THE BELL SYSTEM

Technical Journal

Volume 52 February 1973 Number 2

THE BELL SYSTEM TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL is published ten times a year by the American Telephone and Telegraph Company, J. D. deButts, Chairman and Chief Executive Officer, R. D. Lilley, President, J. J. Scanlon, Vice President and Treasurer, R. W. Ehrlich, Secretary. Checks for subscriptions should be made payable to American Telephone and Telegraph Company and should be addressed to the Treasury Department, Room 2312C, 195 Broadway, New York, N. Y. 10007. Subscriptions $10.00 per year; single copies $1.25 each. Foreign postage $1.00 per year; 15 cents per copy. Printed in U.S.A.

THE BELL SYSTEM

TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 52 February 1973 Number 2

A Fundamental Comparison of Incomplete Charge Transfer in Charge Transfer Devices

By C. N. BERGLUND* and K. K. THORNBER

(Manuscript received September 22, 1972)

Using a small-signal analysis, we present a general comparison of the more important contributions to the incomplete transfer of charge in the following charge transfer devices: (i) the polyphase charge-coupled device and the IGFET bucket-brigade shift register, where individual charge transfers are single-step processes, and (ii) the two-phase charge-coupled device, the conductively connected charge-coupled device, the tetrode bucket brigade, and the stepped-oxide bucket brigade, where individual charge transfers are two-step processes. A recently proposed lumped-charge-model approximation is made in order to estimate the time dependence of the transferred charge including both drift and diffusion. In this calculation we also include the effects associated with the injection of charge from a diffused source into the IGFET channel which modify the current-voltage behavior at low currents. Using this calculation of the time dependence of the transferred charge, the various contributions to incomplete transfer, including those due to trapping in the interface states, are derived and compared for each of the charge transfer devices of interest. The results

* Present address: Bell Northern Research, P. O. Box 3511, Station C, Ottawa, Ontario, Canada. (Mr. Berglund's contribution was made to this paper while he was employed by Bell Laboratories.)

show that highest transfer efficiencies at low frequencies will usually be obtained from suitably designed shift resisters utilizing a two-step transfer process. This is because a two-step process tends to reduce those contributions to incomplete transfer which are due to signal-charge-modulation of the conductances and capacitances governing the charge transfer current. For example, the coefficient of incomplete charge transfer, α, for an individual transfer in a polyphase CCD is found to be 0.6, 0.85, and $1.4 \cdot 10^{-3}$ at 2, 4, and 8 MHz, respectively. By contrast, for the same physical device dimensions, α in a stepped-oxide, two-phase CCD is reduced to 0.35, 0.5, and $0.8 \cdot 10^{-3}$ at the same frequencies, nearly a factor of two reduction. The intrinsic contributions to incomplete charge transfer, which dominate the high-frequency operation, are found to be relatively insensitive to the type of device considered for the dimensions given. An α of $0.3 \cdot 10^{-1}$ at 100 MHz is typical. Among the various two-step process devices, interface-state effects are found to be comparable, $0.3 \cdot 10^{-3}$ being a typical value for this frequency-independent contribution. At low frequencies, a frequency-independent contribution to incomplete transfer in bucket-brigade transfer processes is found. The size of this contribution is about $0.15 \cdot 10^{-3}$. These results further illustrate the utility of analyzing incomplete transfer effects in terms of single-device, small-signal characteristics.

I. INTRODUCTION

Since the introduction of the charge-coupled device[1] (CCD) and the bucket-brigade shift registers,[2–4] there have been a number of proposed modifications to the basic structures in order to achieve specific performance goals.[5–10] It is the purpose of this paper to analyze the mechanisms of charge transfer in these devices and to compare the characteristics of the incomplete charge transfer in each. In this way one can evaluate the relative merits and limitations of these modified charge transfer devices (CTD's).

In a recent article, the authors[11] presented a general analysis of incomplete charge transfer in CTD's. Using a lumped model[12,13] to characterize the dynamics of the charge transfer, it was possible to calculate α, the small-signal coefficient of incomplete transfer, in terms of single-device, small-signal characteristics. It was found that three contributions to incomplete transfer are common to all charge transfer shift registers: an intrinsic transfer rate contribution, an output conductance or feedback contribution, and a storage-capacitance modulation contribution. A significant feature of that calculation was that each of these contributions could be expressed analytically in terms of

the charge transferred as a function of time. Analytic results are, of course, very convenient for comparative purposes. In addition, numerical values for α could be obtained, the accuracy of which depends only on the accuracy to which the transferred charge is known, and *not* on the much lower accuracy obtainable from the *difference* between the transferred charge for slightly different signal sizes. Although originally developed to treat the standard CCD and bucket-brigade[2–4] devices, it is possible by a straightforward generalization to analyze the more complicated device structures considered here with little additional effort.

In this paper we derive in a uniform manner the incomplete transfer properties[11] of the simple bucket brigade, the three-phase CCD, various two-phase CCD's, the conductively connected CCD, and the tetrode and stepped-oxide bucket brigades. These calculations are then used as a basis for comparison of the performance to be expected from each CTD. It will also become clear that other CTD's can be treated in a similar fashion.

Using a lumped-charge model for CTD behavior[13] based on a charge-control concept proposed by Lee and Heller,[12] we first calculate the time-dependence of the transferred charge in a CTD for both a bucket-brigade cell and a CCD cell. This derivation of necessity is somewhat approximate, but it is found to be satisfactory for the purpose of calculating the incomplete transfer properties by our method and is comparable with the other approximations necessary for any such calculation. Being simple, it emphasizes the important characteristics of the transfer event. From the time dependence of the charge transfer we determine the small-signal transfer inefficiency[11] and calculate the clock-frequency dependence of the several contributions to incomplete transfer. Finally, these theoretical results are applied to several specific CTD's and their important operating limitations are discussed.

II. TIME DEPENDENCE OF TRANSFERRED CHARGE

2.1 *Preliminary Considerations*

The starting point in determining the performance characteristics of a charge transfer device is a calculation of the time dependence of the charge transfer process. However, the nonlinearity of the problem, the requirement that several assumptions be made concerning the charge transport and trapping along a semiconductor interface, and the dependence of the solution on clock-voltage waveform and on the details of device design, all combine to make a general solution valid

for all cases an unrealistic goal. Indeed for our purposes here a highly accurate solution is not necessary. We shall, therefore, make assumptions to simplify the problem considerably.

2.2 *General Approach*

Our first assumptions are that the charge transfer in the more complicated modified structures can be treated as a series of simpler transfer steps, and that each charge-transfer step can be characterized by either a "CCD" transfer process or a "bucket-brigade" transfer process. The major difference between the two is that in the "bucket-brigade" transfer process the injection of carriers from a diffused region into a surface channel is taken into account. In the "CCD" transfer process, by contrast, the charge carriers are initially in the surface "channel" (storage well) and no injection over a barrier is necessary. We shall solve for the charge transfer as a function of time for both types of transfer processes. Such solutions will be sufficient to discuss CTD's in which the charge transfer from one storage region to the next are single-step processes. For CTD's with more complicated charge transfer mechanisms involving multiple-step processes, the charge transfer is represented as a series of single-step processes of either the "CCD" or the "bucket-brigade" type as is appropriate. In this way, sufficiently accurate results will be achieved to illustrate most of the important features of the charge transfer and to provide a basis for comparison of the several CTD schemes.

In order to carry out the above program, we must calculate the charge transfer in a "CCD" and in a "bucket-brigade" transfer process. The general approach that will be used will be the charge-control or lumped-charge model. This type of model has usually been used for time-dependent analyses of circuit problems[14–16] as well as for analysis of some CTD's such as the IGFET[17] and bipolar[18] bucket-brigade shift registers. By comparison of results from such a model with exact computer solutions,[19–21] it has been shown that the model is also applicable with surprising accuracy to the charge-coupled device.[12] We shall assume that the gradual channel approximation is valid for all the cases to be treated here. This means that the one-dimensional solution to Poisson's equation will be used throughout and that fringing electric fields will influence only a small fraction of the active device area, and will, by definition, mean that our results will not be applicable to CTD's such as the buried channel CCD,[22] except possibly in the high-frequency limit. For simplicity, it will also be assumed that we are dealing with p-channel devices with a constant carrier mobility, that depletion-

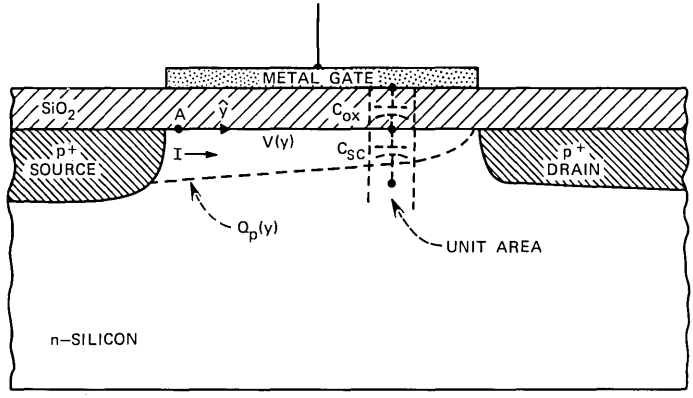

Fig. 1—IGFET gate region illustrating the symbols defined in the text.

layer capacitances are constant, voltage-independent quantities, and that the clock voltage driving the CTD's are ideal square waves.

We begin by calculating the steady-state transport under given boundary conditions for an IGFET gate. This is similar to the conventional IGFET solution except that we include carrier diffusion in addition to drift. The current-voltage relation we obtain will then be used in calculating the "CCD" and "bucket-brigade" transfer processes.

Referring to Fig. 1 and following Sze,[23] the minority carrier charge density $Q_p(y)$ is given by

$$Q_p(y) = C_i V(y) \tag{1}$$

where C_i is the capacitance per unit area of the oxide C_{ox} in parallel with the silicon space charge C_{SC}, and $V(y)$ is the silicon surface potential with reference to the surface potential value in the absence of minority carriers. A lateral electric field can only exist in the gradual channel approximation because of a gradient in Q_p, and will be represented as $-(1/C_i)(dQ_p/dy)$. Given that the current flowing at any value of y must be a constant I, and that current density J is given by

$$J = q\mu pE - qD\frac{dp}{dy} \tag{2}$$

where μ and D are the carrier mobility and diffusion constants respectively (related by the Einstein relationship), p is the hole density, and E is the electric field, one can write for the current flowing from

the source to the drain:

$$I = -\frac{Z\mu}{C_i}\left[Q_p(y) + C_i\frac{kT}{q}\right]\frac{dQ_p(y)}{dy} \tag{3}$$

where Z is the channel width. Placing the boundary condition that the surface potential V on the "drain" side of the channel is approximately zero (i.e., assume the drain voltage is sufficiently negative that we are operating in saturation) and on the "source" side is assumed equal to a voltage V_A, one obtains

$$I = \frac{\beta}{2}V_A\left(V_A + 2\frac{kT}{q}\right) \tag{4}$$

where $\beta = Z\mu C_i/L_c$ is the conventional IGFET gain factor, and L_c is the effective channel length. This expression differs from the usual saturation IGFET current expression (which ignores diffusion) only by the additive term $2kT/q$ in the last factor. Similarly we can calculate the total charge Q stored under the gate for a given voltage V_A. Representing the geometrical capacitance of the gate oxide in parallel with the underlying silicon space-charge capacitance as C_{GO},

$$Q = \tfrac{2}{3}C_{GO}V_A\left[\frac{V_A + \frac{3}{2}\frac{kT}{q}}{V_A + 2\frac{kT}{q}}\right]. \tag{5}$$

Hence, for V_A very large or very small with respect to kT/q, Q is linear in the voltage V_A. Thus an effective gate capacitance C_G can be defined which is $\frac{2}{3}$ or $\frac{1}{2}$ respectively of the geometrical capacitance. For our purposes here, this is sufficiently close to a constant value that we can assume C_G is equal to $\frac{2}{3}C_{GO}$ over the entire range of V_A of interest.

2.3 *The "CCD" Transfer Process*

In the CCD problem treated by several authors,[12,19–21,24] the transient decay of the charge stored under a single capacitor plate is calculated. With the ideal square-wave clocks assumed here, this problem is illustrated in Fig. 2. Given the charge-control or lumped-charged model assumption, the problem reduces to one of discharging a capacitor through a nonlinear resistor derived from eq. (4). In the lumped-charge model as shown above, the total charge under the plate is approximately linear in the voltage V_A (see Fig. 2). Hence, one can

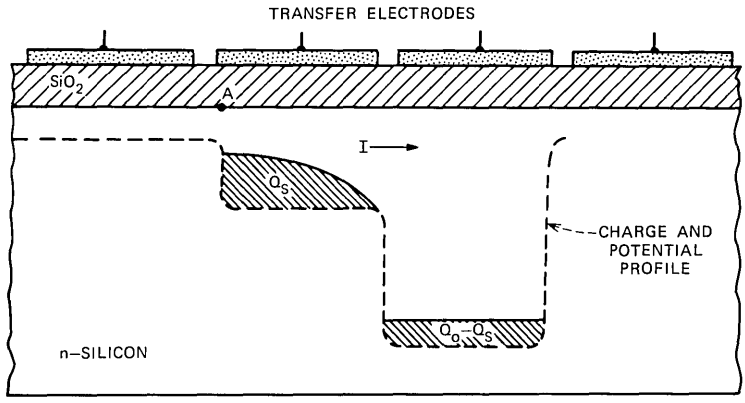

Fig. 2—Transfer of charge in a standard CCD cell with symbols defined in the text.

write from eq. (4)

$$\frac{dQ_S}{dt} = C_G \frac{dV_A}{dt} = -I = -\frac{\beta}{2} V_A \left(V_A + 2\frac{kT}{q} \right) \tag{6}$$

(L_c in β now refers to the effective length of the storage well of the CCD cell). If the initial voltage at A at the beginning of transfer was

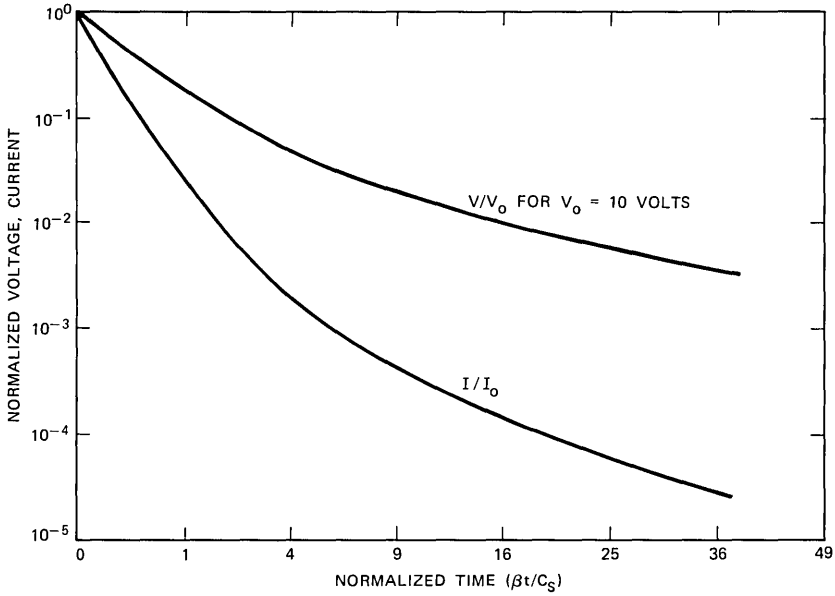

Fig. 3—Normalized voltage and current plotted as a function of normalized time for the transfer of charge from one well to the next in a standard CCD.

V_o and the voltage at time t is V, then eq. (6) gives the solution

$$\frac{V}{V_o} = \frac{\exp\left[-\frac{\beta kT}{qC_G}t\right]}{1 + \frac{qV_o}{2kT}\left[1 - \exp\left(-\frac{\beta kT}{qC_G}t\right)\right]}. \tag{7}$$

Except for the fact that V and V_o are voltages at point A rather than some appropriately defined average voltage under the plate, this result is very similar though not identical to one derived by Lee and Heller[12] and is plotted in Fig. 3 assuming a 10-volt value for V_o. Comparison of eq. (7) with the computer results of Strain and Schryer[19] also gives excellent agreement provided that the voltage V in eq. (7) is scaled to reflect the average voltage as defined by Strain and Schryer.[19]

It will be shown later that in calculating the incomplete transfer properties of CTD's the current I at time t compared to the current I_o at the initial part of the transfer is of importance. From eq. (6)

$$\frac{I}{I_o} = \frac{V}{V_o}\left(\frac{V + \frac{2kT}{q}}{V_o + \frac{2kT}{q}}\right) \tag{8}$$

and this is also plotted in Fig. 3.

2.4 *The "Bucket-Brigade" Transfer Process*

The bucket-brigade transfer problem is very similar to that of the CCD in the sense that the current during transfer is given by eq. (4) with the voltage V_A being measured at point A in the channel near the source (see Fig. 4). The important difference which requires us to treat the transfer process separately is that in the bucket brigade the minority carriers must be injected into the channel from the diffused p-regions, i.e., a large fraction of the charge is stored in the diffused regions and the p-island voltage is different in general from V_A. Representing the voltage on the source p-region as V_p, the storage capacitance associated with the p-island as C_p, and that associated with the channel as C_G, we can write analogous to eq. (6)

$$C_G\frac{dV_A}{dt} + C_p\frac{dV_p}{dt} = -\frac{\beta}{2}V_A\left(V_A + 2\frac{kT}{q}\right). \tag{9}$$

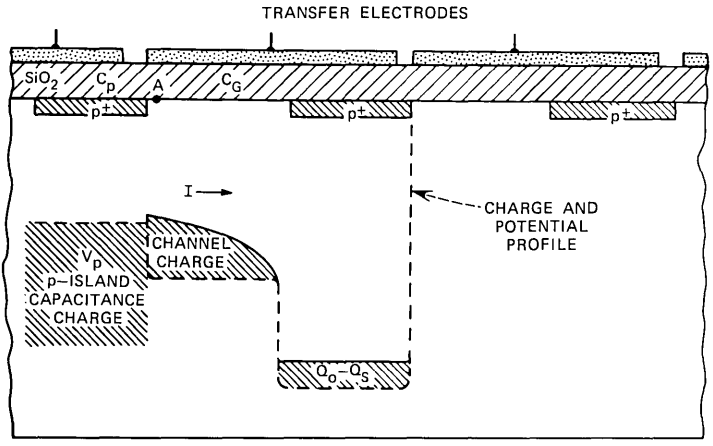

Fig. 4—Transfer of charge in a standard IGFET bucket-brigade cell with symbols defined in the text.

In order to solve this equation, a relation between the p-island voltage and V_A must be obtained. This problem has been treated somewhat differently by Barron.[25] For our purposes here, we shall simply assume that the minority carrier density at point A is related to the source p-island voltage by an expression of the form

$$C_i V_A = Q_p = Q_{po} \exp[q(V_p - V_A)/kT] \tag{10}$$

where Q_{po} is some constant [see eq. (1)]. Taking derivatives with respect to time gives

$$\frac{dV_p}{dt} = \left(1 + \frac{kT}{qV_A}\right)\frac{dV_A}{dt} \tag{11}$$

and substituting in eq. (9) we obtain

$$\frac{dV_A}{dt} = -\frac{\beta}{2C_p} V_A^2 \left\{ \frac{V_A + 2\dfrac{kT}{q}}{V_A\left(1 + \dfrac{C_G}{C_p}\right) + \dfrac{kT}{q}} \right\}. \tag{12}$$

This expression can be solved exactly for V_A as a function of time, and such a calculation has been carried out assuming C_G negligible compared to C_p and is plotted in Fig. 5. In addition to this, however, it is of most value to note that eq. (12) reduces to the CCD transfer

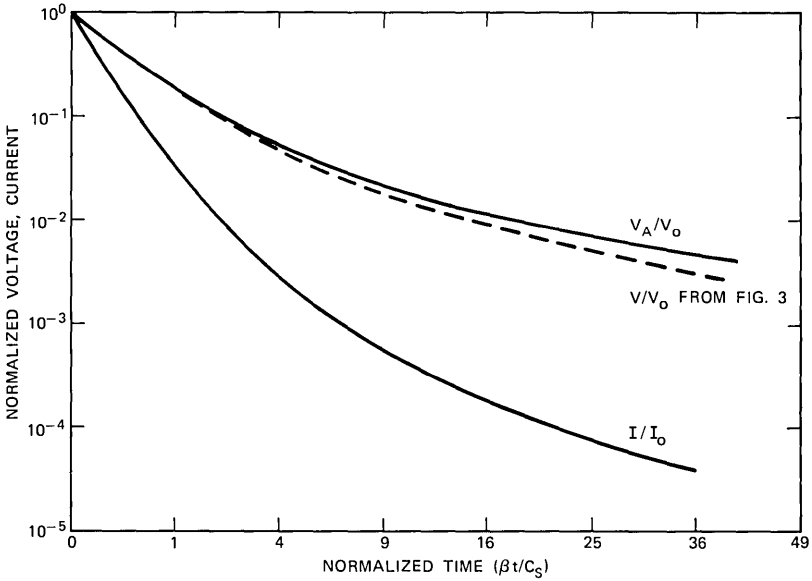

Fig. 5—Normalized voltage and current plotted as a function of normalized time for the transfer of charge from one p-region to the next in a standard IGFET bucket-brigade cell. V/V_o from Fig. 3 is shown for comparison.

result in the limit of C_p equal to zero, so that in that sense eq. (12) is a general expression for single-step charge transfer devices. Also, the last factor in eq. (12) for C_p large compared to C_G is a very slowly varying function with value on the order of one or two depending on V_A. If we assume this quantity to be unity, then the solution to eq. (12) is the previously[4] derived bucket-brigade result assuming an ideal square-law current-voltage relation

$$\frac{V_A}{V_o} = \frac{1}{1 + \dfrac{\beta V_o t}{2C_p}}. \tag{13}$$

Hence we see that for the bucket-brigade transfer process under square-wave clocks, the voltage V_A varies as reciprocal time after long times rather than exponentially as it does for the CCD transfer process. This is also illustrated by the exact solution shown in Fig. 5. Note from eq. (10) that V_p can be related to V_A through

$$V_p - V_{po} = V_A + \frac{kT}{q} \ln \frac{V_A}{V_o} \tag{14}$$

where V_{po} is a constant. Thus V_p follows V_A but differs by a term logarithmic in V_A. Since V_A varies reciprocally with time at long times, V_p ultimately has a logarithmic dependence on time as reported by Buss and Gosney.[26]

For comparison to the CCD transfer process, we have plotted V_A from Fig. 3 in Fig. 5. Also of interest is the time dependence of the transfer current, given by eq. (8) and plotted in Fig. 5.

III. TRANSFER EFFICIENCY IN CHARGE TRANSFER DEVICES

Given that one has calculated the charge left behind as a function of transfer time as in the previous section, the problem of using these results to predict the performance of an n-stage shift register remains. Under most shift register operating conditions, a circulating charge or fat zero will be used, so that a large fraction of the charge left behind after each transfer will be independent of the signal charge. Thus, a definition of transfer inefficiency as the total charge remaining divided by the total charge to be transferred will often be of little value even if we had a highly accurate solution including the net amount of charge trapped in interface states. Further, the amount of degradation in a shift register will depend on the time dependence of the particular signal being used. Since, in addition, the charge transfer process is generally nonlinear, it has been found convenient to define a differential or small-signal parameter to characterize incomplete charge transfer in CTD's; i.e., superimpose a small-signal charge on a larger background charge being continually transferred. It has been shown[27,28] that by linearizing the problem in this way, the signal degradation by an n-stage register can be readily characterized. For this reason we will focus our attention here on the small-signal incomplete transfer parameter α defined as the change in the charge left behind after transfer divided by the small change in signal charge, which was the cause of the change in charge left behind. Using derivatives

$$\alpha = \frac{dQ}{dQ_o} \tag{15}$$

where Q_o is the total charge to be transferred. ($\alpha = \alpha(t)$ since $Q = Q(t)$.) In a recent paper[11] the authors have shown that incomplete transfer in CTD's can be treated in a general way using this small-signal approach, and that α can be separated into three components: an intrinsic transfer rate term, a drain conductance or feedback term, and a storage-capacitance modulation term. Interface states introduce additional terms, but of a similar nature to these three.

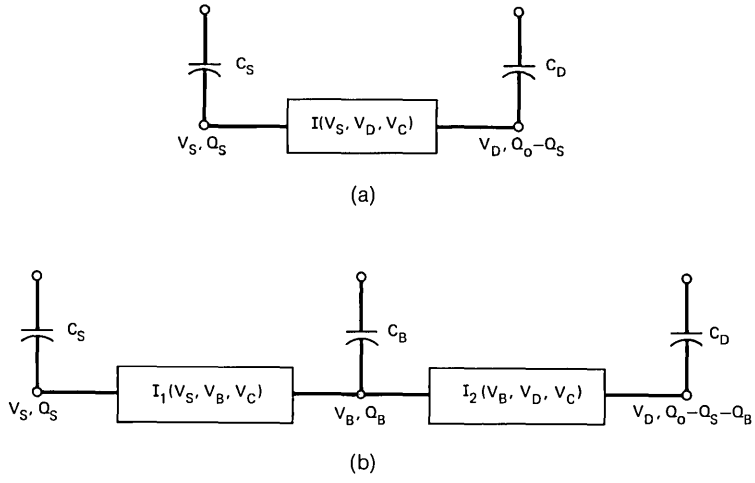

Fig. 6—(a) General representation of a single-step transfer cell of a charge transfer device. (b) General representation of a two-step transfer cell of a charge transfer device.

In this previous paper[11] we described the charge transfer process in a CTD in terms of the model shown in Fig. 6a. Although applicable for treating two-step processes, this model is really best suited for single-step processes. This being the case, it is convenient to treat a two-step process using the model shown in Fig. 6b, which is basically two single-step models in series. Even more complicated charge transfers can be treated by including additional single-step models.

The single-step process is illustrated in Fig. 6a using the same symbols as used previously. The charge from a storage capacitor C_S is transferred through some nonlinear conductance to the drain capacitor C_D as in the simple bucket brigade or the three-phase CCD. (V_C is the clock voltage.) However, in the two-step process illustrated in Fig. 6b, the charge from the storage capacitor C_S is transferred to some intermediate capacitor C_B then to the drain capacitor C_D during a single transfer period. The advantage of this is that channel-length modulation effects will be reduced, as will be described later. Examples of this kind of transfer are the two-phase CCD and the tetrode bucket brigade.

In Appendixes A and B, the various contributions to the incomplete transfer parameter α are calculated for both single-step and two-step CTD's, and in the next section, α is evaluated theoretically for several proposed CTD structures in order to provide a comparison and point

TABLE I—LIST OF ASSUMPTIONS MADE TO SIMPLIFY CTD INCOMPLETE TRANSFER CALCULATIONS

1. All devices are based on the silicon-silicon-dioxide system using n-type silicon substrates doped to a density of 10^{16} cm^{-3}.
2. The silicon dioxide thickness is 1000 Å.
3. All charge-storage regions, gates, and diffusions have a length of 10 microns.
4. Net interface state charge after transfer is constant at 2×10^{-9} coulombs per square centimeter.
5. Minority carrier mobility is constant at 200 cm^2/volt-second.
6. Fringing electric fields penetrate a distance equal to the one-dimensional space-charge width.
7. The background charge on which the signal charge is superimposed corresponds to a voltage at point A of 10 volts, and the drain-to-substrate voltage reaches a value of 10 volts at the end of transfer.
8. Clock voltages are ideal square waves.

out the important incomplete transfer mechanisms. To simplify the calculations and provide specific examples, several assumptions have been made which are listed in Table I. Calculations for conditions other than those listed in Table I can be made using the derivations in the appendix. In Table II we summarize the nature of the transfer processes involved in each type of device discussed in Section IV.

IV. INCOMPLETE TRANSFER PROPERTIES OF SEVERAL CHARGE TRANSFER DEVICES

4.1 *The Simple Bucket Brigade*

The simple bucket-brigade shift register is a two-phase CTD fabricated as illustrated in the insert in Fig. 7. In such a shift register, it has been shown[4,17] that the two dominant terms under most conditions are the drain conductance or feedback contribution, α_D, and the intrinsic transfer rate contribution, α_i. However, the analyses

TABLE II—SUMMARIES OF THE NATURE OF TRANSFER PROCESSES

Type of Device	Mode of Charge Transfer	
Single-Step Transfer Process		
Standard CCD	CCD	
Standard Bucket Brigade	BB	
Two-Step Transfer Process	First Transfer	Second Transfer
Two-Phase CCD	CCD	CCD
C4D	CCD	BB
Stepped-Oxide BB	BB	CCD
Tetrode BB	BB	BB

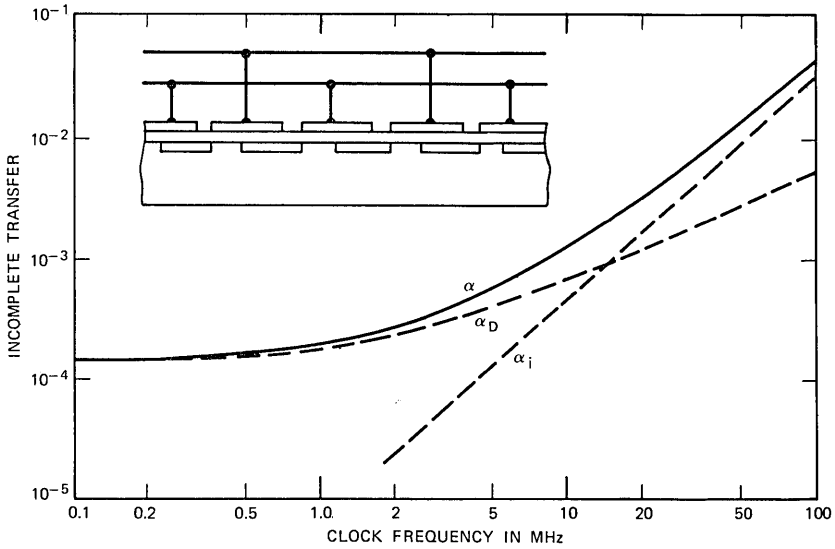

Fig. 7—Contributions to α (incomplete transfer) in the simple bucket-brigade shift register shown in the insert.

have usually assumed a simple square-law IGFET current-voltage characteristic and have ignored the fact that the carriers must be injected into the channel from the diffused regions. Recently[26] it has been pointed out that this injection requirement leads to a charge left behind which is logarithmic with time and an incomplete transfer parameter which tends to a constant value at low clock frequencies. Using the assumptions listed in Table I and assuming in addition that the drain capacitance C_D is equal to C_p, Fig. 7 shows the calculated behavior of the two contributions to α, α_i and α_D, as a function of clock frequency. The tendency for α to saturate at low frequencies is apparent in addition to the previously derived linear behavior of α_D and quadratic behavior of α_i with clock frequency at the high frequencies.

The other components of α, storage capacitance modulation and interface-state capacitance modulation, have not been shown in Fig. 7 because their values depend on the ratio of the channel capacitance to the storage capacitance. Both should be relatively small compared to the sum of α_i and α_D at all frequencies, if C_G is much smaller than C_p; so that qualitatively the bucket-brigade shift register should be well approximated by considering only α_i and α_D. However, if C_G and C_p

are of the same order as will often be the case for small-area devices, the interface-state capacitance modulation will add a frequency-independent term $\alpha_{C,SS}$ of value equal to approximately 8×10^{-4} (C_G/C_p). In that case, the interface-state term may dominate α at low frequencies.

4.2 *The Three-Phase CCD*

As originally proposed,[1] the charge-coupled device was driven by a three-phase clocking scheme as shown in the insert in Fig. 8. If it is assumed that the transfer time from one capacitor plate to the adjacent capacitor plate is one-third of a clock period, then the incomplete transfer parameter for the CCD under the restrictions of Table I is calculated to vary with clock frequency as shown in Fig. 8. In comparison to the simple bucket brigade, the three-phase CCD has a similar clock frequency behavior at high frequencies and a similar upper limit for operation. However, at lower frequencies the two modulation terms α_C and $\alpha_{C,SS}$ become dominant.[29,30] At very low frequencies, α tends to a relatively frequency-independent value due to the interface-state capacitance modulation term. Over most of the range of clock frequencies, the feedback or drain conductance term,

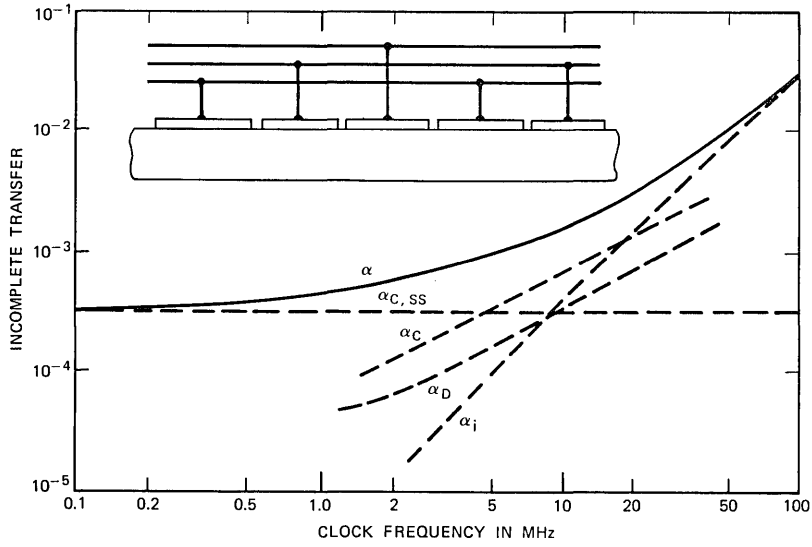

Fig. 8—Contributions to α (incomplete transfer) in the standard three-phase CCD shown in the insert.

of such importance for the bucket brigade, is nearly negligible for the three-phase CCD.

In Fig. 8 the various modulation terms have been calculated assuming that the channel-length modulation can be computed using eq. (69). Care should be used in comparing the modulation terms in Fig. 8 with those in Fig. 7. Owing to the uncertainty in the estimated modulation of the length of the channel or storage region, the magnitudes of the modulation terms should be considered to be approximate. Since α_D, α_C, and $\alpha_{C,SS}$ are each directly proportional to the length modulation, primary emphasis can be placed on their relative magnitudes and on the clock-frequency dependence of the various terms.

4.3 *The Two-Phase CCD*

Since the basic charge-coupled device has no inherent directionality like that of the bucket brigade, two-phase operation can only be achieved if some asymmetry is introduced into the CCD cell. One way to do this is to make each CCD capacitor plate consist of two regions, a transfer or barrier region which prevents carriers from moving in the wrong direction and a storage region. Ion-implanted barriers[7] and two oxide thicknesses[8,9] are two schemes which have been proposed, but two-phase operation can also be achieved by fabricating a four-phase CCD and placing a dc bias between alternate clock lines. Figure 9 illustrates these approaches to two-phase operation.

It is immediately evident from Fig. 9 that a two-phase CCD operates using a two-step transfer process. Hence, as shown in Appendix A, the operational improvement over a three-phase CCD comes primarily from making C_B small compared to C_D and by reducing the channel-length modulation. Given that the barrier length is identical to the storage capacitor length as assumed in this work, this means that best performance should be achieved either with the stepped-oxide device in Fig. 9, since the thicker oxide over the barrier region results in a smaller C_B, or with the ion-implanted device since dL_c/dV_D is reduced. In calculating the incomplete transfer for the stepped-oxide device, we will assume that the geometrical capacitance associated with the barrier and the transconductance are both one-half the values in the storage region.

