THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 55, No. 1, January 1976
Printed in U.S.A.

An Algorithmic Procedure for Designing Hybrid FIR/IIR Digital Filters

By M. R. CAMPBELL, R. E. CROCHIERE, and L. R. RABINER

(Manuscript received July 3, 1975)

An algorithmic procedure for designing hybrid FIR/IIR digital filters is proposed and evaluated in this paper. The design is implemented as a two-stage optimization in which a Hooke and Jeeves optimization procedure is used to optimize the IIR component of the filter and the McClellan et al. optimization procedure is used to optimize the FIR component of the filter. To evaluate this method, a set of eight low-pass filters were designed in which a single complex conjugate pole pair was used as the IIR component. The resulting designs were compared and contrasted with standard IIR low-pass filters and the optimal FIR, linear-phase, low-pass filter in terms of multiplications, storage, and group delay properties.

I. INTRODUCTION

A wide variety of digital filter-design methods have been proposed and studied in the past several years.[1–9] Generally, these design methods can be classified as analytical or algorithmic, depending on the form of solution of the approximation problem which is used. Additionally, the resulting designs are classified as finite impulse response (FIR) or infinite impulse response (IIR), depending on the filter properties. FIR filters have the property that they can be easily designed to have exactly linear phase. Furthermore, linear-phase FIR filters can be designed to approximate an arbitrary magnitude response to within given tolerances by using a sufficiently high order. IIR filters cannot achieve a linear-phase response exactly, but are capable of approximating sharp cutoff filters with considerably lower order filters than are required for the FIR designs which meet identical magnitude specifications.[10] Thus, for many practical filter applications, there is a trade-off between the exact linear-phase response obtainable using an FIR filter and the reduced filter order obtainable using an IIR filter.

In this paper, a hybrid design is proposed which bridges the gap, somewhat, between the FIR and IIR filters. The hybrid filter is a particular class of IIR filter where the degree of the numerator of its system

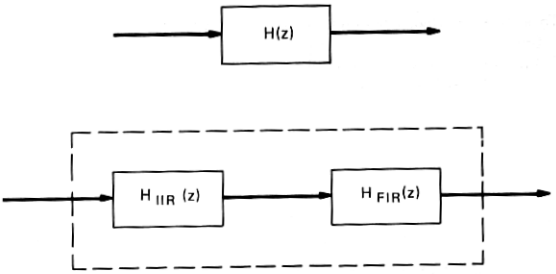


Fig. 1—Hybrid FIR/IIR filter.

function is significantly higher than the degree of the denominator. Thus, it is reasonable to consider the hybrid design as a cascade of an Mth-order IIR filter with Nth-order FIR filter, as shown in Fig. 1, where $N \gg M$. In the designs discussed here, M is set to 2* and N is in the range of 27 to 51. The idea behind this hybrid design is to incorporate the good features of both IIR and FIR filters, yielding a resulting filter which can meet arbitrary design specifications with a significantly smaller filter order than is required by the FIR filter alone, and with a smaller group delay variation (i.e., more phase linearity) than is normally obtained by the IIR filter alone.

Based on the above discussion, the hybrid filter of Fig. 1 can be put in the form

$$H(z) = H_{\text{IIR}}(z)H_{\text{FIR}}(z), \tag{1}$$

where, by assuming a second-order denominator as mentioned above,

$$H_{\text{IIR}}(z) = \frac{1}{(1 - \rho e^{j\theta}z^{-1})(1 - \rho e^{-j\theta}z^{-1})} \tag{2}$$

and

$$H_{\text{FIR}}(z) = \sum_{n=0}^{N-1} h_F(n)z^{-n}, \tag{3}$$

where ρ is the radius of the pole in the z plane and θ is the pole angle. (Although we have used a second-order IIR filter in eq. (2) and in the examples of this paper, the design approach is general and can be applied to other orders of IIR sections as well.) The overall filter is decomposed into a cascade of sections for two reasons. The first is to emphasize the fact that the hybrid filter is most readily realized as a cascade of an IIR filter with an FIR filter. In this way, the direct form structure can be used to realize the FIR section so as to fully utilize the symmetry in the impulse response due to the linear phase condition.

* Cases where $M = 1$ were also studied but have not led to any useful results.

The second reason that this cascade of sections is used is because it separates the design problem into two distinct parts; one which is fairly simple and one which can be solved using a well-known FIR approximation method.[11]

To obtain the best second-order denominator hybrid filter approximation to the desired specifications, the parameters ρ and θ of the IIR filter are systematically varied, and an optimal FIR filter is obtained for each set of parameters using the design program of McClellan et al.[11] An alternative method of obtaining such a hybrid design would be to use a different type of algorithmic procedure (e.g., Ref. 8 or 9) and simultaneously vary both the numerator coefficients and the denominator coefficients. Unfortunately, since the numerator order is so much higher than the denominator order, these optimization methods are not always successful.

In the remainder of this paper we discuss the hybrid filter-design algorithm and present some typical results on low-pass filter designs obtained with this method. Then we compare and contrast the resulting designs with some standard IIR low-pass filters including Butterworth, Chebyshev, and elliptic filters, and with the optimal FIR filter. The bases of comparison are the number of multiplications per sample, the group delay variation, and the storage requirements. We conclude with a discussion of the properties of the hybrid filters.

II. DESIGN ALGORITHM

Since the algorithm is being applied to the design of low-pass filters (although it is equally useful for any arbitrary magnitude function), the desired frequency response of the system $H_I(e^{j\omega})$ is of the form

$$|H_I(e^{j\omega})| = \begin{cases} 1 & 0 \leqq \omega \leqq 2\pi F_p \\ 0 & 2\pi F_s \leqq \omega \leqq \pi, \end{cases} \tag{4}$$

where F_p and F_s are the passband and stopband edges, respectively. The tolerances are δ_p in the passband and δ_s in the stopband. Thus, $|H_I(e^{j\omega})|$, the magnitude response of the composite system of Fig. 1, satisfies the inequalities

$$\begin{aligned} 1 - \delta_p \leqq |H_I(e^{j\omega})| &\leqq 1 + \delta_p \qquad 0 \leqq \omega \leqq 2\pi F_p \\ 0 \leqq |H_I(e^{j\omega})| &\leqq \delta_s \qquad 2\pi F_s \leqq \omega \leqq \pi. \end{aligned} \tag{5}$$

The method in which the individual magnitude responses of the IIR filters and the FIR filters are chosen is shown in Fig. 2. Initial values are chosen for the IIR coefficients ρ and θ, based on the location of the highest Q-pole pair of an elliptic filter which meets the tolerance scheme of eq. (5). The reason for this choice will be discussed later.

