THE BELL SYSTEM
TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 52 July–August 1973 Number 6

Practical Design Rules for Optimum Finite Impulse Response Low-Pass Digital Filters

By O. HERRMANN,* L. R. RABINER, and D. S. K. CHAN

(Manuscript received November 22, 1972)

Although a great deal is known about design techniques for optimum (in a minimax error sense) finite impulse response (FIR) low-pass digital filters, there have not been established any practical design rules for such filters. Thus, a user is unable to easily decide on the (approximate or exact) filter order required to meet his design specifications and must resort to tables or trial and error procedures. In this paper, such a set of design rules is given. In the case of very narrow bandwidth or very wide bandwidth filters, analytic relations between the filter parameters can be readily obtained. In all other cases, exceedingly good linear and nonlinear fits to the data can be obtained over somewhat restricted ranges of the parameters. These fitting procedures lead to a practical set of simple design rules for estimating filter order from the desired specifications.

I. INTRODUCTION

The problem of designing an optimal (in the minimax sense) low-pass FIR digital filter to meet design specifications has been thoroughly investigated[1–5] and may be considered to be solved. Thus, given a

* University of Erlangen—Nuremberg, Germany.

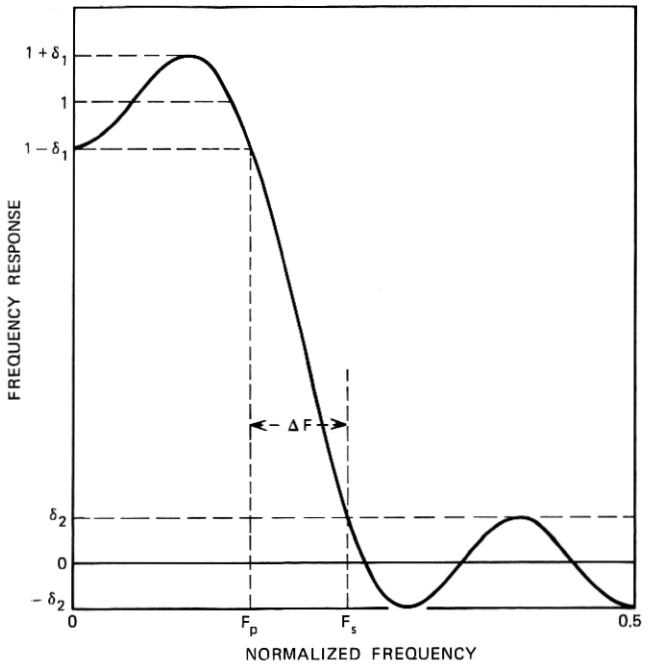

Fig. 1—Definition of low-pass filter parameters.

specified impulse response duration of N samples, a passband cutoff frequency F_p (see Fig. 1), a stopband cutoff frequency F_s, and a ratio of tolerances $K = \delta_1/\delta_2$ (where δ_1 = passband tolerance and δ_2 = stopband tolerance), an optimal approximation to these specifications can be designed. The approximation is optimal in the sense that, for given values of N, F_p, F_s, and K, δ_1 (or equivalently δ_2) is minimum. The nature of the solution is such that there are three distinct classes into which it may belong, depending on the specific design parameters. These classes have been called extraripple filters,[4,6] scaled extraripple filters,[7] and equiripple filters with one less than the maximum possible number of ripples. The differences between these classes lie in the number and amplitude of the ripples in the weighted error curve. The weighted error curve is defined as:

$$E(e^{j2\pi f}) = \begin{cases} \dfrac{1 - H(e^{j2\pi f})}{K} & 0 \leqq f \leqq F_p \\ -H(e^{j2\pi f}) & F_s \leqq f \leqq 0.5, \end{cases} \tag{1}$$

where $H(e^{j2\pi f})$ is the frequency response of the optimal filter. Extra-

ripple filters have $(N + 5)/2$ equal magnitude extrema in their error curves. Scaled extraripple filters also have $(N + 5)/2$ extrema, all but one of which are of equal magnitude. The third class of solution has $(N + 3)/2$ equal amplitude extrema in its error curves. Figure 2 shows plots of curves of transition bandwidth, $\Delta F = F_s - F_p$, versus passband cutoff frequency, F_p, for two sets of conditions.[6] The data in Fig. 2a show the curves for $K = 1$, $\delta_1 = \delta_2 = 0.1$, $N = 9$, and $N = 11$; whereas Fig. 2b shows the curves for $K = 100$, $\delta_1 = 0.01$ ($\delta_2 = 0.0001$), $N = 19$, and $N = 21$. The minima along each curve are the extraripple filters. (For fixed values of δ_1 and δ_2, as in this figure, there are only $(N - 1)/2$ distinct extraripple filters.) The local regions around the minima are the scaled extraripple filters[7] and the remainder of the curve represents equiripple solutions with one less than the maximum number of ripples. As will be shown in the next section, the first and last extraripple filters can be obtained analytically because they are simply related to the Chebyshev polynomial of appropriate degree.[8,9]

As seen from Fig. 2, as F_p varies from 0 to its maximum possible value, the transition width goes through a sequence of minima and maxima. The variation in the transition widths of the minima and maxima decreases as N increases. Furthermore, the variation in transition width between adjacent maxima and minima also decreases as N increases. In fact, except for a narrow region at the beginning and end of the curve, the curve of transition width versus passband cutoff frequency is relatively flat over a wide range of values of F_p, δ_1, and K. Figure 3 shows a sequence of three plots of transition width versus passband cutoff frequency for extraripple filters (i.e., only the minima of the curve are plotted) of length $N = 101$ for various values of δ_1 and K. (The entire curves are not plotted because the amount of computation required for a smooth curve of such high order is inordinately high.) Figure 3a shows a sequence of four curves for $K = 1$, $\delta_1 = 0.1$, 0.01, 0.001, 0.0001, whereas Fig. 3b shows the same sequence for $K = 10$, and Fig. 3c shows the sequence for $K = 100$. Several observations can be made from this figure.

(*i*) For a wide range of values of F_p, and fixed δ_1 and δ_2, the transition width of the extraripple filters is relatively insensitive to F_p. The larger the value of K, or the smaller the value of δ_1, the worse this approximation becomes.

(*ii*) In the regions of either very small or very large values of F_p, the transition width generally decreases.

Based on these curves and on previous results with window designs,[10]

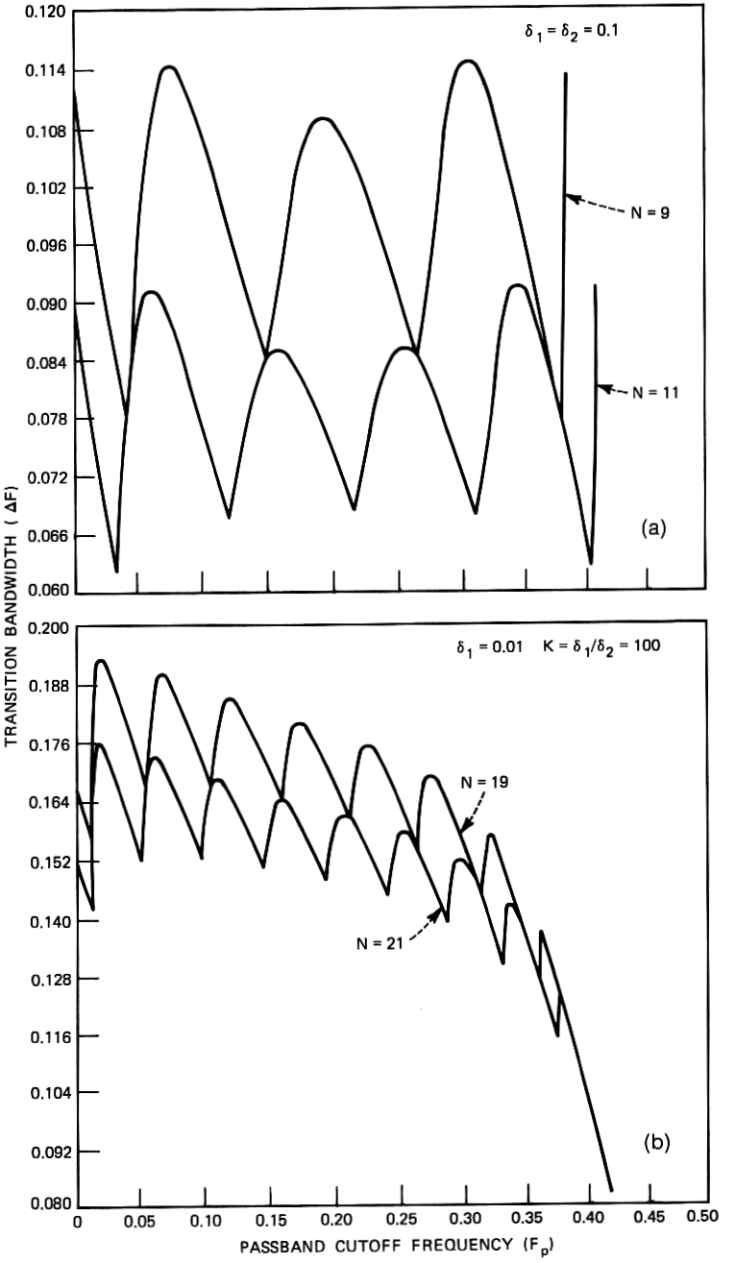

Fig. 2—The transition width as a function of passband cutoff frequency for two sets of filter parameters.