Figure 10 shows the calculated incomplete transfer results for the stepped-oxide, two-phase CCD. Comparing these results to those for the three-phase CCD shown in Fig. 8, it is seen that the intrinsic transfer rate term, thus the high-frequency limitation, is about the same for both. The storage-capacitance modulation and drain con-

ductance terms are both reduced because of the small value of C_B, but the interface-state capacitance modulation term is approximately the same. Hence, operationally, some improvement is gained by using a two-phase CCD rather than a three-phase CCD.

A word of caution concerning the results for α shown in Fig. 10 is in order at this point. Some recent measurements[31] indicate that at

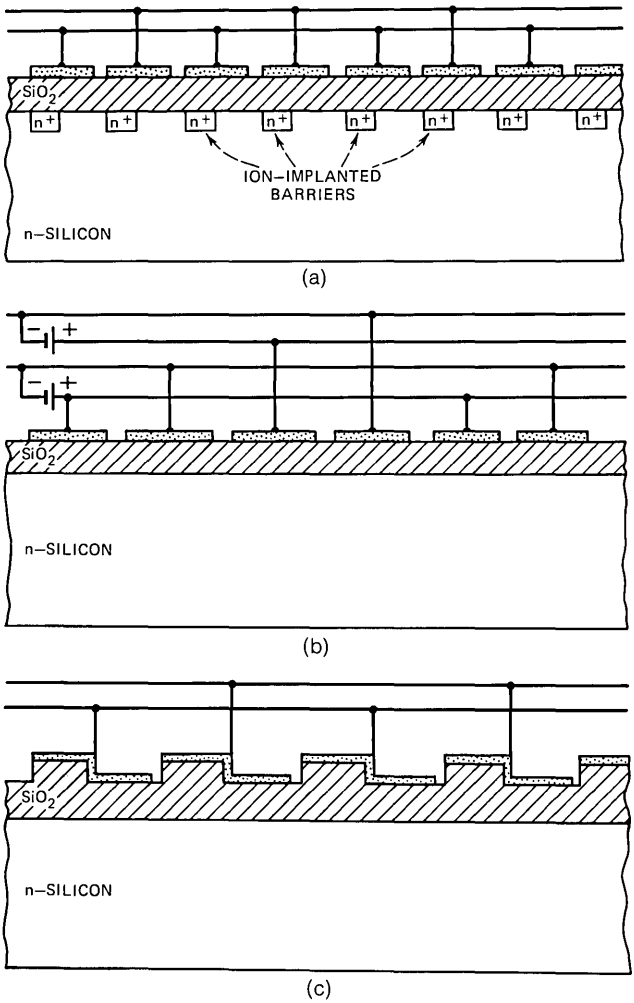

Fig. 9—Several examples of two-phase CCD device structures: (a) Ion-implanted barrier. (b) Standard four-phase run two phase. (c) Stepped-oxide.

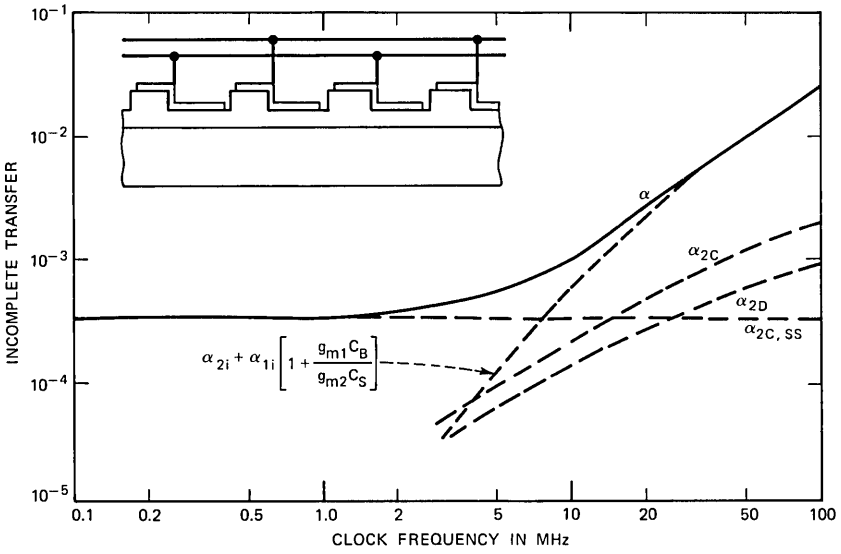

Fig. 10—Contributions to α (incomplete transfer) for the stepped-oxide, two-phase CCD.

low frequencies, α may be *in*dependent of interface-state density. However, as shown in Fig. 10, it is at low frequencies that the interface-state contribution to α is dominant, and this contribution is proportional to the density of interface states. The result shown, about $3 \cdot 10^{-4}$, is for a density of $1 \cdot 10^{10}$ states/cm^2. For $2 \cdot 10^{10}$ states 1 cm^2, one predicts $\alpha \approx 6 \cdot 10^{-4}$, which is reasonably close to the $4 \cdot 10^{-4}$ observed at this density.[32] Tentative results for other devices[31] have yielded similar low-frequency α's for interface-state densities up to $2 \cdot 10^{11}$ states/cm^2. If true, a reexamination of the contribution of interface states would be in order. However, before this is attempted, it is essential to ascertain the interface-state density not at midgap but at V_{SS} (see Appendix B) for the devices whose α's are being measured. Lateral inhomogeneities[33] may also be contributing to incomplete transfer.

4.4 *The Conductively Connected Charge-Coupled Device (C4D)*

The C4D[34] is illustrated in Fig. 11. Diffused p-regions connect the depletion region under the capacitor plates as compared to the usual CCD arrangement of closely spaced capacitor plates. An ion-implanted n^+ barrier is used to provide the directionality so that two-phase operation is achieved. From Fig. 11, it can be seen that transfer occurs

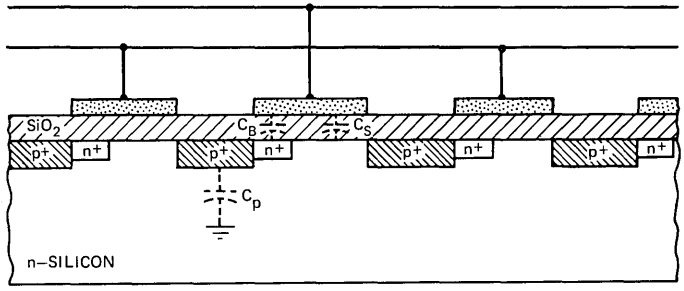

Fig. 11—Structure of the conductively connected, charge-coupled device (C4D).

by a two-step process. The charge is stored under that portion of the capacitor plate which sees the n-silicon substrate (labeled C_S in Fig. 11), and the first step of the transfer, a CCD transfer process, is from this depletion region to the adjacent diffused p-region. The second step of the transfer is from the p-region over the n^+ barrier to the next inversion region, a bucket-brigade transfer process.

In practice the p-regions will have finite capacitance to the substrate as shown dotted in Fig. 11. This has two effects. First, the p-region capacitance is in parallel with the adjacent drain capacitance C_S during transfer and charge must transfer back and forth from the p-region to C_S during clocking. Even if no charge is transferred over the barrier, a charge corresponding to the p-island capacitance multiplied by the voltage change across C_S will flow to C_S, thus assuring that the interface states in the storage region become occupied during each cycle regardless of whether charge has been transferred or not. This has been referred to as an automatic "fat zero."[7,34] However, the fat zero influences only the first step of the two-step process, and, as previously discussed (see also Appendix A), it is primarily the second step which contributes to the incomplete transfer.

The second effect of the p-region capacitance is that it adds to the intermediate capacitor C_B. Since for best operating performance we wish to minimize C_B, it seems that best results will be achieved by making the p-island capacitance as small as possible. In fact, in the limit of negligible p-island capacitance with respect to barrier capacitance, the C4D will have the same performance as a two-phase ion-implanted CCD with zero electrode spacing (see Fig. 9), and this will represent its optimum performance capabilities. However, it is the p-island capacitance when the p-region is at its most positive voltage with respect to the substrate that is of importance, and this corre-

sponds to the largest value of its capacitance. In most cases this will not be negligible compared to the drain capacitance C_S. If we consider the other extreme case, when p-island capacitance is large compared to C_S, then the C4D will perform like a simple bucket brigade but with the modulation terms multiplied by the ratio of p-island capacitance to depletion region capacitance plus p-island capacitance. Also, interface-state effects are similar to those in the ion-implanted two-phase CCD. Hence, C4D operating characteristics will lie between those of the simple bucket brigade and those of the ion-implanted two-phase CCD.

4.5 *The Stepped-Oxide and Tetrode Bucket Brigade*

The tetrode bucket brigade, first proposed by Sangster,[2,3] is shown in Fig. 12a. It was proposed[5] in order to reduce the drain conductance or

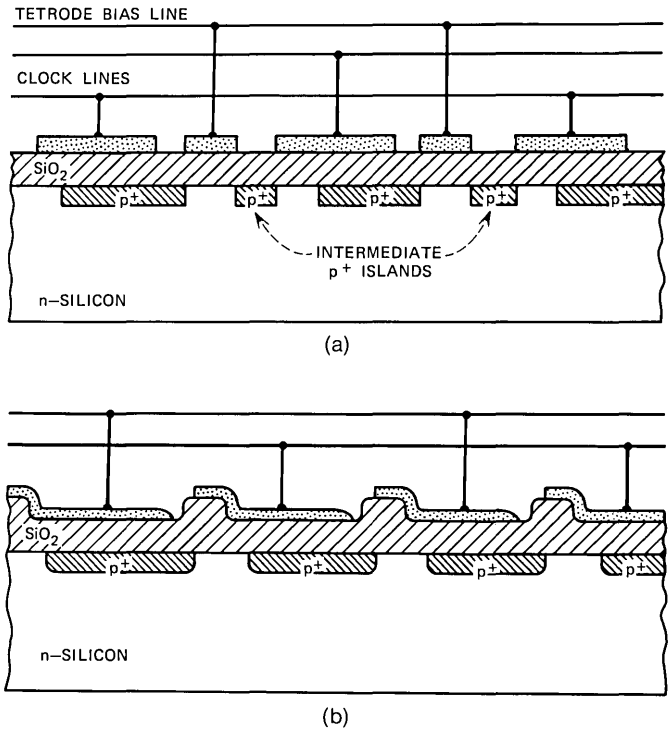

Fig. 12—(a) Structure of the tetrode bucket-brigade device. (b) Structure of the stepped-oxide bucket-brigade device.

feedback contribution to incomplete transfer, the effect known to be the dominant performance-limiting effect for bucket brigade at low frequencies. Sangster[5] pointed out that because of the finite capacitance between the intermediate p-diffusion and the substrate, the improvement in performance was not as great as was hoped, but the drain conductance term α_D was reduced by the ratio of the intermediate p-region capacitance to the storage capacitance.

Recently,[6] one of the authors (CNB) has suggested a similar scheme to reduce α_D referred to as the stepped-oxide bucket brigade and illustrated in Fig. 12b. The gate of each IGFET is separated into two regions, the region nearer the source having a higher threshold than that near the drain. In this way the charge transfer process becomes a two-step process like that of the tetrode bucket brigade, but the intermediate capacitance C_B is made up only of the effective gate capacitance associated with the low-threshold region near the drain. In this way, the first step of the charge transfer is by a "bucket-brigade" transfer process, and the second step is by a "CCD" process for the stepped-oxide bucket brigade but by another "bucket-brigade" process for the tetrode bucket brigade.

Like the other charge transfer devices treated here, the upper frequency limit for both the tetrode bucket brigade and the stepped-oxide bucket brigade will be limited by an intrinsic transfer rate term dependent on the channel lengths. Given the 10-micron assumption for all critical lengths, these two bucket-brigade registers will have the same upper frequency limit as the other CTD's. At lower frequencies, the incomplete transfer parameter should become very small depending on the ratio of C_B to C_D, and experimentally this has been found to be the case with measured values of α below 10^{-4}.[5]

V. DISCUSSION AND CONCLUSIONS

In the preceding section we discussed the coefficient of incomplete transfer, α, for CTD's in which the transfer of charge could be characterized as a single-step process or as a two-step process, and in which each step is either a "CCD" process or a "bucket-brigade" process. All devices were found to have approximately the same high-frequency limitation dominated by the intrinsic transfer rate contribution to α under the assumption that they were all of the same geometrical size. However, the middle- and low-frequency performance of the single-step devices can be improved upon by using the more complicated two-step devices. One finds in two-step devices that the source is effectively isolated from the sink so that the primary contribution to

α comes from the modulation contributions associated with the second transfer. If now the intermediate capacitance is made much less than the storage capacitance, the modulation terms are reduced by approximately the ratio of the intermediate capacitance to the storage capacitance. If this is achieved by making the storage capacitance large and keeping the intermediate capacitance fixed, better low-frequency performance will be accompanied by a lower high-frequency limit.

It was also found that the bucket-brigade transfer process leads to a frequency-independent contribution to α at low frequencies. This results from the requirement that charge must be injected from the diffused region into the channel of the IGFET. Conservation of current leads to a relation between the voltages of both sides of this barrier: V_p in the diffused region and V_A at the beginning of the channel. For long transfer periods (low clock frequency) V_A varies at $1/t$ and V_p varies as log V_A: hence, V_p varies as log t, or essentially independent of time and of clock frequency.

The analytic expressions we have obtained for the contributions to the coefficient of incomplete charge transfer α of both single-step and two-step transfer devices are quite general and provide a good qualitative and a reasonable quantitative prediction of the dependence of α upon the various device parameters. If more accurate quantitative results are desired, particular attention must be paid to channel-length modulation. This, however, is a static problem and should prove to be simpler than the dynamic problem of calculating α directly from a solution of $Q(t)$.

VI. ACKNOWLEDGMENTS

We wish to thank J. R. Brews, G. E. Smith, and R. J. Strain for helpful discussions.

APPENDIX A

Derivation of the Incomplete Transfer Parameter α

A.1 *Single-Step Process*

In a recent paper the authors[11] have shown that the incomplete transfer parameter α for a charge transfer device (ignoring for the moment the contribution of interface states) is made up of three terms

$$\alpha = \alpha_i + \alpha_D + \alpha_C$$

where α_i is the intrinsic transfer-rate contribution, α_D is the drain-conductance contribution, and α_C is the storage-capacitance modulation

term. For a single-step process, the intrinsic term is given by

$$\alpha_i = \exp\left[-\int_0^\tau \frac{g_m}{C_S}\,dt\right] \tag{16}$$

where g_m is the transconductance of the transfer mechanism joining C_S to C_D (see Fig. 6) and τ is the transfer time. As a first approximation for square-wave clocks, $g_m \approx dI/dV_S$ and $dt \approx (C_S/I)dV_S$, so eq. (16) becomes

$$\alpha_i \approx \frac{I}{I_o} \tag{17}$$

where I_o is the transfer current at the beginning of transfer.

The drain conductance term, α_D, is given by

$$\alpha_D = \frac{g_r C_S}{g_m C_D} \tag{18}$$

where g_r is the reverse transconductance of the transfer mechanism, and both g_r and g_m are evaluated at time τ. Since in eq. (4) it is only the constant β which can be dependent on drain voltage V_D, we can write

$$g_r = -\frac{\partial I}{\partial V_D} = -\frac{I}{\beta}\left(\frac{d\beta}{dV_D}\right) = I\left(\frac{1}{L_c}\frac{dL_c}{dV_D}\right) \tag{19}$$

where we have taken $\beta = Z\mu C_i/L_c$ [see eq. (4)].

According to the assumptions made here and listed in Table I, β can only vary through channel-length modulation. However, the exact form of this variation has not been theoretically established. The most common assumption is that the channel length L_G is shortened by an amount equal to the one-dimensional depletion-layer width L_D associated with the voltage on the IGFET drain diffusion. Even for an IGFET with source shorted to the substrate, this assumption is somewhat crude, and for the bucket brigade with the source p-islands reverse biased with respect to the substrate, it is expected to be even less accurate. For the CCD, an additional problem arises in that the transition fringing-field region under high lateral electric field connecting one charge storage region to the adjacent one extends under both capacitor plates so that less "channel-length modulation" will occur than for an equivalent bucket brigade.[19,20] This effect is particularly important when an ion-implanted barrier is used because, owing to the higher doping density in the barrier, channel length associated

with the barrier will change much less than the length of the adjacent storage well. An accurate treatment of this problem of calculating dL_C/dV_D is beyond the scope of this paper. However, in Appendix C we briefly outline an approximation which can be used.

Returning to eq. (19), and inserting I from Eq. (4), we have

$$g_r = \frac{\beta}{2} V_A \left(V_A + \frac{2kT}{q} \right) \frac{1}{L_c} \frac{dL_c}{dV_D}. \tag{20}$$

Similarly, from eq. (4) we can derive g_m to obtain

$$g_m = \frac{\partial I}{\partial V_S} = \beta \left(V_A + \frac{kT}{q} \right) \frac{\partial V_A}{\partial V_S}. \tag{21}$$

For a "bucket-brigade" transfer process, $\partial V_A/\partial V_S$ is $[1+(kT/qV_A)]^{-1}$ from eq. (11). It is also easy to show using eq. (11) that the effective C_S for a bucket brigade-transfer process is (see Fig. 4)

$$C_S = C_p + \frac{V_A}{V_A + \dfrac{kT}{q}} C_G. \tag{22}$$

Hence α_D is given by

$$\alpha_D = \frac{C_p \left(V_A + \dfrac{2kT}{q} \right) + C_G V_A \left[\dfrac{V_A + \dfrac{2kT}{q}}{V_A + \dfrac{kT}{q}} \right]}{2C_D} \cdot \frac{1}{L_c} \frac{dL_c}{dV_D}. \tag{23}$$

Given a calculation of V_A as a function of transfer time, the drain conductance contribution can be determined from eq. (23) by using the appropriate values of C_p and C_G. For a CCD transfer process, C_p is zero, and in the limit as V_A tends to zero (i.e., at long transfer times or low clock-frequency operation) α_D tends to zero linearly with V_A. For a bucket-brigade transfer process, C_p is finite and α_D will tend to a constant value in the limit of small V_A.

The capacitance modulation term, α_C, has been shown[11] to be

$$\alpha_C = Q_S \left(\frac{1}{C_S} \frac{dC_S}{dQ_o} \right) \tag{24}$$

where Q_S is the transferable charge remaining on C_S at time τ. Usually, however, C_S is made up of several contributions only one of which is

modulated by Q_o. In this case if the particular capacitance is C_{Si} with a transferable charge Q_{Si} remaining on it,

$$\alpha_C = \frac{Q_{Si}}{C_D}\left(\frac{1}{C_{Si}}\frac{dC_{Si}}{dV_D}\right). \tag{25}$$

In all the CTD's of interest here it is the gate capacitance C_G which is modulated due to the channel-length modulation effect. Since this capacitance is linear in channel length, we obtain

$$\alpha_C = \frac{C_G V_A}{C_D}\left(\frac{1}{L_C}\frac{dL_C}{dV_D}\right). \tag{26}$$

The one extra contribution to incomplete transfer which will be considered here is that due to interface states. It has been found[11] that such states lead to two terms, one due to interface-state capacitance modulation and the other due to the modulation of the dynamics of trapping and detrapping during transfer. For square-wave clocks and a large circulating charge or fat zero, it has been shown that the latter contribution will be small.[31] Thus we will consider here only the capacitance modulation term $\alpha_{C,SS}$. If Q_{SS} is the net charge trapped in interface states after transfer, then

$$\alpha_{C,SS} = Q_{SS}\left(\frac{1}{C_{SS}}\frac{dC_{SS}}{dQ_o}\right) \tag{27}$$

where C_{SS} is the interface-state capacitance. If we assume that this capacitance is also modulated by the channel-length modulation, then:

$$\alpha_{C,SS} = \frac{Q_{SS}}{C_D}\left(\frac{1}{L_c}\frac{dL_c}{dV_D}\right). \tag{28}$$

We have used this expression in a previous article.[11] On the other hand, Tompsett[29] has reported that variations in edge effects and in capture during transfer are important. These are included in (27); however, whereas the channel-width modulation leading to (28) can be greatly reduced by using two-step rather than single-step transfer processes (see below), edge effects, which depend on the *initial* amount of charge Q_o present, are nearly unaffected by merely increasing the number of transfer steps. Reduction of edge effects can come only by making the effective size of the well less dependent on Q_o, e.g., by ion-implanted barriers[7] or diffused regions, or by keeping the carriers away from the surface, e.g., by storing charge in diffused islands or buried

channels. Changes in capture during transfer can be greatly reduced by using zero-gap technology.[8,9]

It should be noted that apart from the intrinsic contribution α_i, the three contributions given above, α_D, α_C, and $\alpha_{C,SS}$, are all proportional to $(1/L_C) \cdot (dL_C/dV_D)$. Hence, even though our estimates of dL_C/dV_D tend to be crude (see Appendix C), the relative size of these contributions can be ascertained more precisely.

A.2 *Two-Step Process*

In a two-step transfer process as shown in Fig. 6b, the charge first transfers from C_S to an intermediate capacitor referred to as C_B, then to the drain capacitor C_D during a single transfer time τ. In this case transferable charge is left behind on both C_B and C_S, and the derivation of α becomes somewhat more complex. The main advantage of using a two-step process is that all of the contributions to α due to channel-length modulation can be reduced since modulation of C_S by the voltage V_D on C_D is a second-order effect (i.e., variations in V_D modulate slightly the value of V_B, which in turn modulates to a much lesser degree the value of V_S). For this reason, we will ignore here the channel-length modulation effects for the first step of the two-step process.

Referring to Fig. 6b for a definition of terms, the transferable charge on the first storage capacitor at time t is

$$Q_S = \int_{V_{S,0}}^{V_S} C_S dV + \int_{V_{SS1,0}}^{V_{SS1}} C_{SS1} dV. \tag{29}$$

In this expression we shall be able to ignore the effects of channel-length modulation, these being of second order in a two-step process. However, channel-length modulation effects are important for the charge Q_B on the intermediate capacitance C_B. We write

$$Q_B = \int_{V_{B0}}^{V_B} C_B dV + \int_{V_{SS2,0}}^{V_{SS2}} C_{SS2} dV \tag{30}$$

where C_{SS2} is only the interface-state term subject to first-order channel-length modulation. We can solve for the effective α for the first step of the process α_1 (see Appendix B)

$$\alpha_1 = \frac{dQ_S}{dQ_o} = \alpha_{1i} + \alpha_{1C,SS} \tag{31}$$

where

$$\alpha_{1i} = \exp\left[-\int_0^\tau \frac{g_{m1}}{C_S} dt'\right]$$

TABLE III—MAJOR CONTRIBUTIONS TO INCOMPLETE CHARGE TRANSFER IN TERMS OF SMALL-SIGNAL DEVICE PARAMETERS $\alpha = \alpha_i + \alpha_D + \alpha_C + \alpha_{C,SS}$

Type	Single-Step	Two-Step
Intrinsic, α_i	$\exp\left(-\int_0^\tau (g_m/C_S)dt\right)$	$\exp\left(-\int_0 (g_{m2}/C_B)dt\right) + \exp\left(-\int_0^\tau (g_{m1}/C_S)dt\right) \cdot (1 + g_{m1}C_B/g_{m2}C_S)$
Drain Conductance, α_D (Feedback)	g_rC_S/g_mC_D	$g_{r2}C_B/g_{m2}C_D$
Capacitance Modulation, α_C	$Q_S\left(\frac{1}{C_S}\frac{dC_S}{dQ_o}\right)$	$Q_B\left(\frac{1}{C_B}\frac{dC_B}{dQ_o}\right)$
Interface State Capacitance Modulation, $\alpha_{C,SS}$	$Q_{SS}\left(\frac{1}{C_{SS}}\frac{dC_{SS}}{dQ_o}\right)$	$Q_{SS1}\left(\frac{1}{C_{SS1}}\frac{dC_{SS1}}{dQ_o}\right) + Q_{SS2}\left(\frac{1}{C_{SS2}}\frac{dC_{SS2}}{dQ_o}\right)$

TABLE IV—MAJOR CONTRIBUTIONS TO INCOMPLETE CHARGE TRANSFER IN TERMS OF APPROXIMATE EVALUATION OF DEVICE PARAMETERS $\alpha = \alpha_i + \alpha_D + \alpha_C + \alpha_{C,SS}$

Type	Single-Step	Two-Step
Intrinsic, α_i	I/I_0	$\frac{I_1}{I_{10}}\left(1+\frac{g_{m1}C_B}{g_{m2}C_S}\right)+I_2/I_{20}$
Drain Conductance, α_D (Feedback)	$\frac{C_p\left(V_A+\frac{2kT}{q}\right)+C_GV_A\left[\frac{V_A+\frac{2kT}{q}}{V_A+\frac{kT}{q}}\right]}{2C_D}\frac{1}{L_C}\frac{dL_C}{dV_D}$	α_D for second transfer only
Capacitance Modulation, α_C	$\frac{C_GV_A}{C_D}\left(\frac{1}{L_C}\frac{dL_C}{dV_D}\right)$	$\frac{C_BV_A}{C_D}\left(\frac{1}{L_{C2}}\frac{dL_{C2}}{dV_D}\right)$
Interface State Capacitance Modulation, $\alpha_{C,SS}$	$\frac{Q_{SS}}{C_D}\left(\frac{1}{L_C}\frac{dL_C}{dV_D}\right)$	$\frac{Q_{SS2}}{C_D}\left(\frac{1}{L_{C2}}\frac{dL_{C2}}{dV_D}\right)+\alpha_{1C,SS}$

and $\alpha_{1C,SS}$ is given in (65). For the second step, however, the net current into C_B is the difference between I_1 and I_2 so that the equation to be solved is, using eq. (31),

$$\frac{d\alpha_2}{dt} = \frac{g_{m1}}{C_S} \exp\left[-\int_0^t \frac{g_{m1}}{C_S}\, dt\right] - g_{m2}\frac{dV_B}{dQ_o} + g_{r2}\frac{dV_D}{dQ_o} \tag{32}$$

where

$$\alpha_2 = \frac{dQ_B}{dQ_o}. \tag{33}$$

(Details for arriving at an expression similar to (32) are given in Appendix B.) From eq. (30),

$$\frac{dV_B}{dQ_o} = \frac{\alpha_2}{C_B} - \frac{1}{C_B}\int_{V_{B0}}^{V_B} \frac{dC_B}{dQ_o}\, dV - \frac{1}{C_B}\int_{V_{SS2,0}}^{V_{SS2}} \frac{dC_{SS2}}{dQ_o}\, dV \tag{34}$$

and

$$\frac{dV_D}{dQ_o} = \frac{1}{C_D}. \tag{35}$$

Solving for α_2, in a similar manner to that used in Ref. 10,

$$\alpha_2 = \alpha_{2i} + \alpha_{2D} + \alpha_{2C} + \alpha_{2C,SS} + \alpha'_{1i} \tag{36}$$

where

$$\alpha_{2i} = \exp\left[-\int_0^\tau \frac{g_{m2}}{C_B}\, dt\right] \approx \frac{I_2}{I_{20}} \tag{37}$$

is the intrinsic transfer rate term,

$$\alpha_{2D} \approx \frac{g_{r2}}{g_{m2}}\frac{C_B}{C_D} \tag{38}$$

is the drain conductance or feedback term, which can be evaluated using eq. (23),

$$\alpha_{2C} \approx Q_B\left[\frac{1}{C_B}\frac{dC_B}{dQ_o}\right] \approx \frac{Q_B}{C_D}\left(\frac{1}{L_{G2}}\frac{dL_{G2}}{dV_D}\right) \tag{39}$$

is the capacitance modulation term,

$$\alpha_{2C,SS} \approx Q_{SS2}\left[\frac{1}{C_{SS2}}\frac{dC_{SS2}}{dQ_o}\right] \approx \frac{Q_{SS2}}{C_D}\left(\frac{1}{L_{G2}}\frac{dL_{G2}}{dV_D}\right) \tag{40}$$

is the interface-state capacitance modulation term, and

$$\alpha'_{1i} \approx \frac{g_{m1}}{g_{m2}}\frac{C_B}{C_S}\alpha_{1i} \tag{41}$$

reflects the fact that charge transfers from C_S to C_B in addition to the transfer of charge from C_B to C_D. Realizing that the total incomplete transfer parameter is $\alpha_1 + \alpha_2$, the final result for a two-step transfer process can be written in terms of the parameters of the individual transfer steps as

$$\alpha = \alpha_{2i} + \alpha_{2D} + \alpha_{2C} + \alpha_{2C,SS} + \alpha_{1C,SS} + \alpha_{1i}\left(1 + \frac{g_{m1}}{g_{m2}}\frac{C_B}{C_S}\right). \quad (42)$$

Referring to eqs. (38), through (41), it is apparent that a significant improvement in α can be obtained in a two-step process by making C_B much smaller than C_D, provided the edge-effect contribution[29] to $\alpha_{1C,SS}$ can be reduced.

The results of this appendix are summarized in Tables III and IV.

APPENDIX B

General Derivation of the Incomplete Transfer Parameter α

In a previous paper[11] we outlined the details of the derivation of our general expression for α for the single-step process. In this appendix a similar derivation of α is given for the two-step process. The general derivation for an m-step process should then be straightforward to devise if needed.

Referring to Fig. 6b, as the size of the initial, transferable charge Q_o is varied, both $Q_S(t)$ and $Q_B(t)$ will also vary. If $Q_D(t) \equiv Q_o - Q_S(t) - Q_B(t)$ is the transferable charge on the drain, then α for transfer process may be defined as

$$1 - \alpha \equiv \frac{dQ_D}{dQ_o}. \quad (43)$$

It follows that

$$\alpha = \alpha_1 + \alpha_2 \quad (44)$$

where

$$\alpha_1 = \frac{dQ_S}{dQ_o} \quad (45)$$

referring to I_1, the current of the first transfer step, and where

$$\alpha_2 = \frac{dQ_B}{dQ_o} \quad (46)$$

which refers to both I_1 and I_2, I_2 being the current of the second transfer step. As in the case of single-step transfer, we shall derive a differential equation which can then be solved for α.

As before,[11] we first assume that we can write the currents governing the transfer of charge in the following form

$$I_1 = I_1(V_S, V_B, V_C, V_{SS1}) \tag{47}$$

and

$$I_2 = I_2(V_B, V_D, V_C, V_{SS2}) \tag{48}$$

where the additional voltages V_{SSi} $(i = 1,2)$ are just the voltages induced by the trapped charges Q_{SSi} on the effective capacitances C_{SSi} of the traps (V_C is the clock voltage). Q_S, Q_B, and Q_D are related to C_S, C_B, C_{SSi} $(i = 1,2)$, and C_D as follows

$$Q_S = \int_{V_{S,0}}^{V_S} C_S dV + \int_{V_{SS1,0}}^{V_{SS1}} C_{SS1} dV \tag{49}$$

$$Q_B = \int_{V_{B,0}}^{V_B} C_B dV + \int_{V_{SS2,0}}^{V_{SS2}} C_{SS2} dV \tag{50}$$

$$Q_D = \int_{V_{D,0}}^{V_D} C_D dV = Q_o - Q_S - Q_B. \tag{51}$$

From (49) to (51) and using (43) to (46) one can obtain the following relationships:

$$\frac{dV_S}{dQ_o} = \frac{\alpha_1}{C_S} - \frac{1}{C_S}\int_{V_{S,0}}^{V_S} \frac{dC_S}{dQ_o} dV - \frac{\alpha_{SS1}}{C_S} - \frac{1}{C_S}\int_{V_{SS1,0}}^{V_{SS1}} \frac{dC_{SS1}}{dQ_o} dV \tag{52}$$

$$\frac{dV_B}{dQ_o} = \frac{\alpha_2}{C_B} - \frac{1}{C_B}\int_{V_{B,0}}^{V_B} \frac{dC_B}{dQ_o} dV - \frac{\alpha_{SS2}}{C_B} - \frac{1}{C_B}\int_{V_{SS2,0}}^{V_{SS2}} \frac{dC_{SS2}}{dQ_o} dV \tag{53}$$

$$\frac{dV_D}{dQ_o} = \frac{1 - \alpha_1 - \alpha_2}{C_D} - \frac{1}{C_D}\int_{V_{D,0}}^{V_D} \frac{dC_D}{dQ_o} dV \approx \frac{1}{C_D} \tag{54}$$

where we have made use of the following definition:

$$\alpha_{SSi} \equiv C_{SSi} \frac{dV_{SSi}}{dQ_o}, \qquad (i = 1,2). \tag{55}$$

A differential equation for $\alpha = \alpha_1 + \alpha_2$ can now be derived at once. Since $dQ_S/dt = -I_1$ and $dQ_B/dt = I_1 - I_2$ (conservation of charge), it follows that

$$\frac{d\alpha}{dt} = -\frac{dI_2}{dQ_o} = -g_{m2}\frac{dV_B}{dQ_o} + g_{r2}\frac{dV_D}{dQ_o} - \frac{\partial I_2}{\partial V_{SS2}}\frac{dV_{SS2}}{dQ_o} \tag{56}$$

where as before $g_{m2} \equiv \partial I_2/\partial V_B$ and $g_{r2} \equiv -\partial I_2/\partial V_D$. Inserting (53)

to (55) into (56), replacing α_2 by $\alpha - \alpha_1$, and solving for α, one finds that

$$\alpha(t) = \int_0^t \alpha_1(t')E(t,t')dt' + \int_0^t \left(\int_{V_{B,0}}^{V_B} \frac{dC_B}{dQ_o} dV\right) E(t,t')dt' + \int_0^t (g_{r2}C_B/g_{m2}C_D)E(t,t')dt' + \int_0^t \left(\alpha_{SS2}\left(1 - \frac{\partial I_2}{\partial V_{SS2}} \frac{C_B}{C_{SS2}} \frac{1}{g_{m2}}\right) + \int_{V_{SS2,0}}^{V_{SS2}} \frac{dC_{SS2}}{dQ_o} dV\right) E(t,t')dt' \quad (57)$$

where $E(t,t')$, the suppression factor, is defined by

$$E(t,t') \equiv \exp\left(-\int_{t'}^t dt''/\tau_2(t'')\right) \Big/ \tau_2(t') \quad (58)$$

and

$$1/\tau_2(t') \equiv g_{m2}(t')/C_B(t'). \quad (59)$$

The second, third, and fourth integrals in (57) can be evaluated at once assuming the factor multiplying $E(t,t')$ varies more slowly than $E(t,t')$. This yields

$$\alpha(t) \equiv \int_0^t \alpha_1(t')E(t,t')dt' + \int_{V_{B,0}}^{V_B} dC_B/dQ_o dV + g_{r2}C_B/g_{m2}C_D + \alpha_{SS2}\left(1 - \frac{\partial I_2}{\partial V_{SS2}} \frac{C_B}{C_{SS2}} \frac{1}{g_{m2}}\right) + \int_{V_{SS2,0}}^{V_{SS2}} \frac{dC_{SS2}}{dQ_o} dV. \quad (60)$$

These terms are discussed in the text, in Appendix A, and in a previous work.[11]

To evaluate the first term in (57) we must calculate $\alpha_1(t)$. To do this we derive and then solve a differential equation for $\alpha_1(t)$. Since $dQ_S/dt = -I_1$, it follows that

$$\frac{d\alpha_1}{dt} = -\frac{dI_1}{dQ_o} = -g_{m1}\frac{dV_S}{dQ_o} + g_{r1}\frac{dV_B}{dQ_o} - \frac{\partial I_1}{\partial V_{SS1}} \frac{dV_{SS1}}{dQ_o}. \quad (61)$$

Before inserting (52) and (53) into (61), we should note the relative magnitudes of the contributions to α_1. dV_B/dQ_o is bounded by α/C_B

[see (60)]. Hence the contribution to α_1 of the term $g_{r1}dV_B/dQ_o$ is of the order of $\alpha \cdot \alpha_1$, which is quite negligible. As explained in Appendix A, variations in C_S are second order in ΔQ_o and hence can be ignored in (52). In Ref. 9, we also pointed out that dC_{SS}/dQ_o was the dominant interface-state contribution to α, the term in α_{SS} being much smaller. Here dC_{SS1}/dQ_0 and dC_{SS2}/dQ_o are the corresponding dominant trapping terms by the same reasoning. Hence, the third term in (52) as well as the fourth term in (60) can be dropped. This reduces (61) to

$$\frac{d\alpha_1}{dt} = -\frac{g_{m1}}{C_S}\alpha_1 + \frac{g_{m1}}{C_S}\int_{V_{SS1,0}}^{V_{SS1}} \frac{dC_{SS1}}{dQ_o}\,dV \tag{62}$$

the solution of which is

$$\alpha_1(t) = \alpha_{1i}(t) + \alpha_{1C,SS}(t) \tag{63}$$

where

$$\alpha_{1i}(t) = \exp\left(-\int_0^t (g_{m1}/C_S)_{t''}dt''\right) \tag{64}$$

and

$$\alpha_{1C,SS}(t) = \int_0^t dt' \exp\left(-\int_{t'}^t (g_{m1}/C_S)_{t''}dt''\right) \cdot \left(\frac{g_{m1}}{C_S}\int_{V_{SS1,0}}^{V_{SS1}} \frac{dC_{SS1}}{dQ_o}\,dV\right)_{t'} \tag{65}$$

$$\approx \int_{V_{SS1,0}}^{V_{SS1}} \frac{dC_{SS1}}{dQ_o}\,dV \tag{66}$$

$$\approx Q_{SS1}\frac{1}{C_{SS1}}\frac{dC_{SS1}}{dQ_o}. \tag{67}$$

Inserting (63) for $\alpha_1(t)$ into (60) we obtain an explicit expression for $\alpha(t)$ as desired.