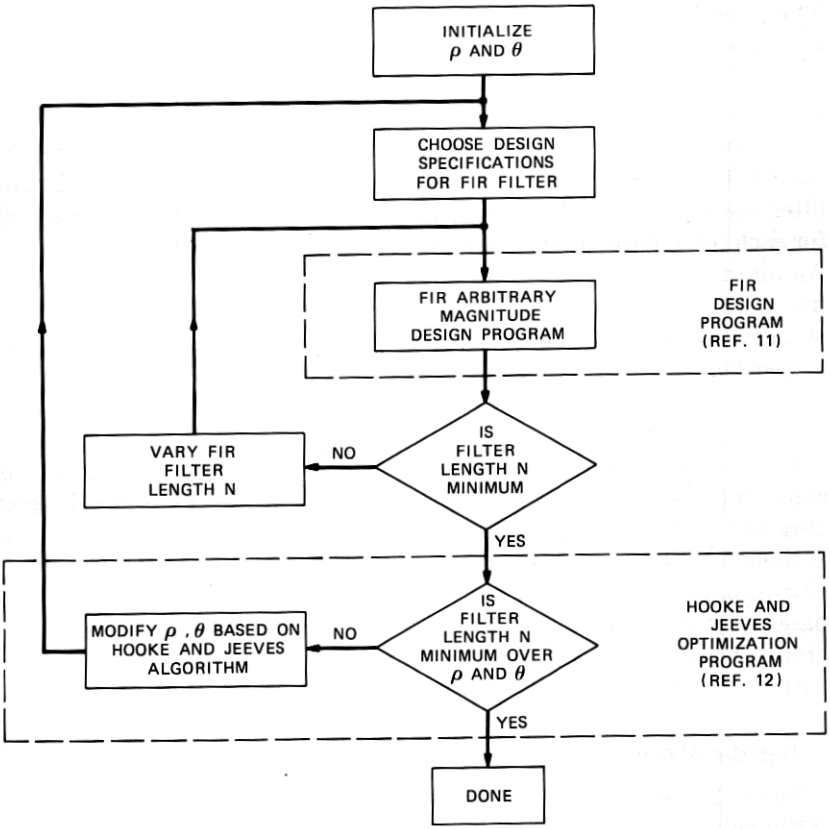


Fig. 2—Block diagram of the design algorithm.

Based on the initial values of ρ and θ, a set of design specifications for the FIR filter is obtained from eqs. (1) and (5) of the form

$$\begin{aligned} \frac{1-\delta_p}{|H_{\text{IIR}}(e^{j\omega})|} \leqq |H_{\text{FIR}}(e^{j\omega})| \leqq \frac{1+\delta_p}{|H_{\text{IIR}}(e^{j\omega})|} \qquad & 0 \leqq \omega \leqq 2\pi F_p \\ 0 \leqq |H_{\text{FIR}}(e^{j\omega})| \leqq \frac{\delta_s}{|H_{\text{IIR}}(e^{j\omega})|} \qquad & 2\pi F_s \leqq \omega \leqq \pi. \end{aligned} \tag{6}$$

Equation (6) specifies the appropriate parameters (the band edges, desired values, and weighting) for the McClellan et al. arbitrary magnitude FIR design program.[11] The final parameter required is N, the filter length. An initial guess of the value of N is made, and the design program yields the optimal FIR filter for the given specifications. A control loop is used to find the smallest value of N which can be used to meet the given input specifications.

Once the minimum N (for the given values of ρ and θ) is obtained, an outer loop optimization program is used to vary ρ and θ to obtain the minimum value of N as a function of ρ and θ. The algorithm used to find the optimum values of ρ and θ is the well-known Hooke and Jeeves optimization.[12] Generally, the Hooke and Jeeves algorithm is capable of optimizing a continuous function of (several) continuous variables. However, for this problem the function $[N(\rho, \theta)]$ is not continuous, but instead is discrete (integer) valued. To handle the problems created by this discrete-valued function, fairly careful control over the variation of ρ and θ had to be maintained. To illustrate this point, Fig. 3 shows a typical contour of the variation of N as a function of ρ.

The minimum of this function occurs at the point labeled A. However, because $N(\rho, \theta)$ is discrete, it is possible for the optimization algorithm to prematurely terminate on flat regions such as those

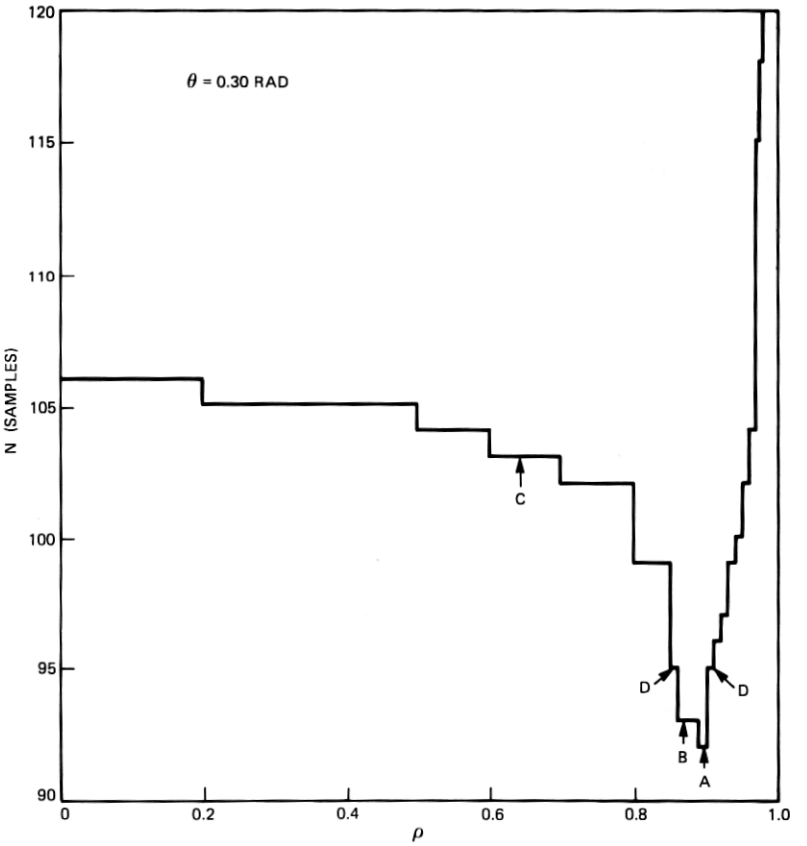


Fig. 3 —Typical contour of N as a function of ρ.

labeled B and C. To minimize these difficulties, a careful choice is required of the initial step sizes, $\Delta\rho$ and $\Delta\theta$, used by the Hooke and Jeeves algorithm to vary ρ and θ. For example, if $\Delta\rho$ is too large, the algorithm may not find the "valley" of the function. Similarly, if $\Delta\rho$ is too small, the algorithm may "hang up" on a flat region as it reduces its step size. Successful choices of $\Delta\rho$ and $\Delta\theta$ appear to be approximately one-half the distance across the "valley" of the function (e.g., one-half the distance between the two points labeled D in Fig. 3). For the examples in the next section, the choices $\Delta\rho = 0.005$, $\Delta\theta = 0.02$ (radians) yielded good results. Figure 4 shows another example of the variation of N as a function of ρ and θ. Here we can observe that there are two reciprocal minima of N as a function of ρ. If the optimum ρ_0 outside the unit circle is found, the corresponding optimum value of ρ inside the unit circle can be found by taking the reciprocal of ρ_0.

A factor that strongly affects the efficiency of the algorithm is the initial choice of ρ and θ. If accurate initial estimates of ρ and θ are used, they can result in a significant speed-up of the design method. As stated earlier, good estimates were found to be the location of the highest Q-pole pair of an elliptic filter which meets the same tolerance scheme. Other good estimates appear to be the highest Q-pole pair locations of the Chebyshev 2 and Chebyshev 1 designs which meet the same tolerance scheme. For greater assurance that an optimum has been found, several of these starting points may be tried to see if the algorithm converges to the same value of ρ and θ.

