THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 53, No. 2, February 1974
Printed in U.S.A.

Some Comparisons Between FIR and IIR Digital Filters

By L. R. RABINER, J. F. KAISER, O. HERRMANN, and M. T. DOLAN

(Manuscript received July 27, 1973)

The purpose of this paper is to make comparisons between optimum, linear phase, finite impulse response (FIR) digital filters and infinite impulse response (IIR) digital filters which meet equivalent frequency domain specifications. The basis of comparison is, for the most part, the number of multiplications per sample required in the usual realizations of these filters—i.e., the cascade form for IIR filters, and the direct form for FIR filters. Comparisons are also made between group-delay equalized filters and linear phase FIR filters. Considerations dealing with finite word-length effects are discussed for both these filter types. A set of design charts is also presented for determining the minimum filter order required to meet given low-pass filter specifications for both digital and analog filters.

I. INTRODUCTION

Although a great deal is known about the properties of different types of digital filters, very little has been done to relate the various designs as to performance and complexity of realization. Thus the filter designer must learn the details of several design procedures before being able to make a wise decision on a suitable filter for his specific application. It is the purpose of this paper to add insight into some of the problems that have been encountered by filter designers by: (*i*) presenting new and useful design curves for digital and analog low-pass filters, and (*ii*) making several comparisons between optimum (quasi-equiripple), linear phase, FIR low-pass filters and equiripple (elliptic) IIR filters which meet equivalent frequency domain specifications. Although these results are presented for low-pass filter designs, they are easily extended to the case of bandpass, bandstop, and high-pass filters by the well-known frequency band transformations.

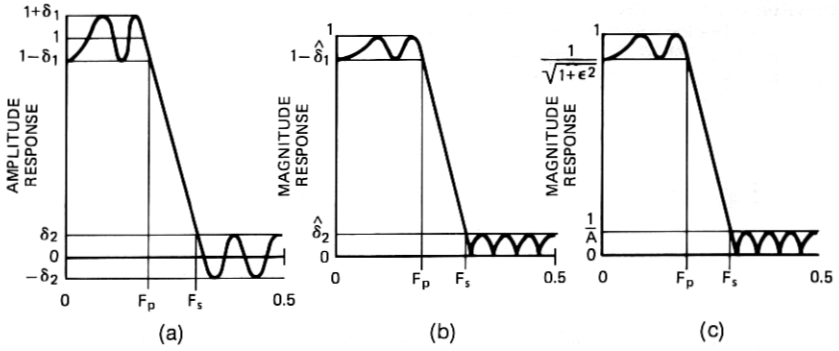

Fig. 1—Terminology used to describe low-pass filter characteristics.

The organization of the paper is as follows. After defining the terminology to be used, the design relationships between the FIR filter parameters are reviewed. The design relationships for the IIR filter parameters are developed and a novel graphical interpretation of these relations is presented in the form of useful filter design charts (applicable to both digital and analog filters). Using the design relationships, the filter orders required to achieve equivalent performance are compared for different ranges of filter parameters.

II. TERMINOLOGY

The design procedures for the two general classes of digital filters, FIR and IIR, have essentially progressed along independent paths. As a result, the terminology used in specifying the filter performance is generally not quite the same. Thus it is instructive to define the most commonly accepted definitions of the filter parameters for these two classes of filters. The relations between these parameter sets for equivalence are then established. Figure 1a shows the amplitude response of a typical optimum FIR low-pass filter and Fig. 1b shows the magnitude response of a typical elliptic low-pass filter. For the FIR case the amplitude response in the passband ($0 \leqq f \leqq F_p$)* generally oscillates between $1 + \delta_1$ and $1 - \delta_1$, where δ_1 is the passband ripple. In the stopband ($F_s \leqq f \leqq 0.5$) the amplitude response oscillates between $+\delta_2$ and $-\delta_2$ where δ_2 is the stopband ripple. For the elliptic case the magnitude response is constrained to always be less

* Throughout this paper the frequency scale has been normalized with respect to the sampling frequency. Thus the normalized sampling frequency is 1.0 and the frequency range graphed is $0 \leqq f \leqq 0.5$ or, equivalently, $0 \leqq \omega \leqq \pi$.

than 1.0. Thus, in the passband ($0 \leqq f \leqq F_p$) the magnitude response oscillates between 1 and $1 - \hat{\delta}_1$. In the stopband the magnitude response oscillates between $\hat{\delta}_2$ and 0. It is straightforward to relate δ_1, δ_2, $\hat{\delta}_1$, and $\hat{\delta}_2$ so the resulting magnitude characteristics are equivalent. If the FIR amplitude characteristic is scaled by $1/(1 + \delta_1)$ and the magnitude of the resulting amplitude response is taken, then the following relationships are obtained:

$$\hat{\delta}_1 = \frac{2\delta_1}{1 + \delta_1} \tag{1}$$

$$\hat{\delta}_2 = \frac{\delta_2}{1 + \delta_1} \tag{2}$$

$$\delta_1 = \frac{\hat{\delta}_1}{2 - \hat{\delta}_1} \tag{3}$$

$$\delta_2 = \frac{2\hat{\delta}_2}{2 - \hat{\delta}_1}. \tag{4}$$

Thus, given either (δ_1, δ_2) or ($\hat{\delta}_1$, $\hat{\delta}_2$), eqs. (1) through (4) can be used to find the equivalent specifications for the other type of filter.

Although the notation of Fig. 1b is acceptable for the magnitude response of an elliptic filter, it is not the most widely used form for these filters. Figure 1c shows the same magnitude response described in terms of passband parameter ϵ and stopband parameter A. Comparing Figs. 1b and 1c it is easy to relate ϵ and A to $\hat{\delta}_1$ and $\hat{\delta}_2$ as

$$\epsilon = \frac{\sqrt{2 - \hat{\delta}_1}\sqrt{\hat{\delta}_1}}{(1 - \hat{\delta}_1)} \tag{5}$$

$$A = \frac{1}{\hat{\delta}_2}. \tag{6}$$

At this point it is convenient to define the additional filter terms E, ATT, and η as

$$E = \text{(in-band) ripple} = 20 \log_{10} \sqrt{(1 + \epsilon^2)} \tag{7}$$

$$\text{ATT} = \text{stopband attenuation} = 20 \log_{10} A \tag{8}$$

$$\eta = \frac{\epsilon}{\sqrt{A^2 - 1}} = \frac{\hat{\delta}_2\sqrt{\hat{\delta}_1}\sqrt{2 - \hat{\delta}_1}}{(1 - \hat{\delta}_1)\sqrt{1 - \hat{\delta}_2^2}} = \frac{2\delta_2(\sqrt{\delta_1})}{(1 - \delta_1)\sqrt{(1 + \delta_1)^2 - \delta_2^2}}. \tag{9}$$

Thus, parameters E and ATT are a third set of parameters which describe the characteristics of the magnitude response of the elliptic

filter. The parameter η has been shown to be a basic analog filter parameter[1] which will be used in the filter design curves given in a later section.

III. FIR DESIGN RELATIONS

The five basic FIR filter parameters are F_p, F_s, δ_1, δ_2, and N, the duration of the filter impulse response in samples. For the general case of optimum, linear phase, low-pass FIR filters, there exist no simple analytical relationships between these five filter parameters, except in special cases, e.g., one passband or one stopband ripple. However, an approximate empirical relationship between the filter parameters has recently been obtained[2] which accurately satisfies known design results for a wide range of values of the filter parameters. The relationship is of the form:

$$N = 1 + \frac{D_\infty(\delta_1, \delta_2)}{\Delta F} - f(\delta_1, \delta_2)\Delta F, \tag{10}$$

where

$$\Delta F = F_s - F_p = \text{relative transition width}, \tag{11}$$

$$D_\infty(\delta_1, \delta_2) = [0.005309\ (\log_{10}\delta_1)^2 + 0.07114\log_{10}\delta_1 - 0.4761]\log_{10}\delta_2 - [0.00266\ (\log_{10}\delta_1)^2 + 0.5941\log_{10}\delta_1 + 0.4278], \tag{12}$$

and

$$f(\delta_1, \delta_2) = 0.51244\log_{10}(\delta_1/\delta_2) + 11.01. \tag{13}$$

Equation (10) can generally predict the value of N required to meet specifications on δ_1, δ_2, F_p, and F_s to within ± 2. In the cases where F_p is very close to 0, or F_s is very close to 0.5, eq. (10) tends to overestimate the required N. It should be noted that eq. (10) shows the estimate of N to be independent of specific values of F_s or F_p, but instead is dependent only on the transition width, $(F_s - F_p)$.