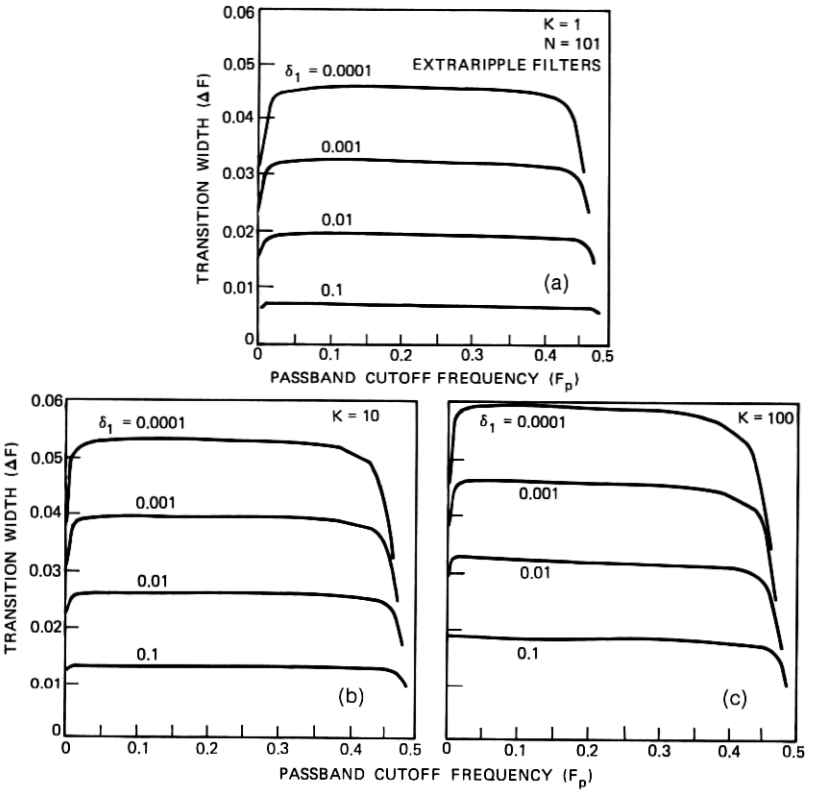

Fig. 3—The transition width as a function of passband cutoff frequency for $N = 101$-point extraripple filters and $K = 1$, 10, and 100.

it seems fairly reasonable to expect some simple design relationships to exist between the five basic filter parameters, N, F_p, F_s, δ_1, and δ_2 (or K), at least in the extreme case of the Chebyshev solution, and also for a reasonably large region near $F_p = 0.25$. Experience in practical situations has shown that the number of terms needed in the optimum FIR low-pass filter to meet design specifications is significantly less than the number of terms estimated by known relationships on windows.[10] Therefore, the goals of this paper are to obtain approximations to the actual design relationships between linear-phase, low-pass filter design parameters and to illustrate their use in actual design examples.

The organization of this paper is as follows. First, the design relationships for the Chebyshev solution are derived, and approximate formulas for the transition width in the limit of large values of N are obtained. Then the results of measurements on a wide range of filters

where the number of passband and stopband ripples are equal are shown. Then minimum mean-square relative error fits to the observed data (for large N) assuming both linear and nonlinear dependency on the basic filter parameters are computed. To apply the design relationships for all values of N, a correction formula is derived, based on consideration of the appropriate transition width of the filter. Finally, a set of rules is presented for going from a set of desired filter parameters to an estimate of the lowest-order filter which meets these specifications.

1.1 *Summary of Design Relationships*

Given the low-pass filter parameters F_p, F_s, δ_1, and δ_2 (or, equivalently, $K = \delta_1/\delta_2$), the minimum filter impulse response duration, N required to meet the above specifications can be estimated from the relation

$$N = \frac{D_\infty(\delta_1, \delta_2) - f(K)(F_s - F_p)^2}{(F_s - F_p)} + 1,$$

where

$$D_\infty(\delta_1, \delta_2) = [5.309 \times 10^{-3} (\log_{10} \delta_1)^2 + 7.114 \times 10^{-2} (\log_{10} \delta_1) - 0.4761] \log_{10} \delta_2 - [2.66 \times 10^{-3} (\log_{10} \delta_1)^2 + 0.5941 (\log_{10} \delta_1) + 0.4278]$$

and

$$f(K) = 0.51244 \log_{10} K + 11.01217.$$

The above relations are valid to within 1.3 percent relative error in N if $\delta_1 \leqq 0.1$ and $\delta_2 \leqq 0.1$.

II. CHEBYSHEV SOLUTIONS

Let $\{h(n), n = -(N-1)/2, \cdots, (N-1)/2\}$ be the impulse response of the desired digital filter. (N is assumed to be odd throughout this paper.) The impulse response satisfies the symmetry condition $h(n) = h(-n)$ to give the desired linear phase. The frequency response of the filter is given by

$$H(e^{j2\pi f}) = h(0) + \sum_{n=1}^{(N-1)/2} 2h(n) \cos(2\pi f n). \tag{2}$$

By making the substitution

$$\cos(2\pi f) = \frac{x - \left(\frac{X_0 - 1}{2}\right)}{\left(\frac{X_0 + 1}{2}\right)} \tag{3a}$$

or

$$x = \left(\frac{X_0 + 1}{2}\right)\cos(2\pi f) + \left(\frac{X_0 - 1}{2}\right) \tag{3b}$$

the interval $0 \leqq f \leqq 0.5$ is mapped to the interval $X_0 \geqq x \geqq -1$. It is easily shown[3,7] that the mapping transforms the trigonometric polynomial of eq. (1) to an algebraic polynomial in x of the form

$$G(x) = \sum_{n=0}^{(N-1)/2} b(n)x^n, \tag{4}$$

where the $\{b(n)\}$ are straightforwardly related to the $\{h(n)\}$.

The basic filter design problem is to find coefficients $b(n)$ [or $h(n)$] such that the weighted error of approximation is equiripple in both the passband and stopband. In the case when either the passband or the stopband has only one ripple, the solution to the filter design problem may be obtained analytically based on the theory of Chebyshev polynomials. In all other cases, alternative techniques must be used to arrive at the appropriate solution.

Consider the Mth degree Chebyshev polynomial, $T_M(x)$, as shown in Fig. 4 for $M = 4$. The standard representation for $T_M(x)$ is

$$T_M(x) = \cos[M \cos^{-1}x], \qquad |x| \leqq 1 \tag{5}$$

and

$$T_M(x) = \cosh[M \cosh^{-1}x], \qquad |x| > 1. \tag{6}$$

In the interval $-1 \leqq x \leqq 1$, $T_M(x)$ oscillates between ± 1 and, beyond the value $x = +1$, $T_M(x)$ grows approximately as x^M. If we define X_0 (see Fig. 4) as the point where $T_M(X_0) = (1 + \delta_1)/\delta_2$, and X_p as the point where $T_M(X_p) = (1 - \delta_1)/\delta_2$, it is readily seen that $\delta_2 \cdot T_M(x)$ is a polynomial of the form of eq. (4) [with $M = (N - 1)/2$] which is equiripple in both the passband $X_p \leqq x \leqq +X_0$, and the stopband $-1 \leqq x \leqq +1$, and hence is an optimal solution to the filter design problem. Of course, this solution is a special case of the general solution in that there is only one ripple in the passband, but at least it is an analytically tractable case from which a great deal can be learned about the relationships between the filter parameters.

It is straightforward to solve for the points X_p and X_0 in terms of δ_1, δ_2, and N. If we set $M = (N - 1)/2$, then at $x = X_0 > 1.0$, we get the relation

$$T_M(X_0) = \frac{1 + \delta_1}{\delta_2} = \cosh[M \cosh^{-1} X_0] \tag{7}$$

or

$$X_0 = \cosh\left[\frac{1}{M}\cosh^{-1}\left(\frac{1 + \delta_1}{\delta_2}\right)\right]. \tag{8}$$

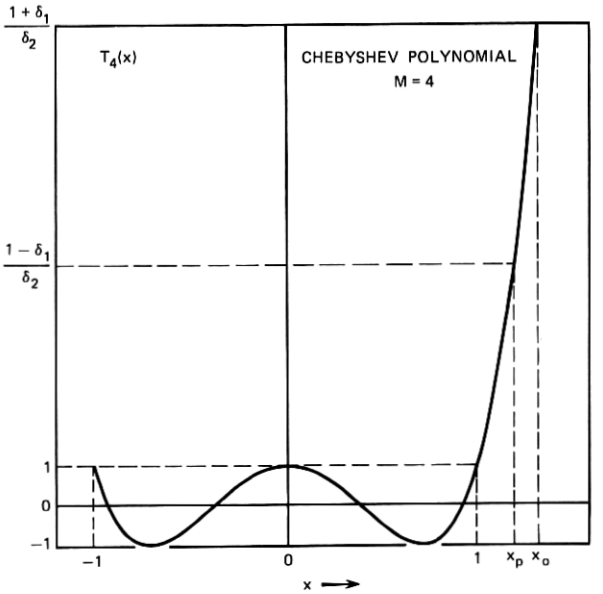

Fig. 4—A fourth-degree Chebyshev polynomial defining the points X_0 and X_p.

At $x = X_p > 1.0$, we get the relation

$$T_M(X_p) = \frac{1 - \delta_1}{\delta_2} = \cosh\left[M \cosh^{-1} X_p\right] \tag{9}$$

or

$$X_p = \cosh\left[\frac{1}{M} \cosh^{-1}\left(\frac{1 - \delta_1}{\delta_2}\right)\right]. \tag{10}$$

The inverse mapping of eq. (3) can be used to determine the filter cutoff frequencies by the relation

$$f = \frac{1}{2\pi} \cos^{-1}\left[\frac{2x - X_0 + 1}{X_0 + 1}\right], \qquad |x| \leqq X_0. \tag{11}$$

Thus, F_p (corresponding to X_p) and F_s (corresponding to $x = +1.0$) can be readily obtained using eq. (11) with the appropriate values for x. In this case, the transition width $(F_s - F_p)$ can be analytically determined for all values of N, δ_1, and δ_2 as:

$$\Delta F = F_s - F_p = \frac{1}{2\pi}\left[\cos^{-1}\left(\frac{3 - X_0}{1 + X_0}\right) - \cos^{-1}\left(\frac{2X_p - X_0 + 1}{1 + X_0}\right)\right]. \tag{12}$$

In the limit of large values of N (or, equivalently, M), eq. (12) can be simplified by the following approximations. First, for sufficiently small values of α,

$$\cosh \alpha \approx 1 + \frac{\alpha^2}{2}, \qquad \alpha \ll 1. \tag{13}$$

Thus, eqs. (10) and (8) can be simplified to the form [replacing M by $(N-1)/2$]

$$X_p \approx 1 + \frac{\left[\left(\frac{2}{N-1}\right)\cosh^{-1}\left(\frac{1-\delta_1}{\delta_2}\right)\right]^2}{2}. \tag{14}$$

$$X_0 \approx 1 + \frac{\left[\left(\frac{2}{N-1}\right)\cosh^{-1}\left(\frac{1+\delta_1}{\delta_2}\right)\right]^2}{2}. \tag{15}$$

The approximation is then made that for sufficiently small ϵ,

$$\cos^{-1}(1-\epsilon) \approx \sqrt{2\epsilon}. \tag{16}$$

Thus, F_s and F_p are well approximated as

$$F_s \approx \frac{1}{2\pi}\left[\frac{2}{N-1}\cosh^{-1}\left(\frac{1+\delta_1}{\delta_2}\right)\right] \tag{17}$$

$$F_p \approx \frac{1}{2\pi}\left(\frac{2}{N-1}\right)\left[\left[\cosh^{-1}\left(\frac{1+\delta_1}{\delta_2}\right)\right]^2 - \left[\cosh^{-1}\left(\frac{1-\delta_1}{\delta_2}\right)\right]^2\right]^{\frac{1}{2}}. \tag{18}$$

Thus, for large N ($N \gg 1$) the approximation to the transition width curve is given by

$$\Delta F = F_s - F_p \approx \frac{1}{\pi(N-1)}\left[\cosh^{-1}\left(\frac{1+\delta_1}{\delta_2}\right) - \left[\left[\cosh^{-1}\left(\frac{1+\delta_1}{\delta_2}\right)\right]^2 - \left[\cosh^{-1}\left(\frac{1-\delta_1}{\delta_2}\right)\right]^2\right]^{\frac{1}{2}}. \tag{19}$$

(This approximation is valid to within 1 percent for $N \geqq 51$.) Equation (19) shows ΔF to be inversely proportional to $(N-1)$. This identical inverse behavior has been noted previously for filters designed by windowing.[10] These and other considerations lead one to consider as a performance measure of a low-pass digital filter the quantity D defined as

$$D = (N-1)\Delta F = (N-1)(F_s - F_p) \tag{20}$$

which, in many cases, depends only on δ_1 and δ_2.