Another factor that strongly affects the efficiency of the algorithm is the strategy of the control loop for varying N (see Fig. 2). Generally, it was found that a good initial guess of N after changing ρ and θ is the previous value of N. An efficient strategy for increasing or decreasing N to find its new minimum then appeared to be a "tree" search (or log search). That is, N is incremented or decremented by an amount ΔN, depending on whether the previous choice of N yields a design which meets the specifications in eq. (6). On the next trial, ΔN is reduced by one-half, and the process is repeated until $\Delta N = 1$. The search can be terminated sooner if at any stage the FIR design is sufficiently close to the tolerance requirements (e.g., within 1 percent) given in (6). Other variations on this strategy are also possible.

III. EXPERIMENTAL RESULTS

Using the hybrid filter-design algorithm of Fig. 2, several low-pass filters were designed, ranging from narrow-band to wide-band designs. Table I gives the filter specifications (band edges and ripple tolerances) for eight low-pass filters, along with the resulting IIR pole position, the length of the FIR section, and the length of an optimal linear-phase

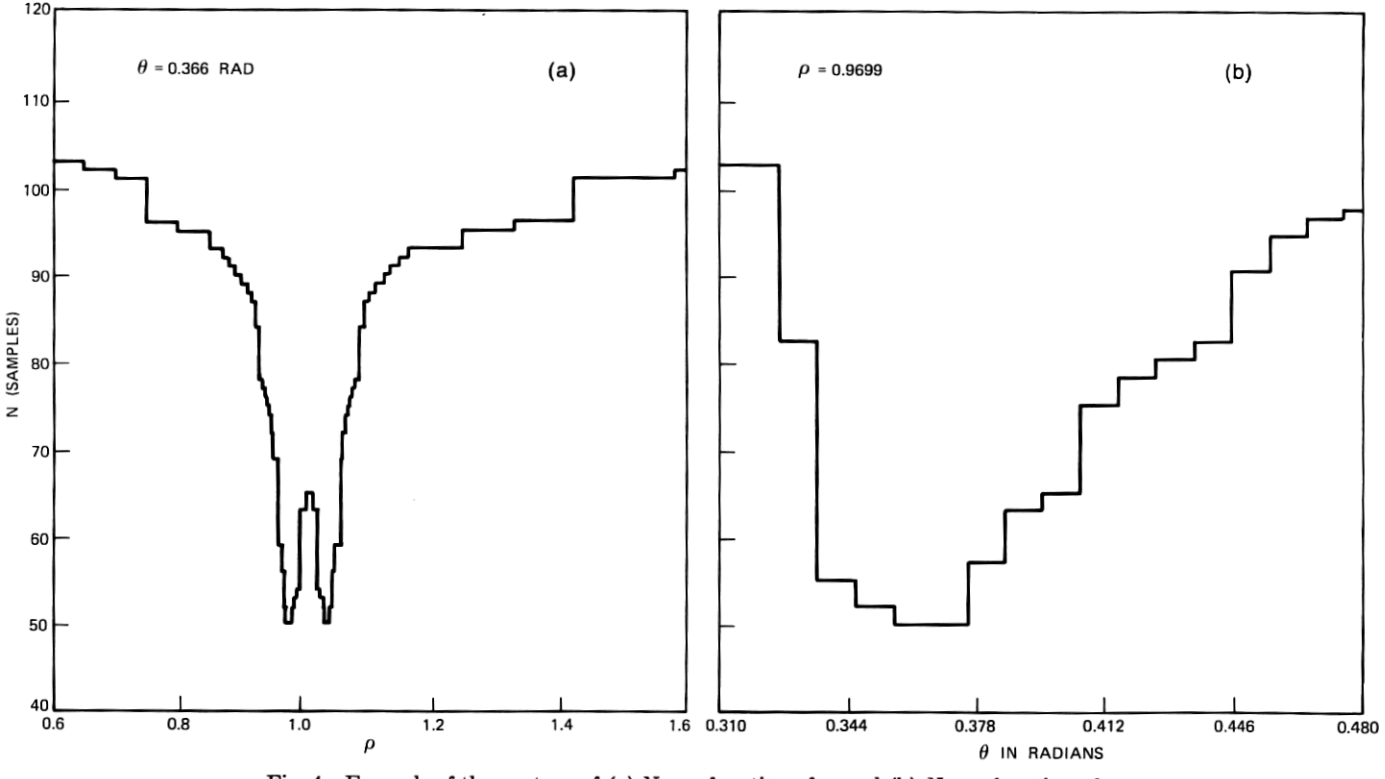


Fig. 4—Example of the contour of (a) N as a function of ρ, and (b) N as a function of θ.

Table I — Low-pass filter design

Filter Example	Filter Specifications				Optimum Hybrid FIR/IIR Design			Equivalent Optimum FIR Design	Highest Q-Pole of Equivalent Elliptic Design	
	F_p	F_s	δ_p	δ_s	ρ	θ(rad)	N	N_{FIR}	ρ_e	θ_e(rad)
1	0.05	0.075	0.01	0.001	0.9636	0.3363	39	109	0.9776	0.3238
2	0.1	0.125	0.01	0.001	0.9761	0.6817	50	107	0.9807	0.6367
3	0.15	0.175	0.01	0.001	0.9787	0.9979	51	106	0.9768	0.9530
4	0.2	0.225	0.01	0.001	0.9859	1.3133	51	105	0.9746	1.2684
5	0.05	0.125	0.005	0.0005	0.9634	0.3601	31	44	0.9534	0.3401
6	0.1	0.175	0.005	0.0005	0.9390	0.6805	27	43	0.9453	0.6605
7	0.15	0.225	0.005	0.0005	0.9582	0.9935	31	43	0.9482	0.9736
8	0.2	0.275	0.005	0.0005	0.9516	1.3127	29	42	0.9416	1.2927

FIR filter which also meets the design specifications. The first four examples are filters with a fairly narrow transition band (0.025), whereas the last four examples are filters with a wider transition band (0.075). As seen in Table I, the reduction in length of the FIR filter in the hybrid, from the optimal FIR linear-phase filter, is on the order of 1.5:1 to 2:1—i.e., there is a fairly significant reduction in FIR filter length.

Figures 5 to 7 show examples of the magnitude responses of filters nos. 1, 5, and 2, respectively. Figure 5a shows the log magnitude response of the IIR filter, Fig. 5b shows the log magnitude response of the FIR section, and Figs. 5c and 5d show the linear and log magnitude responses respectively, of the composite filter. It can be seen from these figures that the response of the IIR filter makes the requirements for the FIR section significantly easier to obtain. For example, the required tolerance near $\omega = \pi$ is on the order of -15 dB in order that the composite response be more than 60 dB down at this frequency.

Figures 6 and 7 show examples of some undesirable characteristics of the frequency response which can occur in the composite filter. In these cases, there is a ripple of the magnitude response in the transition band of the filter.* Since there is no real constraint on the composite filter response in the transition band, this behavior is not unreasonable. The question becomes one of whether or not a filter with a ripple in the transition band is acceptable. In general, the answer to this question is that it depends on the intended application. In some cases, this behavior is acceptable, in others it is not.

Table II gives the approximate height of the transition band ripple for each example of Table I. In example 5 (see Fig. 6d), the transition band ripple is down by 17 dB and may be perfectly acceptable. In example 2 (see Fig. 7d), the ripple peak amplitude is about 1.06—i.e.,

* Note that the pure FIR filter does not have such a peak in the don't care region.

