & \vdots & & \vdots \\ \alpha_{0L} & \alpha_{1L} & \cdots & P_L + \alpha_{LL} \end{bmatrix},$$

$$\beta_l = \int_{-1/2T}^{1/2T} [TM(f)]^{\frac{1}{2}} \frac{\left[ N_{l+1}\left(f - \frac{m}{T}\right) \right]^{\frac{1}{2}}}{\left| A_l\left(f - \frac{m}{T}\right) \right|} df, \qquad l = 0, \cdots, L$$

$$\alpha_{hl} = \int_{-1/2T}^{1/2T} \frac{\left[ N_{l+1}\left(f - \frac{m}{T}\right) \right]^{\frac{1}{2}}}{\left| A_l\left(f - \frac{m}{T}\right) \right|} \frac{\left[ N_{h+1}\left(f - \frac{m}{T}\right) \right]^{\frac{1}{2}}}{\left| A_h\left(f - \frac{m}{T}\right) \right|} df, \quad h, l = 0, \cdots, L.$$

In the above, $m$ is the integer that minimizes the ratios $[N_{l+1}(f - m/T)]^{1/2}/ \mid A_l(f - m/T) \mid$ (notice that $m$ may vary with $f$), $1/T$ is the symbol rate, and $M(f)$ is the spectral density of the stationary random message sequence $\{a_k\}$. Furthermore, as illustrated in Fig. 1 of Part I, $N_l(f)$ is the spectral density of the noise at the input of the $l$th analog repeater $B_l(f)$, $N_{L+1}(f)$ is the spectral density of the noise at the input of the receiving filter $B_{L+1}(f)$, $A_l(f)$ is the transfer function of the transmission medium between $B_l(f)$ and $B_{l+1}(f)$, and $P_l$ is the average output power of $B_l(f)$.

Equation (12) is valid under the two conditions:

($i$) For any frequency $f$ and integer $K$, and for $l = 0, \ldots , L$, we have either

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{\mid A_l(f) \mid} > \frac{\left[N_{l+1}\left(f - \frac{K}{T}\right)\right]^{\frac{1}{2}}}{\left\lvert A_l\left(f - \frac{K}{T}\right)\right\rvert}, \quad \text{for all } l \tag{13}$$

or

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{\mid A_l(f) \mid} < \frac{\left[N_{l+1}\left(f - \frac{K}{T}\right)\right]^{\frac{1}{2}}}{\left\lvert A_l\left(f - \frac{K}{T}\right)\right\rvert}, \quad \text{for all } l. \tag{14}$$

($ii$) Let

$$\Lambda = \begin{bmatrix} \lambda_0^{\frac{1}{2}} \\ \lambda_1^{\frac{1}{2}} \\ \cdot \\ \cdot \\ \lambda_L^{\frac{1}{2}} \end{bmatrix} = Q^{-1}\beta \tag{15}$$

then we must have

$$[TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \frac{\left[N_{l+1}\left(f - \frac{m}{T}\right)\right]^{\frac{1}{2}}}{\left\lvert A_l\left(f - \frac{m}{T}\right)\right\rvert} \geqq 0, \quad \text{for} \quad -\frac{1}{2T} \leqq f \leqq \frac{1}{2T} \tag{16}$$

and

$$[\lambda_l]^{\frac{1}{2}} > 0, \quad l = 0, \cdots , L \tag{17}$$

where $m$, as defined previously, is the integer that minimizes the ratios $[N_{l+1}(f - m/T)]^{\frac{1}{2}}/|A_l(f - m/T)|$.

Notice that condition $i$ guarantees that $m$ does not vary with $l$.

Now we consider the coaxial cable systems. We assume in Section II for the coaxial cable systems that

$$E[a_k] = 0, \tag{18}$$

$$E[a_k^2] = \frac{d^2(\nu^2 - 1)}{12}, \tag{19}$$

$$E[a_i a_{i+j}] = 0, \qquad j \neq 0, \tag{20}$$

$$N_l(f) = N_0, \qquad l = 1, \cdots, L + 1, \tag{21}$$

and

$$P_l = P, \qquad l = 0, \cdots, L. \tag{22}$$

By definition[1]

$$M(f) = E[a_k^2] + 2 \sum_{j=1}^{\infty} E[a_i a_{i+j}] \cos 2\pi f j T. \tag{23}$$

Substituting (19) and (20) into (23) gives

$$M(f) = \frac{d^2(\nu^2 - 1)}{12}. \tag{24}$$

In order to consider repeater spacing, let us define

$S_l$ = length of the cable (in miles) between the repeaters $B_l(f)$ and $B_{l+1}(f)$, $l = 0, \ldots, L$. Over most of the useful frequency range one may assume[2]

$$| A_l(f) | = e^{-S_l(|f|/f_0)^{\frac{1}{2}}}, \qquad l = 0, \cdots, L \tag{25}$$

where $f_0$ is a cable constant.

From (21) and (25),

$$\frac{[N_{l+1}(f)]^{\frac{1}{2}}}{| A_l(f) |} = (N_0)^{\frac{1}{2}} e^{S_l(|f|/f_0)^{\frac{1}{2}}}, \qquad l = 0, \cdots, L. \tag{26}$$

Since the right side of (26) increases monotonically with $f$, condition $i$ is satisfied (that is, for any $f$ and $K$ either (13) holds for all $l$, or (14) holds for all $l$).

Notice from (26) that, in general, the ratios $[N_{l+1}(f)]^{\frac{1}{2}}/| A_l(f) |$ will increase monotonically with $f$ even if the repeater noises are not

white (the variation in the exponent usually outweighs possible variations in the noise spectral densities). Thus, condition $i$ will usually be satisfied even if white noises are not assumed.

The second condition in (16) and (17) serves as a final check. It is not used in any computation, and we only have to show that our results satisfy it. This is done at the end of this appendix.

As already discussed, for each $f$ in $-1/2T \leq f \leq 1/2T$, we should choose the integer $m$ in $\beta_l$ and $\alpha_{hl}$ to minimize the ratios

$$\frac{\left[N_{l+1}\left(f - \frac{m}{T}\right)\right]^{\frac{1}{2}}}{\left|A_l\left(f - \frac{m}{T}\right)\right|}, \qquad l = 0, \cdots, L.$$

It is clear from (26) that these ratios are minimized by choosing

$$m = 0, \qquad \text{for all } f \text{ in } -\frac{1}{2T} \leq f \leq \frac{1}{2T}. \tag{27}$$

Substituting (24), (21), (25), and (27) into the definition of $\beta_l$ gives

$$\beta_l = \psi \frac{1}{\mu_l^{\frac{1}{2}}} [1 + (\mu_l - 1)e^{\mu_l}], \qquad l = 0, \cdots, L \tag{28}$$

where

$$\psi = \left[\frac{N_0 d^2(\nu^2 - 1)}{3T}\right]^{\frac{1}{2}}$$

$$\mu_l = \frac{S_l}{[2Tf_0]^{\frac{1}{2}}}.$$

Substituting (21), (25), and (27) into the definition of $\alpha_{hl}$ gives

$$\alpha_{hl} = \frac{2N_0}{T(\mu_h + \mu_l)^2} [1 + (\mu_h + \mu_l - 1)e^{\mu_h + \mu_l}], \qquad h, l = 0, \cdots, L. \tag{29}$$

Substituting (22) into (12) gives

$$\hat{\mathcal{E}} = \beta'[P\mathbf{I} + \alpha]^{-1}\beta \tag{30}$$

where $\mathbf{I}$ is the identity matrix and

$$\alpha = \begin{bmatrix} \alpha_{00} & \alpha_{10} & \cdots & \alpha_{L0} \\ \alpha_{01} & \alpha_{11} & \cdots & \alpha_{L1} \\ \vdots & \vdots & & \vdots \\ \alpha_{0L} & \alpha_{1L} & \cdots & \alpha_{LL} \end{bmatrix}.$$

The total distance between the two digital repeaters $B_0(f)$ and $B_{L+1}(f)$ is

$$\eta = S_0 + S_1 + \cdots + S_L. \tag{31}$$

Clearly, we should regard $\eta$ as fixed when varying the repeater spacings $S_0$ to $S_L$ to minimize $\hat{\mathcal{E}}$.

It is customary to use uniform repeater spacing, that is,

$$S_l = \frac{\eta}{L+1} = S, \qquad l = 0, \cdots, L \tag{32}$$

where $S$ is the common repeater spacing. It is shown in the following that:

(i) Uniform repeater spacing minimizes the mean square error $\hat{\mathcal{E}}$ if $\alpha$ is negligible in (30).

(ii) $\alpha$ is indeed negligible for the coaxial cable systems considered. (Therefore, uniform repeater spacing is considered in the text.)

Let us prove the first statement. When $\alpha$ is negligible in (30), we have

$$\hat{\mathcal{E}} \cong \mathcal{E}_0 = \mathfrak{g}'[PI]^{-1}\mathfrak{g} \tag{33}$$

where $\mathcal{E}_0$ is an abbreviation.

Substituting (28) into (33) yields

$$\mathcal{E}_0 = \frac{1}{P}\,\psi^2 \sum_{l=0}^{L} \frac{1}{\mu_l^4}\,[1 + (\mu_l - 1)e^{\mu_l}]^2. \tag{34}$$

We now determine the repeater spacings $S_0$ to $S_L$ which minimize $\mathcal{E}_0$ in (34), subject to the constraint in (31). Since there is a one-to-one correspondence between $S_l$ and $\mu_l$, the problem is equivalent to determining the values of $\mu_0$ to $\mu_L$ which minimize $\mathcal{E}_0$ in (34), subject to the constraint

$$\mu_0 + \mu_1 + \cdots + \mu_L = \mu_{\text{total}} \tag{35}$$

where $\mu_{\text{total}}$ is a fixed constant.

A necessary condition for $\mu_l$ to minimize $\mathcal{E}_0$ subject to the constraint in (35) is

$$x_l = \lambda = 0, \qquad l = 0, \cdots, L \tag{36}$$

where $\lambda$ is a Lagrange multiplier and

$$x_l = \frac{2}{\mu_l^{3}}\,[1 + (\mu_l - 1)e^{\mu_l}]e^{\mu_l} - \frac{4}{\mu_l^{5}}\,[1 + (\mu_l - 1)e^{\mu_l}]^2. \tag{37}$$

Clearly (36) requires that

$$x_i = x_j, \qquad i, j = 0, \cdots, L; \qquad i \neq j. \tag{38}$$

We may solve (35) and the $L$ equations in (38) for $\mu_0$ to $\mu_L$. Clearly,

$$\mu_l = \frac{\mu_{\text{total}}}{L + 1}, \qquad l = 0, \cdots, L \tag{39}$$

is a solution. It is also the only solution because since repeater spacing $S_l > 0$, we have $\mu_l > 0$. It can be shown for all $l = 0, \ldots, L$ that

$$x_l = \tfrac{1}{3} \quad \text{when} \quad \mu_l = 0,$$

and that $x_l$ increases montonically with $\mu_l$ when $\mu_l > 0$. From this it is clear that $x_l$'s cannot be all equal as required by (38) if $\mu_l$'s are not all equal. Therefore, (39) is the unique solution of the constraint (35) and the necessary conditions (38). It can be easily established that $\mathcal{E}_0$ is a minimum, not a maximum, at (39). Therefore, (39) minimizes $\mathcal{E}_0$, or, uniform repeater spacing minimizes $\mathcal{E}_0$.

Next we must show that $\alpha$ is so small that for all practical purposes minimizing $\mathcal{E}_0$ minimizes $\hat{\mathcal{E}}$. Notice that it is not necessary to show this for all possible repeater spacings. Clearly we do not have to show this for classes of nonuniform repeater spacings which we know will produce $\hat{\mathcal{E}}$ larger than the $\hat{\mathcal{E}}$ of uniform repeater spacing. One such class is that which calls for

$$S_l > YS, \qquad \text{for at least one } l,$$

or equivalently

$$\mu_l > Y\mu, \qquad \text{for at least one } l,$$

where

$$\mu = \frac{S}{(2Tf_0)^{\frac{1}{2}}}, \tag{40}$$

and $Y$ is given by the equality

$$\frac{\frac{Td^2(\nu^2 - 1)}{3(Y\mu)^4 T^2} [(Y\mu - 1)e^{Y\mu} + 1]^2}{\frac{P}{N_0} + \frac{1}{2(Y\mu)^2 T} [(2Y\mu - 1)e^{2Y\mu} + 1]}$$

$$= \frac{\frac{Td^2(\nu^2 - 1)(L + 1)}{3\mu^4 T^2} [(\mu - 1)e^{\mu} + 1]^2}{\frac{P}{N_0} + \frac{L + 1}{2\mu^2 T} [(2\mu - 1)e^{2\mu} + 1]}. \tag{41}$$

The right side of (41) is the $\hat{\mathcal{E}}$ of uniform repeater spacing (see Appendix C). The left side can be easily shown to be a lower bound of $\hat{\mathcal{E}}$ for the class of nonuniform repeater spacings which calls for $S_l > YS$ for at least one $l$. Hence, this class produces an $\hat{\mathcal{E}}$ larger than that of uniform repeater spacing, and should be ruled out. Consequently, it is only necessary to show that $\alpha$ is negligible when

$$S_l \leqq YS, \qquad \text{for all } l = 0, \cdots, L. \tag{42}$$

This can be easily shown for the coaxial cable systems considered. For example, consider the typical system parameter values (Section III): $f_0 = 5 \times 10^6$ hertz, $P = 0.1$ watt, $N_0 = 10^{-19}$ watts per hertz, $S = 1.25$ miles, $L = 9$, and $1/T = 2.8 \times 10^8$. Substituting these values into (41) gives

$$y = 1.2.$$

Notice that it is not necessary to specify $d$ and $\nu$ because they cancel out in (41). Furthermore, it can be shown that the left side of (41) increases monotonically with $Y$ under the conditions in (16) and (17). Hence, the solution of $Y$ is unique. Substituting the above values into (42) and (29), one can show that the largest element in $\alpha$ cannot exceed 0.00003, which is extremely small compared with 0.1 for the diagonal elements of $P\mathbf{I}$. Thus, $\alpha$ is negligible in (30), and, for all practical purposes, minimizing $\mathcal{E}_0$ minimizes $\hat{\mathcal{E}}$.

Finally, before adopting uniform repeater spacing, we must show that it satisfies (16) and (17) in condition $ii$. As shown in Appendix C, for uniform repeater spacing, (17) is automatically satisfied and (16) is equivalent to (6). As discussed in Section II and demonstrated in Section III, (6) is easily satisfied. Therefore, uniform repeater spacing easily satisfies (16) and (17) in condition $ii$.

APPENDIX B

*Optimum Repeater Characteristics*

Part I showed that,[1] under the same conditions—(13) to (17) in Appendix A—the analog and digital repeater characteristics which minimize the mean square error $\mathcal{E}$ are:

$$| B_0(f) |^2$$

$$= \frac{T[N_1(f)]^{\frac{1}{2}}}{M(f) \mid A_0(f) \mid \lambda_0^{\frac{1}{2}}} \left\{ [TM(f)]^{\frac{1}{2}} - \sum_{l=0}^{L} \lambda_l^{\frac{1}{2}} \frac{[N_{l+1}(f)]^{\frac{1}{2}}}{\mid A_l(f) \mid} \right\} \quad \text{for} \quad f \, \varepsilon \, \mathfrak{F},$$

$$= 0, \quad \text{for} \quad f \notin \mathfrak{F} \tag{43}$$

$$| B_j(f) |^2 = \left[ \frac{\lambda_{j-1}N_{j+1}(f)}{\lambda_j N_j(f)} \right]^{\frac{1}{2}} \frac{1}{| A_{j-1}(f)A_j(f) |}, \quad \text{for} \quad f \, \varepsilon \, \mathfrak{F},$$

$$= \text{arbitrary}, \quad \text{for} \quad f \notin \mathfrak{F}, \qquad j = 1, 2, \cdots, L \qquad (44)$$

$$| B_{L+1}(f) |^2 = \left[ \frac{\lambda_0 \lambda_L}{N_1(f)N_{L+1}(f)} \right]^{\frac{1}{2}} \frac{| A_0(f) |}{| A_L(f) |} \frac{M(f)}{T} | B_0(f) |^2, \quad \text{for all} \quad f$$

$$(45)$$

and the repeater phases need only satisfy the condition

$$e^{-i\theta(f)} = 1, \quad \text{for} \quad f \, \varepsilon \, \mathfrak{F},$$

$$= \text{arbitrary}, \quad \text{for} \quad f \notin \mathfrak{F} \qquad (46)$$

where $\theta(f)$ is the over-all phase of the system that is defined by

$$\prod_{i=0}^{L+1} A_i(f)B_i(f) = \left| \prod_{i=0}^{L+1} A_i(f)B_i(f) \right| e^{-i\theta(f)}.$$

A time delay may always be added to $\theta(f)$. Furthermore, $\theta(f)$ may be distributed arbitrarily among the repeaters. The notations above are all defined in Appendix A after (12) except $\mathfrak{F}$ which is the frequency set

$$\mathfrak{F} = \left\{ f : f = g - \frac{m}{T}, \; -\frac{1}{2T} \leq g \leq \frac{1}{2T} \right\} \qquad (47)$$

where $m$ is also defined after (12).

Now apply these general equations to the coaxial cable systems considered. Substituting $m$ in (27) into (47) gives

$$\mathfrak{F} = \left\{ f : -\frac{1}{2T} \leq f \leq \frac{1}{2T} \right\}. \qquad (48)$$

The best uniform repeater spacing has been discussed in Appendix A. In this and the following appendices, we adopt uniform repeater spacing. Therefore,

$$S_l = S, \qquad l = 0, \cdots, L \qquad (49)$$

and

$$| A_l(f) | = e^{-S(|f|/f_0)^{\frac{1}{2}}}, \qquad l = 0, \cdots, L. \qquad (50)$$

Substituting (49) into (28) and (29) gives

$$\beta_l = \psi \frac{1}{\mu^2} [1 + (\mu - 1)e^{\mu}], \qquad l = 0, \cdots, L \qquad (51)$$

$$\alpha_{hl} = \frac{N_0}{2\mu^2 T} [1 + (2\mu - 1)e^{2\mu}], \qquad h, l = 0, \cdots, L \qquad (52)$$

where

$$\mu = \frac{S}{[2Tf_0]^{\frac{1}{2}}}.$$

Notice that now $\beta_l$ and $\alpha_{hl}$ do not vary with $h$ and $l$. Substituting (51), (52), and (22) into (15), one obtains

$$[\lambda_l]^{\frac{1}{2}} = \frac{\psi \frac{1}{\mu^2} [1 + (\mu - 1)e^{\mu}]}{P + (L + 1)\frac{N_0}{2\mu^2 T} [1 + (2\mu - 1)e^{2\mu}]} \qquad l = 0, \cdots, L. \quad (53)$$

Now $\lambda_l$ does not vary with $l$.

Substituting (21), (24), (48), (50), and (53) into (43) to (45) gives the repeater amplitude characteristics which minimize the mean square error as:

$$| B_0(f) | = \left[\frac{12TN_0^{\frac{1}{2}}e^{S(f/f_0)^{\frac{1}{2}}}}{d^2(\nu^2 - 1)}\right]^{\frac{1}{2}} \left\{\left[\frac{Td^2(\nu^2 - 1)}{12\lambda_0}\right]^{\frac{1}{2}} - (L + 1)N_0^{\frac{1}{2}}e^{S(f/f_0)^{\frac{1}{2}}}\right\}^{\frac{1}{2}},$$

$$\text{for} \quad 0 \leqq f \leqq \frac{1}{2T},$$

$$= 0, \quad \text{for} \quad f > \frac{1}{2T} \qquad\qquad\qquad (54)$$

$$| B_{L+1}(f) | = \left[\frac{\lambda_0 d^2(\nu^2 - 1)}{12TN_0}\right]^{\frac{1}{2}} | B_0(f) |, \quad \text{for all} \quad f \qquad (55)$$

and

$$| B_j(f) | = e^{S(f/f_0)^{\frac{1}{2}}}, \qquad 0 \leqq f \leqq \frac{1}{2T},$$

$$= \text{arbitrary}, \qquad f > \frac{1}{2T} \qquad j = 1, 2, \cdots, L. \qquad (56)$$

Several observations are made from (54) to (56). First, $| B_0(f) |$ and $| B_{L+1}(f) |$ differ only by a multiplicative constant. Therefore, identical filters may be used for the transmitting and receiving filters in the digital repeaters. [As discussed after (46), an all-pass network may be used at any point of the system to adjust overall phase of the system.]

Second, the $|B_j(f)|$, $j = 1, \ldots, L$, do not vary with $j$; hence, identical analog repeaters may be used. Furthermore, $|B_j(f)|$ in (56) is just the reciprocal of $|A_l(f)|$ in (50); therefore the analog repeater

gain-frequency characteristic is shaped to match the loss-frequency characteristic of the coaxial cable.

Third, it is much more difficult to realize the transmitting and receiving filters in the digital repeaters than it is the analog repeater filters. This is because $|B_0(f)|$ and $|B_{L+1}(f)|$ must be zero for $f > 1/2T$ (usually requiring a vertical cutoff, or discontinuity, at $f = 1/2T$), while the analog repeater filters $B_1(f)$ to $B_L(f)$ may have arbitrary amplitudes for $f > 1/2T$ (no discontinuity is required at $f = 1/2T$ and the filters may cut off in any convenient manner).

## APPENDIX C

### Information Rate

Substituting $\beta_l$ in (51), $\alpha_{hl}$ in (52), and (22) into (12) yields

$$\hat{\varepsilon} = \frac{T \dfrac{d^2(\nu^2 - 1)}{3} \dfrac{1}{\mu^4 T^2} [(\mu - 1)e^\mu + 1]^2}{\dfrac{P}{N} + \dfrac{1}{2\mu^2 T} [(2\mu - 1)e^{2\mu} + 1]} \tag{57}$$

where

$$N = (L + 1)N_0 .$$

Solving (57) for $\nu^2$, yields

$$\nu^2 = 1 + \frac{\dfrac{S^2 P}{f_0 N} + (2\mu - 1)e^{2\mu} + 1}{\dfrac{2}{3\mu^2} [(\mu - 1)e^\mu + 1]} \cdot \frac{\hat{\varepsilon}}{d^2}. \tag{58}$$

The bit rate $R$ is therefore given by

$$R = \frac{1}{T} \log_2 \nu$$

$$= \frac{1}{2T} \log_2 \left\{ 1 + \frac{\dfrac{S^2 P}{f_0 N} + (2\mu - 1)e^{2\mu} + 1}{\dfrac{2}{3\mu^2} [(\mu - 1)e^\mu + 1]^2} \cdot \frac{\hat{\varepsilon}}{d^2} \right\}.$$

which is (7) in Section II.

Equation (12) and hence (7) are valid under the conditions stated in (13), (14), (16), and (17). Appendix A showed that the condition in (13) and (14) is satisfied. Furthermore, it can be easily shown

from (53) that (17) is satisfied. Hence, the only condition remaining is (16). From (21), (24), (27), and (50), it is clear that now (16) is satisfied if and only if it is satisfied at $f = 1/2T$, or (rearranging the terms) if and only if

$$\frac{P}{N_0} \geqq \frac{L+1}{2T\mu^2} \left[ 2\mu e^{2\mu} - 3e^{2\mu} + 4e^{\mu} - 1 \right]$$

which is condition (6).

APPENDIX D

*Symbol and Bit Rates*

This appendix proves that under the normal operating condition in (6), the number of bits per symbol, $\log_2 \nu$, decreases monotonically when the baud rate increases.

Clearly, $\log_2 \nu$ decreases if $\nu^2$ in (58) decreases. For a given system, the quantities $S$, $f_0$, $P$, $L$, and $N_0$ do not vary with the baud rate $1/T$. The quantity $\hat{\mathcal{E}}/d^2$ is fixed to obtain an approximately constant error rate. The quantity $\mu$, however, is a function of $1/T$. Therefore, we have

$$\frac{\partial(\nu^2)}{\partial T} = \frac{\partial(\nu^2)}{\partial\mu} \frac{\partial\mu}{\partial T}$$

$$= \frac{\hat{\mathcal{E}}}{d^2} \frac{\partial}{\partial\mu} \left\{ \frac{\dfrac{S^2 P}{f_0 N} + (2\mu - 1)e^{2\mu} + 1}{\dfrac{2}{3\mu^2} \left[ (\mu - 1)e^{\mu} + 1 \right]^2} \right\} \frac{\partial}{\partial T} \frac{S}{(2Tf_0)^{\frac{1}{2}}}$$

$$= -\frac{\hat{\mathcal{E}}}{d^2} \frac{\mu^2}{2T} \left( \mu e^{\mu} - e^{\mu} + 1 \right)^{-3} \cdot \alpha_1 \qquad (59)$$

where

$$\alpha_1 = \left[ -3\mu^2 e^{\mu} + 3\mu e^{\mu} - 3e^{\mu} + 3 \right] \frac{S^2 P}{f_0 N} + \alpha_2 \qquad (60)$$

$$\alpha_2 = 3\mu^2 e^{3\mu} + 6\mu^2 e^{2\mu} - 3\mu^2 e^{\mu}$$
$$\quad - 9\mu e^{3\mu} + 6\mu e^{2\mu} + 3\mu e^{\mu} + 3e^{3\mu} - 3e^{2\mu} - 3e^{\mu} + 3. \qquad (61)$$

Clearly, $\mu = S/(2Tf_0)^{\frac{1}{2}} > 0$. It can be shown that

$$\mu e^{\mu} - e^{\mu} + 1 > 0 \quad \text{for} \quad \mu > 0.$$

Therefore, from (59), if $\alpha_1 < 0$ for $\mu > 0$, then $\partial\nu^2/\partial T > 0$ and $\log_2\nu$ decreases when baud rate $1/T$ increases.

We now prove that, under the normal operating condition in (6), $\alpha_1 < 0$ for $\mu > 0$. It can be shown that

$$-3\mu^2 e^\mu + 3\mu e^\mu - 3e^\mu + 3 < 0 \tag{62}$$

for $\mu > 0$. The condition in (6) can be written as

$$\frac{S^2 P}{f_0 N} \geqq 2\mu e^{2\mu} - 3e^{2\mu} + 4e^\mu - 1. \tag{63}$$

From (62) and (63)

$$\frac{S^2 P}{f_0 N} [-3\mu^2 e^\mu + 3\mu e^\mu - 3e^\mu + 3] + \alpha_3 \leqq 0 \tag{64}$$

where

$$\alpha_3 = [2\mu e^{2\mu} - 3e^{2\mu} + 4e^\mu - 1][3\mu^2 e^\mu - 3\mu e^\mu + 3e^\mu - 3].$$

It can be shown that

$$\alpha_2 < \alpha_3 \quad \text{for } \mu > 0 \tag{65}$$

Therefore, from (60), (65), and (64)

$$\alpha_1 = [-3\mu^2 e^\mu + 3\mu e^\mu - 3e^\mu + 3]\frac{S^2 P}{f_0 N} + \alpha_2$$

$$< [-3\mu^2 e^\mu + 3\mu e^\mu - 3e^\mu + 3]\frac{S^2 P}{f_0 N} + \alpha_3 \leqq 0 \quad \text{for } \mu > 0. \tag{66}$$

Inequality (66) shows that $\alpha_1 < 0$ for $\mu > 0$. From previous discussion, this implies that the number of bits per symbol, $\log_2 \nu$, decreases monotonically when symbol rate $1/T$ increases. The proof is complete.

APPENDIX E

*Error Rates*

The system from one digital repeater to the next (including $L$ analog repeaters—see Fig. 1) is referred to as a "digital repeater section" in the following.

In case $i$ of Section IV, $L = 0$ and each digital repeater section covers a distance of 1.25 miles. There are 100 digital repeater sections in 125 miles. If $d^2/\mathscr{E}$ is set to 144, error rate is approximately $10^{-9}$ for each digital repeater section, or approximately $10^{-7}$ over a distance of 125 miles.

In case $ii$, $L = 9$ and each digital repeater section covers 12.5 miles.

If $d^2/\hat{\varepsilon}$ is set to 126, error rate is about $10^{-8}$ for each digital repeater section, or approximately $10^{-7}$ over a distance of 125 miles.

In case $iii$, $L = 99$ and each digital repeater section covers 125 miles. If $d^2/\hat{\varepsilon}$ is set to 108, error rate is approximately $10^{-7}$ for each digital repeater section, that is, a distance of 125 miles.

The bit rates in Table I are not sensitive to variations in $d^2/\hat{\varepsilon}$. For instance, if one sets $d^2/\hat{\varepsilon}$ to 126 for all three cases (comparing the three cases with the same mean square error at decision circuit inputs), the bit rate of case $i$ increases only about 1.5 percent from that in Table I, the bit rate of case $ii$ is unchanged, and the bit rate of case $iii$ decreases only about 2 percent.

REFERENCES

1. Chang, R. W. and Freeny, S. L., "Hybrid Digital Transmission Systems—Part I: Joint Optimization of Analog and Digital Repeaters," B.S.T.J., this issue, pp. 1663–1686.
2. Pierce, J. R., "Information Rate of a Coaxial Cable with Various Modulation Systems," B.S.T.J., *45*, No. 8 (October 1966), pp. 1197–1207.
3. Norris, W. E., Bell Telephone Laboratories, and Lytle, D. W., University of Washington, unpublished work.
4. Bennett, W. R. and Davey, J. R., *Data Transmission*, New York: McGraw-Hill, 1965.
5. *Transmission Systems for Communications*, by Members of the Technical Staff, Bell Telephone Laboratories, Incorporated, 3rd ed., 1964.

# The Binary Regenerative Channel*

## By RICHARD H. McCULLOUGH

The nature of the errors in a regenerative digital transmission system is such that a memoryless channel is a poor model for predicting the error phenomena. In this paper we present a model which provides a reasonable approximation to observed error phenomena. The memory of the channel is represented by a Markov model. This model is similar to the model developed by E. N. Gilbert, but several important modifications greatly simplify the estimation of parameters, and make the model correspond more closely to the physical phenomena involved.

Bounds for the channel capacity of the binary regenerative channel are obtained. Error separation, block error, and burst statistics are derived.

Error model parameters are derived from available experimental data on the T1 digital transmission line and the switched telephone network. The Markov model is shown to provide a good representation of the observed error phenomena.

## I. INTRODUCTION

The Gilbert burst-noise channel introduced the idea of error states.[1] The error states represent different error processes, each of which generates independent errors. Gilbert's model yields a "renewal error process," that is, an error process for which the gaps between successive errors are independent random variables with the same probability distribution. Elliott[2] introduced a generalization which yields what we shall call a "Markov error process," that is, an error process for which the gaps between errors are dependent random variables with probability distributions which depend only on the last gap between errors. More recently, Elliott used a renewal error process, with component error processes which do not generate independent errors.[3] In order to match experimental data for block

---

error distributions, he found it necessary to introduce three empirically derived error separation distributions.

Berger and Mandelbrot proposed a renewal error process in which the error separation follows a Pareto distribution.[4] Sussman has used the Pareto distribution to model the switched telephone network.[5]

Gilbert's model and Elliott's model both assume that the transition probabilities for the error states are independent of the occurrence of errors. In this paper we drop that assumption to define a general Markov error process. We then consider a particular Markov error process in which transitions between error states are allowed only when an error occurs. We associate this error process with the "binary regenerative channel." Error separation, block error, and burst statistics are derived for this latter process. Error model parameters are calculated from available data for the T1 digital transmission line, and for the switched telephone network. We discuss briefly the possible usefulness of the Pareto distribution for approximating a many-state Markov error process, or for approximating a nonstationary error process.

This author extends this model to apply to a ternary channel.[6]

## II. ERROR MODEL

An error model must be able to reproduce the burst error phenomena which are known to occur in digital channels. Real channels seldom appear to be memoryless, and it is common for a large fraction of the errors to be burst errors. To reproduce the burst phenomena, we have chosen to use a Markov model similar to Gilbert's.[1] Our model differs from his in two important aspects. First, we have attempted to make the model correspond more closely to the physical phenomena involved by introducing several error-producing states, each with different error rates. Second, transitions between states are allowed only immediately following an error. This assumption greatly simplifies estimation of the parameters of the model, since the number of digits between adjacent errors is determined by a single error state.

The similarities and differences between these models are most easily understood by examining the transitions between error states. We shall restrict our present discussion to *two-state* error processes. We define:

$\Sigma_n$ = error state for the $n$th error digit

$Z_n$ = $n$th error digit = 1 for error and 0 for no error

$p_{ij}$ = Prob $\{\Sigma_n = j \mid \Sigma_{n-1} = i, Z_{n-1} = 0\}$

$q_{ij}$ = Prob $\{\Sigma_n = j \mid \Sigma_{n-1} = i, Z_{n-1} = 1\}$

$P_i$ = Prob $\{Z_n = 1 \mid \Sigma_n = i\}$ = average error probability (for the binary symmetric channel) of state $i$.

The state diagram of the general Markov model is shown in Fig. 1. A renewal error process is obtained if $q_{ij}$ is independent of $i$ (so that the next error state is independent of the state which produces the error), or if $P_1$ or $P_2$ is zero (so that only one state produces errors).

The Gilbert burst-error process[1] assumes that $p_{ij} = q_{ij}$ and $P_2 = 0$. The assumption that $P_2 = 0$ makes this process a renewal error process. Elliott's generalization[2] assumes $p_{ij} = q_{ij}$ but $P_2 \neq 0$. This process is a renewal error process only if $q_{11} = q_{21}$ (and $q_{12} = q_{22}$). Our model, the "binary regenerative channel," assumes that $p_{ij} = \delta_{ij}$, the Kronecker delta. This process is a renewal error process only if $q_{11} = q_{21}$ (and $q_{12} = q_{22}$).

Our assumption that state transitions can occur only after errors ($p_{ij} = \delta_{ij}$) seems reasonable for two reasons. First we hold the operational viewpoint that all our information comes from the occurrence of errors, and we might as well assume that nothing changes between errors. This also provides a practical technique for estimating transition probabilities from error separation data. Furthermore, this model seems to be quite "stable" in that extremely small transition probabilities are not encountered in practice, so that statistical estimates are relatively easy to obtain.
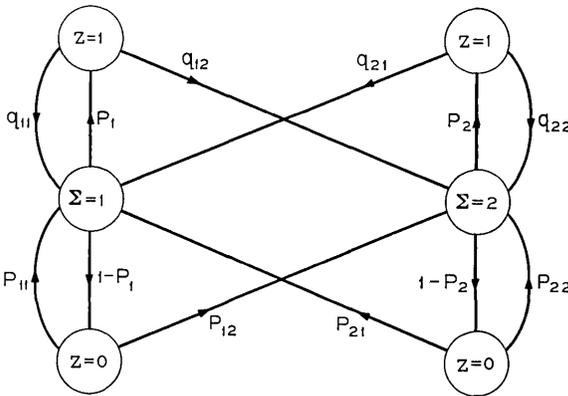


Fig. 1 — State diagram of general Markov model.