A simpler expression giving a less accurate estimate of N is

$$N = \frac{-10\log_{10}(\delta_1\cdot\delta_2) - 15}{14\Delta F} + 1. \tag{14}$$

This expression is a modification of the design relationship for FIR filters designed by windowing techniques (where $\delta_1 = \delta_2$). See Ref. 3, pp. 237–238.

IV. IIR DESIGN RELATIONS

One of the most general procedures for designing IIR digital filters is through the bilinear transformation of an appropriate continuous

filter. There are two equivalent techniques for obtaining the desired digital filter using the bilinear transformation and these are illustrated in Fig. 2. The technique of Fig. 2a begins with an analog low-pass filter with normalized passband cutoff frequency of 1 radian per second, and analog stopband cutoff frequency of Ω_s radians per second. This filter is bilinearly transformed[4] to give a "normalized" digital filter with passband cutoff frequency $\pi/2$ radians per second (or $f = 0.25$ on the normalized scale) and stopband cutoff frequency $\hat{\omega}_s$. The relation between the frequency variables Ω and $\hat{\omega}$ is given by

$$\Omega = \tan(\hat{\omega}/2). \tag{15}$$

Thus Ω_s and $\hat{\omega}_s$ are simply related by eq. (15) with $\Omega = \Omega_s$, and $\hat{\omega} = \hat{\omega}_s$. Finally, a digital all-pass transformation[5] is used to give the desired digital low-pass filter with passband cutoff frequency ω_p and stopband cutoff frequency ω_s. The relation between the frequency variables ω and $\hat{\omega}$ is given by

$$\tan\omega = \frac{(1-\alpha^2)\sin\hat{\omega}}{(1+\alpha^2)\cos\hat{\omega} - 2\alpha}, \tag{16}$$

where

$$\alpha = \frac{\sin(\omega_p/2) - \cos(\omega_p/2)}{\sin(\omega_p/2) + \cos(\omega_p/2)}. \tag{17}$$

The second technique (shown in Fig. 2b) begins with the identical analog low-pass filter as in the first technique and immediately performs a low-pass–to–low-pass transformation to give an analog filter with passband cutoff frequency $\hat{\Omega}_p$ and stopband cutoff frequency $\hat{\Omega}_s$. The relation between the frequency variables Ω and $\hat{\Omega}$ is

$$\Omega = \hat{\Omega}/\hat{\Omega}_p. \tag{18}$$

The resulting low-pass filter is then transformed to a digital filter using the bilinear transformation, giving the same end result as in the first technique (Fig. 2a). The relation between frequency variables ω and $\hat{\Omega}$ is

$$\hat{\Omega} = \tan(\omega/2). \tag{19}$$

In the case where the prototype normalized analog filter is an elliptic filter, it is relatively easy to derive a design formula relating the various filter parameters, both in the analog and digital cases. For either the analog or digital case the order, n, of the elliptic filter is related to the remaining filter parameters by the equation

$$n = \frac{K(k)K(\sqrt{1-k_1^2})}{K(k_1)K(\sqrt{1-k^2})}, \tag{20}$$

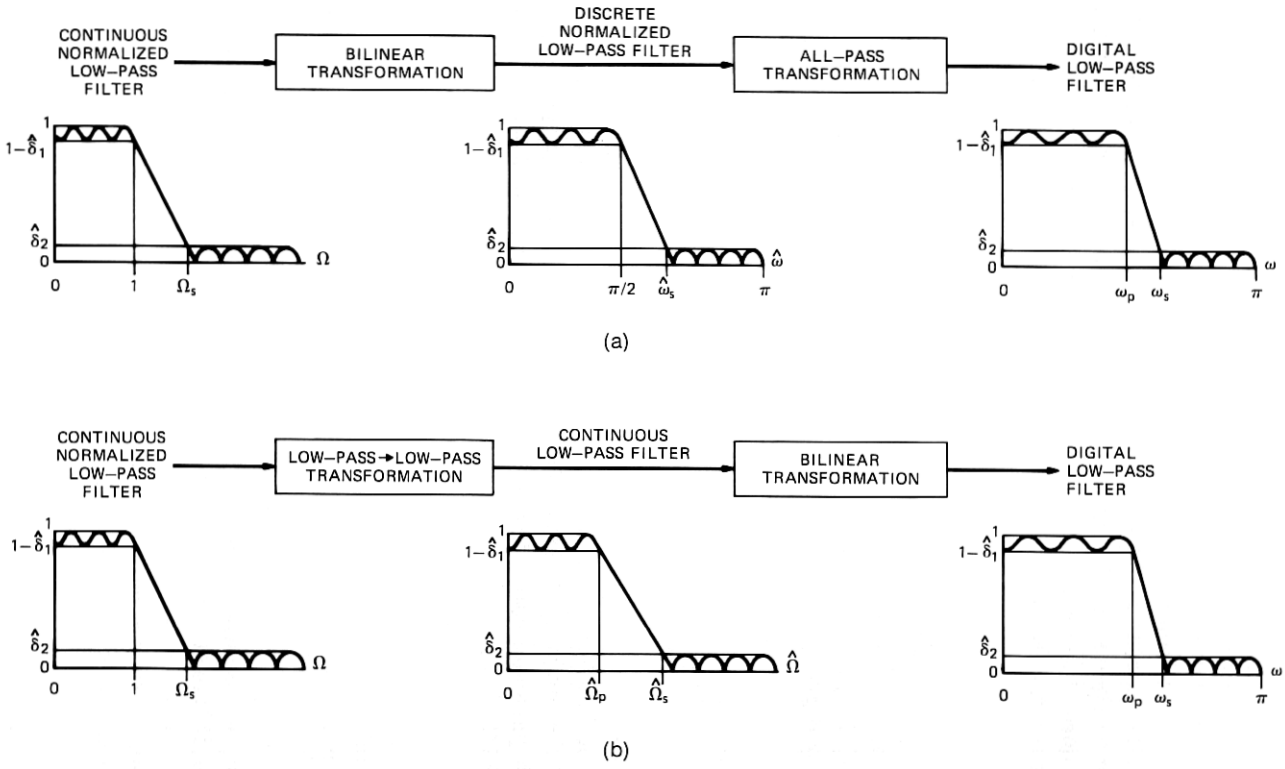

Fig. 2—Two techniques for transforming a continuous normalized low-pass filter to a digital low-pass filter.

where $K(\cdot)$ is the complete elliptic integral of the first kind and

$$k = \text{transition ratio} = \frac{1}{\Omega_s} = \frac{\tan(\omega_p/2)}{\tan(\omega_s/2)} \tag{21}$$

and

$$k_1 = \eta = \frac{\epsilon}{\sqrt{A^2 - 1}} = \frac{2\delta_2\sqrt{\delta_1}}{(1 - \delta_1)\sqrt{(1 + \delta_1)^2 - \delta_2^2}}. \tag{22}$$

Thus eq. (20) relates filter order, n, to the parameters F_p, F_s [through eq. (21)] and δ_1 and δ_2 [through eq. (22)].