Table II — Transition band ripple for Table I examples

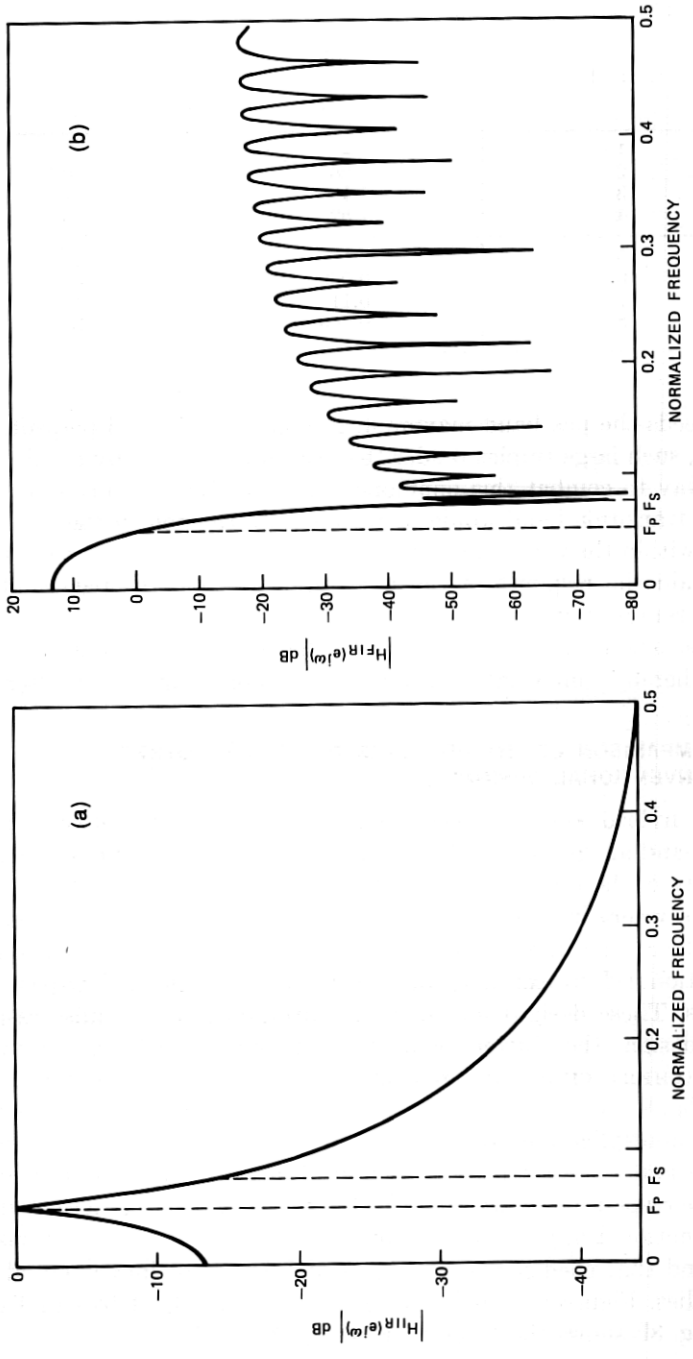
Example	Height of Transition Band Ripple	
	Magnitude	dB
1	—	
2	1.06	0.5
3	1.16	1.29
4	1.52	3.64
5	0.14	−17
6	0.022	−33
7	0.11	−19
8	0.063	−24

it exceeds the passband maximum response by about 6 percent. Generally, such large ripples render the filter useless in many applications. One way to combat this nonmonotonicity of the magnitude response in the transition band would be to constrain the pole of the IIR section to lie within the passband of the filter. Another possibility would be to constrain the response of the FIR section beyond the passband edge to guarantee monotonic behavior of the magnitude response. Either of these alternatives would lead to higher values of the FIR filter duration, thereby somewhat negating the gains of using the IIR filter.

IV. COMPARISON OF HYBRID FILTER DESIGNS TO OTHER CONVENTIONAL DESIGNS

The hybrid FIR/IIR design represents a trade-off between an FIR design and an IIR design. The usefulness of such an approach depends strongly on how it compares with other conventional designs. To illustrate where in this scope of design alternatives a hybrid approach might be competitive, we have compared the hybrid examples designed in Section III to examples of five other conventional low-pass filter designs. These designs included the optimum finite impulse response (FIR) design, the Butterworth (BUT) design, the Chebyshev type 1 and 2 (CHEB1, CHEB2) designs, and the Cauer elliptic (ELLIPT) design.

In Fig. 8, the various designs are compared on the basis of the number of multiplications required for their implementation. In the FIR designs and the FIR parts of the hybrid designs, the symmetry of the impulse response was exploited. In the IIR designs, it was assumed that a conventional implementation of cascaded second-order sections was used and that coefficients of value 1 and 2 are not implemented with multiplies. Figure 8a shows the results for examples 1 to 4 in Table I and Fig. 8b shows the results for examples 5 to 8.


(a)
(b)
$|H_{IIR}(e^{j\omega})|$ dB
$|H_{FIR}(e^{j\omega})|$ dB
NORMALIZED FREQUENCY
F_P F_S
0
0.2
0.3
0.4
0.5
20
10
0
−10
−20
−30
−40
−50
−60
−70
−80

