THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XLIV APRIL 1965 NUMBER 4

Automatic Equalization for Digital Communication

By R. W. LUCKY

(Manuscript received November 24, 1964)

Distortion in transmission channels causes intersymbol interference in digital communication systems. This distortion may be partially corrected at the receiver through the use of a tapped delay line having adjustable tap gain settings (transversal filter). The problem of minimizing distortion with a finite-length transversal filter is examined. In the region of small initial channel distortion where most existing systems operate, the best tap gain settings satisfy a set of simultaneous linear equations. For larger initial distortion, iterative techniques are required to find best gain settings. The distortion is shown to be a convex function of the tap gains, so mathematical programming techniques may be employed for optimization.

The practical problem is that of evolving a logical strategy whereby the tap gains of the transversal filter may be set to optimum values. An easily implemented system for automatic equalization is described which makes use of a steepest-descent technique of minimization. The equalizer is automatically set prior to data transmission in a training period during which a series of test pulses is transmitted. Only polarity information is required, so digital logic may be used in the equalizer. For application to high-speed data transmission, great accuracies are required for the tap gain settings. Thus the problem of noise in the channel during equalization is quite important. The final error due to noise and channel distortion and the equalizer settling time are discussed and evaluated. Finally, the effect of a transversal filter equalizer in terms of the system frequency-domain characteristics is considered.

I. INTRODUCTION

Present data rates on voice telephone channels are limited to about 2400 bits per second. Although the noise margin on these facilities is sufficient to permit much higher rates, the nonuniform transmission characteristics of the channel cause what might be termed a distortion barrier, prohibiting faster transmission. The distortion of data pulses by the channel results in these pulses being smeared out in time so as to overlap other transmitted pulses. This intersymbol interference is one of the chief degrading factors in present systems and becomes the determining factor in the design of higher-rate systems. To alleviate the effects of intersymbol interference it is necessary to equalize the channel.

In the past equalization has generally been accomplished by flattening the amplitude characteristic and linearizing the phase characteristic using fixed amplitude-frequency and phase-frequency networks. Although this type of equalization is adequate for speech transmission requirements, it does not provide the precise control over the channel time response which is necessary for high-speed data transmission. Thus to realize the full transmission capability of the channel there is a need for automatic, time-domain equalization.

Among the basic philosophies for automatic equalization of data systems are pre-equalization at the transmitter and post-channel equalization at the receiver. Since the former technique requires a feedback channel, we will concentrate our efforts here on equalization at the receiver. This equalization can be performed either during a training period prior to data transmission or it can be performed continuously during data transmission. The typical voice channel changes little during an average data call, so pre-call equalization should be sufficient in most cases. Many of the principles and techniques which will be discussed here can be applied to the continuous or adaptive equalizer, although our main concern will be in the pre-call automatic equalizer.

Most engineers are agreed that multilevel vestigial sideband transmission offers the best hope for higher-speed transmission on voice channels. There are good theoretical reasons for this choice of modem on a channel of limited bandwidth and high signal-to-noise ratio. The equalization problem we will discuss is based on the use of VSB transmission or what is equivalent, baseband transmission. The equalizer is to be placed at the receiver directly after the demodulation process. Thus as far as the equalizer is concerned the transmission is baseband.

Suppose now that the equivalent baseband system transmits amplitude a_n at time nT, where a_n is chosen from a set of M possible discrete amplitudes. The single-pulse response of the over-all system, including

channel and equalizer, we call $h(t)$. The received signal $y(t)$ is then

$$y(t) = \sum_{n=-\infty}^{\infty} a_n h(t - nT). \tag{1}$$

At some suitably chosen sampling time t_0 the output voltage $y(t_0)$ may be abbreviated

$$y_0 = \sum_{n=-\infty}^{\infty} a_n h_{-n} \tag{2}$$

where

$$h_n = h(t_0 + nT). \tag{3}$$

This voltage is the sum of the wanted term a_0 plus an intersymbol interference term

$$y_0 = h_0 \left[a_0 + \frac{1}{h_0} \sum_{n=-\infty}^{\infty}{}' a_n h_{-n} \right]. \tag{4}$$

(The prime will be used very frequently on summations to indicate deletion of the $n = 0$ term.)

Now in the second term of (4) the a_n coefficients are chosen by the data system user according to some probabilistic rule. Since in a VSB system the maximum positive and negative of values of a_n are equal in magnitude, say $a_{\max}$, the maximum value the intersymbol interference term can assume is

$$\text{max interference} = \frac{a_{\max}}{h_0} \sum_{n=-\infty}^{\infty}{}' |h_n|. \tag{5}$$

We thus define an interference criterion proportional to (5) called distortion and labelled D

$$D = \frac{1}{h_0} \sum_{n=-\infty}^{\infty}{}' |h_n|. \tag{6}$$

The so-called eye opening for an M-level VSB system is simply and monotonically related to D.

$$I = 1 - (M - 1)D. \tag{7}$$

The eye opening is a widely accepted criterion of a data system's performance. In what follows we shall always use the equalizer to minimize the distortion D defined by (6). A particular advantage of this criterion is that it is not dependent on noise statistics or on the statistical distribution of the customer's data sequence $\{a_n\}$. It is a minimax

criterion in that we seek to maximize the customer's minimum margin against noise over all data sequences. A heuristic argument can be made for a criterion using the sum of the squares of the h_n's. Obviously neither criterion minimizes average probability of error — a mathematically intractable problem even when assuming Gaussian noise and independent, equally likely input symbols. Practically speaking the minimization of either criterion leads to negligibly different results.

Now what we need is a variable filter for an equalizer which can exercise wide control over the time response samples $\{h_n\}$. The transversal filter shown in Fig. 1 is ideal for this purpose.[1,2] This filter consists of a continuous delay line tapped at T-second intervals. Each tap has an associated attenuator which in combination with an inverter is capable of giving variable gain. The filter output is the sum of all the attenuated tap voltages. Such a filter is capable of flexible control of the output time sequence $\{h_n\}$ when used in tandem with the channel to be equalized. In fact, under suitable conditions, derived in the final section of this paper, a transversal filter of infinite length may be used to completely eliminate distortion in a channel. Generally speaking, a transversal filter of finite length cannot eliminate distortion, so we will

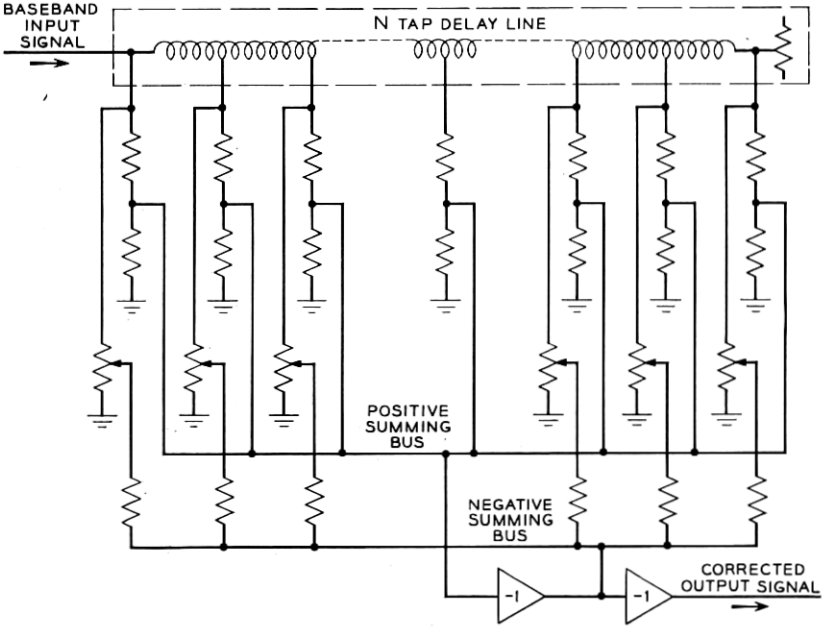

Fig. 1 — Elements of a transversal filter.

study the minimization of distortion D with a transversal filter of finite length. We then show an implementation of an automatic equalizer which achieves minimum distortion under certain conditions generally satisfied on voice telephone channels.

In succeeding sections the theoretical and practical limitations of the equalizer as implemented are studied. In order to effectively eliminate distortion in voice channels great accuracies are required in the setting of the tap gain coefficients. Background noise and settling time for the equalizer become extremely important. Finally we discuss the behavior of the transversal equalizer, essentially a time-domain device using a time-domain criterion, in the frequency domain.

II. THE MINIMIZATION OF DISTORTION

We will assume that the transversal filter has $N + 1$ taps with associated tap gains. One of these taps is taken as the reference with its gain denoted c_0, while the other N taps are placed somewhere at integer multiples of T seconds before or after tap c_0. The positions of these taps are usually the locations $-(N/2)T$, $(-(N/2) + 1)T$, $\cdots$, $(N/2)T$, but since these particular locations are not necessary to any of the theorems we use a more general formulation. Let K_N be the set of integers denoting the positions of the $N + 1$ tap locations

$$K_N = \{n \mid \text{tap location exists at time } nT \text{ from reference}\}.$$

The tap gains will be denoted c_j, $j \subset K_N$.

The impulse response at the input to the transversal filter is denoted $x(t)$ and its samples at times nT form the time sequence $\{x_n\}$. We will assume for convenience that this response is normalized so that $x_0 = 1$. Thus the distortion in the pulse $x(t)$ prior to equalization, called initial distortion D_0, is

$$D_0 = \sum_{n=-\infty}^{\infty}{}' |x_n|. \tag{8}$$

The transversal filter serves as a device to multiply time sequences. In this case the input sequence $\{x_n\}$ is multiplied by the tap gain sequence $\{c_n\}$ according to the rules of polynomial multiplication. It can be seen by inspection that the output sequence $\{h_n\}$ is formed by the rule

$$h_n = \sum_{j \subset K_N} c_j x_{n-j}. \tag{9}$$

The final distortion we seek to minimize is

$$D = \frac{1}{h_0} \sum_{n=-\infty}^{\infty}{}' \mid h_n \mid . \tag{10}$$

The reference response sample h_0 causes somewhat of a practical problem. In a multilevel system the slicing levels are generally fixed so that the gain must be closely controlled. In other words it is necessary to provide a normalizing control to adjust h_0 to unity, the assumed value of x_0. This can be done two ways which lead to slightly different mathematical problems.

Problem 1:

Fix the tap c_0 at unity. Minimize the criterion D over the N variables $\{c_j\}; j \subset K_N, j \neq 0$. The output pulse may be normalized if desired by an over-all gain control outside the transversal filter.

Problem 2:

Let the tap c_0 be variable. Minimize the criterion D over the $N + 1$ variables $\{c_j\}, j \subset K_N$, subject to the constraint $h_0 = 1$.

A study of these two problems reveals that the minimum distortion in either case is the same. Except in isolated cases ($c_0 = 0$ in Problem 2) the optimum tap gains in Problems 1 and 2 are related by a constant factor. There is a practical difference in the range of tap gains required and a mathematical difference in the type of function involved. (Problem 1 is nonlinear, while 2 is piecewise linear.) Since the minima are the same and the optimum gains related, we concentrate on Problem 2 because of its simpler properties.

The constraint on h_0 may be written

$$h_0 = 1 = \sum_{j \subset K_N} c_j x_{-j} . \tag{11}$$

Solving for the gain c_0 we have

$$c_0 = 1 - \sum_{j \subset K_N}{}' c_j x_{-j} . \tag{12}$$

We substitute (12) into (9) to obtain

$$h_n = \sum_{j \subset K_N}{}' c_j (x_{n-j} - x_n x_{-j}) + x_n . \tag{13}$$

Now since h_0 is unity the distortion (10) becomes

$$D = \sum_{n=-\infty}^{\infty}{}' \mid \sum_{j \subset K_N}{}' c_j (x_{n-j} - x_n x_{-j}) + x_n \mid . \tag{14}$$

Now that c_0 has been eliminated to satisfy the constraint on h_0 we desire to minimize D in (14) over the N variables c_j ; $j \subset K_N$, $j \neq 0$.