For the case when the prototype filter is a Chebyshev filter (either type I—equiripple passband, monotone stopband, or type II—maximally flat passband, equiripple stopband) the design equation becomes

$$n = \frac{\cosh^{-1}(1/\eta)}{\ln \beta}, \tag{23}$$

where

$$\beta = \frac{1 + \sqrt{1 - k^2}}{k} \tag{24}$$

and η and k are defined as eqs. (21) and (22). Finally for a prototype Butterworth filter (maximally-flat magnitude, all pole) the design equation is

$$n = \frac{\ln \eta}{\ln k}. \tag{25}$$

Although eqs. (20) through (25) completely describe the design curves for both analog and digital filters, it is generally quite helpful to see the relationships between filter parameters displayed in a meaningful way. Since, in general, there are five filter parameters there is no simple way of presenting these relationships on a single plot, even in terms of well-known nomograph procedures.[6] There is, however, a simple and straightforward way of including all design relations for both digital and analog filters, for any prototype filter, using a sequence of three charts.

The first chart(s) relates the filter design parameter η to the passband and stopband ripple specifications δ_1 and δ_2 or their equivalents. The second chart(s) graphs the filter design equation relating filter order n, design parameter η, and transition ratio k. The third chart(s) relates transition ratio k to passband cutoff frequency F_p and transition bandwidth ν.

Figures 3a through 3d show four possibilities for Chart No. 1. The graphs of Figs. 3a and 3b correspond to digital filters with δ_1 as a param-

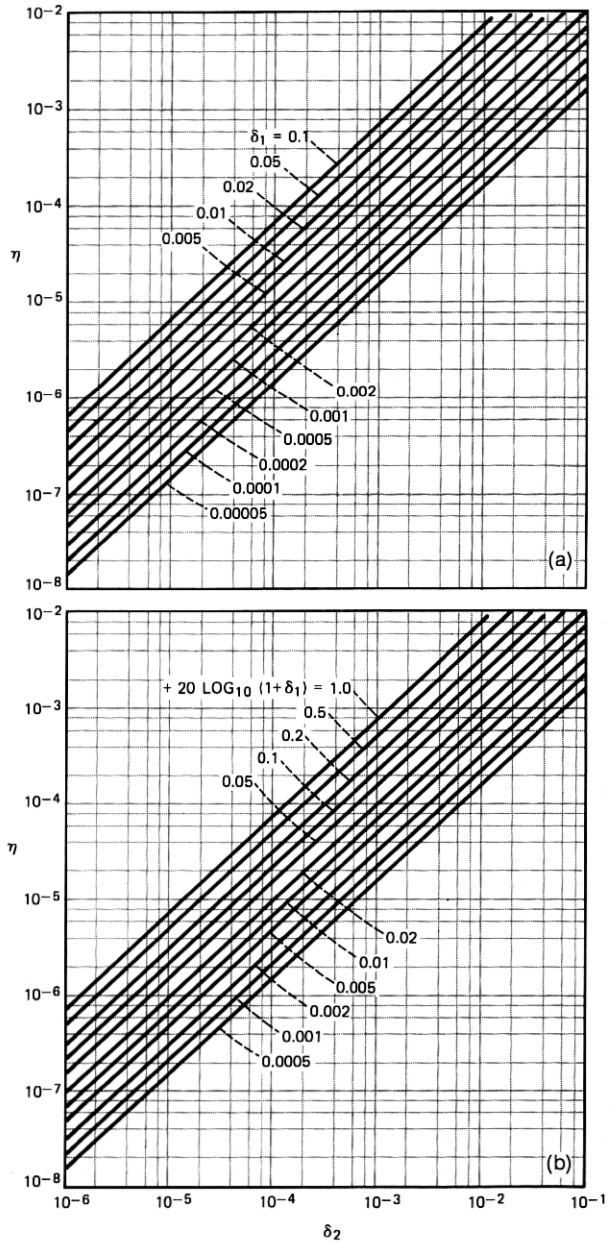

Fig. 3—Plots of η versus stopband specification, with parameter passband specification for low-pass filters.

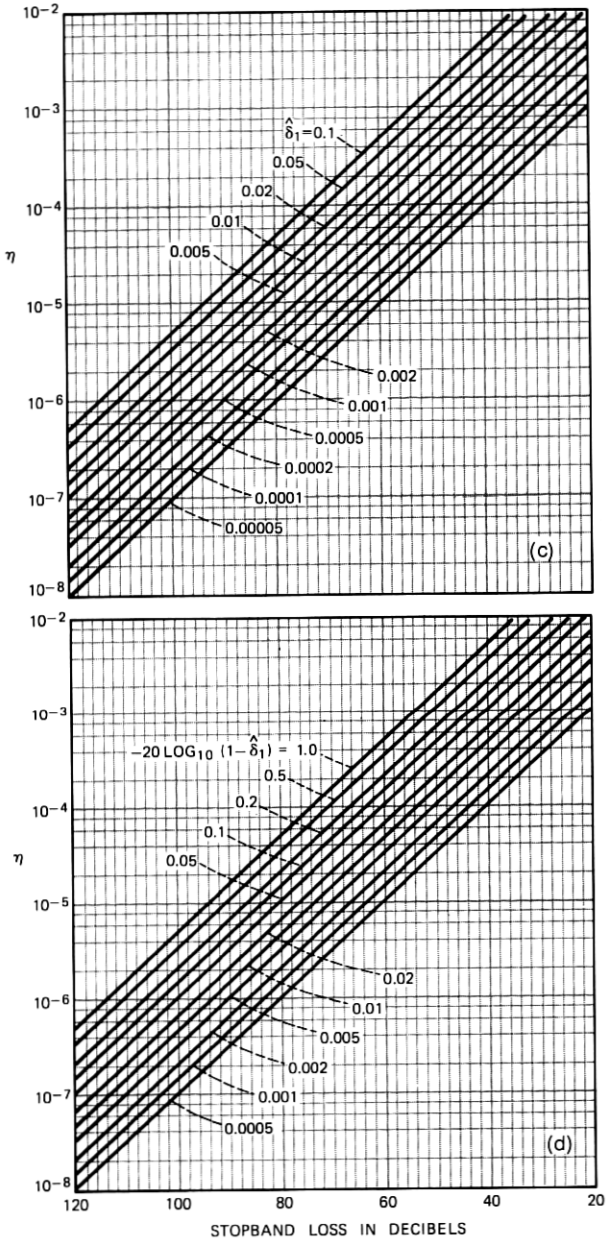

Fig. 3 (continued).

eter (Fig. 3a) or $20 \log_{10} (1 + \delta_1)$(dB) as a parameter (Fig. 3b). The graphs of Figs. 3c and 3d correspond to analog filters with absolute ripple $\hat{\delta}_1$ as a parameter (Fig. 3c) or total ripple $20 \log_{10} [1/(1 - \hat{\delta}_1)]$(dB) as a parameter (Fig. 3d).

Chart No. 2 represents the design relations particular to the prototype filters, i.e., eq. (20) for elliptic filters, eq. (23) for Chebyshev filters, and eq. (25) for Butterworth filters. For these graphs the parameter η is plotted versus transition ratio, k, with filter order, n, as the parameter. Figures 4a through 4c show the resulting graphs for elliptic filters, Chebyshev filters, and Butterworth filters, respectively. The horizontal scale on each of these graphs is a nonuniform scale which was chosen to provide a reasonably good spacing of the curves for the various values of n. The actual nonlinear scale used is represented by the equation

$$x = \frac{k + k^8}{2}, \tag{26}$$

where x is the x-axis coordinate ($0 \leqq x \leqq 1$) and k is the transition width. Thus, the scale is linear for small values of k and highly nonlinear near $k = 1.0$.

Chart No. 3 represents the relation between the transition ratio and the filter cutoff frequencies [eq. (21)]. For these graphs the passband cutoff frequency, F_p, is plotted versus transition ratio, k, for various values of normalized transition width, ν, defined as

$$\nu = F_s - F_p = \frac{\omega_s - \omega_p}{2\pi}. \tag{27}$$

Figures 5a and 5b show the resulting graphs for digital and analog filters. The scale for transition ratio is identical to the scale used for Chart No. 2.

V. USE OF CHARTS

To illustrate how to use the set of charts of Figs. 3 through 5, consider the determination of filter order n required to meet the following specifications:

$\delta_1 = 0.01$ ($\approx \pm 0.086$-dB passband ripple)
$\delta_2 = 0.0001$ (80-dB stopband loss)
passband cutoff frequency = 480 Hz
stopband edge frequency = 520 Hz
sampling frequency = 8000 Hz.