The curves in Fig. 5 (for $N = 127$) show plots of performance D versus $20 \log_{10} (\delta_2)$ both for values of $K = \delta_1/\delta_2$ from $K = 1$ to $K = 1000$ and for values of δ_1 from 0.5 to 0.0001. The behavior of these curves is as predicted from eq. (19) and from intuition about the behavior of D as δ_1 and δ_2 get large. It is clear that when $\delta_1 + \delta_2 = 1.0$ [i.e., $\delta_1 = K/(K + 1)$, $\delta_2 = 1/(K + 1)$], then $D = 0$ since there is no transition band. In this case, the term $\cosh^{-1} (1 - \delta_1)/\delta_2$ vanishes, and the first and second terms in eq. (19) cancel exactly. In the limit of small values of δ_1, the second and third terms in eq. (19) are approximately equal and cancel. Thus, D is approximately of the form

$$D \approx \frac{1}{\pi} \cosh^{-1} \left(\frac{1 + \delta_1}{\delta_2} \right). \tag{21}$$

Since

$$\cosh^{-1} (x) = \ln (x + \sqrt{x^2 - 1}), \tag{22}$$

eq. (21) can be rewritten as (assuming δ_1 negligible)

$$D \approx \frac{1}{\pi} (\ln 2 - \ln \delta_2) \tag{23}$$

which is independent of δ_1. Thus, as seen in Fig. 5b, in the case of small δ_1, D is essentially independent of δ_1.

The curves of Fig. 6 show the behavior of D for various values of $N \leqq 127$, for $K = 1$, 10, and 100. The values of N used for these plots were $N = 3$, 7, 11, 19, 51, and 127. The differences between the data for $N = 127$ and the data for $N = 51$ are relatively small. These curves also exhibit another interesting phenomenon. The curve for $N = 3$ saturates at a value of $D = 1$. This is due to the limitation that the transition width, ΔF, must be less than or equal to 0.5. Thus, the saturation value of D is $(N - 1)/2$, or 1.0 for $N = 3$. The larger the value of K, the larger the value of δ_2 beyond which the curve for $N = 3$ saturates.

In the case of the Chebyshev solution to the optimal filter design problem, a formula can be derived for the impulse response duration N of a filter whose response meets specified values of δ_1, δ_2, and F_s. Since F_p is not specified, this result is useful only as a first guess of a value of N which meets specifications on all four filter parameters. From the discussion given earlier in this section, $f = F_s$ when $x = 1.0$. Thus, eq. (11) can be used to solve for X_0 as

$$X_0 = \frac{3 - \cos (2\pi F_s)}{1 + \cos (2\pi F_s)}. \tag{24}$$

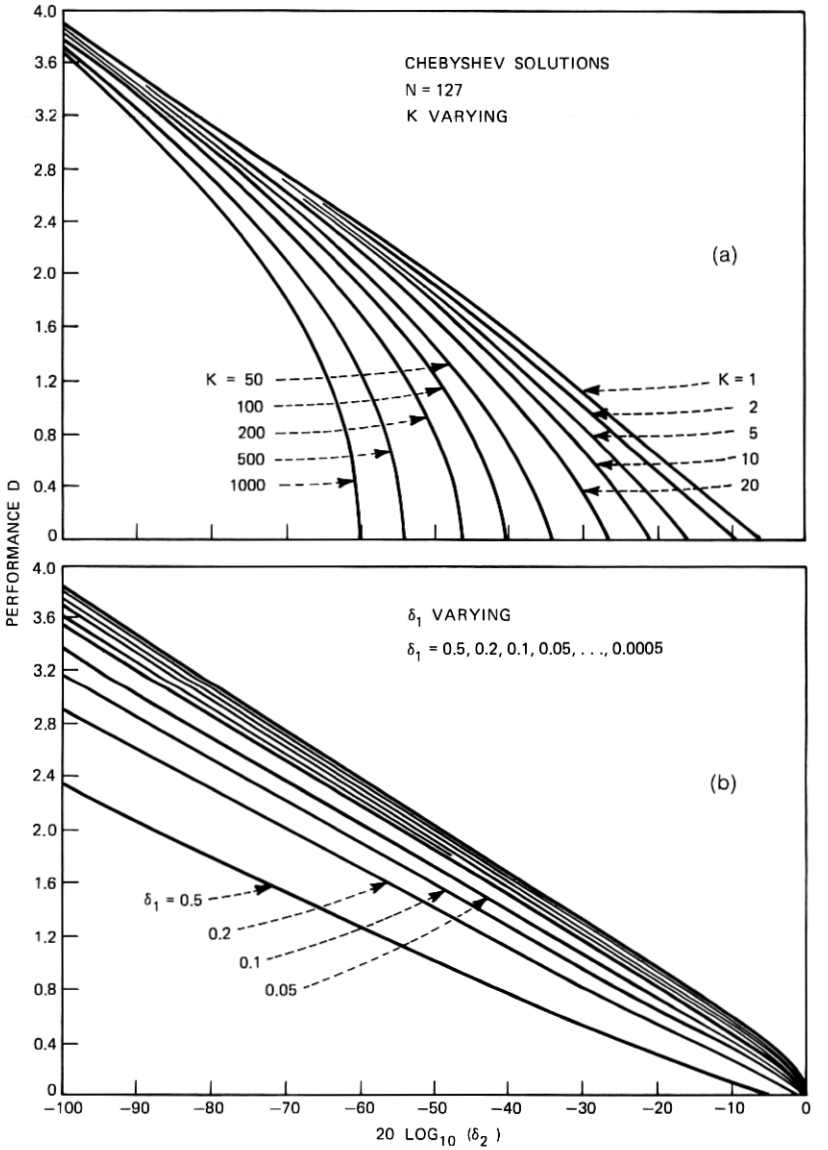

Fig. 5—Plots of D versus $\log_{10} \delta_2$ as a function of K and δ_1 for the $N = 127$ Chebyshev solutions.

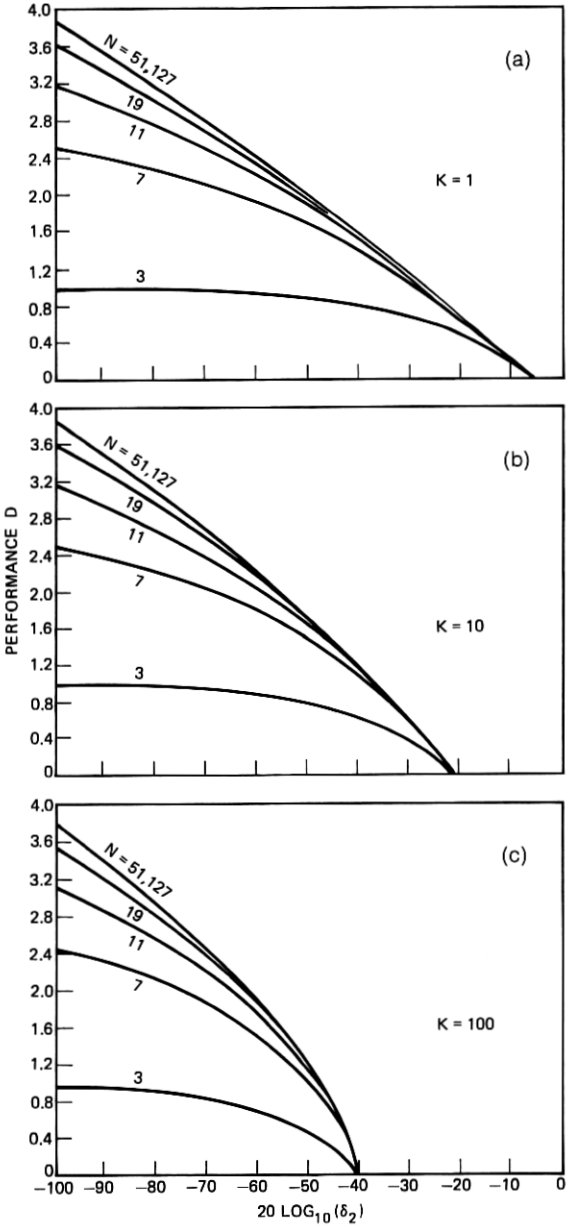

Fig. 6—Plots of D versus $\log_{10} \delta_2$ as a function of N for $K = 1$, 10, and 100, for the Chebyshev solutions.

Equation (8) may now be used to solve for N as

$$N = 1 + \frac{2 \cosh^{-1}\left(\frac{1 + \delta_1}{\delta_2}\right)}{\cosh^{-1}\left[\frac{3 - \cos(2\pi F_s)}{1 + \cos(2\pi F_s)}\right]}. \tag{25}$$

Using trigonometric identities, eq. (25) can be simplified to the form

$$N = 1 + \frac{\cosh^{-1}\left(\frac{1 + \delta_1}{\delta_2}\right)}{\cosh^{-1}\left(\frac{1}{\cos(\pi F_s)}\right)}. \tag{26}$$

For these values of X_0 and N, F_p may be obtained from eqs. (10) and (11). If the transition width obtained is too large (too small), N is then decreased (increased) until the desired specifications are approximately achieved. (Since there is only one Chebyshev solution for fixed δ_1, δ_2, and N, exact values of both F_p and ΔF cannot usually be obtained.) As will be seen later, eq. (26) forms the basis for estimating a lower bound on the filter order required to meet given design specifications on ΔF, δ_1, and δ_2.