Our expression for $\alpha(t)$ can be simplified somewhat. If $\alpha_{1C,SS}$ (66) can be assumed to be slowly varying near turnoff, then we can use the fact that $\int E\,dt = 1$ [see (58) and (59)] to write

$$\alpha(t) = \int_0^t \alpha_{1i}(t)E(t,t')dt' + \int_{V_{B,0}}^{V_B} dC_B/dQ_o dV + g_{r2}C_B/g_{m2}C_D + \int_{V_{SS1,0}}^{V_{SS1}} dC_{SS1}/dQ_o dV + \int_{V_{SS2,0}}^{V_{SS2}} dC_{SS2}/dQ_o dV. \tag{68}$$

In (68), we have dropped terms proportional to α_{SS1} and α_{SS2} as discussed above.

To simplify the first term of (68) we must proceed more carefully. Using (64), we may write this intrinsic term as

$$\alpha_{1i}(t)\int_0^t dt' \exp\left(-\int_{t'}^t dt''(g_{m2}/C_B - g_{m1}/C_S)\right)\cdot(g_{m2}(t')/C_B(t')).$$

This expression can be handled in various ways. In Appendix A we have assumed that $g_{m2}/C_B \gg g_{m1}/C_S$ in which case one can multipy and divide the integrand by $(g_{m2}/C_B - g_{m1}/C_S)$, assume $(g_{m2}/C_p)\cdot(g_{m2}/C_B - g_{m1}/C_S)^{-1}$ is slowly varying, and integrate as before to obtain the intrinsic term of α

$$\alpha_{1i}\cdot(1 - g_{m1}C_B/g_{m2}C_S)^{-1} \approx \alpha_1(1 + g_{m1}C_B/g_{m2}C_S).$$

Alternatively, if $g_{m2}/C_B = a_2 f(t)$ and $g_{m1}/C_S = a_1 f(t)$, then the integral may be performed without approximation, and one finds

$$\alpha_{1i}\frac{a_2}{a_2 - a_1}\left\{1 - \exp\left[-\int_0^t (a_2 - a_1)f(t')dt'\right]\right\}.$$

Finally, in the extreme case that $g_{m2}/C_B = g_{m1}/C_S$ one has for this term

$$\alpha_{1i}\log_e(1/\alpha_{1i})$$

which varies typically from $5\alpha_1$ to $9\alpha_1$ for α_1 from 0.007 to 0.0001. Being the intrinsic term, it dominates in α only for relatively high clock frequencies.[11]

APPENDIX C

In this appendix we consider several approximations which can be used to determine dL_C/dV_D. One approach is to simply define an appropriate one-dimensional depletion layer width L_D assumed to vary as the square root of voltage V_D, and assume that the effective channel length, L_c is $L_G - L_D$. This approximation, valid only when the oxide thickness (times the ratio of the dielectric constant of the oxide to that of the semiconductor) is much less than the space-charge width, we shall use when a channel in the substrate material empties into a heavily doped diffused region, as is the case for the simple, tetrode, and stepped-oxide bucket brigade. Under these conditions we find using eq. (4)

$$\frac{1}{L_c}\frac{dL_c}{dV_D} = \frac{L_D}{2(L_G - L_D)}\frac{1}{|V_D|}. \tag{69}$$

If, on the other hand, one has charge transport over a barrier into an inverted region at the surface of the substrate material, the length modu-

lation will be reduced approximately by a factor of $(1 + N_B/N_S)^{-1}$, where N_B is the doping in the barrier region and N_S is the doping in the substrate. This may be applied to the use of the C4D and the ion-implanted-barrier two-phase CCD. Thus one finds

$$\frac{1}{L_c}\frac{dL_c}{dV_D} = \frac{L_D}{2(L_G - L_D)}\frac{1}{|V_D|}\frac{1}{1 + N_B/N_S} \tag{70}$$

where L_D refers to the substrate material. Estimating dL_c/dV_D by equating it to the modulation in the size of the depletion region should represent an upper limiting case. That is, the actual dL_c/dV_D should be less than the value predicted by (69) or (70), making our resulting prediction for α conservatively larger than the actual result.

For CCD structures in which charge transfers are directly between two inversion regions (e.g., as in the simple polyphase CCD and the stepped-oxide, two-phase CCD), the depletion-width approximation discussed above is clearly very unsatisfactory. We attempted to find a better estimate for dL_c/dV_D in such cases but were unsuccessful.

We hope that better approximations will be motivated by extensions of existing calculations.[19,20]

Note: We point out that dL_c/dV_D might be interpreted as the reciprocal of an electric field. The first field that comes to mind is the average (or peak) field at the silicon-silica interface between inversion regions. For the device dimensions used in examples in the text $dL_c/dV_D \approx (10^6\ \text{V/cm}^2)^{-1}$, whereas computer calculations[21] suggest that the surface electric field is on the order of 10^4 to 10^5 V/cm. This disparity seems to rule out such a simple interpretation.

REFERENCES

1. Boyle, W. S., and Smith, G. E., "Charge Coupled Semiconductor Devices," B.S.T.J., *49*, No. 4 (April 1970), pp. 587–593.
2. Sangster, F. L. J., and Teer, K., "Bucket-Brigade Electronics–New Possibilities for Delay, Time-Axis Conversion, and Scanning," IEEE J. Solid-State Circuits, *SC-4* (June 1969), pp. 131–136.
3. Sangster, F. L. J., presented at the 1970 International Solid-State Circuits Conference, Philadelphia, Pa., February 18–20, 1970.
4. Berglund, C. N., and Boll, H. J., presented at the 1970 International Electron Devices Meeting, Washington, D. C., October 28–30, 1970, and "Performance Limitations of the IGFET Bucket Brigade Shift Register," IEEE Trans. Electron Devices, *ED-19* (1972), pp. 852–860.
5. Sangster, F. L. J., "Progress on Bucket-Brigade Charge-Transfer Devices," presented at the International Solid-State Circuits Conference, Philadelphia, Pa., February 17, 1972.
6. Berglund, C. N., unpublished work.
7. Krambeck, R. H., Walden, R. H., and Pickar, K. A., "Implanted Barrier Two-Phase Charge Coupled Devices," Appl. Phys. Ltrs., *19* (December 1971), pp. 520–522. R. J. Strain, R. H. Krambeck, private communication.

8. Kahng, D., and Nicollian, E. H., U. S. Patent #3651349, issued March 1972.
9. Berglund, C. N., Powell, R. J., Nicollian, E. H., and Clemens, J. T., "Two-Phase Stepped-Oxide CCD Shift Register Using Undercut Isolation," Appl. Phys. Ltrs., *20* (1972), pp. 413–414.
10. Krambeck, R. H., Strain, R. J., and Smith, G. E., unpublished work.
11. Berglund, C. N., and Thornber, K. K., "Incomplete Transfer in Charge Transfer Devices," IEEE J. Solid-State Circuits, to be published. See Refs. 5 to 21 of this paper for a list of prior treatments of incomplete charge transfer.
12. Lee, M. S., and Heller, L. G., "Charge-Control Method of Charge-Coupled Device Transfer Analysis," IEEE Trans. Electron Devices, to be published.
13. Berglund, C. N., and Thornber, K. K., unpublished work.
14. Koehler, D., "The Charge-Control Concept in the Form of Equivalent Circuits," B.S.T.J., *46*, No. 3 (March 1967), pp. 523–576.
15. Beaufoy, R., and Sparkes, J. J., "The Junction Transistor as a Charge-Controlled Device," ATE J. (London), *13*, 1957, pp. 310–324.
16. Gummel, H. K., "A Charge-Control Relation for Bipolar Transistors," B.S.T.J., *49*, No. 1 (January 1970), pp. 115–120.
17. Thornber, K. K., "Incomplete Charge Transfer in IGFET Bucket-Brigade Shift Registers," IEEE Trans. Electron Devices, *ED-18* (October 1971), pp. 941–950.
18. Berglund, C. N., "The Bipolar Bucket Brigade Shift Register," IEEE J. Solid State Circuits, *SC-7*, 1972, pp. 180–184.
19. Strain, R. J., and Schryer, N. L., "A Nonlinear Diffusion Analysis of Charge-Coupled-Device Transfer," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1721–1740.
20. Mohsen, A. M., McGill, T. C., and Mead, C. A., "Charge Transfer in Charge-Coupled Devices," IEEE Solid-State Conference Digest of Technical Papers, *15*, 1972, pp. 248–249.
21. Amelio, G. F., "Computer Modeling of Charge-Coupled Device Characteristics," B.S.T.J., *51*, No. 3 (March 1972), pp. 705–730.
22. Walden, R. H., Krambeck, R. H., Strain, R. J., McKenna, J., Schryer, N. L., and Smith, G. E., "The Buried Channel Charged Coupled Device," B.S.T.J., *51*, No. 7 (September 1972), pp. 1635–1460.
23. Sze, S. M., *Physics of Semiconductor Devices*, New York: John Wiley & Sons, Inc., 1969, Ch. 10, Section 4.
24. Kim, C. K., and Lenzlinger, M., "Charge Transfer in Charge-Coupled Devices," J. Appl. Phys., *42*, 1971, pp. 3586–3594.
25. Barron, M. B., "Low Level Currents in Insulated Gate Field Effect Transistors," Solid-State Electrons, *15*, 1972, pp. 293–302.
26. Buss, D. D., and Gosney, W. M., "The Effect of Subthreshold Leakage on Bucket-Brigade Device Operation," 1972 Device Research Conference, Edmonton, Alberta.
27. Joyce, W. B., and Bertram, W. J., "Linearized Dispersion Relation and Green's Function for Discrete-Charge-Transfer Devices with Incomplete Transfer," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1741–1759.
28. Berglund, C. N., "Analog Performance Limitations of Charge-Transfer Dynamic Shift Registers," IEEE J. Solid-State Circuits, *SC-6* (December 1971), pp. 391–394.
29. Tompsett, M. F., "Quantitative Effects of Interface States on the Performance of Charge-Coupled Devices," IEEE Trans. on Electron Devices.
30. Berglund, C. N., and Strain, R. J., "Fabrication and Performance Consideration of Charge-Transfer Dynamic Shift Registers," B.S.T.J., *51*, No. 3 (March 1972), pp. 655–703.
31. Kahng, D., private communication.
32. Tompsett, M. F., Kosicki, B. B., and Kahng, D., "Measurements of Transfer Inefficiency of 250-Element Undercut-Isolated Charge Coupled Devices," B.S.T.J., *52*, No. 1 (January 1973), pp. 1–7.
33. Brews, J. R., "Surface Potential Fluctuations Generated by Interface Charge Inhomogeneities in MOS Devices," J. Appl. Phys., *43*, (1972), pp. 2306–2313.
34. Strain, R. J., "Properties of an Idealized Traveling Wave Charge Coupled Device," IEEE Trans. Electron Devices, *ED-20* (October, 1972), pp. 1119–1130.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 2, February, 1973
Printed in U.S.A.

Quantizing Noise of ΔM/PCM Encoders

By DAVID J. GOODMAN and LARRY J. GREENSTEIN

(Manuscript received September 11, 1972)

We consider the pulse-code-modulation encoder that contains a delta modulator for analog-to-digital conversion, and a finite impulse response digital filter that suppresses high-frequency components of the delta modulation signal. A PCM word generator produces fixed-length binary code words by rounding and amplitude limiting the filter output samples. The quantizing noise of the resulting PCM signal has four components: delta modulation slope overload noise, filtered delta modulation granular noise, amplitude overload noise, and word generator roundoff noise. We analyze the total quantizing noise for the case where the encoder input is a Gaussian random process and the digital filter impulse response is uniform (all coefficients equal). Such filters possess important implementation advantages and appear to be near optimal with respect to signal-to-noise performance. Our analysis results in curves which show the relationship of signal-to-noise ratio to filter order, delta modulation sampling rate, and PCM word length.

I. INTRODUCTION

A new approach to digital encoding of continuous waveforms employs digital hardware to unite the economy of single-integration delta modulation (ΔM) with the efficiency of pulse code modulation (PCM). A finite impulse response digital filter suppresses the granular noise component of the ΔM representation of a continuous signal, and a word generator truncates the binary coded filter output to produce PCM code words of desired length. This encoding method controls the precision of the digital code by means of the ΔM speed and the filter order rather than with the resolution of the multibit quantizer that appears in conventional PCM encoders. This is a desirable substitution in view of current technology in which the cost of high-speed digital circuitry is rapidly declining.

This method of ΔM/PCM encoding, which was originally proposed by Goodman,[1] has been applied to speech encoding by Freeny, et al.,[2,3]

and to video by Kaneko and Ishiguro.[4] Previous theoretical results[1] focus on the filtering of the ΔM granular noise, but provide little insight into the important influence of ΔM slope overload and PCM amplitude overload on encoder design. Assuming the encoder input is a sample function of a Gaussian random process, the present paper analyzes the effects of the overload components of the quantizing distortion. It demonstrates that amplitude overload noise can be reduced if least significant bits of the filter output are truncated.

We focus our attention on "uniform filter encoders," in which all filter impulse response coefficients are unity. Such encoders offer significant practical advantages, and they appear to be near optimal with respect to signal-to-noise performance. For such encoders, we show how performance varies with filter order, ΔM speed, and PCM word length, and we demonstrate the application of our results to the design of practical encoders.

II. SIGNAL PROCESSING OPERATIONS

The block diagram of Fig. 1 shows the operations involved in transforming the continuous signal $y(t)$ to a uniformly quantized M-bit PCM sequence. Digital logic may be added to convert this sequence to a nonuniform PCM format.[5] The single-integration delta modulator of Fig. 2 converts $y(t)$ to a sequence of pulses with amplitude $+1$ or -1 at the rate $f_s = 1/\tau$ per second. The feedback loop is an ideal integrator with gain factor δ, while the up/down counter obtains a digital replica of $x(t)$, the ΔM approximation signal. The output of the Nth-

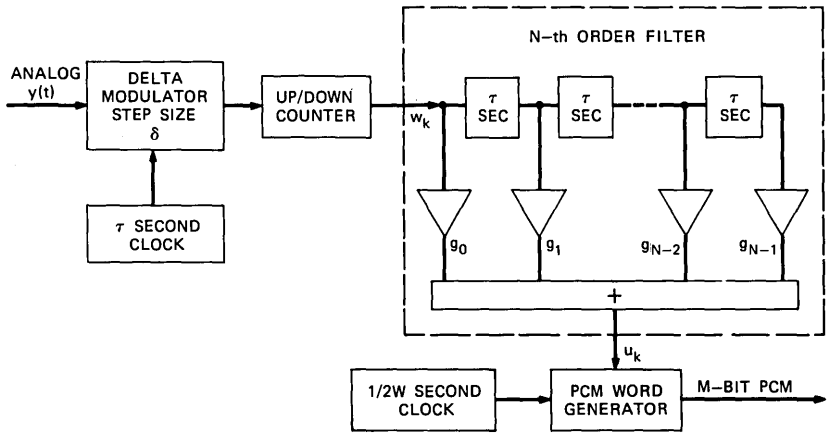

Fig. 1—Encoder block diagram.

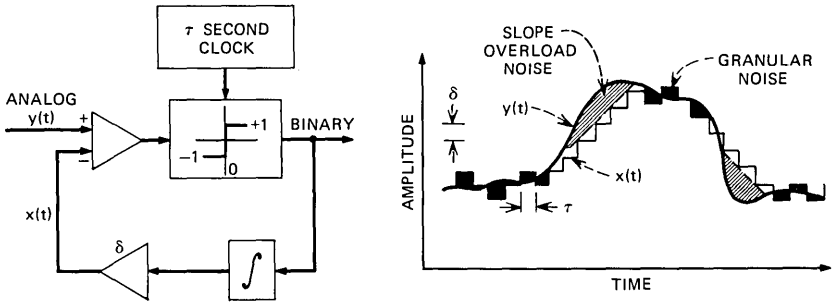

Fig. 2—Delta modulator.

order digital filter is a weighted sum of consecutive ΔM approximations to $y(t)$.

Although the filter outputs range over a discrete set, the number of possible filter outputs is unlimited because there is no fundamental restriction on the amplitude range represented by a delta modulator. It follows that, with the PCM word length prespecified, an additional quantizing operation is required. This quantization is performed by the word generator which restricts to 2^M the number of possible coder output words. The word generator introduces amplitude overload and it may also add to the granular quantization error by rounding off least significant bits of the filter output.

The filter and word generator are controlled by a clock which causes coder output words to be generated at the rate $2W$ per second, where W is the bandwidth of the analog input. Hence, the data rate of the coder is $2MW$ bits/second and the PCM sequence may be decoded as if it were produced by a conventional encoder consisting of a $1/2W$-second sampler and a uniform quantizer with 2^M output levels.

III. THE PCM QUANTIZATION LEVELS

With the filter coefficients, g_i, integers as in a practical realization, the filter output at $t = k\tau$ is the integer

$$u_k = \sum_{i=0}^{N-1} g_i w_{k-i} \tag{1}$$

where $\{w_j\}$ is the sequence of counter outputs. Because $w_j = w_{j-1} \pm 1$, the parity of the filter input alternates between even and odd at each ΔM sampling instant. It follows that if $f_s/2W$, the ratio of ΔM sampling rate to PCM sampling rate, is an even integer, the parities of

w_k, w_{k-1}, $\cdots$, w_{k-N+1} are invariant at the PCM sampling instants. Hence, the parity of u_k is the same at all PCM sampling instants. Because odd-parity filter outputs lead to an easily implemented word generator, we restrict our attention to encoders in which w_k and u_k are both odd integers at the PCM sampling instants. The filter coefficients of these encoders satisfy conditions, derived in the Appendix, which do not severely restrict the set of available filter transfer functions. The conditions do, however, preclude uniform filter encoders of orders 4, 8, 12, etc.

If t_0 is the encoder delay, the odd integer u_k is a scaled approximation to $y(k\tau - t_0)$. To determine the scaling factor, we observe that $x(k\tau)$, the ΔM approximation to $y(k\tau)$, is related to w_k by $x(k\tau) = \delta w_k$. Further, since the filter provides relatively distortionless gain over the signal bandwidth, it expands the amplitude scale of w_k by approximately the amount of the dc gain,

$$I = \sum_{i=0}^{N-1} g_i. \tag{2}$$

Thus, $(\delta/I)u_k$ is an approximation to $y(k\tau - t_0)$ and, with u_k ranging over odd integers, the signal levels represented by the input to the word generator are in the set

$$\cdots, \quad -3\frac{\delta}{I}, \quad -\frac{\delta}{I}, \quad \frac{\delta}{I}, \quad 3\frac{\delta}{I}, \cdots, \tag{3}$$

with quantizing step size $2\delta/I$. Because the scaling by δ/I is approximate, we admit an additional scale factor, γ, which brings the PCM representation optimally close to $y(t)$ in the mean square sense. The actual step size of the filter output is therefore

$$d_0 = \frac{2\delta}{I}\gamma. \tag{4}$$

In Section 7.5, we show that γ, which depends on g_i, is close to unity for encoders of practical interest.

Figure 3 shows the mapping of the filter output into M-bit code words. To eliminate α information bits from the binary representation of u_k, the word generator truncates the $\alpha + 1$ least significant bits. (With u_k odd at PCM sampling instants, the least significant bit always has value one and hence conveys no information.) In the absence of amplitude overload,

$$|u_k| \leqq 2^{M+\alpha} - 1$$

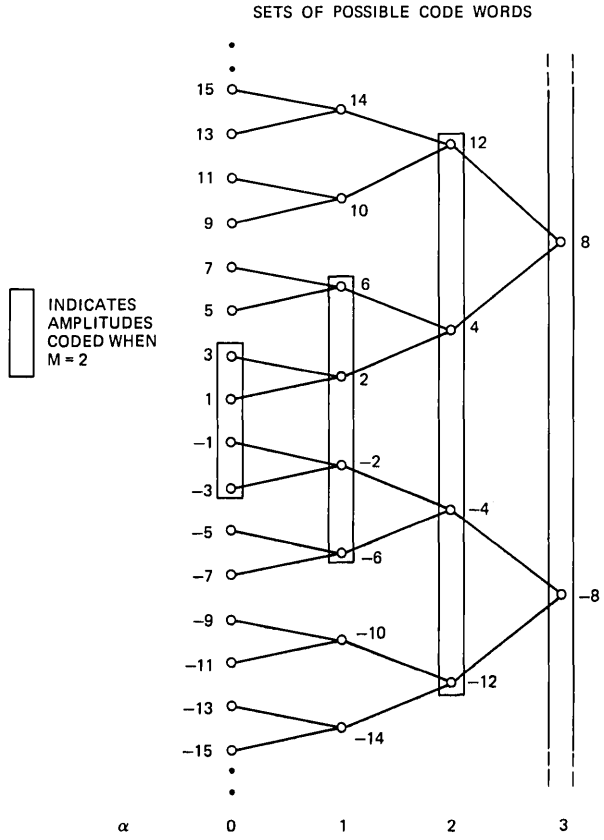

Fig. 3—Word generator roundoff procedure.

and the PCM code word consists of the $M - 1$ least significant bits of the truncated binary representation† of u_k and the sign bit. When

$$|u_k| > 2^{M+\alpha} - 1,$$

the transmitted code word is either the most positive or most negative M-bit word. A decoder recovers the integer code words of Fig. 3 by appending a one and α zeros to the least significant end of the PCM word.

With the truncation of each information bit, the step size increases by a factor 2 so that, with α information bits truncated, the PCM code

† In Section V, we point out an advantage of the twos complement binary format.

words represent signal levels in the set

$$\pm 2^{\alpha}\frac{d_o}{2},\ \pm 3\cdot 2^{\alpha}\frac{d_o}{2},\ \pm 5\cdot 2^{\alpha}\frac{d_o}{2},\cdots,\ \pm\ (2^M-1)\cdot 2^{\alpha}\frac{d_o}{2} \tag{5}$$

with step size

$$d = 2^{\alpha}d_o = 2^{\alpha}\frac{2\delta}{I}\gamma \tag{6}$$

and maximum amplitude $(2^M - 1)2^{\alpha}\delta\gamma/I$.

IV. UNIFORM DIGITAL FILTERS

The value of I, the dc gain of the digital filter, is crucial in determining the character of the overall PCM quantizing noise. With the filter coefficients all integers, I may be regarded as a measure of coefficient quantization. A large value of I corresponds to fine quantization because it allows considerable freedom in choosing g_i. To obtain a filter transfer function that approximates with arbitrary accuracy the optimum transfer function with respect to granular noise,[1] an arbitrarily high value of I is required. On the other hand, amplitude overload noise increases rapidly with I because the dynamic range of the encoder is nearly proportional to I^{-1}.

The rapid increase in amplitude overload noise as a function of I leads us to focus our attention on the uniform filter,

$$g_i = 1;\quad i = 0, 1, \cdots, N-1, \tag{7}$$

for which $I = N$, resulting in the greatest dynamic range attainable with an Nth order filter with all coefficients of the same polarity. (We exclude from consideration filters with $g_i = 0$ for one or more i.) Reference 1 suggests that, for high sampling rates, the coefficients of the optimum filter with respect to granular noise are nearly equal and that the difference in granular noise rejection between this optimum filter and the uniform filter is marginal. This observation suggests that encoders with uniform filters, because they minimize amplitude overload noise and produce near minimal granular noise, are nearly optimal with respect to total quantizing noise. Further support for this speculation is given later.

In the frequency domain, the uniform filter transfer function is $\sin(\pi N f/f_s)/\sin(\pi f/f_s)$ and the filter rejects increasing amounts of ΔM granular noise as N increases. So long as f_s/N is large relative to $2W$, the signal component of $x(t)$ is undistorted by the filter; but, as f_s/N approaches $2W$, distortion of in-band signal components becomes

significant, and overall performance deteriorates with increasing N. Thus, if the advantages of very high-order filtering are sought, designs more sophisticated than eq. (7) are required.

V. IMPLEMENTATION

Besides possessing noise-rejection properties, uniform filters admit considerable hardware economies relative to other designs. With all coefficients unity, no multiplication is required, and each filter output is merely the sum of N successive counter levels. Therefore, one may implement the uniform filter as a resettable accumulator, thereby eliminating the delay line of Fig. 1, as well as the multipliers. To obtain a PCM sample, the coder sets the accumulator to the current level of the up/down counter and adds to the accumulator the next $N - 1$ counter levels.

Because the addition of N numbers is required only once for each PCM sample, and because $f_s/2W$, the number of ΔM samples per PCM sample, is generally much greater than N, it is possible to time-share a single accumulator among many signal channels. With inputs presented to the accumulator at the ΔM rate, the number of channels sharing a single accumulator may be as high as $f_s/2WN$. Hence, in terms of hardware requirements, the filter order, N, determines time-sharing capacity rather than the number of circuits necessary to realize a single encoder.

In addition to adding counter levels into an accumulator and truncating least significant bits of the sum, the encoder must detect amplitude overload and generate the most positive or most negative code word when the word generator is overloaded.

It must also restore the proportionality of the counter level, w_k, to the ΔM approximation, $x(k\tau)$, after each instance of counter overload. The wrap-around property of twos complement arithmetic ensures this proportionality whenever $|x(k\tau)| < (2^{M+\alpha} - 1)\delta$. On the other hand, a saturating counter would require special measures to restore tracking after each instance of counter overload.

VI. ENCODER PERFORMANCE

6.1 *Figure of Merit and Design Specifications*

An ideal decoder of the encoder output sequence obtains $\hat{y}(t)$ (defined in Section 7.1), a delayed, noisy approximation to the analog input $y(t)$. We define the quantizing noise power of $\hat{y}(t)$ to be

$$N_T = E\{[\gamma\hat{y}(t) - y(t - t_0)]^2\} \tag{8}$$

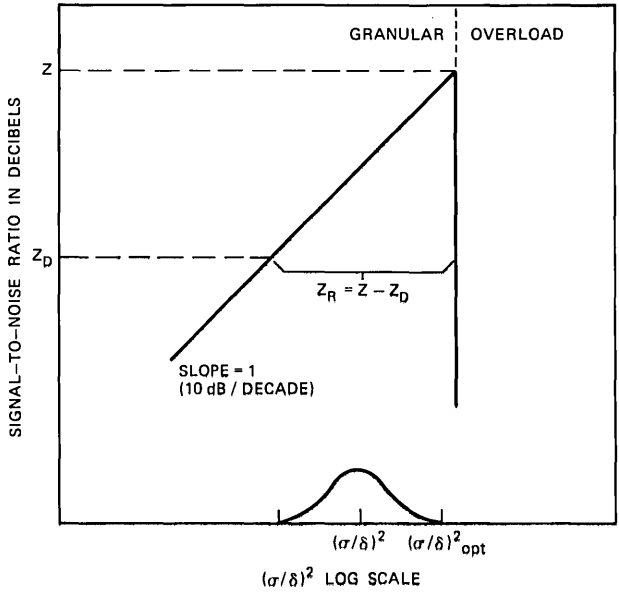

Fig. 4—Relationship of design objectives (Z_D,Z_R) to figure of merit (Z).

where E denotes expectation, t_0 is the encoder delay, and γ is the mean-square optimum scaling factor. For each digital filter, PCM word length, and ΔM sampling rate there is a unique combination of values of δ, the ΔM step size, and α, the word generator parameter, that results in minimal N_T. We choose as a figure of encoder merit the ratio of signal power to this minimum noise power,†

$$Z = \frac{E\{[y(t - t_0)]^2\}}{\min\limits_{\alpha,\delta} N_T}. \tag{9}$$

In the design of a practical encoder, typical specifications include a signal-to-noise ratio design goal, Z_D, and a range, Z_R, of input powers over which the actual signal-to-noise ratio must equal or exceed Z_D. The practical significance of our figure of merit is found in the approximation

$$Z \approx Z_D + Z_R \tag{10}$$

† N_T is a convex function of δ and α. In our numerical work we have used simple search techniques to find $\min_{\alpha,\delta} N_T$.

where each quantity is measured in decibels. For example, an encoder for which $Z = 55$ dB will actually attain this signal-to-noise ratio for a single level of input power, and will maintain a signal-to-noise ratio of 35 dB or better over a range of 20 dB in signal power.

Equation (10) is derived from Fig. 4, an approximation to the dependence of signal-to-noise ratio on input level. If (σ/δ) exceeds $(\sigma/\delta)_{\text{opt}}$, the optimum ratio of rms input to ΔM step size, overload noise predominates in the distortion and the signal-to-noise ratio falls rapidly as σ^2 increases. On the other hand, with $(\sigma/\delta) < (\sigma/\delta)_{\text{opt}}$, granular noise predominates and, with δ fixed, is essentially independent of σ. Hence, the signal-to-noise ratio is proportional to σ^2 in the granular region.

If the ensemble of input power levels is log-normally distributed, as in models used for speech signals,[6] $10 \log (\sigma/\delta)^2$ is a normal random variable, the mean value of which we denote by $10 \log (\bar{\sigma}/\delta)$. Hence, the probability that the signal-to-noise ratio exceeds Z_D is maximum when δ_D, the design value of the step size, is chosen such that

$$10 \log (\bar{\sigma}/\delta_D)^2 = 10 \log (\sigma/\delta)^2_{\text{opt}} - \tfrac{1}{2} Z_R. \tag{11}$$

That is, $10 \log (\bar{\sigma}/\delta_D)^2$ is the midpoint of the design range of length Z_R.

6.2 *Performance Characteristics*

Figure 5 shows a typical set of performance curves, computed according to the theory presented in Section VII. The curves pertain to 11-bit encoding of Gaussian signals having a truncated first-order Butterworth power spectrum, where the ratio of 3-dB frequency to cutoff frequency is 0.25. This type of process has been used to model band-limited speech.[7] The performance curves show the figure of merit, Z, of uniform filter encoders of various orders as a function of $f_s/2W$, the ΔM sampling rate expressed as a multiple of the PCM rate.

The choice of a specific encoder configuration represents a compromise between the advantages of low ΔM speed and low filter order. The nature of this compromise is illustrated in Fig. 6, which shows combinations of ΔM speed and filter order that satisfy two quality objectives. The broken curves relate ΔM speed to the maximum filter order consistent with sharing the accumulator described in Section V among 24, 48, and 96 signal channels, respectively. All design points to the right of a broken line are permissible for the given number of multiplexed channels.

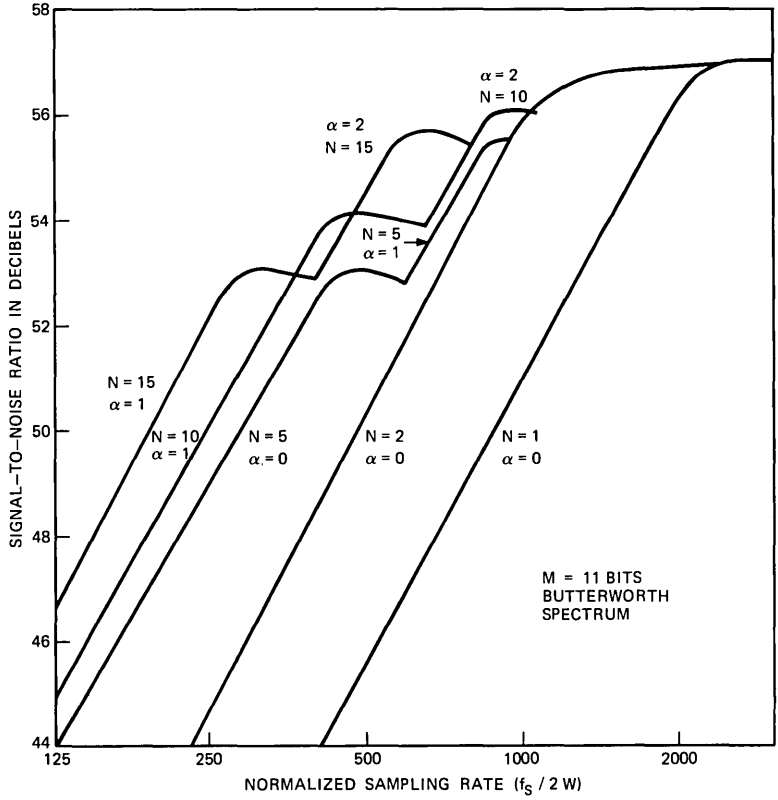

Fig. 5—Performance characteristics, $M = 11$ bits, Butterworth spectrum.

6.3 *Dependence of Performance on Design Parameters*

Figure 5 demonstrates two types of variation of Z with f_s: Z rising with a slope of 20 dB/decade, and Z flat or decreasing slowly with f_s. The first type of behavior occurs when amplitude overload is negligible and slope overload controls the optimum ΔM step size. In this case, the optimum step size varies approximately as $1/f_s$, and the decrease continues until amplitude overload becomes significant. When amplitude overload is the predominant overload noise, the optimum step size is essentially constant and the slightly negative slope of Z indicates that an increase in f_s results in an increase in the granular noise correlation from sample to sample, leading to a greater proportion of the ΔM granular noise power in the passband of the filter.