Normalizing the band-edge frequencies gives

$$F_p = \frac{480}{8000} = 0.06$$

$$F_s = \frac{520}{8000} = 0.065.$$

For the determination of filter order n for a digital filter of the elliptic type, the charts of Figs. 3a, 4a, and 5a are used (Fig. 4a specializes the design to the elliptic type). To obtain the value of η on Fig. 3a, we use the curve $\delta_1 = 0.01$ and find its intersection with the line $\delta_2 = 0.0001$ which yields a value of η approximately equal to 2×10^{-5}. To obtain the transition ratio, we use Fig. 5a by finding the intersection of the curve $\nu = F_s - F_p = 0.005$ with line $F_p = 0.06$; this yields a value of 0.923 for the transition ratio (this agrees nicely with $F_p/F_s = 0.06/0.065 = 0.923$, an alternate way of arriving at the same result). Finally, the filter order, n, can now be determined from Fig. 4a by finding the intersection of the lines $\eta = 2 \times 10^{-5}$ and transition ratio = 0.923; thus the required theoretical elliptic filter order is ≈ 11.5. In order to meet specifications on all four parameters, a 12th-order filter must be used.

However, there are several tradeoffs possible for the final filter specifications. For example, if η is held fixed at 2×10^{-5} and the transition ratio is changed to approximately 0.94 to lie on the $n = 12$ curve, then either F_s or F_p can be varied to match this new value of transition ratio. The tradeoffs here are obtained from Fig. 5a. If the transition ratio is held fixed, then for $n = 12$ we find η is $\approx 1.0 \times 10^{-5}$; from Chart No. 1 (Fig. 3a) we can observe the tradeoff as δ_1 and δ_2 are varied for this new value of η. Finally, both transition ratio and η can be varied, e.g., to 0.93 for transition ratio and 1.5×10^{-5} for η, so as to make their intersection remain on the $n = 12$ curve; now all four filter parameters can be varied to match the new values of η and transition ratio.

It is interesting to note that if a Chebyshev or Butterworth filter type is specified in place of the elliptic, the designer need only substitute Figs. 4b or 4c for Fig. 4a as Chart No. 2 and proceed as before. In both cases of the example given, the required filter order considerably exceeds the maximum limit of 20 of the curves; thus the "efficiency" of the elliptic design is clearly seen.

Clearly, this design procedure presents a tremendous amount of flexibility to the designer—more so than is generally available in most

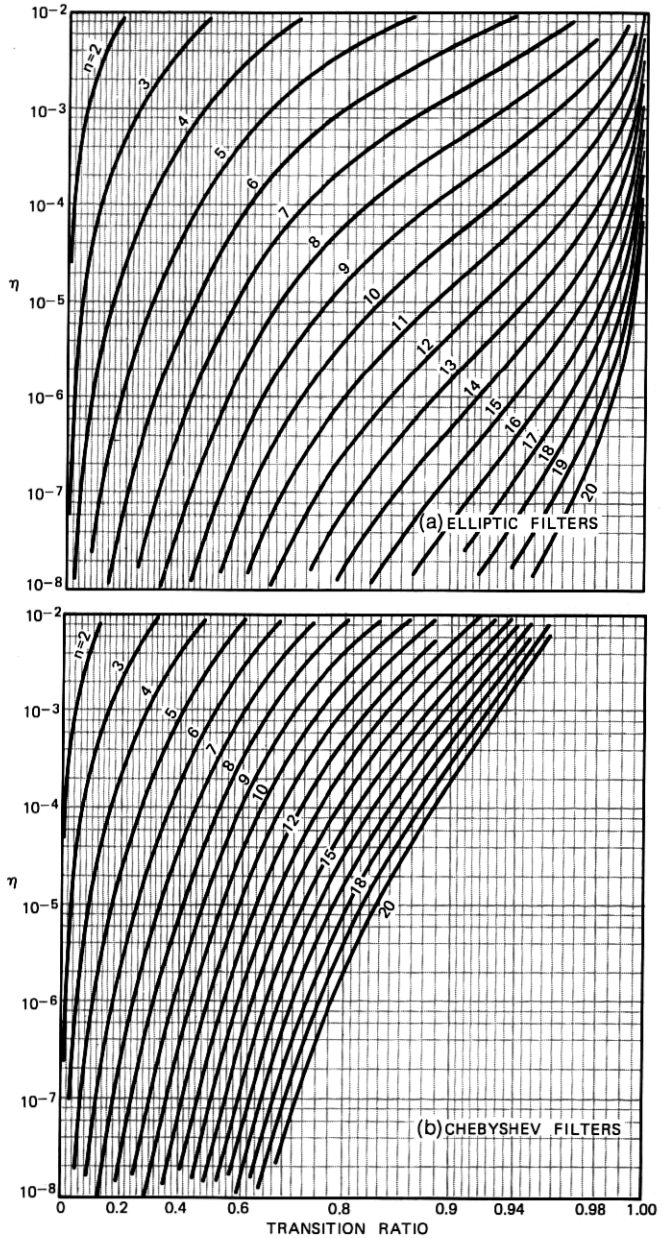

Fig. 4—Plots of η versus transition ratio as a function of filter order n for elliptic, Chebyshev, and Butterworth low-pass filters.

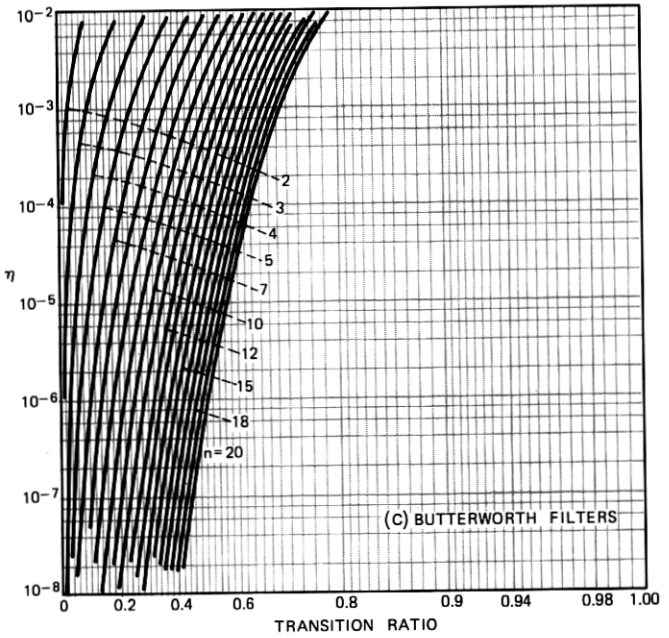

Fig. 4 (continued).

programs for filter order determination. *Furthermore, the insight into the design problem afforded by this graphical technique allows the designer to get a feeling for the way in which small changes in filter specification affect the required filter order.* Quite often the designer is willing to changes his ideas on "required" specifications, especially if it reduces the filter order necessary to meet his specifications.

VI. COMPARISONS BETWEEN OPTIMUM FIR AND ELLIPTIC DIGITAL FILTERS

Based on the design formulas of the preceding sections, it is possible to make some comparisons between optimum FIR low-pass filters and equivalent elliptic filters. The main basis of comparison will be the number of multiplications per input sample* required in the most standard realization of each filter type, i.e., the direct form for the FIR case and the cascade form for the elliptic case.[7] Direct realization of an N-point impulse response filter with linear phase requires

* The number of multiplications per input sample is a useful measure of the computational complexity of the filtering operations as it represents the number of multiply-add operations required for a software implementation of the algorithm as well as for a general hardware implementation.