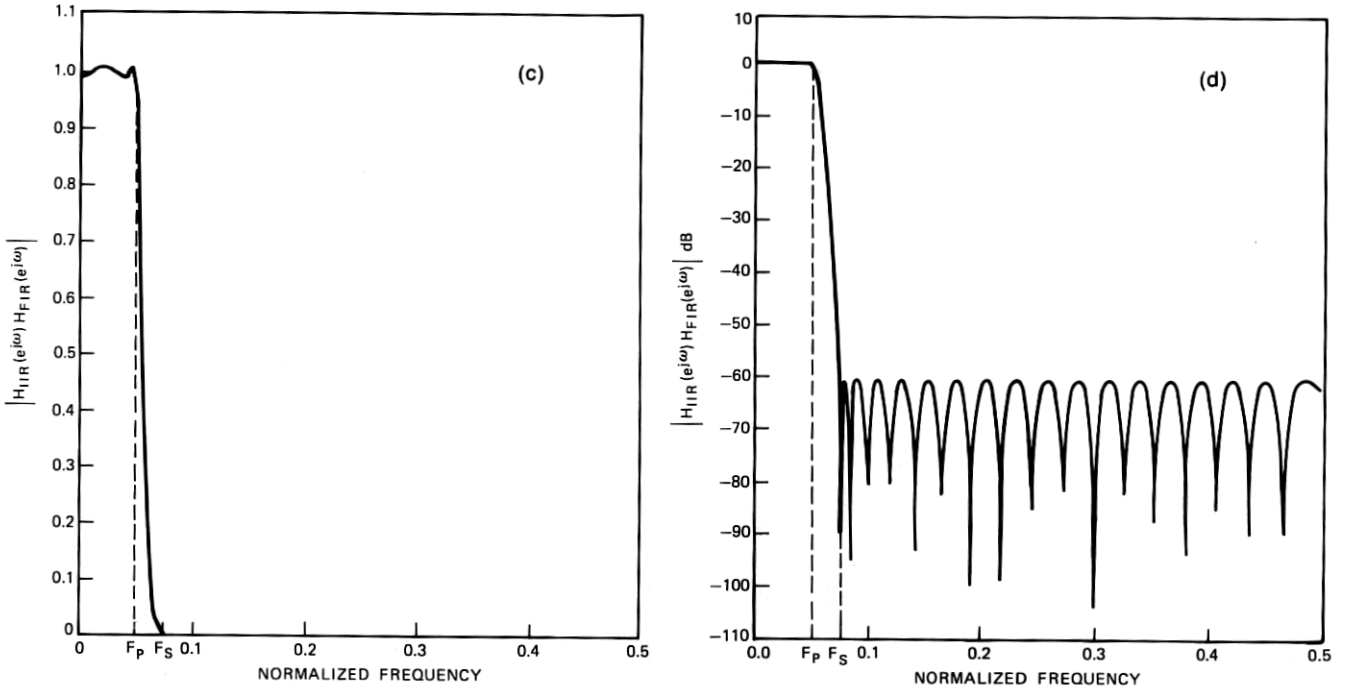
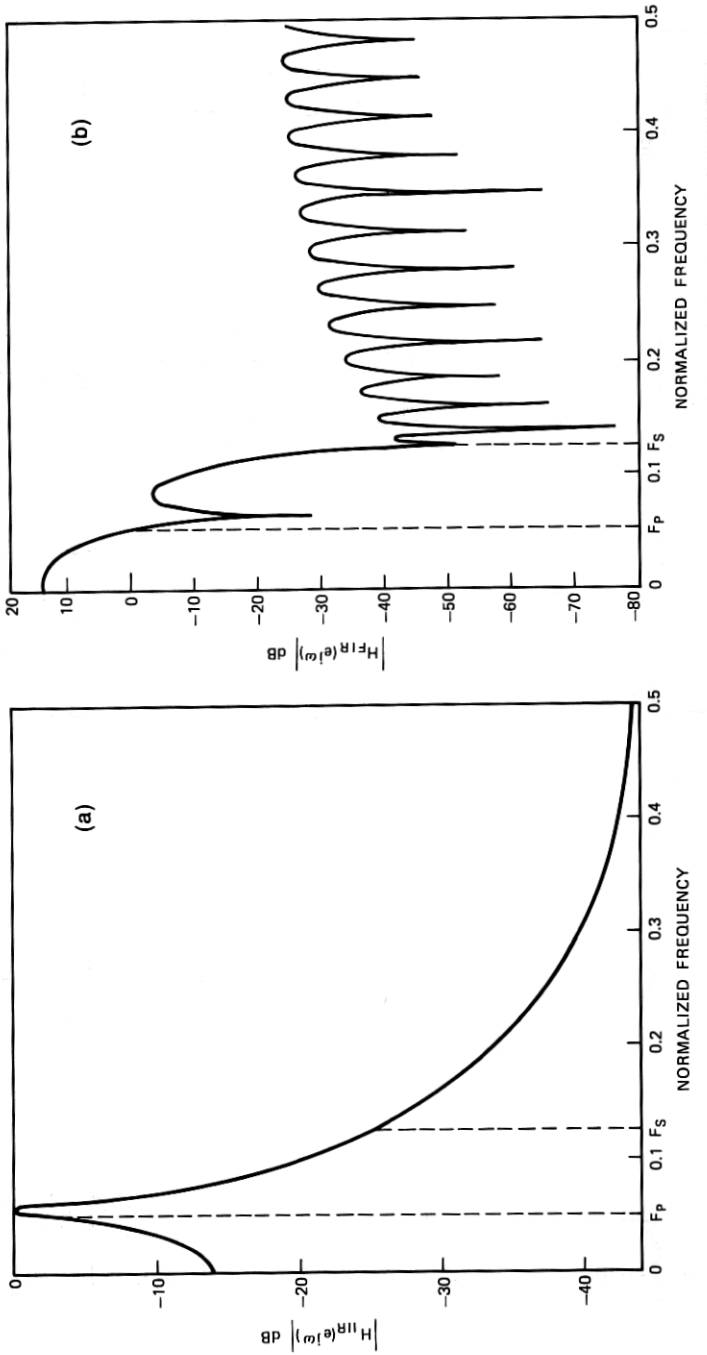


Fig. 5—Frequency responses of example 1. (a) Log magnitude plot of IIR section. (b) Log magnitude plot of FIR section. (c) Linear plot of composite magnitude response of FIR/IIR design. (d) Log magnitude plot of FIR/IIR design.


(a)
(b)
$|H_{IIR}(e^{j\omega})|$ dB
$|H_{FIR}(e^{j\omega})|$ dB
NORMALIZED FREQUENCY
NORMALIZED FREQUENCY
F_P
F_S
0.1
0.2
0.3
0.4
0.5
0
−10
−20
−30
−40
−50
−60
−70
−80
10
20