First observe that D is a continuous, piecewise-linear function of the variables $\{c_j\}$. We can rewrite (14) in the form

$$D = \sum_{j \subset K_N}{}' c_j \sum_{n=-\infty}^{\infty}{}' (x_{n-j} - x_n x_{-j}) \operatorname{sgn} h_n + \sum_{n=-\infty}^{\infty}{}' x_n \operatorname{sgn} h_n \quad (15)$$

$$\operatorname{sgn} h_n = \begin{cases} +1, & h_n \geqq 0 \\ -1, & h_n < 0. \end{cases}$$

In this equation the coefficients of the c_j's are constant over certain regions of the N-dimensional space of definition of $\{c_j\}$. Breakpoints where the coefficients assume new values occur whenever an output sample h_n is zero. A minimum cannot occur between breakpoints where the function is linear; thus at least one value h_{k_1} is zero at the minimum. The equation $h_{k_1} = 0$ may be used to eliminate one of the N variables c_i . The reduced equation is of the same piecewise linear form, requiring at least one more output sample $h_{k_2} = 0$, etc. We arrive at the conclusion that at least N samples of the output time sequence $\{h_n\}$ must be zero at the minimum. But N equations of the form $h_{k_i} = 0$; $i = 1, \cdots, N$ are sufficient to determine the values of the tap gains $\{c_j\}$. We need only solve N simultaneous linear equations using (13) with $n = k_i$; $i = 1, \cdots, N$.

The question remains as to the values of the k_i's, i.e., which N zeros in the output response sequence does one force to achieve minimum distortion? In most cases of interest this question is answered by the following theorem, which is proved in Appendix A.

Theorem I

If $D_0 < 1$, then the minimum distortion D must occur for those N tap gains which simultaneously cause $h_n = 0$ for all $n \subset K_N$, $n \neq 0$.

Another important property of the distortion function which is both useful and descriptive is the following:

Theorem II

If the tap c_0 is used to satisfy the constraint $h_0 = 1$, then the distortion D is a convex function of the N variables c_j ; $j \subset K_N$; $j \neq 0$.

Proof

For convenience denote settings of the equalizer by the N-component vectors $\bar{\alpha}$ and $\bar{\sigma}$. To prove convexity of D it is necessary to show that

for any two settings $\bar{\alpha}$ and $\bar{\sigma}$ and for all λ, $0 \leqq \lambda \leqq 1$,

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] \leqq \lambda D(\bar{\alpha}) + (1 - \lambda)D(\bar{\sigma}). \tag{16}$$

This equation would show that the distortion always lies on or beneath a chord joining values of distortion in N-space.

From (14)

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] = \sum_{n=-\infty}^{\infty}{}' \Big| \sum_{j \subset K_N}{}' [\lambda\alpha_j + (1 - \lambda)\sigma_j] \cdot (x_{n-j} - x_n x_{-j}) + x_n \Big| \tag{17}$$

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] = \sum_{n=-\infty}^{\infty}{}' \Big| \lambda \Big\{ \sum_{j \subset K_N}{}' \alpha_j(x_{n-j} - x_n x_{-j}) + x_n \Big\} + (1 - \lambda) \Big\{ \sum_{j \subset K_N}{}' \sigma_j(x_{n-j} - x_n x_{-j}) + x_n \Big\} \Big| \tag{18}$$

$$D[\lambda\bar{\alpha} + (1 - \lambda)\bar{\sigma}] \leqq \lambda D(\bar{\alpha}) + (1 - \lambda)D(\bar{\sigma}). \tag{19}$$

One of the most important properties of convex functions is that they possess no relative minima other than their absolute minimum. Thus any minimum of D found by systematic search or other mathematical programming methods must be the absolute (or global) minimum of distortion.

In summary, we have shown that D is a continuous, piecewise-linear, convex function of the N tap gains. This function has a single minimum which must occur when N zeros appear in the output time sequence $\{h_n\}$. If the initial distortion is less than 100 per cent ($D_0 < 1$), then the minimum occurs when the N samples of the output time sequence which correspond in location to the N taps on the transversal filter are simultaneously zero. This description is illustrated in two simple cases in Figs. 2 and 3. In Fig. 2 only one tap is variable ($N = 1$), while in Fig. 3 equal distortion contours are plotted for an example of a 2-tap equalizer. In both cases the initial distortion is less than 100 per cent, so the minima are easily located.

III. EQUALIZATION STRATEGIES AND IMPLEMENTATION

3.1 *Strategy When $D_0 < 1$*

The condition $D_0 < 1$, which is sufficient for easy location of the minimum distortion, is equivalent to the condition that the unequalized channel is capable of supporting binary transmission without error in the absence of noise. It may be seen from (7) that $D_0 = 1$ implies a

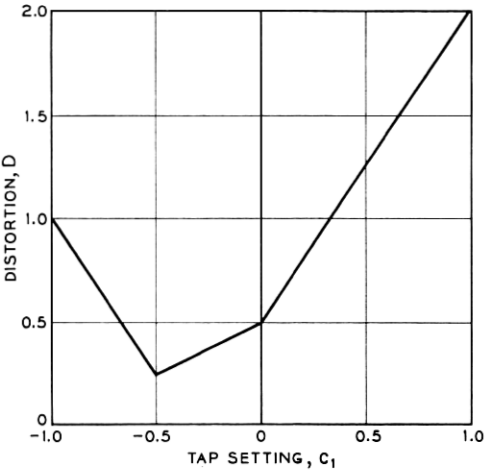

Fig. 2 — Distortion vs tap setting in a one-dimensional example.

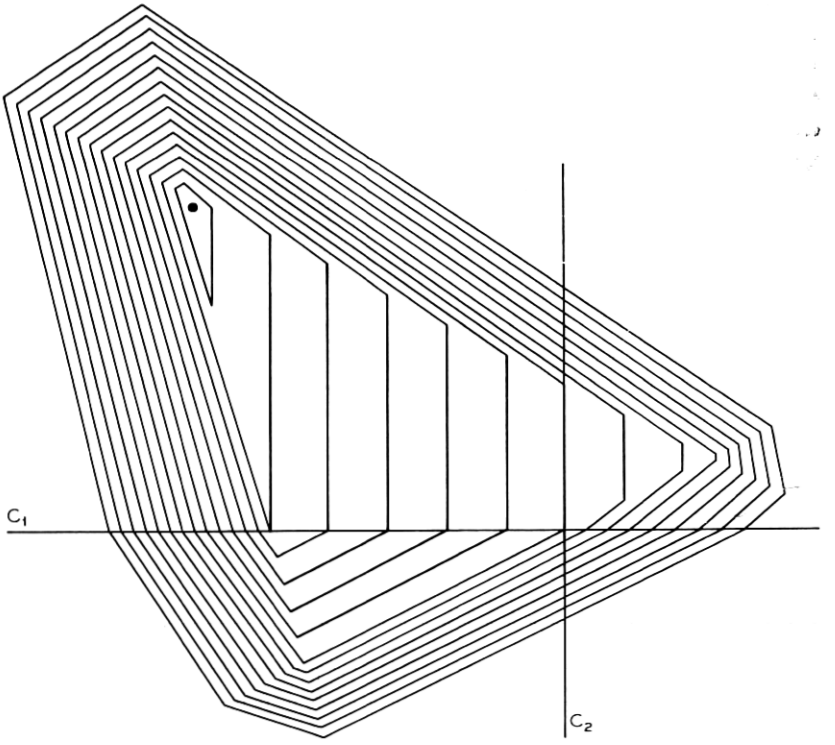

Fig. 3 — Equal distortion contours for a two-tap example.

completely closed binary eye. In almost all cases of interest on voice telephone channels the modem will support binary transmission without equalization, and so the condition is met. In these cases equalization enables multilevel operation to take advantage of a relatively high signal-to-noise ratio. For the present we will deal with the design of an equalizer where $D_0 < 1$.

To make matters more concrete we will henceforth assume that the N adjustable taps of the transversal filter are divided equally before and after the reference tap. In the several equalizer models which have been built the normalization has been carried out sometimes by an outside gain control and sometimes by an adjustable center tap. In either case the task of the automatic tap gain setting apparatus is to zero the N output samples h_n ; $| n | \leqq N/2$, $n \neq 0$. By Theorem I this achieves minimum distortion. One can, of course, derive the tap gains by the simultaneous solution of N linear equations, but from an instrumentation point of view there are many simpler schemes. The condition $D_0 < 1$ ensures a "loosely coupled" system where the interaction between tap gains is weak. Thus there are a number of iterative strategies which converge to the desired settings. In all these schemes a sequence of test pulses is transmitted prior to actual data transmission. After each test pulse the tap gains of the equalizer are readjusted in such a way as to eventually result in the proper N zeros in the output time sequence. The choice between such schemes is dictated by ease of instrumentation, settling time required for equalization and accuracy of the final settings.

It is true that with a special-purpose computer the optimum tap settings could be computed using only a single test pulse; however, this overlooks the presence of background and impulse noise on the facility. Each test pulse is in itself unreliable, so many pulses must be averaged in some way to give accurate settings. The equalization system must also be relatively unaffected by any large bursts or impulses which occur during the setup period.

An equalization strategy has been devised which meets all requirements and is easily instrumented. The motivation for this strategy is based on a steepest-descent technique. Consider the use of an outside gain control for normalization with the center tap gain fixed at unity. Solving the N simultaneous equations $h_n = 0$ for $| n | \leqq N/2$, $n \neq 0$ is equivalent to minimizing the "truncated distortion"

$$D_N = \sum_{n=-N/2}^{N/2}{}' | h_n |. \tag{20}$$

The function D_N is a convex function amenable to solution by steepest-descent techniques. After each test pulse has been received the tap gains are incremented so that the N-dimensional incrementing vector is in a direction opposed to the gradient of D_N. This gradient may be written

$$\bar{\nabla} D_N = \sum_{j=-N/2}^{N/2}{}' \frac{\partial D_N}{\partial c_j} \bar{a}_j \tag{21}$$

where $\bar{a}_j$ is a unit vector in the direction of the c_j coordinate. The components of the gradient are

$$\frac{\partial D_N}{\partial c_j} = \sum_{n=-N/2}^{N/2}{}' \frac{\partial h_n}{\partial c_j} \operatorname{sgn} h_n \tag{22}$$

$$\frac{\partial D_N}{\partial c_j} = \sum_{n=-N/2}^{N/2}{}' x_{n-j} \operatorname{sgn} h_n . \tag{23}$$

Now we approximate (23) by assuming the samples x_n, for $n \neq 0$, are small in comparison to x_0, which is unity.

$$\frac{\partial D_N}{\partial c_j} \approx \operatorname{sgn} h_j \tag{24}$$

$$\bar{\nabla} D_N \approx \sum_{j=-N/2}^{N/2}{}' \operatorname{sgn} h_j \bar{a}_j . \tag{25}$$

The approximation has resulted in an extremely simple expression, since all steps are of equal magnitude and the direction of each step is determined by simply taking the polarity of the corresponding output sample. No analog voltages are involved, so digital logic can be used in the tap gain setting circuitry. Before describing this circuitry, it is necessary to demonstrate that this iterative scheme does indeed converge to the desired minimum. After each test pulse, each tap gain c_j is incremented by an amount $-\Delta \operatorname{sgn} h_j$, so that the new output samples are

$$h_n^* = \sum_{j=-N/2}^{N/2}{}' (c_j - \Delta \operatorname{sgn} h_j) x_{n-j} + x_n \tag{26}$$

$$h_n^* = h_n - \Delta \sum_{j=-N/2}^{N/2}{}' \operatorname{sgn} h_j x_{n-j} \tag{27}$$

$$h_n^* = h_n - \Delta \operatorname{sgn} h_n - \Delta \sum_{\substack{j=-N/2 \\ j \neq n}}^{N/2}{}' \operatorname{sgn} h_j x_{n-j} \tag{28}$$

$$| h_n^* | \leqq | | h_n | - \Delta | + \Delta D_0 . \tag{29}$$

Since $D_0 < 1$, we have

$$| h_n^* | < | | h_n | - \Delta | + \Delta. \tag{30}$$

If $| h_n | > \Delta$, then

$$| h_n^* | < | h_n | \tag{31}$$

and each output sample that we desire to zero is always decreased.

On the other hand, if $| h_n | < \Delta$, then

$$| h_n^* | < 2\Delta. \tag{32}$$

Thus the process must converge to within an error of 2Δ on each output sample. As the step size Δ goes to zero the truncated distortion D_N approaches zero, which implies that the over-all distortion D has been minimized.

3.2 *Equalizer Implementation*

An equalizer implementing this strategy is shown in Fig. 4. This equalizer employs a 13-tap delay line with 6 variable taps on either side of the reference tap. The action of the equalizer is as follows: A succession of test pulses is sent through the transmission line and transversal filter. As each test pulse comes out of the transversal filter it is sliced to retain only polarity information and then sampled at T-second intervals. These polarity samples are stored in a 12-stage shift register. When the shift register is full, a gate is opened, and all taps are simultaneously adjusted one step up or down in attenuation according to the polarities in the shift register. At this time a pulse height or normalization adjustment is also made on the over-all gain by means of the upper slicer-sampler circuit shown in Fig. 4. The method of obtaining electronically controlled steps of attenuation for the equalizer taps is shown in Fig. 5. It can be seen that reversible counters, directed by the 12-stage shift register, control the attenuation of a ladder network by means of relays.