Second, this is exactly the model we would choose for an error process consisting of random errors plus signal-correlated errors, where a certain fraction of the random errors produce a wake of closely spaced errors. In the case of bursts which are not correlated with random errors, our physical intuition suggests something like $p_{11} = 1$, $p_{12} = 0$ and $p_{21} \approx q_{21} \ll 1$ would be more appropriate. Operationally, however, it really does not matter if the first error in a burst is "incorrectly" identified as a random error.

Although the mathematical descriptions of the Gilbert model, the Elliott model, and our new model are different, we suspect that the error separation, block error, and burst statistics obtained from the three models will be quite similar. (We show that the form of the error separation statistics is identical for the Gilbert model and our new model.) We contend, however, that the new model is more useful because the parameters of the model are easily determined from experimental data and are easier to interpret in terms of physical noise processes.

In the above discussion we have considered the two error states to correspond to different physical error processes in a single channel. However, this single channel is clearly equivalent to a two channel transmission system where the "error" state indicates which channel is being used. We use this latter interpretation in the next section. Notice that the two channels are simply binary symmetric channels with different error rates. In practice we have $P_2 \ll P_1 \approx 1/2$; therefore, we refer to state (channel) 1 as the *burst error state* (channel) and state (channel) 2 as the *random error state* (channel).

III. CHANNEL CAPACITY

Closed form expressions for the capacity of the Markov channel have not yet been found* so that we are limited to determining the capacity for specific numerical values of the parameters. On the other hand, we can find reasonably simple and tight bounds on the capacity which are quite useful. Therefore we consider only bounds on the channel capacity.

Let the sequences of input, output, and error digits be denoted by $X_i$, $Y_i$, and $Z_i$, respectively, with $Y_i \equiv (X_i + Z_i) \bmod 2$ and $i = 1, 2, \cdots$. Since the noise sequence is independent of the input sequence,

---

* Note that the method used by Gilbert[1] is valid only for a renewal error process, and did not yield a closed form solution.

the channel capacity is given by

$$C = \max_{p(X)} [H(X) - H(X|Y)]$$

$$= \max_{p(X)} [H(X) - H(Z)] = 1 - H(Z)$$

where $H$ denotes the entropy of the sequence and $p(X)$ is the probability of the sequence $X$. Following Ash[7] we define the entropy of the noise sequence by

$$H(Z) = \lim_{n \to \infty} H(Z_n \mid Z_{n-1}, \cdots, Z_1).$$

We shall find bounds on the channel capacity by bounding the entropy of the noise, using the steady state probabilities of the Markov noise sequence.

Consider the state diagram in Fig. 1. There are really only four states which we shall designate by $S_i = 11, 10, 21,$ or $20,$ $i = 1, 2, \ldots,$ where the first digit indicates which of the two binary symmetric channels is being used ($\Sigma_i = 1$ or $2$) and the second digit gives the value of the error digit ($Z_i = 1$ or $0$). Thus the state diagram may be redrawn as shown in Fig. 2 (using $p_{ij} = \delta_{ij}$). The steady state probabilities $r_{ij}$ are the solutions of the equations

$$[r_{11}r_{10}r_{21}r_{20}] = [r_{11}r_{10}r_{21}r_{20}] \begin{bmatrix} q_{11}P_1 & q_{11}(1 - P_1) & q_{12}P_2 & q_{12}(1 - P_2) \\ P_1 & 1 - P_1 & 0 & 0 \\ q_{21}P_1 & q_{21}(1 - P_1) & q_{22}P_2 & q_{22}(1 - P_2) \\ 0 & 0 & P_2 & 1 - P_2 \end{bmatrix}.$$

It can be shown that

$$r_{11} = R_1 P_1$$

$$r_{10} = R_1(1 - P_1)$$

$$r_{21} = R_2 P_2$$

$$r_{20} = R_2(1 - P_2)$$

where

$$R_1 = P_e \frac{Q_1}{P_1} \qquad Q_1 = \frac{q_{21}}{q_{12} + q_{21}}$$

$$R_2 = P_e \frac{Q_2}{P_2} \qquad Q_2 = \frac{q_{12}}{q_{12} + q_{21}}$$
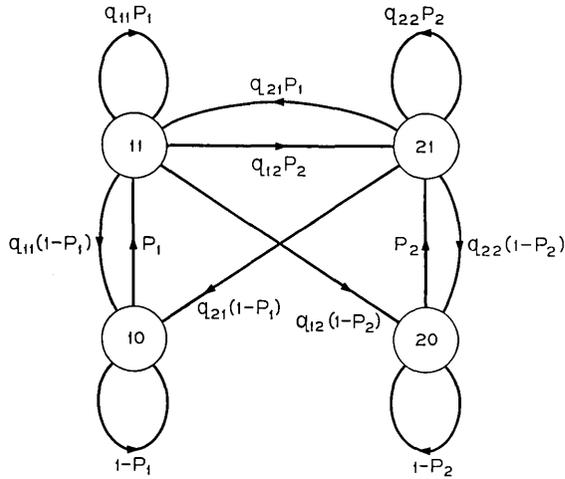
Fig. 2 — State diagram of binary regenerative channel.

$$P_e = \cfrac{1}{\cfrac{Q_1}{P_1} + \cfrac{Q_2}{P_2}} = R_1 P_1 + R_2 P_2 .$$

We notice that $R_1$ and $R_2$ are the steady state probabilities of $\Sigma_i$, $Q_1$ and $Q_2$ are the steady state probabilities of $\Sigma_i$ given that $Z_{i-1} = 1$, and $P_e$ is the steady state probability that $Z_i = 1$ (that is, $P_e$ is the average error rate).

We are now ready to compute upper and lower bounds on $H(Z)$. An upper bound is

$$H(Z) \leqq \lim_{n \to \infty} H(Z_n \mid Z_{n-1})$$

$$= P_e h(Q_1 P_1 + Q_2 P_2) + (1 - P_e) h \left[ \frac{P_e}{1 - P_e} (1 - Q_1 P_1 - Q_2 P_2) \right]$$

where $h(P) = -P \log P - (1-P) \log (1-P)$. A simpler (and looser) upper bound is

$$\lim_{n \to \infty} H(Z_n \mid Z_{n-1}) \leqq \lim_{n \to \infty} H(Z_n) = h(P_e).$$

Since $S_n$ is determined by a first order Markov process, a lower bound is

$$H(Z) \geqq \lim_{n \to \infty} H(Z_n \mid Z_{n-1} , \cdots , Z_1 ; S_{n-1})$$

$$= \lim_{n \to \infty} H(Z_n \mid S_{n-1})$$

$$= R_1 P_1 h(q_{11} P_1 + q_{12} P_2) + R_1 (1 - P_1) h(P_1)$$

$$+ R_2 P_2 h(q_{21} P_1 + q_{22} P_2) + R_2 (1 - P_2) h(P_2).$$

Using the fact that $h(P)$ is a convex function we obtain a simpler (and looser) lower bound

$$\lim_{n \to \infty} H(Z_n \mid S_{n-1}) \geqq R_1 P_1 [q_{11} h(P_1) + q_{12} h(P_2)] + R_1 (1 - P_1) h(P_1)$$

$$+ R_2 P_2 [q_{21} h(P_1) + q_{22} h(P_2)] + R_2 (1 - P_2) h(P_2)$$

$$= R_1 h(P_1) + R_2 h(P_2)$$

$$= \lim_{n \to \infty} H(Z_n \mid \Sigma_n).$$

From the loose bounds we see that the capacity of the binary regenerative channel is greater than the capacity $C = 1 - h(P_e)$ of a binary symmetric channel with the same average error rate, and is less than the capacity $C = R_1 [1 - h(P_1)] + R_2 [1 - h(P_2)]$ which could be achieved if we always knew which component channel was being used.

A convenient way to describe the channel capacity is to give the probability $P_c$ of the binary symmetric channel with capacity $C$, that is, $H(Z) = h(P_c)$. From the bounds given above it follows that

$$h^{-1}[R_1 h(P_1) + R_2 h(P_2)] \leqq h^{-1}[\lim_{n \to \infty} H(Z_n \mid S_{n-1})] \leqq P_c$$

$$\leqq h^{-1}[\lim_{n \to \infty} H(Z_n \mid Z_{n-1})] \leqq P_e.$$

For the practical case where $P_2 \ll P_1 \approx 1/2$ and $Q_1 \approx Q_2 \approx 1/2$ the above inequalities are approximately

$$P_2 \lesssim P_2 \lesssim P_c \lesssim (1 - Q_1 P_1) P_e \lesssim P_e$$

or

$$Q_2 P_e \lesssim P_c \lesssim (1 - Q_1 P_1) P_e.$$

The loose bounds given above can be generalized to apply to any finite number of memoryless, nonsymmetric channels in the form

$$B(R_1 P_1 + \cdots + R_m P_m) \leqq C \leqq R_1 B(P_1) + \cdots + R_m B(P_m)$$

where $R_i$ is the steady state probability of using channel $i$ and $B(P_i)$ is the capacity of channel $i$.

Digital transmission systems may use many digital links, with regeneration of the transmitted signal at the end of each link. For a system of $N$ identical links where $NP_2 \ll P_1$, we can approximate the over-all system by a single digital regenerative link with the same $q_{ij}$ and $P_1$ but with $P_2$ replaced by $P_2' = NP_2$. This substitution yields

$$P_e' \approx NP_e \qquad R_1' \approx NR_1 \qquad R_2' \approx R_2$$

which agrees with one's intuitive notion of how the over-all system should behave. That is, the number of random errors and burst errors both increase by a factor of $N$, and the length of the bursts remains the same.

## IV. ERROR STATISTICS

We shall now derive error separation, block error, and burst statistics for the two-state Markov error process. We assume two component error processes which generate independent errors with different average error rates, $P_i$, $i = 1, 2$. Transitions between error states are allowed only after errors, with the probabilities given by

$$q_{ij} = \text{Prob \{state } i \rightarrow \text{state } j | \text{last digit was an error\}}.$$

For the error separation statistics we shall make use of several results for independent errors. We begin by deriving the basic equations for an independent error process. Let $P$ be the probability that any digit is in error. The error sequence then contains a 1 with probability $P$ and $O$ with probability 1-$P$. Given an error, the probability that the next error occurs on the $k$th digit is

$$p(k) = \text{Prob } \{O^{k-1}1|1\} = \text{Prob } \{O^{k-1}1\} = P(1 - P)^{k-1}.$$

The average error separation is

$$\bar{k} = \sum_{k=1}^{\infty} kp(k) = P \sum_{k=1}^{\infty} k(1 - P)^{k-1} = \frac{1}{P}.$$

The probability that the number of good digits between errors is greater than or equal to $n$ (that is, the error separation is $n+1$ digits or greater) is given by the cumulative distribution

$$Q(n) = \text{Prob } \{k > n\} = 1 - \sum_{k=1}^{n} p(k)$$

$$= 1 - P \sum_{k=1}^{n} (1 - P)^{k-1} = (1 - P)^n.$$

For $P \ll 1$ the following approximation is quite useful

$$(1 - P)^n = e^{n \, \log(1-P)} \approx e^{-nP}.$$

The probability of getting $m$ errors in $n$ digits is

$$P(m, n) = \frac{n!}{m! \, (n - m)!} P^m (1 - P)^{n-m}.$$

The probability that a block of $n$ digits contains an error is

$$P(\geqq 1, n) = 1 - P(0, n) = 1 - (1 - P)^n \approx nP \text{ for } P \ll 1, \; n \ll \frac{1}{P}.$$

Let $Q_j$, $j = 1, 2$, be the unconditional probability of being in state $j$ at the first digit following an error. Making use of the results for the independent error process, we have

$$p(k) = Q_1 P_1 (1 - P_1)^{k-1} + Q_2 P_2 (1 - P_2)^{k-1}$$

$$\bar{k} = \frac{Q_1}{P_1} + \frac{Q_2}{P_2} \equiv \frac{1}{P_e}$$

$$Q(n) = Q_1 (1 - P_1)^n + Q_2 (1 - P_2)^n$$

where $P_e$ is the average error rate.

The expression for $P(m, n)$ for the Markov error process is a very complicated function of the parameters of the process. However, the form of the dependence upon $n$ is easily found through an appropriate set of recurrence relations. The recurrence relations are also useful for computing numerical values on a digital computer.

Let $A_i(m, n)$ be the probability that $m$ errors have occurred (that is, $m$ occurrences of state 11 or 21) in $n$ digits and that channel $i$ is used for the $n+1$st digit (that is, $\Sigma_{n+1} = i$). Then

$$P(m, n) = A_1(m, n) + A_2(m, n).$$

Considering all possible events which may occur at the $n$th digit we obtain the following pair of recurrence relations.

$$A_1(m, n) = A_1(m, n - 1) \cdot (1 - P_1) + A_1(m - 1, n - 1) \cdot P_1 q_{11}$$
$$+ A_2(m - 1, n - 1) \cdot P_2 q_{21}$$

$$A_2(m, n) = A_2(m, n - 1) \cdot (1 - P_2) + A_2(m - 1, n - 1) \cdot P_2 q_{22}$$
$$+ A_1(m - 1, n - 1) \cdot P_1 q_{12}.$$

Solving the equations for successive values of $m$ we find that the

general solution has the form[6]

$$P(m, n) = G_m(n)(1 - P_1)^{n-m} + G_m^*(n)(1 - P_2)^{n-m}$$

where $G_m(n)$ is a polynomial of degree $m$ in the variable $n$, and the asterisk denotes a cyclic permutation of the parameter subscripts, that is, $1 \to 2$ and $2 \to 1$. Assuming that the Markov process is in the steady state at digit "zero," the first two polynomials are

$$G_0(n) = R_1$$

$$G_1(n) = 2\left(\frac{1 - P_1}{P_2 - P_1}\right)R_1 P_1 q_{12} + [R_1 P_1 q_{11}]n.$$

It is possible to determine the functional dependence of $P(m, n)$ upon $n$ because we assumed that the error state changes only after an error occurs. This effectively decouples the set of recurrence relations so that $A_1(m, n)$ and $A_2(m, n)$ can be determined separately. For larger $m$ the explicit expressions for the coefficients of $G_m(n)$ become so complicated that they are of little use. Thus one can only hope to gain some insight into the behavior of $P(m, n)$ as a function of $m$ by numerical evaluation for a typical case.

The average number of digit errors in a block of $n$ digits is

$$\bar{m} = \sum_{i=0}^{n} iP(i, n) = nP_e .$$

Given that the block contains one or more errors, the average number of digit errors is

$$\bar{e} = \frac{\bar{m}}{P(\geq 1, n)} = \frac{n}{Q_1[1 - (1 - P_1)^n]/P_1 + Q_2[1 - (1 - P_2)^n]/P_2}.$$

In practice we usually have $P_2 \ll P_1 \approx 1/2$ and $R_1 \ll R_2 \approx 1$. Therefore, we have

$$P(m, n) \approx G_m^*(n)e^{-(n-m)P_2} \quad \text{for} \quad n - m \gg 1/P_1 .$$

Specifically, we find that

$$P(0, n) \approx e^{-nP_2}$$

$$P(\geq 1, n) \approx 1 - e^{-nP_2}$$

$$\bar{e} \approx \frac{1}{Q_2} \cdot \frac{nP_2}{1 - e^{-nP_2}} \quad \text{for} \quad n \gg 1/P_1 .$$

The burst error behavior of the channel is indicated by the num-

ber of successive occurrences of state 1, the burst error state. Let

$$p(b) = \text{Prob \{leave state 1 after } b \text{ errors|now in state 1\}}$$
$$= q_{11}^{b-1}(1 - q_{11}) \qquad b = 1, 2, \cdots .$$

Notice that $b$ occurrences of state 1 implies $b + 1$ "closely spaced" errors.

In the next section we calculate error model parameters from available data on transmission of binary information over the T1 digital transmission line, and over the switched telephone network. A three-state model is required to provide a reasonable match in some cases. Therefore, we digress briefly to generalize our results to apply to a three-state model. (Actually, we give the results in a form which is suitable for any finite number of states.)

We assume three component error processes which generate independent errors with different average error rates, $P_i$, $i = 1, 2, 3$. Transitions between error states are allowed only after errors, with the probabilities given by

$$q_{ij} = \text{Prob \{state } i \rightarrow \text{state } j | \text{last digit was an error\}}.$$

Let $Q_j$, $j = 1, 2, 3$, be the unconditional probability of being in state $j$ at the first digit following an error. The $Q_j$ are the solutions of the following set of equations:

$$Q_1 = Q_1 q_{11} + Q_2 q_{21} + Q_3 q_{31}$$
$$Q_2 = Q_1 q_{12} + Q_2 q_{22} + Q_3 q_{32}$$
$$Q_3 = Q_1 q_{13} + Q_2 q_{23} + Q_3 q_{33} .$$

Corresponding to the previous results we now have

$$p(k) = Q_1 P_1 (1 - P_1)^{k-1} + Q_2 P_2 (1 - P_2)^{k-1} + Q_3 P_3 (1 - P_3)^{k-1}$$

$$\bar{k} = \frac{Q_1}{P_1} + \frac{Q_2}{P_2} + \frac{Q_3}{P_2} \equiv \frac{1}{P_e}$$

$$Q(n) = Q_1 (1 - P_1)^n + Q_2 (1 - P_2)^n + Q_3 (1 - P_3)^n$$

where $P_e$ is the average error rate. The recurrence relations become

$$A_1(m, n) = (1 - P_1) A_1(m, n - 1) + P_1 q_{11} A_1(m - 1, n - 1)$$
$$+ P_2 q_{21} A_2(m - 1, n - 1) + P_3 q_{31} A_3(m - 1, n - 1)$$
$$A_2(m, n) = (1 - P_2) A_2(m, n - 1) + P_2 q_{22} A_2(m - 1, n - 1)$$

$$+ P_3 q_{32} A_3(m-1, n-1) + P_1 q_{12} A_1(m-1, n-1)$$

$$A_3(m, n) = (1 - P_3) A_3(m, n-1) + P_3 q_{33} A_3(m-1, n-1)$$

$$+ P_1 q_{13} A_1(m-1, n-1) + P_2 q_{23} A_2(m-1, n-1)$$

where $A_i(m, n)$ is the probability that $m$ errors have occurred and the error process is in state $i$ after $n$ digits. The solution of the recurrence relations gives

$$P(m, n) = A_1(m, n) + A_2(m, n) + A_3(m, n)$$

$$= G_m(n)(1 - P_1)^{n-m} + G_m^*(n)(1 - P_2)^{n-m} + G_m^{**}(n)(1 - P_3)^{n-m}$$

where $G_m(n)$ is a polynomial of degree $m$ in the variable $n$. The coefficients of $G_m(n)$ are complicated functions of the model parameters. $G_m^*(n)$ is $G_m(n)$ with the parameter subscripts cyclically permuted, that is, $1 \rightarrow 2$, $2 \rightarrow 3$, $3 \rightarrow 1$. $G_m^{**}(n)$ is $G_n^*(n)$ with the same cyclic permutation. We again have $G_o(n) = R_1$ so that

$$P(0, n) = R_1(1 - P_1)^n + R_2(1 - P_2)^n + R_3(1 - P_3)^n$$

$$P(\geqq 1, n)$$

$$= 1 - P(0, n)$$

$$= P_e \left[ Q_1 \frac{1 - (1 - P_1)^n}{P_1} + Q_2 \frac{1 - (1 - P_2)^n}{P_2} + Q_3 \frac{1 - (1 - P_3)^n}{P_3} \right].$$

The probability of being in state $i$ at any digit is

$$R_i = \frac{Q_i \bar{k}_i}{Q_1 \bar{k}_1 + Q_2 \bar{k}_2 + Q_3 \bar{k}_3} = \frac{\dfrac{Q_i}{P_i}}{\dfrac{Q_1}{P_1} + \dfrac{Q_2}{P_2} + \dfrac{Q_3}{P_3}} = P_e \frac{Q_i}{P_i}.$$

## V. EXPERIMENTAL PARAMETERS

### 5.1 *T1 Digital Transmission Line*

For the T1 digital transmission line* (see Refs. 8 and 9), the error data was obtained by measurements[10] on three different lines, each looped to obtain an equivalent system length of about 24 miles. In total, there were five runs of approximately one hour duration, that is, about $5 \times 10^9$ digits each. The transmitted pattern was 10000000 repeated. Each run produced about 100 errors. The data were proc-

---

* Manufactured for Bell System use only, by Western Electric Co., manufacturing and supply unit of the Bell System.

essed in real time with an IBM 7094 computer equipped with a direct data device. The results (a sequence of numbers $a_1, a_2, \ldots$ where $a_i$ is the number of good digits between the $j$th and $[j + 1]$st errors) were recorded on a magnetic tape.

To determine the parameters of the Markov error process model, we processed the experimental results as follows.

(*i*) Add 1 to each $a$ to get $k$, the error separation.
(*ii*) Classify each $k$ as state
    1 = "burst error state" $1 \leq k < 10$
    2 = "intermediate error state" $10 \leq k < 10^3$
    3 = "random error state" $10^3 \leq k$.
(*iii*) For each state $i$, find the average error separation, $\bar{k}_i = 1/P_i$.
(*iv*) From the sequence of states find the relative frequency of occurrence of state $i$, $Q_i$, and the relative frequency of occurrence of a transition from state $i$ to state $j$, $q_{ij}$.

Steps *iii* and *iv* were carried out for each run individually, and with all runs together (considered as one big sample). Table I lists the parameters (rounded to two significant digits) which were obtained by the above procedure. Notice that the conditions $q_{11} = q_{21} =$

### TABLE I — MARKOV MODEL FOR T1

| Run | $q_{ij}$ | | | $Q_i$ | $P_i$ | $P_e$ |
|---|---|---|---|---|---|---|
| 1 | .35 | .00 | .65 | .23 | .46 | $1.5 \times 10^{-8}$ |
|   | .43 | .07 | .50 | .19 | $3.2 \times 10^{-3}$ | |
|   | .12 | .31 | .57 | .58 | $.86 \times 10^{-8}$ | |
| 2 | .35 | .26 | .39 | .36 | .46 | $1.4 \times 10^{-8}$ |
|   | .67 | .08 | .25 | .19 | $4.0 \times 10^{-3}$ | |
|   | .24 | .17 | .59 | .45 | $.61 \times 10^{-8}$ | |
| 3 | .29 | .33 | .38 | .35 | .37 | $1.0 \times 10^{-8}$ |
|   | .62 | .19 | .19 | .23 | $4.9 \times 10^{-3}$ | |
|   | .24 | .17 | .59 | .42 | $.44 \times 10^{-8}$ | |
| 4 | .53 | .22 | .25 | .38 | .53 | $2.4 \times 10^{-8}$ |
|   | .50 | .15 | .35 | .20 | $3.6 \times 10^{-3}$ | |
|   | .19 | .20 | .61 | .42 | $1.0 \times 10^{-8}$ | |
| 5 | .47 | .13 | .40 | .19 | .26 | $1.6 \times 10^{-8}$ |
|   | .31 | .31 | .38 | .17 | $3.7 \times 10^{-3}$ | |
|   | .08 | .14 | .78 | .64 | $1.0 \times 10^{-8}$ | |
| All runs together | .42 | .21 | .37 | .31 | .43 | $1.6 \times 10^{-8}$ |
|   | .51 | .16 | .33 | .20 | $3.7 \times 10^{-3}$ | |
|   | .16 | .20 | .64 | .49 | $.77 \times 10^{-8}$ | |

$q_{31}$ and $q_{12} = q_{22} = q_{32}$ and $q_{13} = q_{23} = q_{33}$ are not satisfied. Hence, the T1 error process does not appear to be a renewal error process. (This does not mean that we should discard the renewal error process. In essence, it is a first approximation to a real error process, the Markov error process is a second approximation, and higher order Markov processes are higher order approximations. The first approximation may be satisfactory in some applications.) Using the parameters of Table I, the validity of the model was checked in three ways.

First, the theoretical cumulative distribution of the error separation,

$$Q(n) = Q_1(1 - P_1)^n + Q_2(1 - P_2)^n + Q_3(1 - P_3)^n$$

was plotted for each run. The theoretical and experimental curves matched within approximately ±0.05, for all five runs. Typical curves are shown in Fig. 3a (semilog plot) and 3b (log-log plot). Notice that we could have derived rough values for the $Q_i$ and $P_i$ by inspection of the experimental $Q(n)$ curve.
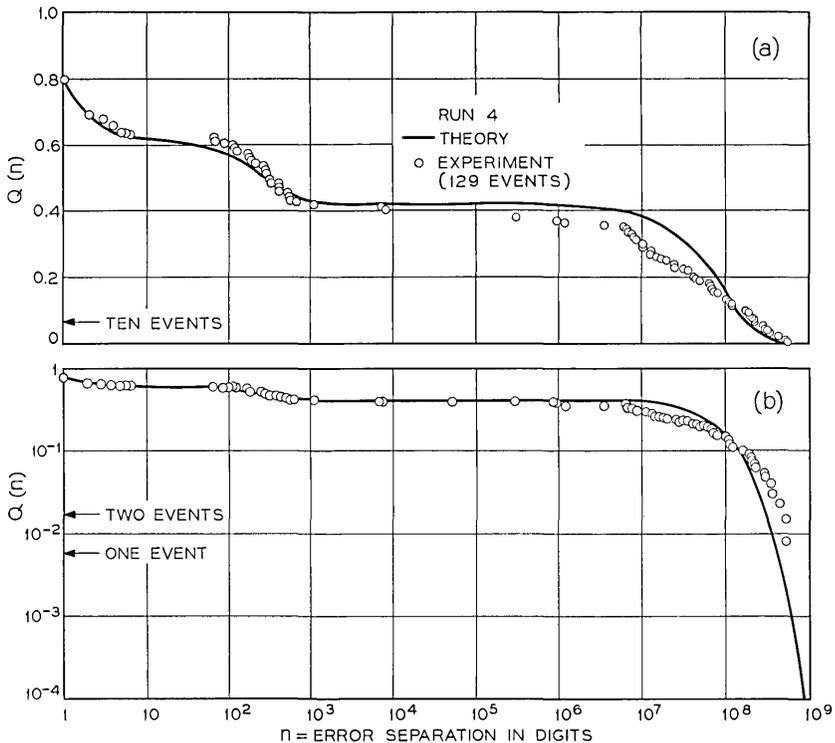


Fig. 3 — Error separation statistics for T1 digital transmission line.
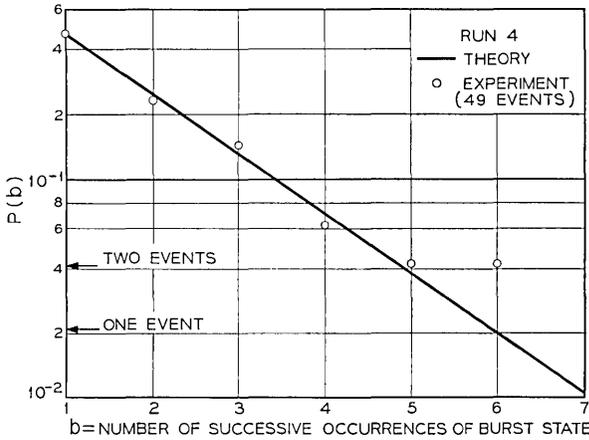
Fig. 4 — Burst statistics for T1 digital transmission line.

Second, to check the burst error behavior of the channel, we compared the experimental and theoretical probability densities, $p(b)$, of successive occurrences of the burst error state

$$p(b) = q_{11}^{b-1}(1 - q_{11}).$$

The agreement is excellent as illustrated in Fig. 4. (Since our sample size is only 49, we should not expect the experimental points to follow the theoretical curve for probabilities of about $1/49 \approx 0.02$.) The experimental and theoretical curves matched within approximately 0.02 in all five runs. Notice that the procedure for calculating $q_{11}$ simply provides an exact match at $b = 1$.

Third, to check the adequacy of the model for predicting block error statistics, we compared the theoretical and experimental (averaged over all possible phases) curves for $P(m, n)$. Figures 5 and 6 show $P(m, n)$ versus $n$ and $m$, respectively. The agreement is excellent for $m \leq 4$. For $m \geq 5$ the experimental curves are somewhat erratic owing to the small sample (the quantum of probability is approximately $2 \times 10^{-10}$ in this case), which happened to contain two unusual error patterns.

The excellent match between the experimental data and the model indicates that a three-state Markov error process with independent transitions is a good representation of the T1 error process. Since this is the case, it is useful to consider a physical interpretation of the mathematical model. The three different error states correspond
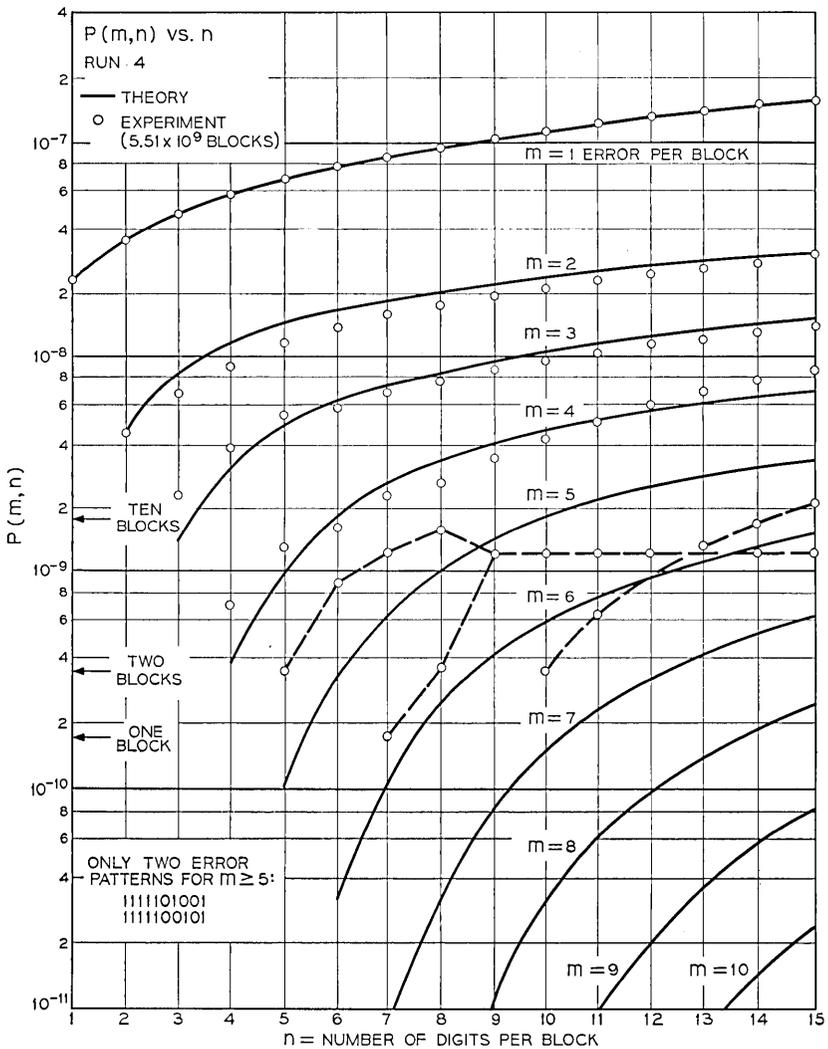
Fig. 5 — Block error statistics (digits per block) for T1 digital transmission line.

to different sources of error, of which only one is controlling at any given time. Allowing state transitions only at error digits corresponds with the fact that we cannot identify the controlling error process except by the error (and error separation) which it produces.

As for the sources of error, we can make several speculations. Fur-
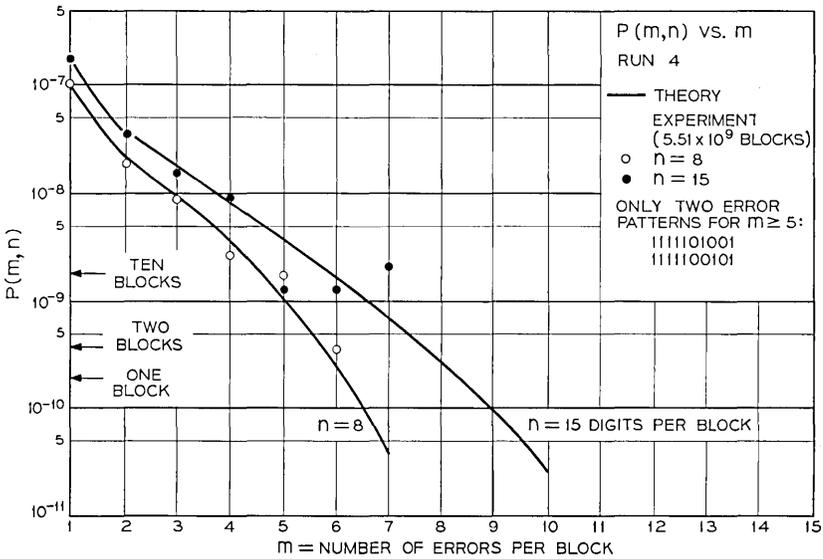
Fig. 6 — Block error statistics (errors per block) for T1 digital transmission line.

ther experimental data will be required to determine which of the suggested possibilities is correct. "Burst errors" may result from signal correlated errors (generated in successive regenerators) following a random error, or from burst-like interferences such as impulse noise (all errors of the burst generated at the same regenerator). "Intermediate errors" may be caused by looping effects (outgoing and incoming regenerators are packaged together and are thus subject to the same interference), or by slowly propagating interferences such

TABLE II — SIGNAL DEPENDENCE OF T1 ERRORS

| Run | Number of errors for each signal digit | | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| 1 | 0 | 31 | 13 | 6 | 8 | 5 | 4 | 7 | 74 |
| 2 | 1 | 19 | 8 | 12 | 6 | 7 | 6 | 6 | 65 |
| 3 | 1 | 26 | 13 | 7 | 5 | 9 | 7 | 2 | 70 |
| 4 | 2 | 63 | 15 | 11 | 10 | 11 | 8 | 10 | 130 |
| 5 | 1 | 54 | 7 | 3 | 4 | 4 | 1 | 4 | 78 |
| All runs | 5 | 193 | 56 | 39 | 33 | 36 | 26 | 29 | 417 |
| Percent all runs | 1.2 | 46.3 | 13.4 | 9.4 | 7.9 | 8.6 | 6.2 | 7.0 | 100.0 |

as teletype or other dc signaling. "Random errors" are assumed to come from thermal noise.

We have thus far ignored the question of whether the error sequence is really independent of the transmitted signal sequence. Table II summarizes the available experimental data concerning this point. We observe that the average probability of error is roughly 5 to 40 times greater when the signal digit $X_i = 0$, depending on the number of digits since the last 1. (Notice, however, that Table II does not really tell us the number of digits since the last 1 because errors cause the "signal" to be different in successive regenerators.) Over all, the average probability of error is roughly 10 times greater when $X_i = 0$.