As mentioned earlier, Chebyshev polynomials can be used as the optimal solution in the case when there is one ripple in either the passband or the stopband. To see how this second case can be handled, consider the Chebyshev polynomial $T_M(x)$ where $T_M(X_0) = (1 + \hat{\delta}_2)/\hat{\delta}_1$ and $M = (N - 1)/2$. As shown earlier, this polynomial represents an optimal filter with passband ripple $\hat{\delta}_2$, stopband ripple $\hat{\delta}_1$, passband cutoff frequency F_p, and stopband cutoff frequency F_s. Consider the function $R_M(x)$ defined as

$$R_M(x) = 1 - \hat{\delta}_1 T_M(-x + X_0 - 1), \tag{27}$$

where $-1 \leqq x \leqq X_0$. An examination of the properties of $R_M(x)$ shows

(*i*) $R_M(x)$ is a polynomial in x of degree M.
(*ii*) In the interval $X_0 - 2 \leqq x \leqq X_0$, $R_M(x)$ oscillates between the values $1 - \hat{\delta}_1$, and $1 + \hat{\delta}_1$.
(*iii*) In the interval $-1 \leqq x \leqq X_0 - 1 - X_p$, $R_M(x)$ goes from $-\hat{\delta}_2$ to $\hat{\delta}_2$.

If we define $\hat{F}_p$ and $\hat{F}_s$ as the equivalent filter cutoff frequencies, then

it is readily shown that

$$\begin{aligned} \hat{F}_p &= 0.5 - F_s \\ \hat{F}_s &= 0.5 - F_p. \end{aligned} \tag{28}$$

Thus $R_M(x)$ is a polynomial with only one stopband ripple which satisfies the filter optimality criterion. In summary, to design Chebyshev approximations with only one *stopband* ripple, one merely makes the substitution

$$\begin{aligned} \delta_1 &= \hat{\delta}_2 \\ \delta_2 &= \hat{\delta}_1 \\ F_p &= 0.5 - \hat{F}_s \\ F_s &= 0.5 - \hat{F}_p \end{aligned} \tag{29}$$

and solves for an equivalent filter with one *passband* ripple using the formulas of this section.

The data on the Chebyshev solutions provide valuable insights into the behavior of the design relationships between filter parameters in more general cases. These data and their design relationships are presented in the next sections.

III. MEASUREMENTS OF D

Earlier, it was shown that the transition width or, equivalently, the performance measure D for the Chebyshev solutions (i.e., either small F_p or large F_s) was generally significantly smaller than for most of the range of values of passband cutoff frequency. To obtain data on a more realistic set of filters, the value of D was measured for a large number of extraripple filters designed with the constraint that the number of passband and stopband ripples were the same. In this manner, the passband and stopband widths were almost equal and, thus, the measured data would characterize the parameter set over as wide a range of values of δ_1, δ_2, and F_p as possible. Six different values of N were used including $N = 3$, 7, 11, 19, 51, and 127. Over 1500 filters were designed to cover the parameter range $0.00001 \leqq \delta_2 \leqq 0.5$, and $1 \leqq K \leqq 500$ ($K = \delta_1/\delta_2$). Figures 7a through 7i show plots of D versus $20 \log_{10} (\delta_2)$ for the nine values of K and the chosen values of N. All these data are presented since they are fairly general and may be useful in a wide variety of contexts other than this paper. (Also, their measurement required almost 3 hours of computer time on a fairly fast processor.)

It is remarkable how similar the plots of D versus $\log \delta_2$ for the more general case of Fig. 7 are to the identical plots for the Chebyshev solutions. Similar behavior for small D is expected, since D tends to 0 as

$\delta_1 + \delta_2$ tends to 1.0 independent of N, F_p, and F_s. However, the approximately linear behavior of D as a function of $\log \delta_2$ and $\log K$ (for large N) is unexpected. Another similarity between the two cases is the independence of D of $(N - 1)$ for large N. The tendency of D to saturate for small values of N is yet another similarity between the curves. The main difference between the sets of curves is that, in the Chebyshev case, D is approximately independent of δ_1 for small δ_1. This behavior is not observed for the extraripple filters of Fig. 7.

A summary of the behavior of D for $N = 127$, as a function of $\log \delta_2$ for various values of either K or δ_1, is presented in Fig. 8. The values of δ_1 used were 0.5, 0.2, 0.1, $\cdots$, 0.00002, whereas K ranged from 1 to 500 as in Figs. 7a through 7i. In some sense, this figure represents a set of design curves for high-degree low-pass filters. In the next section, we show how the data of Fig. 8 can be approximated by linear and nonlinear fits, and how simple modifications can be made to correct the results for values of N less than 127.

IV. DATA-FITTING PROCEDURES

In order to make most efficient use of the data of the previous section in a practical design problem, it is useful to express the relationships between the filter parameters in a simple manner. Since we know of no way of deriving exact analytical formulas, as in the Chebyshev case, a minimum mean-square relative error fit to the data over a restricted but reasonable range was sought. Both a linear and a nonlinear fit to the data of Fig. 8 (N large) were obtained. Corrections for smaller values of N were then obtained giving a complete set of design rules.

The data of Fig. 8 suggest that except in the region $D = 0$ (large values of δ_1) a simple linear fit can be obtained. The curves of Fig. 8 were assumed to be of the form

$$\begin{aligned} D_{\text{LIN}} &= a + b \log_{10} (\delta_2) + c \log_{10} K, \\ a &= -0.803, \\ b &= -1.359, \\ c &= -0.737. \end{aligned} \tag{30}$$

where D_{LIN} is the predicted value of D. The values for a, b, and c were chosen to minimize the sum of the squares of the relative differences between D_{LIN} and D for the $N = 127$ data for values of δ_1 in the range $0.01 \geqq \delta_1$; $0.01 \geqq \delta_2$. The reason relative rather than absolute errors were considered is that a fixed percentage error in D, δD, approximately gives a fixed percentage error in N, δN, when transition width is held

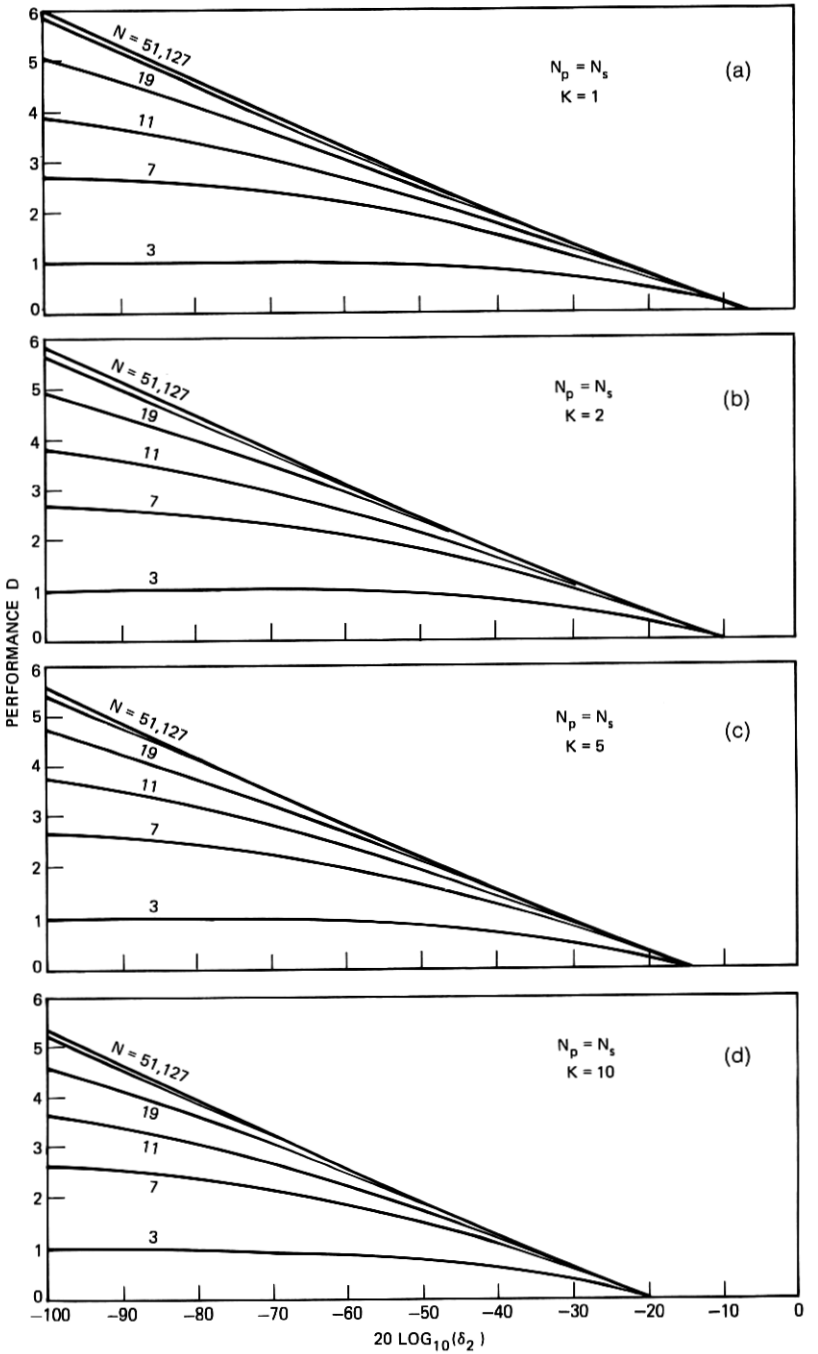

Fig. 7—Plots of D versus $\log_{10} \delta_2$ as a function of N for $K = 1, 2, 5, 10, 20, 50, 100, 200, 500$ for the case of extraripple filters with the same number of passband and stopband ripples.