In the particular equalizer which has been constructed, the quantum steps on the tap gain controls are about 0.25 per cent ($\Delta = 0.0025$). For high-quality voice channels utilized by the VSB data system, about 100 test pulses are required for settling of the tap attenuators to optimum values. This settling time depends on the size of the quantum step, the initial distortion, and the channel noise. The settling time and residual distortion are the subject of the following section on

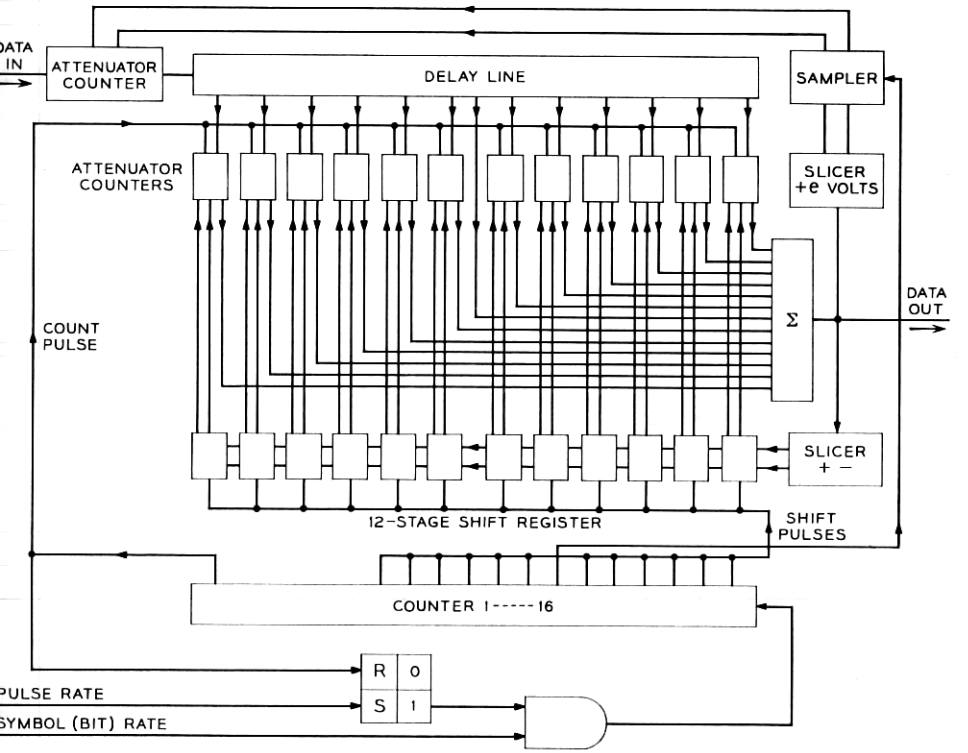

Fig. 4 — Automatic equalizer.

equalizer performance. Note that the equalizer is relatively unaffected by large impulse noise, since the output voltage is sliced and a large impulse can cause only one wrong step, which is subsequently corrected.

Figs. 6 and 7 show the results of equalization in a typical example using the 12-tap automatic equalizer. The single-pulse response together with the binary and 8-level eye patterns are shown before and after equalization. Before equalization only binary transmission was possible in this example. Since the equalizer has enabled 8-level operation in place of binary, a threefold increase in speed capability of the channel has been obtained.

3.3 *The Initial Distortion Limitation*

Thus far we have only considered equalization strategy when the initial distortion is less than 100 per cent. When $D_0 > 1$, there are two

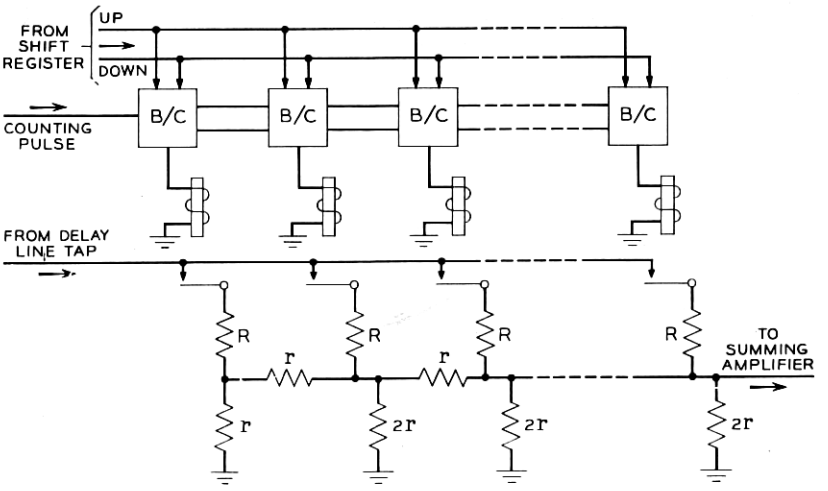

Fig. 5 — Reversible counter-controlled attenuator.

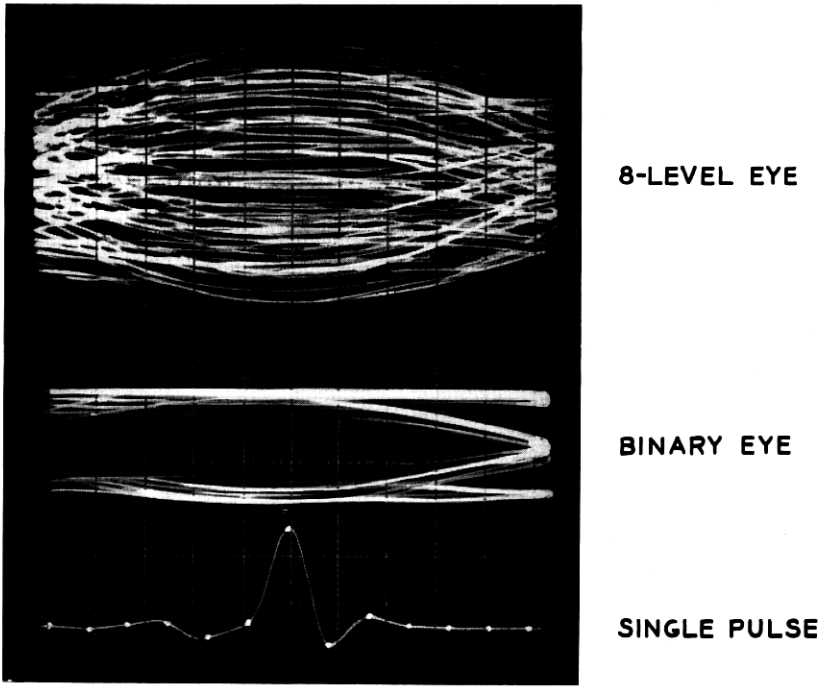

Fig. 6 — Unequalized pulse and corresponding "eye patterns."

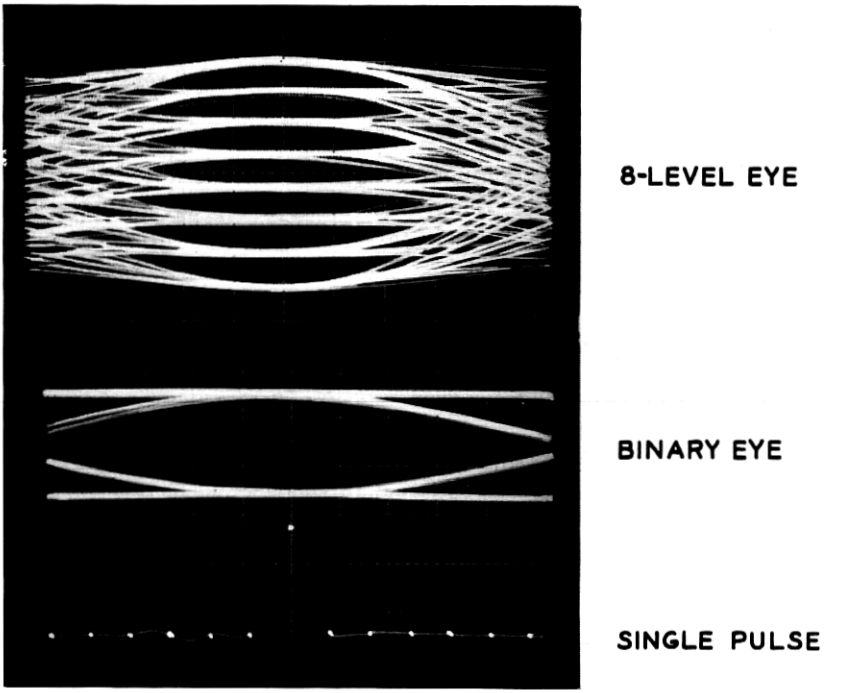

Fig. 7 — Equalized pulse and corresponding "eye patterns."

sources of failure for this equalizer. First, the equalizer convergence algorithm may fail to converge to the tap gains which force zeros in the output response for $t = -(N/2)T, \cdots, (N/2)T$. Second, these settings may not be the optimum (minimum distortion) gains. In either case the condition $D_0 < 1$ is *sufficient* but not *necessary* to prevent failure, and it is quite possible that the equalizer will converge and be optimum over a wider range of inputs.

In computer simulations of the equalizer strategy it has thus far been impossible to induce failure of the convergence algorithm without causing failure in optimality. Examples of the converse are, however, easily constructed. One such example is illustrated in Figs. 8 and 9. The channel to be equalized in this example has a 50 per cent cosine roll-off amplitude characteristic and a linear delay characteristic. The final distortion after equalization with an N-tap equalizer is shown in Fig. 8 for two different values of peak delay. With a peak delay of 5 pulse intervals the initial distortion $D_0 = 1.4$. In this case it happens

that the equalizer converges and is optimum. The more taps N which are used on the equalizer, the lower the final distortion.

When the peak delay in this example is increased to 6 pulse intervals, we encounter a radically different behavior. The equalizer still converges to the tap settings which satisfy $h_n = 0$ for $|n| \leqq N/2, n \neq 0$, but this solution no longer minimizes distortion. The initial distortion here is about 2.1 and after equalization with a 6-tap equalizer the final distortion is over 3. Clearly this is not optimum, since the zero tap settings gave better performance. As is shown, increasing the number of taps results in larger residual distortion.

The equalizer's failure in this latter case is more clearly illustrated in Fig. 9. The impulse response before equalization looks harmless enough, but after the equalizer has forced 3 zeros on either side of the response peak (6-tap equalizer) a large side lobe is created outside the equalizer time range on the right-hand side of the response. This side lobe contains more distortion than was present in the original response. In this

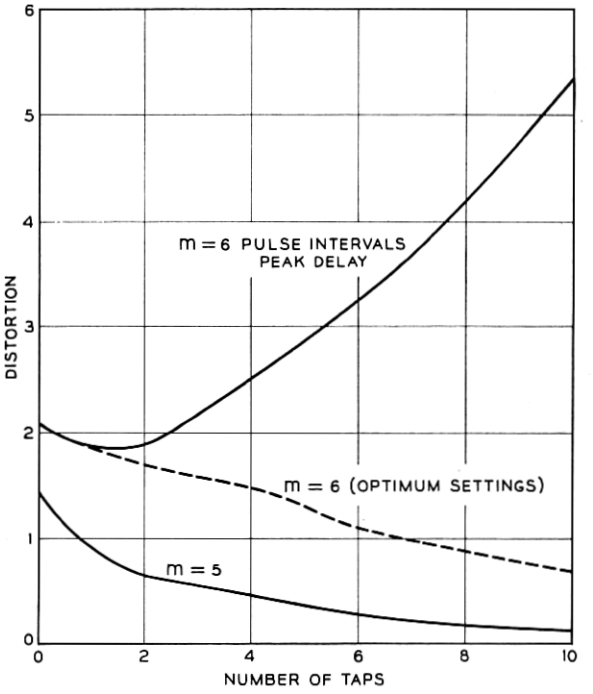

Fig. 8 — Example of equalizer failure—linear delay.