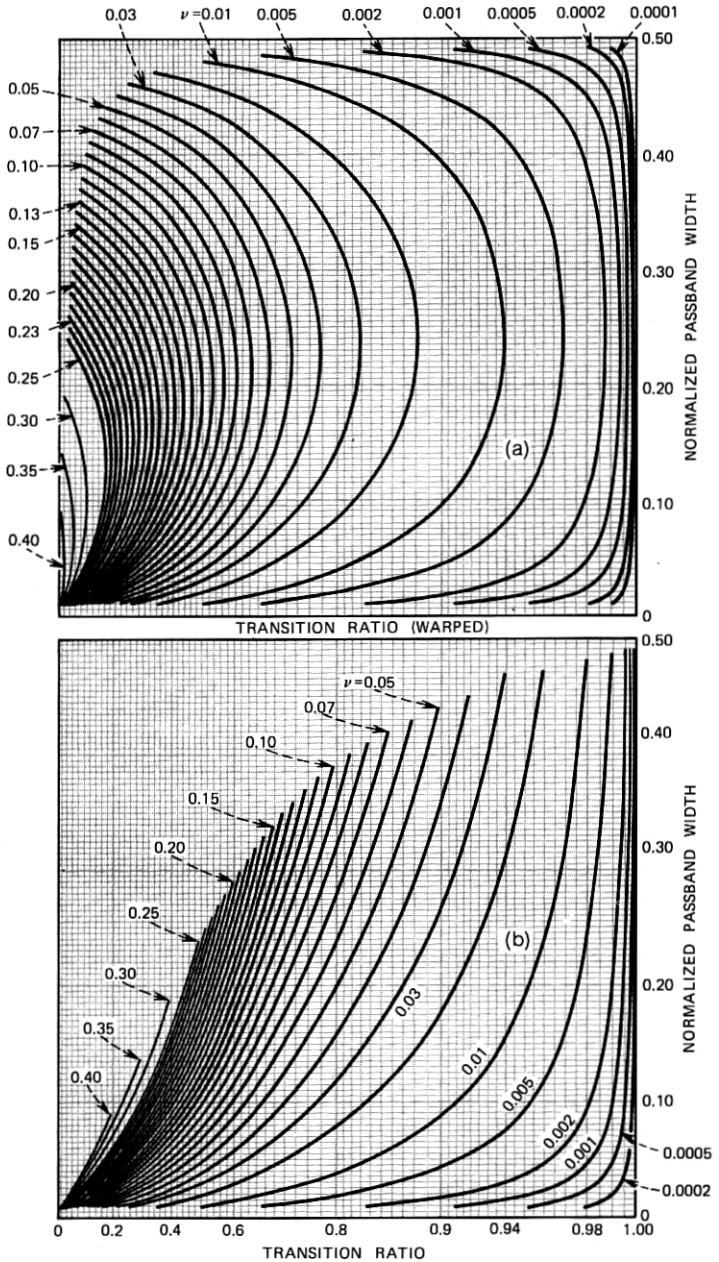

Fig. 5—Plots of passband cutoff frequency versus transition ratio as a function of transition width for discrete and continuous low-pass filters.

$[(N + 1)/2]$ multiplications per sample, whereas cascade realization of an nth-order elliptic filter (all zeros on the unit circle) requires $[(3n + 3)/2]$* multiplications per sample where $[\cdot]$ denotes "integer part of."

Thus, one basis of comparison between equivalent filter designs (i.e., both meeting the same specifications on δ_1, δ_2, F_p, and F_s) is in terms of the efficiency of the respective realizations, i.e., which structure requires fewer multiplications per sample. Equivalence between structures is attained when the condition

$$\left[\frac{3n + 3}{2}\right] = \left[\frac{(N + 1)}{2}\right] \tag{28}$$

or equivalently

$$\frac{N}{n} \approx 3 + \frac{1}{n}. \tag{29}$$

Using the appropriate filter design formulas, we have measured the quantity N/n as a function of n for a large range of values of F_p, δ_1, and δ_2. Figure 6 shows two typical sets of curves which were obtained. Figure 6a shows data for the case $F_p = 0.15$, $\delta_1 = 0.1$, $\delta_2 = 0.1$, 0.01, 0.001, 0.0001, and Fig. 6b shows data for $F_p = 0.35$, $\delta_1 = 0.00001$, and the same range of δ_2 as in Fig. 6a. Also shown in these plots is the line $N/n = 3$ for showing where the data lie with respect to the fixed portion of eq. (29). As seen in this figure, for certain values of F_p, δ_1, and δ_2, the ratio of N/n falls below the equivalence level of eq. (29), i.e., the FIR filter is more efficient than the elliptic filter. However, in general, the elliptic filter is more efficient than the optimum FIR filter, and, in the case of high-order elliptic designs, the ratio of N/n is often in the hundreds or thousands.

Based on our examination of large amounts of data, the following general observation can be made: the most favorable conditions for the FIR design are large values of δ_1, small values of δ_2, and large transition widths (i.e., small transition ratios). One also observes the following behavior:

(*i*) For values of $F_p \geqq 0.3$, the ratio N/n always exceeded $3 + 1/n$ for all values of δ_1, δ_2, and n.

(*ii*) For values of $n \geqq 7$, the ratio N/n always exceeded $3 + 1/n$ for all values of δ_1, δ_2, and F_p.

---

*This number of multiplications per sample for the IIR filter assumes that any scaling between sections is an integer power of 2 and is performed entirely by shifts of the data. If finer scaling multipliers are included between each cascade section, the realization requires $[(4n + 3)/2]$ multiplications per sample.

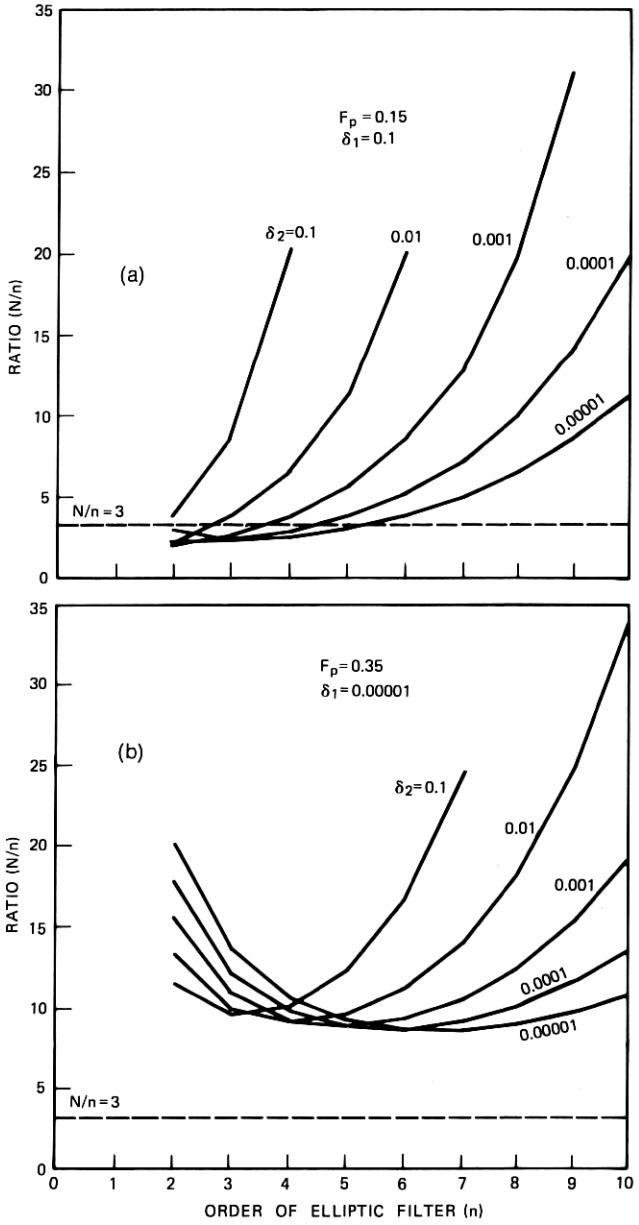

Fig. 6—Plots of the ratio N/n as a function of n for optimum FIR filters and elliptic filters meeting identical specifications on δ_1, δ_2, F_p, and F_s.

(*iii*) The smaller the value of F_p, the larger the range of δ_1, δ_2, and n for which N/n was less than $3 + 1/n$.