What does this imply about our model? First, since there are so few errors when $X_i = 1$, our component channels are nonsymmetric and the model parameters derived above essentially apply only when $X_i = 0$. In fact, it is possible that the error rate for $X_i = 1$ is the same in all three component channels, so that no burst phenomena occurs if $X_i = 1$ for all $i$. Second, the dependence of the error rate on the number of digits since the last 1 probably results from intersymbol interference. This suggests that bursts might very well be signal correlated errors which are generated in successive regenerators, in which case the average length of a burst should increase with the number of regenerators. Unfortunately, the available data are not sufficient to verify or disprove these conjectures.

How do we correct our model to take into account the data presented in Table II? As a first approximation we would replace the three component binary symmetric channels with memoryless nonsymmetric binary channels with error probabilities $P_1$, $P_2$, and $P_3$ for $X_i = 0$, and $P'_1$, $P'_2$, and $P'_3$ for $X_i = 1$. With the limited data available the best we can do is to use the previously calculated values for $P_1$, $P_2$, and $P_3$, and let (using the figures for all runs)

$$P'_1 = P'_2 = P'_3 = \tfrac{5}{417}(1.6 \times 10^{-8}) = 1.9 \times 10^{-10}.$$

The computation of channel capacity and error statistics now becomes more difficult because we must consider the joint probability densities of the source and channel. However, we can still use the bounds for channel capacity given at the end of Section III.

To get any better approximation we must replace the three channels with three nonsymmetric binary channels with memory. The memory would contain $d$, the number of digits since the last 1, and could probably

be limited to three states: $d = 1$, $d = 2$, and $d \geqq 3$. (As already discussed this memory would also generate burst phenomena so that we might possibly require only two channels for this model.) This approach is intuitively appealing for modeling the effects of intersymbol interference, but we should use ternary channels with memory because the T1 digital transmission line actually transmits ternary signals.

The fact that the average error probability is always greater for $X_i = 0$ (that is, even for $d \geq 3$) probably is because of long term intersymbol interference, which in the case of T1 may persist for hundreds of digits. This interference is approximately proportional to the running sum of the digits $W_i = X_1 + X_2 + \ldots + X_i$. The Bipolar Code used in the T1 digital transmission system guarantees that $W_i$ can assume only the values 0 or $-1$ in the absence of errors.* We assume that the output of each regenerator is recoded into Bipolar so that the $W_i$ satisfy the same constraint in successive links, and the channel can be described using a finite memory. Recoding allows one to localize errors to a particular digital link and reduces the error rate in successive links. If the output is not recoded, the $W_i$ are theoretically unbounded which requires an infinite memory to describe the channel.

To summarize our thoughts on the T1 error process, we may say the following. The Markov model analyzed in the preceding sections of this paper provides a good representation of the signal-independent error phenomena, and reproduces all the gross error statistics. The extension of the model suggested in this section shows promise of providing a good representation of the signal-dependent error phenomena, and should reproduce the fine grain error statistics; additional data are required to determine the parameters and validity of the suggested extension. Notice that the signal-dependent memory is realized as a simple Markov process when the source digits are independent random variables.

## 5.2 *Switched Telephone Network*

We now consider the error model for the switched telephone network. Gilbert[1] has shown that a *two-state* Markov model provides a good approximation to the cumulative error separation distribution for an *individual* digital channel. Although Gilbert used a different model, his theoretical results for error separation are identical in

---

* McCullough[6] treats the general class of ternary restricted sum codes for which the digit sum is bounded ($-a \leqq W_i \leqq b$) for every code sequence.

form to the results of Section IV, that is,

$$Q(n) = Q_1(1 - P_1)^n + Q_2(1 - P_2)^n.$$

We notice, however, that if we choose $Q_i$ and $P_i$ so that the error separation distributions are identical, the "equivalent" binary regenerative channel will have a higher average error rate. When $P \ll hq$ Gilbert's equation (14) becomes

$$Q(n) \approx \left(1 - \frac{p}{Q - hq}\right)(hq)^n + \left(\frac{p}{Q - hq}\right)(1 - P)^n$$

so that

$$P_1 \approx 1 - hq$$

$$P_2 \approx P$$

$$Q_2 \approx \frac{p}{Q - hq}$$

or

$$h \approx \frac{1 - P_1}{1 - Q_2 P_1}$$

$$P \approx P_2$$

$$p \approx Q_2 P_1$$

and

$$P(1) \approx \frac{1 - Q_2}{1 - Q_2 P_1} \frac{P_2}{Q_2} \approx \frac{1 - Q_2}{1 - Q_2 P_1} P_e.$$

In our notation, the parameters for his examples (see Gilbert's Fig. 3) are

Channel 1146:  $Q_1 = 0$      $P_1$ arbitrary
              $Q_2 = 1,$     $P_2 = 5.4 \times 10^{-3}$

Channel 1296:  $Q_1 = 0.816$   $P_1 = 0.190$
              $Q_2 = 0.184$   $P_2 = 2.57 \times 10^{-3}.$

For an *average* of many digital channels, a *three-state* Markov model can provide a reasonable numerical fit. $Q_i$ and $P_i$ were determined for samples of the Alexander-Gryb-Nast,[11] Townsend-Watts,[12] and Kelly[13] data on the error performance of the switched telephone network. Table III lists the parameters, which were determined by

TABLE III — MARKOV MODEL FOR SWITCHED TELEPHONE NETWORK

$$Q(n) = Q_1(1 - P_1)^n + Q_2(1 - P_2)^n + Q_3(1 - P_3)^n$$

|  | Alexander-Gryb-Nast | Townsend-Watts | Kelly |
|---|---|---|---|
| $Q_1$ | 0.46 | 0.58 | 0.75 |
| $Q_2$ | 0.22 | 0.10 | 0.10 |
| $Q_3$ | 0.32 | 0.32 | 0.15 |
| $P_1$ | 0.544 | 0.567 | 0.56 |
| $P_2$ | $10^{-2}$ | $10^{-2}$ | $10^{-3}$ |
| $P_3$ | $10^{-4}$ | $\ll 10^{-3}$ | $5 \times 10^{-6}$ |

trial and error matching of the $Q(n)$ curves. It should be obvious that the $Q_i$ and $P_i$ were quantized rather coarsely. Figure 7 shows that the maximum difference between the experimental and theoretical curves is about ±0.05.

Although the numerical fit is reasonably good, it is evident that the sharp transition of a single independent-error process is not a good match to the gradual slope of the experimental curves at larger error separations. However, the experimental curves represent an average over many different channels. The parameters of the model will vary from channel to channel, resulting in an over-all error process which contains *many* states. Each state will have a small probability of occurrence ($Q_i$) and a slightly different average error probability
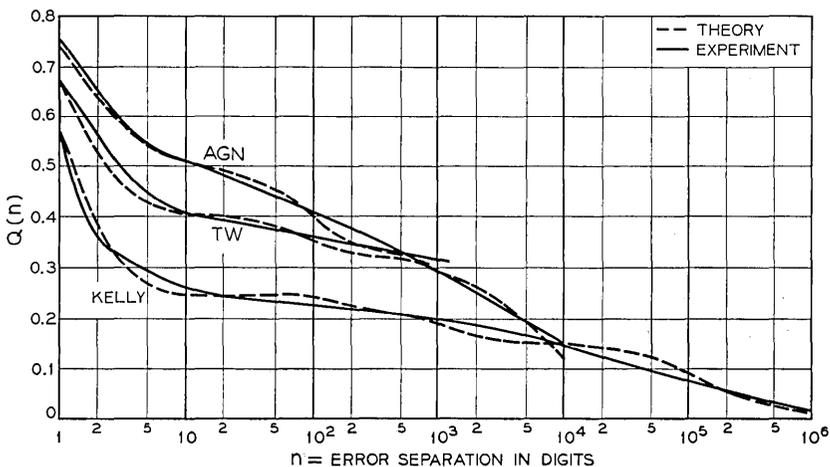


Fig. 7 — Error separation statistics for switched telephone network.

$(P_i)$. $Q(n)$ for such an error process will exhibit a gradual slope at larger error separations.

It is obvious that we could use a larger number of states in the Markov model, and match the experimental $Q(n)$ curves to any desired degree of accuracy. At this point we should consider whether it is an individual channel or an average of many channels that we wish to match. Usually it will be the former, in which case a three-state (or perhaps even a two-state) Markov model will be satisfactory. If it is the latter, it makes sense to seek a single (noninde-pendent) error process which provides a better match for the gradual slope observed at larger error separations. A likely candidate is the Pareto distribution proposed by Berger and Mandelbrot.[4] They also give statistical evidence which supports the renewal error process hypothesis. Sussman[5] has shown that the Pareto distribution pro-vides a good fit to the Alexander-Gryb-Nast data. It is interesting that Sussman hypothesized that the Pareto distribution may be the limiting form of "the superposition of many unrelated error-causing events," which is exactly what our model suggests.

To incorporate the Pareto distribution into our model, we would represent the cumulative distribution of error separation as

$$Q(n) = Q_1(1 - P_1)^n + Q_2(n + 1)^{-\alpha}$$

where $\alpha$ is a parameter which would be chosen so as to give the best match to the experimental data. It should be recognized that the above distribution will not be a good approximation for the Markov error process for very large values of $n$. As $n \to \infty$ the Markov dis-tribution approaches

$$Q(n) \to Q_\infty(1 - P_\infty)^n$$

where $Q_\infty$ and $P_\infty$ describe the channel with the smallest average random error rate.

In some situations the Pareto distribution may also be a good representation of an individual channel. We have implicitly assumed stationary channels. A nonstationary channel whose parameters vary rapidly with time is essentially equivalent to the average of a large number of stationary channels, each with different parameters. Such a model may be appropriate for digital communication systems using radio links. On the other hand, a slowly varying nonstationary chan-nel is essentially equivalent to a single stationary channel, since the parameters will not change appreciably during any message of rea-

sonable length. This kind of model appears to be appropriate for digital communication systems using paired cable or coaxial cable.

To summarize, we feel that the Markov model is a good representation for the error process of an individual digital channel. The Markov model also explains the observed measurements for the average of a large number of digital channels, and leads naturally to the idea of using the Pareto distribution to approximate the behavior of a Markov error process with many states.

## VI. ACKNOWLEDGMENTS

## REFERENCES

1. Gilbert, E. N., "Capacity of a Burst-Noise Channel," B.S.T.J., *39*, No. 5 (September 1960), pp. 1253–1266.
2. Elliott, E. O., "Estimates of Error Rates for Codes on Burst-Noise Channels," B.S.T.J., *42*, No. 5 (September 1963), pp. 1977–1998.
3. Elliott, E. O., "A Model of the Switched Telephone Network for Data Communications," B.S.T.J., *44*, No. 1 (January 1965), pp. 89–109.
4. Berger, J. M., and Mandelbrot, B., "A New Model for Error Clustering in Telephone Circuits," IBM J. Res. and Development, *7*, No. 3 (July 1963), pp. 224–236.
5. Sussman, S. M., "Analysis of the Pareto Model for Error Statistics of Telephone Circuits," IEEE Trans. Commun. Syst., *11*, No. 2 (June 1963), pp. 213–221.
6. McCullough, R. H., "Ternary Codes for Regenerative Digital Transmission Systems," Ph.D. dissertation, Polytechnic Institute of Brooklyn, June 1967.
7. Ash, R., *Information Theory*, New York: Interscience Publishers, 1965.
8. Fultz, K. E., and Penick, D. B., "The T1 Carrier System," B.S.T.J., *44*, No. 7 (September 1965), pp. 1405–1452.
9. Travis, L. F., and Yaeger, R. E., "Wideband Data on T1 Carrier," B.S.T.J., *44*, No. 8 (October 1965), pp. 1567–1604.
10. DeWitt, R. G., and Forde, J. P., unpublished work.
11. Alexander, A. A., Gryb, R. M., and Nast, D. W., "Capabilities of the Telephone Network for Data Transmission," B.S.T.J., *39*, No. 3 (May 1960), pp. 431–476.
12. Townsend, R. L., and Watts, R. N., "Effectiveness of Error Control in Data Communications over the Switched Telephone Network," B.S.T.J., *43*, No. 6 (November 1964), pp. 2611–2638.
13. Kelly, J. P., "Test of Data Transmission via Switched Message Networks for SAGE and BUIC," Mitre Corp. report, March 1964.

# Quantizing Noise and Data Transmission

## By JAMES E. MAZO

(Manuscript received February 20, 1968)

*Methods for calculating the power in the quantizing noise on digital transmission facilities have been known for some time. A more difficult but unavoidable problem is the effect that this noise has on data signals intended for analog transmission. This paper demonstrates that to assume that the noise will behave as a white Gaussian noise process will always (except for a simple factor) yield an upper bound on the probability of error when no companding is present. We assume that linear detection will be used, as for a PAM system, and the result is true whether or not filtering or demodulation is involved. Results are illustrated by applying them to a model of an existing VSB modem whereby the additional degradations resulting from data set imperfections are included as added baseband noise.*

*A modem operating perfectly would make no errors at all at the higher transmission levels. For example, with no companding, a set with an eight-level eye closed by even 30 percent would not yield errors for input powers down to $-15$ dBm. Thus quantizing noise is not a basic limiting factor in the error rate for all input levels. A similar rigorous theory is not available for compandored systems, but for special situations reasonable estimates can be made. For logarithmic companding and eight-level VSB transmission, worst case estimates indicate error rates about $10^{-6}$ for one link of $T1$ carrier.*

## I. INTRODUCTION AND SUMMARY

The T1 carrier system is a digital transmission scheme for analog signals.[1] Even though the digits in the coded bit stream might be transmitted without error, when the analog signal (which may in fact be a data signal designed for analog facilities) is reconstructed at the receiving terminal, quantizing noise is inevitably added and can be large enough to cause errors in the customer's data.

We show that, under some simple constraints between sampling rates and bandwidths which are satisfied in practice, and independent of the particular data signal used, an upper bound on error rates is obtained if the quantizing noise is assumed to be a white Gaussian process of power $\Delta^2/12$ and bandwidth $1/2T_1$.* We assume that linear detection will be used, and the result is true even if additional filtering is done (as one might do with a receiving filter). And it is true whether or not a demodulation process takes place. Using the model in Fig. 1 for the digital transmission scheme, results are specialized to obtain error rates for eight-level VSB transmission (Fig. 2). Imperfections of the data set are included as added baseband noise. If it were not for these imperfections, error free transmission would result over an appreciable range of power levels (see Table I). For a logarithmic compandor and VSB data, even using worst case estimates, the error rate for one link is quite low, about $10^{-6}$.

## II. QUANTIZED TRANSMISSION SCHEME

Let us consider a transmission scheme for a single channel that, for our purposes, typifies the T1 carrier system. As suggested in Fig. 1, the signal to be transmitted is assumed not to have any power beyond $B$ Hz. The signal is sampled at the Nyquist rate $T_1 = 1/(2B)$ and these samples are passed through an instantaneous nonlinear device with characteristic $v_{out} = F(v_{in})$. The compressed samples are then quantized by a uniform quantizer of step size $\Delta$, and coded into binary sequences. The binary sequences are assumed to be transmitted without error and the process is reversed: sequences are decoded into pulses, expanded according to the inverse function $F^{-1}(x)$ and the impulses are used to excite an ideal filter of bandwidth $B$ and amplitude $T_1$.† A receiving filter $G$ generally follows the ideal filter and we include this in our description, although it would not be part of a T1 transmission system. If the bandwidth of $G$ is entirely contained in $B$ then one may consider the impulses to excite $T_1G$ directly.

To be more specific, we are concerned with two particular compandor characteristics $F(x)$. One is $F(x) = x$, that is, quantizing

---

* Here $\Delta$ is the quantizer step size and $1/T_1$ is the sampling rate. Also this statement is true only modulo a simple factor given in the text.
† The amplitude gain of the ideal output filter for the carrier system is chosen to be $T_1$ in order that the signal component will undergo no gain relative to its sampled values at the transmitter.
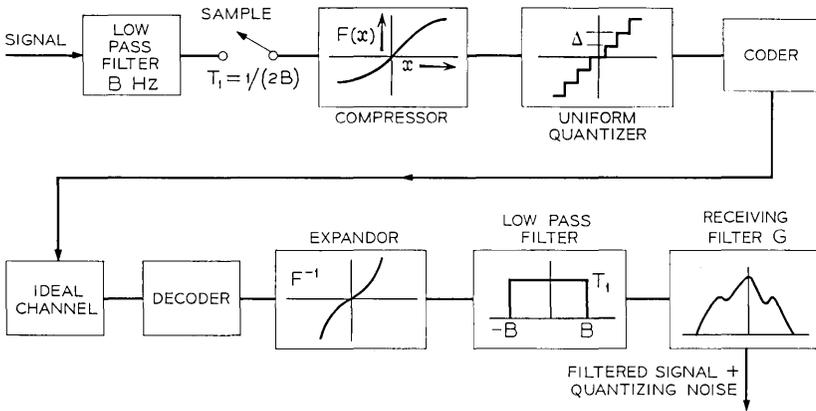
Fig. 1 — Essential elements of quantized transmission scheme.
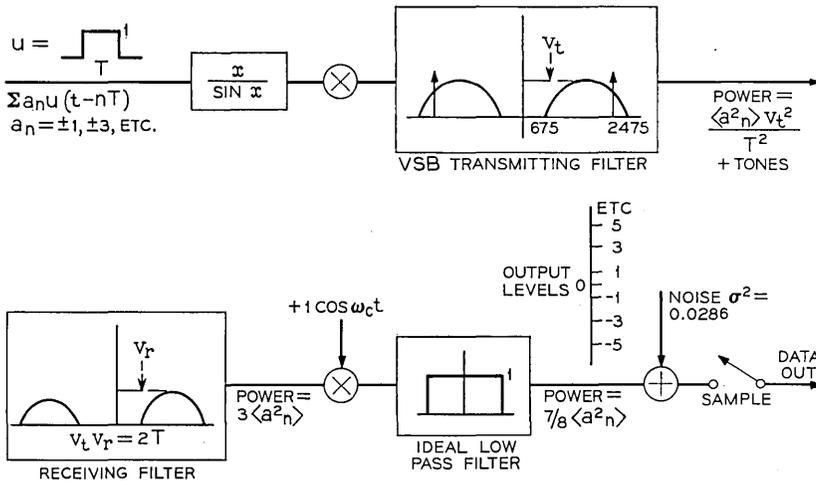


Fig. 2 — Model of back-to-back transmission of VSB modem. The noise added at baseband represents imperfections of the modem.

TABLE I — QUANTIZING NOISE*

| Mean square $P$(dBm) | $n_{max}$ |
|---|---|
| − 5 | 0.221 |
| −10 | 0.394 |
| −15 | 0.698 |
| −20 | 1.245 |
| −25 | 2.21 |
| −30 | 3.94 |
| −35 | 6.98 |
| −40 | 12.45 |
| −45 | 22.1 |

* Peak values of quantizing noise as a function of input power $P$. The noise scale is such that a perfect receiver would make no errors for $n_{max} < 1$. The power scale is such that the quantizer overloads at an instantaneous power of +6 dBm.

without compandoring. The other case is (in normalized units)

$$F(x) = -F(-x)$$

$$F(x) = \frac{\ln (1 + \mu x)}{\ln (1 + \mu)}, \qquad 0 \leqq x \leqq 1 \tag{1}$$

$$= 1 \qquad\qquad x > 1,$$

where $\mu$, the degree of compandoring, is large. Typically, $\mu = 100$ for a good approximation to existing devices.

Finally, when specific values are required, we assume 7 bit coding to be used for the quantized samples and use $\Delta = 1/63$.

We hasten to add that quantizing noise is not the only degrading factor for the existing $T_1$ facilities. Apparently mismatch and mistracking of compressor and expandor cause nonlinearities which are responsible for peculiar behaviors of error rate versus signal power curves.[2]

III. GENERAL THEORY

Let us represent the signal $l(t)$ which is to be sampled and quantized by

$$l(t) = x(t) \cos \omega_c t - y(t) \sin \omega_c t, \tag{2}$$

and the sampling wave as

$$\sum_k \delta(t - kT_1 - \tau), \tag{3}$$

where the random timing phase is uniformly distributed over the interval $0 \leqq \tau \leqq T_1$. The pulse trains representing (2) immediately after sampling, compression, and quantization are given by expressions (4), (5), and (6), respectively.

$$\Sigma l(kT_1 + \tau)\delta(t - kT_1 - \tau); \tag{4}$$

$$\Sigma l_{\text{comp}}(kT_1 + \tau)\delta(t - kT_1 - \tau); \tag{5}$$

$$\Sigma \hat{l}_{\text{comp}}(kT_1 + \tau)\delta(t - kT_1 - \tau), \tag{6}$$

where

$$l_{\text{comp}}(kT_1 + \tau) = F[l(kT_1 + \tau)]$$

is the compressed sample value and $\hat{l}_{\text{comp}}(t)$ is the particular one of the $(2^N - 1)$ levels that the quantizer output gives as the value for $l_{\text{comp}}(t)$. If we let the subscript "exp" stand for the result of operation of the expandor at the receiving terminal, then the impulse associated with time $(kT_1 + \tau)$ has area

$$[\hat{l}_{\text{comp}}(kT_1 + \tau)]_{\text{exp}} = l(kT_1 + \tau) + \epsilon(kT_1 + \tau). \tag{7}$$

Because the expandor has as its input an estimate of the compressed pulse area, the error term $\epsilon(kT_1 + \tau)$ is not zero but may take any value in an interval, that is,

$$\epsilon(t) \; \varepsilon \; \left[ \frac{-\Delta(t)}{2} , \frac{\Delta(t)}{2} \right]. \tag{8}$$

The spread $\Delta(t)$ that the quantizing error may take is not necessarily equal to the quantizer step size $\Delta$ when companding is present, but is given by the formula (see Appendix A)

$$\Delta(t) = \frac{\Delta}{|F'[l(t)]|}. \tag{9}$$

In (9), $F'[l(t)]$ is the derivative of the compressor characteristic evaluated at that input amplitude of the signal at the time of the sampling. The error signal generated at the output of the receiving filter is obtained by convolving the impulse train

$$\Sigma \epsilon(kT_1 + \tau)\delta(t - kT_1 - \tau) \tag{10}$$

with the impulse response $T_1 g(t)$ of the receiving filter.* Denoting

---

* Again, $g(t)$ is associated with the receiving filter of the data set and the constant $T_1$ is the gain of the ideal output filter of the carrier system.

this noise by $n_l(t)$ we have

$$n_l(t) = T_1 \Sigma \epsilon(kT_1 + \tau) g(t - kT_1 - \tau). \tag{11}$$

To proceed further we make the assumption that the quantizing error $\epsilon(t)$ of the output sample in (10) is uniformly distributed over the interval

$$\left[ \frac{-\Delta(t)}{2}, \frac{\Delta(t)}{2} \right],$$

thus having mean zero and variance $\Delta^2(t)/12$, and that different sample errors are independent. Notice that the latter assumption is not the same as assuming that different sample *values* are independent.

## IV. SUMS OF UNIFORM VARIATES

As (11) illustrates, a basic problem which must be dealt with is the probability distribution of sums of independent and uniformly distributed random variables. We will obtain an upper bound on the tail probabilities of interest by applying the technique of the Chernoff bound.[3, 4] This bounding technique states that if a probability density $p(x)$ has a moment generating function (*mgf*) $M(s)$, where

$$M(s) = \int_{-\infty}^{\infty} [\exp(sx)] p(x) \, dx, \tag{12}$$

then

$$Q = \text{Prob} [x \geq a] \leq M(s) \exp(-sa), \qquad s \geq 0. \tag{13}$$

Thus to obtain an upper bound one simply multiplies the moment generating function by an exponential, both evaluated at an arbitrary positive $s$. Actually it is known that there is a best $s$ to choose, and it is that one, if it exists, which satisfies the equation

$$\frac{d}{ds} \ln M(s) = a. \tag{14}$$

Equation (14) assures a stationary value for the right side of (13) and it can be shown that such an $s$ in fact minimizes $M(s)e^{-sa}$.

For example, for a Gaussian variate of mean $m$ and variance $\sigma^2$, the moment generating function is well known to be given by

$$M(s) = \exp \left[ ms + \frac{s^2 \sigma^2}{2} \right]. \tag{15}$$

Thus the best $s$ to choose is, using (14),

$$s = \frac{(a - m)}{\sigma^2}. \tag{16}$$

Notice that only if $a \geqq m$ is this $s \geqq 0$. Thus, as long as $a$ is greater than the mean, we have for the Gaussian case

$$Q \leqq \exp\left[-\frac{(a - m)^2}{2\sigma^2}\right], \tag{17}$$

where (17) results from using (15) and (16) in (13). For the Gaussian variate under discussion the exact answer is also well known to be given by

$$Q = \tfrac{1}{2} \operatorname{erfc} \frac{(a - m)}{(2)^{\frac{1}{2}}\sigma}, \tag{18}$$

where erfc $x$ is the coerror function.[5] In addition, equation (7.1.13) of Ref. 5 states that

$$\exp(-x^2) \leqq (\pi)^{\frac{1}{2}}[x + (x^2 + 2)^{\frac{1}{2}}][\tfrac{1}{2} \operatorname{erfc} x], \tag{19}$$

and hence the difference between the Chernoff answer (17) and the exact answer (18) for the Gaussian case is no more than the multiplicative factor $(\pi)^{\frac{1}{2}}[(\rho)^{\frac{1}{2}} + (\rho + 2)^{\frac{1}{2}}]$ where $(\rho)^{\frac{1}{2}} = (a - m)/[(2)^{\frac{1}{2}}\sigma]$.

We modify this procedure for our problem with the following obvious lemma.

*Lemma 1: If $G(s)$ is an upper bound for the moment generating function, that is, $M(s) \leqq G(s)$ for all $s$, then*

$$Q \leqq e^{-sa}G(s), \qquad s \geqq 0. \tag{20}$$

In particular, a positive $s = s_0$ which satisfies

$$\frac{d}{ds} \ln G(s) \bigg|_{s=s_0} = a \tag{21}$$

is legitimate.

Next consider a random variable $x$ which is uniformly distributed over $[-\Delta/2, \Delta/2]$. The variance of this variable is $\Delta^2/12$, and it has a moment generating function $M(s)_{\text{unif}}$

$$M(s)_{\text{unif}} = \frac{\sinh\frac{s\Delta}{2}}{\frac{s\Delta}{2}} = \sum_{n=0}^{\infty} \left(\frac{s\Delta}{2}\right)^{2n} \frac{1}{(2n + 1)!}. \tag{22}$$

Now the $n$th term of the sum in (22) is positive and upper bounded by

$$\left[ \left(\frac{s\Delta}{2}\right)^2 \frac{1}{6} \right]^n \frac{1}{n!}.$$

Hence

$$M(s)_{\text{unif}} \leqq \sum_{n=0}^{\infty} \left[ \left(\frac{s\Delta}{2}\right)^2 \frac{1}{6} \right]^n \frac{1}{n!} = \exp \left[ \frac{\Delta^2}{12} \frac{s^2}{2} \right]. \tag{23}$$

Thus we have shown that the moment generating function of a zero-mean uniform density is upper bounded by that of a zero-mean Gaussian having the same variance.* If the uniform variable has mean $m$ the theorem is still true if we use instead the moment generating function of a Gaussian with mean $m$.

We are now ready to write down a whole class of random variables which have moment generating functions upper bounded by those of a Gaussian of the same variance. Suppose the result is true for two independent random variables, $x$ and $y$, of variances $\sigma_x^2$ and $\sigma_y^2$, namely

$$M_x(s) \leqq \exp \left[ \frac{s^2 \sigma_x^2}{2} \right]$$

$$M_y(s) \leqq \exp \left[ \frac{s^2 \sigma_y^2}{2} \right]. \tag{24}$$

Then using the theorem that the moment generating function of a sum of two independent random variables is the product of their individual moment generating functions, we have

$$M_{x+y}(s) = M_x(s) M_y(s) \leqq \exp \left[ \frac{s\sigma_x^2}{2} \right] \exp \left[ \frac{s\sigma_y^2}{2} \right]$$

$$= \exp \left[ \frac{s^2}{2} (\sigma_x^2 + \sigma_y^2) \right] = \exp \left[ \frac{s^2 \sigma_{x+y}^2}{2} \right],$$

where

$$\sigma_{x+y}^2 = \sigma_x^2 + \sigma_y^2$$

is the variance of $(x + y)$. Thus the moment generating function of a sum of any number of independent uniforms of arbitrary means and variances is upper bounded by the appropriate Gaussian one (same mean and variance as the sum), and thus use of (17) through (21) provides a rigorous upper bound for tail probabilities of the sum.

* A similar theorem was discussed by Saltzberg for the case of equally spaced delta functions.[6] We have followed his method of proof here.

## V. FURTHER ANALYSIS WITHOUT COMPANDING

When no companding is present, the independent random variables $\epsilon(kT_1 + \tau)$ have variance $\Delta^2/12$ and the variance of the noise (11) is

$$\sigma_l^2 \equiv \langle n_l^2(t) \rangle = T_1^2 \cdot \frac{\Delta^2}{12} \cdot \sum_{k=-\infty}^{\infty} g^2(t - kT_1 - \tau). \tag{25}$$

We evaluate the infinite sum in (25) by using the Poisson sum formula, namely

$$\sum_{k=-\infty}^{\infty} g^2(t - kT_1 - \tau) = \frac{1}{T_1} \sum_{m=-\infty}^{\infty} \exp\left[\frac{2\pi i m(t - \tau)}{T_1}\right] G_2\left(m\frac{2\pi}{T_1}\right), \tag{26}$$

where $G_2(\omega)$ is the Fourier transform of $g^2(t)$. Now since the bandwidth of the filter $G$ is assumed not to exceed $1/2T_1$ Hz, only the $m = 0$ term of (26) contributes and we obtain

$$\sigma_l^2 = T_1 \cdot \frac{\Delta^2}{12} \cdot G_2(0). \tag{27}$$

Equation (27) implies that the noise power measured before the receiving filter is $\Delta^2/12$. This result has been obtained by Bennett[7] who also showed that the spectrum of this noise is flat across the band. Further, equation (27) is consistent with filtering white noise since

$$G_2(0) = \int_{-\infty}^{\infty} g^2(t) \, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |G(\omega)|^2 \, d\omega. \tag{28}$$

An important fact about (27) is that the received passband noise power without companding is independent of many properties of the signal. Thus it is independent of signal power and multilevel structure. It is not independent of rate, however, since this enters implicitly into the factor $G_2(0)$, and likewise it is not independent of roll-off. By halving the speed and doubling the number of levels, one decreases the noise by 3 dB, but loses 6 dB in noise margin, thus leaving one with a net loss of 3 dB in noise margin. Thus it is best to use as few levels as possible consistent with given speed objectives, at least if the quantizing noise behaves anything like Gaussian noise.

Let us discuss further some statistical aspects of the quantizing noise at baseband. The "line" signal must be demodulated as in VSB transmission by multipling the (filtered) received signal by $\cos \omega_c t$ and eliminating double frequency components. We represent the impulse response $g(t)$ of the passband receiving filter $G$ by

$$g(t) = g_x(t) \cos \omega_c t - g_y(t) \sin \omega_c t. \tag{29}$$

Further we specialize to the practical constraints* $\omega_c > B_i$ and $2\pi/T_1$ $> 2(\omega_c + B_i)$. The demodulated noise is

$$n_b(t) = \frac{T_1}{2} \Sigma \epsilon(kT_1 + \tau)\{g_x(t - kT_1 - \tau) \cos (\omega_c kT_1 + \omega_c \tau)$$

$$+ g_y(t - kT_1 - \tau) \sin [\omega_c kT_1 + \omega_c \tau]\}. \qquad (30)$$

The general expression (29) can be simplified for a VSB receiving filter which is symmetric about midband frequency $\omega_1$, and linear phase characteristic, by writing

$$g(t) = g_1(t) \cos \omega_1 t$$

$$\omega_c - \omega_1 = \pi/2T, \qquad (31)$$

where $1/T$ is the symbol repetition frequency. Of course a $g_x$ and a $g_y$ may be immediately written down from (31).

From (30) we derive in Appendix B, equation (32) for the baseband variance $\sigma_b^2(t)$:

$$\sigma_b^2(t) = \frac{1}{2} \left(\frac{1}{4} \frac{\Delta^2}{12}\right) T_1 [G_{2x}(0) + G_{2y}(0)], \qquad (32)$$

where $G_{2i}(\omega)$ is the Fourier transform of $g_i^2(t)$. We now will show that this result is identical to the baseband noise power that would appear if flat Gaussian noise of power $\Delta^2/12$ were on the line. We do not regard this as obvious; in fact it is not true that the *signal* power at baseband is the same as if one had Gaussian noise of the same power and spectrum on the line that the signal has. The proof depends on a few simple observations. If passband Gaussian noise is represented by

$$n(t) = n_x(t) \cos \omega_c t - n_y(t) \sin \omega_c t, \qquad (33)$$

then

$$\sigma_n^2 \equiv \langle n^2 \rangle = \langle n_x^2 \rangle = \langle n_y^2 \rangle, \qquad (34)$$

and so baseband noise power is $\sigma_n^2/4$. Next we notice that white Gaussian noise, having same total power as quantizing noise over the band $(-1/2T_1, 1/2T_1)$ Hz, has two sided spectral density

$$N(\omega) = N_0/2 = \frac{\Delta^2}{12} \cdot T_1 \text{ watts per cycle.} \qquad (35)$$

---

* $B_i$ is the bandwidth of $g_i(t), i = x, y$.

Thus the Gaussian noise power out of the receiving filter $G$ would be

$$
\begin{aligned}
\sigma_n^2 &= \frac{\Delta^2}{12} \cdot T_1 \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \mid G(\omega) \mid^2 \\
&= \frac{\Delta^2}{12} T_1 \int_{-\infty}^{\infty} g(t)^2 \, dt \\
&= \frac{\Delta^2}{12} T_1 \bigg[ \int_{-\infty}^{\infty} \frac{g_x^2(t) + g_y^2(t)}{2} \, dt + \int_{-\infty}^{\infty} \frac{g_x^2(t) - g_x^2(t)}{2} \cos 2\omega_c t \, dt \\
&\qquad\qquad + \int_{-\infty}^{\infty} g_x(t) g_y(t) \sin 2\omega_c t \, dt \bigg].
\end{aligned}
\tag{36}
$$

Since neither $g_x(t)$ nor $g_y(t)$ are assumed to have any frequencies as high as $\omega_c$ the last two integrals above vanish. The final remark that completes the proof is

$$
\int_{-\infty}^{\infty} g_i^2(t) \, dt = G_{2i}(0).
$$

Thus noncompandored quantizing noise behaves, at least concerning, its power, as zero mean white Gaussian noise, flat over the band $(-1/2T_1, 1/2T_1)$ Hz, and total power $\Delta^2/12$. This statement is true with or without demodulation.