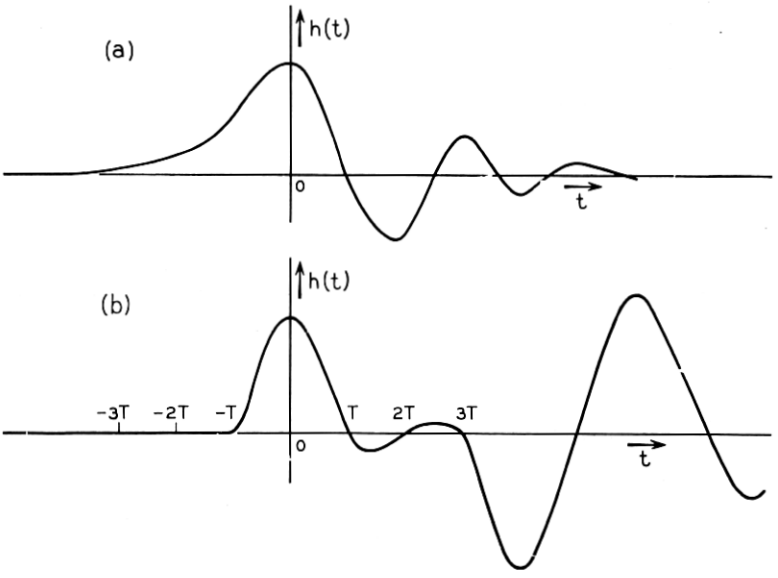

Fig. 9 — (a) Impulse response for channel with linear delay, m = 6; (b) after equalization with 6-tap equalizer.

example minimum distortion settings of the equalizer do not set zeros at the samples h_n for $| n | \leqq (N/2)$, $n \neq 0$. The actual optimum settings may be obtained by methods we will discuss presently, and for comparison purposes the minimum obtainable distortion for this example is plotted in Fig. 8.

3.4 *Strategy When* $D_0 > 1$

We described the situation when the initial distortion was less than 100 per cent as "loosely coupled." In this range many iterative schemes can be devised which converge to the optimum. Conversely, when $D_0 > 1$ the tap gains become more strongly interdependent. Also, the optimum may shift so that it is impossible to simply instruct the automatic equalizer to force N zeros in the impulse response at the locations of the N variable taps of the equalizer. Thus the equalization strategy for the general case becomes more subtle and complicated than the simple strategy previously explained.

Fortunately, Theorem II describes D (using center-tap normalization) as a convex function of the tap gains. Therefore the distortion has a single minimum, and mathematical programming methods may be

used to locate the optimum tap gains. A steepest-descent method may be applied to D [see (14)]. The gradient of D is

$$\bar{\nabla} D = \sum_{j=-N/2}^{N/2}{}' \frac{\partial D}{\partial c_j} \bar{a}_j \tag{33}$$

$$\frac{\partial D}{\partial c_j} = \sum_{n=-\infty}^{\infty}{}' (x_{n-j} - x_n x_{-j}) \operatorname{sgn} h_n . \tag{34}$$

If each tap gain c_j is incremented proportional to its gradient component (34), the system will eventually approach the optimum tap gains. The dotted lines showing minimum distortion in Fig. 8 were calculated using this strategy on a digital computer.* Such a strategy can be implemented with a modification of the circuitry previously described. In this implementation the term involving $x_n x_{-j}$ in (34) is neglected as small in comparison to x_{n-j}. The remainder of (34) may be derived physically by a two-pulse cycle. The first test pulse is sliced to obtain the sequence $\{\operatorname{sgn} h_n\}$ which is stored in the shift register. This sequence is then used to control the polarity of the gains at each tap. A second test pulse is transmitted and the transversal filter serves to multiply the input sequence $\{x_n\}$ by the tap gain sequence $\{\operatorname{sgn} h_n\}$. The output voltage is approximately the sequence $\{\partial D/\partial c_j\}$ from (34) and may be used to digitally increment the tap attenuators.

In any such scheme it is now necessary to have a transversal filter about twice as long as the number of variable taps to be used. This is because the sum in (34) is infinite, but practically speaking an N-tap transversal filter will affect the impulse response for not more than $2NT$ seconds. Thus the test pulses must go through a $2N$-tap delay line, all of whose taps are equipped to handle the ± 1 gains which store the sequence $\{\operatorname{sgn} h_n\}$. However, only the inner N taps need have associated variable attenuators. This complication comes about because the distortion depends on what happens outside the NT range (witness Fig. 9!) as well as what happens inside.This information must be measured by "listening posts" established by taps on the delay line outside the normal range and then taken into consideration in incrementing the variable tap gains.

IV. THEORETICAL PERFORMANCE OF THE EQUALIZER

4.1 *Design and Performance Parameters*

Our attention will now be confined to the equalizer described in Section 3.2 and illustrated in Fig. 4. In this fairly simple system there

* The distortion minimization can also be formulated as a linear programming problem, so that using a digital computer the exact optimum can be reached in a finite number of steps.

are only two parameters which the designer has under his control to affect the performance and cost of the system. These are:

N — the number of variable taps on the transversal filter.

Δ — the step spacing on each electronically controlled attenuator. (More generally this is Δ_j, since different tap positions may take different-sized steps.)

The economics of the choice of these two parameters may be readily appreciated. The cost of the equalizer is nearly directly proportional to N, since not only does a larger N entail a proportionally longer delay line, but also proportionally more logic circuitry. A major portion of this logic circuitry is taken by the reversible counters for setting the attenuator tap coefficients. To decrease the step spacing Δ, each of the counters must be augmented by additional stages.

In judging the performance of the equalizer we shall be interested in the following two parameters;

D_r — the final distortion or residual distortion after equalization.

T_s — the settling time or the time required to set the equalizer during the training period.

Obviously, the smaller the residual distortion D_r the better the data system will perform, but on the other hand it is too expensive to attempt to reduce this residual distortion much below the required tolerance for a given system. The settling time T_s on an experimental implementation of the automatic equalizer has been on the order of a second. When compared with the time required for establishing the call and acquiring timing synchronization this seems negligible. However, we shall find that when greater accuracy is required the time T_s can become quite appreciable.

The performance parameters D_r and T_s depend on the design parameters N and Δ and upon the channel to be equalized. Since it has been common to describe the channel by frequency-domain characteristics we shall do so here. The channel characteristics of interest are:

$A(\omega)$ — the amplitude characteristic of the system, including transmitter shaping filter and the attenuation characteristic of the channel.

$\beta'(\omega)$ — the delay characteristic of the channel.

S/N — the signal-to-noise ratio of the channel. The noise is assumed to be Gaussian and white since the effects of impulse noise are not particularly important in the automatic equalization system.

Fig. 10 shows how the parameters N and Δ affect the performance of a noiseless system example. In these curves the distortion is plotted as a

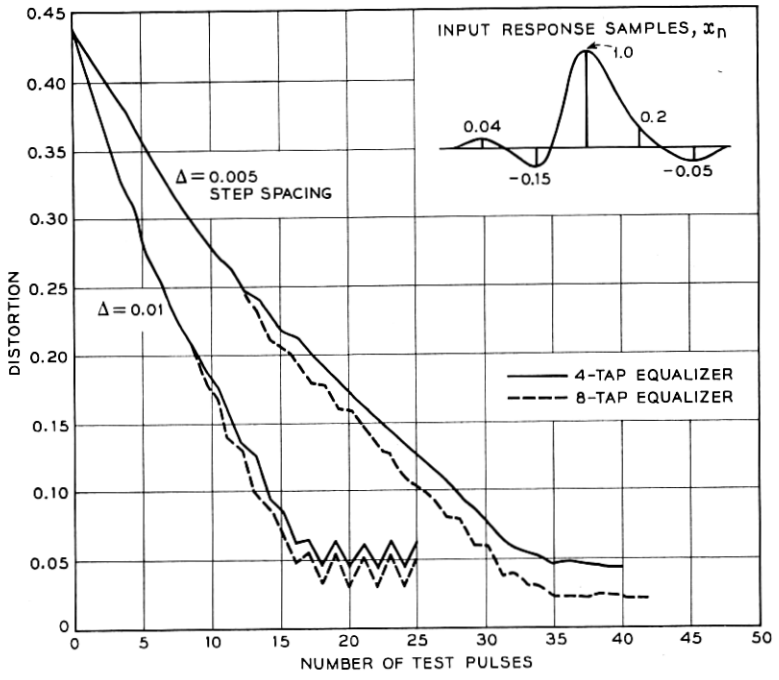

Fig. 10 — Distortion vs time.

function of time. It can be seen that increasing the number of taps N results in a lower residual distortion without affecting the settling time of the system. After the required settling time has elapsed, the system reaches a limit cycle where all taps oscillate one step up and down about the optimum values. Decreasing the step spacing Δ decreases the oscillation of distortion and increases the settling time required. When noise is introduced into the model, the curve of distortion versus time becomes a random walk with final values of distortion becoming random variables.

The residual distortion is composed basically of two components. One component is the theoretical minimum distortion for the given channel and length N of transversal filter available. This corresponds to the case where $S/N \rightarrow \infty$ and $\Delta \rightarrow 0$. No settings of the filter can reduce the residual distortion any further. This distortion component will be designated D_c. It is a function only of the channel characteristics $A(\omega)$ and $\beta'(\omega)$ and of the number of taps N.

The second component of residual distortion arises from our inability

to reach optimum tap settings because of the oscillation of the final tap settings due to noise and finite step spacing Δ. This component is approximately independent from tap to tap. On a per-tap basis it is a function only of step spacing Δ_j and signal-to-noise ratio. This component will be designated D_s, where D_s is the contribution from a single tap. Thus for an N-tap equalizer the distortion may be resolved as follows.

$$D_r = D_c(N,A,\beta') + \sum_{j=-N/2}^{N/2}{}' D_s(\Delta_j, S/N). \tag{35}$$

The subscripts "c" and "s" stand for channel distortion and system distortion respectively.

In the next section we will discuss generally D_c, the channel distortion, and show curves relating D_c and N for various shapes of amplitude and delay distortion. The subsequent section then deals with the system distortion D_s. Curves are shown relating D_s and Δ for various signal-to-noise ratios. The question of settling time is also discussed in this latter section.

4.2 *Residual Channel Distortion*

When $D_0 < 1$, the minimum distortion is obtained by setting $h_n = 0$ for $|n| \leqq N/2$, $n \neq 0$. The remaining distortion we describe as D_c, the residual channel distortion.

$$D_c = \frac{1}{h_0} \sum_{|n|>N/2} |h_n|. \tag{36}$$

This expression for the residual channel distortion is easily calculated on a digital computer from N, $A(\omega)$, and $\beta'(\omega)$ by first computing the response samples $\{x_n\}$ from the Fourier transform of the channel's frequency characteristic and then solving the N simultaneous equations $h_n = 0$ for $|n| \leqq N/2$, $n \neq 0$. A number of curves obtained by this procedure are presented in Figs. 11 through 13. In all of these figures the transmitter shaping characteristic was raised cosine with a 50 per cent roll-off. (Full raised cosine shaping gives very similar results.) The reference time was taken at the peak of the response $x(t)$. This is not necessarily optimum, but usually is fairly close to optimum.