Since the design formula for N for the optimum FIR case is not exact but only an estimate, measurements were also made of the required theoretical value of n (elliptical filter order) to meet the specifications of optimum FIR filters which had already been designed. Typical results of these measurements are shown in Fig. 7. Figure 7a shows the theoretical order n (n need not be an integer) required to match specifications on F_p, F_s, for $\delta_1 = 0.1$, $\delta_2 = 0.1$, 0.01, 0.001, 0.0001, and 0.00001, as a function of F_p for a set of optimum FIR filters with $N = 21$. (It should be noted that as F_p varies, F_s also varies so as to achieve the desired specifications on δ_1 and δ_2.) Figure 7b shows similar measurements for $N = 41$. In Fig. 7a the theoretical point of equivalence is $n = 6.3$, whereas in Fig. 7b it is $n = 13$. From this figure it is seen that for these cases the elliptic filter is always more efficient than the equivalent FIR filter, as anticipated by the discussion in the preceding paragraphs.

In summary, elliptic filters are generally more efficient in achieving given specifications on the frequency response than optimum FIR filters. However, the FIR filters have the additional useful property that their phase is exactly linear, i.e., there is no group delay distortion. For the elliptic filter, however, there is generally a large amount of group delay distortion (concentrated primarily near the band edge). A question of both theoretical and practical importance is whether, in cases when the additional requirement of a flat delay is specified, it is more desirable to equalize the delay of an elliptic filter or to use the equivalent optimum FIR filter (with its constant group delay). In the next section we discuss various aspects of this question. It should be noted that the above alternatives are not the only possibilities for obtaining a digital filter which meets frequency domain specifications on both magnitude and group delay responses. For example, a filter can be designed, using modern optimization procedures, where the number of poles and zeros are unequal. In such cases, the comparisons between FIR and IIR filters are quite distinct from those to be discussed in the next section.

VII. COMPARISONS OF OPTIMUM FIR FILTERS AND DELAY-EQUALIZED ELLIPTIC FILTERS

Recently developed optimization procedures[8] make it possible to design an all-pass equalizer which can equalize the group delay of any

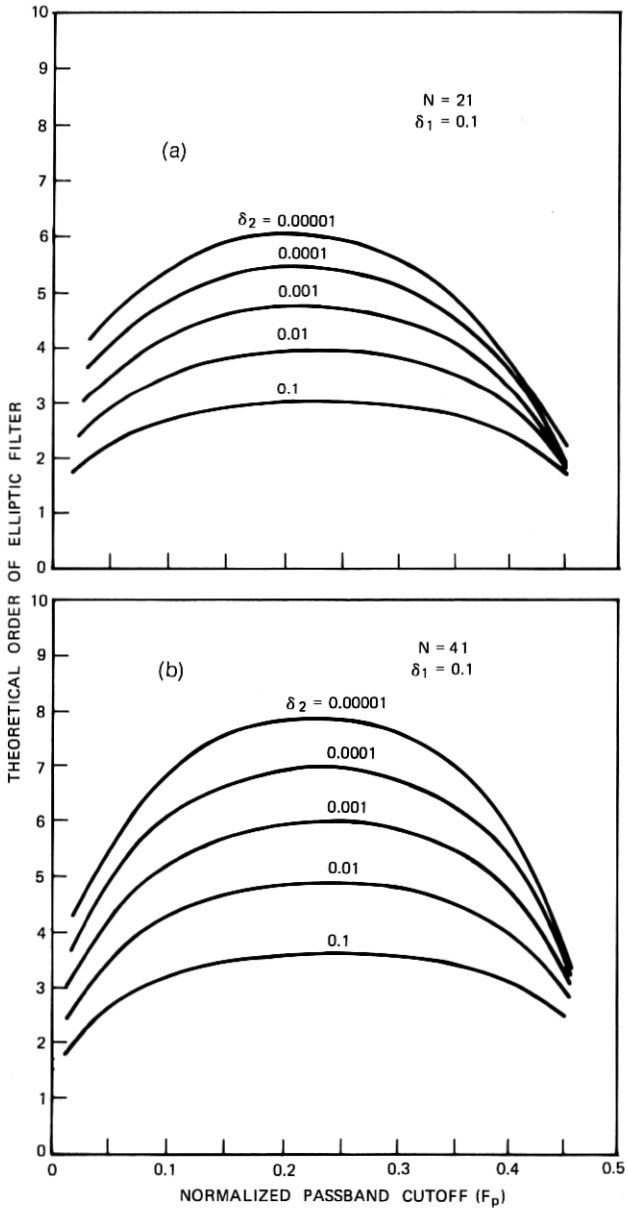

Fig. 7—Theoretical order of elliptic filters required to meet given specification on δ_1, δ_2, F_p, and F_s as a function of F_p for various values of δ_2. Optimum FIR filters with $N = 21$ meet the specifications for all filters of (a), whereas $N = 41$ is required for all filters of (b).

digital filter to any desired accuracy over a restricted band of frequencies. As an example of the use of this procedure, Fig. 8 shows plots of the group delay of a 6th-order (unequalized) elliptic filter (with parameters $\delta_1 = 0.01$, $\delta_2 = 0.0001$, $F_p = 0.24163$, $F_s = 0.34842$) and the equalized group delay using a 10th-order all-pass filter. The relative error in the equalized delay curve is 3.6 percent of the average delay in the passband. In this case the equalized elliptic filter requires 20 multiplications per sample, whereas an optimum FIR filter which achieves the same specifications requires only 11 multiplications per sample.

The difficulty with trying to equalize the group delay of a filter lies in the fact that the equalized filter must have a total delay greater than the largest delay in the unequalized filter which always occurs near the passband cutoff frequency. Thus, in the example of Fig. 8, even though the delay throughout most of the passband is between 2 and 6 samples, the delay at the edge of the band is about 15 samples. It can be shown that an all-pass equalizer of degree n_e has the

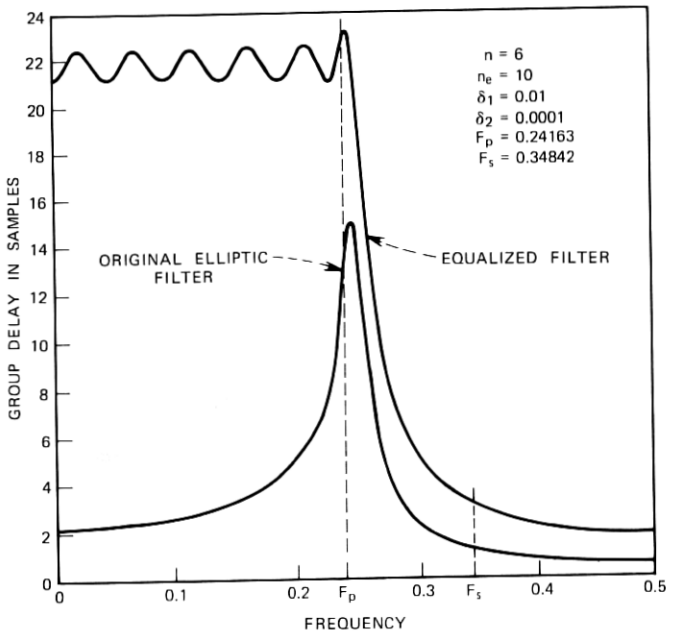

Fig. 8—The group delay of an unequalized and an equalized elliptic filter. The equalizer is of 10th degree and the elliptic filter is of 6th degree.