One would like to go further and treat the baseband noise as zero mean Gaussian of variance given by (32). There is a justification for making this additional step. Recall the result of Section IV, which stated that if

$$
z = \Sigma \Delta_i
$$

$$
\langle z \rangle = \Sigma \langle \Delta_i \rangle
\tag{37}
$$

$$
\sigma^2 = \operatorname{var} z = \tfrac{1}{12} \Sigma \Delta_i^2
$$

is a sum of independent and uniformly distributed variates $\Delta_i$, then (provided $\sigma^2 = 1/12 \Sigma \Delta_i^2 < \infty$) for all $A$ such that $A > \Sigma \langle \Delta_i \rangle$,

$$
\operatorname{Prob}(z > A) \leqq (\pi)^{\frac{1}{2}} [(\rho)^{\frac{1}{2}} + (\rho + 2)^{\frac{1}{2}}] P_g(A).
\tag{38}
$$

In (38), $(\rho)^{\frac{1}{2}} = A/[(2)^{\frac{1}{2}} \sigma]$, and $P_g(A)$ is the probability that a Gaussian variate of the same mean and variance as $z$ is greater than $A$. Since $P_g(A)$ depends exponentially on $\rho$, the coefficient structure in (38) is not nearly as important as $P_g(A)$. We would like to argue (but not prove) that ignoring the coefficient in (38), that is, simply

assuming Gaussian behavior, is quite accurate for the baseband noise (30) for the error rates of practical interest.

Thus consider eight-level, 50 percent roll off, transmission over our hypothetical "noncompandored $T_1$" transmission facility. From (30) and (31)

$$n_b(t) = \frac{T_1}{2} \sum_k \epsilon(kT_1 + \tau)h_1(t - kT_1 - \tau)$$

$$\cdot \cos [(\omega_c - \omega_1)(t - kT_1 - \tau) + \omega_c(kT_1 + \tau)], \qquad (39)$$

where, according to an appropriate normalization,

$$h_1(t) = \frac{4v_r}{\pi T} \frac{\cos \frac{\pi t}{T}}{1 - \left(\frac{2t}{T}\right)^2}. \qquad (40)$$

Notice that since $h_1 \approx 1/t^2$, $t$ large, the sum in (39) is bounded. A computer study of (39) for various values of $t$ and $\tau$ shows this bound to be not too sensitive (about 5 percent variations) to choices of $t$ and $\tau$. Numerically we find

$$| n_b(t) | \leqq \frac{T_1}{2} \cdot \frac{\Delta}{2} \cdot \left(\frac{4v_r}{\pi T}\right)(5.31). \qquad (41)$$

From the sum formula (26) the variance of (39) is obtained. We calculate

$$\sigma_b^2 = \frac{1}{4} \frac{\Delta^2}{12} \frac{T_1 v_r^2}{T}. \qquad (42)$$

Thus a peak-to-mean square ratio of the baseband noise power may easily be shown to be 15 dB. To obtain some insight from this value, consider the question of how many $(N)$ identical independent, zero-mean uniform densities one would have to convolve to get a peak to rms value of 15 dB; the answer is $N = 10$. Ten uniforms generate, we feel, a reasonable approximation to a Gaussian curve. As a check, consider that our ten uniform densities each have range $[-0.5, 0.5]$. To check (not prove) the approximation on the tails we calculate

$$\text{Prob [sum} \geqq 4.5] = \frac{1}{10!} = 2.75 \times 10^{-7}.$$

The Gaussian assumption gives $4.46 \times 10^{-7}$. Thus we will assume that for error rates $> 10^{-7}$ the Gaussian assumption will yield reasonably accurate answers, not just being a bound in the sense discussed above.

The above theory showing that noncompandored quantizing noise may be considered to be additive white Gaussian noise with zero mean, variance $\Delta^2/12$, of 4 kHz bandwidth has been compared with the experimental results of Gustafson on the performance of the VSB (203) data set which operates at 5400 bits per second. Fortunately an experimental curve is available for error rate versus signal-to-noise ratio without companding and this is shown in Fig. 3 along
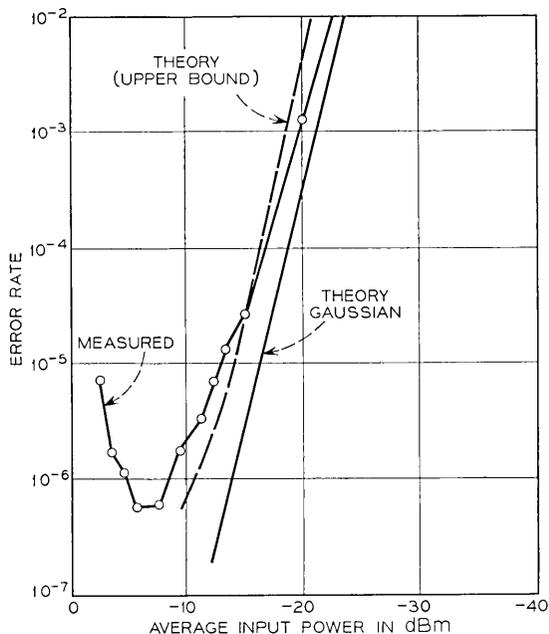


Fig. 3 — Comparison of experimental and theoretical error rates for one link of noncompandored transmission. Theory neglects overload distortion. Instantaneous input power of +6 dBm is the onset of overload.

with the results of present theory* (for one link of T1.) The rise in the experimental curve at high input signal power results from overload distortion of the quantizer which has been neglected for the present analysis. Overload occurs at a peak power of +6 dBm on the scale used in Fig. 3, and thus the peak power to average power for the eight-level VSB set (including pilot tones) appears to be around 11 dB. In general the observed error rate is higher than the theoret-

---

* To model the performance of the actual 203 receiver, an additional noise source is included at baseband, as suggested by Saltzberg[8] and shown in Fig. 2. The baseband $S/N$ for this noise is chosen to be 28.08 dB. This noise alone would yield an error rate of $2.5 \times 10^{-9}$.

ical prediction, and for low $P_e$ is even higher than the theoretical upper bound. Perhaps this is caused by other distortions in the system not considered here.

For multiple links of transmission one should take the quantizing noise to have the same properties as above, but the total noise power is $N \cdot \Delta^2/12$, where $N$ is the number of links.

We wish to emphasize that the curve drawn in Fig. 3 does not represent any theoretical ideal; we have tried to understand the performance of an existing data set and its imperfections. Actually, if the data set were functioning perfectly, there is a range of input power where no errors would be made. We have normalized units so that no errors can be made if the baseband noise is less than unity in magnitude. Table I shows the peak value of quantizing noise calculated from equation (41) as a function of input power measured in the same units as in Fig. 3. An input power of $-15$ dB would be near typical operating levels. If the data set were imperfect but the eight-level eye were no more than 30 percent closed (but one had perfect timing), then one would still not make errors down to $-15$ dBm. In general we see that quantizing noise is not a *basic* limiting factor on the error rate for all input power levels.

VI. ANALYSIS WITH COMPANDING

Equation (9) indicates that the derivative of the compressor characteristic is an important quantity. For the logarithmic curve given in equation (1),

$$F'(x) = \frac{\mu}{\ln(1 + \mu)} \frac{1}{1 + \mu \mid x \mid}.$$ (43)

The average of $\Delta^2(t)$ now is not $\Delta^2/12$ but is

$$\sigma_{av}^2 = \langle \Delta^2(t) \rangle = \frac{\Delta^2}{12} \cdot \left\langle \frac{1}{\mid F'(x) \mid^2} \right\rangle$$

$$= \frac{\Delta^2}{12} \left[ \frac{\ln(1 + \mu)}{\mu} \right]^2 (1 + 2\mu\langle \mid x \mid \rangle + \mu^2 P),$$ (44)

where the average power $P = \langle x^2 \rangle$. Now $\langle \mid x \mid \rangle$ cannot be less than zero nor more than $(P)^{1/2}$. Hence

$$\kappa(1 + \mu^2 P) \leqq \sigma_{av}^2 \leqq \kappa[1 + \mu(P)^{\frac{1}{2}}]^2$$ (45)

where

$$\kappa = \frac{\Delta^2}{12} \left[ \frac{\ln(1 + \mu)}{\mu} \right]^2.$$ (46)

The lower and upper bounds in (45) indicate that for large $\mu(P)^{\frac{1}{2}}$ the average noise power is not a sensitive function of the probability density of the input signal. The knowledge of $\sigma_{av}^2$ cannot be used here to obtain a strict upper bound for the probability of error as was done in the uncompandored case, for the "instantaneous" noise variance is correlated with signal values. Thus large input signals "see" bigger step sizes, in effect, than smaller inputs would. One concludes from this that for multilevel transmission the outermost levels would have the greatest noise associated with them.

To make exact calculations on this matter is a difficult task, and we confine ourselves to some estimates of the effects. Estimates can be obtained by restricting attention to special sequences. Thus for an eight-level PAM system let an arbitrary sequence consisting only of the outer levels ($\pm 7$) be transmitted, and compare this with another sequence consisting of ($\pm 5$) transmitted in place of ($\pm 7$). Then the quantizing noise will be—considering the $\mu^2$ term in (44) to be of principal importance—in the ratio $7^2/5^2$. Thus the outer level will have, in this circumstance, 3 dB more noise than the next inner level. The contrast between these levels will be somewhat lessened in a random sequence using all levels, but it is clear that the 3 dB number quoted here provides an upper bound to the difference.

Worst case estimates of error rate in the compandored case may be made by replacing $\Delta^2/12$ in (42) by the upper bound for $\sigma_{av}^2$ given in (45), and finally using peak power instead of average power in (45).

For the eight-level VSB system considered previously, operating on $T1$ facilities, this procedure yields error rates of $10^{-5} - 10^{-6}$ over one transmission link (for the interesting ranges of input power).

## VII. ACKNOWLEDGMENT

## APPENDIX A

*Derivation of Quantization Error*

Equation (9) of the text relating the output sample error $\Delta(t)$ to step size $\Delta$, compressor characteristic $F(x)$, and signal amplitude $l(t)$ at time of sampling is easy to derive if the chain rule is used to differentiate the relation

$$F^{-1}[F(x)] = x \qquad (47)$$

to obtain

$$\frac{dF^{-1}_{(u)}}{du}\bigg|_{u=F(x)} \frac{dF(v)}{dv}\bigg|_{v=x} = 1. \tag{48}$$

Now clearly, if the error made when $l_{\text{comp}}$ is quantized is small, that is, if $\Delta$ is small, then

$$\begin{matrix}\text{output} \\ \text{sample}\end{matrix} = F^{-1}(l_{\text{comp}}) + \Delta \frac{dF^{-1}}{du}\bigg|_{u=l_{\text{comp}}}, \tag{49}$$

or

$$\begin{matrix}\text{output} \\ \text{sample}\end{matrix} = l + \frac{\Delta}{F'(l)}, \tag{50}$$

where (48) has been applied to (49) to obtain (50).

APPENDIX B

*Derivation of $\sigma_b$*

Squaring (30) and averaging over $\{\epsilon\}$ gives

$$\sigma_b^2(t) = \frac{T_1^2}{4}\cdot\frac{\Delta^2}{12}\cdot \sum \{g_x^2(t - kT_1 - \tau)$$

$$\cdot\cos^2[\omega_c kT_1 + \omega_c \tau] + g_y^2(t - kT_1 - \tau)$$

$$\cdot\sin^2[\omega_c kT_1 + \omega_c \tau] + 2g_x(t - kT_1 - \tau)g_y(t - kT_1 - \tau)$$

$$\cdot\sin[\omega_c kT_1 + \omega_c \tau)]\cos[\omega_c kT_1 + \omega_c \tau]\} \tag{51}$$

or,

$$\sigma_b^2(t) = \frac{1}{2}\left(\frac{T_1^2}{4}\frac{\Delta^2}{12}\right)\sum [g_x^2 + g_y^2]$$

$$+ \frac{1}{2}\left(\frac{T_1^2}{4}\frac{\Delta^2}{12}\right)\{\sum (g_x^2 - g_y^2)\cos[2\omega_c kT_1 + 2\omega_c \tau]$$

$$+ 2\sum g_x g_y \sin[2\omega_c kT_1 + 2\omega_c \tau]\}. \tag{52}$$

All the sums in (52) may be evaluated using the Poisson sum formula quoted in equation (26). The first term on the right of (52) is simplest to handle. Since $g_i$ has no frequencies higher than $1/2T_1$, the Fourier transform $G_{2i}(\omega)$ of $g^2(t)$ has support contained in $[-2\pi/T_1, 2\pi/T_1]$, and further, since it is a convolution, $G_{2i}(\pm 2\pi/T_1) = 0$. Thus

$$\frac{1}{2}\left(\frac{T_1^2}{4}\frac{\Delta^2}{12}\right)\sum (g_x^2 + g_y^2) = \frac{1}{2}\left(\frac{1}{4}\frac{\Delta^2}{12}\right)T_1[G_{2x}(0) + G_{2y}(0)]. \tag{53}$$

The other sums in (52) are all zero, for a typical sum is (where $\theta = 2\omega_c T_1$)

$$\sum f(t - nT - \tau) \cos [n\theta + \varphi]$$

$$\equiv \cos \left[ t \frac{\theta}{T_1} + \varphi - \frac{\tau\theta}{T_1} \right] \sum f(t - nT - \tau) \cos (t - nT_1 - \tau) \frac{\theta}{T_1}$$

$$+ \sin \left[ t \frac{\theta}{T_1} + \varphi - \frac{\theta}{T_1} \right] \sum f(t - nT - \tau) \sin (t - nT_1 - \tau) \frac{\theta}{T_1} \cdot$$

$$(54)$$

The sum formula is now directly applicable to the functions $f(t) \cos 2\omega_c t$ and $f(t) \sin 2\omega_c t$. The functions have Fourier transforms which, according to the discussion following equation (29), vanish at $\omega = \pm 2\pi/T_1 \cdot k$, where $k$ is any integer, *including* zero. The results claimed follow.

REFERENCES

1. Fultz, K. E. and Penick, D. B., "The $T1$ Carrier System," B.S.T.J., *44*, No. 7 (September 1965), pp. 1405–1451.
2. McNees, K. G., unpublished work.
3. Chernoff, K., "A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on a Sum of Observations," *Ann. Math. Stat. 23*, No. 4 (December 1952), pp. 493–507.
4. Wozencraft, J. M. and Jacobs, I. M., *Principles of Communication Engineering*, New York: John Wiley and Sons, 1965, pp. 97–106.
5. National Bureau of Standards, *Handbook of Mathematical Functions*, Washington, D. C.: U. S. Government Printing Office, 1964, Chapter 7.
6. Saltzberg, B. R., "Intersymbol Interference Error Bounds with Application to Ideal Bandlimited Signaling," IEEE Trans. Inform. Theory, *14*, No. 4 (July 1968), pp. 563–568.
7. Bennett, W. R., "Spectra of Quantized Signals," B.S.T.J., *27*, No. 3 (July 1948), pp. 446–472.
8. Saltzberg, B. R., unpublished work.

# On the Solutions of Equations for Nonlinear Resistive Networks

By A. N. WILLSON, JR.

*Several theorems are proved concerning the solutions of equations that arise in the study of resistive nonlinear electrical networks. The first, an existence and uniqueness theorem, applies to equations describing an interesting class of networks which includes certain active and nonreciprocal networks for which the existence and uniqueness of solutions has not previously been established. A method of computing bounds on the location of the solutions is given, and two iterative techniques are presented for computing the solutions. It is proved that the iterative techniques converge for a subclass of the equations which also includes equations describing certain active and nonreciprocal networks. Finally, the rate of convergence of the iterative techniques is compared with that of another well-known iterative technique and some practical computational aspects are pointed out. Computations for two example problems, not reported here, were carried out to show the practicality of applying these iterative techniques to the equations of specific networks.*

## I. INTRODUCTION

In this paper we consider the solution of the equation

$$F(x) + Ax = B \tag{1}$$

where $x \equiv \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ is a point in the $n$-dimensional Euclidean space $E^n$,

$F(x) \equiv \begin{bmatrix} f_1(x_1) \\ \vdots \\ f_n(x_n) \end{bmatrix}$ is a nonlinear function mapping $E^n$ into $E^n$, $A$ is an

$n \times n$ matrix of real numbers, and $B \equiv \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$ is an arbitrary point

in $E^n$. We prove (Theorem 1) that there is a unique solution of (1) if:

   (*i*) Each $f_i$ is a strictly monotone increasing function mapping $E^1$ onto $E^1$,
and

   (*ii*) The elements $a_{ij}$ of the matrix $A$ satisfy the inequality

$$a_{ii} \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|, \quad \text{for} \quad i = 1, \cdots, n.$$

We then demonstrate a straightforward method of computing bounds on the location of this solution. Finally, we present two iterative techniques for computing the solution; and prove (Theorem 3) that the two additional assumptions:

   (*iii*) Either all of the functions $f_i$ are convex, or else all $f_i$ are concave,
and

   (*iv*) $a_{ij} \leq 0$ if $i \neq j$,
are sufficient to guarantee that the iterations converge to the solution.

   Equations of type (1) occur often in the study of nonlinear electrical networks. For example, if a linear $n$-port containing resistors, independent sources, and dependent sources has a two-terminal device whose $V$ vs $I$ curve is specified by $I_i = f_i(V_i)$, for $i = 1, \cdots, n$, connected across each port, then the port voltages may often be expressed as the solution of an equation of type (1). In this case the matrix $A$ will be the $y$-parameter matrix of the $n$-port, the constant vector $B$ will account for the presence of the independent sources, and the components of the vector $x$ will be the desired port voltages.

II. ACTIVE AND NONRECIPROCAL $n$-PORTS

   In case the $n$-port of the above example contains no dependent sources and the functions $f_i$ satisfy condition (*i*) above, the existence and uniqueness of a solution of (1) follows immediately from the well-known result of R. J. Duffin.[1] In fact, with the additional assumption that the slope of each $f_i$ is bounded by positive constants the computational technique of J. Katzenelson and L. H. Seitelman may be used to compute the solution.[2] This computational technique is based upon a theorem of I. W. Sandberg which relies upon the contraction-mapping fixed point theorem.[3]

Sandberg's theorem may, in fact, be used to prove the existence and uniqueness of a solution of (1), and to construct a convergent iteration process for computing this solution, whenever the matrix $A$ is positive semidefinite* and the slope of each $f_i$ is bounded by positive constants. Other theorems which do not require that the slopes of each strictly monotone increasing $f_i$ be bounded by positive constants also exist. (For example, see Ref. 4.) These theorems guarantee existence and uniqueness of a solution of (1) whenever $A$ is positive semidefinite but do not specify a procedure for computing it.

Suppose, however, that the matrix $A$ is not positive semidefinite; that is, suppose the $n$-port in the above example is active. Then the above results no longer apply. It may often happen that the matrix $A$ is not positive semidefinite but still satisfies condition (ii) above. The matrix

$$ A = \begin{bmatrix} 1 & 1 \\ 5 & 7 \end{bmatrix}, $$

for example, has this property. It is interesting to notice that in this case the matrix $A$ will necessarily also be nonsymmetric (the corresponding $n$-port will be nonreciprocal). This follows from the fact that for symmetric matrices $A$, condition (ii) implies that $A$ is a dominant matrix[5] which, in turn, implies that $A$ is positive semidefinite. It is for this class of active nonreciprocal $n$-ports that our work provides entirely new results. Even for the passive case, however, notice that our computational techniques do not require that the slopes of the functions $f_i$ be bounded. Also, there is reason to believe that for certain problems our iteration schemes may converge more rapidly than the ones based upon the contraction mapping theorem. More is said about this in Section VII.

## III. EXISTENCE AND UNIQUENESS

Before proving the existence and uniqueness theorem we first prove a lemma which is used many times in this and the following section.

*Lemma 1:* Let the $n \times n$ matrix $A$ of real numbers satisfy condition (ii) of Section I. For $j = 1, \cdots, n$ let $p_j$ denote the $j$th component of $p \varepsilon E^n$. Let $k \varepsilon \{1, \cdots, n\}$ be chosen such that $|p_k| = max \{|p_j| : j = 1, \cdots, n\}$. Then,

---

* The $n \times n$ matrix $A$ is said to be positive semidefinite if $\langle x, Ax \rangle \geq 0$ for all $x$ in $E^n$.

$$p_k > 0 \implies \sum_{j=1}^{n} a_{kj} p_j \geqq 0,$$

and

$$p_k < 0 \implies \sum_{j=1}^{n} a_{kj} p_j \leqq 0.$$

*Proof:*

$$a_{kk} \mid p_k \mid \geqq \sum_{\substack{j=1 \\ j \neq k}}^{n} \mid a_{kj} \mid \cdot \mid p_k \mid \geqq \sum_{\substack{j=1 \\ j \neq k}}^{n} \mid a_{kj} p_j \mid \geqq \left| \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} p_j \right|.$$

Thus,

$$a_{kk} \mid p_k \mid \geqq \pm \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} p_j .$$

But then,

$$p_k > 0 \implies a_{kk} p_k \geqq - \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} p_j \implies \sum_{j=1}^{n} a_{kj} p_j \geqq 0,$$

and,

$$p_k < 0 \implies -a_{kk} p_k \geqq \sum_{\substack{j=1 \\ j \neq k}}^{n} a_{kj} p_j \implies \sum_{j=1}^{n} a_{kj} p_j \leqq 0. \qquad \square$$

*Theorem 1: There exists a unique solution of (1) whenever conditions (i) and (ii) of Section I are satisfied.*

*Proof:* We first prove that if a solution exists it is unique. Let $x^1$ and $x^2$ be solutions of (1). Then,

$$F(x^2) - F(x^1) = A(x^1 - x^2).$$

For $j = 1, \cdots, n$ let $x_j^1$ and $x_j^2$ denote the $j$th components of $x^1$ and $x^2$, respectively, and choose $k \, \varepsilon \, \{1, \cdots, n\}$ such that

$$\mid x_k^1 - x_k^2 \mid = \max \{\mid x_j^1 - x_j^2 \mid : j = 1, \cdots, n\}.$$

If $x_k^1 > x_k^2$ then, by Lemma 1,

$$f_k(x_k^2) - f_k(x_k^1) = \sum_{j=1}^{n} a_{kj}(x_j^1 - x_j^2) \geqq 0.$$

If $x_k^1 < x_k^2$ then, by Lemma 1,

$$f_k(x_k^2) - f_k(x_k^1) = \sum_{j=1}^{n} a_{kj}(x_j^1 - x_j^2) \leqq 0.$$

Both of these conclusions contradict the fact that $f_k$ is strictly monotone increasing. Thus, $x_k^1 = x_k^2$ and hence $x_j^1 = x_j^2$ for $j = 1, \cdots, n$. That is, the solution of (1) is unique, if it exists.

We prove existence of a solution by induction. For $k = 1, \cdots, n$ let

$$F_k(x) \equiv \begin{bmatrix} f_1(x_1) \\ \vdots \\ f_k(x_k) \end{bmatrix}, \qquad A_k \equiv \begin{bmatrix} a_{11} \cdots a_{1k} \\ \cdots \cdots \cdots \\ a_{k1} \cdots a_{kk} \end{bmatrix}, \qquad B_k \equiv \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}.$$

Clearly, the matrix $A_k$ satisfies condition (ii) of Section I. Also, it is clear that there exists a unique solution* of $F_1(x) + A_1 x = B_1$ for every strictly monotone increasing function $f_1$ mapping $E^1$ onto $E^1$.

Assume that there exists a solution of $F_k(x) + A_k x = B_k$ for arbitrary strictly monotone increasing functions $f_i$, $i = 1, \cdots, k$ mapping $E^1$ onto $E^1$. Then, for every real number $x_{k+1}$, the equation

$$F_k(x) + A_k x + \begin{bmatrix} a_{1,k+1} \\ \vdots \\ a_{k,k+1} \end{bmatrix} x_{k+1} = B_k$$

has a (unique) solution; since for $i = 1, \cdots, k$ the function $f_i(x_i) + a_{i,k+1} x_{k+1}$ is strictly monotone increasing from $E^1$ onto $E^1$. Let the components of this solution be denoted by $x_i = m_i(x_{k+1})$ for $i = 1, \cdots, k$. We have thus defined $k$ functions $m_i$ on $E^1$.

We now prove that for every $x_{k+1}^1, x_{k+1}^2 \, \varepsilon \, E^1$,

$$| x_{k+1}^2 - x_{k+1}^1 | \geq | m_j(x_{k+1}^2) - m_j(x_{k+1}^1) |, \quad \text{for} \quad j = 1, \cdots, k. \quad (2)$$

This inequality, incidentally, implies that each $m_i$ is continuous.

Let $x_{k+1}^1, x_{k+1}^2 \, \varepsilon \, E^1$ and choose $l \, \varepsilon \, \{1, \cdots, k\}$ such that

$$| m_l(x_{k+1}^2) - m_l(x_{k+1}^1) |$$
$$= \max \{ | m_j(x_{k+1}^2) - m_j(x_{k+1}^1) | : j = 1, \cdots, k \}.$$

Assume that $| m_l(x_{k+1}^2) - m_l(x_{k+1}^1) | > | x_{k+1}^2 - x_{k+1}^1 |$. Clearly, then, $m_l(x_{k+1}^2) - m_l(x_{k+1}^1) \neq 0$. If $m_l(x_{k+1}^2) - m_l(x_{k+1}^1) > 0$ then,

$$f_l[m_l(x_{k+1}^2)] - f_l[m_l(x_{k+1}^1)] > 0.$$

---

* We take the liberty of using the same symbol $x$ to denote points in any of the spaces $E^k$, $1 \leq k \leq n$. No confusion should arise since the subscripts on $F$ and $A$ will make our choice clear.

Also, since the matrix $A_{k+1}$ satisfies condition $(ii)$ of Section I, letting

$$p_j = m_j(x^2_{k+1}) - m_j(x^1_{k+1}), \quad \text{for} \quad j = 1, \cdots, k,$$

$$p_{k+1} = x^2_{k+1} - x^1_{k+1},$$

we have, by Lemma 1,

$$\sum_{j=1}^{k} a_{lj}[m_j(x^2_{k+1}) - m_j(x^1_{k+1})] + a_{l,k+1}(x^2_{k+1} - x^1_{k+1}) \geqq 0.$$

Thus,

$$f_l[m_l(x^2_{k+1})] + \sum_{j=1}^{k} a_{lj}m_j(x^2_{k+1}) + a_{l,k+1}x^2_{k+1}$$

$$> f_l[m_l(x^1_{k+1})] + \sum_{j=1}^{k} a_{lj}m_j(x^1_{k+1}) + a_{l,k+1}x^1_{k+1}, \qquad (3)$$

which is a contradiction since the quantity on each side of this inequality is equal to $b_l$. If $m_l(x^2_{k+1}) - m_l(x^1_{k+1}) < 0$ then,

$$f_l[m_l(x^2_{k+1})] - f_l[m_l(x^1_{k+1})] < 0.$$

By applying Lemma 1 again, as above, one arrives again at (3) with $>$ replaced by $<$. This is also a contradiction. Thus, we must have

$$| x^2_{k+1} - x^1_{k+1} | \geqq | m_l(x^2_{k+1}) - m_l(x^1_{k+1}) |,$$

and hence (2) is proved

Now, consider the function

$$\sum_{j=1}^{k} a_{k+1,j}m_j(x_{k+1}) + a_{k+1,k+1}x_{k+1}. \qquad (4)$$

Let $x^1_{k+1}, x^2_{k+1} \; \varepsilon \; E^1$ with $x^1_{k+1} < x^2_{k+1}$. Then,

$$a_{k+1,k+1} \geqq \sum_{j=1}^{k} | a_{k+1,j} |$$

implies

$$a_{k+1,k+1}(x^2_{k+1} - x^1_{k+1})$$

$$= a_{k+1,k+1} | x^2_{k+1} - x^1_{k+1} | \geqq \sum_{j=1}^{k} (| a_{k+1,j} | \cdot | x^2_{k+1} - x^1_{k+1} |).$$

But, using (2),

$$a_{k+1,k+1}(x^2_{k+1} - x^1_{k+1}) \geqq \sum_{j=1}^{k} | a_{k+1,j}[m_j(x^2_{k+1}) - m_j(x^1_{k+1})] |$$

$$\geqq -\sum_{j=1}^{k} a_{k+1,j}[m_j(x_{k+1}^2) - m_j(x_{k+1}^1)],$$

which implies

$$\sum_{j=1}^{k} a_{k+1,j} m_j(x_{k+1}^1) + a_{k+1,k+1}x_{k+1}^1 \leqq \sum_{j=1}^{k} a_{k+1,j} m_j(x_{k+1}^2) + a_{k+1,k+1}x_{k+1}^2 \ .$$

That is, the function (4) is monotone increasing. Clearly (4) is continuous. It follows, therefore, that if $f_{k+1}$ is a strictly monotone increasing function mapping $E^1$ onto $E^1$, then so is the function

$$f_{k+1}(x_{k+1}) + \sum_{j=1}^{k} a_{k+1,j} m_j(x_{k+1}) + a_{k+1,k+1}x_{k+1} \ .$$

Thus, there exists a unique solution of the equation

$$f_{k+1}(x_{k+1}) + \sum_{j=1}^{k} a_{k+1,j} m_j(x_{k+1}) + a_{k+1,k+1}x_{k+1} = b_{k+1} \ .$$

If $x_{k+1}^0$ denotes this solution then

$$x^0 \equiv \begin{bmatrix} m_1(x_{k+1}^0) \\ \vdots \\ m_k(x_{k+1}^0) \\ x_{k+1}^0 \end{bmatrix}$$

is the (unique) solution of

$$F_{k+1}(x) + A_{k+1}x = B_{k+1} \ .$$

Thus, we have proved that there exists a unique solution of (1). $\square$

IV. BOUNDS ON THE SOLUTION

Having established the existence and uniqueness of a solution of (1) a natural question to arise is: What can one say about the location of this solution? It turns out that we can say quite a bit (again assuming that conditions (i) and (ii) of Section I are satisfied). One can, in fact, with little effort (compared with the effort required, in general, to actually compute the solution) determine a finite region $R$ in $E^n$, in which the solution must lie. This region is the cartesian product of finite intervals $I_i \subset E^1$, for $i = 1, \cdots, n$, each of which has the property that if

$$x^0 \equiv \begin{bmatrix} x_1^0 \\ \cdot \\ \cdot \\ \cdot \\ x_n^0 \end{bmatrix}$$

is the solution of (1) then $x_i^0 \; \varepsilon \; I_i$ and, as

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} \mid a_{ij} \mid \; \to 0,$$

the length of $I_i$, $l(I_i) \to 0$. Thus, when the off-diagonal elements of $A$ are small, the region $R$ will also be small.

In many applications it may be sufficient to know only that there exists a unique solution of (1) and to know the region $R$ in which it must lie. If, however, one actually does want to compute the solution by some iterative technique, the knowledge of $R$ should be useful in determining a starting point for the iteration. In fact, it will be shown that if the point $x^*$ is the solution of

$$F(x) + \text{diag} \, [a_{11}, \, \cdots, \, a_{nn}] \, x \; = \; B, \tag{5}$$

then $x^*$ is also in $R$ and thus might be a reasonable starting point for an iterative computation of $x^0$.

The computation of bounds for the solution of (1) proceeds in two steps. First, one solves each of the equations

$$f_i(x_i) \; = \; b_i \, , \quad \text{for} \quad i \; = \; 1, \, \cdots, \, n. \tag{6}$$

Letting $\alpha_i$ denote the solutions of (6), and defining

$$\alpha \; = \; \max \, \{\mid \alpha_i \mid \; : i \; = \; 1, \, \cdots, \, n\},$$

$$B' \; = \; \begin{bmatrix} \sum\limits_{\substack{j=1 \\ j \neq 1}}^{n} \mid a_{1j} \mid \\ \cdot \\ \cdot \\ \cdot \\ \sum\limits_{\substack{j=1 \\ j \neq n}}^{n} \mid a_{nj} \mid \end{bmatrix},$$

one then solves each of the equations

$$F(x) + \text{diag} \, [a_{11}, \, \cdots, \, a_{nn}] \, x \; = \; B \, - \, \alpha B', \tag{7a}$$

$$F(x) + \text{diag} \, [a_{11}, \, \cdots, \, a_{nn}] \, x \; = \; B \, + \, \alpha B'. \tag{7b}$$

Denoting the solutions of (7a) and (7b) by

$$
\eta \equiv \begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix} \quad \text{and} \quad \xi \equiv \begin{bmatrix} \xi_1 \\ \vdots \\ \xi_n \end{bmatrix},
$$

respectively, one has $R = I_1 \times \cdots \times I_n$, where

$$
I_i = [\eta_i, \xi_i], \quad \text{for} \quad i = 1, \cdots, n.
$$

It is clear from the fact that each component of the vector $\alpha B'$ is a nonnegative number and from the monotone nature of the left-hand sides of (7) that $x^*$, the solution of (5), is (as claimed) always in $R$. It is also clear that, for $i = 1, \cdots, n$, as

$$
\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \to 0,
$$

the $i$th components of both $B - \alpha B'$ and $B + \alpha B'$ approach $b_i$, and hence $\eta_i \to x_i^*$ and $\xi_i \to x_i^*$. Thus, $l(I_i) \to 0$. We now prove that the solution of (1) is in $R$.

*Theorem 2: If $R$ is constructed as described above, then the solution of (1) is contained in $R$ whenever conditions (i) and (ii) of Section I are satisfied.*

*Proof:* Let $x^0$ be the solution of (1) and let $k \varepsilon \{1, \cdots, n\}$ be chosen such that $|x_k^0| = \max\{|x_i^0| : i = 1, \cdots, n\}$. Then, by Lemma 1, if $x_k^0 > 0$, $\sum_{j=1}^{n} a_{kj} x_j^0 \geqq 0$ and hence,

$$
0 = f_k(x_k^0) + \sum_{j=1}^{n} a_{kj} x_j^0 - b_k \geqq f_k(x_k^0) - b_k
$$

or $f_k(x_k^0) \leqq b_k$.

Thus, because of the monotonicity of $f_k$,

$$
|x_k^0| = x_k^0 \leqq \alpha_k \leqq \alpha,
$$

and hence $|x_i^0| \leqq \alpha$ for $i = 1, \cdots, n$. Similarly, by Lemma 1, if $x_k^0 < 0$, $\sum_{j=1}^{n} a_{kj} x_j^0 \leqq 0$ and hence,

$$
f_k(x_k^0) \geqq b_k,
$$

and thus

$$| x_k^0 | = -x_k^0 \leq -\alpha_k \leq \alpha,$$

and hence $| x_i^0 | \leq \alpha$, for $i = 1, \cdots, n$. Thus, in any case, $| x_i^0 | \leq \alpha$, for $i = 1, \cdots, n$.