Fig. 11 shows the residual channel distortion for parabolic delay over a wide range of equalizer size N. Fig. 12 is a similar presentation of the residual distortion resulting from linear delay. In Fig. 13 the channel has no delay distortion, but has an attenuation characteristic which

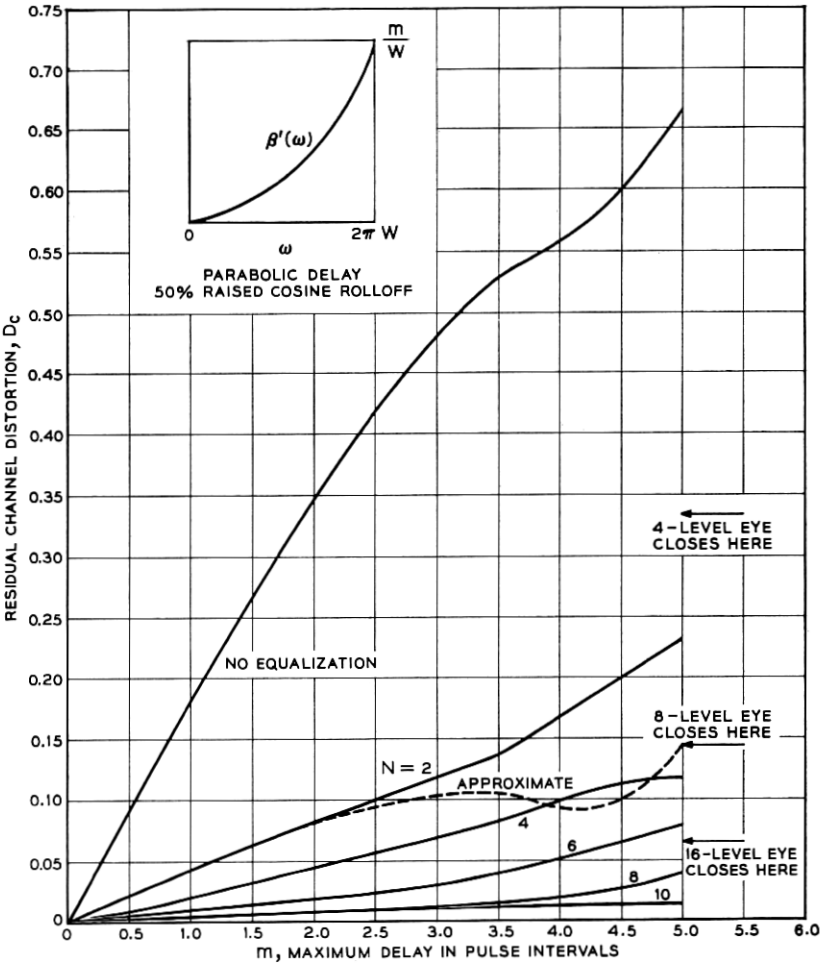

Fig. 11 — Residual channel distortion for an N-tap equalizer with parabolic delay.

falls off with various slopes. These curves are intended to answer the question of how well a given transversal filter can equalize a given type of distortion. These curves represent theoretical minima. To see how well an actual equalizer can approach these values, the system distortion D_s for the given Δ and S/N must be added. From (36) the residual distortion consists of all the response samples h_n removed from the center of the response by more than $nT/2$ seconds. This is approxi-

mately composed of two components: what distortion was already out there in the original channel response $x(t)$, and what distortion has been squeezed out into this range by the equalization process. We know from echo theory that the former component results from delay and amplitude ripples of greater than $N/2$ cycles in the bandwidth and increases approximately linearly with the amplitude of the ripple content. The latter component of the residual channel distortion which is squeezed outside the transversal filter range is more difficult to conceptually visualize.

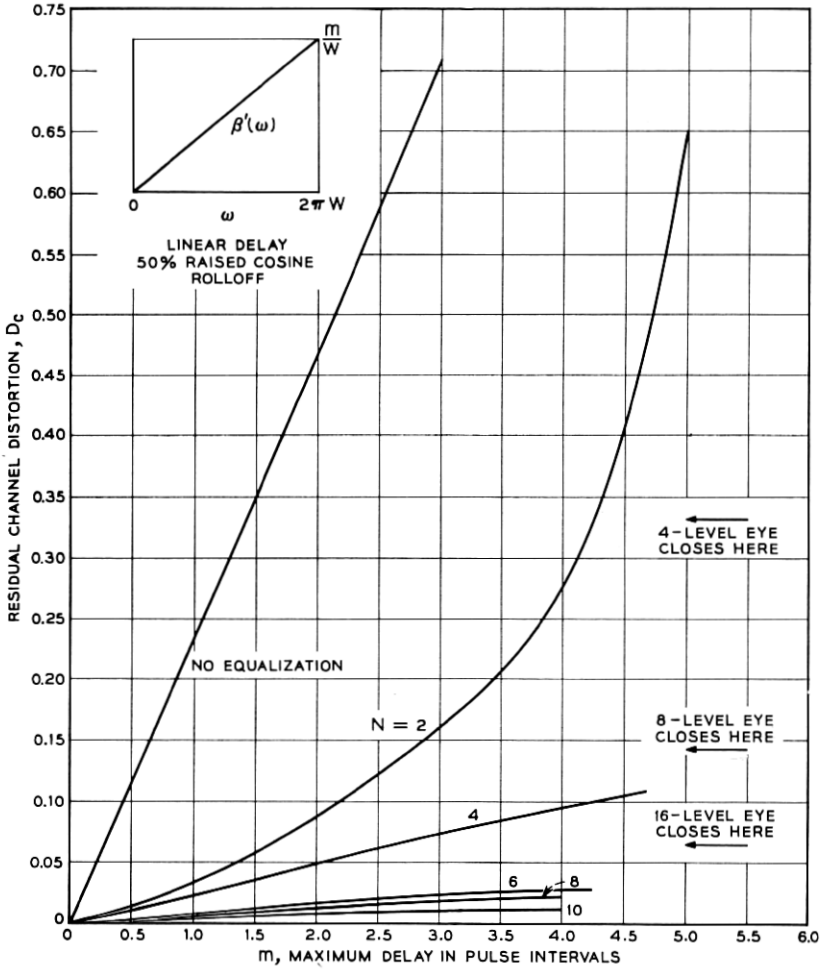

Fig. 12 — Residual channel distortion for an N-tap equalizer with linear delay.

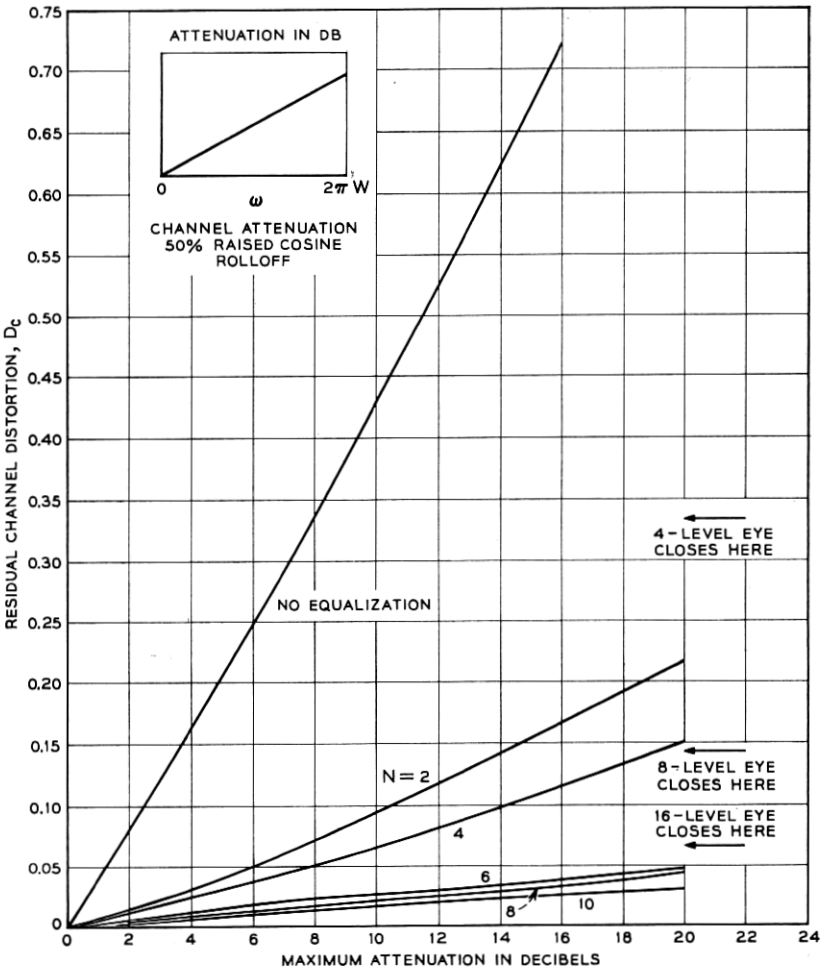

Fig. 13 — Residual channel distortion for an N-tap equalizer with attenuation distortion present in the transmission channel.

We can show that this component of D_c increases approximately with the *square* of the delay or amplitude variation in the channel.

Removing the large terms involving c_0 and x_0, which are unity, from under the summation in (9) we arrive at

$$h_n = x_n + c_n + \sum_{\substack{j=-N/2 \\ j\neq 0,n}}^{N/2} c_j x_{n-j} . \tag{37}$$

In order to make $h_n = 0$ for $-N/2 \leqq n \leqq N/2$, $n \neq 0$, it can be seen that a first approximation for the tap setting c_j is

$$c_j \approx -x_j . \tag{38}$$

After equalization we will have $h_n = 0$ inside the range of the filter. Outside the range we will have approximately, using (37) and (38)

$$\underset{|n|>N/2}{h_n} \approx x_n - \sum_{\substack{j=-N/2 \\ j\neq 0,n}}^{N/2} x_j x_{n-j} . \tag{39}$$

The first term in (39) is the initial distortion term and the second is distortion outside the range of the filter generated by the equalization process. Since the samples x_n vary approximately linearly with amplitude and delay ripple amplitude, it can be seen that the two components of h_n vary linearly and quadratically with ripple amplitude respectively.

The curves in Figs. 11 through 13 show that the linear term dominates in the case of parabolic or linear delay in the channel. This indicates that very little additional distortion in the range $|t| > NT/2$ is being pushed out by the equalization process. In Fig. 11 directly beneath the $N = 2$ curve is plotted the initial distortion outside the range of this particular equalizer. This dotted curve marked "approximate" indicates the quantity

$$D_r(\text{approx.}) = \sum_{|n|>1} |x_n| . \tag{40}$$

In the upper range of the $N = 2$ curve the second component involving "squeezed-out" distortion becomes important.

As long as the linear term dominates we can think of the transversal filter as completely equalizing any delay or amplitude variation of less than $N/2$ cycles in frequency content. For example, a Fourier expansion of the parabolic delay reveals the major content at low frequencies; thus few taps are required to equalize parabolic delay. The higher-frequency content of the parabola is unequalizable by a short transversal filter.

Frequently the delay and amplitude characteristic of a channel consist of approximately sinusoidal ripples. This is generally the case when an equalization network has been incorporated in the channel. A good rule of thumb is that it takes twice as many taps on the equalizer as there are ripples in the bandwidth. Thus the better a channel has been previously equalized the longer a transversal filter will be required to further improve equalization. Frequency- and time-domain equaliza-

tion are very similar in the respect that they both leave ripples of delay and amplitude of higher frequency after equalization.

4.3 *Residual System Distortion*

The term D_s, residual system distortion, results from the inability of the equalizer to reach the actual optimum tap settings. In the case of very high signal-to-noise ratio the value of D_s is obvious. The tap c_j will end in a limit cycle of one step Δ_j about the optimum tap setting. In one position

$$c_j = c_j(\text{optimum}) + \epsilon \tag{41}$$

while in the other position

$$c_j = c_j(\text{optimum}) + \epsilon - \Delta_j . \tag{42}$$

Assuming that $\Delta_j > |\epsilon|$ and neglecting the last term in (37), we obtain the average system distortion for this case.

$$D_s(\Delta_j , (S/N) = \infty) = \tfrac{1}{2}\Delta_j . \tag{43}$$

Thus the system distortion is one-half step for high S/N. However, a system will very seldom operate in a true high-S/N environment because the tap steps themselves are usually well within the noise. For example, with a Δ of 0.01 the step-to-noise ratio is 40 db below the signal-to-noise ratio. With noise the taps end in a random walk instead of the limit cycle of the noiseless case, and the average system distortion will be considerably higher than that given by (43).

Let's examine the behavior of an individual sample h_j when noise is present in the channel. The receiver then bases its decision on whether to advance or retard tap c_j by the amount Δ on the sign of the quantity $(h_j + n_j)$, where n_j is Gaussian with mean zero and variance σ^2, and is independent over index j and from test pulse to test pulse. Assuming that the sample h_j is affected only by tap setting c_j [equivalent to dropping the last term in (37)] we write

$$h_j \approx x_j + c_j . \tag{44}$$

Since c_j takes on only values which are integral multiples of Δ, we can without any particular loss of generality quantize h_j in steps of Δ. The behavior of each of the samples h_j is similar, so in order to keep confusion to a minimum we shall drop the "j" index and define

$$\text{Prob } [h = k\Delta, \text{ after } m \text{ test pulses}] = p(k,m). \tag{45}$$

Now we can write a difference equation for this probability as follows

$$p(k,m+1) = p(k-1,m)P(k-1) + p(k+1,m)Q(k+1) \quad (46)$$

where $P(k)$ and $Q(k)$ are the probabilities of noise being less than or greater than $-k\Delta$ respectively.

$$P(k) = \int_{-\infty}^{-k\Delta} \frac{1}{\sqrt{2\pi}\sigma} \exp -(x^2/2\sigma^2)\, dx = \tfrac{1}{2}[1 - \text{Erf}\,(k\Delta/\sqrt{2}\sigma)] \quad (47)$$

$$Q(k) = 1 - P(k). \quad (48)$$

Initially the value h_j starts at position $x_j \approx l\Delta$ so that we can use the initial condition

$$p(k,0) = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l. \end{cases} \quad (49)$$

The difference equation (46) then defines a probability distribution which spreads out as time (m) progresses and eventually ends in a stable symmetrical distribution centered at zero. [Actually there are two final distributions reached: one for m even and one for m odd. Since the system could be turned off in either state with equal likelihood, these distributions must be averaged. This effect is equivalent to the averaging of (41) and (42) in the noiseless case.]