property

$$\frac{1}{2\pi}\int_0^{\pi} \tau_g(\omega)d\omega = 0.5n_e, \tag{30}$$

where $\tau_g(\omega)$ is the equalizer group delay and the integral is taken over half the sampling interval ($0 \leqq \omega \leqq \pi$). Since $\tau_g(\omega) \geqq 0$, i.e., group delays add, to justify eq. (30) it is sufficient to show that a first-degree all-pass equalizer has the required property. The z-transform of a first-degree all-pass equalizer is

$$H(z) = \frac{1 - z^{-1}/a}{1 - az^{-1}}, \tag{31}$$

where a is the pole position and $1/a$ is the zero position in the z-plane. The group delay is commonly defined as

$$\tau_g(\omega) = -\frac{d[\measuredangle H(e^{j\omega})]}{d\omega}, \tag{32}$$

where $\measuredangle H(e^{j\omega})$ is the phase of the transfer function. Using eqs. (31) and (32) we obtain

$$\tau_g(\omega) = \frac{1 - a^2}{1 + a^2 - 2a\cos\omega} \tag{33}$$

for the first-degree equalizer. Integrating eq. (33) from 0 to π and normalizing by 2π gives

$$\frac{1}{2\pi}\int_0^{\pi} \frac{1 - a^2}{1 + a^2 - 2a\cos\omega}\, d\omega = \frac{1}{\pi}\tan^{-1}\left[\frac{(1 - a^2)\tan(\omega/2)}{(1 - a)^2}\right]\Bigg|_0^{\pi} = \frac{\pi}{2\pi} = 0.5.$$

The significance of eq. (30) is that one can estimate the minimum-order equalizer required to equalize a given group delay characteristic by determining the area between the line $\tau = \tau_{max}$ and the curve $\tau_g(\omega)$ and dividing by π, where τ_{max} is the maximum value of $\tau_g(\omega)$ in the passband. Thus in the example of Fig. 8, the estimated order of the equalizer is approximately $(13 \times \pi/2)/\pi = 6.5$. Of course, the required order of the equalizer must be greater than the estimate given above, since this estimate assumes the delay of the equalizer exactly compensates the delay of the unequalized filter. As the degree of the equalizer is increased over the estimate, the peak error of approximation decreases monotonically.

We have used the above algorithm, along with initial estimates of equalizer order, to equalize three sets of elliptic filters. The data for

Table I — Comparisons between optimum FIR and equalized elliptic digital filters
(Set 1: $\delta_1 = 0.01$, $\delta_2 = 0.0001$)

F_p	F_s	n	N	n_e	$\bar{\tau}_g$	r	N_1*	N_2*
0.0502	0.13999	5	21	2	28.7	12.1	11	11
				4	42.7	3.4		13
0.09846	0.24111	5	21	2	14.5	11.6	11	11
				4	22.2	4.1		13
				6	29.4	0.8		15
0.14722	0.25297	6	21	4	17.6	13.1	11	14
				6	23.0	6.3		16
				8	28.5	2.6		18
0.19507	0.30647	6	21	4	13.8	16.0	11	14
				6	17.8	8.7		16
				8	22.0	4.2		18
0.24163	0.34842	6	21	6	14.5	11.1	11	16
				8	18.3	7.0		18
				10	21.8	3.6		20
0.28664	0.41668	5	21	6	11.6	8.4	11	15
				8	14.5	3.8		17
				10	17.3	1.6		19
0.33014	0.43727	5	21	6	10.7	14.7	11	15
				8	13.1	8.3		17
				10	15.7	4.5		19
0.37254	0.47479	4	21	6	8.7	19.1	11	13
				8	11.1	6.5		15
				10	13.4	3.2		17
0.41665	0.49417	3	21	8	9.6	6.3	11	14
				10	11.8	3.2		16

* N_1 is the number of multiplications per sample for the optimum FIR filter; N_2 is the number of multiplications per sample for the equalized elliptic filter.

these three sets of filters are given in Tables I through III. Included in the table are the filter specifications (δ_1, δ_2, F_p, F_s); the required elliptic order n; the required FIR filter duration N; the equalizer order n_e; the average passband delay, $\bar{\tau}_g$ (in samples), of the equalized filter; the percentage ripple, r, in the passband group delay of the equalized filter; and a comparison between the number of multiplications per sample required in both the optimum FIR filter and the equalized elliptic filters. The data in these tables indicate that to achieve equalization to within about a 3-percent error requires on the order of 30 percent more multiplications per sample for the equalized filter than for the optimum FIR design, although in most cases the unequalized elliptic filter was more efficient than the optimum FIR designs. Thus it would appear that, at least for these restricted results, if constant group delay is required in addition to the equiripple magni-

Table II — Comparisons between optimum FIR and equalized elliptic digital filters
(Set 2: $F_p = 0.25$, $\delta_1 = 0.02$, $\delta_2 = 0.001$)

F_s	n	N	n_e	$\bar{\tau}_g$	r	N_1*	N_2*
0.4893	2	11	2	3.3	1.2	6	6
			4	5.6	0.1		8
0.44816	3	13	2	4.5	9.4	7	8
			4	7.3	1.0		10
0.39146	4	19	2	5.9	25.1	10	9
			4	8.8	8.0		11
			6	11.9	2.2		13
0.34153	5	29	2	8.4	37.4	15	11
			4	10.6	21.6		13
			6	13.7	11.6		15
			8	16.9	5.6		17
			10	20.3	2.4		19
0.30639	6	45	4	13.8	34.7	23	14
			6	16.0	25.0		16
			8	18.7	16.9		18
			10	22.0	11.7		20
			12	25.5	7.9		22
			14	29.4	5.2		24
			16	32.8	3.2		26
			18	36.3	1.8		28

* N_1 is the number of multiplications per sample for the optimum FIR filter; N_2 is the number of multiplications per sample for the equalized elliptic filter.

tude characteristics, then the optimum FIR filter is always more efficient than an equalized elliptic filter. It should also be noted that the delay of the optimum FIR filter [$(N - 1)/2$ samples] was *always* less than the delay of the equalized elliptic filter.

The examples of Tables I through III considered filters where the order of the unequalized elliptic filter was six or less. It can be argued that, for higher-order elliptic designs, the relative efficiency of the elliptic filter over the optimum FIR filter is far greater than for lower-order designs; hence in these cases perhaps the equalized filter may still be more efficient than the optimum FIR design. This conjecture turns out to be untestable because high-order elliptic filters have a peak passband delay τ_{max} which is much larger than for low-order filters, hence the order required for the equalizer becomes extremely large and thus is not even practical to consider if equalization over the entire passband is required. To illustrate this point, Fig. 9 shows the group delay of a 10th-order elliptic low-pass filter with $F_p = 0.25$. Using eq. (30) to get an estimate of n_e we arrive at a value of $n_e = 45$. Since this value of n_e is only an underbound on the actual order of the

Table III — Comparisons between optimum FIR and equalized elliptic digital filters
(Set 3: $F_p = 0.25$, $\delta_1 = 0.02$, $\delta_2 = 0.0001$)

F_s	n	N	n_e	$\bar{\tau}_g$	r	N_1*	N_2*
0.49661	2	11	2	3.3	1.2	6	6
			4	5.6	0.1		8
0.47564	3	11	2	4.5	9.1	6	8
			4	7.3	1.0		10
0.43591	4	17	2	5.8	23.3	9	9
			4	8.8	7.0		11
			6	11.8	1.7		13
0.38983	5	21	2	8.0	33.4	11	11
			4	10.3	18.0		13
			6	13.5	8.7		15
			8	16.7	3.7		17
			10	20.0	1.4		19
0.34878	6	31	4	12.8	28.9	16	14
			6	15.5	19.2		16
			8	18.2	11.8		18
			10	22.0	7.7		20
			12	25.3	4.3		22
			14	28.8	2.2		24

* N_1 is the number of multiplications per sample for the optimum FIR filter; N_2 is the number of multiplications per sample for the equalized elliptic filter.

equalizer, it is clear that it is not practical to try to obtain such a high-degree equalizer.

Another interesting question which arises when one considers the idea of equalizing an IIR filter is how does the cascade combination of an elliptic filter and an all-pass equalizer compare to the optimum IIR filter which best approximates both the desired magnitude and group delay characteristics? It is clear that the optimum IIR filter can be no worse than the cascade; the question remains as to how much

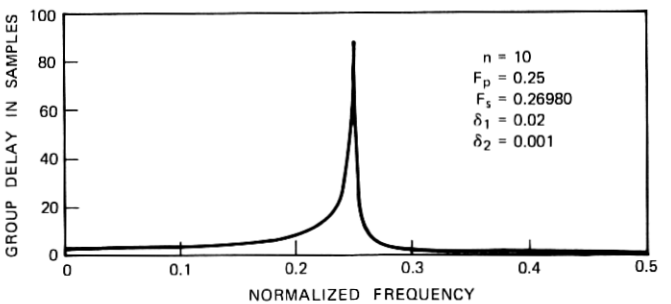

Fig. 9—The group delay of a 10th-order elliptic filter with $F_p = 0.25$.

better it can be. There is no clear-cut answer to this question. However, based on our experience with equalized elliptic filters, several observations can be made. (We shall use the z-plane pole-zero plot of a typical equalizer filter, shown in Fig. 10, to aid in understanding the nature of the equalized filter.)