The flat portions of the curves represent transition regions to higher

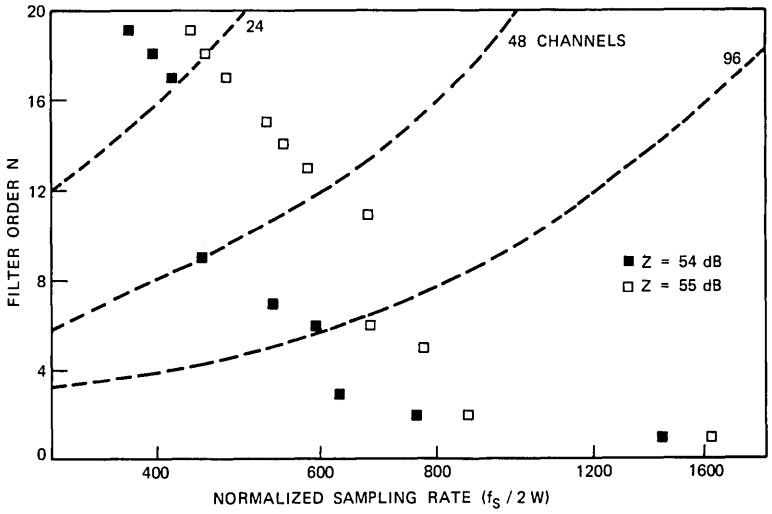

Fig. 6—Design alternatives derived from Fig. 5, $M = 11$ bits, Butterworth spectrum.

values of α, at which the increased roundoff noise of the word generator is offset by improved immunity to amplitude overload. As f_s increases indefinitely, Z approaches the maximum signal-to-noise ratio associated with uniform PCM encoding of Gaussian signals.

In Figs. 7 and 8, we see that the shapes of the characteristic curves are essentially invariant with the number of bits in the PCM code. In Fig. 7, which pertains to 11-bit encoding of signals with a flat spectrum, amplitude overload effects occur at points that are approximately $10 \log (2^{11}/2^8) = 18$ dB higher in signal-to-noise ratio and further to the right by the factor $2^{11}/2^8$ in sampling rate, relative to the corresponding points in Fig. 8, which pertains to 8-bit encoding of the same input process.

Figures 5, 7, and 8 also demonstrate the effect of filter order. When f_s is quite low and amplitude overload is negligible, Z increases monotonically with N. However, the value of f_s at which amplitude overload becomes significant decreases as N increases, and the earlier transitions from $\alpha = 0$ to $\alpha = 1$, $\alpha = 1$ to $\alpha = 2$, etc., lead to the crossovers.

Figures 5 and 7 relate to the same PCM word length but different signal spectra. The principal difference between the two sets of curves is a scale change of the horizontal axis. In Fig. 5, the axis is shifted to the left relative to Fig. 7 by the factor 1.6, which is the ratio by which

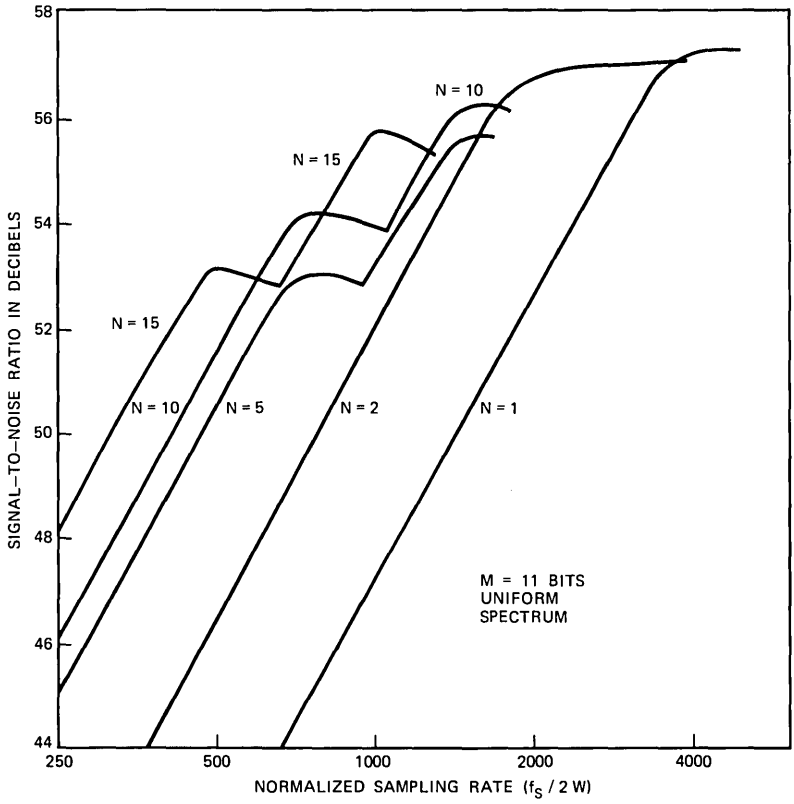

Fig. 7—Performance characteristics, $M = 11$ bits, uniform spectrum.

the rms slope of signals having the uniform spectrum exceeds the rms slope of signals having the Butterworth spectrum.

VII. QUANTIZING NOISE ANALYSIS

7.1 *Noise Components*

To reconstruct an analog signal from the sequence of word generator outputs, we first recover one of the integers shown in the αth column of Fig. 3 by appending a one and α zeros to the least significant end of each code word. We next multiply the sequence of integers by the nominal scale factor δ/I and denote the resulting sequence by $\{\hat{y}_j\}$. Finally, we perform ideal interpolation of $\{\hat{y}_j\}$ to obtain the continuous waveform

$$\hat{y}(t) = \sum_{j=-\infty}^{\infty} \hat{y}_j \frac{\sin 2\pi W(t - j/2W)}{2\pi W(t - j/2W)}. \tag{12}$$

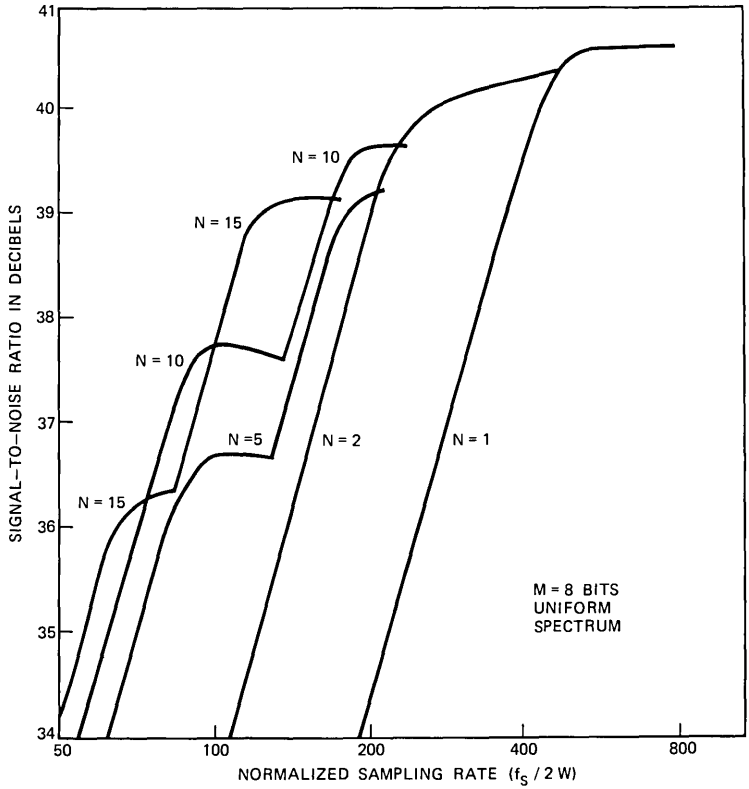

Fig. 8—Performance characteristics, $M = 8$ bits, uniform spectrum.

Our purpose, in this section, is to investigate the difference between $\hat{y}(t)$ and the encoder input, $y(t)$, when this input is a sample function of a zero-mean stationary Gaussian random process. Our measure of distortion is the total quantizing noise power, N_T, defined in eq. (8). Because $y(t)$ is stationary, N_T is independent of time, and in the sequel we omit time arguments and subscripts from the notation of signals when there is no risk of ambiguity. Although γ in eq. (8) is a complicated function of signal statistics and encoder design parameters, the introduction of the preliminary scaling factor, δ/I, leads to $\gamma \approx 1$ in situations of greatest interest. (See Section 7.5.) The other constant in eq. (8) is t_0, and we observe that if the filter coefficients have even symmetry ($g_i = g_{N-1-i}$), the filter delay is

$$t_0 = (N - 1)\tau/2, \tag{13}$$

one-half the filter memory span.

To derive N_T as a function of the encoder design parameters, we recognize $\hat{y}$ as the sum of a signal term and four noise terms. We begin by writing

$$\hat{y} = \frac{\delta}{I}(u + n_w) \tag{14}$$

where u is the filter output and n_w represents the difference between the input and output of the word generator. In studying $(\delta/I)u$, the filtered ΔM signal, it is customary to identify granular and slope overload components of the ΔM quantizing noise, in the manner indicated in Fig. 2. If we rewrite eq. (1) as $u = g * (x/\delta)$, with $*$ denoting convolution and x denoting the approximation signal in the delta modulator feedback loop, we obtain

$$\frac{\delta}{I}u = \frac{g}{I} * x = \frac{g}{I} * (y + n_G + n_S) \tag{15}$$

where n_G and n_S are ΔM granular and slope overload noise, respectively.

The distortion introduced by the word generator, $(\delta/I)n_w$, may itself be resolved into two components, namely, n_A, which accounts for amplitude overload, and n_R, which represents the roundoff effect. These observations lead us to a representation of the quantizing noise signal as the sum of four noise components:

$$(\gamma\hat{y} - y) = \left[\frac{\gamma g}{I} * (y + n_G) - y\right] + \frac{\gamma g}{I} * n_S + \gamma n_A + \gamma n_R. \tag{16}$$

The term in square brackets is the filtered granular noise, and the remaining terms are slope overload noise, amplitude overload noise, and word generator roundoff noise, respectively. In this paper we evaluate the expected square of eq. (16) by assuming that the expected product of each pair of terms is negligible relative to the sum of the two mean square values. Thus we express the total quantizing noise as the sum of four noise powers

$$N_T = N_G + \gamma^2 N_S + \gamma^2 N_A + \gamma^2 N_R \tag{17}$$

in which each term is the expected square of a term in eq. (16).

When the total quantizing noise is low, we are justified in approximating the average cross products of eq. (16) by zero because: (*i*) granular noise and roundoff noise are zero during overload bursts; (*ii*) each type of overload occurs with low probability and the probability of their joint occurrence is negligible; and (*iii*) we have found that $|E(n_G n_R)|$ is many orders of magnitude lower than $N_G + N_R$ when $M \geqq 4$.

7.2 *ΔM Granular Noise and Slope Overload Noise*

Expanding the square of the term in brackets in eq. (16), we obtain

$$N_G = E(y^2) - 2\frac{\gamma}{I}\sum_{i=0}^{N-1} g_i R_{xy}(i\tau - t_0) + \left(\frac{\gamma}{I}\right)^2 \sum_{i=0}^{N-1}\sum_{j=0}^{N-1} g_i g_j R_{xx}(i\tau - j\tau) \quad (18)$$

in which the ΔM correlation functions,

$$R_{xy}(\tau) = E[y(t)x(t+\tau)], \qquad R_{xx}(\tau) = E[x(t)x(t+\tau)]$$

are derived in Ref. 8, under the assumption that overload never occurs. To use the results of Ref. 8 and also account for overload, we should multiply N_G in eq. (18) by the probability that overload is absent. For the applications that interest us, the probability is greater than 0.99, and we adopt eq. (18) as an approximation to N_G that overestimates the granular noise component of N_T by no more than one percent. In Section 7.4, we similarly overestimate N_R.

To compute N_S, we adopt the assumption of previous authors[7,9] that essentially all of the ΔM slope overload noise power is in the signal band of $y(t)$ so that $N_S = E\{[(g/I) * n_S]^2\} \approx E(n_S^2)$. In our numerical analysis, we have followed Protonotarios,[9] who derives $E(n_S^2)$ as a function of

$$S = \frac{\delta}{\tau}\left\{E\left[\frac{dy}{dt}\right]^2\right\}^{-\frac{1}{2}}, \quad (19)$$

the ratio of the maximum slope of $x(t)$ to the rms slope of $y(t)$. For high values of S, N_S is proportional to $S^{-5}\exp[-\frac{1}{2}S^2]$.

7.3 *Amplitude Overload Noise*

The maximum output of the word generator is $2^\alpha(2^M - 1)$. Assuming there is no granular or slope overload noise during amplitude overload intervals,

$$\begin{aligned} n_A &= \frac{g}{I} * y - \frac{\delta}{I}2^\alpha(2^M - 1); && g * y > 2^\alpha(2^M - 1)\delta \\ &= \frac{g}{I} * y + \frac{\delta}{I}2^\alpha(2^M - 1); && g * y < -2^\alpha(2^M - 1)\delta \\ &= 0; && |g * y| \leqq 2^\alpha(2^M - 1)\delta. \end{aligned} \quad (20)$$

Because $(g/I) * y$ is a sample function of a zero-mean Gaussian process

with variance

$$\sigma_F^2 = \frac{1}{I^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g_i g_j R_{yy}(i\tau - j\tau), \tag{21}$$

the mean square value of eq. (20) is

$$N_A = (2/\pi)^{\frac{1}{2}} \int_A^\infty \sigma_F^2 (x - A)^2 \exp(-\tfrac{1}{2}x^2) dx$$

$$= \sigma_F^2 \{(1 + A^2) \operatorname{erfc}(2^{-\frac{1}{2}}A) - (2/\pi)^{\frac{1}{2}} A \exp(-\tfrac{1}{2}A^2)\} \tag{22}$$

in which A is the amplitude overload factor,

$$A = \frac{2^\alpha (2^M - 1)\delta}{I\sigma_F}. \tag{23}$$

7.4 *Word Generator Roundoff Noise*

With u the filter output and $z(\cdot)$ the mapping shown in Fig. 3,

$$N_R = \frac{\delta^2}{I^2} E[(z - u)^2]. \tag{24}$$

Because u is an odd integer, we have the binary number representation of $u > 0$,

$$u = \sum_{i=1}^{\infty} b_i 2^i + 1$$

where $b_i = 0$ or 1. The word generator truncates $b_\alpha b_{\alpha-1} \cdots b_1 1$ from this representation and z is obtained by replacing these digits with $10 \cdots 0 = 2^\alpha$.

Hence

$$z = \sum_{i=\alpha+1}^{\infty} b_i 2^i + 2^\alpha \tag{25}$$

and

$$z - u = 0; \qquad \alpha = 0$$

$$= 2^\alpha - 1 - \sum_{i=1}^{\alpha} b_i 2^i; \qquad \alpha \geqq 1. \tag{26}$$

For $u < 0$ the odd symmetry of Fig. 3 implies $z(u) = -z(-u)$. Hence $z - u$ is an odd integer in $[-(2^\alpha - 1), 2^\alpha - 1]$ and the ex-

pectation in eq. (24) is a weighted average of the integers $1^2, 3^2, \cdots, (2^\alpha - 1)^2$. From this observation, we immediately obtain the bounds

$$\frac{\delta^2}{I^2} \leqq N_R \leqq \frac{\delta^2}{I^2}(2^\alpha - 1)^2. \tag{27}$$

With $\alpha = 1$, the bounds are equal and $N_R = \delta^2/I^2$.

For $\alpha > 1$, we evaluate N_R only for coders with uniform digital filters. All odd integers are possible outputs of such filters. That is, $\Pr\{u = 2n + 1\} > 0$ for all n and, for low-noise encoding, this probability is quite nearly constant over a set of $2^{\alpha-1}$ consecutive integers. When scaled to the amplitude range of $y(t)$, such a set of filter outputs lies in an interval of length

$$\frac{2\delta}{I} 2^{\alpha-1}\gamma = d/2, \tag{28}$$

a small fraction of σ (typically of the order of $4\sigma/2^M$). For $M \geqq 5$, the envelope of $\Pr\{u = 2n + 1\}$ has approximately the Gaussian shape of the probability density of $y(t)$ and a piecewise constant approximation to this density over intervals of length d or less leads to highly accurate expressions for quantizing noise power.[10]

Over intervals of length $d/2$, $(z - u)^2$ takes on all the values $1^2, 3^2, \cdots (2^\alpha - 1)^2$ either in ascending or descending order. Hence, the piecewise constant approximation to $\Pr\{u = 2n + 1\}$ reduces the expectation in eq. (24) to an unweighted average of these $2^{\alpha-1}$ integers,

$$N_R = \frac{\delta^2}{I^2} 2^{-(\alpha-1)} \sum_{i=1}^{2^{\alpha-1}} (2i - 1)^2 = \frac{\delta^2}{I^2} \frac{(4^\alpha - 1)}{3}. \tag{29}$$

Noting that eq. (6) admits the expression

$$\gamma^2 \left(\frac{\delta}{I}\right)^2 = \frac{d^2}{4(4^\alpha)},$$

we summarize the results of this section as follows:

$$\begin{aligned} \gamma^2 N_R &= 0; && \alpha = 0 \\ &= \frac{d^2}{16}; && \alpha = 1 \\ &= \frac{d^2}{12}(1 - 4^{-\alpha}); && \alpha \geqq 0, \quad \text{uniform filter encoders.} \end{aligned} \tag{30}$$

The last line indicates that $N_R \to d^2/12$ as $\alpha \to \infty$. This limit is the granular noise associated with instantaneous PCM encoding of samples ranging over the continuum.

7.5 *The Optimum Scale Factor*

To complete our quantizing noise analysis and establish the validity of prescaling the word generator output by δ/I, we show that γ, the additional scaling factor that brings the amplitude of $\hat{y}(t)$ optimally close to that of $y(t)$ in the mean square sense, is nearly unity in designs of practical interest. Specifically, we derive the inequalities,

$$(1-b)\gamma_0 < \gamma < \gamma_0 \tag{31}$$

where b is of the order of magnitude of the noise-to-signal ratio and

$$\gamma_0 = \frac{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} g_i g_j R_{yy}(i\tau - t_0)}{\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} g_i g_j R_{yy}(i\tau - j\tau)}. \tag{32}$$

Clearly as $f_s \to \infty$, all of the covariances in eq. (32) approach σ^2 and hence $\gamma_0 \to 1$. In all of the numerical examples considered in Section VI, the sampling rates are so high that γ_0 is quite nearly unity. For example, all of the points plotted in Fig. 6 correspond to $0.99 < \gamma_0 < 1.01$.

By definition, γ is the value of c that minimizes

$$E[(c\hat{y} - y)^2] = E\left[\left(\frac{cg}{I} * (y + n_G) - y + cn_0\right)^2\right]. \tag{33}$$

The expression in square brackets on the right side is identical to eq. (16) with c replacing γ and cn_0 replacing the last three terms. Because we have assumed the correlation of n_0 and $[(cg/I) * (y + n_G) - y]$ to be zero, we may rewrite

$$E[(c\hat{y} - y)^2] = E\left[\frac{cg}{I} * (y + n_G) - y\right]^2 + c^2E(n_0^2). \tag{34}$$

Differentiating this equation with respect to c, and equating to γ the value of c that causes the derivative to be zero, we obtain

$$\gamma = \frac{E\left[y \frac{g}{I} * (y + n_G)\right]}{E\left[\frac{g}{I} * (y + n_G)\right]^2 + E(n_0^2)}$$

$$= \frac{\frac{1}{I} \sum_{i=0}^{N-1} g_i R_{xy}(i\tau - t_0)}{\frac{1}{I^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g_i g_j R_{xx}(i\tau - j\tau) + N_S + N_A + N_R}. \quad (35)$$

Reference 8 demonstrates that, for low-noise encoding, the approximations $R_{xy} \approx R_{yy}$ and $R_{xx} \approx R_{yy} + R_{ee}$ [where $R_{ee}(\cdot)$ is the autocorrelation function of the unfiltered granular quantizing noise] are extremely precise. Thus eq. (35) becomes

$$\gamma = \frac{\frac{1}{I} \sum_{i=0}^{N-1} g_i R_{yy}(i\tau - t_0)}{\frac{1}{I^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g_i g_j R_{yy}(i\tau - j\tau) + N_G' + N_S + N_A + N_R} \quad (36)$$

where

$$N_G' = \frac{1}{I^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} g_i g_j R_{ee}(i\tau - j\tau)$$

and is of the order of magnitude of N_G.[1] We now recognize that the first term in the denominator of eq. (36) is σ_F^2 [eq. (21)], the power of the filtered signal component of the ΔM approximation. If we divide numerator and denominator by this term (and substitute $\sum_i g_i$ for I) we obtain

$$\gamma = \frac{\gamma_0}{1 + b} \quad (37)$$

where we have defined

$$b = (N_G' + N_A + N_S + N_R)/\sigma_F^2 \quad (38)$$

which is the order of magnitude of N_T/σ^2. Equation (31) follows immediately from eq. (37).

VIII. SUMMARY AND CONCLUSION

Section VII presents the analytical steps enabling us to compute the figure of merit, Z, of a ΔM/PCM encoder. The rationale for using

this figure of merit is presented in Section VI (Fig. 4) along with results for some special cases of interest (Figs. 5, 7, and 8). The ultimate utility of these results is that they enable the designer to determine tradeoffs between ΔM sampling speed and digital filter order for specified values of encoder quality (e.g., Fig. 6).

We should reiterate the conditions assumed for the encoder in deriving our results. First, we have assumed that the digital filter output is at odd parity at every PCM sampling instant. Aside from simplifying the roundoff noise analysis, this condition appears to correspond to the simplest possible implementation of the PCM word generator. The primary design constraint it imposes is the prohibition of digital filters of orders 4, 8, 12, etc.

Second, we have assumed that the digital filter has uniform coefficients. This condition makes a complete noise analysis relatively straightforward and also leads to a simple filter implementation. Furthermore, it corresponds to a robust design that appears to be near optimal in all cases of practical interest. Attempts to demonstrate the latter point quantitatively have foundered on the difficulty of assessing the roundoff noise power (N_R) when the filter coefficients are nonuniform. If we assume that, for any step size, d, the roundoff noise

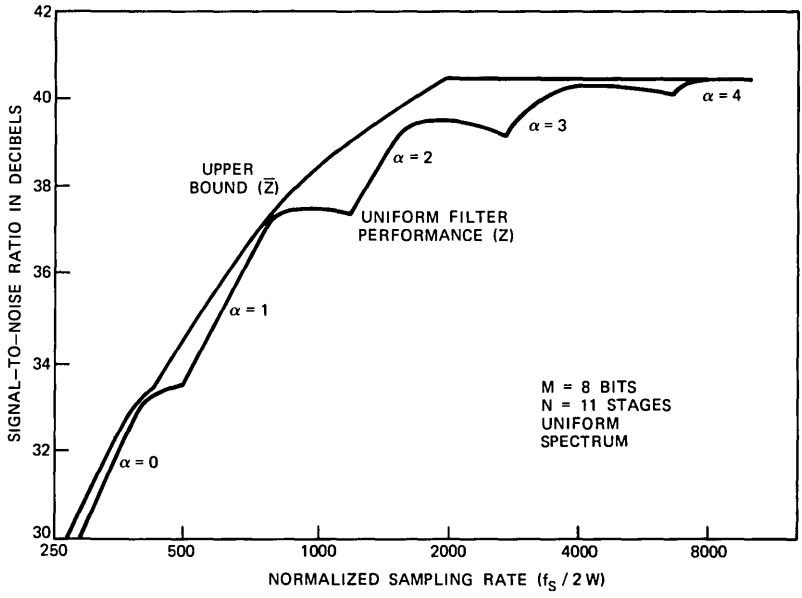

Fig. 9—Uniform filter assessment.

power is nondecreasing with the number of bits truncated, we can derive an upper bound on Z. A comparison of this upper bound, $\bar{Z}$, with the figure of merit resulting from uniform coefficients is shown in Fig. 9 for one case. Comparable results can be expected for other combinations of signal spectrum, PCM word length, and filter order.

APPENDIX

Odd-Parity Filter Outputs

With the encoder delay equal to one-half the filter memory span [eq. (13)], we consider symmetric coefficient sets

$$\begin{aligned} g_i &= g_{N-1-i}, \qquad i = 0, 1, \cdots, \tfrac{1}{2}N - 1; \qquad N \text{ even} \\ g_i &= g_{N-1-i}, \qquad i = 0, 1, \cdots, \tfrac{1}{2}(N-3); \qquad N \text{ odd}. \end{aligned} \tag{39}$$

Such coefficients give equal weight to counter levels that are equally advanced or retarded with respect to $y(k\tau - t_0)$, the input value estimated at $t = k\tau$. Equation (39), when combined with eq. (1), implies

$$u_k = \sum_{i=0}^{\frac{1}{2}N-1} g_i[w_{k-i} + w_{k-(N-1)+i}]; \qquad N \text{ even} \tag{40}$$

$$u_k = g_{\frac{1}{2}(N-1)} w_{k-\frac{1}{2}(N-1)} + \sum_{i=0}^{\frac{1}{2}(N-3)} g_i[w_{k-i} + w_{k-(N-1)+i}]; \qquad N \text{ odd}. \tag{41}$$

With the counter levels, w_i, alternating in parity, the two counter levels in square brackets in eq. (40) are of opposite parity because the difference in subscripts, $N - 1 - 2i$, is an odd number. Hence, their sum is odd. On the other hand, the two corresponding counter levels in eq. (41) have the same parity, and thus an even sum, because $N - 1 - 2i$ is even with N odd. These observations lead to the following necessary and sufficient conditions for u_k ranging over the set of odd integers:

Condition A: With N even, u_k is odd if and only if there is an odd number of odd coefficients in the set $g_0, g_1, \cdots, g_{\frac{1}{2}N-1}$.

Condition B: With N odd, u_k is odd at a PCM sampling instant if and only if $g_{\frac{1}{2}(N-1)}$ is odd and the low-speed clock is synchronized so that PCM sampling instants occur when $w_{k-\frac{1}{2}(N-1)}$ is odd. This synchronization can be achieved if the ratio of ΔM sampling rate to PCM sampling rate is an even integer.

In uniform filter encoders, Condition A is always satisfied when $N = 2, 6, 10$, etc. It can never be satisfied with $N = 4, 8, 12$, etc. For Condition B to be satisfied, the encoder must be synchronized such that w_{k-N+1} (the first term entering the accumulator described in Section V) is odd when $N = 1, 5, 9$, etc; w_{k-N+1} must be even when $N = 3, 7, 11$, etc.

REFERENCES

1. Goodman, D. J., "The Application of Delta Modulation to Analog-to-PCM Encoding," B.S.T.J., *48*, No. 2 (February 1969), pp. 321–343.
2. Freeny, S. L., Kieburtz, R. B., Mina, K. V., and Tewksbury, S. K., "Design of Digital Filters for an All Digital Frequency Division Multiplex-Time Division Multiplex Translator," IEEE Trans. Circuit Theory, *CT-18*, No. 6 (November 1971), pp. 702–711.
3. Freeny, S. L., Kieburtz, R. B., Mina, K. V., and Tewksbury, S. K., "Systems Analysis of a TDM-FDM Translator/Digital A-Type Channel Bank," IEEE Trans. Com. Tech., *COM-19*, No. 6 (December 1971), pp. 1050–1059.
4. Ishiguro, T., and Kaneko, H., "A Nonlinear DPCM Codec Based on ΔM/ΔPCM Code Conversion with Digital Filter," ICC Conference Record, Montreal (June 1971), pp. 1–27 to 1–32.
5. Kaneko, H., "A Unified Formulation of Segment Companding Laws and Synthesis of Codecs and Digital Compandors," B.S.T.J., *49*, No. 7 (September 1970), pp. 1555–1588.
6. Purton, R. F., "A Survey of Telephone Speech Signal Statistics and Their Significance in the Choice of a PCM Companding Law," Proc. IEEE, *109*, Part B (January 1962), pp. 60–66.
7. O'Neal, J. B., Jr., "Delta Modulation Quantizing Noise Analytical and Computer Simulation Results for Gaussian and Television Input Signals," B.S.T.J., *45*, No. 1 (January 1966), pp. 117–141.
8. Goodman, D. J., "Delta Modulation Granular Quantizing Noise," B.S.T.J., *48*, No. 5 (May–June 1969), pp. 1197–1218.
9. Protonotarios, E. N., "Slope Overload Noise in Differential Pulse Code Modulation," B.S.T.J., *46*, No. 9 (November 1967), pp. 2119–2162.
10. Cattermole, K. W., *Principles of Pulse Code Modulation*, New York: American Elsevier Publishing Co., 1969, Chapter 3.1.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 2, February, 1973
Printed in U.S.A.

Scattering Losses Caused by the Support Structure of an Uncladded Fiber

By D. MARCUSE

(Manuscript received September 5, 1972)

We consider an uncladded dielectric waveguide core that is held by dielectric supports. The radiation losses caused by the support structure are being considered. The analysis is simplified by using a slab waveguide model held by slab-shaped supports. Only order-of-magnitude estimates are attempted. The radiation losses can be reduced by reducing the refractive index contrast between the supports and the surrounding medium with the help of an index matching liquid. The radiation losses remain large unless the index match is sufficiently close.

I. INTRODUCTION

A typical optical fiber consists of a core, the refractive index of which is larger than that of the cladding material surrounding the core.[1] The cladding serves the purpose of keeping any outside influence, such as dust, at a safe distance from the core. The requirement of a lower refractive index for the cladding makes it difficult to find suitable cladding materials for one of the most promising core materials–fused silica. Fused silica is particularly useful as fiber core material because of its inherently low absorption loss. However, the refractive index of fused silica $n = 1.46$, is lower than that of most other glasses. In particular, there are as yet no low-loss materials suitable for fiber claddings, the refractive indices of which are lower than that of fused silica. The few available materials have absorption losses that rule out their use as claddings for low-loss optical fiber waveguides.

It appears natural to ask whether a fiber without cladding could not be made. In principle a dielectric fiber waveguide works just as well without a cladding if it could be suspended in air or in vacuum. But since no method of levitating the fused silica fiber without mechanical supports has yet been devised, the necessity exists of holding the uncladded core by some kind of supporting structure. If the supports of

the fiber core have a refractive index lower than that of the naked core, the amount of light scattered out of the core by contact with the supports may be tolerable. Since the supports touch the naked core only occasionally, their dielectric losses need not be exceptionally small. Solid materials of lower index than fused silica, but with much higher losses, do exist.

The scattering losses of the supports could be reduced by submerging the core and the support structure in a low-loss index matching liquid.[2] If it were possible to match the refractive index of the supports perfectly, the submersion technique could eliminate all scattering losses from the supports so that only the losses of the index matching liquid and the heat loss of the fiber supports would remain. The heat losses of the supports would equal their average value averaged over the entire length of the guide. Because of the low filling factor, this average loss could be tolerably small. However, even this index matching technique encounters certain problems. It is unlikely that a low-loss liquid could be found that matches the refractive index of the support structure perfectly. But even if this were possible, the index match would work only at one fixed temperature so that scattering losses would still occur if the ambient temperature drifts from the design value.

In order to explore the requirements for low scattering losses of a partially index-matched support structure, a model calculation is carried out in this paper. For simplicity we limit the discussion to the TE modes of the slab waveguide in the hope of obtaining order-of-magnitude estimates. The model to be investigated is shown in Fig. 1.

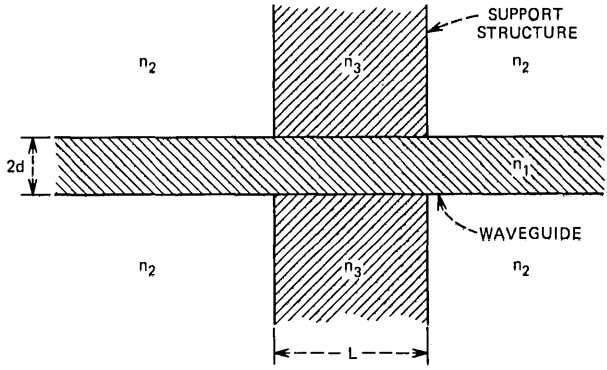

Fig. 1—Dielectric slab waveguide core of refractive index n_1 supported by two slabs of index n_3.

The dielectric slab of refractive index n_1 is embedded in a medium of refractive index n_2. This surrounding medium can be thought of either as air or as a suitable index matching liquid. The support structure is simulated by two dielectric slabs attached at right angles to the waveguide core. The refractive index of these support pieces is n_3. The shape of the model supports does not resemble the shape to be expected of actual supports. However, it is hoped that this model will provide insight into the scattering losses to be expected from an actual support structure. Since only order-of-magnitude estimates are expected from this model, we can use rather crude approximations of the mathematical expressions. The idea for using a naked fiber core held by a dielectric support structure originated with R. Kompfner.

II. SCATTERING LOSS THEORY

Our calculation of the scattering losses caused by the waveguide supports is based on the usual expansion of the field of the dielectric waveguide in terms of normal modes.[1] We restrict ourselves to TE modes of the slab waveguide whose only electric field component E_y is tangential to the surface of the slab. The field is expanded in terms of normal guided and radiation modes of the slab waveguide

$$E_y = \sum_\nu C_\nu \mathcal{E}_\nu + \int_0^\infty q(\rho)\mathcal{E}(\rho)d\rho. \tag{1}$$

The expansion coefficients in (1) are obtained from[1]

$$C_\mu = C_\nu \int_0^L K_{\mu\nu}(z)e^{-i(\beta_\nu-\beta_\mu)z}dz. \tag{2}$$

An analogous expression holds for $q(\rho)$ with $K_{\rho\nu}$ instead of $K_{\mu\nu}$. It can be shown [see eqs. (9.2–21) and (9.2–30) of Ref. 1] that the coupling coefficient $K_{\mu\nu}$ is given by

$$K_{\mu\nu} = \frac{\omega\epsilon_0}{4P}\int_{-\infty}^{\infty}[n^2(x) - n_0^2(x)]\mathcal{E}_\nu\mathcal{E}_\mu^* dx. \tag{3}$$

$n_0(x)$ is the refractive index distribution of the ideal waveguide, $n_0(x) = n_1$ inside of the waveguide core and $n_0(x) = n_2$ outside of the core region. The index distribution $n(x)$ is defined as being $n(x) = n_1$ in the waveguide core and $n(x) = n_3$ in the region of the waveguide support. L is the width of the support and P is the power carried by the incident mode. ω and ϵ_0 are, respectively, the radian frequency of the field and the permittivity of vacuum.

The relative power loss from mode ν to mode μ caused by the support is given by[1]

$$\frac{\Delta P_{\mu\nu}}{P_\nu} = \frac{|C_\mu|^2}{|C_\nu|^2} = |K_{\mu\nu}|^2 \frac{4 \sin^2 (\beta_\nu - \beta_\mu) \dfrac{L}{2}}{(\beta_\nu - \beta_\mu)^2}. \tag{4}$$

In particular, this formula applies to the reflection loss R of the mode ν which is obtained by setting $\beta_\mu = -\beta_\nu$. The radiation loss of mode ν is similarly obtained from the formula[1]

$$\frac{\Delta P_s}{P_\nu} = \int_0^\infty \frac{|q(\rho)|^2}{|C_\nu|^2} d\rho = 4 \int_{-n_2 k}^{n_2 k} |K_{\rho\nu}|^2 \frac{\sin^2 (\beta_\nu - \beta) \dfrac{L}{2}}{(\beta_\nu - \beta)^2} \frac{|\beta|}{\rho} d\beta. \tag{5}$$

The integration variable in the integral on the extreme right-hand side was changed from ρ to β.