A number of these final distributions were computed using (46), (47) and (48) on a digital computer. The contribution to the residual distortion owing to oscillation of tap c_j is the random variable $|h_j|$. The system distortion $D_s(\Delta_j, S/N)$ is defined as the expected value of $|h_j|$

$$D_s = E[|h_j|] = \Delta_j \sum_k |k|\, p(k,\infty). \quad (50)$$

A number of curves relating D_s, Δ and signal-to-noise ratio are shown in Fig. 14. Observe that these curves are nearly piecewise linear with a break point approximately where $\Delta = \sigma$. Above this point the high signal-to-noise condition prevails and $D_s \approx \frac{1}{2}\Delta$. Below this point the steps Δ are within the noise and the slope of the curve changes. By making a low signal-to-noise approximation we can derive equations for these curves.

For Δ/σ small we use the first term of the series

$$\text{Erf}\,(X) = \frac{2X}{\sqrt{\pi}}\left[1 - \frac{X^2}{1!3} + \frac{X^4}{2!5} - \frac{X^6}{3!7} + \cdots\right] \quad (51)$$

and obtain

$$P(k) = \tfrac{1}{2} - \frac{k\Delta}{\sqrt{2\pi}\,\sigma}. \quad (52)$$

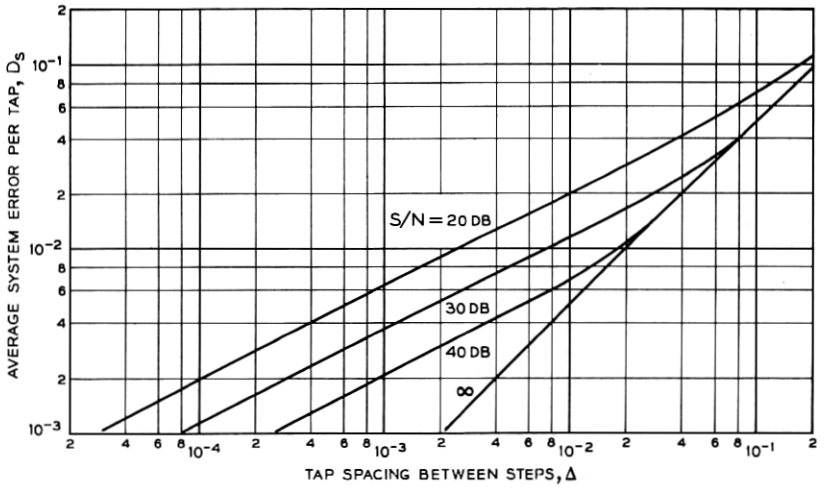

Fig. 14—D_s vs Δ for various S/N ratios.

Using this transition probability in the difference equation (46) we obtain an equation similar to the classical problem of the random walk of an elastically bound particle:

$$p(k,m+1) = p(k-1,m)\left[\frac{1}{2} - \frac{(k-1)\Delta}{\sqrt{2\pi}\,\sigma}\right] + p(k+1,m)\left[\frac{1}{2} + \frac{(k+1)\Delta}{\sqrt{2\pi}\,\sigma}\right]. \tag{53}$$

An exact solution to this difference equation assuming the initial condition (49) is given by M. Kac in Ref. 3. Since this solution is more complicated than the asymptotic distribution obtained as $\Delta \to 0$ and is no more useful for our purposes, we will not repeat Kac's formula here.

It is simple to derive expressions for the mean and variance of h_j after m test pulses have been transmitted. Using (53) we obtain for the mean value

$$E[h(m)] = \Delta \sum_k kp(k,m) \tag{54}$$

$$E[h(m)] = \Delta\left[\frac{1}{2} + \frac{\Delta}{\sqrt{2\pi}\,\sigma}\right]\sum_k \cdot k[p(k-1,m-1) + p(k+1,m-1)] + \frac{\Delta^2}{\sqrt{2\pi}\,\sigma}\sum_k k^2[p(k+1,m-1) - p(k-1,m-1)]. \tag{55}$$

The various summations in (55) can be rearranged with changes in variable to give

$$E[h(m)] = E[h(m-1)]\left(1 - \frac{2\Delta}{\sqrt{2\pi}\,\sigma}\right). \tag{56}$$

Since the initial value of h_j is $x_j \approx l\Delta_j$, the average value of h_j after m test pulses is

$$E[h_j(m)] = l\Delta_j\left(1 - \frac{2\Delta_j}{\sqrt{2\pi}\,\sigma}\right)^m. \tag{57}$$

One can derive a similar expression for the variance of h_j after m test pulses

$$E[h_j^2(m)] = (l\Delta_j)^2\left(1 - \frac{4\Delta_j}{\sqrt{2\pi}\,\sigma}\right)^m + \frac{\sqrt{2\pi}}{4}\Delta_j\sigma\left[1 - \left(1 - \frac{4\Delta_j}{\sqrt{2\pi}\,\sigma}\right)^m\right]. \tag{58}$$

As $m \to \infty$ the final distribution is obtained and

$$\lim_{m\to\infty} E[h_j(m)] = 0 \tag{59}$$

$$\lim_{m\to\infty} E[h_j^2(m)] = \frac{\sqrt{2\pi}}{4}\Delta_j\sigma. \tag{60}$$

Equation (60) is particularly interesting, since the variance of h_j is proportional to the square root of the noise variance σ^2.

Now suppose the tap-to-noise ratio Δ/σ tends to zero and test pulses are sent at a rate $1/\tau$ which tends to infinity. Using the difference equation (53) we can derive a differential equation for the probability density of the sample h_j after t seconds have elapsed:

$$\frac{\partial p(h,t)}{\partial t} = \frac{2\Delta}{\sqrt{2\pi}\,\sigma\tau}\left[p(h,t) + h\frac{\partial p(h,t)}{\partial h}\right] + \frac{\Delta^2}{2\tau}\frac{\partial^2 p(h,t)}{\partial h^2}. \tag{61}$$

A solution to this equation, attributed to Lord Rayleigh, is given in Ref. 3. Assuming h_j starts at x_j, the probability density after t seconds is

$$p(h_j, t) = \frac{1}{\sqrt{2\pi}\,\alpha(1 - e^{-2\gamma t})}\exp\left[-(h_j - x_j e^{-\gamma t})^2/2\alpha^2(1 - e^{-2\gamma t})\right], \tag{62}$$

where α^2 is the same as $E[h_j^2(\infty)]$ in (60)

$$\alpha^2 = \frac{\sqrt{2\pi}}{4}\Delta_j\sigma \tag{63}$$

and

$$\gamma = \frac{2\Delta_j}{\sqrt{2\pi}\,\sigma_T}. \tag{64}$$

As $t \to \infty$ the density (62) tends to Gaussian of mean zero and variance α^2. We can thus easily derive the low signal-to-noise ratio value of D_s.

$$D_s = E[|\,h_j\,|] = 2\int_0^\infty \frac{h_j}{\sqrt{2\pi}\,\alpha} \exp -(h_j^2/2\alpha^2)\, dh_j \tag{65}$$

$$D_s = \sqrt{2/\pi}\,\alpha = (\sigma\Delta_j/\sqrt{2\pi})^{\frac{1}{2}} \tag{66}$$

$$D_s \approx 0.633\sqrt{\sigma\Delta_j}. \tag{67}$$

The straight-line portions of the curves D_s versus Δ in Fig. 14 match nearly exactly with (67).

In summary, we have found

$$D_s \approx \begin{cases} \Delta_j/2 & \text{if} \quad \Delta_j > \sigma \\ 0.633\sqrt{\sigma\Delta_j} & \text{if} \quad \Delta_j < \sigma. \end{cases} \tag{68}$$

The important fact about (68) is that the residual system error goes down as the square root of both tap spacing Δ and the standard deviation of the noise σ. Thus once the step spacing and noise are comparable, cutting the steps Δ by a factor of ten cuts the residual system error due to tap uncertainty by about 3.16. An interesting sidelight is that cutting Δ by ten might seem equivalent to averaging ten samples of h_j before taking one step of Δ. In the latter case the noise σ would be cut by $\sqrt{10} = 3.16$ and the residual system distortion D_s cut by only $\sqrt{3.16} = 1.77$. However, as might be guessed this latter technique will require less settling time than the former technique using 0.1Δ and moving ten times as fast.

This leads us into an abbreviated discussion of settling time. The time required to settle is usually determined by the largest distorting sample x_j for $j \neq 0$. If the step-to-noise ratio Δ/σ is large, then the settling time is obviously

$$T_s = x_{j\max}/\Delta_j. \tag{69}$$

When the noise becomes important, then we need to make some sort of arbitrary definition of settling time. This definition could be based on the time required for the mean of h_j or the variance of h_j to reach some predetermined position or percentage, or it could be based on the time

constant $1/\gamma$ in the density given by (62). All these types of definition lead to similar expressions differing chiefly by constants, so we'll use the simplest and define

$$T_s = 1/\gamma = \sqrt{2\pi}\,\sigma\tau/2\Delta_j\,. \tag{70}$$

Recall that τ is the interval between test pulses. This expression is independent of $x_{j\max}$ for low tap-to-noise ratios. Using a binomial expansion of the term to the mth power in (57), (70) can be derived as the point where the first two terms of the expansion for $E[h_j(m)]$ cancel.

Going back to our previous example, where we compared decreasing the step spacing Δ with averaging several samples of h_j, we can now show that cutting the tap spacing Δ by a factor of N is exactly equivalent in settling time and system error to averaging N^2 consecutive samples of h_j. This is evident from the proportionalities involved in (69) and (70). In either case the same amount of time is required to achieve a given level of accuracy in equalization.

V. FREQUENCY-DOMAIN CONSIDERATIONS

5.1 *Frequency-Domain Relationship*

The equalizer that has been described is intended for the correction of distortion in digital data transmission. This equalizer is strictly a time-domain device which corrects the pulse response of the channel to the best ability of a finite-length tapped delay line. The question which most frequently arises concerning its operation asks what happens to the frequency characteristics of the channel as a result of the time-domain equalization. This question is asked not only out of curiosity and because engineers tend to think in terms of the frequency domain, but also because the time-domain equalizer is sometimes considered for the equalization of analog channels.

In this section we will develop a general formula for the frequency-domain characteristics of an equalized channel in terms of the unequalized characteristics. This formula allows us to compute final characteristics and to make a few general observations about the relationship between initial and final characteristics. In addition to these results we are able to derive conditions as to when a channel may be equalized successfully, and we find one additional technique of computing optimum tap settings for long equalizers.

Henceforth we will assume an infinite-length transversal filter. If the equalizer is long enough to do its job properly, the final characteristics will closely approximate the infinite length characteristics we will

derive here. An outline of this derivation is as follows. If the equalized pulse response is to have zero distortion, it must have Nyquist frequency characteristics implying symmetry in real and imaginary components. Knowing the channel characteristics and the type of characteristic capable of being assumed by the equalizer, we show there is only one such Nyquist characteristic the product can assume. This characteristic must be the final frequency response of the equalized channel.