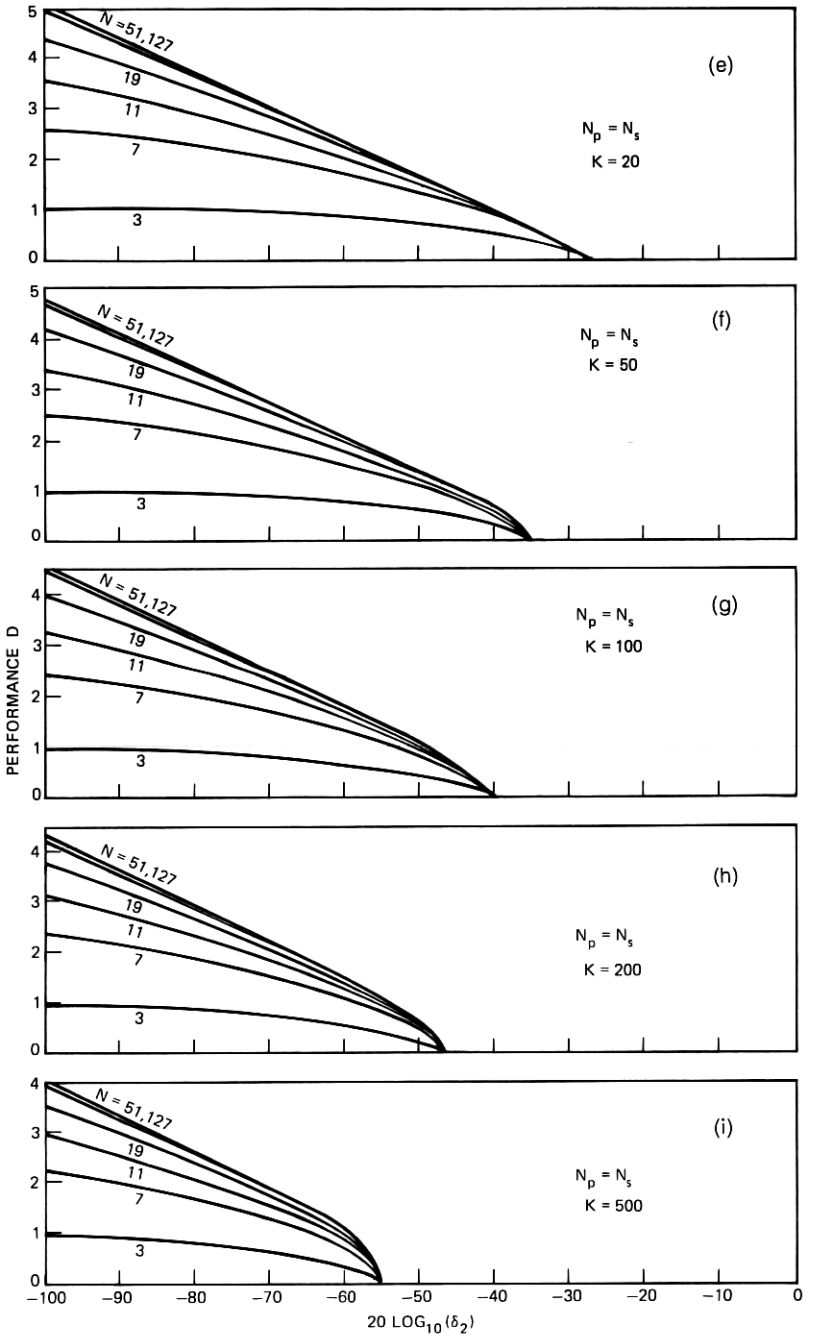

Fig. 7 (continued).

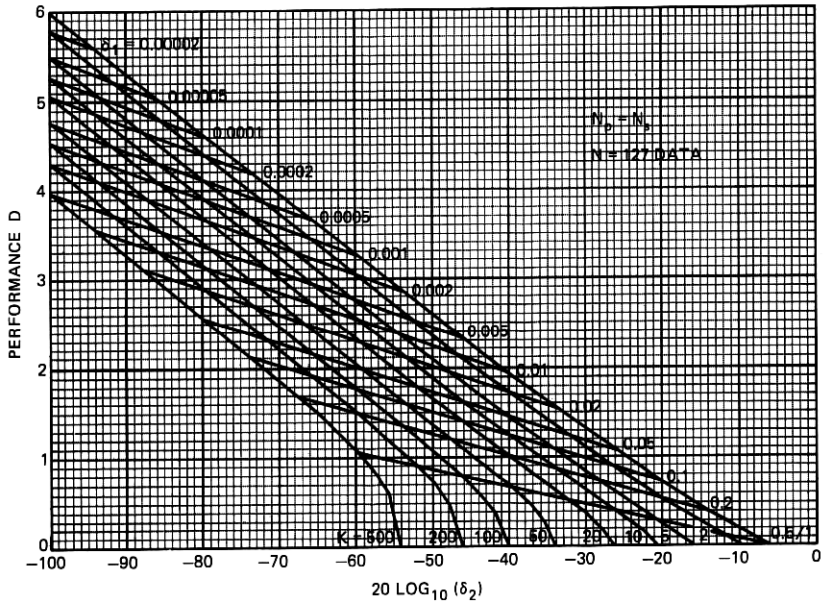

Fig. 8—Plot of D versus $\log_{10} \delta_2$ for the $N = 127$ data of Fig. 7 showing curves for various values of K and δ_1.

fixed, as in most design problems, i.e.,

$$\frac{\delta D}{D} = \frac{\delta(N-1)}{N-1} = \frac{\delta N}{N-1}. \tag{31}$$

Figure 9 shows a plot of the relative error of the predicted values of D versus $\log \delta_2$ for $\delta_1 = 0.01, 0.005, 0.002, \cdots, 0.00002$. Except for a small region on the curve $\delta_1 = 0.01$, the relative error is less than 1.0 percent for the entire range of δ_2 and δ_1 considered. Based on a value of $N = 127$, a relative error of 1.0 percent in D is equivalent to an error of 1.26 samples in N, or approximately one-half a filter order off from the correct order. Errors of this magnitude are generally considered to be quite small, i.e., the prediction is reasonably good.

In an effort to improve the fit and extend the range of applicability of the approximation, a nonlinear formula was chosen for D. Based on the data of Fig. 8, it was observed that the slope of the curve of D versus $\log_{10} \delta_2$ changes nonlinearly with $\log \delta_1$. The simplest approximation was to try a fit which was linear with respect to $\log \delta_2$ and

quadratic in log δ_1. Such a fit is of the form

$$D_{NL} = [a_1(\log_{10}\delta_1)^2 + a_2\log_{10}\delta_1 + a_3]\log_{10}\delta_2 + [a_4(\log_{10}\delta_1)^2 + a_5\log_{10}\delta_1 + a_6]. \quad (32)$$

The constants a_1 to a_6 were chosen to minimize the mean-square relative error for the $N = 127$ data over the range $0.1 \geqq \delta_1 \geqq 0.000001$, $0.1 \geqq \delta_2 \geqq 0.000001$, and turned out to be

$$\begin{aligned} a_1 &= 5.309 \times 10^{-3} \\ a_2 &= 7.114 \times 10^{-2} \\ a_3 &= -4.761 \times 10^{-1} \\ a_4 &= -2.660 \times 10^{-3} \\ a_5 &= -5.941 \times 10^{-1} \\ a_6 &= -4.278 \times 10^{-1}. \end{aligned}$$

Figure 10 shows the relative error of the predicted value of D as a function of log δ_2 for values of δ_1 from 0.1 to 0.00002. The peak percentage error is 1.3 percent, and over most of the range the percentage error is

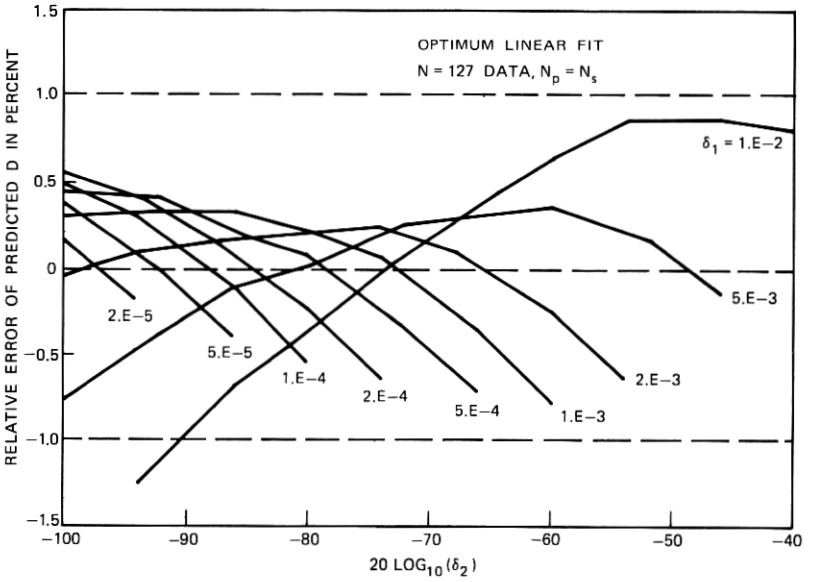

Fig. 9—The relative errors in fitting the data of Fig. 8 with a linear curve over the range $\delta_1 \leqq 0.01$, for various values of δ_1 ($N = 127$).

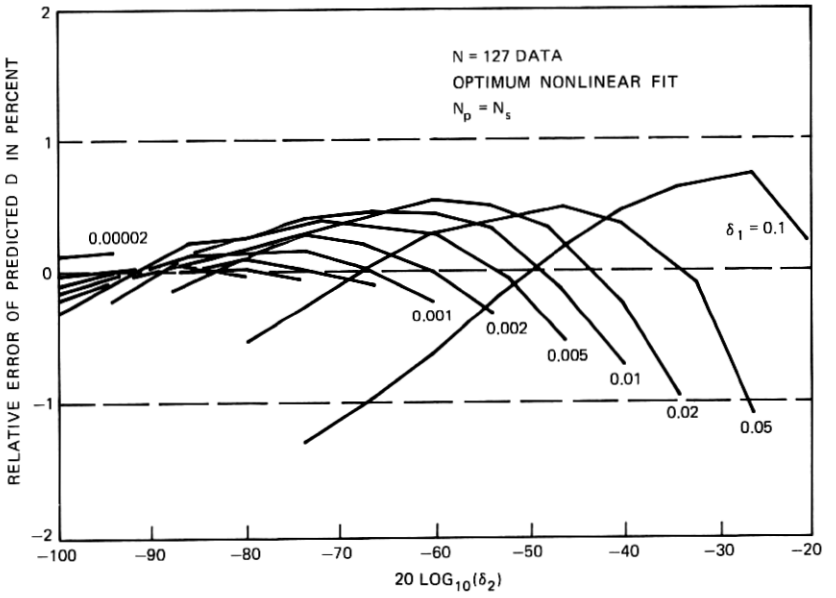

Fig. 10—The relative errors in fitting the data of Fig. 8 with a nonlinear curve over the range $\delta_1 \leqq 0.1$, for various values of δ_1 ($N = 127$).

much smaller. Clearly, this prediction formula is acceptable for almost any design application.