In order to evaluate the coupling coefficients, we need the field expressions for the guided and radiation modes only outside of the core region, because the integrand in (3) vanishes inside of the waveguide core. We have for $|x| > d$

$$\mathcal{E}_\nu = \left\{ \frac{2\gamma_\nu P}{(n_1^2 - n_2^2)\beta_\nu (1 + \gamma_\nu d)\omega\epsilon_0} \right\}^{\frac{1}{2}} k_{K_\nu} e^{-\gamma_\nu (|x| - d)}. \tag{6}$$

Equation (6) for the guided TE modes follows from eqs. (8.3–12) and (8.3–18) of Ref. 1, with the help of (8.6–16). The parameters γ_ν and κ_ν are related to the free-space propagation constant $k = 2\pi/\lambda$ in the following way

$$\gamma_\nu^2 = \beta_\nu^2 - n_2^2 k^2 \tag{7}$$

and

$$\kappa_\nu^2 = n_1^2 k^2 - \beta_\nu^2. \tag{8}$$

The field of the TE radiation modes is obtained from eqs. (8.4–4), (8.4–9), and (8.4–18) of Ref. 1.

$$\mathcal{E}_\rho = \left\{ \frac{\rho^2 k^2 P}{2\pi\omega\epsilon_0 |\beta| (\rho^2 \cos^2 \sigma d + \sigma^2 \sin^2 \sigma d)} \right\}^{\frac{1}{2}} \cdot \left\{ \left(\cos \sigma d - i \frac{\sigma}{\rho} \sin \sigma d \right) e^{-i\rho(|x| - d)} + \left(\cos \sigma d + i \frac{\sigma}{\rho} \sin \sigma d \right) e^{i\rho(|x| - d)} \right\}. \tag{9}$$

The coefficient for coupling between the guided modes ν and μ follows from (3) and (6)

$$K_{\mu\nu} = \frac{n_3^2 - n_2^2}{n_1^2 - n_2^2} \frac{\kappa_\nu \kappa_\mu (\gamma_\nu \gamma_\mu)^{\frac{1}{2}}}{(\gamma_\nu + \gamma_\mu)[|\beta_\nu \beta_\mu|(1 + \gamma_\nu d)(1 + \gamma_\mu d)]^{\frac{1}{2}}}. \tag{10}$$

The power reflection coefficient R_ν for the νth mode is obtained from (4) and (10) by setting $\beta_\mu = -\beta_\nu$, $\kappa_\mu = \kappa_\nu$, and $\gamma_\mu = \gamma_\nu$

$$R_\nu = \frac{(n_3^2 - n_2^2)^2 \kappa_\nu^4 \sin^2 \beta_\nu L}{4(n_1^2 - n_2^2)^2 \beta_\nu^4 (1 + \gamma_\nu d)^2}. \tag{11}$$

It is interesting to consider the reflection coefficient in the limit $\gamma_\nu = 0$ at the cutoff point of the guided mode. We obtain from (11) with $\beta_\nu = n_2 k$ and $\kappa_\nu^2 = (n_1^2 - n_2^2)k^2$

$$R_\nu = \frac{(n_3^2 - n_2^2)^2}{16 n_2^4} (4 \sin^2 n_2 k L). \tag{12}$$

The sine factor (multiplied by 4) on the right-hand side of this equation describes the interference between the reflection from the front and back surface of the support structure. If we omit this factor we obtain the reflection from only one of the two interfaces in the form

$$\bar{R}_\nu = \frac{(n_3^2 - n_2^2)^2}{16 n_2^4} = \frac{(n_3 + n_2)^2}{(2n_2)^2} \frac{(n_3 - n_2)^2}{(2n_2)^2}. \tag{13}$$

Comparison of eq. (13) with the correct expression for the power reflection coefficient from a dielectric interface [see eq. (1.6–36) of Ref. 1]

$$R = \left(\frac{n_3 - n_2}{n_3 + n_2}\right)^2 \tag{14}$$

provides an indication of the accuracy of our approximation. Instead of solving the infinite system of coupled equations, we obtained the coefficient of the reflected wave and hence the power reflection coefficient by considering only the incident and the reflected wave alone. The resulting equation (13) agrees with the correct equation (14) in the limit $(n_3 - n_2) \to 0$. The approximate solution of the infinite equation system is thus a good approximation only for small index differ-

ences. However, it is certainly quite satisfactory as an order-of-magnitude estimate.

The coupling coefficient for mode ν and a radiation mode characterized by the parameter ρ follow from (3), (6), and (9)

$$K_{\rho\nu} = \frac{n_3^2 - n_2^2}{(n_1^2 - n_2^2)^{\frac{1}{2}}} \cdot \frac{k\kappa_\nu(\gamma_\nu)^{\frac{1}{2}}\rho(\gamma \cos \sigma d - \sigma \sin \sigma d)}{(\beta_\nu^2 - \beta^2)[\pi|\beta\beta_\nu|(1 + \gamma_\nu d)(\rho^2 \cos^2 \sigma d + \sigma^2 \sin^2 \sigma d)]^{\frac{1}{2}}}. \quad (15)$$

The radiative power loss caused by one support follows from (5) and (15). The integral cannot be solved exactly. Instead of resorting to numerical integration, I worked out an approximate solution which holds only as an order-of-magnitude estimate.

$$\frac{\Delta P}{P} = \frac{2}{3\pi} \frac{(n_3^2 - n_2^2)^2}{(n_1^2 - n_2^2)^{\frac{3}{2}}} \frac{k\kappa_\nu^2\gamma_\nu^3}{\beta_\nu(1 + \gamma_\nu d)(\beta_\nu + n_2 k)^2(\beta_\nu - n_2 k)^3}. \quad (16)$$

The radiation loss depends on the width L of the support structure. For large values of L/λ, the width dependence disappears. The approximation (16) holds in this limit. Comparison of (16) with a few sample values of the numerical integration has shown that this approximation can depart from a few percent to as much as 50 percent from the value obtained by integrating (5).

III. DISCUSSION AND NUMERICAL RESULTS

The radiation losses caused by the waveguide supports are particularly high for single-mode operation. Our discussion is thus directed towards multimode applications. However, Fig. 2 shows the dependence of the radiation loss on the width L of the supports for a single-mode case. It is apparent that the interference of radiation originating at the front and back surface of the support slab causes the radiation loss curve to oscillate. These oscillations die out with increasing width of the supports. The parameter V, that is used to label the curve in Fig. 1 and all remaining figures, is defined by

$$V = (n_1^2 - n_2^2)^{\frac{1}{2}} kd. \quad (17)$$

V is a parameter that combines frequency, slab width, and refractive index difference. Its values determine the number of modes that can

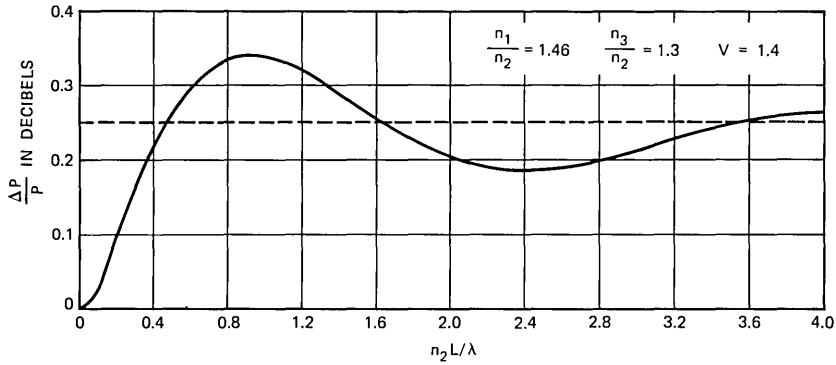

Fig. 2—Relative scattered power as a function of support thickness L. The dotted line represents eq. (16), which holds approximately for $L \to \infty$.

propagate on the waveguide. The lowest-order even TE mode has no cutoff, the cutoff value of V for this mode is thus $V_c = 0$. The cutoff of the next mode, the first odd mode, is given by $V_c = \pi/2 = 1.57$. In general, we obtain the cutoff condition for all the even and odd slab modes from the formula[1]

$$V_c = \nu \frac{\pi}{2}. \tag{18}$$

For even modes, ν assumes the values 0, 2, 4, etc.; for odd modes we have $\nu = 1, 3, 5$, etc.

The solid curve of Fig. 2 was obtained by numerical integration of (5) and (15). The dotted line also shown in the figure results from the approximation (16). This approximation holds for $L \to \infty$ and is thus independent of the variable $n_2 L/\lambda$.

The following figures apply to multimode operation and show the radiation losses caused by one set of waveguide supports as a function of the mode angle θ. Each guided mode can be decomposed into two plane waves, the propagation vectors of which form angles $+\theta$ and $-\theta$ with the waveguide axis. The angle θ is defined by

$$\cos \theta_\nu = \frac{\beta_\nu}{n_1 k}. \tag{19}$$

The mode angles assume discrete values θ_ν corresponding to the discrete values β_ν of the propagation constant of the guided modes. Figures 3 through 8 show θ as a continuous variable. It is important to remember that only certain discrete values of θ are allowed. The number of modes that exist below a certain value of θ is approximately proportional to

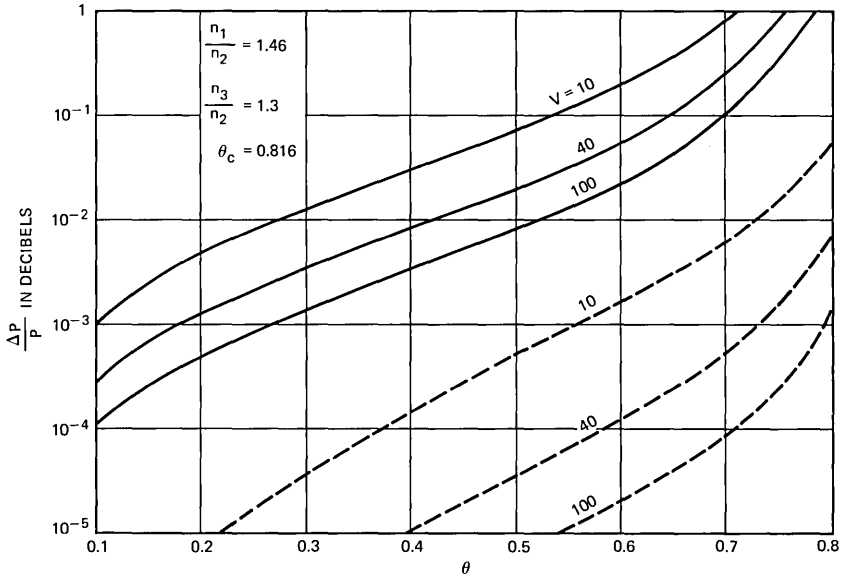

Fig. 3—Relative scattered power (solid lines) as a function of mode angle θ. The waveguide core and the supports are in an air environment. The dotted lines represent the reflection loss of each mode.

this angle in case of the slab waveguide but it is proportional to the square of θ in case of the round optical fiber. For the slab, the number of TE and TM modes with angles smaller than a given value θ is approximately given by ($\theta \ll 1$ is assumed)

$$N_\theta = \frac{4}{\pi} \theta n_1 k d. \tag{20}$$

The total number of guided TE and TM modes that can be supported is approximately given by

$$N_{\max} = \frac{4}{\pi} V. \tag{21}$$

For the round fiber we have[3] (d = fiber radius)

$$N_\theta = \tfrac{1}{2}(\theta n_1 k d)^2 \tag{22}$$

modes with angles less than θ. The maximum number of modes that the fiber can support is[3]

$$N_{\max} = \tfrac{1}{2} V^2. \tag{23}$$

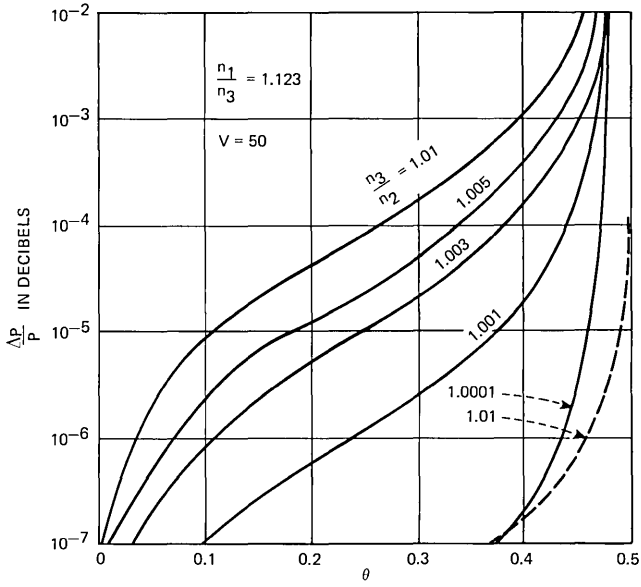

Fig. 4—Scattering losses as a function of mode angle θ. The core index is assumed to be $n_1 = 1.46$, the support index is $n_3 = 1.3$. Core and supports are immersed in an index matching liquid. The dotted line is the reflection loss for $n_3/n_2 = 1.01$.

Figure 3 shows the relative power loss as a function of mode angle θ for three different values of V. The refractive indices are chosen to represent a quartz fiber in air, $n_1/n_2 = 1.46$. It was assumed that the supports consist of a material with refractive index $n_3 = 1.3$. The solid curves represent the radiation losses while the dotted curves show the reflection losses for the modes with mode angle θ. It is apparent that the reflection losses are much smaller than the radiation losses. Mode conversion from each guided mode to all the other guided modes has not been considered. The maximum mode angle θ in this and all the following figures (with the exception of Fig. 4) coincides approximately with the right-hand edge of the graph.

Figure 4 shows the radiation losses for a fixed value $n_1/n_3 = 1.123$ of the ratio of core index to the index of the supports. This figure was drawn for the case that the core index is again $n_1 = 1.46$ and $n_3 = 1.3$, but allows for the fact that an index matching liquid is used in an attempt to reduce the scattering losses. The ratio n_3/n_2, that indicates the degree of index matching, is used as a curve parameter. The dotted curve is the reflection loss for $n_3/n_2 = 1.01$. The reflection losses for all the other index ratios are much smaller. The critical angle is different for each curve of this figure.

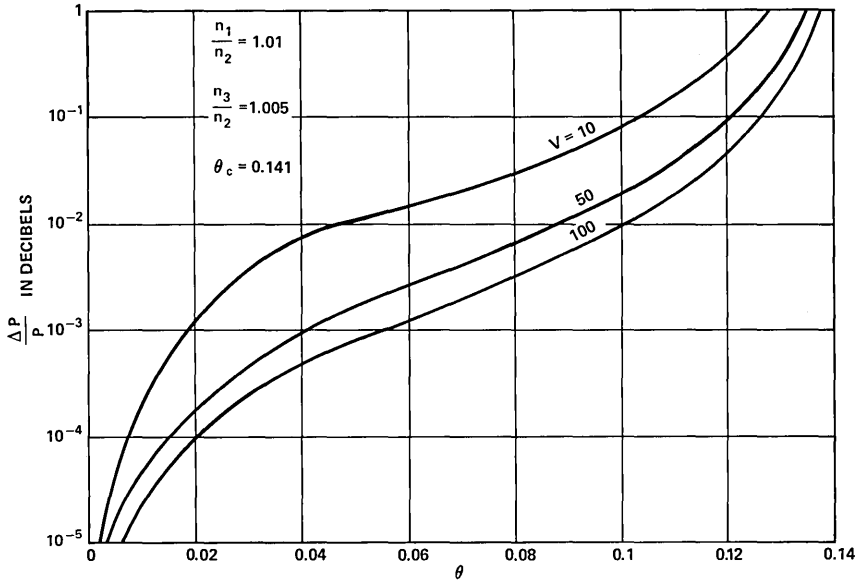

Fig. 5—Relative scattering losses as a function of mode angle θ, $n_1/n_2 = 1.01$, $n_3/n_2 = 1.005$.

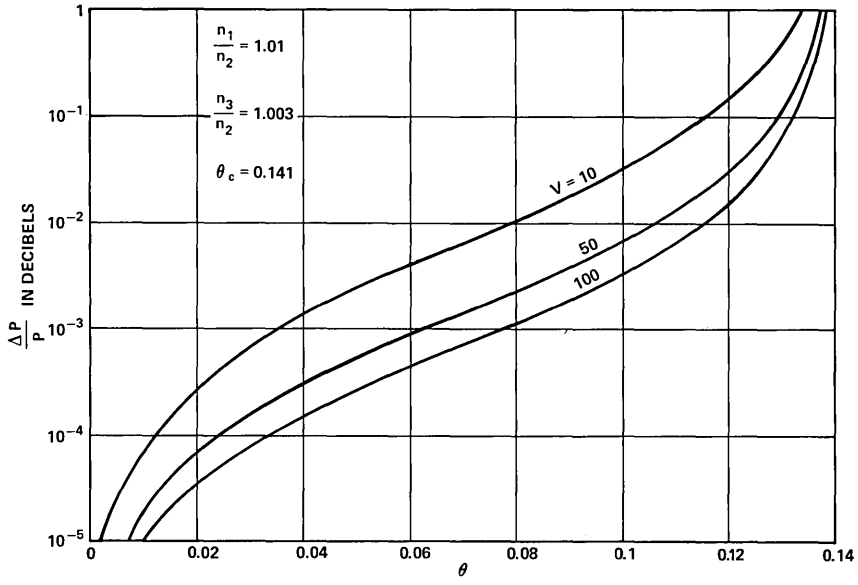

Fig. 6—Same as Fig. 5, $n_3/n_2 = 1.003$.

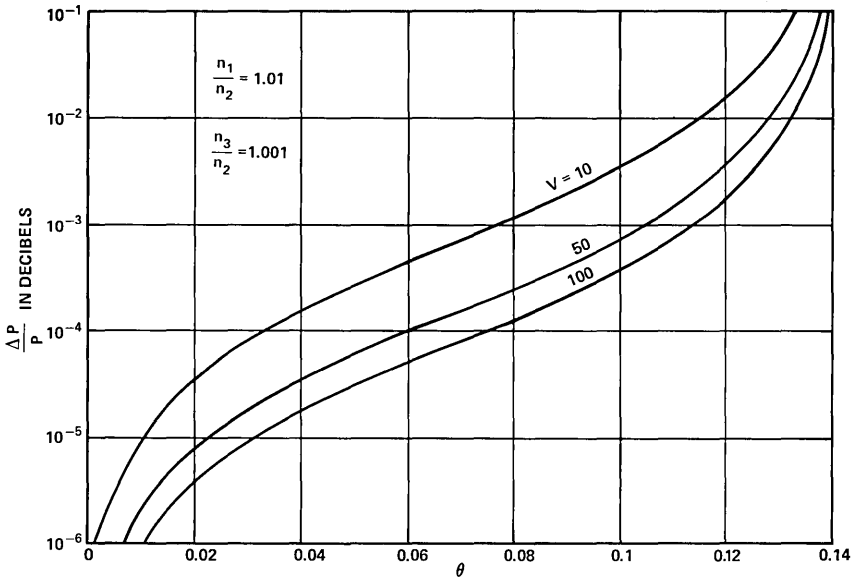

Fig. 7—Same as Fig. 5, $n_3/n_2 = 1.001$.

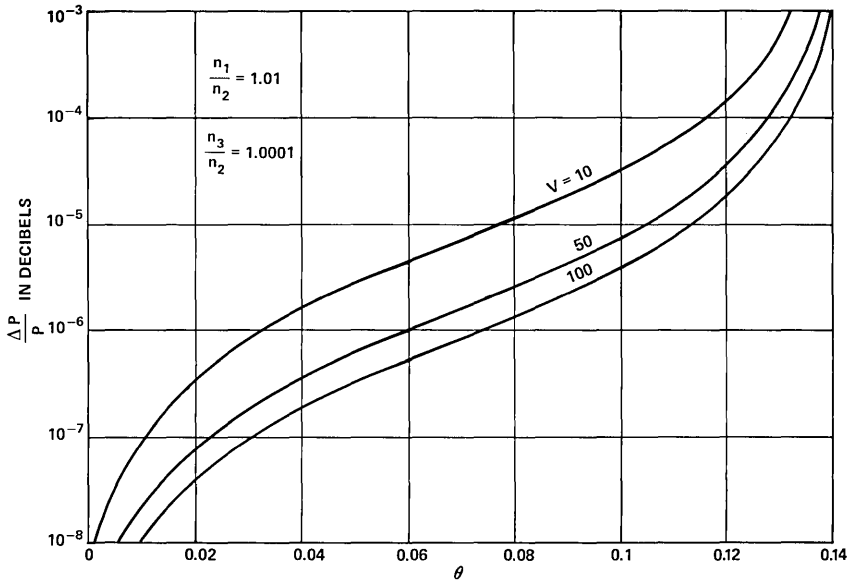

Fig. 8—Same as Fig. 5, $n_3/n_2 = 1.0001$.

The remaining four figures are all similar to each other. All show the radiation losses as functions of mode angle for three different values of V and for different values of the ratio n_3/n_2, but fixed value $n_1/n_2 = 1.01$. It is thus assumed that the index of the matching fluid remains the same, while the index of the supports is changed slightly. The reflection losses remain below the scale of all these figures.

For an evaluation of the total radiation losses to be expected from a fiber held by many supports, it is necessary to know the total number of supports. Let us assume for simplicity that we have one support per centimeter. A fiber of 1-km length is thus held by a total of 10^5 supports. In order to stay below a radiation loss of 10 dB/km, we must remain below the $\Delta P/P = 10^{-4}$ dB line of the figures. As a first crude approximation, we can assume that all modes, the loss of which remains below this line, are received at the end of the fiber while the modes that exceed this loss are lost. For a slab waveguide, the number of modes that can travel through the fiber is directly proportional to the mode angle θ, so that the ratio of transmitted to dissipated modes can be read off the horizontal axis. For the $n_3/n_2 = 1.01$ curve of Fig. 4, modes with angles less than 0.26 radian stay below the 10^{-4} line and can be considered as being transmitted through the fiber of 1-km length. This means that roughly half the modes are lost while half of them are being transmitted. However, if the waveguide is a round fiber (and if we accept the applicability of the loss curves for this case) we find that only one-quarter of all the modes can be transmitted while three-quarters of the modes and hence three-quarters of the power (for uniform initial power distribution) is lost. These estimates ignore mode conversion between the guided modes. The small values of the reflection loss suggest that conversion between guided modes is reasonably small. However, some power is converted from modes with small angles to large angle modes so that the loss estimate made on the basis of the individual mode losses alone must be optimistic.

A comparison of Fig. 4 and Fig. 5 shows clearly that it is advantageous to try to match the refractive index of the supports but keep the core index as large, compared to the surrounding liquid, as possible. The curves of Fig. 5 show much larger loss values because the difference in refractive index of core and index matching liquid is small, so that the fields extend farther into the liquid and thus are more effectively scattered by the supports. In order to achieve losses as low as those of the $n_3/n_2 = 1.01$ curve of Fig. 4, with a guide whose core-to-liquid index ratio is only 1.01, requires that the index ratio of support and matching liquid be better than 1.001.

IV. CONCLUSIONS

The problem of radiative power loss caused by light scattering from the supports of an unclad fiber has been investigated. The study was based on a slab waveguide model with dielectric slabs used as supports. The study comes to the conclusion that tolerably low scattering losses are obtained only if the supports are made less visible to the wave by index matching with a suitable matching liquid. The matching liquid must itself have very low dielectric losses. It is important to make the index difference between core and supports as large as possible (with the core index being larger than that of the supports) and match the support index as closely as possible. For a core-to-support index ratio of $n_1/n_3 = 1.123$, an index ratio of $n_3/n_2 = 1.01$ between supports and matching liquid is sufficient to allow one-quarter of all the modes of a round fiber to travel with acceptably low radiation losses. If the core index is more nearly equal to the refractive index of the supports, a much better index match for the supports is required. To achieve conditions comparable to the last example requires an index match of the supports of better than one tenth of a percent if $n_1/n_2 = 1.01$. It may be difficult to maintain such a good index match over the whole range of expected operating temperatures.

REFERENCES

1. Marcuse, D., *Light Transmission Optics*, New York: Van Nostrand Reinhold Company, 1972.
2. Stone, J., "Optical Transmission in Liquid-Core Quartz Fibers," Appl. Phys. Ltrs., *20*, No. 7 (April 1972), pp. 239–240.
3. Gloge, D., "Weakly Guiding Fibers," Appl. Opt., *10*, No. 10 (October 1971), pp. 2252–2258.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 2, February, 1973
Printed in U.S.A.

Passband Equalization of Differentially Phase-Modulated Data Signals

By R. D. GITLIN, E. Y. HO, and J. E. MAZO

(Manuscript received July 7, 1972)

A new approach to the automatic equalization of differentially phase-modulated data signals passing through a linear distorting medium is presented. The equalizer, which is of the transversal filter type, operates in the frequency passband and contains two sets of taps–in-phase and quadrature. A tap-rotation property of the equalizer is used to establish an absolute phase reference at the equalizer output, and once this crucial step is taken, the passband equalizer output can be used to automatically (as well as adaptively) adjust the tap weights so as to minimize a mean-square distortion function. The resulting algorithm requires correlating an error signal with the tap voltages, and thus it is possible to use a structure similar to that employed when equalizing baseband PAM.

The generality of the approach makes it applicable to the equalization of any double-sideband data signal.

I. INTRODUCTION

Recently much attention has been given to high-speed synchronous data transmission via differential phase modulation and comparison detection. High-speed (above 2400 bits per second) transmission usually requires equalization[1] to compensate for the linear distortion introduced by the channel. This study will be concerned with the automatic and adaptive equalization of such data signals. At first glance, the nonlinear nature of the modulated signal would seem to preclude linear compensation. Upon closer examination it becomes apparent that a digitally phase-modulated signal, when resolved into in-phase and quadrature components, is linear in each component. Since this property is preserved after transmission through a linear medium, simultaneous linear equalization of each signal component becomes feasible. Due, however, to the purely quadratic nature of the channel-

dependent terms present in the differential detector output, a linear equalizer should† precede the detector–hence the equalizer operates in the frequency passband.

Resolving the channel output into in-phase and quadrature components suggests that satisfactory performance will be obtained if the equalized in-phase signal is approximately Nyquist and the equalized quadrature signal has most of its samples close to zero. This is accomplished by choosing an appropriate cost function (of the equalized samples) and using an appropriately structured equalizer capable of minimizing this function. We will focus our attention on a mean-square cost function and on a synchronous tapped delay line (TDL) equalizer. The equalizer has two branches–an in-phase and quadrature branch–which together perform the passband compensation. The delays are separated by a symbol interval, and each delay output is first multiplied by a tap weight (each signal in the quadrature branch is also shifted by 90 degrees) and then added to the other delayed tap signals. The generality of the approach makes it applicable to the equalization of any double-sideband modulated data signal (e.g., combined amplitude and phase modulation, quadrature amplitude modulation).

Though the differential detector output is independent of any absolute phase reference, it is convenient to select a cost function which depends on such a quantity. The awkwardness of this situation is resolved by noting the tap-rotation property of the equalizer. Simply put, this means that any assumed reference phase angle can be "matched" by a rotation of the equalizer taps. This, in effect, permits the establishment of an arbitrary phase reference at the equalizer output. Equalization is accomplished with respect to this arbitrary reference while detection is done in an incoherent fashion. The tap adjustment algorithm is similar to that done in baseband PAM, i.e., tap signals are correlated with error signals. In fact, once the notion of a tap rotation is introduced, there is a convenient analogy to baseband PAM equalization.

The basic system equations are indicated in Section II, and the equalizer cost function and structure are described in Section III. Section IV describes methods for adjusting the equalizer taps when an ideal data reference sequence is available and also when the adjustments are made using random data. In Section V we consider the effects of frequency offset and phase jitter on the equalizer.

---

† In order that the mean-square distortion be a quadratic, and thus easily minimized, function of the in-phase and quadrature pulse samples.

II. BASIC SYSTEM EQUATIONS

In this section we indicate the response of a linear passband channel to a phase-modulated data signal. We also make some observations concerning the choice of phase reference that will be useful in our discussion of passband equalization.

The channel input is a phase-modulated (PM) data signal. This signal is given by

$$s_i(t) = \sum_n p(t - nT) \cos(\omega_c t + \theta_n) \tag{1a}$$

$$= u_{\text{in}}(t) \cos \omega_c t - v_{\text{in}}(t) \sin \omega_c t \tag{1b}$$

$$= r_{\text{in}}(t) \cos(\omega_c t + \Psi_{\text{in}}(t)), \tag{1c}$$

where θ_n is the information symbol, $1/T$ is the symbol rate, ω_c is the carrier frequency, $p(t)$ is the impulse response of the transmitter shaping filter,† and

$$u_{\text{in}}(t) \equiv \sum_n p(t - nT) \cos \theta_n \tag{1d}$$

$$v_{\text{in}}(t) \equiv \sum_n p(t - nT) \sin \theta_n \tag{1e}$$

are respectively the in-phase and quadrature signal components. The envelope and phase of $s_i(t)$ are given by

$$r_{\text{in}}(t) = \sqrt{u_{\text{in}}^2(t) + v_{\text{in}}^2(t)} \tag{1f}$$

$$\Psi_{\text{in}}(t) = \tan^{-1} \frac{v_{\text{in}}(t)}{u_{\text{in}}(t)}. \tag{1g}$$

By letting

$$a_n = \cos \theta_n \tag{1h}$$

$$b_n = \sin \theta_n, \tag{1i}$$

we can write

$$s_i(t) = \left[\sum_n a_n p(t - nT)\right] \cos \omega_c t - \left[\sum_n b_n p(t - nT)\right] \sin \omega_c t. \tag{1j}$$

From eq. (1j), we can see that the bandlimited signal $s_i(t)$ is linear in both a_n and b_n.‡ As we have previously remarked, it is this linear

† We assume the usual Nyquist shaping in the sense that $p(kT) = 0$, $k \neq 0$.

‡ Note that any double-sideband signal, such as quadrature amplitude modulation (QAM), can be written in the form of (1j). What we have to say in the sequel is sufficiently general to apply to any such double-sideband signals.

representation that makes linear compensation first plausible then possible.

We note that the transmitter changes the value of θ_n once every T seconds, and that for differentially coded data, the customer information is the quantity $\theta_n - \theta_{n-1}$. Applying $s_i(t)$ to a bandpass channel with impulse response

$$2F_1(t) \cos \omega_c t - 2F_2(t) \sin \omega_c t,$$

produces an output signal

$$s_o(t) = u_o(t) \cos \omega_c t - v_o(t) \sin \omega_c t, \tag{2a}$$

where

$$\begin{aligned} u_o(t) &= F_1(t) \circledast u_{\text{in}}(t) - F_2(t) \circledast v_{\text{in}}(t) \\ v_o(t) &= F_2(t) \circledast u_{\text{in}}(t) + F_1(t) \circledast v_{\text{in}}(t), \end{aligned} \tag{2b}$$

with $\circledast$ denoting the convolution operation. To obtain a more compact representation, we define the in-phase channel pulse

$$x(t) = p(t) \circledast F_1(t), \tag{3a}$$

and the quadrature channel pulse

$$y(t) = p(t) \circledast F_2(t). \tag{3b}$$

In terms of the data components

$$\begin{aligned} a_n &= \cos \theta_n, \\ b_n &= \sin \theta_n, \end{aligned}$$

and the above notation, the channel output is given by

$$\begin{aligned} s_o(t) = \cos \omega_c t [\sum_n a_n x(t - nT) - b_n y(t - nT)] \\ - \sin \omega_c t [\sum_n a_n y(t - nT) + b_n x(t - nT)]. \end{aligned} \tag{4}$$

The above in-phase and quadrature representation clearly indicates the effect of the channel on the transmitted signal. A "good" channel would be one which has a small quadrature pulse† (ideally all its samples would be zero) and an in-phase pulse which is Nyquist. In any event, it is important to note that in passband the channel distortion appears in a linear fashion.

† If the channel has even amplitude symmetry and odd phase symmetry about the carrier, then $y(t)$ will be identically zero for all t.

Since the carrier phase reference is arbitrary, the output signal can also be written as

$$s_o(t) = \cos(\omega_c t + \phi)\Big[\sum_n a_n x^{(\phi)}(t - nT) - b_n y^{(\phi)}(t - nT)\Big] - \sin(\omega_c t + \phi)\Big[\sum_n a_n y^{(\phi)}(t - nT) + b_n x^{(\phi)}(t - nT)\Big] \quad (5)$$

where $x^{(\phi)}(t)$ and $y^{(\phi)}(t)$ denote the in-phase and quadrature pulses with respect to the reference phase† ϕ. Comparing the right-hand sides (RHS) of (4) and (5) gives

$$\begin{bmatrix} x^{(\phi)}(t) \\ y^{(\phi)}(t) \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}. \quad (6)$$

Equation (6) indicates that changing the phase reference by ϕ radians corresponds to *rotating* the in-phase and quadrature signal vector $[x(t), y(t)]$ by this amount.‡ We wish to emphasize the above interpretation, since we will want to consider the effect of such a rotation on several quantities of interest. We indicate the incoherent nature of the system by including an arbitrary phase angle when describing the channel output, i.e., the absolute phase of the received signal is unknown to the receiver. For incoherent reception, it is common to use differential coding of the data and differentially coherent (comparison) detection. We will assume this mode of transmission and detection.

III. MEAN-SQUARE PASSBAND EQUALIZATION

We wish to develop an adaptive equalizer which is capable of compensating for the linear (in-phase and quadrature) channel distortion, while keeping the equalizer adaption (settling) time small. "Mean-square" adaptive equalization[1,2] has been successfully used in systems using linear modulation to meet the above objectives. Our approach will be to extend, with appropriate modification, this technique to the problem at hand. After specifying the equalizer structure, we select an appropriate cost function and indicate how the equalizer processes the received random signal to adapt to the optimum configuration. We begin by considering the output of the differential detector.

† We reserve the right to suppress the phase reference when there is no possibility of confusion. Of course the phase reference can be chosen as convenience dictates.

‡ Clearly the same relation also holds for the sampled values of $x(t)$ and $y(t)$.

3.1 *The Detector Output*

The equalizer is to be compatible with the standard comparison detector[1] which is shown in Fig. 1. It is straightforward to show that the detector outputs at the nth sampling instant are†

$$I(nT + t_o) = \sum_i \sum_j \cos(\theta_i - \theta_j)[x_{n-i}x_{n-j-1} + y_{n-i}y_{n-j-1}] + \sum_i \sum_j \sin(\theta_i - \theta_j)[x_{n-i}y_{n-j-1} - y_{n-i}x_{n-j-1}] \quad (7)$$

$$Q(nT + t_o) = \sum_i \sum_j \cos(\theta_i - \theta_j)[y_{n-i}x_{n-j-1} - x_{n-i}y_{n-j-1}] + \sum_i \sum_j \sin(\theta_i - \theta_j)[x_{n-i}x_{n-j-1} + y_{n-i}y_{n-j-1}] \quad (8)$$

where

$$x(nT + t_o) = x_n$$

and

$$y(nT + t_o) = y_n.$$

We note that the intersymbol interference at the detector output is of a quadratic nature and thus is quite different than that encountered in linear modulation. It should also be noted that the channel-dependent terms in (7) are inner products, and hence are independent of the phase reference. The terms in (8) are components of cross products and hence also independent of phase. This is a manifestation of the differential nature of the receiver, i.e., only quantities which are invariant with respect to a change in phase reference can affect the detector output.

If we are fortunate enough to have a distortionless channel, i.e., for all integers n,

$$x_n = \delta_{no} \quad (9a)$$

$$y_n = 0, \quad (9b)$$

then detection can be simply accomplished by noting that

$$I(nT + t_o) = a_n a_{n-1} + b_n b_{n-1} = \cos(\theta_n - \theta_{n-1}) \quad (10a)$$

$$Q(nT + t_o) = b_n a_{n-1} - a_n b_{n-1} = \sin(\theta_n - \theta_{n-1}). \quad (10b)$$

Unfortunately, (9) will not be satisfied for an arbitrary channel. As mentioned above, the function of the proposed equalizer is to linearly

† We have, of course, neglected noise and have suppressed the phase reference. We also wish to emphasize that a desirable sampling epoch t_o should be determined. As usual, this is a problem in its own right, but is not considered further in this paper.