The impulse response of the channel in terms of the amplitude, $A(\omega)$, and phase, $\beta(\omega)$, is

$$h(t) = \frac{1}{\pi}\int_0^\infty A(\omega)\cos[\omega t - \beta(\omega)]d\omega. \tag{71}$$

We assume that the time base has been adjusted by the removal of a flat delay (linear phase) component from $\beta(\omega)$ so that the peak of the output response occurs at time zero. For zero distortion we require

$$h_n = 0 = \frac{1}{\pi}\int_0^\infty A(\omega)\cos[n\omega T - \beta(\omega)]d\omega \qquad \text{all } n,\ n \neq 0. \tag{72}$$

Changing (72) to use real and imaginary components we arrive at the equivalent condition

$$\int_0^\infty A_x(\omega)\cos n\omega T d\omega = 0 \qquad \text{all } n,\ n \neq 0 \tag{73}$$

$$\int_0^\infty A_y(\omega)\sin n\omega T d\omega = 0 \qquad \text{all } n \tag{74}$$

where

$$A_x(\omega) = A(\omega)\cos\beta(\omega) \tag{75}$$

$$A_y(\omega) = A(\omega)\sin\beta(\omega). \tag{76}$$

Let us assume that $A(\omega) = 0$ for $\omega > 2\pi/T$, i.e., the channel has no frequency component higher than twice the Nyquist band for its signaling rate. It would be an unusual case if this were not true. We now make a change in variables to shift the origin to the Nyquist frequency π/T and redefine the characteristics about this frequency.

$$\hat{A}_x(\omega) = A_x\left(\omega + \frac{\pi}{T}\right) \tag{77}$$

$$\hat{A}_y(\omega) = A_y\left(\omega + \frac{\pi}{T}\right). \tag{78}$$

(In general, a circumflex on any variable indicates that it is defined about the frequency π/T as origin.) With this change the conditions (73) and (74) become

$$\int_{-\pi/T}^{\pi/T} \hat{A}_x(\omega) \cos n\omega T d\omega = 0 \qquad \text{all } n,\ n \neq 0 \tag{79}$$

$$\int_{-\pi/T}^{\pi/T} \hat{A}_y(\omega) \sin n\omega T d\omega = 0 \qquad \text{all } n. \tag{80}$$

We also have the normalizing condition for the center sample.

$$h_0 = 1 = \frac{1}{\pi} \int_{-\pi/T}^{\pi/T} \hat{A}_x(\omega) d\omega. \tag{81}$$

Equation (79) says that $\hat{A}_x(\omega)$ must be orthogonal to all $\cos n\omega T$ between $-\pi/T$ and π/T. This implies that its Fourier expansion consists of a constant to satisfy (81) plus any arbitrary sine components. In other words it must be a constant plus any arbitrary odd function. Similarly, $\hat{A}_y(\omega)$ must be any arbitrary even function to satisfy condition (80). Simply stated, the conditions are then

$$\hat{A}_x(\omega) = (T/2) + \text{odd function} \tag{82}$$

$$\hat{A}_y(\omega) = \text{even function}. \tag{83}$$

These conditions were first derived by Nyquist,[4] although engineers are generally more familiar with the case of flat delay where (82) and (83) reduce to the statement that $A(\omega)$ has Nyquist symmetry.

Now that we see the conditions that the product of the equalizer and channel responses must meet for perfect equalization, let's look at the type of response the equalizer alone can have. The Fourier transform of the equalizer's time response is

$$C(\omega) = \sum_{n=-\infty}^{\infty} c_n e^{-jn\omega T} \tag{84}$$

where the c_n's are the tap gain settings. As before we shift the origin of definition for $C(\omega)$ to π/T. The change in variables and definition gives

$$\hat{C}(\omega) = \hat{C}_x(\omega) + j\hat{C}_y(\omega) \tag{85}$$

with

$$\hat{C}_x(\omega) = c_0 + \sum_{n=1}^{\infty} (-1)^n (c_n + c_{-n}) \cos n\omega T \tag{86}$$

$$\hat{C}_y(\omega) = \sum_{n=1}^{\infty} (-1)^n (c_n - c_{-n}) \sin n\omega T. \tag{87}$$

Observe that $\hat{C}_x(\omega)$ is an even function and $\hat{C}_y(\omega)$ is an odd function.

Now let us suppose that the unequalized channel frequency response in terms of real and imaginary components defined about $\omega = \pi/T$ is

$$\hat{A}(\omega) = \hat{A}_x(\omega) + j\hat{A}_y(\omega). \tag{88}$$

The product of the unequalized frequency response $\hat{A}(\omega)$ from (88) and the equalizer response $\hat{C}(\omega)$ from (85) must satisfy the odd and even requirements on its real and imaginary components as given in (82) and (83). After separating the product $\hat{A}(\omega)\hat{C}(\omega)$ into real, imaginary, odd, and even components, it is possible to arrive at simultaneous equations for $\hat{C}_x(\omega)$ and $\hat{C}_y(\omega)$. The details of this process are quite straightforward and have been relegated to Appendix B.

The final equalizer response is found in terms of the even (labelled with an "e" subscript) and odd (labelled with an "o" subscript) components of the unequalized channel response. The equations derived are

$$\hat{C}_x(\omega) = \frac{\frac{T}{2}\hat{A}_{xe}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} \tag{89}$$

$$\hat{C}_y(\omega) = \frac{-\frac{T}{2}\hat{A}_{yo}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)}, \tag{90}$$

where

$$\hat{A}_{xe}(\omega) = \tfrac{1}{2}[\hat{A}_x(\omega) + \hat{A}_x(-\omega)] \tag{91}$$

$$\hat{A}_{xo}(\omega) = \tfrac{1}{2}[\hat{A}_x(\omega) - \hat{A}_x(-\omega)]. \tag{92}$$

These equations determine the real and imaginary components of the equalizer frequency response.

Using this response for the equalizer, the equalized channel response becomes

$$\begin{aligned}\frac{2}{T}\hat{C}(\omega)\hat{A}(\omega) = &\left[1 + \frac{\hat{A}_{xe}(\omega)\hat{A}_{xo}(\omega) + \hat{A}_{yo}(\omega)\hat{A}_{ye}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)}\right] \\ &+ j\left[\frac{\hat{A}_{xe}(\omega)\hat{A}_{ye}(\omega) - \hat{A}_{yo}(\omega)\hat{A}_{xo}(\omega)}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)}\right]\end{aligned} \tag{93}$$

which is the result we have been seeking.

5.2 *Interpretation of the Frequency-Domain Relationship*

The frequency response of the equalizer [from (89) and (90)] is

$$\hat{C}(\omega) = \frac{T/2}{\hat{A}_{xe}^2(\omega) + \hat{A}_{yo}^2(\omega)} [\hat{A}_{xe}(\omega) - j\hat{A}_{yo}(\omega)]. \tag{94}$$

Recall that $\hat{A}_{xe}(\omega)$ is the *even* component of the *real* part of the channel response taken about the Nyquist frequency, π/T. Similarly $\hat{A}_{yo}(\omega)$ is the *odd* component of the *imaginary* part of the response. Notice that $\hat{C}(\omega)$ does not depend on $\hat{A}_{xo}(\omega)$ or $\hat{A}_{ye}(\omega)$, so that these components may be specified arbitrarily without changing the equalizer settings. This points up the difference between frequency-domain equalization and time-domain equalization. Our data distortion criterion is such that we don't care about the components $\hat{A}_{xo}(\omega)$ and $\hat{A}_{ye}(\omega)$ of the frequency response. However, these components affect the shape of the final equalized characteristic, as can be seen from (93).

It is clear from (94) that the channel cannot be equalized if $\hat{A}_{xe}(\omega)$ and $\hat{A}_{yo}(\omega)$ are both zero for some ω, $0 \leqq \omega \leqq \pi/T$, in which case $\hat{C}(\omega)$ is unbounded. If this does not happen, practical considerations indicate that the channel is capable of being equalized. (These practical considerations require that both $[A(\omega)C(\omega)]$ and its Fourier transform are continuous and absolutely integrable.)

Equalization Condition:

A channel with complex gain $\hat{A}(\omega)$ is capable of being perfectly equalized with an infinite transversal filter if and only if the even component of the real part of $\hat{A}(\omega)$ and the odd component of the imaginary part of $\hat{A}(\omega)$ do not simultaneously vanish for some ω, $0 \leqq \omega \leqq \pi/T$.

Note that a channel can be equalized even if it transmits no energy in some interval $[\omega_1, \omega_2]$, $\omega_2 < \pi/T$, so long as the missing energy components are replaced in a symmetrically located region above π/T, i.e., $[(2\pi/T) - \omega_2, (2\pi/T) - \omega_1]$.

Now let's go back to (93) to try to get some feel for the shape of the final response. The equation itself is not complicated, but it unfortunately requires us to break $A(\omega)$ into even and odd components about the Nyquist frequency π/T and then further to break these components into real and imaginary parts. By this time we have almost no idea what the final response will be when we put everything back together as in (93). To take a special case which simplifies this process, suppose that the channel has perfectly flat delay, so all the "y" components are zero. Then

$$\hat{A}(\omega)\hat{C}(\omega) = \frac{T}{2}\left[1 + \frac{\hat{A}_{xo}(\omega)}{\hat{A}_{xe}(\omega)}\right]. \tag{95}$$

The first term represents a rectangular characteristic and the second term is an odd function about the frequency π/T. Thus we have a familiar Nyquist characteristic which is easily determined from the initial characteristic $A(\omega)$. Fig. 15 shows a sample case of this type having flat delay (linear phase). The components $\hat{A}_{xe}(\omega)$ and $\hat{A}_{xo}(\omega)$ and the equalized spectrum are shown for this example characteristic. With a little study it is possible to get a good feel for the sort of final characteristic which is obtained by equalization in this purely real case. However, the introduction of nonlinear phase results back in the complicated expression (93).

Typically, the channel to be equalized cuts off somewhere before $\omega = 2\pi/T$, so there are no energy components in the interval $[(\pi/T) + \omega_c, (2\pi/T)]$. In this case we have

$$\begin{aligned} \hat{A}_{xe}(\omega) &= -\hat{A}_{xo}(\omega) \\ \hat{A}_{ye}(\omega) &= -\hat{A}_{yo}(\omega) \end{aligned} \qquad \pi/T \geqq \omega \geqq \omega_c \tag{96}$$

$$\begin{aligned} \hat{A}_{xe}(\omega) &= \hat{A}_{xo}(\omega) \\ \hat{A}_{ye}(\omega) &= \hat{A}_{yo}(\omega) \end{aligned} \qquad -\pi/T \leqq \omega \leqq -\omega_c. \tag{97}$$

Substitution of (96) and (97) into (93) shows that

$$\hat{A}(\omega)\hat{C}(\omega) = T/2 \quad \text{for} \quad -\pi/T \leqq \omega \leqq -\omega_c. \tag{98}$$

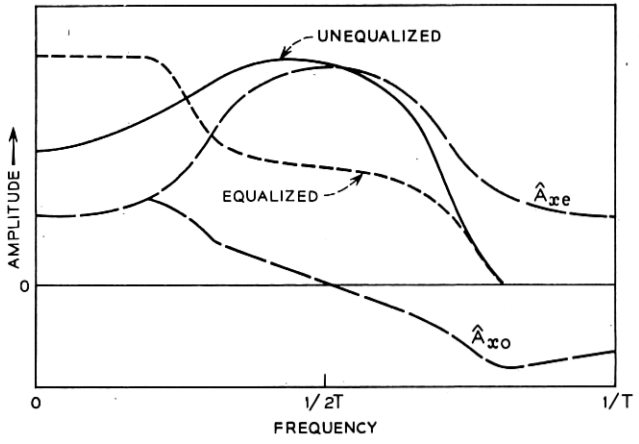

Fig. 15 — Equalization of amplitude distortion.

Which means

$$A(\omega)C(\omega) = \frac{T}{2} \qquad \text{for} \quad 0 \leqq \omega \leqq \frac{\pi}{T} - \omega_c . \tag{99}$$

We can boil this equation down into the most significant observation we can make about the frequency-domain relationship:

If the channel response cuts off at the frequency $(\pi/T) + \omega_c$, then the channel must be equalized to constant amplitude and linear phase in the region $0 \leqq \omega \leqq (\pi/T) - \omega_c$.

This observation is important in the equalization of analog channels to flat amplitude and linear phase. In order to get the largest possible interval of perfect analog equalization we should arrange the tap spacing T so as to approach the Nyquist interval corresponding to the channel cutoff. However, if the channel cuts off *before* π/T then by the previous results the channel is unequalizable. The closer T is to a Nyquist interval the more taps will generally be required to effect a good equalization. One final note to add to the confusion — if ω_c is large very little of the spectrum will be equalized flat. If the tap spacing T cannot be changed we can increase this region of flat equalization by inserting a low-pass filter during the equalization period which has a cutoff frequency close above π/T. The filter may then be removed after the equalizer tap gains have been set.

Some of the principles involved are exemplified in Fig. 16, which shows the equalization of a channel which has a linear delay characteristic (quadratic phase). Initially the amplitude response of this channel is a 50 per cent raised cosine response cutting off at $\omega = 3\pi/2T$. This amplitude characteristic is a Nyquist shape in the absence of delay distortion. After equalization both amplitude and phase are flat from $\omega = 0$ to $\omega = \pi/2T$, as required by our previous observation. From $\omega = \pi/2T$ to $\omega = 3\pi/2T$ the phase still appears parabolic. In this region there has been an interaction between phase and amplitude which has resulted in a change in the amplitude characteristic away from its Nyquist shape. The combination of equalized phase and amplitude in this region is such as to satisfy conditions (82) and (83), although this is not evident from casual inspection.