Now, for all $x$ with $| x_j | \leq \alpha$ for $j = 1, \cdots, n$, and for each $i \varepsilon \{1, \cdots, n\}$ we have,

$$\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} | = \sum_{\substack{j=1 \\ j \neq i}}^{n} (| a_{ij} | \alpha) \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} x_j |,$$

which implies

$$\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} | \geq \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j,$$

and

$$-\alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} | \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j.$$

Thus,

$$a_{ii} x_i - \alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} | \leq \sum_{j=1}^{n} a_{ij} x_j \leq a_{ii} x_i + \alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} |.$$

In particular, for $x = x^0$, we have

$$f_i(x_i^0) + a_{ii} x_i^0 - \alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} | \leq b_i \leq f_i(x_i^0) + a_{ii} x_i^0 + \alpha \sum_{\substack{j=1 \\ j \neq i}}^{n} | a_{ij} |.$$

Comparing this result with (7) we have, as a consequence of the monotonicity of the functions on the left-hand sides of (7),

$$\eta_i \leq x_i^0 \leq \xi_i, \quad \text{for} \quad i = 1, \cdots, n.$$

Hence, $x^0 \varepsilon R$.  □

Since in the above proof it was shown that $| x_i^0 | \leq \alpha$ for $i = 1, \cdots, n$ it might seem to some readers that the intervals $I_i$ might be reduced in length if we simply define them to be: $I_i = [-\alpha, \alpha] \cap [\eta_i, \xi_i]$. This, however, is unnecessary since it is easily shown that $-\alpha \leq \eta_i \leq \xi_i \leq \alpha$, for $i = 1, \cdots, n$.

## V. EXAMPLE

We now give an example of the use of the above method for the computation of solution bounds. Consider the equation

$$\begin{bmatrix} f_1(x_1) \\ f_2(x_2) \end{bmatrix} + \begin{bmatrix} 5 & 4 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 15 \\ 13 \end{bmatrix} ,$$

where $f_1$ and $f_2$ are defined by

$$f_1(x_1) = \begin{cases} 4^{x_1} - 2, & x_1 \geqq 0 \\ \frac{1}{16}x_1 - 1, & x_1 < 0, \end{cases}$$

and

$$f_2(x_2) = \begin{cases} x_2 + 9, & x_2 \geqq 3 \\ 4x_2 , & -3 < x_2 < 3 \\ x_2 - 9, & x_2 \leqq -3. \end{cases}$$

Figure 1 shows the graphs of $f_1$ and $f_2$. Since we know that the region $R$ will be small if the off-diagonal terms of $A$ are small enough, we have intentionally chosen an example in which these terms are rather large.

The computation of $\alpha$ by solving (6) may be done by inspection for this example. One finds that $4^{\alpha_1} = 17$ implies that $\alpha_1$ is slightly greater than 2, and since $\alpha_2 = 4$ we have $\alpha = 4$. Using this result in (7) one readily computes

$$\eta = \begin{bmatrix} 0 \\ 0.125 \end{bmatrix} , \qquad \xi \approx \begin{bmatrix} 2.23 \\ 3.2 \end{bmatrix} .$$
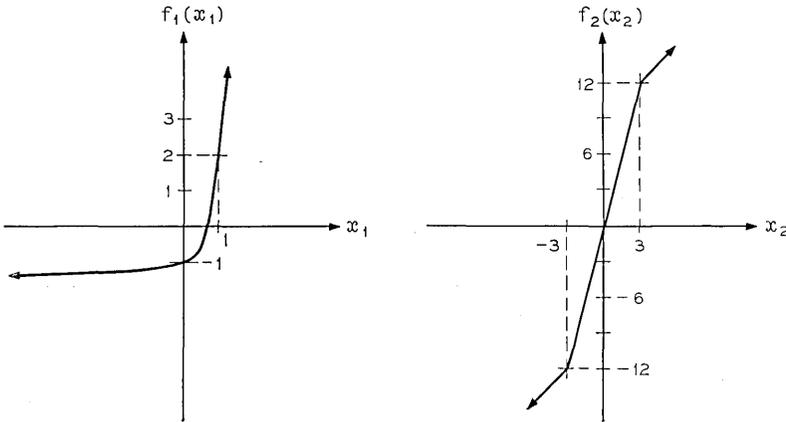


Fig. 1 — The nonlinear functions $f_1$ and $f_2$ for the example.

Actually, it is easily verified that the solution of this example is $x^0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

## VI. COMPUTATION OF THE SOLUTION

For $i = 1, \cdots, n$ we denote by $f'_i(x_i)$ the right-hand derivative of $f_i$ at the point $x_i \ \varepsilon \ E^1$. For each $x \ \varepsilon \ E^n$ we denote by $F'(x)$ the following matrix:

$$F'(x) = \text{diag} \ [f'_1(x_1), \cdots, f'_n(x_n)].$$

It is easy to prove that if $F$ satisfies condition $(i)$ of Section I then $F'(x)$ exists for all $x \ \varepsilon \ E^n$. Also, it is clear that each element of the main diagonal of $F'(x)$ is nonnegative for all $x \ \varepsilon \ E^n$. (Each element is in fact positive if, in addition, $F$ satisfies condition $(iii)$ of Section I.) Finally, we note that

$$F^{-1}(y) \equiv \begin{bmatrix} f_1^{-1}(y_1) \\ \vdots \\ f_n^{-1}(y_n) \end{bmatrix}$$

is defined for all $y \ \varepsilon \ E^n$, assuming again that $F$ satisfies condition $(i)$ of Section $I$.

The following two iteration schemes are proposed for the computation of the solution of (1):

*Scheme 1:* For given $x^1 \ \varepsilon \ E^n$ the sequence $x^1, x^2, x^3, \cdots$ of points in $E^n$ is constructed by use of the formula

$$x^{k+1} = [F'(x^k) + A]^{-1}(B - F(x^k) + F'(x^k)x^k). \tag{8}$$

*Scheme 2:* For given $x^1 \ \varepsilon \ E^n$ the sequence $x^1, x^2, x^3, \cdots$ of points in $E^n$ is constructed by use of the formula

$$x^{k+1} = [F'(F^{-1}(y^k)) + A]^{-1}(B - y^k + F'(F^{-1}(y^k))F^{-1}(y^k)), \tag{9}$$

where $y^k = -Ax^k + B$.

In order to explain the origin of (8) and (9) we make the following observations: If for $i = 1, \cdots, n$ $(x_i^k, y_i^k)$ is a given point in $E^2$, and if we draw the graph of each of the functions $f_i$, then each of the points in the sets $\{(x_i^k, f_i(x_i^k)): i = 1, \cdots, n\}$ and $\{(f_i^{-1}(y_i^k), y_i^k): i = 1, \cdots, n\}$ lies on the graph of the corresponding function $f_i$. Suppose we now replace (approximate) each $f_i$ by the straight line which is tangent to it at the corresponding point in one of the above sets.* Choosing the

---

\* Our definition of *tangent* coincides with the usual one, except that the right-hand derivative is used at those points where the derivative fails to exist.

first set of points we approximate $F$ by

$$\hat{F}(x) \equiv F'(x^k)x + F(x^k) - F'(x^k)x^k.$$

Choosing the second set gives

$$\tilde{F}(x) \equiv F'(F^{-1}(y^k))x + y^k - F'(F^{-1}(y^k))F^{-1}(y^k).$$

If we now define $y^k = - Ax^k + B$ and compute the solution of the equation

$$\hat{F}(x) + Ax = B$$

and call it $x^{k+1}$, we obtain (8). Calling $x^{k+1}$ the solution of

$$\tilde{F}(x) + Ax = B,$$

yields (9).

The above remarks have a very meaningful interpretation for problems arising from nonlinear electrical networks of the type described in Section I. Iteration Scheme 1 implements the following procedure: Given the vector $x^k$ of port voltages for the linear $n$-port, replace each two-terminal nonlinear device with a linear Thévenin's "equivalent" circuit whose $V$ vs $I$ curve is a straight line, tangent to the given curve at the point $(x_i^k, f_i(x_i^k))$. Compute the port voltages in the resulting linear network to obtain $x^{k+1}$.

Iteration Scheme 2 has a similar interpretation; this time, however, the vector of port currents, $y^k = -Ax^k + B$, is used to determine the linear equivalent circuit replacing the nonlinear devices at each step.

In view of the above remarks it is apparent that if one has some facility for solving linear network problems (a computer program, for example) then it might easily be adapted to solve many nonlinear problems as well.

We finally remark that the use of the first iteration scheme is, in essence, the same as using the Newton-Raphson technique to compute the root of (1).

We now prove a theorem which specifies conditions which are sufficient to ensure convergence of each of the above iteration schemes. We emphasize, however, that these iteration schemes will converge for many problems in which the conditions of the theorem are not satisfied—especially if a good enough starting point is provided.

In the following we denote the origin in $E^n$ by $\theta$ and, for the points $x, y \in E^n$, the notation $x \leqq y$ means $x_i \leqq y_i$ for $i = 1, \cdots, n$. The

relations $x < y$, $x \geqq y$, $x > y$ are defined similarly. We also make use of the concept of a *matrix of monotone kind*.[6] The matrix $A$ is said to be of monotone kind if $x \, \varepsilon \, E^n$, $Ax \geqq \theta \Rightarrow x \geqq \theta$. It is easy to show that $A$ is of monotone kind if and only if $A^{-1}$ contains only nonnegative elements. It is also easy to show that if $A$ is of monotone kind and $x, y \, \varepsilon \, E^n$ with $Ax \leqq y$, then $x \leqq A^{-1}y$. Ref. 6 shows that if the strict inequality $>$ holds in condition (*ii*) of Section I, then conditions (*ii*) and (*iv*) are sufficient to ensure that $A$ is of monotone kind.

*Theorem 3: For an arbitrary starting point $x^1$, both of the above iteration schemes will converge to the solution of (1) if conditions (i) through (iv) of Section I are satisfied.*

*Proof:* We give here only the proof for the second iteration scheme, assuming that all of the functions $f_i$ are convex. The other three cases are quite similar and it will be apparent to the reader how this proof may easily be modified to take care of them.*

We first remark that the iteration scheme is well defined. The fact that for every $y^k \, \varepsilon \, E^n$, $F'(F^{-1}(y^k))$ is a diagonal matrix containing all positive numbers on the main diagonal, and the fact that $A$ satisfies conditions (*ii*) and (*iv*) of Section I, assures us that the matrix $[F'(F^{-1}(y^k)) + A]$ is nonsingular (it is, in fact, of monotone kind—see Ref. 6, p. 376).

Let $x^1$ be an arbitrary point in $E^n$. Then, since for $i = 1, \cdots, n$ and $k = 2, 3, 4, \cdots$ each of the points $(x_i^k, y_i^k)$ lies on some straight line, tangent to the corresponding function $f_i$, and since each $f_i$ is strictly monotone increasing and convex, we have that $F^{-1}(y^k) \leqq x^k$ for $k = 2, 3, 4, \cdots$. We now show that $F^{-1}(y^k) \leqq x^k$ implies that $x^{k+1} \leqq x^k$. Obviously,

$$F'(F^{-1}(y^k))(x^k - F^{-1}(y^k)) \geqq \theta.$$

But, by definition, $Ax^k + y^k - B = \theta$; hence,

$$F'(F^{-1}(y^k))(x^k - F^{-1}(y^k)) + Ax^k + y^k - B \geqq \theta,$$

which implies

$$[F'(F^{-1}(y^k)) + A]x^k \geqq B - y^k + F'(F^{-1}(y^k))F^{-1}(y^k).$$

But then, since $[F'(F^{-1}(y^k)) + A]$ is a matrix of monotone kind,

$$x^k \geqq [F'(F^{-1}(y^k)) + A]^{-1}(B - y^k + F'(F^{-1}(y^k))F^{-1}(y^k)),$$

or, $x^k \geqq x^{k+1}$. Thus, the sequence $x^2, x^3, x^4, \cdots$ has the property

$$x^2 \geqq x^3 \geqq x^4 \geqq \cdots .$$

We now show that for $k = 2, 3, 4, \cdots , x^k \geqq x^0$, where $x^0$ is the solution of (1). For each $x^k$, $k = 2, 3, 4, \cdots$ , there is some point $p \ \varepsilon \ E^n$ $(p \equiv F^{-1}(y^{k-1}))$ such that

$$Ax^k - B = F'(p)p - F'(p)x^k - F(p). \tag{10}$$

Furthermore, from the convexity of each $f_i$, it is clear that for every pair of points $q^1, q^2 \ \varepsilon \ E^n$,

$$F(q^1) \geqq F(q^2) + F'(q^2)(q^1 - q^2).$$

In particular,

$$F(x^0) \geqq F(p) + F'(p)(x^0 - p).$$

Hence,

$$F'(p)p - F(p) + F(x^0) \geqq F'(p)x^0$$

which implies

$$F'(p)(p - x^k) - F(p) + F(x^0) \geqq F'(p)(x^0 - x^k).$$

Using (10) we have, therefore,

$$Ax^k - B + F(x^0) \geqq F'(p)(x^0 - x^k).$$

But, $F(x^0) = -Ax^0 + B$, hence

$$A(x^k - x^0) \geqq F'(p)(x^0 - x^k)$$

or,

$$[F'(p) + A](x^k - x^0) \geqq 0.$$

But then, since $[F'(p) + A]$ is of monotone kind, $x^k - x^0 \geqq 0$, or $x^k \geqq x^0$. Thus, we have shown that each sequence $x_i^2, x_i^3, x_i^4, \cdots$ is a bounded monotone sequence and hence the sequence $x^2, x^3, x^4, \cdots$ converges to some point $x^*$ in $E^n$. We now prove that $x^* = x^0$; that is, we show that $x^*$ satisfies (1).

Let $y^* = -Ax^* + B$. Then, as $k \to \infty$, $x^k \to x^*$ and $y^k \to y^*$. Thus, $F^{-1}(y^k) \to F^{-1}(y^*)$ and each element of the matrix $F'(F^{-1}(y^k))$ approaches the corresponding element of $F'(F^{-1}(y^*))$. Now, from (9), we have

$$Ax^{k+1} + F'(F^{-1}(y^k))x^{k+1} = Ax^k + F'(F^{-1}(y^k))F^{-1}(y^k)$$

which implies

$$F'(F^{-1}(y^k))(F^{-1}(y^k) - x^{k+1}) = A(x^{k+1} - x^k)$$

and hence

$$F'(F^{-1}(y^k))(F^{-1}(y^k) - x^* = A(x^{k+1} - x^k) - F'(F^{-1}(y^k))(x^* - x^{k+1}).$$

But as $k \to \infty$, $(x^{k+1} - x^k) \to \theta$ and hence $A(x^{k+1} - x^k) \to \theta$; also, $(x^* - x^{k+1}) \to \theta$ and hence $F'(F^{-1}(y^k))(x^* - x^{k+1}) \to F'(F^{-1}(y^*))\theta = \theta$. Thus, as $k \to \infty$,

$$F'(F^{-1}(y^k))(F^{-1}(y^k) - x^*) \to \theta$$

which implies

$$F^{-1}(y^k) - x^* \to \theta$$

or

$$F^{-1}(y^k) \to x^*$$

and therefore

$$y^k \to F(x^*).$$

Hence, $y^* = F(x^*)$, and thus,

$$F(x^*) + Ax^* = B.$$

Thus, the iteration converges to the solution of (1).    □

Although Theorem 3 states that both of our iteration schemes will converge for the same class of problems, only one of the schemes might converge for some problems for which all of the conditions (i) through (iv) of Section I are not satisfied. Also, for some problems a prior knowledge of the region in which the solution lies might dictate the choice of one iteration scheme over the other. For example, if it is known that some of the functions $f_i$ are quite steep in the neighborhood of the solution then perhaps $F^{-1}$ may be evaluated in this region more accurately than $F$. In this case Scheme 2 might be preferred to Scheme 1.

VII. SPEED OF CONVERGENCE

Section II mentions that in certain situations our iteration schemes may converge to the solution of (1) more rapidly than those based upon the contraction-mapping fixed point theorem. To illustrate this property we have chosen to compare the rate of convergence of Sandberg's iteration scheme to that of our schemes.[3]

If we define the operator $G$ mapping $E^n$ into $E^n$ by

$$G(x) \equiv F(x) + Ax,$$

then, as a special case of Sandberg's Theorem I, we have the result: If there are positive constants $k_1$ and $k_2$ such that

$$\langle G(x) - G(y), x - y \rangle \geqq k_1 \parallel x - y \parallel^2, \tag{11}$$

and

$$\parallel G(x) - G(y) \parallel^2 \leqq k_2 \parallel x - y \parallel^2, \tag{12}$$

for all $x$, $y$ $\varepsilon$ $E^n$, then there is a unique solution of (1) and the solution is given by $\lim_{k \to \infty} x^k$, where $x^1$ is an arbitrary point in $E^n$, and

$$x^{k+1} = \frac{k_1}{k_2} [B - G(x^k)] + x^k,$$

for $k = 1, 2, 3, \cdots$ . The proof of this theorem consists of showing that the mapping

$$H(x) \equiv \frac{k_1}{k_2} [B - G(x)] + x$$

is a contraction.

It is interesting to observe that if the inequalities (11) and (12) are satisfied then positive constants $k_3$ and $k_4$ exist, such that

$$\langle G(x) - G(y), x - y \rangle \leqq k_3 \parallel x - y \parallel^2, \tag{13}$$

and

$$\parallel G(x) - G(y) \parallel^2 \geqq k_4 \parallel x - y \parallel^2, \tag{14}$$

for all $x$, $y$ $\varepsilon$ $E^n$. In fact, a simple application of the Schwarz inequality to (11) and (12) yields (13) and (14) with $k_3 = (k_2)^{\frac{1}{2}}$ and $k_4 = k_1^2$. Now (13) and (14) are of the same form as (11) and (12), except that the inequalities are reversed. Thus, if one uses (13) and (14) in the proof of Sandberg's theorem, reversing all inequalities, one obtains:

$$\parallel H(x) - H(y) \parallel^2 \geqq K \parallel x - y \parallel^2,$$

where,

$$K = 1 - 2(k_1^2/k_2)^{\frac{1}{2}} + (k_1^2/k_2)^2.$$

It is readily seen that if $4k_1^2 < k_2$, then $K$ is positive. If we let $x^0$ denote the solution of (1), and hence $H(x^0) = x^0$, we have, for $k = 1, 2, \cdots$,

$$\parallel x^{k+1} - x^0 \parallel^2 = \parallel H(x^k) - H(x^0) \parallel^2 \geqq K \parallel x^k - x^0 \parallel^2.$$

Thus, $(K)^{\frac{1}{2}}$ represents, in this case, a lower bound on the rate of convergence of the iteration scheme. It is true that $(K)^{\frac{1}{2}}$ is always in the interval (0,1), for indeed Sandberg has proved that the sequence $x^k$ does converge to $x^0$. However, as $k_1$ becomes small, and as $k_2$ becomes large, $K$ approaches 1 and the sequence converges quite slowly. For (1) the largest value that may be used for $k_1$ and the smallest value that may be used for $k_2$ will many times be dictated by the positive constants which are bounds on the slopes of the functions $f_i$. If, for example, the slopes of the $f_i$ become so large for large $x_i$, and so small for large negative $x_i$ that one must choose $k_1 = 10^{-1}$ and $k_2 = 10^2$, then one easily computes $(K)^{\frac{1}{2}} \approx 0.99$. Thus, no matter how close any iterate is to the solution, the next iterate will be no more than about one percent closer.

It is of course true that Sandberg's iteration scheme is applicable to a much more general class of problems than we consider in this paper. If, however, for any problem to which it is applied, the constants $k_1$ and $k_2$ must be restricted such that $k_1/k_2$ is quite small, then the rate of convergence will always be adversely affected. In the Katzenelson-Seitelman application of Sandberg's iteration scheme, their "heuristic refinement" (see Ref. 2) attempts to overcome this difficulty.

Although the classes of equations to which our iteration schemes and the Katzenelson-Seitelman algorithm may be applied are not identical, in those cases where both techniques may be used the advantage that our schemes offer is now clear. From (8) and (9) one easily obtains

$$x^{k+1} - x^0 = [F'(x^k) + A]^{-1}(F(x^0) - F(x^k) - F'(x^k)(x^0 - x^k)),$$

and

$$x^{k+1} - x^0 =$$
$$[F'(F^{-1}(y^k)) + A]^{-1} (F(x^0) - y^k - F'(F^{-1}(y^k))(x^0 - F^{-1}(y^k))),$$

respectively. These equations show that $\| x^{k+1} - x^0 \|$ will be small (even if $\| x^k - x^0 \|$ is rather large) so long as for $i = 1, \cdots, n$,

$$\frac{f_i(x_i^0) - f_i(x_i^k)}{x_i^0 - x_i^k} \approx f_i'(x_i^k),$$

for Scheme 1, or

$$\frac{f_i(x_i^0) - y_i^k}{x_i^0 - f_i^{-1}(y_i^k)} \approx f_i'(f_i^{-1}(y_i^k)),$$

for Scheme 2. That is, as soon as the $k$th iterate comes close enough to the solution that each of the functions $f_i$ is approximately linear, the rate of convergence of our iterations becomes quite rapid. In fact, the rate of convergence increases without bound as the iterates approach the solution. It is also clear that if each of the functions $f_i$ is piecewise linear then our iterations will converge in a finite number of steps.

From the standpoint of computational efficiency it is, of course, the amount of time required to compute an approximate solution that is the major concern. For those problems to which both our iteration schemes and the Katzenelson-Seitelman algorithm may be applied, it can happen that our methods might still be slower than theirs even in the case when the convergence rate of our methods is faster. This can happen because, for some problems, the equation with which we are concerned may be of a higher order than theirs, and also because we must compute the inverse of a matrix at each iteration step. On the other hand, it is clear that for many problems, even from the standpoint of total computation time, our techniques will be more efficient.

## VIII. ACKNOWLEDGMENT

REFERENCES

1. Duffin, R. J., "Nonlinear Networks II," Bull. Amer. Math. Soc., *53* (October 1947), pp. 963–971.
2. Katzenelson, J. and Seitelman, L. H., "An Iterative Method for Solution of Networks of Nonlinear Monotone Resistors," IEEE Trans. Circuit Theory, *CT-13*, No. 3 (September 1966), pp. 317–323.
3. Sandberg, I. W., "On the Properties of Some Systems That Distort Signals-I," B.S.T.J., *42*, No. 5 (September 1963), pp. 2033–2046.
4. Minty, G. J., "Two Theorems on Nonlinear Functional Equations in Hilbert Space," Bull. Amer. Math. Soc., *69*, No. 5 (September 1963), pp. 691–692.
5. Weinberg, L., *Network Analysis and Synthesis,* New York: McGraw-Hill, 1962.
6. Collatz, L., *Functional Analysis and Numerical Mathematics* (tr. from German), New York: Academic Press, 1966.
7. Vandergraft, J. S., "Newton's Method for Convex Operators in Partially Ordered Spaces," SIAM J. Numerical Anal., *4*, No. 3 (September 1967), pp. 406–432.

# Convergence Criteria for Transversal Equalizers

By D. W. LYTLE*

(Manuscript received March 22, 1968)

*Two basic problems in the equalization of data channels using pulse-amplitude modulation are considered. The first of these is to determine just what pulses can be equalized and what the general equalization solution is when using a transversal filter. The second problem is to determine if a simple iterative search routine will converge to a solution if its exists.*

*The unequalized channel impulse response is represented by a polynomial whose coefficients are the sample values of the impulse response. If no roots of this polynomial lie on the unit circle, the channel can be equalized. The transversal filter which equalizes the pulse has tapweight values given by weighted sums of powers of the polynomial roots.*

*Various necessary and sufficient conditions for iterative convergence are developed. Iterative convergence can be guaranteed if the proper linear weighting of the output sample errors is used in adjusting the tap-weights.*

## I. INTRODUCTION

This paper is concerned with certain aspects of the automatic equalization of low-noise, linear channels which are to be used for multi-level pulse-amplitude modulated (PAM) signals. The purpose of the equalizer is to compensate for the channel transfer characteristics in such a way that the over-all impulse response of the channel is a Nyquist-I type of pulse,[1,2,3] that is, as is illustrated in Fig. 1, a pulse with a central peak and uniformly spaced zeros with period $T$. If such an impulse response is achieved, a sequence of amplitude modulated impulses with period $T$ can be transmitted and the sequence of amplitudes can be recovered at the receiver by simply sampling in synchronism with period $T$.

Certain obvious questions such as how to achieve synchronism and

---

* University of Washington, Seattle.
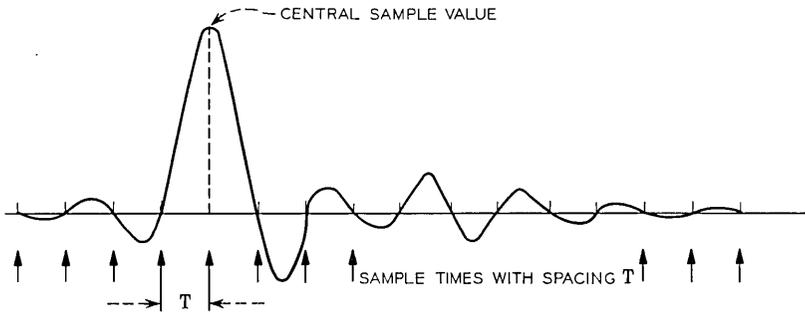
Fig. 1 — A Nyquist-I pulse with central sample and zeros at periodic sample times.

the effects of sampling jitter, nonlinear distortion, and additive noise, although of great importance, are neglected in this presentation in order to concentrate on the methods of adjusting the equalizer. Thus, we assume a perfectly synchronized, noiseless, linear channel with an ideal sampler. Some further constraints which should simplify and clarify the presentation are as follows. We specify that the equalizer is to be a transversal (tapped delay-line) filter with tap weights $\{\alpha_k\}$ which can be adjusted. If the input to this filter, illustrated in Fig. 2, is $\beta(t)$, then the output, $\gamma(t)$, is

$$\gamma(t) = \sum_{i=-n}^{N} \alpha_i \beta(t + jT).$$ (1)

Notice the tap-weight numbering convention and the treatment of the delay-line as being composed of negative as well as positive delay. These conventions will simplify the notation in future derivations.

Although the equalizer may be placed at many points within the communication system, for convenience we will consider it to be the final component other than the final sampler. Thus, the objective is
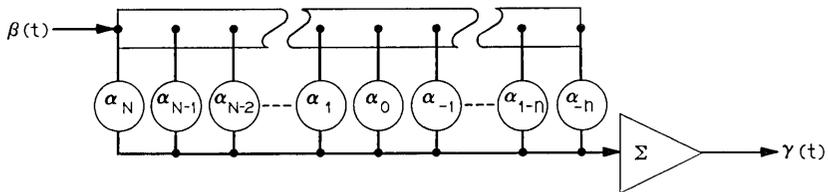


Fig. 2 — A transversal filter with input $\beta(t)$, tap weights $\alpha_j$, and output $\gamma(t)$.

to adjust the tap weights so that the output, $\gamma(t)$, is a Nyquist-I pulse when the input, $\beta(t)$, is the impulse response of the system before equalization.

As a final simplification we assume that the tap-weight adjustment is to be carried out when the impulse response is available at the input. This "training period" assumption avoids the added complexities of extracting channel characteristics when data are being transmitted.[4] However, the properties to be developed can be readily extended to continuously adaptive equalization.

## II. TIMING AND THE ROOTS OF THE IMPULSE RESPONSE

Since the output of the equalizer is to be sampled, we need consider only its effect on the output sample values $\{\gamma_k\}$. Ideally, with an impulse applied to the channel input, one of the output samples, say $\gamma_0$, should have unit amplitude while all the others are zero. The tap weights are to be adjusted in order to approach this goal. An additional parameter which will affect the equalization is the timing.

Let the sample values of the impulse response at the equalizer input be the set $\{\beta_j\}$ where

$$\beta_j = \beta(jT) \tag{2}$$

and

$$\beta(t) = h(t + \tau). \tag{3}$$

Equation (3) is to indicate that the sampling times are arbitrary. This is, if $h(t)$ is the channel response to an impulse applied at $t = 0$, then the sample set $\{\beta_j\}$ is a function of the factor $\tau$. Notice that our assumption of perfect synchronization means that the periodicity factor, $T$, in equation (2) is the proper value. But it does not imply that $\tau$ is prescribed. We shall see that the operation of the equalizer depends very strongly upon the value of $\tau$.

This impulse response is to be equalized by the transversal filter with tap weights $\{\alpha_k\}$. The transversal filter output sample set is $\{\gamma_k\}$ where

$$\gamma_k = \sum_i \alpha_i \beta_{k+i} . \tag{4}$$

The sample set $\{\beta_j\}$ will be considered finite in extent, that is,

$$\beta_j = 0 \quad \text{for} \quad j < -m \quad \text{and} \quad j > M. \tag{5}$$

This is a reasonable approximation for any actual channel.

This sample set may be treated as the coefficients of a polynomial $B(z)$.

$$B(z) = \beta_M z^{M+m} + \beta_{M-1} z^{M+m-1} + \cdots$$
$$+ \beta_0 z^m + \cdots + \beta_{-m+1} z + \beta_{-m}. \qquad (6)$$

In factored form $B(z)$ may be written as

$$B(z) = \beta_M (z - \theta_1)(z - \theta_2) \cdots (z - \theta_I)(z - \phi_1)(z - \phi_2) \cdots (z - \phi_\Omega) \qquad (7)$$

where the roots inside the unit-circle are denoted by the $\theta$ values and the roots outside by the $\phi$ values.

$$| \theta | < 1 < | \phi | \qquad (8)$$

$$I + \Omega = M + m. \qquad (9)$$

As we have noted, the sample set $\{\beta_j\}$ is a function of the factor $\tau$. Thus, for any particular channel, the roots of equation (7) will wander as $\tau$ is varied. Each root will wander on some cyclic path which has a period $T$. That is

$$\theta_i(\tau + nT) = \theta_i(\tau) \qquad (10)$$

This is illustrated in Fig. 3 where a pulse shape is shown and in Fig. 4 where the root loci are shown for variations in $\tau$. Notice that as $\tau$ increases from 0 to $T$, at least one of the roots, regardless of the pulse shape, will cross the unit circle.

Perhaps the periodicity of the roots can be better understood if the sampling is thought of as multiplication by a comb of impulses with spacing $T$. Each impulse has associated with it a power of $z$. For example, in equation (6), we see that the impulse yielding the earliest nonzero sample ($\beta_{-m}$) is associated with the zero power of $z$, the next impulse yielding $\beta_{-m+1}$ is associated with the first power, and so on. As the comb is moved relative to the pulse, $\beta(t)$, the impulses produce different samples and when moved a whole period $T$, the comb will reproduce the original samples again. However, each sample would be paired with a one-higher or one-lower power of $z$ than previously. Thus, for example, if the comb were shifted by $T$ so that the powers of $z$ were one higher, the factorization of equation (7) would be obtained with the same roots except for an additional root at $z = 0$ since the original polynomial is multiplied by the first power of $z$. As the comb is shifted along, the additional root, which must eventually go to $z = 0$, comes in from $z = \infty$ and at some particular shift crosses the unit circle.
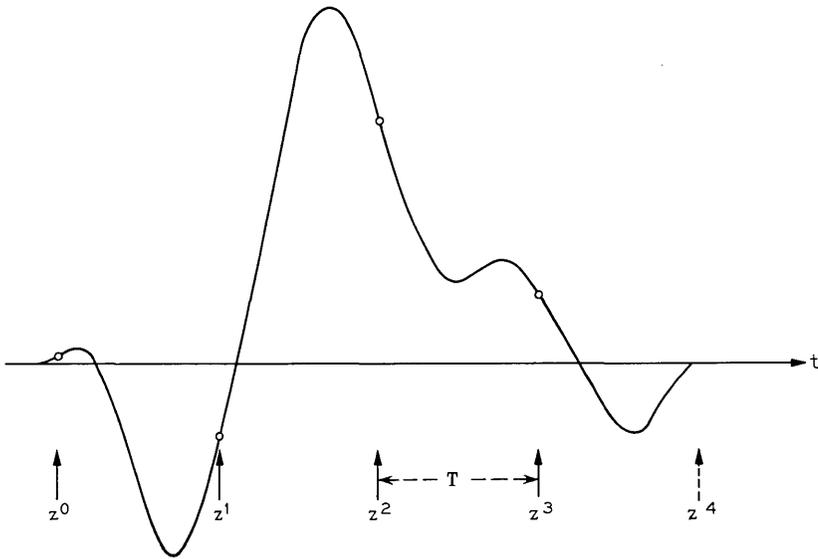
Fig. 3 — A pulse and four nonzero sample positions which yield a polynomial with roots marked 1 in Fig. 4. Nine additional sets of roots are obtained by moving the positions above to the right in increments of $T/10$.
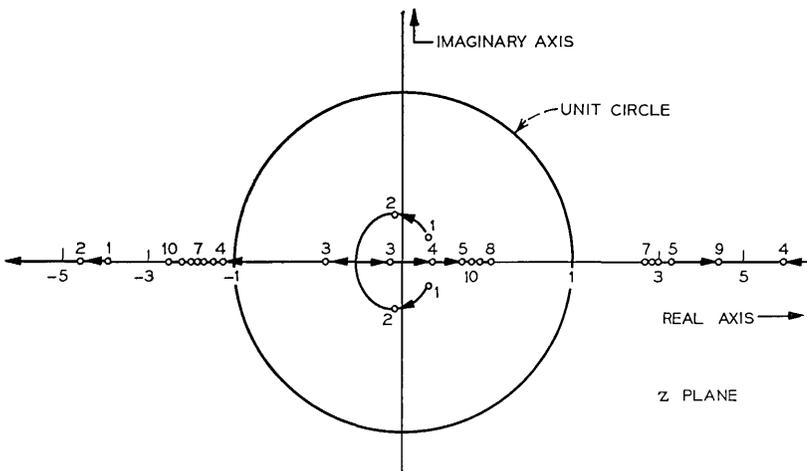


Fig. 4 — Root loci for pulse of Fig. 3. Scale inside circle is magnified four times.