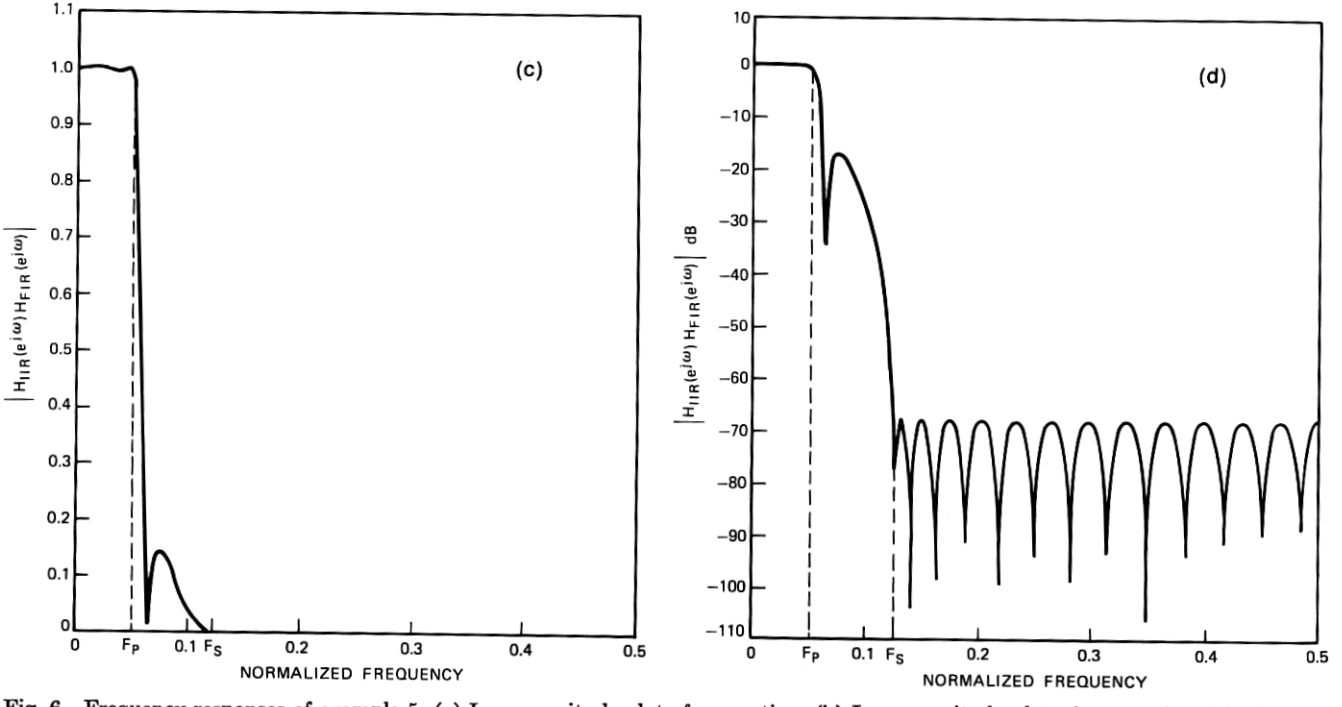
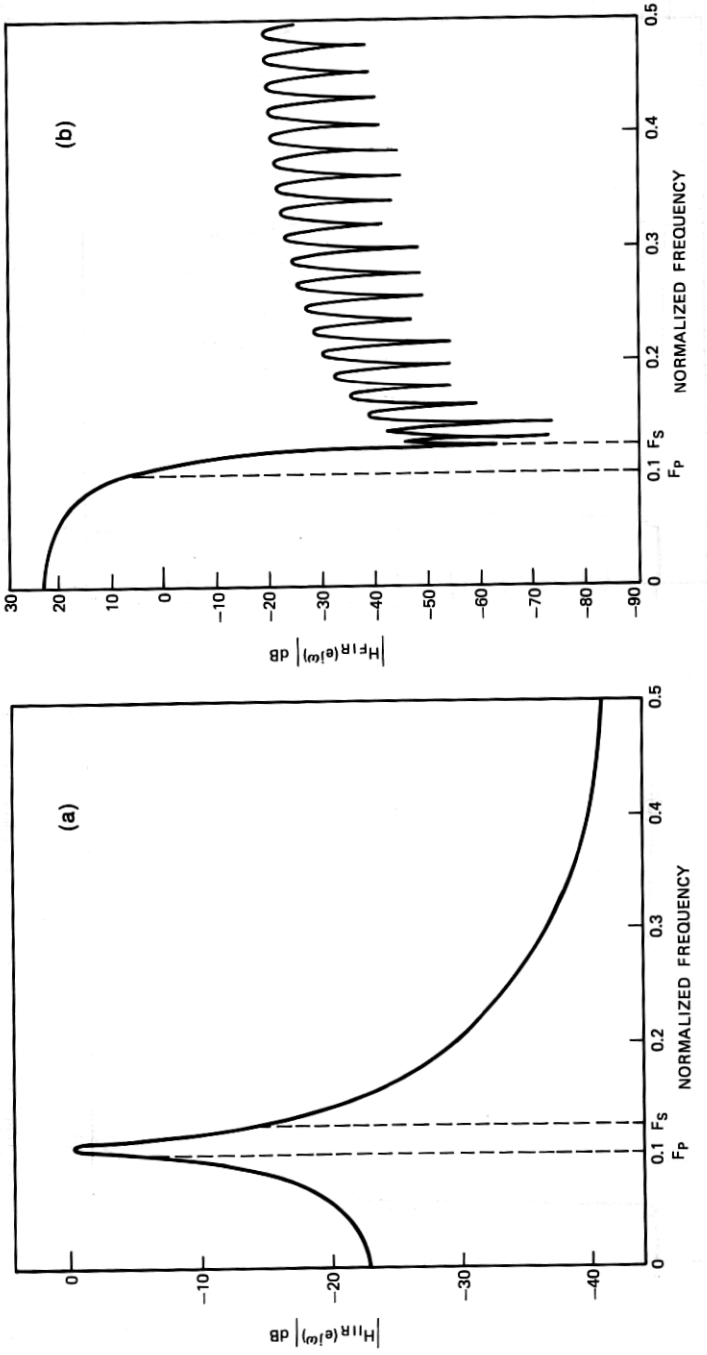


Fig. 6—Frequency responses of example 5. (a) Log magnitude plot of IIR section. (b) Log magnitude plot of FIR section. (c) Linear plot of FIR/IIR cascade. (d) Log magnitude plot of FIR/IIR cascade.


(a)
(b)
$|H_{IIR}(e^{j\omega})|$ dB
$|H_{FIR}(e^{j\omega})|$ dB
NORMALIZED FREQUENCY
F_P
F_S
0
0.1
0.2
0.3
0.4
0.5
30
20
10
0
−10
−20
−30
−40
−50
−60
−70
−80
−90

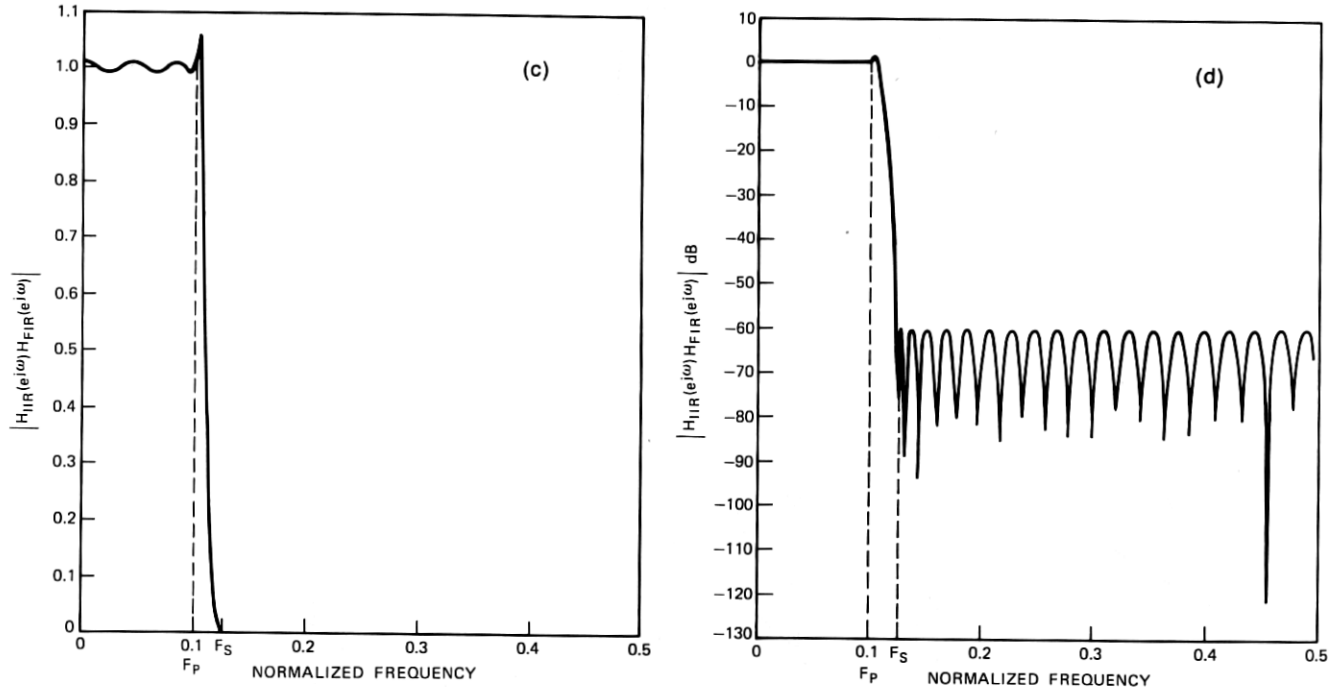


Fig. 7—Frequency responses for example 2. (a) Log magnitude plot of IIR section. (b) Log magnitude plot of FIR section. (c) Linear plot of FIR/IIR cascade. (d) Log plot of FIR/IIR cascade.