One additional important usage of (93) for $\hat{C}(\omega)$ is in the calculation of optimum tap settings for the equalizer. Starting from the unequalized channel pulse time response samples $\{x_n\}$ we construct $\hat{A}_{xe}(\omega)$ and $\hat{A}_{yo}(\omega)$ exactly as in (86) and (87) for $\hat{C}_x(\omega)$ and $\hat{C}_y(\omega)$, i.e., let

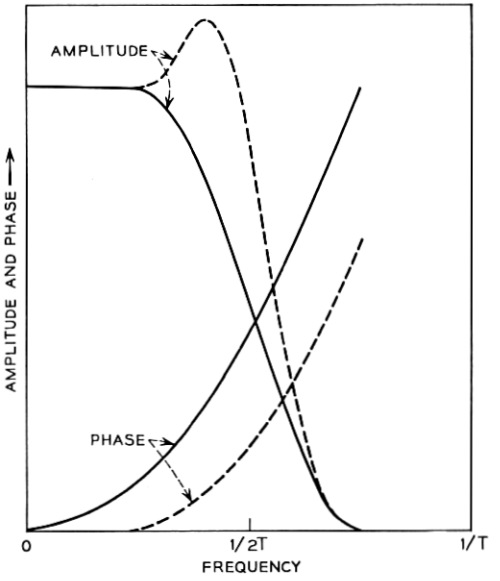

Fig. 16 — Equalization of linear delay distortion.

$$\hat{A}_{xe}(\omega) = x_0 + \sum_{n=1}^{\infty} (-1)^n (x_n + x_{-n}) \cos n\omega T \tag{100}$$

$$\hat{A}_{yo}(\omega) = \sum_{n=1}^{\infty} (-1)^n (x_n - x_{-n}) \sin n\omega T. \tag{101}$$

The other components, $\hat{A}_{xo}$ and $\hat{A}_{ye}$, are of no concern to us, since they affect neither $\{x_n\}$ nor the equalizer settings. Now calculate $\hat{C}(\omega)$ from (93) and evaluate its Fourier transform at time nT. Each of these values may be identified with tap settings through (96) and (97). This procedure may seem complex, but it is by far the fastest way to calculate tap settings for long transversal filters on a digital computer. Time-domain techniques involve the solution of simultaneous equations or iterative methods depending on the input pulse. The time required for this type of minimization usually increases with the square of the number of taps involved. As each new tap is added, all the previous tap settings change to readjust for the new minimum for the increased filter length.

On the other hand, using the frequency-domain technique the time required for computing tap settings increases linearly with the number of taps involved and each new tap setting does not change any of the

previous settings. The catch is that the settings calculated by the frequency-domain method apply only to an infinite-length line. For lines of finite length the settings calculated by frequency- and time-domain methods will differ. The time-domain settings are calculated on the basis of minimum distortion — a quantity defined on the basis of time samples. The frequency-domain settings for a finite line are a root-mean-square approximation to the final equalizer frequency characteristic for an infinite line. As the number of taps goes to infinity the tap settings computed by each method approach the same value. For any finite line the time-domain settings are better (lead to less distortion), since they are by definition optimum. However, for equalizers of about 12 taps or more the results are practically the same. (Indeed, one should remember that the distortion criterion itself is somewhat arbitrary.) Thus the frequency-domain technique described here is a fast and accurate way of computing tap settings for long equalizers. Unfortunately the method does not lend itself to implementation, since it involves a number of difficult operations, i.e., division, Fourier inversion.

VI. ACKNOWLEDGMENTS

The author is particularly indebted to F. K. Becker and E. Port, who are largely responsible for the equalization system shown in Fig. 4. The counter-attenuators illustrated in Fig. 5 were designed by L. N. Holzman.

APPENDIX A

Statement and Proof of Theorem I

Theorem I

If $D_0 < 1$, *then the minimum distortion* D *must occur for those* N *tap gains which simultaneously cause* $h_n = 0$ *for all* $n \subset K_N$, $n \neq 0$.

Proof

We prove this theorem by assuming a minimum has occurred at some point other than that specified in the theorem and showing this assumption leads to a contradiction. Thus we assume that D is at a minimum for some sequence of tap gains $\{c_j\}$ and that $h_k \neq 0$ for some $k \subset K_N$, $k \neq 0$. Now we will show that there exists another sequence $\{c_j^*\}$ for which $D^* < D$ and hence the contradiction.

Let

$$\begin{aligned} c_j^* &= c_j \qquad j \subset K_N, j \neq 0, j \neq k \\ c_k^* &= c_k - \Delta \operatorname{sgn} h_k \end{aligned} \tag{102}$$

where

$$\tfrac{1}{2} | h_k | > \Delta > 0 \tag{103}$$

which is possible, since by hypothesis $h_k \neq 0$. The value of distortion corresponding to the tap gains $\{c_j^*\}$ is, from (14)

$$\begin{aligned} D^* = \sum_{n=-\infty}^{\infty}{}' \Big| \sum_{\substack{j \subset K_N \\ j \neq k}}{}' c_j (x_{n-j} - x_n x_{-j}) + x_n \\ + (c_k - \Delta \operatorname{sgn} h_k)(x_{n-k} - x_n x_{-k}) + x_k \Big|. \end{aligned} \tag{104}$$

Using (13) this reduces to

$$D^* = \sum_{n=-\infty}^{\infty}{}' | h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) | \tag{105}$$

$$\begin{aligned} D^* = \sum_{\substack{n=-\infty \\ n \neq k}}^{\infty}{}' | h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) | \\ + | h_k - \Delta \operatorname{sgn} h_k (1 - x_k x_{-k}) |. \end{aligned} \tag{106}$$

We concentrate for the moment on the second term of (106) and use $h_k = | h_k | \operatorname{sgn} h_k$ to obtain

$$| h_k - \Delta \operatorname{sgn} h_k (1 - x_k x_{-k}) | = | | h_k | - \Delta (1 - x_k x_{-k}) |. \tag{107}$$

But $| x_k x_{-k} | < 1$ since

$$D_0 = \sum_{n=-\infty}^{\infty}{}' | x_n | < 1. \tag{108}$$

Therefore $2 > (1 - x_k x_{-k}) > 0$ and, by (103),

$$| h_k | - \Delta (1 - x_k x_{-k}) > 0. \tag{109}$$

Thus we are able to drop the absolute value brackets around the second term in (106) and write

$$\begin{aligned} D^* = \sum_{\substack{n=-\infty \\ n \neq k}}^{\infty}{}' | h_n - \Delta \operatorname{sgn} h_k (x_{n-k} - x_n x_{-k}) | \\ + | h_k | - \Delta (1 - x_k x_{-k}) \end{aligned} \tag{110}$$

$$D^* \leqq \sum_{n=-\infty}^{\infty}{}' |h_n| + \Delta \sum_{\substack{n=-\infty \\ n\neq k}}^{\infty}{}' |x_{n-k} - x_n x_{-k}| - \Delta(1 - x_k x_{-k}) \tag{111}$$

$$D^* \leqq D + \Delta\Big\{ \sum_{\substack{n=-\infty \\ n\neq k}}^{\infty}{}' |x_{n-k}| + |x_{-k}| \sum_{\substack{n=-\infty \\ n\neq k}}^{\infty}{}' |x_n| - (1 - x_k x_{-k})\Big\} \tag{112}$$

$$D^* \leqq D + \Delta\big\{D_0 - |x_{-k}| + |x_{-k}|\,(D_0 - |x_k|) - 1 + |x_k|\,|x_{-k}|\big\} \tag{113}$$

$$D^* \leqq D - \Delta(1 - D_0)(1 + |x_{-k}|) \tag{114}$$

and since $D_0 < 1$

$$D^* < D \tag{115}$$

which is a contradiction, since we assumed that D was originally at a minimum. This completes the proof of Theorem I.

APPENDIX B

Derivation of Equalizer Frequency Response in Terms of Unequalized Channel Frequency Response

The equalizer frequency response taken about the Nyquist frequency π/T is

$$\hat{C}(\omega) = \hat{C}_{xe}(\omega) + j\hat{C}_{yo}(\omega) \tag{116}$$

where we use the subscripts "x" and "y" to denote real and imaginary components and the subscripts "e" and "o" to denote even and odd components. Breaking the unequalized channel frequency response similarly into its components gives

$$\hat{A}(\omega) = \hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega) + j\hat{A}_{ye}(\omega) + j\hat{A}_{yo}(\omega) \tag{117}$$

with

$$\hat{A}_{xe}(\omega) = \tfrac{1}{2}[\hat{A}_x(\omega) + \hat{A}_x(-\omega)] \tag{118}$$

$$\hat{A}_{xo}(\omega) = \tfrac{1}{2}[\hat{A}_x(\omega) - \hat{A}_x(-\omega)] \tag{119}$$

etc.

We are now ready to put together the equalizer response and the

channel response to get the equalized channel response. Using (117) and (116) for this purpose we obtain

$$\begin{aligned}\hat{A}(\omega)\hat{C}(\omega) = {} & [\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{xe}(\omega) - [\hat{A}_{ye}(\omega) \\ & + \hat{A}_{yo}(\omega)]\hat{C}_{yo}(\omega) + j\{[\hat{A}_{xe}(\omega) \\ & + \hat{A}_{xo}(\omega)]\hat{C}_{yo}(\omega) + [\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)]\hat{C}_{xe}(\omega)\}.\end{aligned} \tag{120}$$

This equation must meet the conditions (82) and (83). If δ_o is an arbitrary odd function, condition (82) dictates that

$$\text{Re}\,[\hat{A}(\omega)\hat{C}(\omega)] = (T/2) + \delta_o \tag{121}$$

$$\begin{aligned}[\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{xe}(\omega) - [\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)]\hat{C}_{yo}(\omega) \\ = (T/2) + \delta_o\end{aligned} \tag{122}$$

$$\hat{A}_{xe}(\omega)\hat{C}_{xe}(\omega) - \hat{A}_{yo}(\omega)\hat{C}_{yo}(\omega) = T/2 \tag{123}$$

$$\hat{A}_{xo}(\omega)\hat{C}_{xe}(\omega) - \hat{A}_{ye}(\omega)\hat{C}_{yo}(\omega) = \delta_o\,. \tag{124}$$

Since δ_o is an arbitrary function, (124) is automatically satisfied.

Let δ_e be an arbitrary even function. Condition (83) now requires

$$\text{Im}\,[\hat{A}(\omega)\hat{C}(\omega)] = \delta_e \tag{125}$$

$$[\hat{A}_{xe}(\omega) + \hat{A}_{xo}(\omega)]\hat{C}_{yo}(\omega) + [\hat{A}_{ye}(\omega) + \hat{A}_{yo}(\omega)]\hat{C}_{xe}(\omega) = \delta_e\,. \tag{126}$$

The even and odd equalities from (126) are respectively

$$\hat{A}_{xo}(\omega)\hat{C}_{yo}(\omega) + \hat{A}_{ye}(\omega)\hat{C}_{xe}(\omega) = \delta_e \tag{127}$$

$$\hat{A}_{xe}(\omega)\hat{C}_{yo}(\omega) + \hat{A}_{yo}\hat{C}_{xe}(\omega) = 0. \tag{128}$$

Again (127) is satisfied trivially and we are left with two equations, (128) and (123), in the two unknowns $\hat{C}_{xe}(\omega)$ and $\hat{C}_{yo}(\omega)$. These equations may be solved simultaneously and we arrive at

$$\hat{C}_x(\omega) = \frac{\frac{T}{2}\hat{A}_{xe}(\omega)}{\hat{A}_{xe}^{\,2}(\omega) + \hat{A}_{yo}^{\,2}(\omega)} \tag{129}$$

$$\hat{C}_y(\omega) = \frac{-\frac{T}{2}\hat{A}_{yo}(\omega)}{\hat{A}_{xe}^{\,2}(\omega) + \hat{A}_{yo}^{\,2}(\omega)}\,. \tag{130}$$

REFERENCES

1. Blumlein, A. D., British Patent No. 517516.
2. Kallman, Heinze E., Transversal Filters, Proc. IRE, *28*, July, 1940.
3. Kac, M., Random Walk and the Theory of Brownian Motion, in *Selected Papers on Noise and Stochastic Processes*, Nelson Wax, ed., Dover Press, New York, N. Y.
4. Nyquist, H., Certain Topics in Telegraph Transmission Theory, Trans. AIEE, *47*, 1928.