Figure 11 shows a summary of the predicted values of D as a function of stopband attenuation for a wide range of values of passband ripple. In this case, passband ripple in dB is defined as

$$\text{Passband ripple} = 20 \log_{10}\left(\frac{1}{1 - \delta_p}\right), \tag{33}$$

$$\text{Stopband attenuation} = -20 \log_{10}(\delta_s), \tag{34}$$

where

$$\delta_p = \frac{2\delta_1}{1 + \delta_1} \tag{35}$$

and

$$\delta_s = \frac{\delta_2}{1 + \delta_1}. \tag{36}$$

These data correspond to standard design data for continuous-time filters where the frequency response magnitude is constrained to be less than or equal to 1.0.

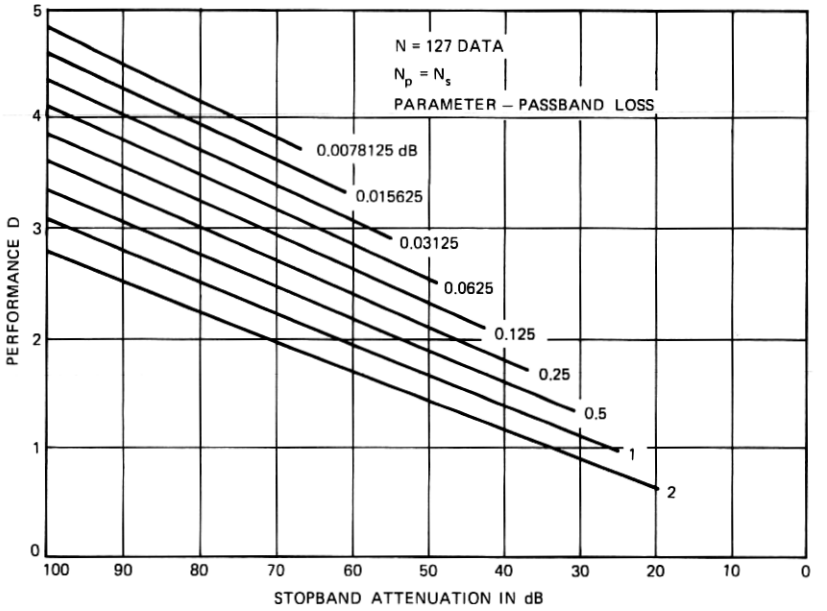

Fig. 11—Plots of D versus $20 \log_{10} \delta_2$ for various values of the parameter $20 \log_{10} (1/1 - \delta_p)$, where $\delta_p = 2\delta_1/(1 + \delta_1)$, as calculated from the optimum nonlinear fits to the data of Fig. 8 ($N = 127$).

V. CORRECTIONS FOR SMALL VALUES OF N

The formulas in the previous section are accurate for predicting D (or, equivalently, N) for values of N greater than about 51. As seen from the curves in Fig. 7, as N decreases, D decreases for fixed values of δ_2 and K. It is also seen from Fig. 7 that the differences increase with decreasing δ_2 or, equivalently, increasing transition width. An examination of the relative deviation of D_{NL}, the predicted value of D, from its true value as a function of transition bandwidth showed that, independent of N, the deviations could be simply approximated by a curve of the form

$$D_{cor} = f(K)(\Delta F)^2, \tag{37}$$

where D_{cor} is the correction term and $f(K)$ is of the form

$$f(K) = (0.51244 \log_{10} K + 11.01217). \tag{38}$$

(The constants in eq. (38) were again obtained by a minimum mean-square relative error data-fitting procedure.) Thus, using eqs. (37) and (38), a formula for D which depends on N, K, and δ_2 can be obtained.

Adopting the notation

(*i*) $D_\infty(\delta_2, K) = D_{NL}$ of eq. (32)*

(*ii*) $\hat{D}(\delta_2, K, N)$ = predicted value of D as a function of N as well as δ_2 and K

(*iii*) $D(\delta_2, K, N)$ = true value of D,

we obtain the relation

$$D_\infty(\delta_2, K) - \hat{D}(\delta_2, K, N) = f(K)(\Delta F)^2. \tag{39}$$

Thus, in a design case where δ_2, K (or δ_1), F_p, and F_s (i.e., ΔF) are specified and the problem is to estimate N, the impulse response duration required to meet these specifications, eq. (39) can be used directly since

$$\hat{D}(\delta_2, K, N) = (N - 1)\Delta F. \tag{40}$$

Thus, combining eqs. (37) and (38) and solving for N gives

$$N = \frac{D_\infty(\delta_2, K) - f(K)(\Delta F)^2}{\Delta F} + 1. \tag{41}$$

A closed-form expression for $\hat{D}(\delta_2, K, N)$ may be obtained by substituting eq. (40) for $\hat{D}(\delta_2, K, N)$ in eq. (39) and solving the quadratic equation for ΔF. Equation (40) is then used to give $\hat{D}(\delta_2, K, N)$. Thus, we get

$$f(K)(\Delta F)^2 + (N - 1)\Delta F - D_\infty(\delta_2, K) = 0 \tag{42}$$

$$\Delta F = \frac{(N-1)}{2f(K)}\left(\sqrt{1 + \frac{4f(K)D_\infty(\delta_2, K)}{(N-1)^2}} - 1\right) \tag{43}$$

$$\hat{D}(\delta_2, K, N) = \frac{(N-1)^2}{2f(K)}\left(\sqrt{1 + \frac{4f(K)D_\infty(\delta_2, K)}{(N-1)^2}} - 1\right). \tag{44}$$

In the limit, as N tends to infinity, eq. (44) shows $\hat{D}(\delta_2, K, N)$ tends to $D_\infty(\delta_2, K)$ as expected.

Using eq. (44), the relative error of the predicted value of D from the true value was measured for the data for $N = 3, 7, 11, 19, 51$, and 127 with $\delta_1 \leqq 0.1$. The relative errors for values of K from 1 to 500 are plotted in Figs. 12a through 12f. In all cases, the worst relative error in D is sufficiently small that the equivalent error in N is less than *one sample*. Thus, for all practical purposes, the design equations above serve as a useful guide for estimating the order of the filter required to meet design specifications.

* Eq. (32) gives D_{NL} as a function of δ_1 and δ_2 but since $\delta_1 = K\delta_2$, it is also implicitly a function of δ_2 and K.

VI. SUMMARY OF DESIGN PROCEDURES AND EXAMPLES

Based on the results presented in this paper, it is now possible to give a set of rules for estimating the filter impulse response duration required to meet given design specifications. These rules are as follows:

(*i*) Check if either δ_1 or δ_2 is greater than 0.1, in which case the graphical data of Fig. 7 are used directly to estimate N.

(*ii*) Calculate a value of N, call this N_1, corresponding to the extraripple case where $N_p = N_s$, from the equation

$$N_1 = \frac{D_\infty(\delta_1, \delta_2) - f(K)(\Delta F)^2}{\Delta F} + 1,$$

where $D_\infty(\delta_1, \delta_2)$ is the optimum nonlinear fit to the data for large N and is given by:

$$D_\infty(\delta_1, \delta_2) = [5.309 \times 10^{-3} [\log_{10} \delta_1]^2 + 7.114 \times 10^{-2} \log_{10} \delta_1 - 0.4761] \log_{10} \delta_2 - [2.66 \times 10^{-3} (\log_{10} \delta_1)^2 + 0.5941 \log_{10} \delta_1 + 0.4278]$$

and $f(K)$ is given by

$$f(K) = 0.51244 \log_{10} (K) + 11.01217.$$

(*iii*) If the desired value of F_p is less than or equal to 0.04 (let us call this case 1), or if the desired value of F_s is greater than or equal to 0.46 (case 2), then the estimate of N is obtained in rule (*iv*). Otherwise, the value N_1 of rule (*ii*) is used as the estimate of N.

(*iv*) To obtain the value of N_c for the Chebyshev solution for case 1, eq. (26) is used to get a first approximation to N_c. N_c is then systematically varied until a Chebyshev solution is obtained which meets specifications on δ_1, δ_2, and ΔF. (As discussed earlier, it is not generally possible to find a Chebyshev solution which meets specifications exactly on all four filter parameters.) For case 2, δ_1 and δ_2 are interchanged, and F_s is replaced by $0.5 - F_p$, in order to solve for N_c from eq. (26). In cases where this rule is applied, the value N_c obtained is a lower bound to the true value of N_1. No upper bound may be given in this case. A discussion of this problem is given below.

Several comments are necessary about these rules before proceeding to some examples. The discussion in this paper has concentrated on two regions of the curve of ΔF versus F_p – the Chebyshev solutions and the case of extraripple filters with an equal number of passband and stop-

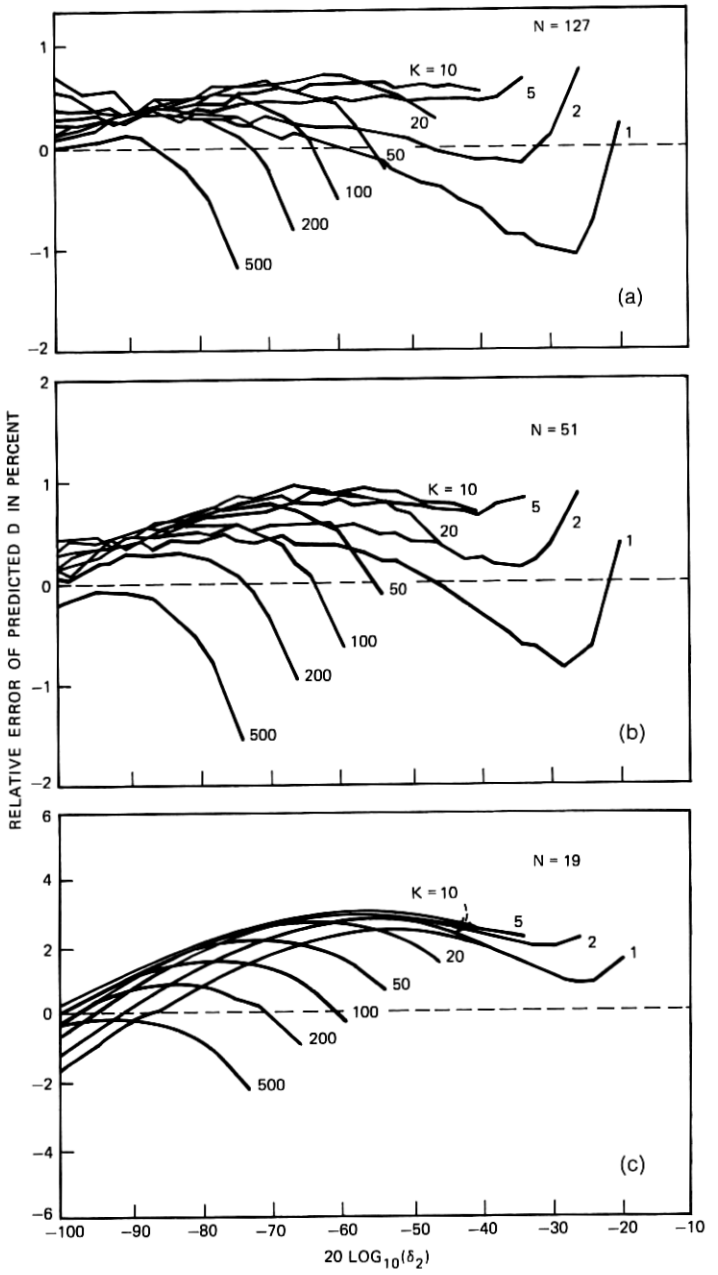

Fig. 12—Plots of the relative errors in fitting the data of Fig. 7 using the corrected values of D for N = 3, 7, 11, 19, 51, and 127.