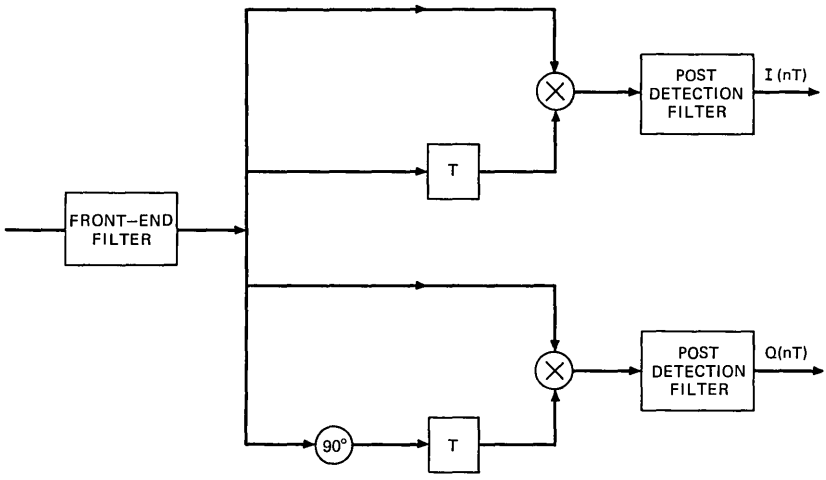

Fig. 1—Comparison (differential) detector.

operate on the in-phase and quadrature samples (the x_n's and y_n's) to produce an overall system response the samples of which (denoted by g_n's and h_n's) approximately satisfy eq. (9). The quality of this approximation is measured by the value of an appropriate cost function. We will have more to say about this later.

3.2 *Equalizer Structure*

Since we would like to equalize both the in-phase and quadrature channel samples, a structure which combines two signals in quadrature is suggested. The equalizer, which is shown in Fig. 2, consists of a tapped delay line (TDL) where each of the $2N + 1$ delay outputs is fed into two branches. In each branch the delay outputs are multiplied by tap weights (c_i and d_i) and then summed. The equalizer output $q(t)$ is the sum of the upper output $q_1(t)$ and the Hilbert transform of the lower output $q_2(t)$. The 90-degree phase shift at the output of the lower branch provides the quadrature signal. Thus the taps are to be adjusted to minimize an appropriate (cost) function of the equalized in-phase and quadrature samples. We remark that a linear equalizer placed in the passband can compensate for linear channel distortion, while a linear equalizer placed after the comparison detector cannot hope to compensate for the quadratic distortion present in the detector output. Thus the equalizer precedes the detector and operates in the frequency passband. We now describe the equalizer operation in detail.

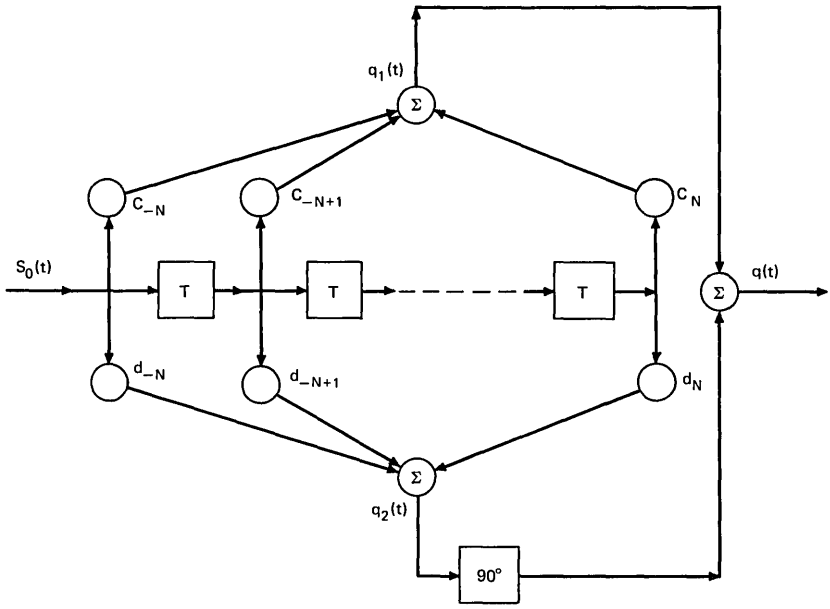

Fig. 2—The passband equalizer.

We can write the equalizer output, $q(t)$, as†

$$q(t) = \cos \omega_c t[\sum_n a_n g(t - nT) - \sum_n b_n h(t - nT)] - \sin \omega_c t[\sum_n a_n h(t - nT) + \sum_n b_n g(t - nT)] \quad (11)$$

where the in-phase and quadrature (equalized) signals, $g(t)$ and $h(t)$, are given by‡

$$g(t) = \sum_{n=-N}^{N} c_n x(t - nT) - \sum_{n=-N}^{N} d_n y(t - nT) \quad (12a)$$

$$h(t) = \sum_{n=-N}^{N} d_n x(t - nT) + \sum_{n=-N}^{N} c_n y(t - nT). \quad (12b)$$

We emphasize that $s_i(t)$ and $q(t)$ are defined with respect to the same

† Assuming that ω_c and T are chosen such that $\omega_c T$ is an integer multiple of 2π. If for some reason this is not convenient (perhaps a particular carrier frequency is desired), then $\omega_c T$ can be chosen to be any convenient angle. Some additional bookkeeping is then required at the receiver.

‡ The time reference has been taken at the center tap of the equalizer.

phase reference, but that the phase reference has not been indicated in the above equations. For a sampling epoch t_o, we denote the equalized samples by

$$g_j = g(jT + t_o), \quad h_j = h(jT + t_o). \tag{13}$$

Thus with the reference phase taken to be ϕ, the equalized samples are related to the channel samples and tap weights by

$$g_j^{(\phi)} = \sum_{-N}^{N} c_n x_{j-n}^{(\phi)} - \sum_{-N}^{N} d_n y_{j-n}^{(\phi)} \tag{14a}$$

$$h_j^{(\phi)} = \sum_{-N}^{N} d_n x_{j-n}^{(\phi)} + \sum_{-N}^{N} c_n y_{j-n}^{(\phi)}. \tag{14b}$$

Again we reserve the right to suppress the representation phase in writing (14). We now make some observations which indicate the effects of a change in reference phase and of a "tap-rotation" on the equalized samples.

Denoting the channel phase reference by ϕ, we can write the equalizer output as

$$\begin{aligned} q(t) = {} & \cos(\omega_c t + \phi)\Big[\sum_n a_n g^{(\phi)}(t - nT) - \sum_n b_n h^{(\phi)}(t - nT)\Big] \\ & - \sin(\omega_c t + \phi)\Big[\sum_n a_n h^{(\phi)}(t - nT) + \sum_n b_n g^{(\phi)}(t - nT)\Big], \end{aligned} \tag{15}$$

and it is easy to see that

$$\begin{bmatrix} g^{(\phi)}(t) \\ h^{(\phi)}(t) \end{bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} g(t) \\ h(t) \end{bmatrix}. \tag{16}$$

Thus we again observe that changing the reference phase by ϕ radians corresponds to rotating the in-phase and quadrature signals (and samples) by this amount, i.e., each two-tuple $(g_i^{(\phi)}, h_i^{(\phi)})$ is rotated by ϕ radians. Suppose that while keeping the phase reference fixed, we transform the taps (c_i and d_i) to new values ($\tilde{c}_i$ and $\tilde{d}_i$) via the rotation

$$\begin{bmatrix} \tilde{c}_i \\ \tilde{d}_i \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} c_i \\ d_i \end{bmatrix} \qquad i = -N, \cdots, 0, \cdots N. \tag{17}$$

We can now express the equalized samples $\tilde{g}_i$ and $\tilde{h}_i$ (corresponding to $\tilde{c}_i$ and $\tilde{d}_i$), using (14) and (17), as

$$\begin{bmatrix} \tilde{g}_i \\ \tilde{h}_i \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} g_i \\ h_i \end{bmatrix} \qquad i = -N, \cdots, 0, \cdots N. \tag{18}$$

Thus rotating the equalizer taps by θ will rotate the equalized samples by the same amount.† It is important to note that the original and (tap) rotated equalized samples are defined with respect to the *same* phase reference (ϕ). This is in contrast to the rotation introduced when the reference phase is changed. We will use these notions in describing the equalizer adjustment algorithm, in fact, the key to the simplicity of the equalization system lies in recognizing that the possibility of tap-rotation allows us to fix the reference phase and still keep a rotational degree of freedom between all (g_i, h_i) pairs. The reference phase will be suppressed in the sequel.

It remains to choose an appropriate cost function (of the equalized samples) and to describe an iterative procedure for determining the tap weights that minimize this function.

3.3 *The Cost Function*

The cost function should be such that after minimization, the equalized channel is approximately ideal (i.e., the intersymbol interference has in some sense been suppressed). There are, of course, many such functions–thus our selection is influenced by additional considerations.

We begin by noting that one possible task to set for the equalizer would be the minimization of

$$D = \sum_{i \neq 0} g_i^2 + \sum_{i \neq 0} h_i^2, \tag{19}$$

subject to appropriate constraint on g_0 and h_0.‡ We observe that since D is independent of an equalizer tap rotation, the minimum value of D is the same for any constraint that satisfies

$$g_0^2 + h_0^2 = 1, \tag{20}$$

including the choices

$$g_0 = 1, \quad h_0 = 0. \tag{21}$$

Clearly a choice such as (21) fixes the reference phase used to describe the equalizer output, and a "rotation" of the equalizer taps can assure that it is met. Ultimately, the information needed to adjust the equalizer taps will indeed be obtained with respect to a particular

† It should be clear that if there were no d_o tap, we could not perform this rotation.

‡ We note that this cost function would also be appropriate for a data signal which employs combined AM and PM. In fact, it is a simple matter to extend all of our results (i.e., the equalizer structure and algorithm) to this signaling format.

phase reference [in fact, we use the one implied by (21)]. Hence the tap adjustments will be made using coherently obtained information.

An additional interpretation of (21) is that the tap vectors are always required to lie on the hyperplanes $g_0 = 1$, $h_0 = 0$. Unfortunately, minimizing a cost function subject to this (hard) constraint cannot be done in real-time. This is due to the awkward requirement that the tap vectors be adjusted while constrained to lie on hyperplanes. For example, suppose we were using a steepest descent[3] algorithm to adjust the taps, then we would need the projected (onto the constraint hyperplanes) gradient of D at every iteration.† At present it is not known how to easily generate such a gradient from the available circuit voltages.

We circumvent this difficulty by imposing a quadratic penalty (a soft constraint) when the variables do not satisfy the constraints.[3] Thus the equalizer will try to minimize the unconstrained mean-square distortion

$$\mathcal{E} = \sum_{i \neq 0} g_i^2 + \sum_{i \neq 0} h_i^2 + (1 - g_0)^2 + h_0^2. \tag{22}$$

It should be noted that for small distortion, the optimum D [subject to (21)] and $\mathcal{E}$ will be essentially identical.‡

A useful consequence of the notions of phase-reference rotation and tap rotation will now be demonstrated. Suppose the phase reference has been chosen and is fixed at this value. Then $\mathcal{E}$, which can be rewritten as§

$$\mathcal{E} = \sum_{\text{all } i} (g_i^2 + h_i^2) + 1 - 2g_0, \tag{23}$$

clearly depends upon the chosen phase reference, since g_0 does. However, since we have shown that a simple *tap rotation* can be used to effect any desired phase, we are guaranteed that the minimization of $\mathcal{E}$ will be over *all* reference phase angles. That is, while (23) is in general not reference-phase invariant, the minimum of (23) over all tap settings is reference-phase invariant. A further manifestation of the tap-rotation property is the following necessary condition for the optimum tap setting: The optimum tap setting is such that $h_0 = 0$. To see that this must be the case, suppose that the taps have settled down and

† A more detailed discussion of this point is available in Ref. 4.
‡ See Ref. 4 for a discussion of this type of subject for baseband PAM.
§ Since the phase reference has been fixed, we do not indicate this quantity when writing the equalized samples.

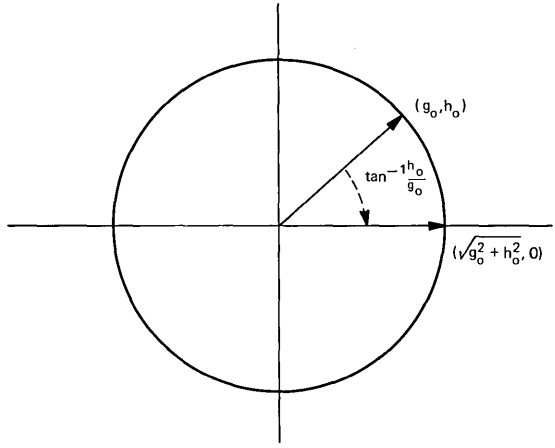

Fig. 3—Illustration of the tap-rotation property.

that $h_0 \neq 0$. Since $\sum_i g_i^2 + h_i^2$ is unchanged under a tap rotation, the equalizer taps could now be rotated so as to further minimize $\mathcal{E}$ by maximizing g_0 (i.e., making $h_0 = 0$). The rotation of the vector (g_0, h_0) into the vector $(\sqrt{g_0^2 + h_0^2}, 0)$ is shown in Fig. 3; thus the required rotation is $\tan^{-1}(h_0/g_0)$.

Recalling that a differential detector is only affected by phase-invariant quantities, we observe that if it were not possible to perform the above sort of minimization, then the utility of the structure and cost function would be suspect.

IV. EQUALIZER TAP ADJUSTMENT

Now that we have selected a cost function, we wish to demonstrate how the equalizer taps should be adjusted to minimize this function. We proceed in a manner reminiscent of mean-square equalization for baseband PAM–in fact it is shown that the tap increments are obtained by cross correlating error signals with delayed equalizer inputs.

4.1 *Deterministic Aspects*

We first wish to show that our criteria, $\mathcal{E}$, is an easily minimized function of the tap weights. We denote the tap vectors by

$$\begin{aligned} \mathbf{c}^T &= (c_{-N}, \cdots, c_0, \cdots, c_N) \\ \mathbf{d}^T &= (d_{-N}, \cdots, d_0, \cdots, d_N) \end{aligned} \tag{24}$$

where the superscript denotes transpose. It is convenient to introduce

the channel correlation matrices†

$$[X]_{ij} = \sum_n x_{n-i}x_{n-j} \tag{25a}$$

$$[Y]_{ij} = \sum_n y_{n-i}y_{n-j} \tag{25b}$$

$$[K]_{ij} = \sum_n (y_{n-i}x_{n-j} - x_{n-i}y_{n-j}), \tag{25c}$$

and the truncated channel vectors

$$\begin{aligned} \mathbf{x}^T &= (x_N, \cdots x_0, \cdots x_{-N}) \\ \mathbf{y}^T &= (y_{-N}, \cdots y_0, \cdots y_N). \end{aligned} \tag{26}$$

We can now write the mean-square distortion, in matrix notation as

$$\mathcal{E} = (\mathbf{c}^T, \mathbf{d}^T) \begin{bmatrix} X + Y & K \\ -K & X + Y \end{bmatrix} \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} - 2(\mathbf{x}^T, -\mathbf{y}^T) \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix} + 1. \tag{27}$$

A few comments are in order on the above representation. We first note that X and Y are the in-phase and quadrature channel correlation matrices. Thus if the quadrature component is zero, we need only the $\mathbf{c}$ taps. The cross-correlation between the in-phase and quadrature channel is measured by the skew-symmetric matrix K. By introducing the bandpass channel correlation matrix

$$A = \begin{bmatrix} X + Y & K \\ -K & X + Y \end{bmatrix}, \tag{28}$$

the augmented channel vector

$$\mathbf{z}^T = (x^T, -y^T), \tag{29}$$

and the composite tap vector

$$\mathbf{b}^T = (\mathbf{c}^T, \mathbf{d}^T), \tag{30}$$

the mean-square distortion can be expressed as

$$\mathcal{E} = \mathbf{b}^T A \mathbf{b} - 2\mathbf{b}^T \mathbf{z} + 1. \tag{31}$$

Conceptually at least, the optimum tap settings are given by setting the gradient of $\mathcal{E}$, with respect to $\mathbf{b}$, to zero. This gives‡

$$A\mathbf{b} - \mathbf{z} = 0 \tag{32a}$$

† $[X]_{ij}$ denotes the ijth element of the matrix X.
‡ Positive definiteness of A is assumed. For the optimum tap setting it is easy to see that $\mathcal{E} = 1 - \mathbf{z}^T A^{-1} \mathbf{z}$.

or

$$\mathbf{b}^* = A^{-1}\mathbf{z} \tag{32b}$$

as the desired tap settings.

If a gradient (steepest-descent) algorithm is used to minimize† $\mathcal{E}$ by iteratively adjusting $\mathbf{b}$, then the speed of convergence of the algorithm (also referred to as the equalizer settling time) is determined by the matrix A. A simple bound on the speed of convergence is given in Ref. 4.

$$\mathcal{E}_n \leqq \frac{1}{\lambda_{\min}}\left(1 - \frac{\lambda_{\min}}{\lambda_{\max}}\right)^n \mathcal{E}_0,$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote, respectively, the maximum and minimum eigenvalues of A, and $\mathcal{E}_n$ is the mean-square error after the nth iteration. The ratio $\rho = (\lambda_{\max}/\lambda_{\min})$ was computed for several voice-grade telephone channels (note that the closer ρ is to unity, the smaller the bound on settling time). Typically, the effect of the quadrature channel was small and the A matrix was diagonally dominant with ρ in the range from 2 to 8. Chang[5] has computed ρ for baseband PAM transmission, and comparing our numerical values with his indicates that for most channels the passband equalizer will settle rapidly. In other words, there are no special phenomena arising in the equalization of a bandpass channel which might lead one to expect a larger settling time than that observed for baseband channels. Though (32b) tells us what the optimum settings should be, we have not as yet indicated how these settings can be generated from available circuit voltages.

4.2 *Real-Time Tap Adjustments Using an Ideal Reference*

Ultimately, the cost function actually minimized is determined by the ease with which the optimum tap settings can be obtained in real time. We first note that if, as assumed, A is positive definite, then $\mathcal{E}$ is a convex function of the tap vector $\mathbf{b}$. Since, as is well known, a convex function has a single minimum, a gradient (steepest-descent) algorithm can be used to adjust the taps. If $\mathbf{b}^{(k)}$ denotes the tap vector after the kth adjustment, then the next tap setting is given by

$$\mathbf{b}^{(k+1)} = \mathbf{b}^{(k)} - \alpha^{(k)}\nabla\mathcal{E}(\mathbf{b}^{(k)}), \tag{33}$$

where $\alpha^{(k)}$ is a gain (or step size) and $\nabla\mathcal{E}$ is the gradient of $\mathcal{E}$. Note

† The details of how one uses such an algorithm in this context are described in the next section. The reader not familiar with this type of algorithm should read Section 4.2 before reading the present discussion.

that the algorithm "turns itself off" only when the gradient is zero. For a careful choice of the sequence $\{\alpha^{(k)}\}$, the above algorithm will converge to the optimum tap setting. Clearly the major problem associated with a real-time implementation of (33) is generating the gradient–or a good (statistical) approximation. To obtain an estimate of the gradient, we parallel the approach taken in baseband PAM.

The gradient of $\mathcal{E}$ can be rewritten in component form, as

$$\frac{\partial \mathcal{E}}{\partial c_i} = \sum_n (g_n x_{n-i} + h_n y_{n-i}) - x_{-i} \tag{34a}$$

$$\frac{\partial \mathcal{E}}{\partial d_i} = \sum_n (h_n x_{n-i} - g_n y_{n-i}) + y_{-i} \qquad i = -N, \cdots 0, \cdots N. \tag{34b}$$

We remark that a reasonable statistical approximation (i.e., an estimate) of the true gradient would be one whose average value is the right-hand side of (34). To do this, we first use the sampled equalizer output and an ideal data reference to generate a quantity, the average value of which is precisely our cost function. Once we have done this, we interchange the (linear) operations of ensemble averaging and differentiation to obtain the desired estimate.

The equalizer output samples are given by†

$$q(jT + t_0) \equiv q_j = \cos(\omega_c t_0 + \phi)\Big[\sum_n (a_n g_{j-n} - b_n h_{j-n})\Big] - \sin(\omega_c t_0 + \phi)\Big[\sum_n (a_n h_{j-n} + b_n g_{j-n})\Big], \tag{35}$$

and for perfect equalization these samples would be

$$q_j^* = a_j \cos(\omega_c t_0 + \phi) - b_j \sin(\omega_c t_0 + \phi) = \cos(\omega_c t_0 + \phi_j + \phi). \tag{36a}$$

If we have an ideal reference (i.e., a_j and b_j are known at the receiver), then since the tap-rotation property permits the equalizer to obtain any phase, we have the liberty to choose ϕ. In other words, we establish a phase reference at the equalizer output, and we force (via the cost function penalty) the equalizer to find the phase angle of the vector $(g_0^{(\phi)}, h_0^{(\phi)})$ which minimizes the cost function. The equalizer does this by rotating the taps until $h_0^{(\phi)} = 0$. For convenience we choose the reference phase, ϕ, so that

$$q_j^* = (a_j - b_j), \tag{36b}$$

† Again, since the reference phase is arbitrary, we are free to choose ϕ as convenience dictates.

i.e., $\omega_c t_0 + \phi$ has been taken to be $\pi/4$. Note that we do not need an oscillator to generate the ideal reference. By analogy with baseband PAM, we next consider the mean-squared difference between the sampled equalizer output (q_j) and the ideal (or desired) sequence (q_j^*), i.e., $E[(q_j - q_j^*)^2]$. With

$$a^2 \equiv E[a_n^2] = E[b_n^2], \tag{37a}$$

the independence of the data sequences is used to show that

$$E[(q_j - q_j^*)^2] = a^2[\sum_n (g_n^2 + h_n^2) + 1 - 2g_0] = a^2\mathcal{E}. \tag{37b}$$

This is an extremely useful observation, since we now have available a quantity $(q_j - q_j^*)^2$, the average value of which is proportional to the cost function. Interchanging the linear operations of expectation and differentiation allows us to write

$$\frac{\partial}{\partial c_i} E[(q_j - q_j^*)^2] = 2E\left[(q_j - q_j^*)\frac{\partial q_j}{\partial c_i}\right] \tag{38a}$$

and

$$\frac{\partial}{\partial d_i} E[(q_j - q_j^*)^2] = 2E\left[(q_j - q_j^*)\frac{\partial q_j}{\partial d_i}\right]. \tag{38b}$$

Note that $q_j - q_j^*$ is the error (at the equalizer output) at the jth sampling instant. It is easy to see that the differentiated terms can be written as

$$\frac{\partial q_j}{\partial c_i} = \cos(\omega_c t_0 + \phi)[\sum_n a_n x_{j-i-n} - \sum_n b_n y_{j-i-n}] - \sin(\omega_c t_0 + \phi)[\sum_n a_n y_{j-i-n} + \sum_n b_n x_{j-i-n}], \tag{39a}$$

and

$$\frac{\partial q_j}{\partial d_i} = -\cos(\omega_c t_0 + \phi)[\sum_n a_n y_{j-i-n} + \sum_n b_n x_{j-i-n}] - \sin(\omega_c t_0 + \phi)[\sum_n a_n x_{j-i-n} - \sum_n b_n y_{j-i-n}]. \tag{39b}$$

We recognize that $\partial q_j/\partial c_i$ is the jth channel output sample delayed by iT seconds, and thus is available at the ith delay element of the equalizer. We also note that $\partial q_j/\partial d_i$ is precisely $\partial q_j/\partial c_i$, but with the carrier phase delayed by $\pi/2$ radians, and thus is available at the output of the ith delay followed by a $\pi/2$ phase shifter. The estimated

gradient, which is given by the bracketed terms on the RHS of (38), can thus be obtained by correlating (muliplying and perhaps then averaging) the error signal $(q_j - q_j^*)$ and the tap outputs.† As the equalization becomes better, the ideal reference can be replaced by decisions, i.e., the receiver operates in a decision-directed mode. This then gives a complete description of the passband equalization of the PM data signal when an ideal reference is available.

At this juncture we wish to make the following remark. Suppose the channel is initially perfect [in the sense of eq. (9)], but the phase reference is such that h_0 is not zero:‡ What does the equalizer do? We first observe that the equalizer will rotate the taps until h_0 is zero, and thus lock in on the established reference phase. However, the differential detector output, which is insensitive to such a rotation, will have been error free from the outset.

4.3 *A Decision-Directed Tap Adjustment Procedure*

Suppose an ideal reference has been used to start-up the equalizer. It is desired that the equalizer then be capable of using decisions in place of an ideal reference. This mode of operation has been dubbed adaptive or decision-directed equalization.[1] The correction term needed for the tap adjustment algorithm is the product of the equalizer error signal $(q_j - q_j^*)$ and a delayed equalizer input. For decision-directed operation, q_j^* (which is $a_j - b_j$) would be replaced by the decided value of q_j. For four-phase operation, a_j and b_j are binary, thus $a_j - b_j = \cos(\phi_j + \pi/4)$ takes on the values -0.707 and 0.707. Thus $\hat{q}_j$ can be obtained as the output of a threshold device (with a threshold at zero).

V. FREQUENCY OFFSET AND PHASE JITTER

In this section we consider the effect of two common transmission impairments, frequency offset and phase jitter, on the operation of the equalizer.

5.1 *Frequency Offset*

Frequently, the transmission media perturbs the received carrier frequency ω_c so that it differs from the transmitted carrier frequency

† Remembering that to obtain $\partial q_j/\partial d_i$ the ith tap signal has been shifted by $\pi/2$ radians.

‡ This could come about, for example, by a phase-hit impinging on an otherwise ideal channel.

$2\pi/T$ by Δ Hz.† As a result of this frequency offset, each tap signal has a different carrier phase. To obtain a more detailed understanding of this effect on the equalizer, we introduce the following rotated equalized samples

$$\begin{bmatrix} g_n(\Delta) \\ h_n(\Delta) \end{bmatrix} = \begin{bmatrix} \cos n\Delta & \sin n\Delta \\ -\sin n\Delta & \cos n\Delta \end{bmatrix} \begin{bmatrix} g_n \\ h_n \end{bmatrix} \tag{40}$$

and

$$\begin{bmatrix} \tilde{g}_{m,n}(\Delta) \\ \tilde{h}_{m,n}(\Delta) \end{bmatrix} = \begin{bmatrix} \cos m\Delta & \sin m\Delta \\ -\sin m\Delta & \cos m\Delta \end{bmatrix} \begin{bmatrix} g_n(\Delta) \\ h_n(\Delta) \end{bmatrix}, \tag{41}$$

where g_n and h_n are the equalized samples in the absence of frequency offset.

It is straightforward to show that the equalizer output, at the mth sampling instant, can be written in terms of the above quantities as

$$q(mT) = \sum_n a_{m-n}\tilde{g}_{m,n}(\Delta) - \sum_n b_{m-n}\tilde{h}_{m,n}(\Delta). \tag{42}$$

The equalized samples [which are precisely the $\tilde{g}_{m,n}(\Delta)$'s and the $\tilde{h}_{m,n}(\Delta)$'s] are thus a cascade of these two rotations. The rotation described by (40) is time-invariant and amounts to presenting the equalizer with a slightly different channel which is obtained by rotating the vector (g_n, h_n) by $n\Delta$ degrees. By observing that

$$\begin{aligned} g_n^2(\Delta) + h_n^2(\Delta) &= g_n^2 + h_n^2 \\ g_0(\Delta) &= g_0, \end{aligned} \tag{43}$$

and

it is clear that this effect can be neglected since the cost function is invariant under such a transformation.

The rotation described by (41) is *time-varying* and indicates that at the mth sampling instant, each channel pair $(g_n(\Delta),h_n(\Delta))$ is rotated by $m\Delta$ radians. Since each channel pair is rotated by the same amount, an equalizer tap-rotation can compensate for this rotation, the only requirement being that the equalizer settling time be much smaller than the period of rotation $1/\Delta$. Thus if the settling time of the equalizer is small, the equalizer automatically tracks the frequency offset by rotating each tap pair (c_i,d_i) at the offset frequency Δ.

5.2 *Phase Jitter*

Phase jitter is an additive random component $\theta(t)$ often present in the phase angle of the channel output, and is characterized as a low-

† Typically Δ is less than 0.5 Hz.

pass process with energy up to 150 Hz. In this section we give only a suggestive description of the effect of phase jitter on the equalizer operation, since a detailed study appears to be difficult and could well be the subject of a separate investigation. If one were to assume the jitter to be constant along the length of the equalizer, i.e.,

$$\theta(mT - iT) = \theta(mT) \quad i = -N, \cdots, 0, \cdots, N,$$

then at the mth sampling instant, each equalized sample pair (g_n, h_n) is rotated by $\theta(mT)$ radians. Such an assumption could only be valid for the extreme low-frequency components of the phase jitter, of course, and these would be successfully "tracked" by slow to-and-fro motions of the tap settings. The higher-frequency components could not be followed by the equalizer, but might be modeled by noise sources at each tap causing fluctuations of the individual settings about the optimum values. In such a model, it would be realistic to include correlations among these fluctuations at each spectral component and one would expect the correlations to decrease with increasing frequency.

VI. SUMMARY AND CONCLUSIONS

An approach to the mean-square equalization of a differentially phase-modulated data signal has been presented. An equalizer, of the transversal filter type, has been proposed which operates in the frequency passband and contains two sets of taps–in-phase and quadrature branches. By exploiting the tap-rotation property of the equalizer, a phase reference is established at the output of the equalizer. Among the manifestations of the tap-rotation property are the ability to control the equalizer by using coherently obtained output samples and the ability to track small amounts of frequency offset and phase jitter. The equalized output is used to automatically (as well as adaptively) adjust the equalizer taps so as to minimize a mean-square distortion function. The required operations (correlating an error signal with tap voltages) are those performed when equalizing baseband PAM. Thus much of the existing knowledge concerning the technology of baseband PAM equalization can be applied to the equalization of phase-modulated data signals.

We make two final remarks. The equalizer structure and tap adjustment algorithm can be applied, with minor modification, to the equalization of any double-sideband modulated data signals, e.g., combined amplitude and phase modulation. As is the case with all equalization systems, the precise dynamic behavior of the proposed equalizer can only be studied, at present, by experiment.

VII. ACKNOWLEDGMENT

The authors would like to thank J. R. Sheehan for introducing us to this subject.

REFERENCES

1. Lucky, R. W., Salz, J., and Weldon, E. J., Jr., *Principles of Data Communications*, New York: McGraw-Hill, 1968.
2. Gersho, A., "Adaptive Equalization of Highly Dispersive Channels for Data Transmission," B.S.T.J., *48*, No. 1 (January 1969), pp. 55–70.
3. Luenberger, D. G., *Optimization by Vector Space Methods*, New York: John Wiley, 1969.
4. Gitlin, R. D., and Mazo, J. E., "Comparison of Some Cost Functions for Automatic Equalization," to be published in IEEE Trans. Commun., *Com-21*, March 1973.
5. Chang, R. W., "A New Equalizer Structure for Fast Start-Up Digital Communication," B.S.T.J., *50*, No. 6 (July–August 1971), pp. 1969–2014.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 2, February, 1973
Printed in U.S.A.

Selectively Faded Nondiversity and Space Diversity Narrowband Microwave Radio Channels

By G. M. BABLER

(Manuscript received August 14, 1972)

The spectral characteristics of nondiversity and space diversity narrowband radio channels subject to multipath fading were estimated from a detailed sampling of channel loss variations as measured on two vertically separated antennas. The data base for this analysis was obtained during a 93-day experiment in which the amplitudes of a set of coherent tones spanning a band of 33.55 MHz and centered at 6034.2 MHz were continuously monitored. The most significant observations were:

(*i*) *For the nondiversity channel, the frequency selectivity of the received transmission loss generally exceeded linear and quadratic components (in frequency) of amplitude distortion for fade depths greater than 30 dB.*

(*ii*) *For the diversity channel constructed by switching between a pair of narrowband channels received on two vertically spaced antennas, the frequency selectivity of the transmission loss was significantly reduced.*

I. INTRODUCTION

During periods of multipath propagation on line-of-sight radio links, deep and selective fading can severely corrupt the desired uniform transmission characteristics of a narrowband microwave channel received on a single antenna. Fortunately, because the multipath processes are sensitive functions of space,[1] the likelihood of simultaneously encountering both deep and frequency-selective fading on two vertically spaced antennas is very small. Earlier studies[2] were directed toward quantification of the spectral and temporal behavior of the amplitude distortion occurring in a nondiversity narrowband radio channel derived from a single receiving antenna. In the present paper, we describe a second experiment undertaken in 1971 to extend the

nondiversity measurements and to examine the amplitude distortions occurring in a diversity channel derived from the narrowband signals received on a pair of vertically spaced antennas.

The data used in this analysis were taken between June 16 and September 16, 1971 (93 days), and included the amplitude measurement of 62 uniformly spaced, coherent tones spanning 33.55 MHz at 6 GHz transmitted over a 26.4-mile radio path and received on two antennas of vertical separation of 19 feet 3 inches. The tone fields received on the two antennas were sampled five times per second and the results recorded whenever significant amplitude activity was occurring anywhere in either channel.

The paper is organized in the following order: (*i*) description of the experiment, (*ii*) a discussion of the occurrence of fading throughout the 1971 fading season, and an identification of the data base for analysis, (*iii*) characterizations of the nondiversity, and (*iv*) diversity amplitude distortions.

II. SUMMARY

A condensed listing of the findings follows:

(*i*) The fade depth distributions for individual tones had the expected slopes[3] of a decade of probability of occurrence per 10 dB change in fade depth with the lower antenna undergoing somewhat less fading.

(*ii*) The statistical distributions of linear and quadratic amplitude distortion of the nondiversity channel exhibited slopes of a decade of probability per 10 dB change in distortion. For a narrow channel bandwidth of 19.8 MHz, the linear and quadratic distortion exceeded 18.5 and 7 dB, respectively, for 10^{-5} of the observation time.

(*iii*) For the nondiversity channel, the amplitude distortion generally exceeded second order for fade depths greater than 30 dB and contained higher-order components in frequency selectivity.

(*iv*) The statistical distributions of linear and quadratic amplitude distortion of the diversity channels derived by threshold or comparative switching exhibited slopes of a decade of probability of occurrence per 5 dB change in distortion. For a bandwidth of 19.8 MHz, and for 10^{-5} of the time, a -30-dB threshold diversity channel experienced linear and quadratic distortions exceeding 10 and 2.8 dB, respectively, and a comparative diversity channel experienced distortions of 8.6 and 2.2 dB, respectively.

(*v*) For the diversity channels, the amplitude distortion generally exceeded zeroth-order for fade depths greater than 10 dB.