(*i*) The zeros of the elliptic filter lie on the unit circle to give good stopband attenuation.
(*ii*) The zeros of the equalizer lie outside the unit circle to give positive delay.
(*iii*) The poles of the elliptic filter are constrained by the transition width requirements of the low-pass filter.
(*iv*) The poles of the equalizer lie approximately on a circle of fixed radius, and are approximately equally spaced in the passband.

If the zeros of the optimum filter are not constrained to lie on the unit circle, then each second-order section will require four multiplications per sample, rather than the three multiplications for each second-order section of the elliptic design and the two multiplications for each second-order section of the all-pass equalizer. Based on the

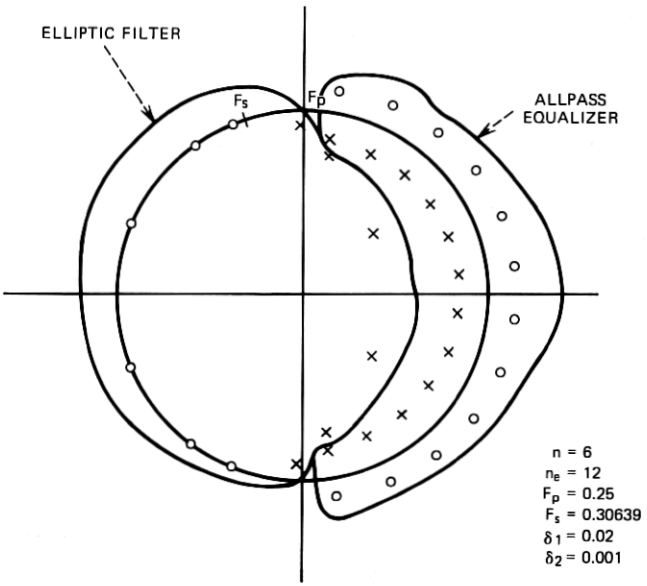

Fig. 10—The pole-zero positions of an equalized elliptic filter.

above observations, it seems unlikely that there is much to gain by using the optimum IIR filter over the equalized filter.

VIII. GENERAL DISCUSSION

In this paper we have considered only one basis for comparison between optimum FIR filters and equivalent IIR designs, that measure being the number of multiplications per sample required in the standard method of realization for each of these filter types. The justification for this measure is that in hardware (and generally in software) the number of multiplications per sample is an excellent measure of the complexity required in the implementation as well as the factor which determines the maximum throughput rate of the system.[9] However, there are many other ways for comparing these filter types when one takes into consideration the various finite word-length effects which occur in a practical design situation. In this section we review several of these design issues.

Among the various finite word-length effects are roundoff noise, both uncorrelated and correlated (e.g., limit cycles), and coefficient quantization sensitivity. For direct-form FIR realization, the peak roundoff noise can easily be made to be less than $\frac{1}{2}$ of the least significant bit by accumulating partial sums in an extended length register and then rounding the final result. For cascade IIR filters realized with fixed-point arithmetic, the roundoff noise problem is inherently related to the dynamic range problem,[7] and involves the concepts of pole-zero pairing and section ordering. Jackson[10] has shown that with reasonable pairing and ordering the uncorrelated roundoff noise variance can be minimized. However, even in the best of situations, the roundoff noise is equivalent to several bits. In terms of correlated roundoff noise, i.e., limit cycles, the direct-form FIR realization has no zero-input limit cycles (because no feedback is present), whereas the cascade IIR realization will generally exhibit zero-input limit cycles. Kaiser[11] has extensively studied these limit cycles and has developed bounds and estimates for their amplitude and frequency.

The coefficient quantization problem is one of the most difficult finite word length effects to treat analytically. Rounding of infinite precision filter coefficients to a fixed number of bits alters the overall frequency response of the filter in a complicated manner. Avenhaus[12] has shown that straight rounding of the infinite precision filter coefficients is generally inferior to optimizing the filter performance over the finite set of fixed precision filter coefficients. However, there are

no general procedures for performing this optimization, nor are there any guarantees of convergence of the existing methods. Furthermore, in many cases the advantage of optimizing finite precision coefficients over straight rounding of the infinite precision coefficients is small. Thus for the case of coefficient quantization neither direct-form realization of FIR filters nor cascade realization of IIR filters seems to offer a relative advantage here.

Thus it is difficult, if not impossible, to be quantitative in comparing FIR and IIR filters based on anything other than number of multiplications per sample. This is why we have used this measure throughout this paper.

IX. SUMMARY

In this paper some comparisons were made between equivalent FIR and IIR digital filters based on the number of multiplications per sample required to realize these filters. In the case of low-pass filters with quasi-equiripple magnitude characteristics, IIR elliptic filters could generally be realized more efficiently than equivalent linear phase FIR filters. When the additional requirement of constant group delay in the passband was added to the specifications, comparisons showed the linear phase FIR filters to be more efficient than group-delay-equalized elliptic IIR filters.

Additionally, a novel set of design charts for determining the minimum filter order required to meet given filter specifications for both digital and analog elliptic, Chebyshev, and Butterworth low-pass filters was presented. Explanation of how to use these charts to gain insight into the various filter parameter tradeoffs was also given.

REFERENCES

1. J. E. Storer, *Passive Network Synthesis*, New York: McGraw-Hill Book Co., 1957.
2. O. Herrmann, L. R. Rabiner, and D. S. K. Chan, "Practical Design Rules for Optimum Finite Impulse Response Low-Pass Digital Filters," B.S.T.J., *52*, No. 6 (July-August 1973), pp. 769–799.
3. J. F. Kaiser, "Digital Filters," in *System Analysis by Digital Computer*, edited by F. F. Kuo and J. F. Kaiser, New York: John Wiley and Sons, 1966.
4. R. M. Golden and J. F. Kaiser, "Design of Wideband Sampled Data Filters," B.S.T.J., *43*, No. 4, Pt. 2 (July 1964), pp. 1533–1546.
5. A. G. Constantinides, "Spectral Transformations for Digital Filters," Proc. IEE, *117*, No. 8, 1970, pp. 1585–1590.
6. E. Christian and E. Eisenmann, *Filter Design Tables and Graphs*, New York: John Wiley and Sons, 1966.
7. L. B. Jackson, "On the Interaction of Roundoff Noise and Dynamic Range in Digital Filters," B.S.T.J., *49*, No. 2 (February 1970), pp. 159–184.
8. A. G. Deczky, "Synthesis of Recursive Digital Filters Using the Minimum p-Error Criterion," IEEE Trans. Audio and Electroacoustics, *AU-20*, No. 4 (October 1972), pp. 257–263.

9. L. B. Jackson, J. F. Kaiser, and H. S. McDonald, "An Approach to the Implementation of Digital Filters," IEEE Trans. Audio and Electroacoustics, *AU-16*, No. 3 (September 1968), pp. 413–421.
10. L. B. Jackson, "Roundoff-Noise Analysis for Fixed-Point Digital Filters Realized in Cascade or Parallel Form," IEEE Trans. Audio and Electroacoustics, *AU-18*, No. 2 (June 1970), pp. 107–122.
11. J. F. Kaiser, "An Overview on Digital Filters," Newsletter of IEEE Circuit Theory Group, *6*, No. 1 (March 1972).
12. E. Avenhaus, "On the Design of Digital Filters with Coefficients of Limited Word Length," IEEE Trans. Audio and Electroacoustics, *AU-20*, No. 3 (August 1972), pp. 206–212.