## III. IDEAL TAP-WEIGHT ADJUSTMENT

Let us now consider the problem of forcing the $\gamma_k$ to zero. The desired conditions are

$$\gamma_0 = \sum_i \alpha_i \beta_i = 1 \tag{11}$$

and

$$\gamma_k = \sum_i \alpha_i \beta_{i+k} = 0 \qquad k \neq 0. \tag{12}$$

Equations (11) and (12) in the form of a matrix array is as follows. We shall assume temporarily that the number of taps is infinite.

$$
\begin{bmatrix}
& & & & \vdots & & & & & & & \\
\cdots\ 0\ 0\ \beta_M & \cdots & \cdots & \beta_2\ \beta_1\ \beta_0\ \beta_{-1}\ \beta_{-2} & \cdots & \cdots & \beta_{-m}\ 0 & \\
0\ 0\ \beta_M & \cdots & \beta_3\ \beta_2\ \beta_1\ \beta_0\ \beta_{-1}\ \beta_{-2} & \cdots & \cdots & \beta_{-m}\ 0\ \cdots & \\
0\ \beta_M & \cdots & \beta_3\ \beta_2\ \beta_1\ \beta_0\ \beta_{-1}\ \beta_{-2} & \cdots & \cdots & \beta_{-m}\ 0\ \cdots & \\
0\ \beta_M & \cdots & \beta_3\ \beta_2\ \beta_1\ \beta_0\ \beta_{-1}\ \beta_{-2} & \cdots & \cdots & \beta_{-m}\ \cdots & \\
& & & \vdots & & & &
\end{bmatrix}
$$

$$
\cdot
\begin{bmatrix}
\vdots \\
\alpha_2 \\
\alpha_1 \\
\alpha_0 \\
\alpha_{-1} \\
\alpha_{-2} \\
\alpha_{-3} \\
\vdots
\end{bmatrix}
=
\begin{bmatrix}
\vdots \\
0 \\
0 \\
1 \\
0 \\
0 \\
0 \\
\vdots
\end{bmatrix}. \tag{13}
$$

Now let us consider a typical equation of the form,

$$\sum_{i=-m-k}^{M-k} \alpha_i \beta_{k+i} = 0 \tag{14}$$

where we will take either the case where $k < -m$ or $k > M$. If these conditions on $k$ are satisfied, then equation (14) involves $\alpha$ values with only positive subscripts or only negative subscripts. Equation (14) is a linear homogeneous difference equation in the variable $\alpha_j$. Such

equations have solutions of the form

$$\alpha_j = z^j. \tag{15}$$

Substitution of equation (15) into equation (14) yields

$$\beta_M z^{M-k} + \beta_{M-1} z^{M-k-1} + \cdots + \beta_1 z^{1-k}$$

$$+ \beta_0 z^{-k} + \beta_{-1} z^{-1-k} + \cdots + \beta_{-m} z^{-m-k}$$

$$= z^{-m-k}[\beta_M z^{M+m} + \beta_{M-1} z^{M+m-1} + \cdots + \beta_0 z^m + \beta_{-m}]$$

$$= z^{-m-k} B(z) = 0. \tag{16}$$

Equation (16), and hence equation (14), will be satisfied for $z$ equal to one of the roots of the impulse response polynomial $B(z)$. And since we have a linear homogeneous equation, any linear combination of solutions is a solution. Thus, the general solution is

$$\alpha_j = C_1 \theta_1^i + C_2 \theta_2^i + \cdots + C_I \theta_I^i + D_1 \phi_1^i + \cdots + D_\Omega \phi_\Omega^i \tag{17}$$

The constants, $C_1, C_2, \ldots, C_I, D_1, D_2, \ldots, D_\Omega$, are arbitrary and will be adjusted to meet the boundary conditions. One boundary condition which should be imposed is the following. Eventually, we must approximate this infinite delay line with one of finite extent. This truncation should throw away only taps of small magnitude, and thus the tap weights should decrease in magnitude away from the center tap. Consequently, we demand that $\alpha_j \to 0$ as $|j| \to \infty$. Thus,

$$C_1 = C_2 = \cdots = C_I = 0 \quad \text{for} \quad j \text{ negative.}$$

$$D_1 = D_2 = \cdots = D_\Omega = 0 \quad \text{for} \quad j \text{ positive.} \tag{18}$$

In effect, we have two solutions; one for taps with negative subscripts and one for taps with positive subscripts.

$$\alpha_j = C_1 \theta_1^i + C_2 \theta_2^i + \cdots + C_I \theta_I^i \quad \text{for} \quad j > 0$$

$$\alpha_j = D_1 \phi_1^i + D_2 \phi_2^i + \cdots + D_\Omega \phi_\Omega^i \quad \text{for} \quad j < 0. \tag{19}$$

The region of overlap in equation (13), that is, the region where the equations involve tap weights with both positive and negative subscripts, will determine the arbitrary constants. To illustrate this, let us consider the impulse response of Fig. 2. For one set (No. 8) of samples, the samples are $-2, 3, 11, -6$ and the roots are $0.5, 3, -2$. Thus

$$\alpha_j = C_1 (0.5)^j \qquad \text{for} \quad j > 0$$

$$\alpha_j = D_1 (3)^j + D_2 (-2)^j \quad \text{for} \quad j < 0. \tag{20}$$

Now, $C_1$, $D_1$, and $D_2$ may be found by using equation (20) in equation (13). However, first we must assign an origin in our pulse. That is, we must decide whether $-2$, or 3, or 11, or $-6$ is to be called $\beta_0$. Solutions exist for each of these possibilities, but only one of these has certain desirable properties which we will discuss later. In this case, $\beta_0$ should be the sample of magnitude 11. With this assignment, equation (13) may be written

$$\vdots$$

$$-2[C_1(.5)^4] + 3[C_1(.5)^3] + 11[C_1(.5)^2] - 6[C_1(.5)] = 0 \tag{21a}$$

$$-2[C_1(.5)^3] + 3[C_1(.5)^2] + 11[C_1(.5)] - 6\alpha_0 = 0 \tag{21b}$$

$$-2[C_1(.5)^2] + 3[C_1(.5)] + 11\alpha_0 - 6[D_1(3)^{-1} + D_2(-2)^{-1}] = 1 \tag{21c}$$

$$-2[C_1(.5)] + 3\alpha_0 + 11[D_1(3)^{-1} + D_2(-2)^{-1}]$$
$$- 6[D_1(3)^{-2} + D_2(-2)^{-2}] = 0 \tag{21d}$$

$$-2\alpha_0 + 3[D_1(3)^{-1} + D_2(-2)^{-1}] + 11[D_1(3)^{-2} + D_2(-2)^{-2}]$$
$$- 6[D_1(3)^{-3} + D_2(-2)^{-3}] = 0 \tag{21e}$$

$$-2[D_1(3)^{-1} + D_2(-2)^{-1}] + 3[D_1(3)^{-2} + D_2(-2)^{-2}]$$
$$+ 11[D_1(3)^{-3} + D_2(-2)^{-3}] - 6[D_1(3)^{-4} + D_2(-2)^{-4}] = 0 \tag{21f}$$

$$\vdots$$

Equations (21a) and (21f) and all others above and below these two are automatically satisfied for any choice of $C_1$, $D_1$, $D_2$. Consider equation (21b). In order for it to be satisfied, $\alpha_0$ must be equal to $C_1$ $(0.5)^0$. Similarly, for equation (21e) to be satisfied, $\alpha_0$ must equal $D_1(3)^0 + D_2(-2)^0$,

$$C_1 = D_1 + D_2 = \alpha_0 \tag{22}$$

and equation (21) yields the following values for the $C$ and $D$ constants.

$$\left.\begin{aligned} D_1 &= 2/95 \\ D_2 &= 5/95 \\ C_1 &= 7/95 \end{aligned}\right\} \beta_0 = 11. \tag{23}$$

If we had chosen to let $\beta_0$ be the sample of magnitude 3, we would obtain

$$\left.\begin{aligned} D_1 &= -10/95 \\ D_2 &= \phantom{-}15/95 \\ C_1 &= \phantom{-}5/95 \end{aligned}\right\} \beta_0 = 3. \tag{24}$$

If the sample of magnitude 2 is made $\beta_0$, then

$$\left.\begin{aligned} D_1 &= 9/25 \\ D_2 &= 4/25 \\ C_1 &= 1/50 \end{aligned}\right\} \beta_0 = -2. \tag{25}$$

In these last two examples, the solutions may be considered inferior because the tap weights away from the center tap will be larger than in the first example. However, stronger objections to the last two choices for $\beta_0$ will be raised shortly.

A point of considerable interest is apparent in the development above. If any of the roots fall on the unit circle, then no solutions exist in which the tap weights decay in magnitude in both directions away from the center tap.

## IV. TRUNCATION EFFECTS

In any practical equalizer, the number of taps available is not infinite. Thus, we must investigate the effects of limiting the number of taps to some reasonable finite value. For example, let us suppose that we have $N + n + 1$ taps.

$$\alpha_j = 0 \quad \text{for} \quad \begin{cases} j < -n \\ j > N \end{cases}. \tag{26}$$

Let us consider two schemes for setting the truncted tap weights. A more or less obvious way is simply to take the infinite solution [for example, equation (23)] for all available taps. This can be represented in matrix form:

$$[B_\infty][\alpha_T] = [\gamma]. \tag{27}$$

The matrix $[B_\infty]$ is the infinite matrix of $\beta$ values shown in equation (13). The truncated tap set is $[\alpha_T]$,

$$[\alpha_T] = \begin{bmatrix} \vdots \\ 0 \\ 0 \\ \alpha_N \\ \alpha_{N-1} \\ \alpha_1 \\ \alpha_0 \\ \alpha_{-1} \\ \vdots \\ \alpha_{-n} \\ 0 \\ 0 \\ \vdots \end{bmatrix} \tag{28}$$

and $[\gamma]$ is the matrix of output sample values. Notice that $[\alpha_T]$ may be written as

$$[\alpha_T] = [\alpha_\infty] - [\alpha_t] \tag{29}$$

where $[\alpha_\infty]$ is the infinite set of tap weights which give us the desired output and

$$[\alpha_t] = \begin{bmatrix} \vdots \\ \alpha_{N+2} \\ \alpha_{N+1} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ \alpha_{-n-1} \\ \alpha_{-n-2} \\ \vdots \end{bmatrix} . \tag{30}$$

Thus, equation (27) becomes

$$[B_\infty][\alpha_T] = [B_\infty][\alpha_\infty] - [B_\infty][\alpha_i] = [\mu] - [\epsilon]. \tag{31}$$

The desired output $[\mu]$, all zero samples except the central one with unit magnitude, is given by $[B_\infty][\alpha_\infty]$ while the error set $[\epsilon]$ is given by $[B_\infty][\alpha_t]$.

It is possible to calculate this error set exactly, but an approximate bound should suffice. Assume that $N$ and $n$ are large so that the $\alpha$ values in $[\alpha_t]$ depend only on the roots which are the nearest to the unit circle. Let the magnitude of nearest inside and outside roots be $\theta_u$ and $\phi_u$. Then approximately

$$\alpha_i \cong \begin{matrix} \theta_u^i & \text{for} & j > N \\ \phi_u^i & \text{for} & j < -n \end{matrix}. \tag{32}$$

Consequently the largest component of any error term in $[\epsilon]$ will be about

$$\text{Max comp} \cong \beta_{\max} \cdot C_u \theta_u^{N+1} \cong \beta_{\max} D_u \phi_u^{-n-1} \tag{33}$$

and each error term is the sum of less than $N + n + 1$ components. Thus, an upper bound on the error terms is

$$|\epsilon_{\max}| < |(N + n)C_u \theta_u^{N+1}| \cong |(N + n) D_u \phi_u^{-n-1}| \tag{34}$$

which vanishes as $N$ and $n$ are made very large.

## V. A SECOND METHOD OF TRUNCATION

This second method offers no improvement in ultimate equalization over the method just discussed. However, it does lend itself to iterative adjustment techniques whereas the first method tacitly assumes a computation which provides the proper infinite solution to begin with. In this second method we require that all the output samples (excluding $\gamma_0$) corresponding to the $N + n + 1$ taps, that is, $\gamma_N, \gamma_{N-1}, \ldots, \gamma_2, \gamma_1, \gamma_{-1}, \gamma_{-2}, \ldots, \gamma_{-n+1}, \gamma_{-n}$, be zero. This criterion may be called the Lucky criterion since it is the one R. W. Lucky has used in his work.[5]

What does this criterion mean in terms of the solutions (powers of impulse response roots) we discussed for the infinite tap case? We are essentially constraining our system further by another set of boundary conditions. We will call these boundaries the positive boundary at $\alpha_N$ and the negative boundary at $\alpha_{-n}$, in addition to the central boundary around $\alpha_0$ where we have already discussed satisfying boundary con-

ditions (those specifying $C_1, C_2, \ldots, D_1, D_2, \ldots$). As an example, let us consider the equations at the positive boundary.

$$\beta_0 \alpha_N + \beta_{-1} \alpha_{N-1} + \cdots \cdots + \beta_{-m} \alpha_{N-m} = \gamma_N \qquad\qquad = 0 \quad (35a)$$

$$\beta_1 \alpha_N + \beta_0 \alpha_{N-1} + \cdots + \beta_{-m} \alpha_{N-(m+1)} = \gamma_{N-1} \qquad = 0 \quad (35b)$$

$$\vdots \qquad\qquad\qquad\qquad\qquad\qquad \vdots$$

$$\beta_{M-1} \alpha_N + \beta_{M-2} \alpha_{N-1} + \cdots + \beta_{-m} \alpha_{N-(m+M-1)} = \gamma_{N-(M-1)} = 0 \quad (35c)$$

$$\beta_M \alpha_N + \beta_{M-1} \alpha_{N-1} + \cdots \cdots + \beta_{-m} \alpha_{N-(m+M)} = \gamma_{N-M} \quad = 0 \quad (35d)$$

Equation (35d) can be satisfied with the solutions determined by the central boundary, that is

$$\alpha_i = C_1 \theta_1^i + C_2 \theta_2^i + \cdots + C_I \theta_I^i . \qquad (36)$$

However, this solution will not satisfy the $M$ equations above equation (35d) since the complete set of $\beta$ values is missing in these equations. The exponentially growing solutions which were discarded earlier must be used now. Thus,

$$\alpha_i = C_1 \theta_1^i + \cdots + C_I \theta_I^i + c_1 \theta_1^i + c_2 \theta_2^i + \cdots + c_\Omega \phi_\Omega^i \quad \text{for} \quad j > 0 \qquad (37)$$

$$\alpha_i = D_1 \phi_1^i + \cdots + D_\Omega \phi_\Omega^i + d_1 \theta_1^i + d_2 \theta_2^i + \cdots + d_I \theta_I^i \quad \text{for} \quad j < 0 \qquad (38)$$

must be used in order to satisfy all the boundary equations. If $\Omega = M$, then the lower case $c$'s in equation (37) provide just enough constants to satisfy the $M$ equations of equations (35a through c). Furthermore, $I$ will equal $m$ and the $I$ lower case $d$'s will provide just enough constants to satisfy the $m$ boundary equations at the negative boundary.

To illustrate the preceding discussion, let us return to the specific example discussed previously. We use the results of equation (23) with

$$\beta_2 = -2, \quad \beta_1 = 3, \quad \beta_0 = 11, \quad \beta_{-1} = -6. \qquad (39)$$

The positive boundary equations are

$$11[C_1(.5)^N + c_1(3)^N + c_2(-2)^N]$$
$$- 6[C_1(.5)^{N-1} + c_1(3)^{N-1} + c_2(-2)^{N-1}] = 0 \qquad (40)$$
$$3[C_1(.5)^N + c_1(3)^N + c_2(-2)^N] + 11[C_1(.5)^{N-1} + c_2(-2)^{N-1}]$$
$$- 6[C_1(.5)^{N-2} + c_1(3)^{N-2} + c_2(-2)^{N-2}] = 0$$

$$\begin{bmatrix} 3^{N-1}(33 - 6) & -2^{N-1}(-22 - 6) \\ 3^{N-2}(27 + 33 - 6) & -2^{N-2}(12 - 22 - 6) \end{bmatrix}\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$= \begin{bmatrix} .5^N(11 - 12)C_1 \\ .5^N(3 + 22 - 24)C_1 \end{bmatrix}. \qquad (41)$$

If we assume that $N$ is large enough so that the value of the $C$ and $D$ variables are unaffected by inclusion of the other roots, then these $C$ and $D$ values may be used to determine the $c$ and $d$ values. With $C_1 = 7/95$, equation (41) becomes

$$\begin{bmatrix} 9 & 14 \\ 6 & -4 \end{bmatrix}\begin{bmatrix} c_1' \\ c_2' \end{bmatrix} = \begin{bmatrix} -\mathcal{C} \\ \mathcal{C} \end{bmatrix} \qquad (42)$$

where

$$c_1' = 3^N c_1 \qquad c_2' = (-2)^N c_2 \qquad \mathcal{C} = \left(\frac{7}{95}\right)(.5)^N. \qquad (43)$$

Thus,

$$c_1' = \mathcal{C}/12 \qquad\qquad c_2' = -\mathcal{C}/8$$

$$c_1 = \frac{\left(\dfrac{7}{95}\right)(.5)^N}{(3)^N} \qquad c_2 = \frac{\left(\dfrac{7}{95}\right)(.5)^N}{(-2)^N}. \qquad (44)$$

Similarly, $d_1$ can be found by the single boundary equation at the negative boundary.

$$d_1 = -\left[\frac{2\left(\dfrac{2}{95}\right)(3)^{-n} + 9\left(\dfrac{5}{95}\right)(-2)^{-n}}{12}\right](.5)^n. \qquad (45)$$

The results worked out above can be roughly represented graphically as in Fig. 5 where the magnitudes of the roots to the tap-number power are illustrated. This figure shows what will be called a "good" solution. That is, the decaying solutions predominate with the growing solutions contributing only a small amount at the positive and negative boundaries. The residual errors, that is, the $\gamma_k$ values for $k > N$ and $k < -n$, will be of the same order of magnitude as those of the first truncation method.

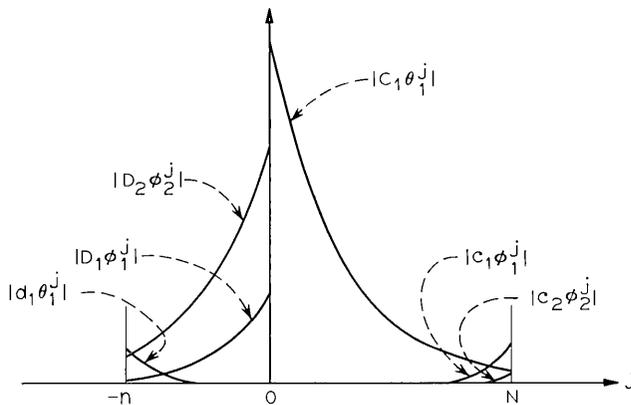In order to have a good solution as demonstrated above, the proper

Fig. 5 — Solution behavior for truncated equalizer.

number of arbitrary constants must be available to satisfy the positive and negative boundary equations. The number of equations at the positive boundary is $M$, and the number of available constants (outside roots) is $\Omega$; similarly, there are $I$ constants (inside roots) available for the $m$ equations at the negative boundary. Thus, the necessary conditions for a good solution are

$$\left.\begin{array}{l} \text{no. outside roots} = \Omega = M = \text{no. of samples following } \beta_0 \\ \text{no. inside roots} = I = m = \text{no. of samples preceding } \beta_0 \end{array}\right\}. \quad (46)$$

We may use the same simple pulse to demonstrate a "bad" solution. If we go back to equation (24), we see that an infinite (non-truncated) solution exists for the situation in which we decided to let $\beta_0 = 3$. The number of inside and outside roots remain the same in this case, but $M$ and $m$ are both changed. Now $M = 1$ and $m = 2$, and equations (46) are no longer satisfied. All the boundary conditions can still be satisfied, but not in such a simple manner. That is to say, in the preceding example, satisfying equations (46), a separation of solutions is possible. The central boundary specifies the values of the upper-case constants, then the lower-case constants are set to compensate for the truncation effect at the positive and negative boundaries. Since the necessary compensation is small, and since the effect of the lower-case constants dies out towards the central boundary, only minor or negligible corrections to the upper-case constants are necessary to keep the central boundary conditions satisfied.

Now let us consider what takes place when this step-by-step solu-

tion is attempted when equations (46) are not satisfied. Suppose $C_1$, $D_1$, and $D_2$ are set to satisfy the central boundary according to equation (24). At the positive boundary, only one of the lower-case $c$'s is needed so no difficulty arises and the positive tap solutions will, to this point, be little different from that illustrated in Fig. 5. However, at the negative boundary, the single available arbitrary constant $d_1$ is not enough to satisfy the two boundary equations which occur in this case. All the upper and lower case $d$-constants can be adjusted to satisfy the negative boundary conditions, but this will destroy the equilibrium of the central boundary solution since any change in the upper-case $D$'s does effect the central equations. If one of the upper-case $D$'s is constrained by the negative boundary, then the remaining unused lower-case $c$ can be brought in to provide enough arbitrary constants to satisfy the central boundary conditions. The net effect of all this will generally be that a growing solution must be made to have a nonnegligible contribution at the central boundary. Consequently, it will be large at the positive boundary. This is illustrated in Fig. 6.

When the situation discussed above and represented in Fig. 6 occurs, the residual values of $\gamma_k$ outside the equalization region will be large and will generally grow larger as the number of taps is increased.

As an actual example of a bad solution of the type discussed above, consider the pulse illustrated on page 563 of Ref. 5. The polynomial
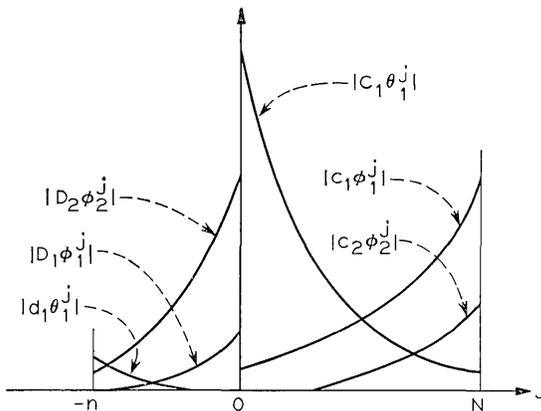


Fig. 6 — Typical solution behavior when the conditions of equations (46) are not satisfied.

which represents this pulse with $\beta_o$ set as the peak value is

$$B(z) = 15z^8 - 20z^7 + 40z^6 - 60z^5 + 0z^4 + 115z^3$$

$$+ 60z^2 + 20z + 10. \qquad (47)$$

Notice that with this choice for $\beta_o$,

$$M = 5 \quad m = 3. \qquad (48)$$

The root locations for this polynomial are illustrated in Fig. 7. There are four inside roots $(I = 4)$ and four outside roots $(\Omega = 4)$. Consequently, equations (46) are not satisfied and no good truncated solution exists. This is verified by the fact that when Lucky attempted to equalize this particular pulse, the equalizer gave an output pulse with large $\gamma_k$, $k > N$. This indicates that a solution such as that shown in Fig. 6 has been approached.
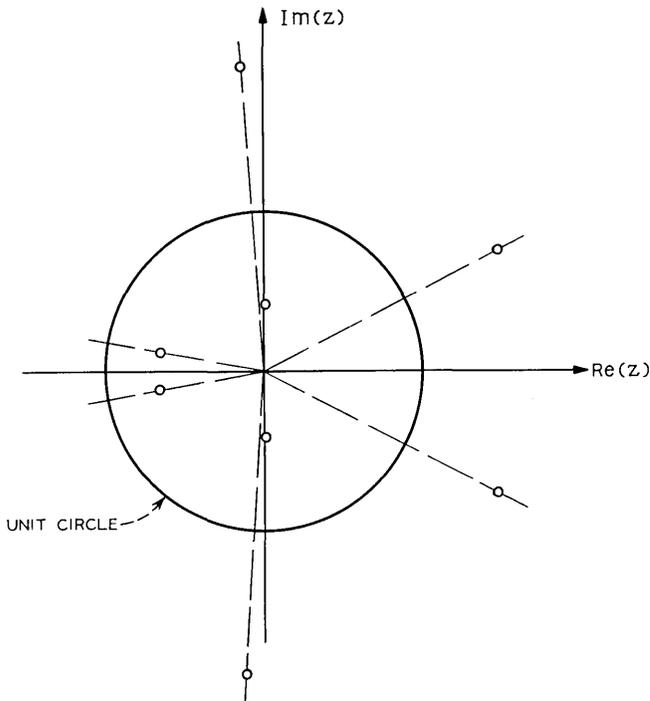


Fig. 7 — Root locations of pulse page 563 of Ref. 5. $B(z) = 15z^8 - 20z^7 + 40z^6 - 60z^5 + 0z^4 + 115z^3 + 60z^2 + 20z + 10$ $\theta_1$, $\theta_1{}^* = 0.404 \angle \pm 89.8° \theta_2$, $\theta_2{}^* = 0.652 \angle \pm 169.9° \phi_1$, $\phi_1{}^* = 1.642 \angle \pm 27.7° \phi_2$, $\phi_2{}^* = 1.894 \angle \pm 94.4°$.

If $\beta_o$ were shifted (delayed) one sample so that it would be the coefficient of the fourth-degree term of $B(z)$, then equations (46) would be satisfied and a good truncated solution would exist. Notice however, that in this case $\beta_o$ would be zero which presents a severe problem in iteratively searching for this good solution.

## VI. ITERATIVE SEARCH FOR EQUALIZATION

The discussion has indicated that, except for singular cases where one or more roots of $B(z)$ are on the unit circle, equalization solutions exist (for the infinite tap case) in which tap weights decrease exponentially in both directions from the center tap. Furthermore, good truncated solutions exist which force all $\gamma_k$ $(-n \leq k \leq N, k \neq 0)$ to zero if $\beta_0$ is selected to satisfy equations (46), which can always be done. The question which now arises is whether a simple iterative search routine will lead to a desirable equalization.

When presented with an impulse response to be equalized using a truncated equalizer and the Lucky criterion of forcing output samples to zero, a hierarchy of questions must be considered:

(*i*) Is the pulse equalizable? That is, are all the roots off the unit circle? (If there are roots on or very close to the unit circle, a change of timing, that is, varying $\tau$ in equation (3), will usually move the roots off the unit circle.)

(*ii*) If the pulse is equalizable, does a good truncated solution exist? A shifting of $\beta$ subscripts can always guarantee the existence of a good solution by satisfying equations (46), but will sometimes create convergence problems.

(*iii*) If the pulse is equalizable and a good truncated solution exists, will a simple iterative search find this solution?

The iterative method of searching for a solution which we consider first consists of measuring the value $(\gamma_k)$ of the $k$th output sample, then subtracting some part of this from the $k$th tap weight.[*] That is

$$\alpha_k^{(r+1)} = \alpha_k^{(r)} - \Delta\gamma_k^{(r)} \tag{49}$$

where the superscripts indicate the iteration number and $\Delta$ is a positive number less than one. This iterative process is not identical to the method presented in Ref. 5 which increments the tap according

---

[*] This is not the only possible iterative search method, but it is one of the simplest.

to only the sign of $\gamma_k$.

$$\alpha_k^{(r+1)} = \alpha_k^{(r)} - \Delta \operatorname{sgn} [\gamma_k^{(r)}].$$     (50)

However, the two methods are very similar, and convergence of one will almost always guarantee convergence of the other.

In matrix notation, the linear iterative search of equation (49) can be written as

$$\alpha^{(r+1)} = \alpha^{(r)} - \Delta\epsilon^{(r)}$$     (51)

$$\epsilon^{(r)} = B\alpha^{(r)} - \mu$$     (52)

where

$$\alpha = \begin{bmatrix} \alpha_N \\ \alpha_{N-1} \\ \vdots \\ \alpha_1 \\ \alpha_0 \\ \alpha_{-1} \\ \vdots \\ \alpha_{-n} \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdot$$

$$B = \begin{bmatrix} \beta_0 & \beta_{-1} & \beta_{-2} & \cdots\cdots\cdots & \beta_{-m} & 0 & 0 & 0 & \cdots & 0 \\ \beta_1 & \beta_0 & \beta_{-1} & \beta_{-2} & \cdots\cdots\cdots & \beta_{-m} \\ \beta_2 & \beta_1 & \beta_0 & \beta_{-1} & \beta_{-2} & \cdots\cdots\cdots & \beta_{-m} \\ \cdots & \beta_2 & \beta_1 & \beta_0 & \beta_{-1} & \beta_{-2} & \cdots\cdots\cdots & \beta_{-m} \\ & & & \beta_0 \\ & & & & \beta_0 \\ & & & & & \beta_0 \\ & & & & & & \ddots \\ & & & & & & & \ddots \\ & & & & & & & & \beta_0 \end{bmatrix}$$

(53)

Equation (52) can be modified by using equation (51),

$$\epsilon^{(r)} = B[\alpha^{(r-1)} - \Delta\epsilon^{(r-1)}] - \mu$$

$$= (I - \Delta B)\epsilon^{(r-1)}$$

$$= (I - \Delta B)^r \epsilon^{(0)} \tag{54}$$

where $I$ is the identity matrix.

Clearly, the iterative routine will converge if

$$\lim_{r \to \infty} \epsilon^{(r)} = 0. \tag{55}$$

However, it does not quarantee convergence to a good solution as illustrated by the pulse of equation (47) and Fig. 7. There, convergence does occur as was illustrated in Ref. 5 and hence, equation (55) is satisfied but convergence is to a bad solution. In that case no good solution exists to which the routine can converge.

The necessary and sufficient condition for convergence is that

$$\lim_{r \to \infty} (I - \Delta B)^r = 0. \tag{56}$$

The matrix of equation (56), that is, the matrix $I - \Delta B$ will converge to zero if and only if all its eigenvalues are less than one in magnitude.[6] This is equivalent to the condition, illustrated in Fig. 8, that the eigenvalues of $B$ lie within a circle of radius $1/\Delta$ centered at $1/\Delta$ on the real axis. Assuming that $\Delta$ can be made as small as necessary, *a necessary and sufficient condition for iterative convergence is that all*
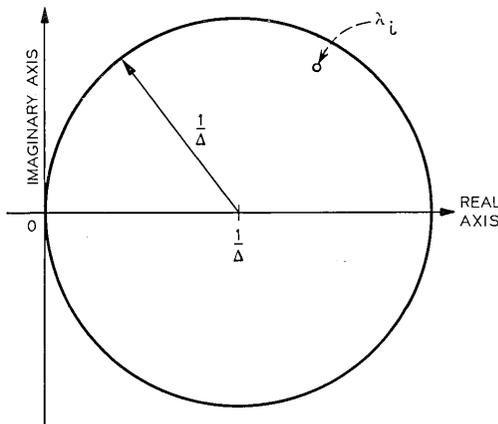


Fig. 8 — Necessary relation of $\Delta$ to location of eigenvalue in complex plane.

*eigenvalues of B should have a positive real part.* If this condition is not satisfied, $(I-\Delta B)^r$ will diverge, and even if the initial error vector $\epsilon^{(0)}$ is very small, the final error vector will grow without bound.

The italicized condition above, coupled with the root location condition of equation (46), guarantees convergence to a good solution. Certain other more restrictive sufficient (but not necessary) conditions can be derived. These conditions may be easier to check for a channel impulse response.

### 6.1 *Necessary and Sufficient Condition for Monotonic Convergence.*

The convergence of the iterative process will be called monotonic if for any starting error vector $\epsilon^{(0)}$, the following inequalities hold (a superscript $t$ indicates the transpose of a matrix).

$$\| \epsilon^{(0)} \| > \| \epsilon^{(1)} \| > \| \epsilon^{(2)} \| > \cdots > \| \epsilon^{(r)} \| > \| \epsilon^{(r+1)} \| > \cdots \tag{57}$$

$$\| \epsilon^{(r)} \| = \epsilon^{(r)\,t}\epsilon^{(r)} = \sum_{i=-n}^{N} [\epsilon_i^{(r)}]^2. \tag{58}$$

Since

$$\epsilon^{(r+1)} = (I - \Delta B)\epsilon^{(r)}, \tag{59}$$

it follows that

$$\| \epsilon^{(r+1)} \| = \epsilon^{(r+1)\,t}\epsilon^{(r+1)} = \epsilon^{(r)\,t}(I - \Delta B)^t(I - \Delta B)\epsilon^{(r)}$$

$$= \epsilon^{(r)\,t}\epsilon^{(r)} + \Delta^2[B\epsilon^{(r)}]^t[B\epsilon^{(r)}] - \Delta\epsilon^{(r)\,t}[B^t + B]\epsilon^{(r)}$$

$$= \| \epsilon^{(r)} \| + \Delta^2 \| B\epsilon^{(r)} \| - \Delta\epsilon^{(r)\,t}[B^t + B]\epsilon^{(r)}. \tag{60}$$

Thus, equation (57) will be satisfied for all possible initial error vectors if an only if

$$\epsilon^t[B^t + B]\epsilon > \Delta \| B\epsilon \| = \Delta(B\epsilon)^t(B\epsilon). \tag{61}$$

Since the right side of the inequality is always positive, the inequality can be satisfied for all possible nonzero $\epsilon$ only if $[B^t + B]$ is positive-definite. If this is true and

$$\min_{\epsilon} \frac{\epsilon^t[B^t + B]\epsilon}{\| B\epsilon \|} > \Delta \tag{62}$$

then equation (61) will be satisfied for all $\epsilon$ and the iterative process will be monotonically convergent.

If $[B^t + B]$ is not positive-definite, convergence can still occur. However, stating the conditions for convergence becomes more difficult. For example, if $[B^t+B]$ is not positive-definite then for some vector, say $\epsilon^{(r-1)}$,

$$\epsilon^{(r-1)\,t}[B^t + B]\epsilon^{(r-1)} < 0 \qquad (63)$$

and

$$\| \epsilon^{(r-1)} \| < \| \epsilon^{(r)} \| . \qquad (64)$$

But if for $\epsilon = \epsilon^{(r-1)}$

$$(B\epsilon)^t[B^t + B_1^\cdot](B\epsilon) > \frac{1}{2^3} \{\Delta^4 \| B^2\epsilon \| + 2\Delta^2 \| B\epsilon \|$$

$$+ \Delta^2 \| [B^t + B]\epsilon \| - 2\Delta\epsilon^t[B^t + B]\epsilon\} \qquad (65)$$

then

$$\| \epsilon^{(r-1)} \| > \| \epsilon^{(r+1)} \| . \qquad (66)$$

Very roughly speaking, equations such as (65) which can be developed indicate that $[B^t+B]$ should yield a predominantly positive quadratic form in order to have convergence. What we mean by this can best be illustrated graphically, as in Fig. 9.

### 6.2 *Further Sufficient Conditions for Monotonic Convergence*

The positive-definiteness of $[B^t+B]$ is a necessary and sufficient condition for monotonic convergence. Being somewhat more restrictive will yield other sufficient (but not necessary) conditions. We notice that $[B^t + B]$ is a Toeplitz matrix, that is,

$$[B^t + B] = [b_{ij}], \quad b_{ij} = b_{|i-j|} = \beta_{(i-j)} + \beta_{(j-i)} . \qquad (67)$$



VARIATION IN DIRECTION OF VECTOR $\epsilon$ WITH
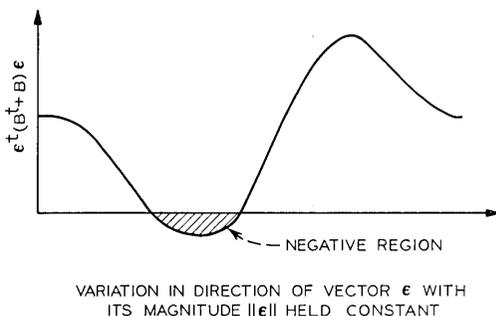ITS MAGNITUDE $\|\epsilon\|$ HELD CONSTANT

Fig. 9 — Illustrating a predominately positive quadratic form.