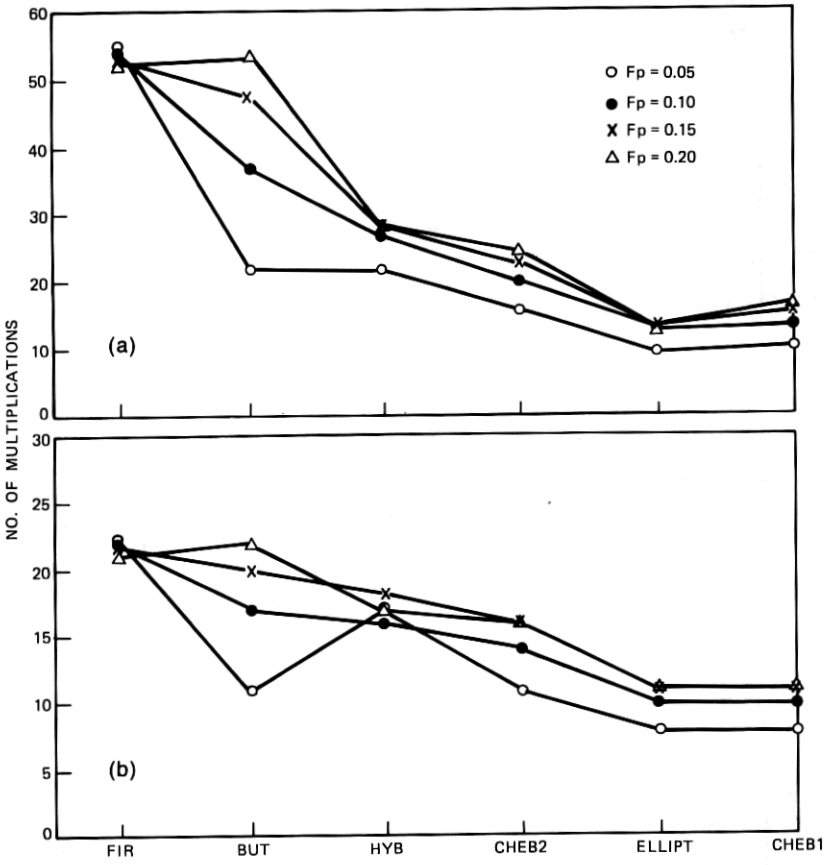


Fig. 8—Comparison of filter designs on the basis of the number of required multiplications.

The results of the comparison in Fig. 8 indicate that the hybrid design is often more efficient than the Butterworth design and is slightly less efficient than a Chebyshev 2 design for the same magnitude specifications.

A second criterion used for comparison was the group delay. Figure 9 shows plots of the group delay for each filter design for example 8. The FIR filter had the largest fixed delay at zero frequency, but its group delay was exactly flat (i.e., zero dispersion) across all frequencies. The hybrid design had a relatively flat component across most of the passband with most of its dispersion near the edge of the passband and within the transition band. The next least dispersive design appeared to be the Chebyshev 2 design, followed by the Butterworth, elliptic, and Chebyshev 1 designs.

Another measure of dispersion used to compare the designs was the maximum minus the minimum group delay across the passband. These results are plotted in Fig. 10. Obviously, the FIR design had exactly zero dispersion in these examples. The next best contenders appeared to be the hybrid and the Chebyshev 2 designs, followed by the elliptic, Butterworth, and Chebyshev 1 designs. In three examples, the order of the Butterworth filter was larger than what could be accommodated by the available design program, so these results were not included. It was essentially the large required order of the Butterworth filters that prevented them from having favorable group delay characteristics.

A final comparison was made on the number of data storage locations (i.e., state variables) necessary for the implementation of the

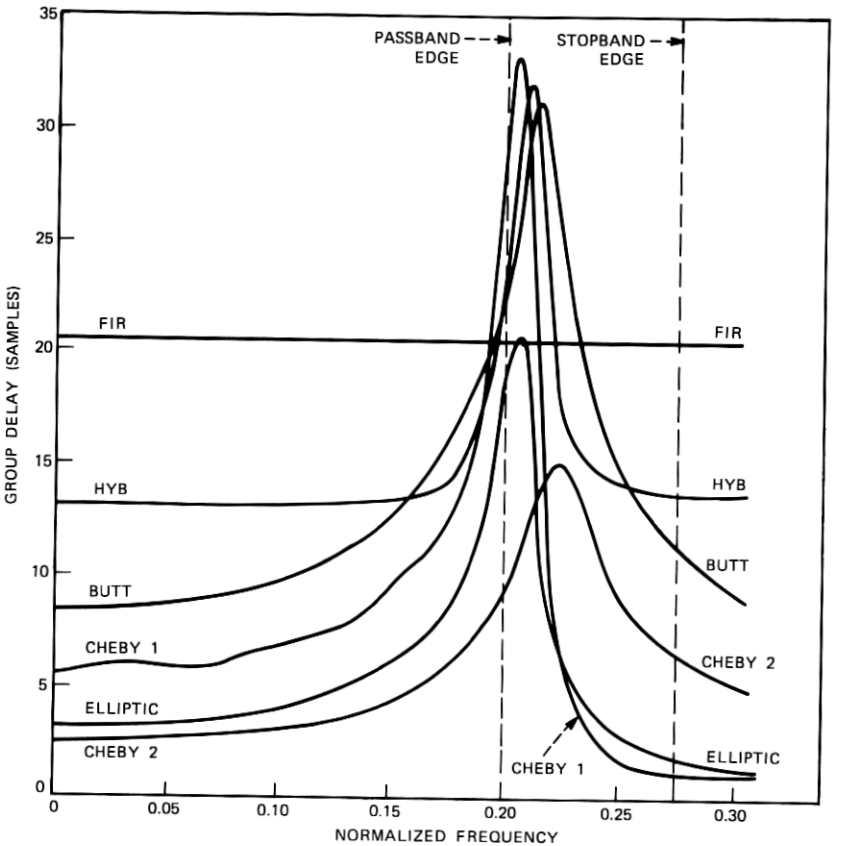


Fig. 9—Comparison of group delays of various filter designs for example 8.

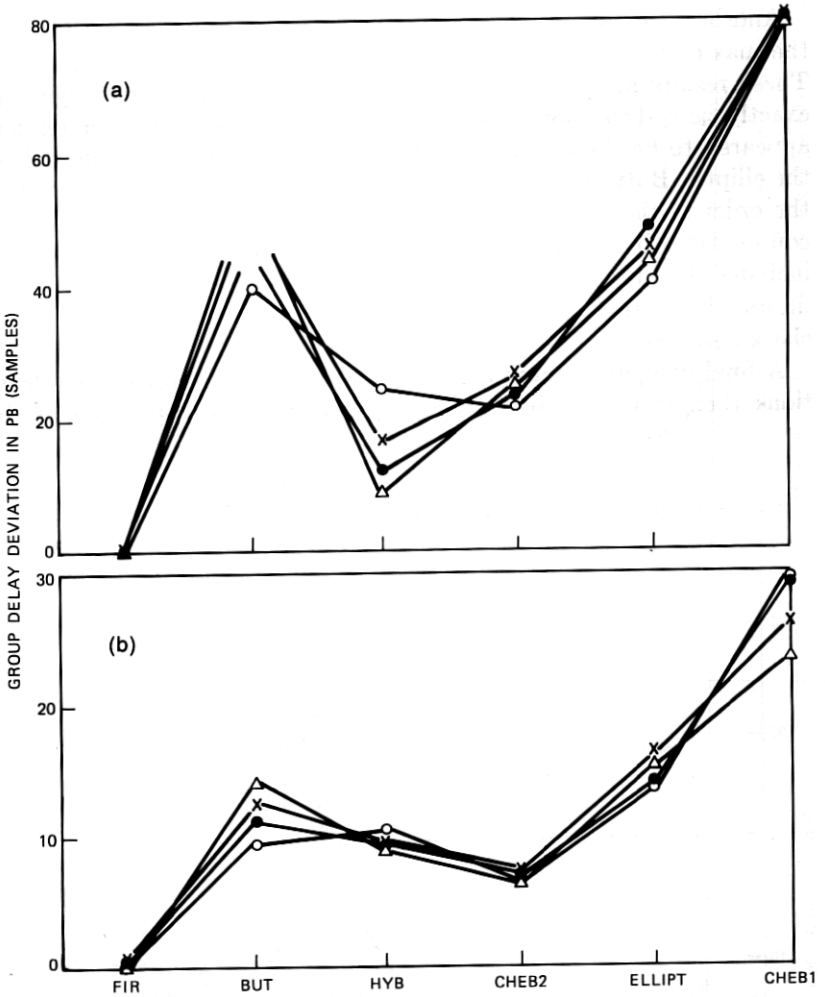


Fig. 10—Comparison of filter designs on the basis of group delay deviation (max.–min.) in the passband.