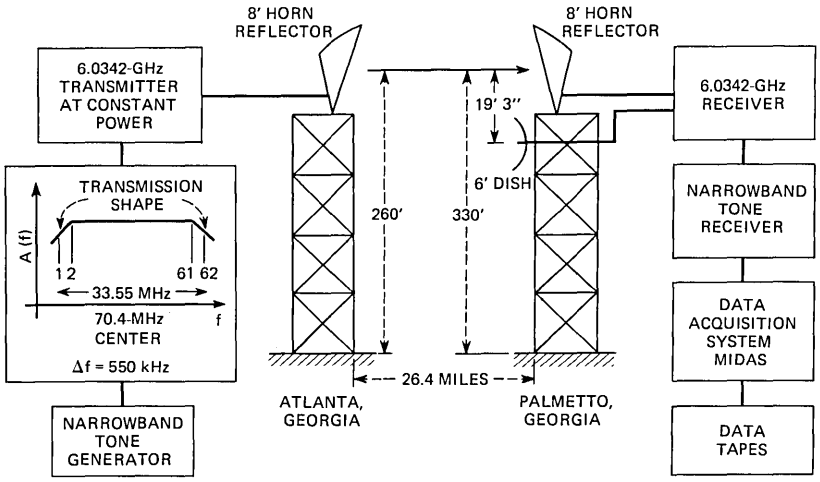

Fig. 1—Experimental arrangement on the Atlanta to Palmetto, Georgia, radio link.

III. EXPERIMENTAL DESCRIPTION

3.1 *The Experiment*

The 1971 measuring arrangement shown in Fig. 1 was essentially the same as the previous year which has been described elsewhere.[2] To review briefly, a flat and constant in time narrowband field of 62 coherent tones spaced 550 kHz apart and centered at a radio frequency of 6.0342 GHz was radiated from a standard horn reflector antenna at a microwave radio relay tower outside of Atlanta, Georgia. The signal, after propagating 26.4 miles along a line-of-sight path, was

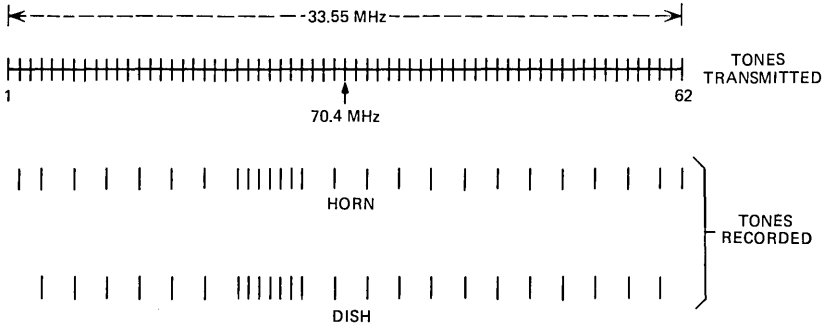

Fig. 2—Subset of tones transmitted and recorded serially on the horn reflector and dish antennas at Palmetto.

intercepted by two antennas: a standard horn reflector with a 1.25-degree half-power beamwidth and a 6-foot dish with a 2-degree half-power beamwidth separated 19 feet 3 inches apart and mounted on a microwave tower at Palmetto, Georgia. Both images of the narrowband tone fields were translated to an intermediate frequency where a tone receiver selected tones for a power measurement. The subset of transmitted tones measured and recorded are shown in Fig. 2; the amplitude quantization was 1 dB, the time quantization was 0.2 second. A multiple input data acquisition system (MIDAS) supervised the measurement and recording of amplitude levels and time as well as controlled the recording rates according to the tone activity.

The reference levels for both tone fields were determined by statistical studies of tone amplitudes during midday, nonfading periods. The rms variation in the reference levels was less than the amplitude quantization size of 1 dB.

IV. OCCURRENCE OF FADING ACTIVITY

The fading activity of the tone fields received on both antennas was continuously monitored from June 16 to September 16, 1971, and recorded for almost all of the 93 days (8.0352×10^6 seconds). The recorded data base, written by MIDAS and stored on 34 magnetic tapes, was condensed to include all periods with any fading in excess of 10 dB with respect to midday normal. This process compressed the time span of the data base to 1.28×10^6 seconds.

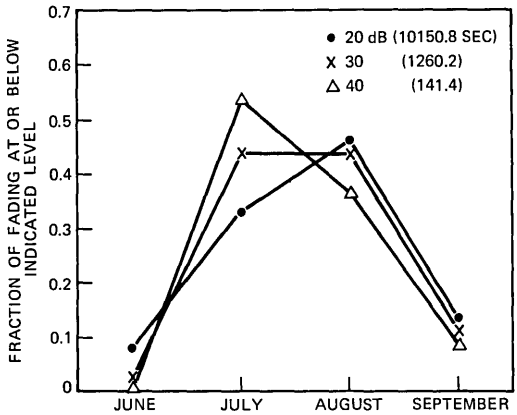

Fig. 3—Occurrence of fading for midchannel tone 24 received on the horn reflector antenna throughout the 1971 measurement period.

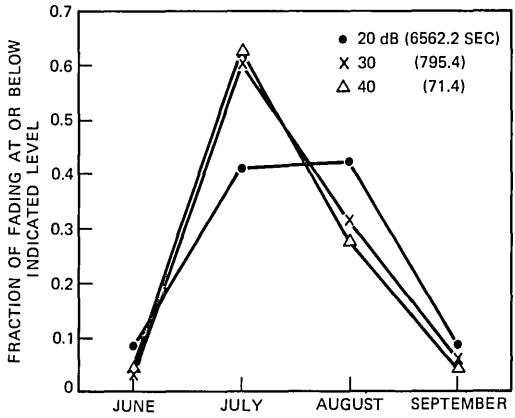

Fig. 4—Occurrence of fading for midchannel tone 24 received on the dish antenna throughout the 1971 measurement period.

The data for a selected set of tones in both narrowband channels was processed to determine the total time during which any tone was faded below signal levels. An overview of the fading activity measured

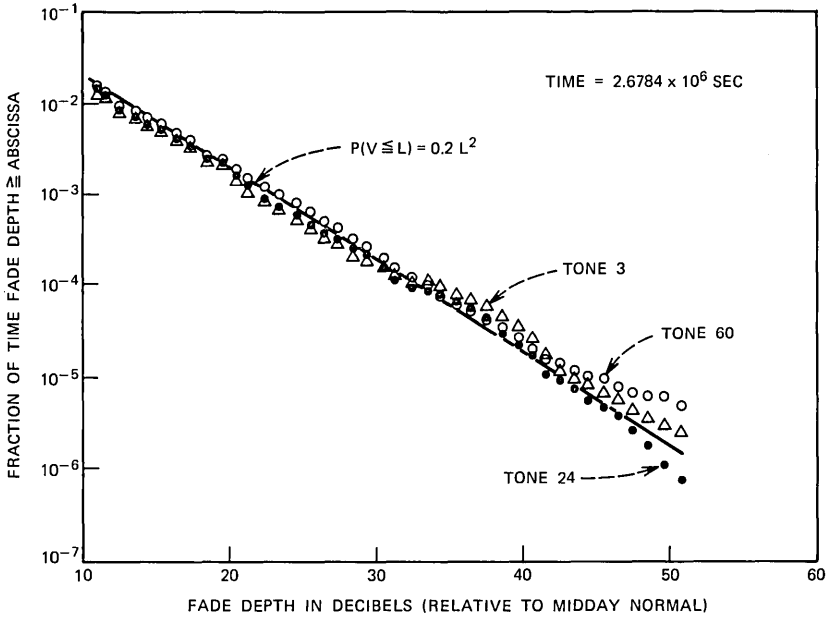

Fig. 5—Fade depth distributions for tones 3, 24, and 60 received on the horn reflector antenna for the month of August.

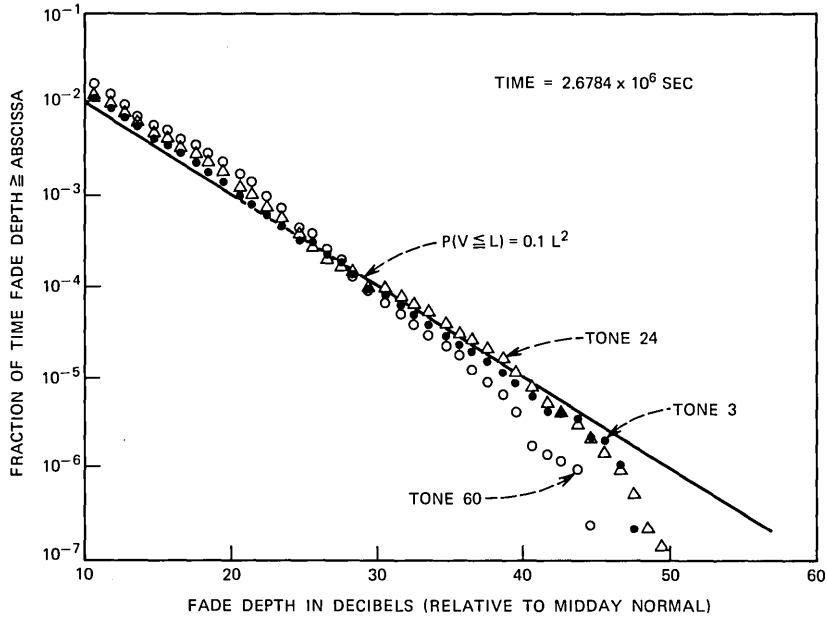

Fig. 6—Fade depth distributions for tones 3, 24, and 60 received on the dish antenna for the month of August.

during the fading season in Georgia and lumped into month categories of a single midchannel tone, tone 24, received on the horn and dish antennas is shown in Figs. 3 and 4, respectively. The occurrence of midchannel fading for both antennas is far less in the June and September periods than for July or August. In addition, we observed more midchannel fading for both antennas at and below 30 dB in July than in August, but the relative occurrence of fading at the 20, 30, and 40 dB levels and intermediate levels was more evenly balanced in August than in July. Because we desire to have a set of fading samples uniformly spread across all fade depths with no particular levels dominating the transmission loss statistics on either antenna, the month of August was selected as the working data base to be used in the characterization of the frequency-selective fading occurring in both channels.

The fade depth distributions for tones 3, 24, and 60 received on the horn and dish antennas are given in Figs. 5 and 6, respectively. The ordinate is the fraction of 2.6784×10^6 seconds (number of seconds in August) that the tones were faded the amount indicated on the

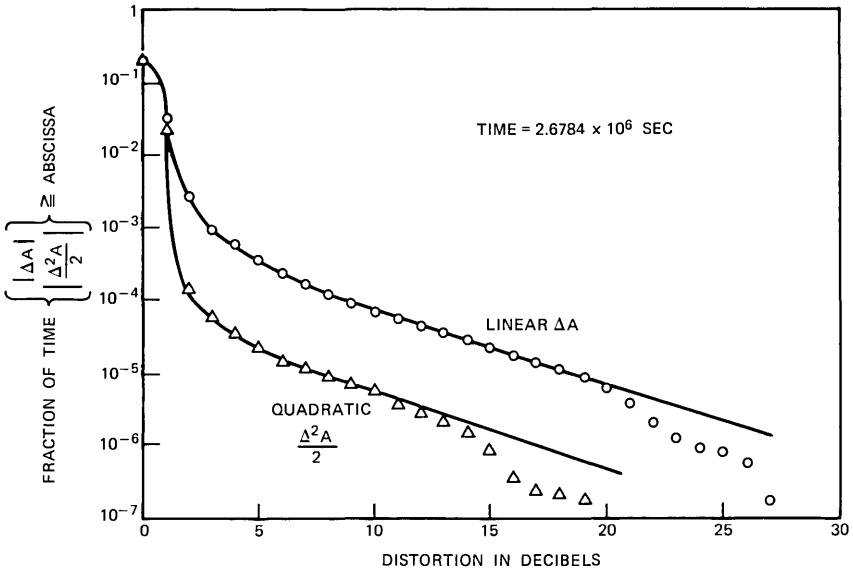

Fig. 7—Unconditional linear and quadratic distortion distributions for the 19.8-MHz nondiversity (horn reflector antenna) channel.

abscissa. The amplitudes, V, of the tones received on both antennas faded below signal levels, L, according to the expected fade depth distribution relation[3]

$$P(V \leqq L) = aL^2 \quad \text{for} \quad L \leqq 0.1 \ (20 \text{ dB}) \tag{1}$$

where fade depth $A = -20 \log L$ and where, for the August period, the proportionality constants a (functions of the radio paths' electromagnetic environment) were found to be $a = 0.2$ for the horn antenna and $a = 0.1$ for the dish antenna. The fact that, throughout the month, the lower dish antenna experienced less fading than the horn antenna reflects the spatial sensitivity of the fading processes and not the small differences in antenna beamwidths. Calculations based on an empirical formulation of the occurrence of multipath fading[4] indicate for the Atlanta-Palmetto 26.4-mile path at 6 GHz, $a \cong 0.27$. The conclusion is that, whereas the fading activity observed on both antennas was somewhat less than expected, the data base for the horn antenna is improved over the previous year.[2]

A second and important observation is that the fade depth distributions of all three tones on both antennas are essentially the same, indicating a nonpreferential amount of fading in the channels which

was not the case in the samples of the 1970 study. This more uniform set of fade samples across the narrowband channels, as well as the somewhat greater number of events, result in an improved understanding of frequency-selective fading in single and pairs of narrowband radio channels.

V. THE NONDIVERSITY CHANNEL SELECTIVITY CHARACTERIZATION

The degree of frequency selectivity which occurred in the nondiversity channel was characterized by statistical distributions of linear ($\Delta A = A(f_3) - A(f_1)$) and quadratic ($\Delta^2 A/2 = (A(f_1) + A(f_3))/2 - A(f_2)$) amplitude distortion constructed from three uniformly spaced samples of transmission loss in dB ($A(f_1)$, $A(f_2)$, $A(f_3)$). The distributions for the linear and quadratic distortions occurring across a 19.8-MHz nondiversity channel (the signal from the horn antenna) are shown in Fig. 7. The abscissa is the amount of distortion in dB, and the ordinate is the fraction of 2.6784×10^6 seconds (August). The distributions exhibit slopes of a decade of probability of occurrence per 10 dB change in distortion with the linear distortion $\Delta A \geqq 18.5$ dB and the quadratic distortion $\Delta^2 A/2 \geqq 7$ dB for 10^{-5} of the time (about 27 seconds). The roll-off of the curves below 10^{-5} is a too-few-samples effect. For high-performance microwave radio systems the smallest

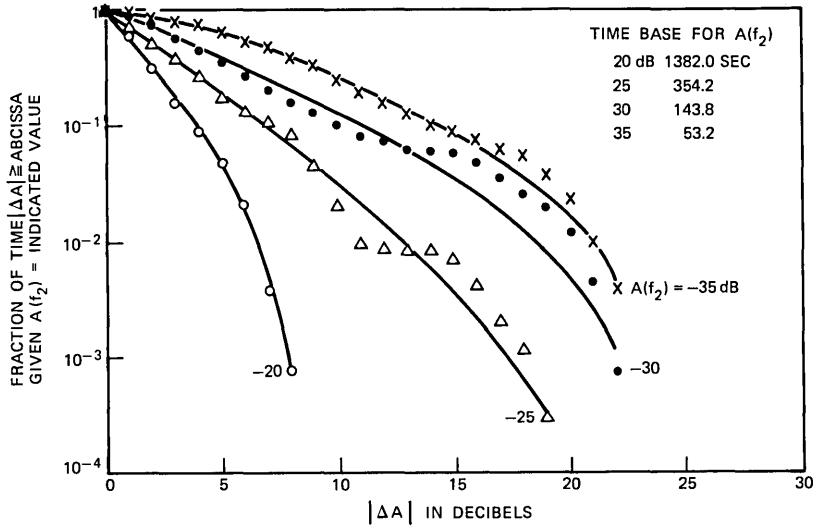

Fig. 8—Conditional linear distortion distributions for the 19.8-MHz nondiversity channel.

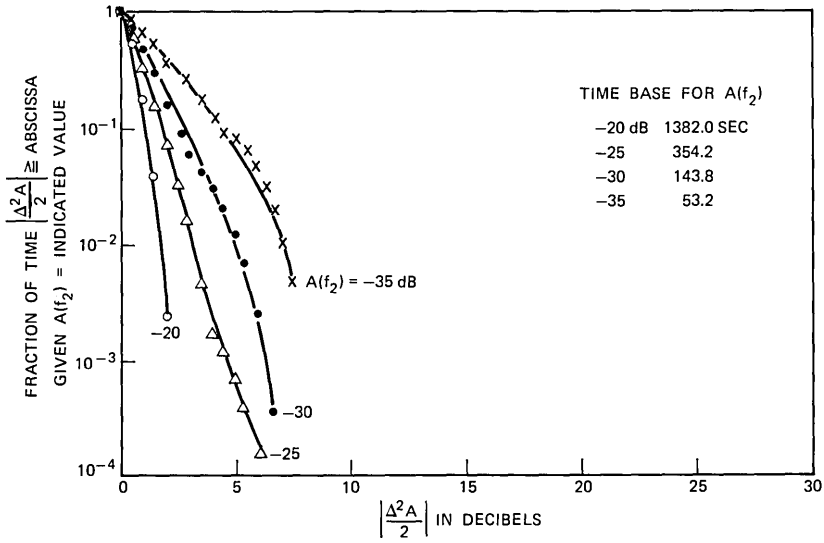

Fig. 9—Conditional quadratic distortion distributions for the 19.8-MHz nondiversity channel.

fraction of time for single-link consideration is approximately 10^{-5}, or a few tens of seconds per month. For larger fractions of time the data show excellent point-to-point consistency.

The linear and quadratic distortion distributions conditioned on fade depth of the midchannel tone $[A(f_2)]$ are shown in Figs. 8 and 9, respectively. The time base for each curve is indicated by the fade depth. Although the distortion curves have some point scatter, the curve-to-curve placement is improved over the 1970 sample.[2] In addition, we note that at each fade depth the channel experienced more linear and less quadratic distortion in 1971 as compared to the 1970 sample. Because there were more deep fading events in 1971 than 1970, and because there was more linear distortion at each fade depth in 1971 than 1970, we can conclude, as shown in Fig. 7 of this paper and Fig. 12 of Ref. 2, that there was more linear distortion in the channel. No similar arguments can be made for the quadratic distortion because of conflicting occurrences of distortion and fade depth for both years, although we do note that the 1971 channel suffered less quadratic distortion. The rate of growth of the linear and quadratic distortion which occurred in the nondiversity channel with increasing bandwidth and fade depth was essentially the same as the 1970 data (and is shown in Fig. 16).

Higher-order selective effects were again studied by accumulating the distributions of the maximum amplitude difference, called MAX|ERROR|, between the observed channel loss samples $A^M(f)$ and a three-term power series quadratic approximation constructed from the linear and quadratic distortion parameters. Thus

$$\text{MAX}|\text{ERROR}| = \text{MAX}\left|\left\{\underbrace{A^M(f)}_{\substack{\text{Measured Samples}\\ \text{of Channel Loss}}} - \underbrace{\left[A(f_2) + \frac{\Delta A}{2}\left(\frac{f-f_2}{\Delta f}\right) + \frac{\Delta^2 A}{2}\left(\frac{f-f_2}{\Delta f}\right)^2\right]}_{\text{Quadratic Power Series Approximation}}\right\}\right|.$$

Fig. 10 shows the occurrence of MAX|ERROR| events for all fade depths (an unconditional distribution) and Fig. 11 shows the conditional distributions of MAX|ERROR| events for individual fade depths. The unconditional distribution of MAX|ERROR| for a bandwidth of 19.8 MHz is approximately the same as the 1970 data for a bandwidth of 20.35 MHz. The conditional distributions indicate

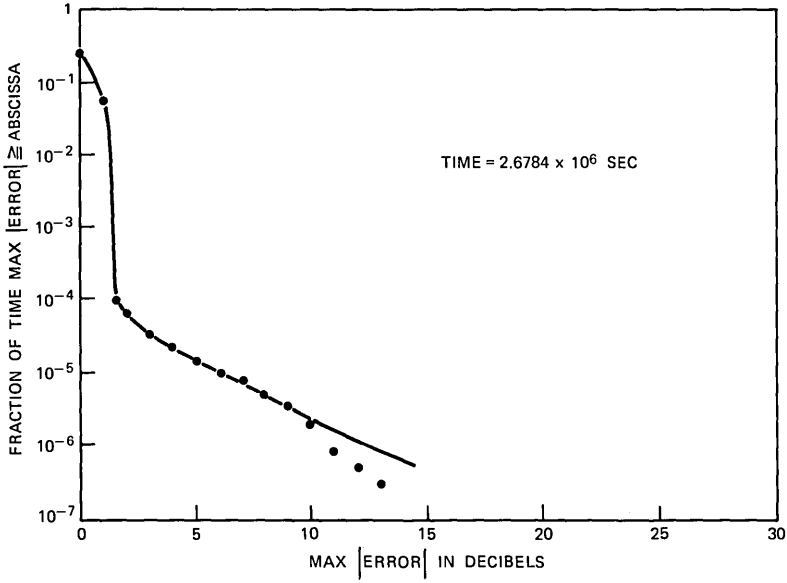

Fig. 10—Unconditional distribution of the MAX|ERROR| for the 19.8-MHz nondiversity channel.

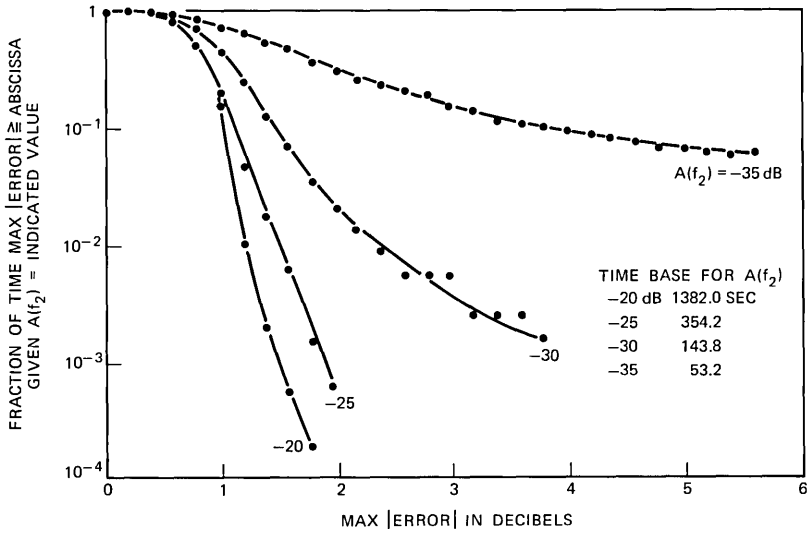

Fig. 11—Conditional distributions of MAX|ERROR| for the 19.8-MHz nondiversity channel.

rapid growth in the difference between the three-term power series approximation and the actual channel loss at and below the 30-dB fade level. For fade depths greater than 30 dB, the observed amplitude-frequency selectivity structure of the nondiversity narrowband radio channel's transmission loss exceeded linear and quadratic components (in frequency) of amplitude distortion by 2 dB as was observed previously in the 1970 narrowband data.

VI. THE DIVERSITY CHANNEL SELECTIVITY CHARACTERIZATION

One form of protection against deep and selective fading in a narrowband radio channel during periods of multipath propagation is to derive a diversity channel from the narrowband signals received on a pair of vertically spaced antennas. For the deeper fades, the reduction in number and duration of fading events increases significantly, since deep fades rarely occur simultaneously on two vertically spaced antennas.[1] Because amplitude distortion (linear, quadratic, and higher orders of structure) is present for only the deeper fades, the likelihood of simultaneously encountering deep fading with large degrees of amplitude distortion on two antennas is indeed rare.

To demonstrate the variety and nature of the measured transmission loss as simultaneously observed in the space diversity pair of narrow-

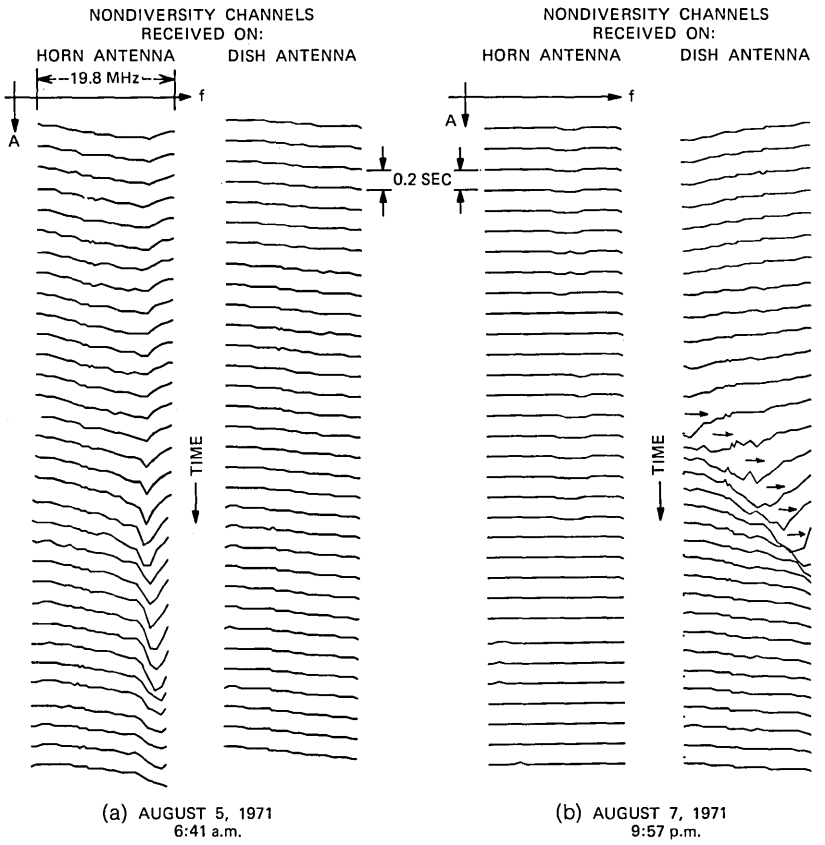

Fig. 12—Time sequential plots of the measured transmission losses simultaneously occurring in the narrowband channels received on the horn reflector and dish antennas.

band channels during the 1971 period, and to motivate the diversity channel construction algorithms, two fading periods are presented in Figs. 12a and 12b. Displayed is the simultaneous time sequential state of the channel transmission loss of the tone fields as measured on the horn and dish antennas. An arbitrary 0-dB transmission loss reference has been employed for figure compactness; the time between scans was 0.2 second.

In event (a), the selectivity develops at the upper edge of the narrowband channel received on the horn antenna and then moves out of channel. The narrowband channel as received on the dish antenna was simultaneously undergoing about 5 dB of linear distortion super-

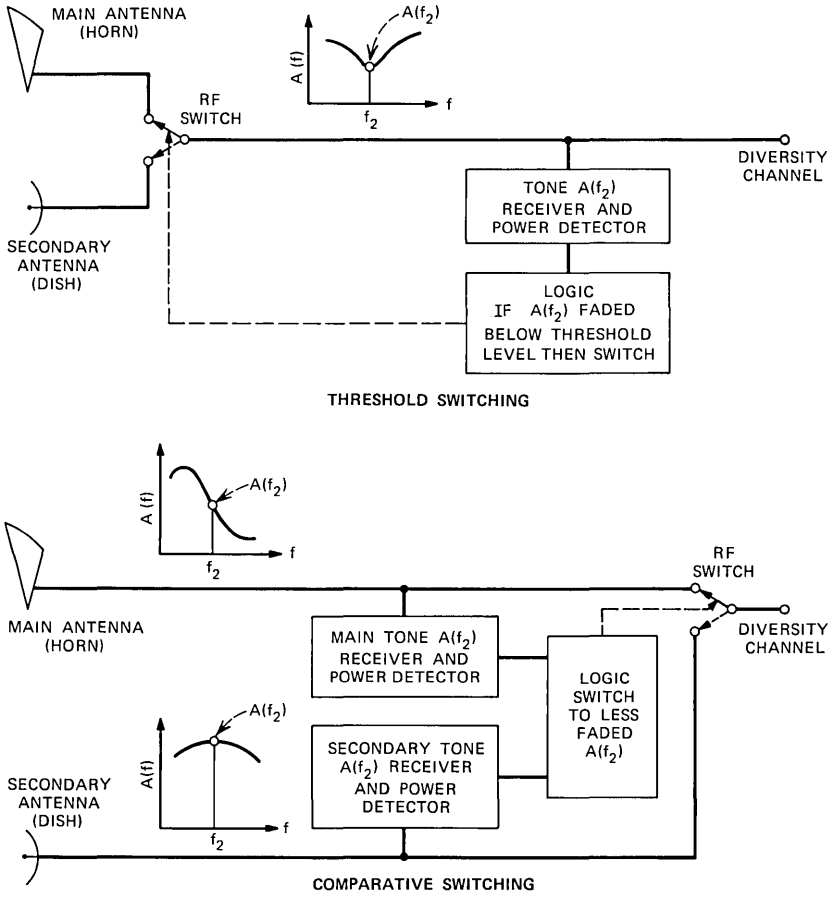

Fig. 13—Diversity channel construction schemes.

imposed on a broader-band nonselective fade. In event (b), the horn antenna suffers only slight nonselective transmission loss, while the dish antenna experiences a rapidly moving selective fade. Most of the fading events observed were similar to events (a) or (b); generally, when there was deep fading with appreciable amplitude distortion in one channel, the alternate channel was undergoing only shallow fading.

In Section 6.1 we describe how narrowband diversity channels were constructed from the space-diversity measurements of the 1971 period and the characterization of the observed diversity narrowband channel amplitude distortions.

6.1 *The Diversity Channel Construction*

The objective of diversity construction is to derive a new narrowband channel undergoing less fading and consequently less amplitude distortion. Two straightforward idealized switching algorithms (schemes), threshold switching and comparative switching, have been examined and are presented schematically in Fig. 13.

In threshold switching, a single narrowband threshold detector measures the power at the center of the narrowband tone field [say at $A(f_2)$] received on the presently connected antenna, and only if

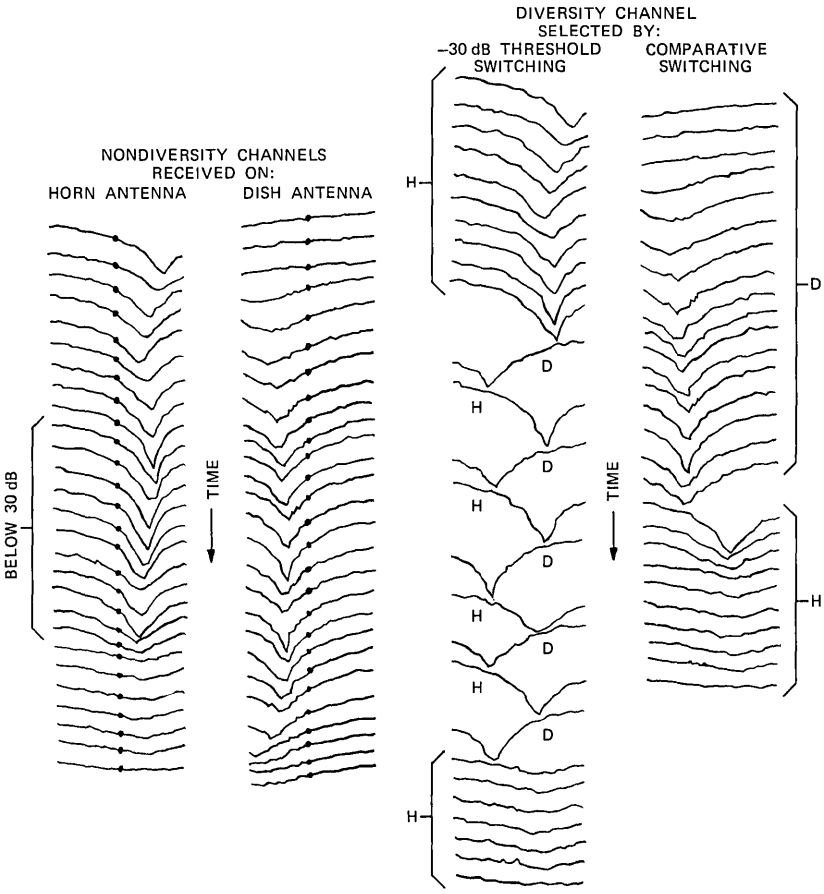

Fig. 14—Time sequential plots of the nondiversity and diversity channel transmission losses.

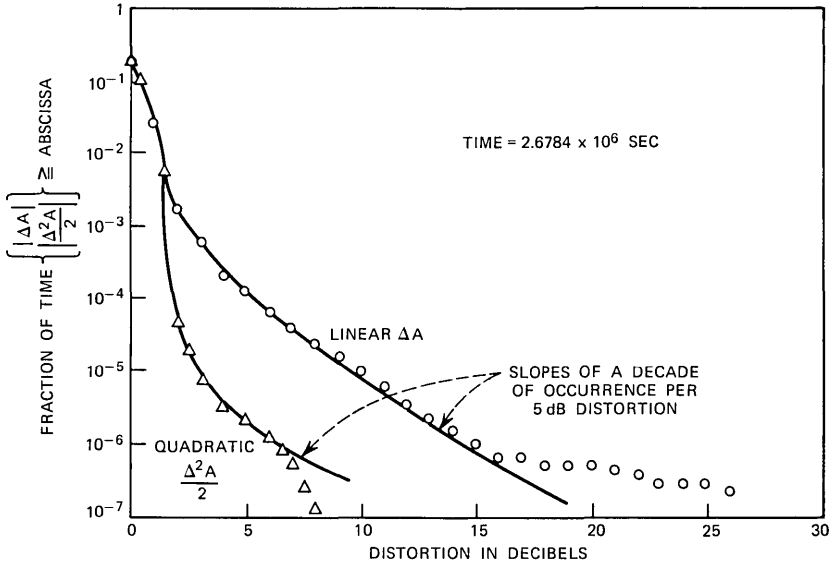

Fig. 15—Unconditional linear and quadratic distortion for the 19.8-MHz, −30-dB threshold diversity channel.

this power falls below some fixed threshold level, system logic directs a radio frequency (RF) switch to select the alternate antenna. If the midchannel power on the secondary antenna is (or falls) below the threshold level, the main antenna is reconnected. The second switching scheme studied, comparative switching, employs two separate receivers which simultaneously monitor the midchannel power of the narrow bands received on both antennas and direct the RF switch to connect in the antenna with greater midchannel power. Both forms of channel construction, threshold and comparative switching, result in a newly derived diversity channel undergoing less fading and consequently less amplitude distortion.

For the selective events displayed in Fig. 12, both switching schemes would result in the selection of the less selectively faded antenna. The resulting diversity channel would therefore be the dish in event (a) and the horn in event (b). To serve as an additional instructive example, in Fig. 14 is shown the channel activity of a diversity channel constructed from the nondiversity channels which are simultaneously undergoing selective fading. The threshold diversity channel is the result of 10 switches between antennas beginning and ending the displayed period on the horn antenna. Because both antennas' mid-

channel power levels are below the −30-dB threshold level, the cycle time between nondiversity channels is 0.2 second. For this particular set of events, contrary to those of Fig. 12, the threshold diversity channel's amplitude distortion is not significantly improved over either nondiversity channel because of the simultaneously occurring selective fading in both antennas. The comparative diversity channel, the result of a single switch between antennas, experiences less amplitude distortion than either nondiversity channel. The events as indicated in Fig. 12 are much more frequent than the highly selective and rare events as displayed in Fig. 14.

By using the real-time measurements of transmission loss simultaneously occurring in both narrowband channels and computer programs to simulate the switching schemes, the linear and quadratic amplitude distortion distributions for the diversity channels were accumulated both for unconditional and conditional fade depths [conditioned on the diversity channel's tone, $A(f_2)$] as well as for different bandwidths. These results follow.