Thus, it has some of the necessary attributes of a correlation matrix. A correlation matrix is positive-definite, so if we can show that $[B^t+B]$ is a correlation matrix, then this is sufficient for positive-definiteness. If $[B^t + B]$ is a correlation matrix then the polynomial $B(z) + B^t(z) = P(z)$, where

$$B(z) = \beta_M z^{M+m} + \cdots + \beta_{-m}$$

$$B^t(z) = \beta_{-m} z^{2m} + \cdots + \beta_M z^{m-M} \tag{68}$$

must have no odd order roots on the unit circle. Root locus methods may be applied here, that is, the loci of

$$\frac{B(z)}{B^t(z)} = -k \tag{69}$$

must not cross the unit circle. Notice that this is a necessary condition that $[B^t + B]$ be a correlation matrix and it is sufficient but *not* necessary for $[B^t + B]$ being positive-definite.

A more restrictive condition leads to another sufficiency condition. If (Lucky condition, Ref. 5)

$$\beta_0 > \sum_{k \neq 0} | \beta_k |, \quad \text{that is,} \quad D_0 < 1 \tag{70}$$

then it is easy to see that $B^t(z) + B(z)$ has no roots on the unit circle and thus $[B^t + B]$ is a correlation matrix and positive-definite. Notice that equation (70) is not a necessary condition for $[B^t + B]$ being a correlation matrix.

The various conditions discussed above are summarized in Fig. 10.
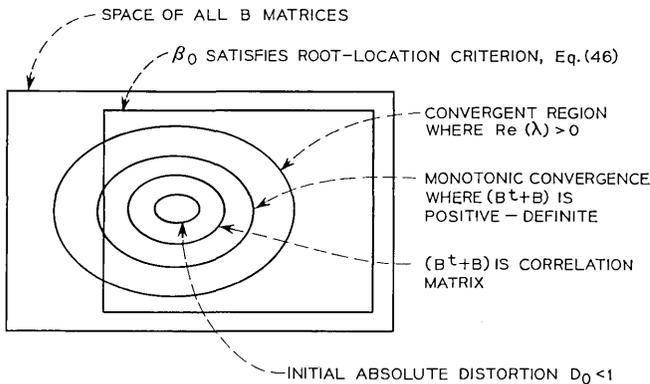


Fig. 10 — Illustrating the different convergence regions.

## VII. MODIFIED ITERATION PROCESSES

In the sample iteration process which has been discussed to this point, the change in tap weight $\alpha_j$ depends only on the output sample value $\gamma_j$. Let us generalize so that the change depends on a linear combination of output samples.

$$\left.\begin{aligned}
B\alpha^{(r)} &= \gamma^{(r)} \\
\epsilon^{(r)} &= \gamma^{(r)} - \mu = B\alpha^{(r)} - \mu \\
\alpha^{(r+1)} &= \alpha^{(r)} - \Delta V \epsilon^{(r)}
\end{aligned}\right\} . \tag{71}$$

That is to say, the $(r + 1)$st value of tap $j$ is

$$\alpha_j^{(r+1)} = \alpha_j^{(r)} - \Delta \sum_k v_{jk}\epsilon_k^{(r)} . \tag{72}$$

Manipulation of equations (71) leads to a result similar to equation (54)

$$\epsilon^{(r)} = [I - \Delta BV]^r \epsilon^{(0)} . \tag{73}$$

Consequently, all the sufficient and necessary conditions which have been developed on the previous pages can now be applied to the matrix $BV$. Now, however, we have considerably more latitude since we are free to specify $V$.

As an example, let us suppose that $V$ is chosen to equal $B^t$. Then

$$BV = BB^t = \text{positive-definite} \tag{74}$$

and monotonic convergence is guaranteed. This is a particularly appealing way of selecting $V$ since the sample weighting can be determined directly from the initial channel impulse response.

$$v_{ik} = v_{i-k} = \beta_{k-i} . \tag{75}$$

It is very interesting that the weighting suggested above is very nearly equivalent to inserting a tapped-delay-line matched filter ahead of the equalizer. A matched filter, whose tap weights are equal to the $\beta$ values in reverse order, will yield an output whose samples will form a $B$ matrix which is a correlation matrix. Thus, the iterative search will be monotonically convergent in this case also. The weighting suggested above yields the same matrix except for "edge effects."

This can be illustrated by the following example. Suppose there are just three $\beta$ values: $\beta_{-1} = 1$, $\beta_0 = 2$, $\beta_1 = -2$. Then, with $V$ set

equal to $B^t$, and assuming a six-tap equalizer so that $B$ is a $6 \times 6$ matrix, we obtain

$$
BV = BB^t = \begin{bmatrix}
5 & -2 & -2 & 0 & 0 & 0 \\
-2 & 9 & -2 & -2 & 0 & 0 \\
-2 & -2 & 9 & -2 & -2 & 0 \\
0 & -2 & -2 & 9 & -2 & -2 \\
0 & 0 & -2 & -2 & 9 & -2 \\
0 & 0 & 0 & -2 & -2 & 8
\end{bmatrix}. \tag{76}
$$

On the other hand, the modified $B$ matrix following a matched filter would be

$$
B_m = \begin{bmatrix}
9 & -2 & -2 & 0 & 0 & 0 \\
-2 & 9 & -2 & -2 & 0 & 0 \\
-2 & -2 & 9 & -2 & -2 & 0 \\
0 & -2 & -2 & 9 & -2 & -2 \\
0 & 0 & -2 & -2 & 9 & -2 \\
0 & 0 & 0 & -2 & -2 & 9
\end{bmatrix} \tag{77}
$$

Setting $V = B^t$ will guarantee monotonic convergence. However, since in most cases $B$ is such that $[B^t + B]$ is close to being positive-definite, it is probable that a less extensive $V$ would be sufficient to guarantee monotonic convergence. As an example, suppose that $V$ is chosen to be a small deviation on the standard iteration of equation (58)

$$
V = I + \delta B^t, \qquad \delta > 0. \tag{78}
$$

Now,

$$
BV = B + \delta BB^t \tag{79}
$$

will have a quadratic form which is greater for every vector $\epsilon$ than the quadratic form for the matrix $B$ alone. Thus, if the negative region such as is illustrated in Fig. 8 is small, then $BV$ can become positive-definite for relatively small $\delta$.

Perhaps a more reasonable way of selecting a $V$ which approximates $B^t$ is to modify equation (75) in the following manner.

$$v_{jk} = v_{j-k} = \begin{cases} \beta_{j-k} & \text{if} \quad | \beta_{k-j} | \geqq L \\ 0 & \text{if} \quad | \beta_{k-j} | < L \end{cases}. \qquad (80)$$

Thus, only the more significant values of $\beta$ are used in weighting the errors. In general such a $V$ will force $B$ towards positive-definiteness. However, there seems to be no general rule for selecting the critical value of $L$, that is the maximum $L$ which will just permit positive-definiteness.

## VIII. CONCLUDING REMARKS

Virtually all input pulses are equalizable in the sense that there exist tap-weight adjustments which will force the output samples in the adjustment interval to zero while the output samples out of the adjustment interval remain small. Furthermore, the residual samples outside of the adjustment interval will become smaller as the number of taps (length of transversal filter) is increased.

There are just two necessary conditions in order for the preceding statement to hold. The first is that the polynomial representing the input pulse have no roots on the unit circle. Although the singular case where roots are exactly on the unit circle is highly improbable, roots very near the unit circle lead to relatively larger residual errors and greater potential for instability.

The second necessary condition is that the selection of the central sample value must be such that equation (46) is satisfied.

Although a pulse may be equalizable, the simple first-order iterative search for the proper tap weights given by equations (51) and (52) may not be convergent. If it is convergent, and assuming the two conditions above are satisfied, it will converge to the proper tap-weight settings. If it is not convergent, it will be divergent with increasing errors in the adjustment interval. The convergence or divergence is independent of the initial tap settings. Thus, even though the tap weights might be set to optimum initially, if the system is in the iterative search mode and is divergent, it will eventually diverge.

The necessary and sufficient condition for convergence is given in Section VI along with a hierarchy of more stringent sufficient conditions. In general, convergence will be dependent upon the absolute timing of the sampling. Consequently, a particular pulse which is equalizable for two different timings might be convergent for one timing and divergent for the other.

If the first-order iterative procedure is divergent, a more complex

weighting of the output errors in adjusting the tap weights can improve the situation. At least one weighting given by equation (75) will guarantee convergence.

REFERENCES

1. Nyquist, H., "Certain Topics in Telegraph Transmission Theory," AIEE Trans. Commun. and Elec., 47, No. 2 (April 1928), pp. 617–644.
2. Bennett, W. R. and Davey, J. R., Data Transmission, New York: McGraw-Hill, 1965, p. 61.
3. Gibby, R. A. and Smith, J. W., "Some Extensions of Nyquist's Telegraph Transmission Theory," B.S.T.J., 46, No. 7 (September 1965), pp. 1487–1510.
4. Lucky, R. W., "Techniques for Adaptive Equalization of Digital Communication Systems," B.S.T.J., 45, No. 2 (February 1966), pp. 255–286.
5. Lucky, R. W., "Automatic Equalization for Digital Communication," B.S.T.J., 44, No. 4 (April 1965), pp. 547–588.
6. Householder, A. S., The Theory of Matrices in Numerical Analysis, New York: Blaisdell Publishing Co., 1964, pp. 53–55.

# On Solutions for Two Waves with Periodic Coupling

By S. E. MILLER

An exact solution for the coupling effects between two waves with a particular complex periodic coupling function is presented; the particular coupling function gives the same wave interactions as constant coupling but at a translated value of differential phase constant. A transformation is given which permits known theory for constant coupling to be applied to the periodic coupling case.

Approximate solutions are given for periodically reversed coupling (sinusoidal or square wave) between two waves, and calculations are presented which indicate the solutions are valid for arbitrarily long coupling regions or arbitrarily large integrated coupling strengths. The region of validity for earlier perturbation theory is defined and proved to include the cases of interest for multimode circular electric waveguides.

## I. INTRODUCTION

This paper describes some solutions for two waves with periodic coupling. Coupled waves have been important in a wide variety of communication devices: transmission lines, directional couplers, amplifiers, and in describing mode interchange phenomena generally.[1] Multimode transmission lines have been advantageously described through coupled wave equations, and a particular situation of importance exists in the circular electric waveguide.

As first shown by H. E. Rowe and W. D. Warters,[2] periodic straightness variations cause periodically reversed coupling from the circular electric wave to several other waves, and this interaction results in the most difficult tolerances on the fabrication and installation of the waveguide itself. Publications by H. E. Rowe and W. D. Warters have provided a comprehensive understanding of the fundamentals involved and have given explicit expressions based on perturbation theory for calculating the loss versus frequency varia-

tion resulting from such periodic mode conversion.[3] A subsequent publication by D. T. Young[4] has indicated how the approximate solution of Rowe and Warters may be transformed mathematically to show explicitly the way differential attenuation smooths out the mode-coupling effects on the loss versus frequency characteristic. Young's solution depends upon a valid perturbation solution. A discussion of the accuracy of the perturbation solution is given by H. E. Rowe;[5] when the differential attenuation is too small the perturbation solution breaks down and it is of interest to know exactly where and quantitatively how this occurs.

We show in this paper that a certain periodic coupling function has exactly the same effect on waves of unequal phase constant as uniform coupling between waves of identical phase constant. A transformation is given to allow the use of earlier theory for periodically coupled waves.

Also presented here is an approximate solution for the periodic coupling distribution sketched in Fig. 2 valid for any value of differential attenuation. It is true that known solutions for uniform coupling, as in Fig. 1, can be applied to Fig. 2 by simplying solving for the output values at $x = \lambda_m/2$ and using these as the input boundary conditions for the transmission region starting at $x = \lambda_m/2$. The resulting exact expression representing conditions at $x = \lambda_m$ can be expressed as a matrix and raised to the $n^{\text{th}}$ power to represent the solution at $x = n\lambda_m$. We seek here a simpler form of expression in which the functional interrelations can be visualized without extensive numerical calculations.

## II. EXACT SOLUTION FOR TWO PERIODICALLY COUPLED WAVES

We start with the following equations for two coupled waves:

$$\frac{d}{dz} E_1(z) = -\gamma_1 E_1 + c_{21}(z)E_2 \tag{1}$$

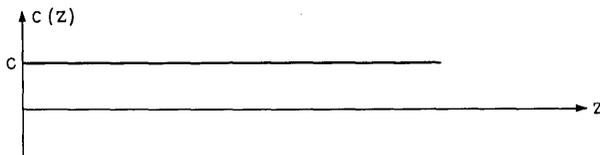$$\frac{d}{dz} E_2(z) = c_{12}(z)E_1 - \gamma_2 E_2 \tag{2}$$
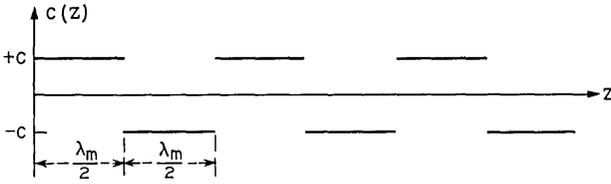
Fig. 1 — Constant coupling.

Fig. 2 — Square-wave coupling.

in which $\gamma_1$ and $\gamma_2$ are the complex propagation constants and $c_{12}$ and $c_{21}$ are coupling functions. Appendix A shows that the coupling functions

$$c_{21} = jc_p \exp(jk_c z) = jc_p[\cos k_c z + j \sin k_c z] \tag{3}$$

and

$$c_{12} = jc_p \exp(-jk_c z) = jc_p[\cos k_c z - j \sin k_c z] \tag{4}$$

give the same solutions previously found[6] for $k_c = 0$ provided that

$$k_c = (\beta_2 - \beta_1) = \Delta\beta \tag{5}$$

where $c_p$ is a constant and

$$\gamma_m = \alpha_m + j\beta_m . \tag{6}$$

When (5) holds, complete transfer of power between waves can occur. When (5) does not hold, the resulting wave interactions can be calculated using previously developed theory for $k_c = 0$ and substituting $\Delta B$ for $\Delta\beta$ in the $k_c = 0$ solutions, where

$$\Delta B = \Delta\beta + k_c \tag{7}$$

III. PHYSICAL REALIZATION OF IDEAL PERIODICALLY COUPLED WAVES

We describe here a physical realization of waves coupled according to equations (3) and (4), and cite an advantage in mode selective directional couplers.

Figure 3 shows a mode-selective coupler between $TE_{10}^{\square}$ of rectangular guide and $TE_{01}^{\circ}$ of round guide. The thin dielectric lining is used to break the $TE_{01}^{\circ} - TM_{11}^{\circ}$ degeneracy. The longitudinal magnetic intensity $h_z^{\circ}$ of the $TE_{01}^{\circ}$ wave is coupled to the longitudinal magnetic intensity $h_z^{\square}$ of $TE_{10}^{\square}$ in the off-axis longitudinal slots and is also coupled to the transverse magnetic intensity $h_t^{\square}$ of $TE_{10}^{\square}$ in the 45° slots on-axis. In each case the magnitude of the coupling is set by the length and width of the slot.
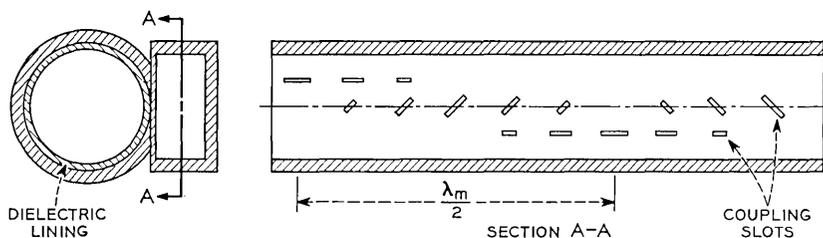
Fig. 3 — TE$^\circ_{10}$-TE$^\circ_{01}$ coupler using unequal phase constants in the rectangular and round waveguides.

The phase reversal of the longitudinal $h^\square_z$ coupling is accomplished by reversing the slot position relative to the centerline, and the phase reversal of the transverse $h^\square_t$ coupling is accomplished by reversing the slant angle of the slot. One set of slots represents the sine term. It is not necessary that any particular fraction of $\lambda_m$ be used in a coupler, since the coupling is always in the same phase relative to the desired waves in the two guides. To accomplish the desired mode selectivity the $\Delta\beta$ between the $TE^\square_{10}$ and $TE^\circ_{01}$ waves is made equal to $k_c = 2\pi/\lambda_m$ as in equation (5).

An advantage of a coupler of this form, compared with one in which constant coupling is used with $\Delta\beta = 0$, is that the waveguides can have the standard dimensions set by other considerations.

Other illustrations of useful coupling between waves of unequal phase constants will be given in another paper which the author is preparing.

IV. SQUARE-WAVE OR SINUSOIDAL COUPLING

We present here the results of Appendices B and C which discuss approximate solutions for the cases in which the coupling is a square wave as in Fig. 2, or the corresponding sinusoidal

$$c(z) = \sin(2\pi z/\lambda_m). \tag{8}$$

Both solutions are expressed in the form and notation of a previous publication[6] giving the solution when the coupling is constant, and the boundary conditions $E_1(0) = 1.0$ and $E_2(0) = 0$ are impressed:

$$E_1(z) = e^{-\gamma_1 z}\{Ae^{r_1 z} + Be^{r_2 z}\} \tag{9}$$

$$E_2(z) = \frac{e^{-\gamma_1 z}}{2\sqrt{\phantom{x}}} \{e^{r_1 z} - e^{r_2 z}\} \tag{10}$$

in which

$$A = \left[ \frac{1}{2} - \frac{1}{2} \frac{\left(\frac{\Delta\beta_*}{2c_*}\right) - i\left(\frac{\Delta\alpha}{2c_*}\right)}{\sqrt{\phantom{xx}}} \right] \qquad (11)$$

$$B = \left[ \frac{1}{2} + \frac{1}{2} \frac{\left(\frac{\Delta\beta_*}{2c_*}\right) - i\left(\frac{\Delta\alpha}{2c_*}\right)}{\sqrt{\phantom{xx}}} \right] \qquad (12)$$

$$r_{\frac{1}{2}} = \frac{\Delta\gamma_*}{2} \pm ic_* \sqrt{\phantom{xx}} \qquad (13)*$$

$$\sqrt{\phantom{xx}} = \sqrt{1 + \left(\frac{\Delta\beta_*}{2c^*}\right)^2 - \left(\frac{\Delta\alpha}{2c_*}\right)^2 - i2\left(\frac{\Delta\beta_*}{2c^*}\right)\left(\frac{\Delta\alpha}{2c^*}\right)} \qquad (14)$$

$$\Delta\alpha = \alpha_1 - \alpha_2$$

$$\Delta\gamma_* = \Delta\alpha + i\Delta\beta. \ .$$

The wave interactions are described by the above equations provided the following values of $c_*$ and $\Delta\beta_*$ given in Table I are used.

TABLE I — VALUES OF $c_*$ AND $\Delta\beta_*$

| Coupling function (see Fig. 2) | $c_*$ | $\Delta\beta_*$ |
|---|---|---|
| Constant $= c$ | $c$ | $\Delta\beta = (\beta_1 - \beta_2)$ |
| Square-wave of magnitude $c$ and period $\lambda_m$ | $\frac{2}{\pi} c$ | $\Delta\beta - \frac{2\pi}{\lambda_m}\sqrt{1 - \left(\frac{c\lambda_m}{\pi}\right)^2}$ |
| $c \sin\left(\frac{2\pi z}{\lambda_m}\right)$ | $\frac{c}{2}$ | $\Delta\beta - \frac{2\pi}{\lambda_m}$ |

These solutions, equations (9) and (10), have been obtained in Appendix B by relating the rate of transfer of power (that is, transfer over a short length interval) for the periodically reversed coupling to that for constant coupling, and noting the effective value of coupling $c_*$ and effective differential phase constant $\Delta\beta_*$. The solutions are correct for $z$ equal to an integral multiple of $\lambda_m/2$, and may be in error by less than approximately $0.2c\lambda_m/\pi$ at intermediate values of $z$.

There probably should be a correction factor in $\Delta\beta_*$ for sinusoidal

---

* $r_1$ corresponds to the $+$ sign and $r_2$ corresponds to the $-$ sign.

coupling similar to the radical shown in Table I for the square wave coupling; the work done thus far has not defined what it should be, but for small $c\lambda_m$ the radical is negligible for most purposes.

The values of $\Delta\beta_*$ in Table I bear marked resemblance to equation (7); however, because the simple sinusoid or square wave coupling phase is two-valued (versus $z$) instead of continually progressing (equations 3 and 4) to provide coupling continuously in step with the phase changes of waves 1 and 2, there are local maxima* in coupling effects at other values of $\Delta\beta$ for sinusoidal or square wave coupling. Appendix B shows that the wave interaction effects are properly described for square wave coupling in the regions near

$$\Delta\beta\lambda_m = 2\pi p \tag{15}$$

$$p = 1, 3, 5, \cdots$$

by the transformations

$$\Delta\beta_* = \Delta\beta - p\frac{2\pi}{\lambda_m}\sqrt{1 - \left(\frac{c\lambda_m}{p\pi}\right)^2} \tag{16}$$

$$c_* = \frac{2c}{p\pi}. \tag{17}$$

Because $c_*$ drops off rapidly with increasing $p$ the corresponding wave interaction effects drop off also.

### 4.1 Numerical Comparison of Approximate and Exact Solutions

A few calculations have been made to find quantitatively the error resulting from the approximations made in equations (9) and (10) for coupling as in Fig. 2. An "exact" solution is obtained by using exact uniform coupling theory on each interval of 0.5 $\lambda_m$, the output of one interval being taken as the input to the next interval.

We take first the simplest case,

$$\Delta\alpha = 0.$$

Then equation (9) becomes

$$E_1 \mid_{\Delta\alpha=0} = \exp\left[i\frac{(\beta_1 + \beta_2)_z}{2}\right]$$

$$\cdot\left\{\cos\left[\sqrt{-\frac{2}{\pi}cz}\right] - i\frac{\left(\frac{\Delta\beta_*}{4c/\pi}\right)}{\sqrt{}}\sin\left[\sqrt{-\frac{2}{\pi}cz}\right]\right\} \tag{18}$$

---

* Rowe and Warters noted this in their work recorded in Ref. 3.

$$E_2 \mid_{\Delta\alpha=0} = -\exp\left[i\frac{(\beta_1 + \beta_2)z}{2}\right]\left\{\frac{ic}{\sqrt{\phantom{-}}}\sin\left[\sqrt{\phantom{-}}\frac{2}{\pi}cz\right]\right\} \qquad (19)$$

in which

$$\sqrt{\phantom{-}} = \sqrt{\left(\frac{\Delta\beta_*}{4c/\pi}\right)^2 + 1} \qquad (20)$$

and $\Delta\beta_*$ is as given in Table I for square wave coupling. We notice in passing that for $\Delta\beta_* = 0$

$$E_1 = \cos\left(\frac{2}{\pi}cz\right) \qquad (21)$$

$$E_2 = i\sin\left(\frac{2}{\pi}cz\right) \qquad (22)$$

and the power exchanges completely back and forth between the two waves as a function of $z$; this is of course identical to the behavior in uniformly coupled waves, but with a modified period given by the $2/\pi$ factor.

We take for the first numerical comparison the condition

$$\frac{2}{\pi}cz = \frac{\pi}{2}$$

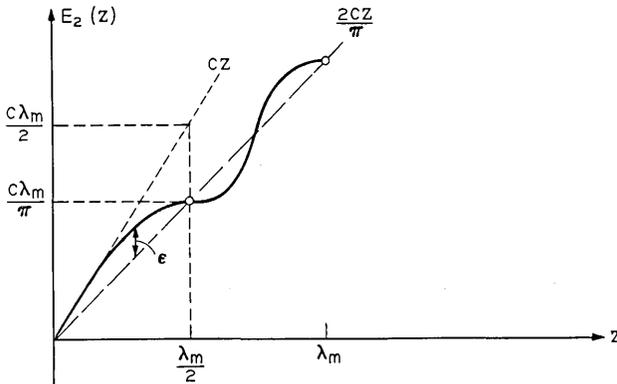so that at $\Delta\beta_* = 0$, $E_1 = 0$ and $\mid E_2 \mid = 1.0$. The additional specific numbers used are



Fig. 4 — Undriven-wave amplitude versus distance for lowest order square-wave coupling.

$$\lambda_m = 2 \text{ feet}$$

$$z = 1000 \text{ feet}$$

$$c = 2.46740 \times 10^{-3} \text{ ft}^{-1}.$$

Fig. 6 shows a plot of the loss $20 \log | E_1 |$ (labelled $| \Delta\alpha/c | = 0$) versus $\Delta\beta_* \lambda_m/2$ and Table II shows the comparison of the exact versus approximate calculations, the latter obtained from equation (9) and associated expressions. It may be kept in mind that $\Delta\beta$ and $\Delta\beta_*$ are inversely proportional to frequency in many cases of interest, so that Fig. 6 is a loss versus frequency plot associated with a particular periodic coupling component. The radical in the expression for $\Delta\beta_*$ was ignored in the comparison of Table II and this might account for the consistent positive difference (approximate—exact) for $\Delta\beta\lambda_m/2$ greater than $\pi$. Aside from the pole at $\Delta\beta\lambda_m/2 = \pi$, the two calculations agree to better than 1 percent in dB even when the loss is a few tenths of a dB.

Figs. 6 and 7 and Tables III, IV, and V show similar comparisons for $\Delta\alpha = - | c |$, $\Delta\alpha = -5 | c |$, and $\Delta\alpha = -50 | c |$. Excellent agreement is obtained in all cases.

Table VI shows a comparison of the phase angle on $E_1$, computed by the two approaches. For the points shown and for the other points (not shown) corresponding to the amplitude values of Tables II and V the agreement is excellent. Fig. 8 shows a plot of the phase, where odd symmetry about $\Delta\beta_* = 0$ is understood.

A check has also been made on the accuracy of equation (9) in the region near $\Delta\beta\lambda_m = 6\pi$, corresponding to $p = 3$ in equations (15) through (17). The same parameters were used as in the calculations for Figs. 6 and 7. The results are plotted in Figs. 9 and 10 which represent *both* the approximate calculation from equation (9) and the exact calculation. The differences are on the same order as given in Table II, and are too small to show in the figures. Figures 9 and 10 may be compared directly with Figs. 6 and 7 to see the "third harmonic" loss (labelled $E_1^3$ in Figs. 9 and 10) in relation to the "fundamental" loss, Figs. 6 and 7.

4.2 *Interpretation and Further Simplification*

Consider first the shape of the loss versus $\Delta\beta_*$ (equivalent to loss versus frequency) curves. In the limit
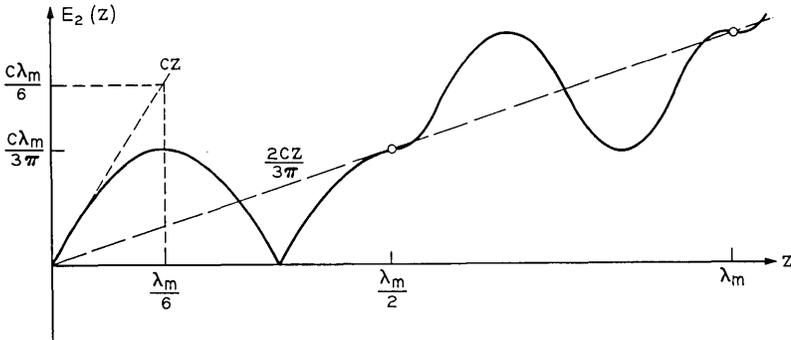
$$\left| \frac{\Delta\alpha}{c} \right| \gg 1 \tag{23}$$

Fig. 5 — Undriven-wave amplitude versus distance for "third harmonic" in square-wave coupling.

TABLE II — EXACT AND APPROXIMATE CALCULATIONS
OF $E_1$ FOR $|\Delta\alpha/c| = 0$

| $\Delta\beta\left(\dfrac{\lambda_m}{2}\right)$ | 20 log $|E_1|$ | | Approximate −exact |
| --- | --- | --- | --- |
| | Approximate | Exact | |
| $\pi - 0.0157079$ | − 0.1662db | − 0.1680db | + 0.0018db |
| $\pi - 0.01413716$ | − 0.1389 | − 0.1403 | + 0.0014 |
| $\pi - 0.01256637$ | − 0.0094 | − 0.0096 | + 0.0002 |
| $\pi - 0.01099557$ | − 0.0951 | − 0.0960 | + 0.0009 |
| $\pi - 0.009424777$ | − 0.4270 | − 0.4300 | + 0.0030 |
| $\pi - 0.0078539816$ | − 0.4973 | − 0.4999 | + 0.0026 |
| | | | |
| $\pi - 0.006283185$ | − 0.1156 | − 0.1161 | + 0.0005 |
| $\pi - 0.0047123889$ | − 0.1260 | − 0.1267 | + 0.0007 |
| $\pi - 0.00314159264$ | − 1.6530 | − 1.6582 | + 0.0052 |
| $\pi - 0.0015707963$ | − 6.4362 | − 6.4540 | + 0.0178 |
| | | | |
| $\pi$ | − 122.73 | − 58.54 | |
| $\pi + 0.0015707963$ | − 6.4362 | − 6.4184 | − 0.0178 |
| $\pi + 0.00314159264$ | − 1.6530 | − 1.6481 | − 0.0049 |
| $\pi + 0.0047123889$ | − 0.1260 | − 0.1255 | − 0.0005 |
| $\pi + 0.006283185$ | − 0.1156 | − 0.1154 | − 0.0002 |
| $\pi + 0.0078539816$ | − 0.4973 | − 0.4949 | − 0.0024 |
| | | | |
| $\pi + 0.00942477$ | − 0.4270 | − 0.4243 | − 0.0027 |
| $\pi + 0.01099557$ | − 0.0951 | − 0.0944 | − 0.0007 |
| $\pi + 0.01256637$ | − 0.0094 | − 0.00964 | + 0.0002 |
| $\pi + 0.01413716$ | − 0.1389 | − 0.1378 | − 0.0011 |
| $\pi + 0.0157079$ | − 0.1662 | − 0.1646 | − 0.0016 |
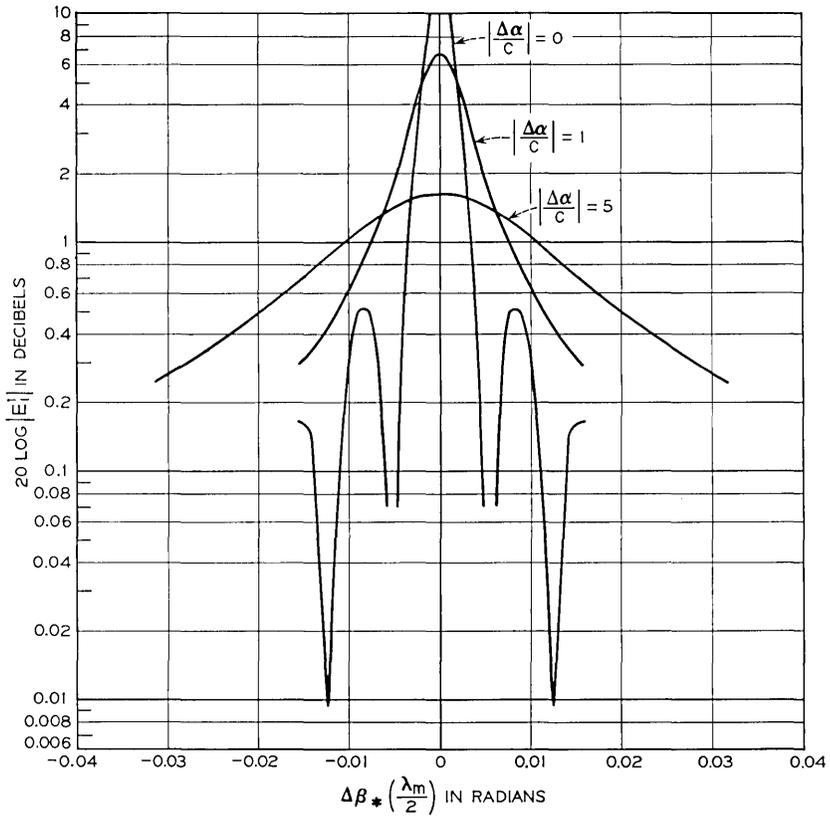
Fig. 6 — Driven-wave loss versus $\Delta\beta_*$ near $\Delta\beta\lambda_m = 2\pi$ (fundamental) for $|\Delta\alpha/c| = 0,1,5$.

earlier perturbation theory[4] has shown that the fractional frequency interval between half-height points on the loss curve is

$$\frac{\Delta f}{f} = \frac{\Delta\beta_{*\frac{1}{2}}}{\Delta\beta} = 2 \left| \frac{\Delta\alpha}{\Delta\beta} \right| \tag{24}$$

or

$$\Delta\beta_{*\frac{1}{2}} = 2 \mid \Delta\alpha \mid. \tag{25}$$

Table VII shows a comparison between that limiting value and the true value for the numerical cases above, including $\mid \Delta\alpha/c \mid$ from one to 50. Even at $\mid \Delta\alpha/c \mid = 1$ there is only a 30 per cent error.