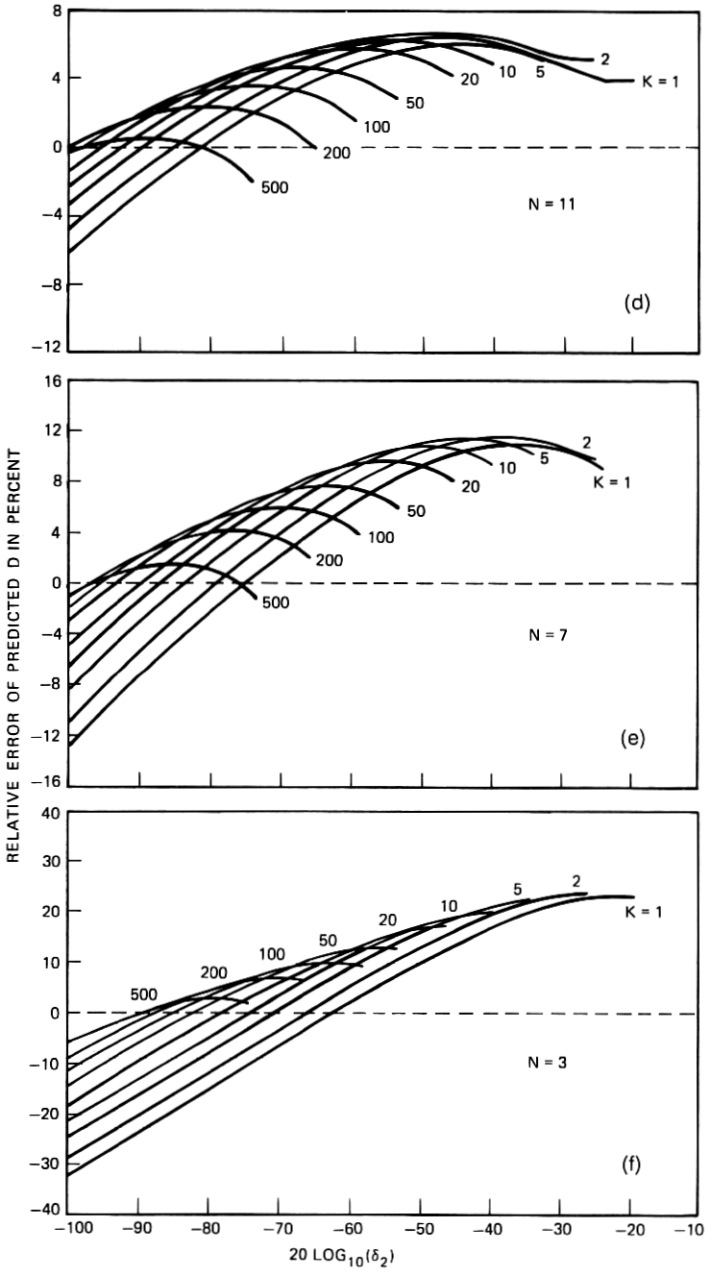

Fig. 12 (continued).

band ripples. The justification for such emphasis was the results shown in Figs. 2 and 3 which indicated that the extraripple solution for $N_p = N_s$ was fairly representative of a large region of the curve of ΔF versus F_p, except in the case where F_p was either very small or very large, in which case the Chebyshev solution became important. Also, as seen from Fig. 2, between extraripple solutions, the curve of ΔF versus F_p peaks up. However, it is seen that in many cases one can "approximately" bound the maximum between extraripple solutions by the next lower-degree extraripple solution. Since the design equations can predict ΔF for this case (the next lower-degree extraripple solution), a good bound on N can be obtained for a reasonably large region of the curve of ΔF versus F_p. In these cases, the value given by rule (*ii*) is good to within ± 4 in the worst cases, i.e., large values of K, and generally to within ± 2.

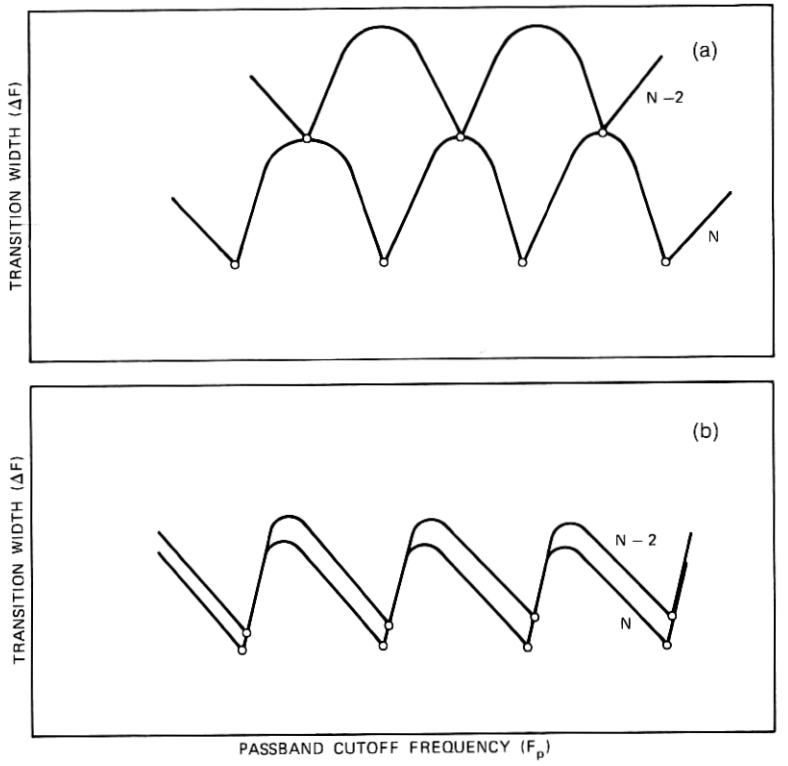

Fig. 13—Explanation of the types of behavior of curves of transition width versus passband cutoff frequency.

In the region of the Chebyshev solutions, however, such bounding procedures no longer are valid. An explanation of the difficulties which may be encountered is given in Fig. 13, which shows two types of curves of ΔF versus F_p. Fig. 13a shows the case, discussed above, where the extraripple solutions of impulse response duration $N - 2$ are approximately midway between extraripple solutions of impulse response duration N. In this case, the next degree solution bounds the maximum ΔF between extraripple solutions. Figure 13b shows the case where the extraripple solutions of impulse response duration $N - 2$ have approximately the same values for ΔF and F_p as the extraripple solutions of impulse response duration N. In these cases, there is no good bound on the maximum value of ΔF between adjacent extraripple solutions. The case of Fig. 13b corresponds to regions of F_p near 0.0 and F_s near 0.5, i.e., in the regions of the Chebyshev solutions. In these cases, as discussed in rule (*iv*), there is only an underbound on N, and no overbound.

The choice of a value of 0.04 in rule (*iii*) as the width of the region during which the behavior of N can only be underbounded was obtained from the data of Fig. 3 which shows that, beyond this region, the variation in the values of ΔF for extraripple solutions is small.

Figures 14 through 16 illustrate typical behavior of the curve of minimum value of impulse response duration N, to meet given specification on δ_1, δ_2, and ΔF as a function of F_p. Figure 14 shows data for the case $\delta_1 = 0.01$, $\delta_2 = 0.0001$, $\Delta F = 0.158$. The value of N_1 from

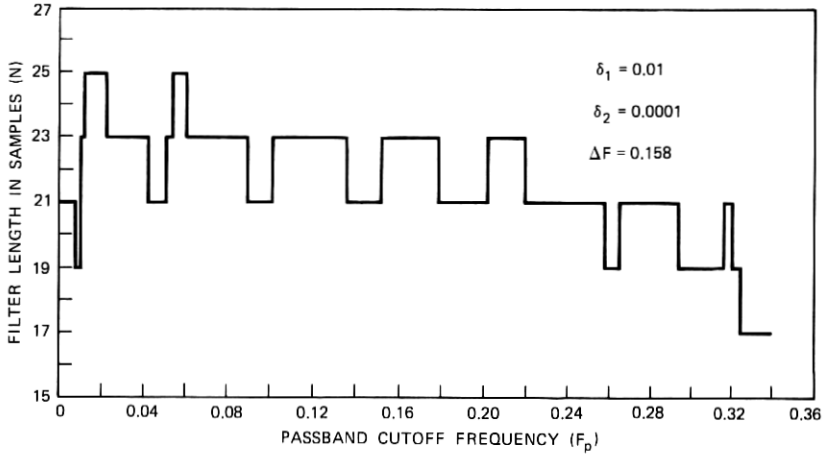

Fig. 14—Optimum values of N to meet given design specifications on δ_1, δ_2, and ΔF, as a function of F_p.