filters. These results are given in Fig. 11. In this respect, the FIR designs and the hybrid designs did not compare as favorably to the recursive designs.

By cross-comparing the above characteristics, it is possible to obtain a good insight into the various trade-offs involved in each filter design. Obviously, no single approach stands out as being superior over all other designs in all respects. The choice of a given design must be weighted according to the needs of a specific application.

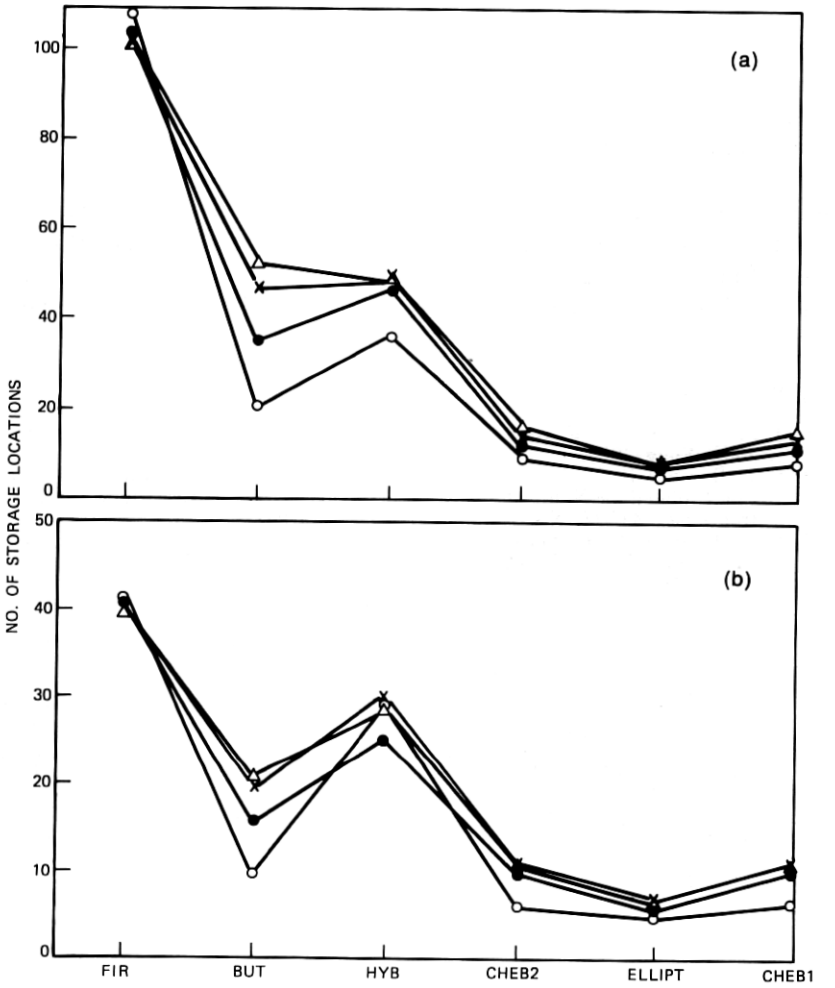


Fig. 11—Comparison of filter designs on the basis of required number of storage locations for state variables.

V. CONCLUSIONS

An algorithmic procedure has been proposed for designing hybrid FIR/IIR digital filters. The procedure is based on the use of a well-known FIR design algorithm for designing the FIR part of the filter, and it is coupled with a well-known optimization algorithm which is used to design the IIR part of the filter. A set of low-pass filters designed in this way were found to have characteristics in between those of optimal FIR designs and conventional IIR designs.

Several properties of the design algorithm have been discussed in detail as well as methods of choosing initial starting points and techniques for speeding up the algorithm. Because of the nature of the FIR/IIR design, it was found that ripples could occur in the transition band. In some examples, these ripples were found to be objectionable, i.e., their amplitude exceeded that of the passband gain although the constraints imposed by the algorithm for the passband and stopband regions were completely met. Thus, an issue which needs further investigation is that of incorporating additional constraints in the algorithm to control the amplitude of the transition band ripples which can occur in such a hybrid filter design.

REFERENCES

1. J. F. Kaiser and F. F. Kuo, *Systems Analysis by Digital Computer*, New York: John Wiley, 1966.
2. L. R. Rabiner, B. Gold, and C. A. McGonegal, "An Approach to the Approximation Problem for Nonrecursive Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-18*, No. 2 (June 1970), pp. 83–106.
3. L. R. Rabiner, J. H. McClellan, and T. W. Parks, "FIR Digital Filter Design Techniques Using Weighted Chebyshev Approximation," Proc. IEEE, *63*, No. 4 (April 1975), pp. 595–610.
4. C. M. Rader and B Gold, "Digital Filter Design Techniques in the Frequency Domain," Proc. IEEE, *55*, No. 2 (February 1967), pp. 149–171.
5. C. S. Burrus and T. W. Parks, "Time Domain Design of Recursive Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-18*, No. 2 (June 1970), pp. 137–141.
6. K. Steiglitz, "Computer-Aided Design of Recursive Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-18*, No. 2 (June 1970), pp. 123–129.
7. A. G. Deczky, "Synthesis of Recursive Digital Filters Using the Minimum P-Error Criterion," IEEE Trans. Audio and Electroacoust., *AU-20*, No. 4 (October 1972), pp. 257–263.
8. L. R. Rabiner, N. Y. Graham, and H. D. Helms, "Linear Programming Design of IIR Digital Filters With Arbitrary Magnitude Function," IEEE Trans. Acoust., Speech, and Signal Processing, *ASSP-22*, No. 2 (April 1974), pp. 117–123.
9. D. E. Dudgeon, "Recursive Filter Design Using Differential Corrections," IEEE Trans. Acoust., Speech, and Signal Processing, *ASSP-22*, No. 6 (December 1974), pp. 443–448.
10. L. R. Rabiner, J. F. Kaiser, O. Herrmann, and M. T. Dolan, "Some Comparisons Between FIR and IIR Digital Filters," B.S.T.J., *53*, No. 2 (February 1974), pp. 305–331.
11. J. H. McClellan, T. W. Parks, and L. R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters," IEEE Trans. Audio and Electroacoust., *AU-21*, No. 6 (December 1973), pp. 506–526.
12. J. L. Kuester and J. H. Mize, *Optimization Techniques With Fortran*, New York: McGraw-Hill, 1973, pp. 309–319.