Consider the limiting case $\Delta\alpha = 0$. Then

$$E_1 = [1 - (E_2)^2]^{\frac{1}{2}} \tag{26}$$

and nulls in the loss curve occur when equation (19) representing $E_2$ is zero. This gives

$$\Delta\beta_* \Big|_{\text{loss null}} = \frac{4c}{\pi}\sqrt{\left(\frac{\pi^2}{2cz}\right)^2 - 1}. \tag{27}$$

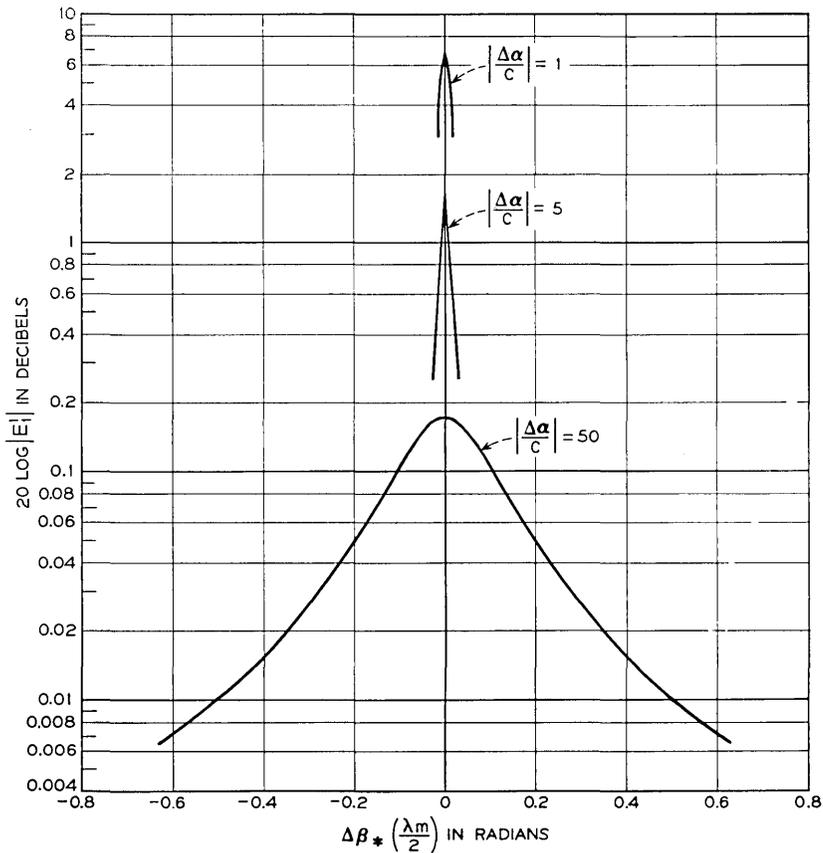When we also have the perturbation condition, $cz \ll 1$, equation (27)



Fig. 7 — Driven-wave loss versus $\Delta\beta_*$ near $\Delta\beta\lambda_m = 2\pi$ (fundamental) for $|\Delta\alpha/c| = 1,5,50$.

becomes

$$\Delta\beta_* \mid_{\text{null}} \cong \frac{2\pi}{z} \qquad (28)$$

which agrees with previously known perturbation theory.[3] For this

TABLE III — EXACT AND APPROXIMATE CALCULATIONS
OF $E_1$ FOR $\mid \Delta\alpha/c \mid = 1$

| $\Delta\beta\left(\dfrac{\lambda m}{2}\right)$ | 20 log $\mid E_1 \mid$ | | Approximate −exact |
|---|---|---|---|
| | Approximate | Exact | |
| $\pi - 0.0157079$ | − 0.2898 | − 0.2910db | + 0.0012db |
| $\pi - 0.01413716$ | − 0.3438 | − 0.3450 | + 0.0012 |
| $\pi - 0.01256637$ | − 0.4152 | − 0.4164 | + 0.0012 |
| $\pi - 0.01099557$ | − 0.5452 | − 0.5440 | + 0.0012 |
| $\pi - 0.009424777$ | − 0.7300 | − 0.7319 | + 0.0019 |
| $\pi - 0.0078539816$ | − 0.9601 | − 0.9620 | + 0.0019 |
| $\pi - 0.006283185$ | − 1.3106 | − 1.3126 | + 0.0020 |
| $\pi - 0.0047123889$ | − 2.0216 | − 2.0243 | + 0.0027 |
| $\pi - 0.00314159264$ | − 3.3781 | − 3.3823 | + 0.0042 |
| $\pi - 0.0015707963$ | − 5.3475 | − 5.3523 | + 0.0048 |
| $\pi$ | − 6.5701 | − 6.5705 | + 0.0004 |
| $\pi + 0.0015707963$ | − 5.3475 | − 5.3434 | + 0.0041 |
| $\pi + 0.00314159264$ | − 3.3781 | − 3.3746 | − 0.0035 |
| $\pi + 0.0047123889$ | − 2.0216 | − 2.0194 | − 0.0022 |
| $\pi + 0.006283185$ | − 1.3106 | − 1.3092 | − 0.0014 |
| $\pi + 0.0078539816$ | − 0.9601 | − 0.9588 | − 0.0013 |
| $\pi + 0.00942477$ | − 0.7300 | − 0.7289 | − 0.0011 |
| $\pi + 0.01099557$ | − 0.5425 | − 0.5416 | − 0.0009 |
| $\pi + 0.01256637$ | − 0.4152 | − 0.4146 | − 0.0006 |
| $\pi + 0.01413716$ | − 0.3438 | − 0.3430 | − 0.0008 |
| $\pi + 0.0157079$ | − 0.2898 | − 0.2892 | − 0.0006 |

TABLE IV — EXACT AND APPROXIMATE CALCULATIONS
OF $E_1$ FOR $\mid \Delta\alpha/c \mid = 5$

| $\Delta\beta\left(\dfrac{\lambda m}{2}\right)$ | 20 $\log_{10} \mid E_1 \mid$ | | Approximate −exact |
|---|---|---|---|
| | Approximate | Exact | |
| $\pi - 0.0314159262$ | − 0.2446db | − 0.2452db | + 0.0006db |
| $\pi - 0.01884955$ | − 0.5337 | − 0.5344 | + 0.0007 |
| $\pi$ | − 1.6195 | − 1.6197 | + 0.0002 |
| $\pi + 0.01884955$ | − 0.5337 | − 0.5344 | + 0.0007 |
| $\pi + 0.0314159$ | − 0.2446 | − 0.2445 | + 0.0001 |

TABLE V — EXACT AND APPROXIMATE CALCULATIONS
OF $E_1$ FOR $|\Delta\alpha/c| = 50$

| $\Delta\beta\left(\dfrac{\lambda_m}{2}\right)$ | $20 \log_{10}\|E_1\|$ | | Approximate −exact |
| --- | --- | --- | --- |
| | Approximate | Exact | |
| $\pi - 0.314159$ | $- 0.0234\text{db}$ | $- 0.0238\text{db}$ | $+ 0.0004\text{db}$ |
| $\pi - 0.1884955$ | $- 0.0523$ | $- 0.0525$ | $- 0.0002$ |
| $\pi$ | $- 0.1724$ | $- 0.1727$ | $+ 0.0003$ |
| $\pi + 0.1884955$ | $- 0.0523$ | $- 0.0525$ | $+ 0.0002$ |
| $\pi + 0.219911$ | $- 0.0418$ | $- 0.0421$ | $+ 0.0003$ |

case, from (26)

$$E_1 \cong 1 - \tfrac{1}{2}\,|\,E_2\,|^2$$

$$= 1 - \tfrac{1}{2}\left\{\frac{\sin\left[\sqrt{\left(\dfrac{\Delta\beta_*}{4\pi/c}\right)^2 + 1}\,\dfrac{2}{\pi}\,cz\right]}{\sqrt{\left(\dfrac{\Delta\beta_*}{4\pi/c}\right)^2 + 1}}\right\}^2 \tag{29}$$

and the loss has the form (just as in the case of sinusoidal coupling[3])

$$\left(\frac{\sin au}{u}\right)^2$$

which in our terminology has half-peak loss (with $\Delta\beta_*$ the variable) at

$$\Delta\beta_{*\frac{1}{2}}\,|_{\Delta\alpha=0} = \frac{1.8\pi}{z}. \tag{30}$$

For the numerical case of this paper, $z = 1000$ feet and

$$\Delta\beta_{*\frac{1}{2}}\,|_{\Delta\alpha=0} = 0.00565.$$

TABLE VI — PHASE ANGLE OF $E_1$ FOR $|\Delta\alpha/c| = 0$

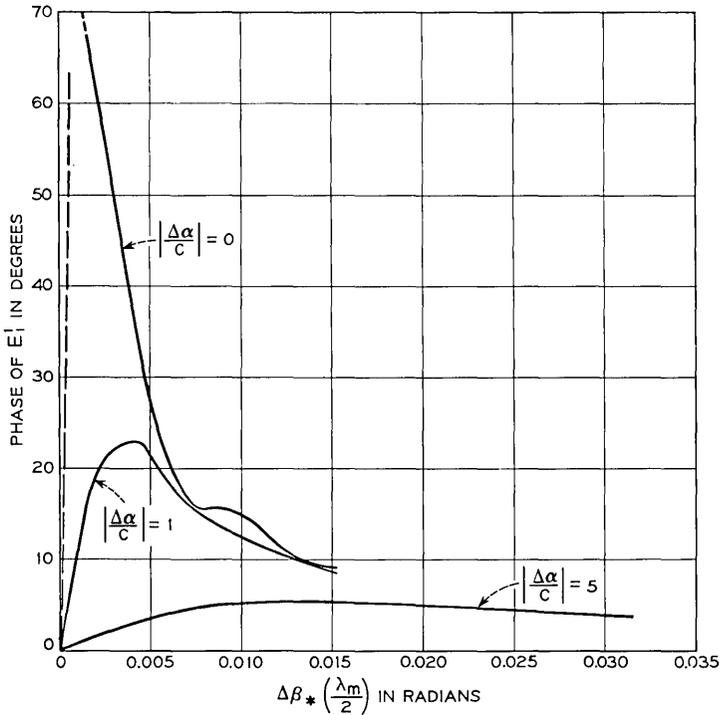| $\Delta\beta\left(\dfrac{\lambda_m}{2}\right)$ | Angle for $E_1$ | | Approximate −exact |
| --- | --- | --- | --- |
| | Approximate | Exact | |
| $\pi + 0.0015707963$ | $- 67.753°$ | $- 67.726°$ | $- 0.027°$ |
| $\pi + 0.00314159264$ | $- 47.111$ | $- 47.096$ | $- 0.015°$ |
| $\pi + 0.00471238891$ | $- 30.086$ | $- 30.092$ | $+ 0.006°$ |
| $\vdots$ | | | |
| $\pi + 0.0157079$ | $- 9.085$ | $- 9.104$ | $+ 0.019°$ |

Fig. 8 — Phase of driven wave versus $\Delta\beta \cdot \lambda_m/2$ near $\Delta\beta\lambda_m = 2\pi$ for $|\Delta\alpha/c| = 0,1,5$.

In Table VII, the calculated value 0.0064 at $|\Delta\alpha/c| = 1$ is reasonable, since D. T. Young's[4] work based on the perturbation theory indicates that the true loss peak is the *convolution* of the shape for $|\Delta\alpha/c| \gg 1$ with the shape for $|\Delta\alpha/c| = 0$.

Consider now the peak loss at $\Delta\beta_* = 0$. When $|\Delta\alpha/c| \gg 1$, and $e^{\Delta\alpha z} \ll 1$, it can be shown that equation (9) simplifies to

$$E_1| = e^{r_3 z} \tag{31}$$

$$\left|\frac{\Delta\alpha}{c}\right| \gg 1$$

$$e^{\Delta\alpha z} \ll 1$$

$$\Delta\beta_* = 0$$

where

$$r_3 = -\frac{4}{\pi^2} c \left| \frac{c}{\Delta\alpha} \right|. \tag{32}$$

In Table VIII we compare the loss computed from (32) with the actual loss, and see that even for $| \Delta\alpha/c | = 1$ the error is only $\approx 30$ percent. A consideration of the terms of equation (9) indicates that these errors would be approximately constant with increasing $z$.

A further simplification of the calculation of loss components now seems justified. For $| \Delta\alpha/c | \geqq 1$ it would appear that (31) and (32) can be used to calculate the peak loss due to a single square-wave coupling
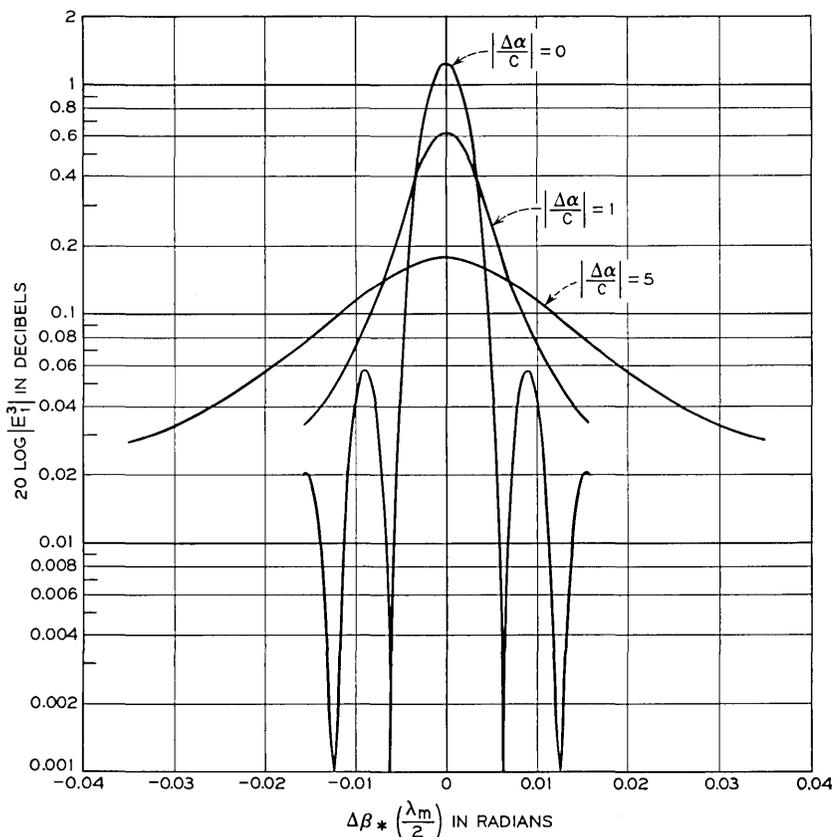


Fig. 9 — Driven-wave loss versus $\Delta\beta_*$ near $\Delta\beta\lambda_m = 6\pi$ (third harmonic) for $| \Delta\alpha/c | = 0,1,5$.
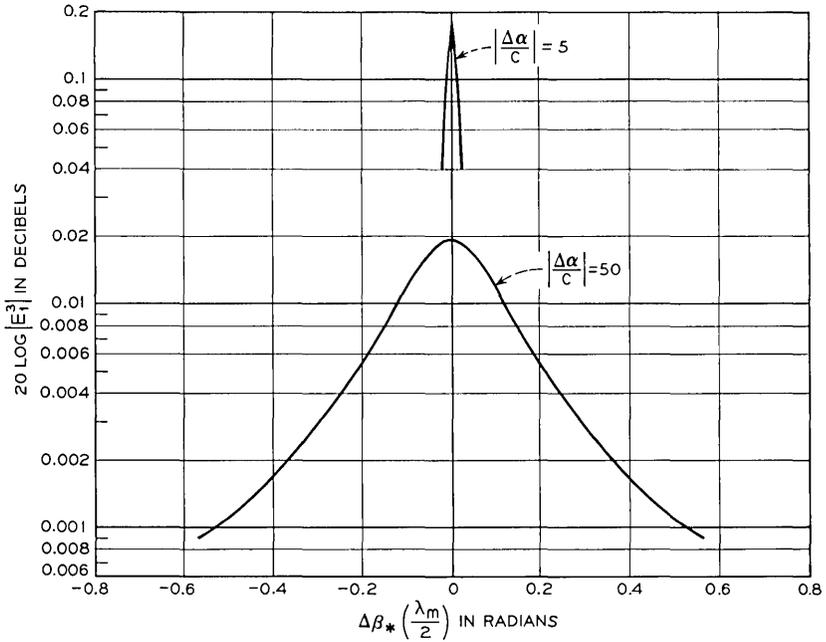
Fig. 10 — Driven-wave loss versus $\Delta\beta_*$ near $\Delta\beta\lambda_m = 6\pi$ (third harmonic) for $|\Delta\alpha/c| = 5,50$.

component. And the shape versus $\Delta\beta_*$ can be calculated from the work of D. T. Young, leading to an over-all attenuation defined by

$$|E_1| = e^{r_4 z} \tag{33}$$

where

$$r_4 = -\frac{4}{\pi^2} c \left| \frac{c}{\Delta\alpha} \right| \frac{1}{1 + \left(\dfrac{\Delta\beta_*}{\Delta\alpha}\right)^2} \tag{34}$$

in which we require $z$ and $\Delta\alpha/c$ such that $e^{\Delta\alpha z} \ll 1$, $|\Delta\alpha/c| \geqq 1$.

TABLE VII — LIMITING AND TRUE VALUES

| (1) $\left|\dfrac{\Delta\alpha}{c}\right|$ | (2) Limiting value of $\Delta\beta_{*1/2}$ from equation (25) | (3) Calculated $\Delta\beta_{*1/2}$ from Figs. 6 and 7 | Ratio of Col. (3) / Col. (2) |
|---|---|---|---|
| 1 | 0.00493 | 0.0064 | 1.3 |
| 5 | 0.0246 | 0.0265 | 1.077 |
| 50 | 0.246 | 0.250 | 1.016 |

### TABLE VIII — LOSS COMPARISON

| | $\Delta\beta_* = 0$ | |
|---|---|---|
| $\left\|\dfrac{\Delta\alpha}{c}\right\|$ | Loss (dB) from equation (32) | Actual loss (dB) equation (9) |
| 1 | − 8.68 | − 6.57 |
| 5 | − 1.74 | − 1.62 |
| 50 | − 0.1737 | − 0.1727 |

APPENDIX A

*Coupling Functions*

We use the following equations to represent two coupled waves of amplitude $E_1$ and $E_2$:

$$\frac{d}{dz} E_1(z) = -\gamma_1 E_1 + c_{21}(z)E_2 \tag{35}$$

$$\frac{d}{dz} E_2(z) = c_{12}(z)E_1 - \gamma_2 E_2 \tag{36}$$

where $\gamma_n = \alpha_n + j\beta_n$ = complex propagation constant. Let

$$c_{21}(z) = jc_p e^{jk_c z}. \tag{37}$$

Energy conservation leads to:

$$c_{12}(z) = jc_p e^{-jk_c z}. \tag{38}$$

Also let

$$E_1 = e^{-\gamma_1 z} V_1 \tag{39}$$

$$E_2 = e^{-\gamma_2 z} V_2. \tag{40}$$

Then (35) and (36) become:

$$\frac{dV_1}{dz} = jc_p V_2 e^{[\gamma_1 - \gamma_2 + jk_c]z} \tag{41}$$

$$\frac{dV_2}{dz} = jc_p V_1 e^{-[\gamma_1 - \gamma_2 + jk_c]z}. \tag{42}$$

We now see that the solution for $k_c = 0$ holds for $k_c$ nonzero provided that we make the change

$$\Delta B = \beta_1 - \beta_2 + k_c$$

and put $\Delta B$ in place of $\Delta\beta = \beta_1 - \beta_2$ in the solution for $k_c = 0$.

Complete transfer of power between waves can occur when $\Delta B = 0$ or

$$k_c = (\beta_2 - \beta_1). \tag{43}$$

APPENDIX B

*Square-Wave Coupling*

We start with equations (1) through (4) representing uniformly coupled waves with $k_c = 0$ (see Fig. 1). Equations (9) through (14) with $c_* = c$ and $\Delta\beta_* = \Delta\beta$ give the amplitude of the output waves at $z$ for the conditions $E_1 = 1.0$ and $E_2 = 0$ at $z = 0$. We apply these equations to the coupling distribution of Fig. 2 and find that the output amplitude for the undriven wave at $z = \lambda_m/2$ is:

$$E_2\big|_{z=\lambda_m/2} = \exp\left[-j(\beta_1 + \beta_2)\frac{\lambda_m}{4}\right]\left\{\frac{+j \sin\left[\dfrac{c\lambda_m\sqrt{1 + \left(\dfrac{\Delta\beta}{2c}\right)^2}}{2}\right]}{\sqrt{1 + \left(\dfrac{\Delta\beta}{2c}\right)^2}}\right\}. \tag{44}$$

In the above, the simplification $\Delta\alpha = 0$ has been assumed.

Using the output waves at $z = \lambda_m/2$ as input conditions to the following coupling region (Fig. 2), the amplitude of the undriven wave at $z = \lambda_m$ is found to be:

$$E_2\big|_{z=\lambda_m} = (-1)\frac{2\left(\dfrac{\Delta\beta}{2c}\right)}{\left[\left(\dfrac{\Delta\beta}{2c}\right)^2 + 1\right]}$$

$$\cdot \sin^2\left[\frac{c\lambda_m}{2}\sqrt{1 + \left(\frac{\Delta\beta}{2c}\right)^2}\right]\exp\left[-j\frac{(\beta_1 + \beta_2)}{2}\lambda_m\right]. \tag{45}$$

The ratio of the undriven wave amplitude at $z = \lambda_m$ to the undriven wave amplitude at $z = \lambda_m/2$ is [from (44) and (45)]

$$\frac{E_2\ |_{\lambda_m}}{E_2\ |_{\lambda_{m/2}}} = j\ \frac{2\left(\dfrac{\Delta\beta}{2c}\right)}{\sqrt{1 + \left(\dfrac{\Delta\beta}{2c}\right)^2}}$$

$$\cdot \sin\left[\frac{c\lambda_m}{2}\ \sqrt{1 + \left(\frac{\Delta\beta}{2c}\right)^2}\right] \exp\left[-j\ \frac{(\beta_1 + \beta_2)}{4}\ \lambda_m\right]. \tag{46}$$

By imposing the condition

$$\frac{c\lambda_m}{2}\ \sqrt{1 + \left(\frac{\Delta\beta}{2c}\right)^2} = \frac{\pi}{2} \tag{47}$$

we notice that the magnitude of (46) approaches two for $\Delta\beta \gg 2c$, that is, for very small coupling. Fig. 4 sketches the undriven wave amplitude as a function of $z$ during the first two coupling intervals. We will now express the approximate or average coupling between the waves by using the linear approximation $E_2(z) = 2cz/\pi$. We notice that uniform coupling without phase reversal would have resulted in the relation $E_2(z) = cz$. We therefore arrive at the transformation

$c$ (constant coupling) becomes

$2c/\pi$ (for periodically reversed coupling).

$$\tag{48}$$

The associated transformation of the condition for maxmium energy transfer, from (47), is

$$\Delta\beta\ |_{\text{maximum conversion}} = \frac{2\pi}{\lambda_m}\ \sqrt{1 - \left(\frac{c\lambda_m}{\pi}\right)^2}. \tag{49}$$

We might notice here that the departure of the actual amplitude in Fig. 4 from the straight-line approximation, shown by $\epsilon$ in Fig. 4, has a maximum value which can be shown to be $0.21(c\lambda_m)/\pi$. Thus the deviation between our straight-line approximation and the actual amplitude becomes smaller for diminished values of coupling per unit $\lambda_m$ .

We now specify $c_*$ and $\Delta\beta_*$ in equations (9) through (14) to represent the waves with periodically reversed coupling using the average coupling approach. Since the in-phase build-up of power in the undriven wave is a maximum for $\Delta\beta$ specified by equation (49), we define a new differential phase parameter to give a departure from this condition:

$$\Delta\beta_* = \Delta\beta - \frac{2\pi}{\lambda_m}\ \sqrt{1 - \left(\frac{c\lambda_m}{\pi}\right)^2}. \tag{50}$$

By comparison with equation (7), which represents an exact solution for complex coupling, this definition of $\Delta\beta_*$ seems reasonable. Numerical checks reported in the text verify this presumption. The value of $c_*$ is given by (48).

It has already been noted, near equation (46) that two successive $\lambda_m/2$ coupling intervals give twice the undriven line amplitude compared with the first $\lambda_m/2$ interval, which is just what occurs for small constant coupling (no reversal) and $\Delta\beta = 0$. We also find that $\Delta\beta_*$ and $c_*$ correctly give the wave amplitudes for periodically reversed coupling at $z$ equal to integral multiples of $\lambda_m/2$, even when $\Delta\beta$ is not large compared with $2c$. For example, letting $E_2$ at $z = \lambda_m/2$ be 0.707 requires $(\Delta\beta/2c) = 1$ (see equation 44); we maintain equation (47) and (48) as before. Then, from (45), $| E_2 |$ becomes unity at $z = \lambda_m$ which is also predicted by $\Delta\beta_*$ and $c_*$ in (9) through (14) and which is analogous to the behavior of two conventional 3 dB directional couplers in cascade.

The above discussion represents the changes in wave propagation introduced by coupling for $\Delta\beta$ in the vicinity of the value given by equation (49).

The perturbation solution[3] for *sinusoidal* periodic coupling is known to yield coupling absorption peaks when

$$\Delta\beta\lambda_m = 2\pi p$$

$$p = 1, 3, 5 \cdots .$$

(51)

Similarly, there are other regions of strong interaction for the square-wave coupling of Fig. 2. For example, consider Fig. 5, which represents the situation when

$$\frac{kc\lambda_m}{2} = \frac{3\pi}{2}.$$

(52)

There is another region of $\Delta\beta$ defined by

$$\Delta\beta \mid_{\text{maximum conversion}} = 3\cdot\frac{2\pi}{\lambda_m} \sqrt{1 - \left(\frac{c\lambda_m}{3\pi}\right)^2}$$

(53)

where there is a local maximum of conversion. As diagramed in Fig. 5, the average conversion coefficient is $2c/3\pi$. Thus the appropriate values of $\Delta\beta_*$ and $c_*$ for equations (9) through (14) are

$$c_* = \frac{2c}{3\pi}$$

$$\Delta\beta_* = \Delta\beta - 3\cdot\frac{2\pi}{\lambda_m} \sqrt{1 - \left(\frac{c\lambda_m}{3\pi}\right)^2}.$$

(54)

More generally there are absorption peaks at

$$\Delta\beta_* = \Delta\beta - p\,\frac{2\pi}{\lambda_m}\sqrt{1 - \left(\frac{c\lambda_m}{p\pi}\right)^2} \tag{55}$$

$$c_* = \frac{2c}{p\pi} \tag{56}$$

where $p = 1, 3, 5, \ldots$

APPENDIX C

*Sine Wave Coupling*

We start with equations (1) and (2) with the coupling defined

$$c_{21} = c_{12} = jc\sin\left(\frac{2\pi z}{\lambda_m}\right). \tag{57}$$

Using perturbation theory[3], and letting $\Delta\alpha = 0$,

$$\vec{E_2} = c\int_0^z \sin\left(\frac{2\pi s}{\lambda_m}\right)e^{-i\Delta\beta s}\,ds \tag{58}$$

which yields

$$E_2 = \frac{c}{\left[\left(\frac{2\pi}{\lambda_m}\right)^2 - \Delta\beta^2\right]}\left\{\frac{2\pi}{\lambda_m} - e^{-i\Delta\beta z}\left[\frac{2\pi}{\lambda_m}\cos\left(\frac{2\pi z}{\lambda_m}\right) + j\Delta\beta\sin\left(\frac{2\pi z}{\lambda_m}\right)\right]\right\}. \tag{59}$$

Evaluating at $\Delta\beta = 2\pi/\lambda_m$ and $z = n\lambda_m/2$ yields

$$E_2\,|_{\Delta\beta=2\pi/\lambda_m} = -j\frac{c}{2}\frac{n\lambda_m}{2} \tag{60}$$

with $n = 1, 2, 3. \ldots$
It can be verified that $\Delta\beta = 2\pi/\lambda_m$ yields the maximum value of $E_2$ at $z = n\lambda_m/2$.

Thus the equivalent uniform coupling value for sinusoidal coupling is

$$c_* = c/2 \tag{61}$$

which appears in Table I.

REFERENCES

1. Pierce, J. R., "Coupling of Modes of Propagation," J. Appl. Phys., *25* (February 1954), pp. 179–183.
2. Rowe, H. E. and Warters, W. D., "Transmission Deviations in Waveguide due to Mode Conversion: Theory and Experiment," presented at the Conference on Long Distance Transmission by Waveguide, January 29, 1959; Proc. IEE, *106*, Part B, Supplement 13 (1959).
3. Rowe, H. E. and Warters, W. D., "Transmission in Multimode Waveguide with Random Imperfections," B.S.T.J., *41*, No. 3 (May 1962), pp. 1031–1170.
4. Young, D. T., "Effect of Differential Loss on Approximate Solutions to the Coupled Line Equations," B.S.T.J., *42*, No. 6 (November 1963), pp. 2741–2794.
5. Rowe, H. E., "Approximate Solution for the Coupled Line Equations," B.S.T.J., *41*, No. 3 (May 1962), pp. 1001–1029.
6. Miller, S. E., "Coupled Wave Theory and Waveguide Applications," B.S.T.J., *33*, No. 3 (May 1954), pp. 661–719.

# Contributors to This Issue

ROBERT W. CHANG, B.S.E.E., 1955, National Taiwan University; M.S.E.E., 1960, North Carolina State College; Ph.D., 1965, Purdue University; Bendix Corporation, 1960-1963; Bell Telephone Laboratories, 1965—. Mr. Chang has been concerned with problems in data transmission and communication theory. Member, Eta Kappa Nu, Sigma Xi, Phi Kappa Phi, IEEE.

I. DANYLCHUK, B.E.E., 1956, Rensselaer Polytechnic Institute; M.E.E., 1965, New York University; Bell Telephone Laboratories 1956—. He is a member of Solid State Device Laboratory and has been engaged in exploratory development of magnetic memories. Member, Tau Beta Pi, Eta Kappa Nu, IEEE.

CHARLES R. ELLISON, JR. BSEE, 1960, Villanova University; MSEE, 1964, Stevens Institute of Technology; Bell Telephone Laboratories, 1964—. He has been collecting and analyzing transmission performance data on Bell System trunks and carrier facilities with emphasis on those parameters that significantly influence the quality of data transmission. He has served as a member of the United States of America Standards Institute sponsored industrial committee to review the engineering installation standards for data transmission in the defense communication system. He is involved in evaluating the quality and reliability of data transmission facilities on the NASA private line network. Member Tau Beta Pi.

STANLEY L. FREENY, BEE, 1958, Georgia Institute of Technology; MEE, 1960, New York University; Bell Telephone Laboratories, 1958—. Mr. Freeny has done work on a variety of problems relating to pulse transmission. Member, IEEE, Tau Beta Pi, Eta Kappa Nu.

CHARLES A. FRITSCH, B.M.E., 1958, University of Dayton; M.S.M.E., 1960, Ph.D., 1962, Purdue University; Bell Telephone Laboratories, 1961—. Dr. Fritsch has worked on problems in the thermal sciences associated with hardening structures to withstand nuclear weapon effects, cooling electronic equipment, and developing of gas lenses. He is

Supervisor of the Fluid Mechanics and Heat Transfer Group of the Engineering Mechanics and Physics Department. Member, A.S.M.E., A.P.S., Sigma Xi.

U. F. GIANOLA, B.Sc., 1948, and Ph.D., 1951 (physics), University of Birmingham, England; postdoctorate research fellowship, University of British Columbia, 1951–53; Bell Telephone Laboratories, 1953—. He heads the Fundamental Memory Components Department and is responsible for the exploratory development of digital memory and logic devices. Senior member, IEEE; member, American Physical Society, Research Society of America.

ROBERT HOLMSTROM, B.S.E.E., 1962, Wayne State University; M.E.E., 1964, New York University; Bell Telephone Laboratories, 1962—. Mr. Holmstrom has been concerned with surveys to gather data on the transmission performance of Bell System toll facilities. He has been involved in both planning surveys and analyzing data gathered. Member, Tau Beta Pi, Eta Kappa Nu.

DEAN W. LYTLE, B.S. and M.S., University of California; Ph.D., 1957, Stanford University; Bell Telephone Laboratories, summers of 1966 and 1967. While with Bell Telephone Laboratories, he was a consultant, working on problems relating to automatic equalization of high-speed, hybrid, multilevel PCN systems. He is now Associate Professor of Electrical Engineering at the University of Washington where he teaches signal theory. Member, IEEE, Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.

JAMES E. MAZO, B.S., 1958, Massachusetts Institute of Technology; M.S., 1960, and Ph.D., 1963, Syracuse University; Research Associate, University of Indiana, 1963–64; Bell Telephone Laboratories, 1964—. Mr. Mazo was engaged in work on quantum scattering theory at Indiana University. Now he is doing theoretical analysis of data systems. Member, American Physical Society, IEEE, Sigma Xi.

RICHARD H. McCULLOUGH, S.B.E.E. and S.M.E.E., 1960, Massachusetts Institute of Technology; Ph.D.E.E., 1967, Polytechnic Institute of Brooklyn; Bell Telephone Laboratories, 1960—. At Bell Laboratories, Mr. McCullough worked on circuit and system design of analog and digital transmission systems, including TL radio, *Tel-*

*star*® communications satellite, and experimental pulse code modulation systems. Since 1965 he has been concerned with scientific and systems programming, and algorithms for automatic processing of sonar signals. Member, Sigma Xi, Tau Beta Pi, Eta Kappa Nu, IEEE.

STEWART E. MILLER, B.S. and M.S., 1941, Massachusetts Institute of Technology; Bell Telephone Laboratories, 1941—. Mr. Miller first worked on coaxial carrier repeaters and then on 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 circular electric waveguide communication, microwave ferrite devices, and other components for microwave radio systems. As Director, Guided Wave Research Laboratory, he heads a group engaged in research on communication techniques for the millimeter wave and optical regions. Fellow, IEEE; member, Eta Kappa Nu, Sigma Xi, Tau Beta Pi.

INGEMAR NåSELL, Civilingenjör, 1955, Royal Institute of Technology, Stockholm, Sweden; M.E.E., 1962, and M.S. (mathematics), 1965, both from New York University; Research Institute of National Defense, Stockholm, Sweden, 1955–1960; Bell Telephone Laboratories, 1960—. Mr. Nåsell is concerned with characterizing the transmission performance of the Bell System toll network for systems engineering purposes. He is Supervisor of the Transmission Surveys and Models Group. Member, Svenska Teknologföreningen, American Statistical Association, Eta Kappa Nu.

CLYDE L. RUTHROFF, B.S.E.E., 1950, and M.A., 1952, University of Nebraska; Bell Telephone Laboratories, 1952—. Mr. Ruthroff has published contributions on the subjects of FM distortion theory, broadband transformers, FM limiters, threshold extension by feedback, and microwave radio systems for satellite and terrestrial use. He is interested in the extension of radio communication into the millimeter and optical wavelengths. Member, A.A.A.S., I.E.E.E., Sigma Xi.

JOHN T. SIBILIA, B.S. (Physics), 1955, M.A. 1957, and Ph.D., 1962, Princeton University; Bell Telephone Laboratories, 1960—. He has been engaged in maser development, exploratory magnetic memory

development, and is now supervising a group concerned with tantalum thin film circuit process instrumentation.

ALAN N. WILLSON, JR., B.E.E., 1961, Georgia Institute of Technology; M.S.E.E., 1965, Ph.D., 1967, Syracuse University; International Business Machines Corporation, 1961–64; Bell Telephone Laboratories, 1967—. Mr. Willson is interested in network and systems theory. Member, IEEE, Eta Kappa Nu, Tau Beta Pi, Sigma Xi.

H. ZUCKER, Dipl.-Ing. 1950, Technische Hochschule, Munich, Germany; M.S.E.E., 1954, Ph.D., 1959, Illinois Institute of Technology; Bell Telephone Laboratories, 1964—. He has been concerned with problems related to satellite communication antennas and optical resonators. Member, IEEE, Eta Kappa Nu, Sigma Xi.