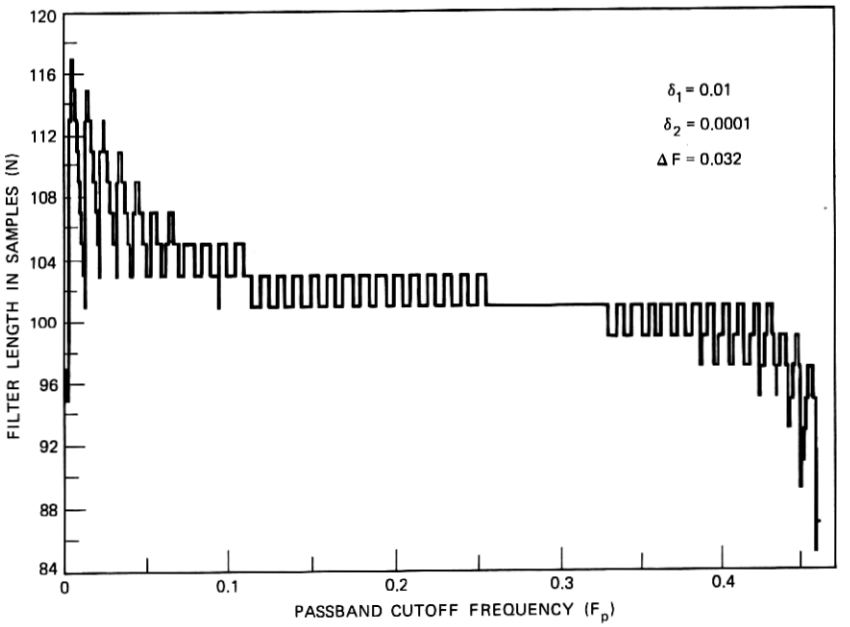

Fig. 15—Optimum values of N to meet given design specifications on δ_1, δ_2, and ΔF, as a function of F_p.

rule (*ii*) is 21, and the values of N_c are 19 (for small F_p) and 13 (for large F_p). Over the region $0.06 \leqq F_p \leqq 0.32$, N is within 2 of the nominal value of $N_1 = 21$. Over the entire range of F_p, N is within 4 of the value $N_1 = 21$. The data of Fig. 14 correspond to the case of Fig. 13a.

Figure 15 shows data for the case $\delta_1 = 0.01$, $\delta_2 = 0.0001$, $\Delta F = 0.032$. The value of N_1 from rule (*ii*) is 101, and the values of N_c are 95 (for small F_p) and 55 (for large F_p). Over the region $0.1 \leqq F_p \leqq 0.38$, N is within 2 of the nominal value of $N_1 = 101$. However, in the region $0 \leqq F_p \leqq 0.036$, the value of N fluctuates from a minimum of 95 (the Chebyshev lower bound) to a maximum of 117. The explanation of this erratic behavior of N is seen in Fig. 16. Figure 16a shows the data of Fig. 15 on an expanded horizontal scale, and Fig. 16b shows a plot of the approximate curves of ΔF versus F_p for all values of N from 95 to 117. The heavily traced parts of these curves show the lowest-order solution which just meets specifications on ΔF. From Fig. 16b, it is clear that in the vicinity of the first few extraripple solutions, the curves of ΔF versus F_p are exceedingly steep. Hence, a slight change in F_p greatly increases the required order solution to meet specifications. In

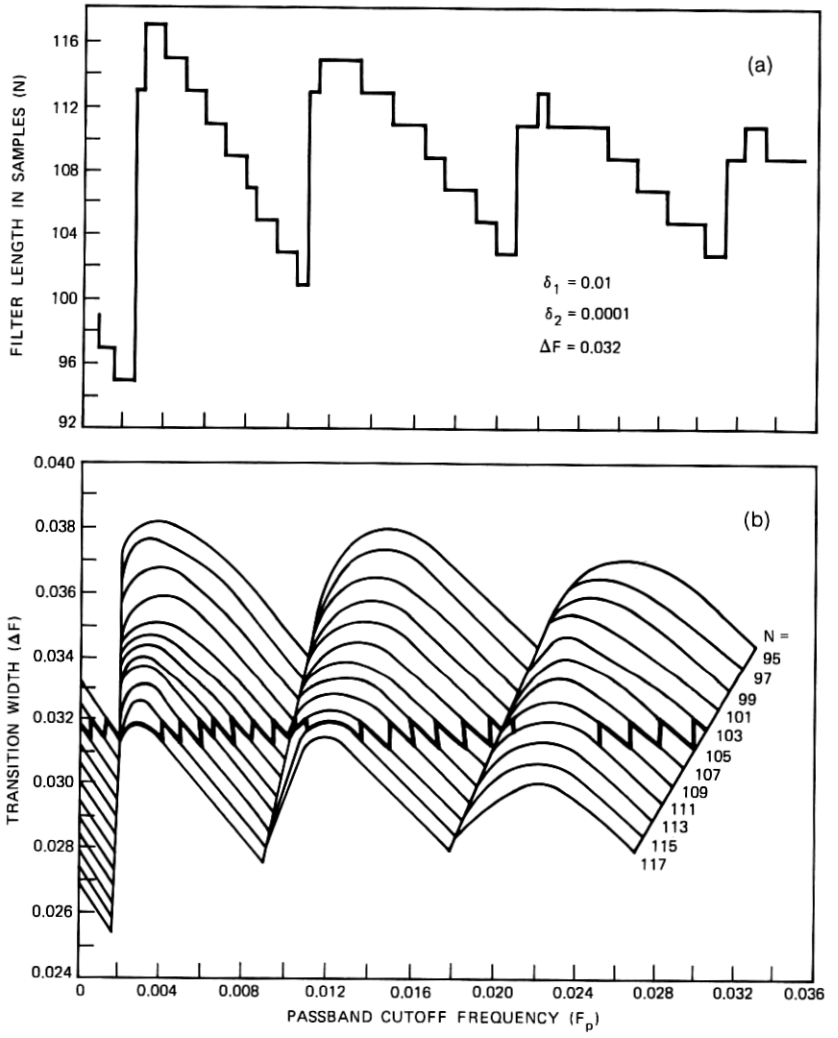

Fig. 16—Explanation of the behavior of the data of Fig. 15 in the region $0 \leqq F_p \leqq 0.036$.

these cases, it is impossible to estimate the exact filter impulse response duration which is required. Instead, only a lower bound can be given. Fortunately, as seen in Fig. 15, the regions in which this erratic behavior can occur are limited.

We conclude this section with a set of examples which illustrate the use of the design rules.

Example 1: Find the minimum value of N required to meet the specifications $\delta_1 = 0.05$, $\delta_2 = 0.0001$, $F_p = 0.19$, $F_s = 0.21$.

From rule (*ii*) we get $N_1 = 129.7$ which is rounded to 129. (Herein, all values of N will be rounded to the nearest odd integer.) Since F_p is far from the range for the Chebyshev solutions, the value of 129 is used as the appropriate estimate. The actual filter impulse response duration required is $N = 131$, although the $N = 129$ filter just missed meeting specifications.

Example 2: Find the minimum value of N required to meet the specifications $\delta_1 = 0.01$, $\delta_2 = 0.0001$, $F_p = 0.213$, $F_s = 0.373$.

From rule (*ii*) we get $N_1 = 19$. Since F_p is again out of the range of the Chebyshev solutions, the value 19 is used as the estimate of N. The actual filter impulse response duration required is $N = 19$.

Example 3: Find the minimum value of N required to meet the specifications $\delta_1 = 0.1$, $\delta_2 = 0.1$, $F_p = 0.12$, $F_s = 0.19$.

From rule (*ii*) we get $N_1 = 11$. Since F_p is out of the range of the Chebyshev solutions, the value 11 is used as the estimate of N. The actual value of N is 11. In this case, it is interesting to note that the value of N_c for the Chebyshev solution is also 11. This example points out that, for filter specifications leading to small values of N, there is very little variation in the actual value of N as F_p varies. This observation has been made earlier with respect to Fig. 14.

Example 4: Find the minimum value of N required to meet the specifications $\delta_1 = 0.01$, $\delta_2 = 0.0001$, $F_p = 0.36$, $F_s = 0.497$.

From rule (*ii*) we get $N_1 = 23$. Since F_s is within the bounds of rule (*iii*) (case 2), N_c is computed from rule (*iv*) as 13. The actual value of N is 19. In this case, the lower bound is within three filter orders of the true solution.

VII. SUMMARY

This paper has presented a wide variety of data on the relationships between design parameters for optimal low-pass FIR linear-phase digital filters. Analytical formulas were derived for the Chebyshev solutions, i.e., when there was only one passband or stopband ripple. Approximate fits to the data using nonlinear relationships between ΔF, N, δ_1, and δ_2 were given in the case where the number of passband and stopband ripples was equal, and it was argued that these relationships

were valid over a wide range of values of the filter parameters. Finally, a simple set of rules for estimating the minimum value of N, which meets given specifications on F_p, F_s, δ_1, and δ_2, was discussed. Examples were given to illustrate the application of the rules.

VIII. ACKNOWLEDGMENT

The authors would like to acknowledge the most helpful comments and criticisms provided by James Kaiser of Bell Laboratories.

REFERENCES

1. Rabiner, L. R., "Techniques for Designing Finite-Duration Impulse-Response Digital Filters," IEEE Trans. Commun. Tech., *COM-19* (April 1971), pp. 188–195.
2. Herrmann, O., "On the Design of Nonrecursive Digital Filters with Linear Phase," Elec. Letters, *6*, No. 11, 1970, pp. 328–329.
3. Hofstetter, E., Oppenheim, A. V., and Siegel, J., "A New Technique for the Design of Nonrecursive Digital Filters," Proc. Fifth Annual Princeton Conf. Inform. Sci. Syst., 1971, pp. 64–72.
4. Parks, T. W., and McClellan, J. H., "Chebyshev Approximation for Nonrecursive Digital Filters with Linear Phase," IEEE Trans. Circuit Theory, *DT-19* (March 1972), pp. 89–194.
5. Rabiner, L. R., "The Design of Finite Impulse Response Digital Filters Using Linear Programming Techniques," B.S.T.J., *51*, No. 6 (July-August 1972), pp. 1177–1198.
6. Parks, T. W., Rabiner, L. R., and McClellan, J. H., "On the Transition Width of Finite Impulse Response Digital Filters," IEEE Trans. Audio and Electroacoustics, *21*, No. 1 (February 1973), pp. 1–4.
7. Rabiner, L. R., and Herrmann, O., "The Predictability of Certain Optimum Finite Impulse Response Digital Filters," IEEE Trans. on Circuit Theory, *20*, No. 4 (July 1973), pp. 401–408.
8. Helms, H. D., "Nonrecursive Digital Filters: Design Methods for Achieving Specifications on Frequency Response," IEEE Trans. Audio and Electroacoustics, *16*, No. 3 (September 1968), pp. 336–342.
9. Heyliger, G. E., "Simple Design Parameters for Chebyshev Arrays and Filters," IEEE Trans. Audio and Electroacoustics, *18*, No. 4 (December 1970), pp. 502–503.
10. Kaiser, J., "Digital Filters," Chapter 7 in *System Analysis by Digital Computer*, edited by F. F. Kuo and J. F. Kaiser, New York: J. Wiley & Sons, 1966, pp. 230–238.