6.2 *The Threshold Diversity Narrowband Channel*

The unconditional linear and quadratic distortion distributions for the 19.8-MHz-wide threshold diversity channel are shown in Fig. 15. A threshold level of −30 dB was chosen because it is at and below this level that the amplitude distortion of a nondiversity channel exceeds second-order selectivity and it is these higher-order distortion events we wish to avoid by the antenna selecting process. For a band-

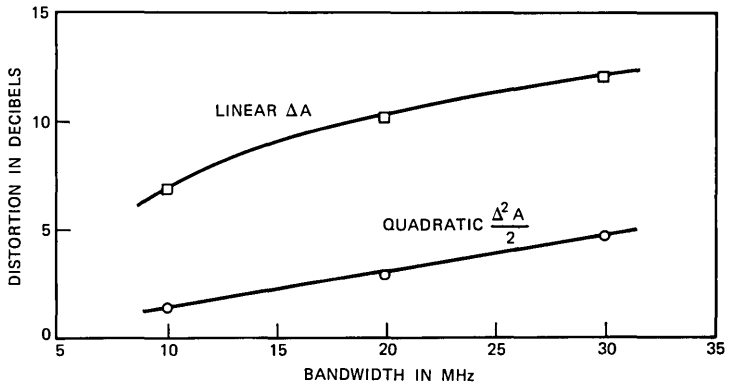

Fig. 16—Growth of linear and quadratic distortion with bandwidth for the nondiversity channel and the −30-dB threshold diversity channel for a fixed 10^{-5} fraction of the time.

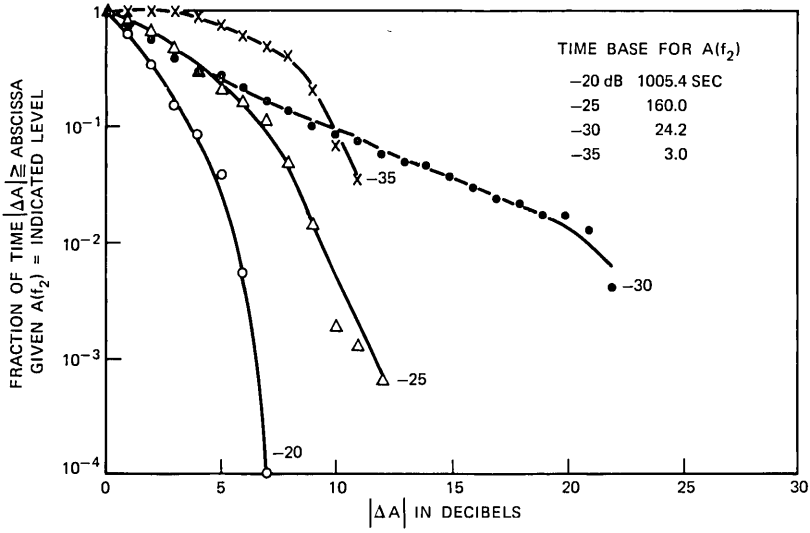

Fig. 17—Conditional linear distortion distributions for the 19.8-MHz, −30-dB threshold diversity channel.

width of 19.8 MHz, the linear and quadratic distortions exceeded 10 and 2.8 dB, respectively, for 10^{-5} of the time. Contrasting Figs. 7 and 15, we find for 10^{-5} fraction of the time that the −30-dB threshold switching procedure has reduced the linear and quadratic amplitude distortions by 46 and 60 percent, respectively, over the distortions occurring in the nondiversity channel. The distributions of distortion for the −30-dB threshold channel exhibit slopes of a decade of probability of occurrence per 5 dB change in distortion. Although the data points below 10^{-6} are less reliable because of fewer samples, the roll-up at the tail of the linear distortion distribution is real and represents a few events spanning a few seconds during the 2.6784×10^6 seconds of measurement where severe fading below 30 dB was simultaneously occurring in both narrowband radio channels. The distributions for other bandwidths of the threshold channel were similar in shape to those presented in Fig. 15; the rate of growth in distortion with bandwidth, as indicated in Fig. 16, was about half the nondiversity channel's rate. The −30-dB threshold diversity channel was the result of 103 switches between antennas.

The conditional distributions of distortion are shown in Figs. 17 and 18. The distributions show the expected reduction in linear distortion as compared to the nondiversity channel, Fig. 8, for fades not exceeding

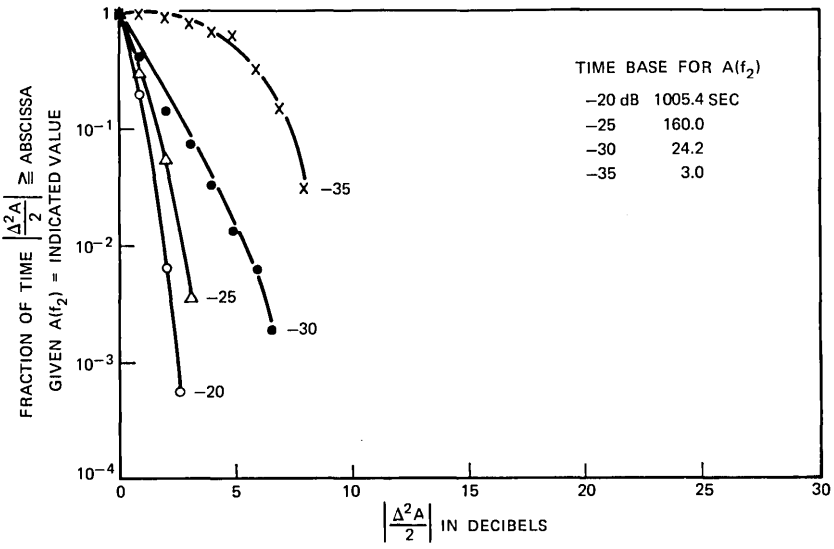

Fig. 18—Conditional quadratic distortion distributions for the 19.8-MHz, −30-dB threshold diversity channel.

30 dB. The roll-up of the distribution for 30-dB fades is the result of two effects: depletion of shallow structure (small linear distortion) events due to the threshold switching action and the over-abundance of a few high structure events at the 30-dB level. The roll-off of the 35-dB fade distribution is the result of too few samples. The distributions of the observed quadratic distortion at the various fade depths, Fig. 18, show a more uniform curve-to-curve behavior and indicate approximately the same amounts of distortion at each fade depth as the nondiversity channel. The 35-dB fade depth curve again has insufficient samples.

These results show that a diversity channel selected according to the −30-dB threshold scheme outlined above undergoes significantly less overall amplitude distortion and presents to the radio sytem a more uniform and desirable transmission loss characteristic. Higher-order structure of the threshold diversity channel will be discussed in Section 6.4

6.3 *The Comparative Diversity Narrowband Channel*

The unconditional linear and quadratic distortion distributions for the 19.8-MHz-wide comparative diversity channel are shown in Fig. 19. Contrasting Figs. 15 and 19, we observe that the comparative di-

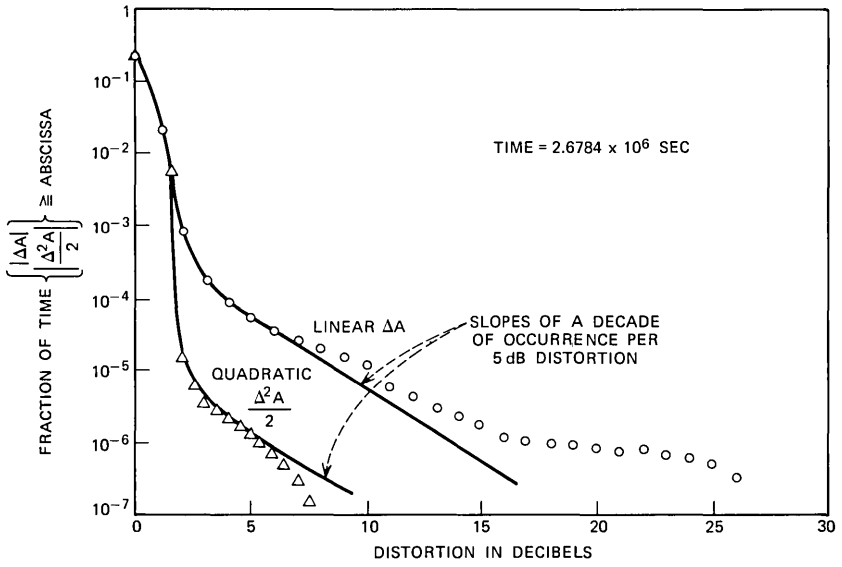

Fig. 19—Unconditional linear and quadratic distortion distributions for the 19.8-MHz comparative diversity channel.

versity channel experienced less linear and quadratic distortion than the threshold diversity channel. For a bandwidth of 19.8 MHz, the linear and quadratic distortions of the comparative diversity channel exceeded 8.7 and 2.2 dB, respectively, for 10^{-5} of the time. Again we note that the distributions for the diversity channel exhibit slopes

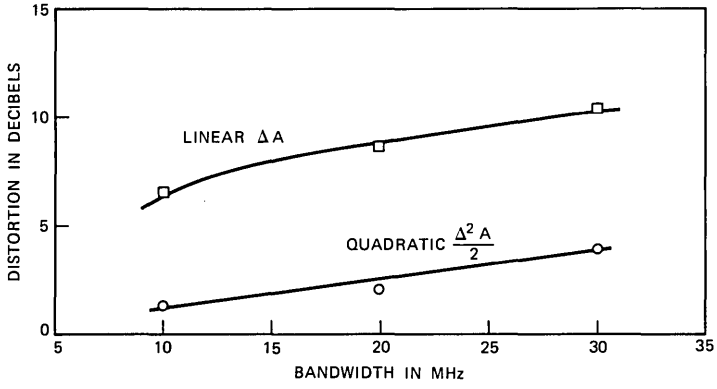

Fig. 20—Growth of linear and quadratic distortion with bandwidth for the comparative diversity channel for 10^{-5} of the time.

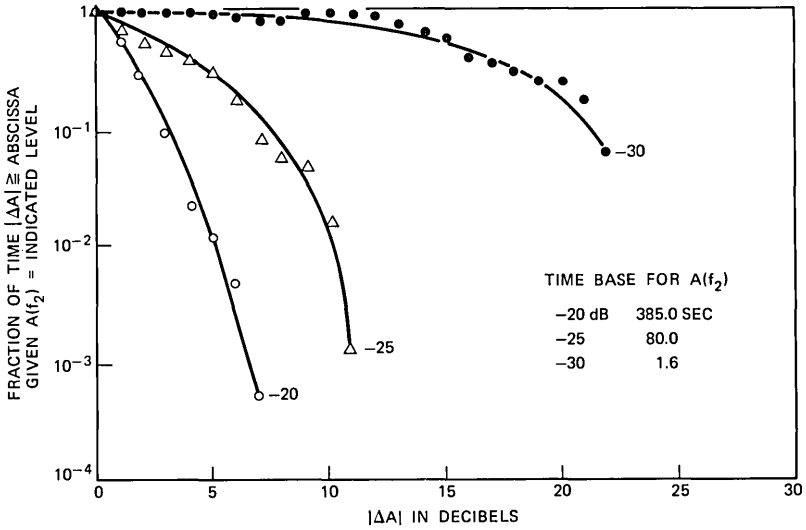

Fig. 21—Conditional linear distortion distributions for the 19.8-MHz comparative diversity channel.

of a decade of probability of occurrence per 5 dB change in distortion. The scatter and roll-up at the tail for the linear distribution represents several seconds of fading events at which time both channels were undergoing significant linear amplitude distortion. The distributions for other bandwidths of the comparative diversity channel were similar

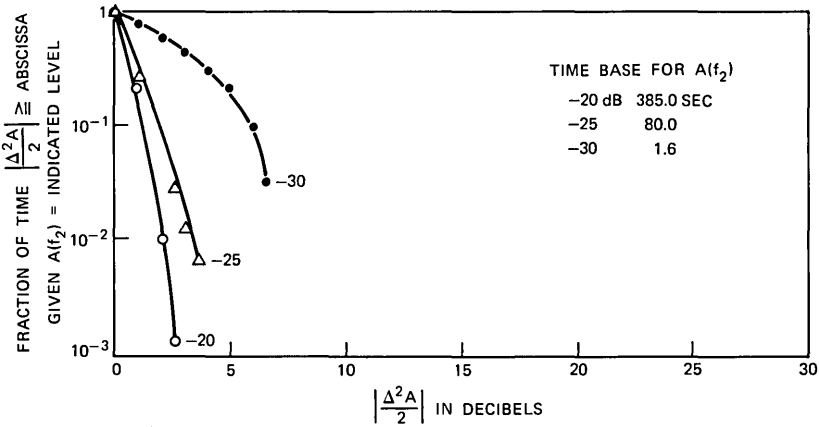

Fig. 22—Conditional quadratic distortion distributions for the 19.8-MHz comparative diversity channel.

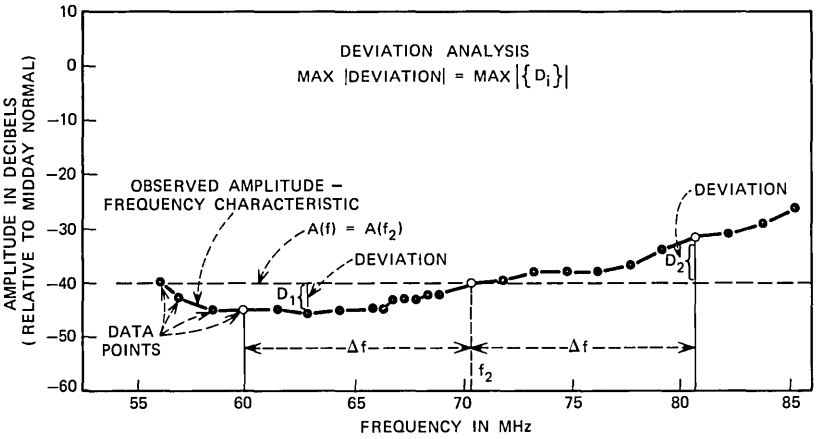

Fig. 23—Definition of deviations between the observed channel loss (solid line) and the zeroth degree, midchannel level approximation (dashed line).

in shape to those presented in Fig. 19 and again the rate of growth in distortion with bandwidth was about half the nondiversity channel's rate as indicated in Fig. 20. The comparative diversity channel was the result of 5873 switches between antennas.

The conditional distributions of distortion are shown in Figs. 21 and 22 and show approximately the same amounts of amplitude distortion as the nondiversity channel for fade depths not exceeding 30 dB. The distributions for the 30-dB level have been included for completeness but are less reliable because of too few samples.

6.4 *Error Analysis of the Higher-Order Distortion*

As previously discussed in Section V, the selectivity structure within a narrowband channel increases with fade depth, and for the nondiversity channel, the selectivity structure exceeds second order for fades in excess of about 30 dB. For the diversity channels constructed by threshold and comparative switching, we found that those fading events with large linear and quadratic amplitude distortion were significantly reduced. Clearly, if a nondiversity channel does not exceed second order for fades not in excess of 30 dB, a diversity channel not undergoing fades in excess of 30 dB will also not exceed second order. Thus, rather than compute and construct the statistical distributions for the maximum amplitude difference between the observed channel loss and a three-term power series quadratic approximation as we did for the nondiversity study, we choose to redefine the errors as

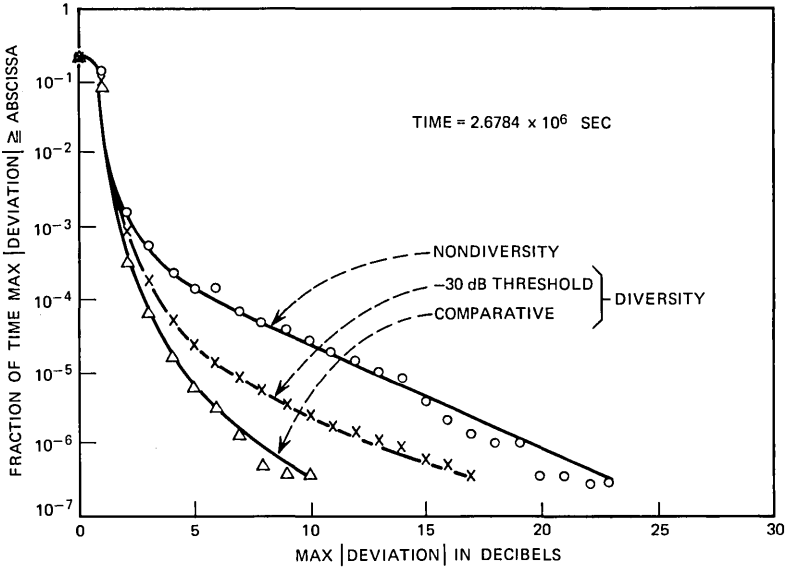

Fig. 24—Unconditional distributions of MAX|DEVIATION| for the nondiversity and diversity channels of bandwidth of 19.8 MHz.

the deviations, D, between the measured samples of transmission loss, $A^M(f_i)$, and a one-term power series approximation constructed from the midchannel loss $A(f_2)$. That is,

$$D_i = D(f_i) = \underbrace{A^M(f_i)}_{\substack{\text{Measured Samples of Transmission Loss}}} - \underbrace{A(f_2).}_{\substack{\text{Midchannel Loss}}} \tag{2}$$

Figure 23 shows the observed selectivity, the zeroth-degree, midchannel level approximation, and two of the deviations. The maximum of the inband amplitude deviation, called MAX|DEVIATION|, was monitored for both forms of diversity channels and is now presented in Fig. 24. For completeness and for comparison, the nondiversity channel's MAX|DEVIATION| distribution is included. Both forms of diversity channels show significant reduction in maximum deviation as compared to the deviation experienced by the nondiversity channel. The marked reduction in deviations for the diversity channels was a result of less deep fading events in the diversity channels.

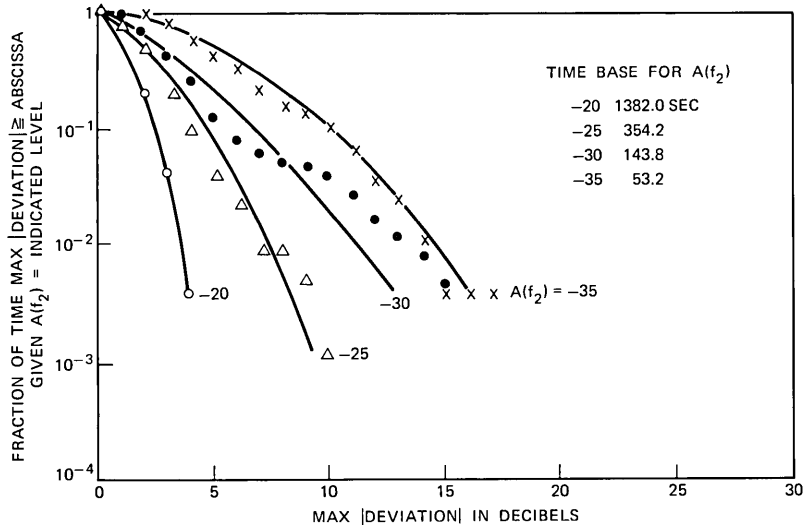

Fig. 25—Conditional distributions of MAX|DEVIATION| for the 19.8-MHz nondiversity channel.

The amplitude deviation distributions conditioned on the mid-channel fade depth for the diversity channels were similar to those for the nondiversity channel which are presented in Fig. 25. For fade depths greater than 10 dB, the observed amplitude selectivity structure of the diversity (and nondiversity) narrowband channels exceeded zeroth order by 2 dB. This is in agreement with the observation that both forms of diversity switching may significantly protect the radio channel from deep selective fading events but not from the less faded channels which experience linear amplitude distortion.

VII. ACKNOWLEDGMENTS

Indebtedness for the experimental data is extended to my colleagues W. T. Barnett, G. A. Zimmerman, and C. H. Menzel. The interest and support of E. E. Muller were invaluable.

REFERENCES

1. Vigants, A., "The Number of Fades in Space-Diversity Reception," B.S.T.J., *49*, No. 7 (September 1970), pp. 1513–1530.
2. Babler, G. M., "A Study of Frequency Selective Fading for a Microwave Line-of-Sight Narrowband Radio Channel," B.S.T.J., *51*, No. 3 (March 1972), pp. 731–757.
3. Lin, S. H., "Statistical Behavior of a Fading Signal," B.S.T.J., *50*, No. 10 (December 1971), pp. 3311–3270.
4. Barnett, W. T., "Multipath Propagation at 4, 6, and 11 GHz," B.S.T.J., *51*, No. 2 (February 1972), pp. 321–361.

Contributors to This Issue

G. M. BABLER, B.S., 1963, M.S., 1965, and Ph.D. (Physics), 1968, University of Missouri; Bell Laboratories, 1968—. Mr. Babler has done modeling and data analysis work on various aspects of electromagnetic wave propagation in random media. He presently is performing studies to more precisely quantify the atmospheric propagational constraints on line-of-sight microwave radio communication channels.

C. N. BERGLUND, B.Sc. (E.E.), 1960, Queen's University, Kingston, Ontario; M.S.E.E., 1961, Massachusetts Institute of Technology; Ph.D. (E.E.), 1964, Stanford University. Research Assistant, M.I.T., 1960–61; Research Associate, Department of Electrical Engineering, Queen's University, Kingston, 1961–62; Research Assistant, Stanford Electronics Laboratories, 1962–64. Bell Laboratories, 1964–72. At Bell Laboratories, Mr. Berglund was a supervisor in the Semiconductor Device Laboratory. Presently he is with Bell Northern Research, Ottawa, Canada. Member, APS.

RICHARD D. GITLIN, B.E.E., 1964, City College of New York; M.S., 1965, and D.Eng.Sc., 1969, Columbia University; Bell Laboratories, 1969—. Mr. Gitlin is presently concerned with problems in data transmission. Member, IEEE, Sigma Xi, Eta Kappa Nu, Tau Beta Pi.

DAVID J. GOODMAN, B.E.E., 1960, Rensselaer Polytechnic Institute; M.E.E., 1962, New York University; Ph.D., 1967, University of London; Bell Laboratories, 1960–62, 1967—. A member of the Acoustics Research Department, Mr. Goodman has studied principles of digital signal processing including analog-to-digital conversion and the statistical approach to digital filter design. Member, IEEE, Eta Kappa Nu, Tau Beta Pi.

LARRY J. GREENSTEIN, B.S.E.E., 1958, M.S.E.E., 1961, and Ph.D. (E.E.), 1967, Illinois Institute of Technology; Bell Laboratories, 1970—. Since joining Bell Laboratories, Mr. Greenstein has been engaged in studies of digital encoding, processing, and transmission. His current activities are in the area of radio communication. Member, AAAS, IEEE.

E. Y. Ho, B.S.E.E., 1964, The National Taiwan University; Ph.D., 1969, University of Pennsylvania; Bell Laboratories, 1969—. Mr. Ho has been engaged in developing and analyzing automatic equalizers for data transmission systems. Member, IEEE.

DIETRICH MARCUSE, Diplom Vorpruefung, 1952, Dipl. Phys., 1954, Berlin Free University; D.E.E., 1962, Technische Hochschule, Karlsruhe, Germany; Siemens and Halske (Germany), 1954–57; Bell Laboratories, 1957—. At Siemens and Halske, Mr. Marcuse was engaged in transmission research, studying coaxial cable and circular waveguide transmission. At Bell Laboratories, he has been engaged in studies of circular electric waveguides and work on gaseous masers. He spent one year (1966–1967) on leave of absence from Bell Laboratories at the University of Utah. He is presently working on the transmission aspect of a light communications system. Mr. Marcuse is the author of two books. Member, IEEE, Optical Society of America.

J. E. MAZO, B.S. (Physics), 1958, Massachusetts Institute of Technology; M.S. (Physics), 1960, and Ph.D. (Physics), 1963, Syracuse University; Department of Physics, University of Indiana, Bloomington, 1963–1964; Bell Laboratories, 1964—. Mr. Mazo has worked on various theoretical problems concerned with data transmission. Since 1972, he has worked at the Mathematical Research Center of Bell Laboratories, Member, American Physical Society, IEEE.

K. K. THORNBER, B.S., 1963, M.S. (E.E.), 1964, Ph.D. (E.E.), 1966, California Institute of Technology; Research Associate, Stanford Electronics Laboratories, 1966–68; Research Assistant, Physics Department, University of Bristol, 1968–69; Bell Laboratories, 1969—. Mr. Thornber is a member of the Semiconductor Device Laboratory. Member, Sigma Xi, Tau Beta Pi.

THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 2, February, 1973
Printed in U.S.A.

B. S. T. J. BRIEF

A New Optical Fiber

By P. KAISER, E. A. J. MARCATILI, and S. E. MILLER

(Manuscript received November 20, 1972)

Currently there is strong interest in optical fibers for use as a transmission medium, analogous to the use of coaxial or wire pairs in the low-frequency region. Most work is devoted to a fiber structure consisting of a central glass core surrounded by a cylindrical glass cladding having a slightly lower index of refraction. This in turn requires that the chemical composition of the core glass differs from that of the cladding glass, leading to undesired effects at the core-cladding interface and perhaps limiting the minimum fiber losses achievable.

The Nippon Sheet Glass Company and the Nippon Electric Company together have developed a fiber (which they call SELFOC) with an index of refraction decreasing parabolically from the fiber axis to its outer boundary. This fiber requires a continuous variation in chemical composition from the fiber axis outward, with attendant complications in the fabrication process. A related guide requiring a film very thin compared to the wavelength has just been reported.[1]

The unique property of the new fiber is that a viable, handleable transmission medium is created by a structural form that uses only a single low-loss material.

The conception was stimulated by the findings of P. Kaiser, et al., who fabricated unclad round fibers and measured their spectral losses in up to 32-meter unsupported lengths.[2] Recently he found total losses as low as 2.5 dB/km at wavelengths near 1.1 μm using selected samples of low OH content fused silica. Similarly low losses have been measured at 1.06 μm in bulk fused silica by T. C. Rich, et al.[3] It appeared attractive to use material of this kind without the need for modifying the composition to alter the refractive index as is necessary with conventional core-cladding fibers or with graded-index fibers.

Figure 1 shows section views of two possible forms of the single-material (SM) fiber. The usefully guided energy is concentrated pri-

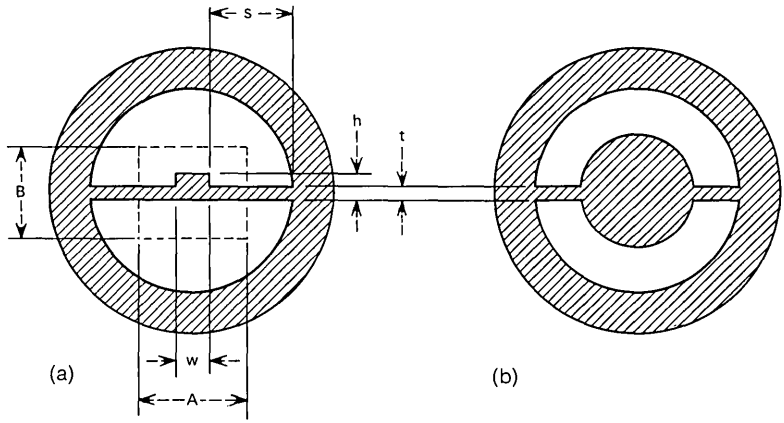

FIG. 1—Cross section of an SM fiber with (a) rectangular and (b) cylindrical core.

marily in the central enlargement, shown rectangular in Fig. 1a and round in Fig. 1b. In Fig. 1a, the central body has dimensions w and h for single-mode operation or dimensions A and B for multimode operation. There is an exponentially decaying field extending outward from the central member in the slab of thickness t; with appropriate spacing between the central enlargement and outer cylinder the guided-wave field at the outside surface can be made negligibly small and the fiber can be handled exactly as can the conventional core-cladding type fiber. Slab modes are possible on the supporting structure, but these are strongly coupled to the outer shell and are readily lost to the surrounding medium. Not shown in the figure is the possibility of adding an absorbing coating on the outer surface for avoidance of crosstalk in a multifiber cable.

The SM fiber structure can have a single propagating mode for any supporting slab thickness t and for any shape of the central enlargement, provided the size of the central enlargement is properly chosen. Practically, though, t must be limited in order to keep the slab dimension s, and consequently the overall size of the guide, reasonably small and still have the exponentially decaying slab field at the outer supporting cylinder small enough.

Analysis has been carried out for both single- and multimode SM fibers of several geometries. A few of the results are abstracted here. For the rectangular-guide case, Fig. 1a, and $t \gg \lambda$, there will be a single propagating mode provided

$$\frac{1}{h^2} + \frac{1}{w^2} \geqq \frac{1}{t^2}. \tag{1}$$

Note that wavelength does not appear in this expression, correct to first order. More exact analysis shows that for $t = 4.89\ \mu m$ and $h = 7.0\ \mu m$, the limiting width w for single-mode operation is $7.07\ \mu m$ at $\lambda = 1.0\ \mu m$ and $6.94\ \mu m$ at $\lambda = 0.6\ \mu m$. In these structures the slab field decays by $1/e$ in 2.80 and 2.75 μm at λ equal to 1.0 and 0.6 μm, respectively.

The wave propagation effects of the slab support can be represented by a uniform-index side support having the same height as the core and an equivalent index $n_e = n_c(1 - \Delta_s)$, where n_c is the index of the w-by-h core. Then, from the equality condition of eq. (1), it can be shown that

$$\Delta_s = \frac{1}{8}\left[\frac{\lambda}{wn_c}\right]^2. \tag{2}$$

Arbitrarily small values of Δ_s can be achieved by making w [and according to (1), also h and t] appropriately large.

For the multimode rectangular guide the number of guided modes may be shown to be

$$N = \frac{\pi}{2}\frac{AB}{t^2}\left[\frac{1}{1 + \left(\frac{\pi}{2v}\right)^2}\right] \tag{3}$$

where

$$v = \frac{\pi t}{\lambda}\sqrt{n_c^2 - n^2}, \tag{4}$$

and n is the index of the unshaded region outside the dotted region of Fig. 1a, and A and B are defined in the figure. Note that the number of modes, eq. (3), is (to first order) independent of wavelength–a unique property. For all modes the field decays exponentially along the slab as noted above; for the highest-order mode the field penetration is the largest and decays by $1/e$ in a length l, where

$$l = \frac{\sqrt{2}t}{\pi}\sqrt{1 + \left(\frac{\pi}{2v}\right)^2}. \tag{5}$$

For the multimode SM fiber the equivalent full-height support has an equivalent refractive index $n_c(1 - \Delta_m)$ where

$$\Delta_m = \frac{1}{8}\left(\frac{\lambda}{tn_c}\right)^2. \tag{6}$$

The value of Δ_m can be used to calculate numerical aperture, the tol-

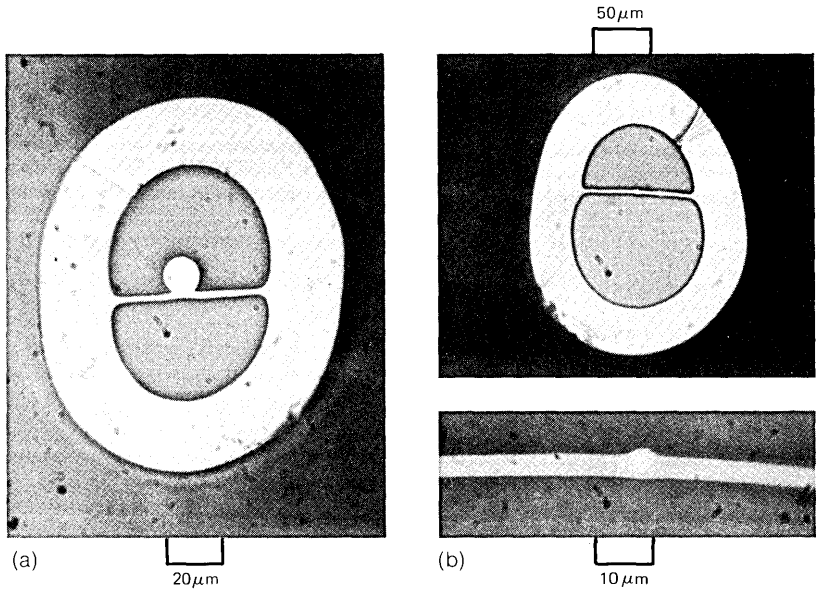

FIG. 2—Photographs of an experimental (a) multimode SM fiber and (b) single-mode SM fiber (top), with magnified core region (bottom).

erable radius of curvature, and modal dispersion. We give here only the numerical aperture,

$$\text{N.A.} = n_c\sqrt{2\Delta_m} = \frac{\lambda}{2t}. \tag{7}$$

SM fibers intended to approximate the geometries shown in Fig. 1 were drawn in an oxygen-hydrogen torch from 6.5 mm i.d., 10 mm o.d. fused quartz tubes containing thin, polished plates and small-diameter rods supported in the center of the tubes. Plates of about 0.2 mm thickness, and core rods of approximately 0.2 mm and 1 mm diameters, resulted in single- and multimode fibers, respectively, whose cross sections are shown in Figs. 2a and b.

For 15-μm-core multimode fibers, a slab thickness t varying between 3 and 4 μm resulted in numerical apertures ranging between 0.11 and 0.08 ($\lambda = 0.6328\ \mu$m), which agree excellently with the predicted values of 0.106 and 0.079, respectively [see eq. (7)]. The h/t ratio of the single-mode guide was about 6.5 μm/4 μm, or 1.625, with the width w amounting to 5 μm.

The spectral losses of the SM fibers were expected to closely approximate those of the unclad fibers drawn from the same material. Whereas

this was true for the general shape of the spectral loss curve which was determined between 0.5 and 1.15 μm, the minimum losses were generally higher. For a 300-m-long, Suprasil 2 multimode fiber they amounted to 39, 50, and 28 dB/km at 0.66, 0.80, and 1.06 μm, respectively. Lowest losses of a single-mode fiber having a slightly different geometry than that shown in Fig. 2b were 55 dB/km at 1.06 μm. We believe that residual contamination of the preform elements is the source of the excess losses.

Total scattering losses in the order of 7.5 dB/km at 0.6328 μm demonstrate that the approximately 30 modes (3) of the multimode fiber are well guided and do not lose power into the surrounding cladding to any significant degree.

Other applications of the SM-fiber principle appear promising. Active fiber guides can be created by putting the active material in the central core or by putting it in a liquid surrounding the central member. Integrated optical circuits can utilize the same structure. For example, for a core and slab of index 1.472, slab thickness $t = 0.98$ μm, and a surround index 1 percent less than 1.472, we find the single-mode limit at $h = 1.10$ μm and $w = 6.75$ μm at $\lambda = 0.6328$ μm; the field decays transversely in the slab by $1/e$ in 2.67 μm. Thicker slabs allow larger w and h with single-mode guidance. In early research on optical integrated circuits, J. E. Goell observed wave propagation in a curved guide of the above general form, now understood as another verification of the principle of the SM fibers.

The assistance of H. W. Astle in the fabrication of the SM fibers is gratefully acknowledged.

REFERENCES

1. Nishizawa, J., and Otsuka, A., "Solid-State Self-Focusing Surface Waveguide (Microguide)," Appl. Phys. Lett., *21*, No. 2 (July 15, 1972).
2. Kaiser, P., Tynes, A. R., Cherin, A. H., and Pearson, A. D., "Loss Measurements of Unclad Optical Fibers," presented at the Topical Meeting on Integrated Optics–Guided Waves, Materials and Devices, in Las Vegas, Nevada, February 7–10, 1972.
3. Rich, T. C., and Pinnow, D. A., "Total Optical Attenuation in Bulk Fused Silica," Appl. Phys. Lett., *20*, No. 7 (April 1, 1972), pp. 264–266.

Bell System