4 Finite Impulse Response Filter Design

TAPIO SARAMÄKI
Department of Electrical Engineering
Tampere University of Technology
Tampere, Finland

This chapter reviews several design techniques for finite impulse response (FIR) filters along with computationally efficient realization methods. The outline of this chapter is given in Section 4-2 after introducing the filter design problem for both FIR and infinite impulse response (IIR) filters.

4-1 DIGITAL FILTER DESIGN PROBLEM

4-1-1 Digital Filter Design Process

Digital filter design involves usually the following basic steps:

1. Determine a desired response or a set of desired responses (e.g., a desired magnitude response and/or a desired phase response).
2. Select a class of filters for approximating the desired response(s) (e.g., linear-phase FIR filters or IIR filters being implementable as a parallel connection of two allpass filters).
3. Establish a criterion of "goodness" for the response(s) of a filter in the selected class compared to the desired response(s).
4. Develop a method for finding the best member in the filter class.
5. Synthesize the best filter using a proper structure and a proper implementation form, for example using a computer program, a signal processor, or a VLSI chip.
6. Analyze the filter performance.

Handbook for Digital Signal Processing, Edited by Sanjit K. Mitra and James F. Kaiser.
ISBN 0-471-61995-7

In most cases, the desired response is the given magnitude response or the given phase (delay) response or both. The desired magnitude response is usually specified by determining the frequency region(s) where the input signal components should be preserved and the region(s) where the signal components should be rejected. The phase response, in turn, is often desired to be linear in those frequency intervals where the signal components are preserved. In certain cases, time-domain conditions may be included, for example, in the design of Nyquist filters where some of the impulse-response values are restricted to be zero-valued. There are also applications where constraints on the step response are imposed.

The second step consists of determining a proper class of filters to approximate the given response(s). First, it must be decided whether to use FIR filters or IIR filters. After this, a proper class of FIR or IIR filters is selected. For many computationally efficient or low-sensitivity FIR and IIR filter structures, there are constraints on the transfer function. In these cases, the design of the transfer function and the filter implementation cannot be separated, and the desired filter structure determines the class of filters under consideration.

In order to find the best member in the selected filter class, an error measure is needed by which the nearness of the approximating response(s) to the given response(s) is determined. There are several error measures, as will be seen in Section 4-1-3. In many cases, the maximum allowable value of the error measure, for example, the maximum allowable deviation from the given desired response, is specified. In this case, the problem is to first determine the minimum complexity of a filter (e.g., the minimum filter order) required to meet the criteria. The remaining problem is to find the best member in the class of filters with this complexity to minimize the error measure.

The fourth step is to find or develop a method for finding this best member. This chapter and the following chapter describe several design methods for FIR and IIR digital filters. The fifth step involves synthesizing the filter designed at the previous step. The final step is to test whether the resulting filter meets all the given criteria. Also the performance under finite-precision arithmetic is studied. The last two steps are considered in Chapters 6 and 7.

The above design process is often used iteratively. If the resulting filter does not possess all the desired properties, then the desired response(s), the filter class, or the error measure should be changed and the overall process repeated until a filter is obtained with a satisfactory overall performance.

4-1-2 Filter Specifications

The requirements for a digital filter are normally specified in the frequency domain in terms of the desired magnitude response and/or the desired phase (delay) response. In the lowpass case, the desired magnitude response is usually given by

$$D(\omega) = \begin{cases} 1 & \text{for } \omega \in [0, \omega_p] \\ 0 & \text{for } \omega \in [\omega_s, \pi] \end{cases} \tag{4.1}$$

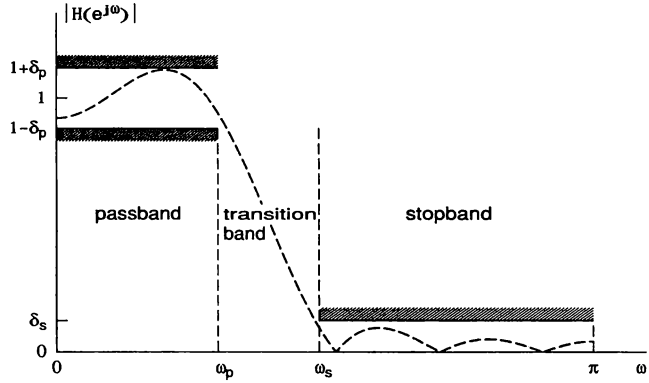


FIGURE 4-1 Tolerance limits for approximation of an ideal lowpass filter magnitude response.

and the specifications are given for the realizable magnitude response $|H(e^{j\omega})|$ as shown in Figure 4-1. It is desired to preserve signal components in the region $[0, \omega_p]$, called the *passband* of the filter, and to reject signal components in the region $[\omega_s, \pi]$, called the *stopband* of the filter. ω_p and ω_s are called, respectively, the passband edge and stopband edge angular frequencies. The permissible errors in the passband and in the stopband are δ_p and δ_s, respectively. The dashed line represents an acceptable magnitude response staying within the limits $1 \pm \delta_p$ in the passband and being less than or equal to δ_s in the stopband. To make it possible to approximate the desired function as close as possible, the specification includes a *transition band* of nonzero width $(\omega_s - \omega_p)$ in which the filter response changes from unity in the passband to zero in the stopband. Note that, because of the symmetry and periodicity of the magnitude response $|H(e^{j\omega})|$, it is sufficient to give the specifications only for $0 \leq \omega \leq \pi$.

Usually, the amplitudes of the allowable ripples for the magnitude response are given logarithmically (i.e., in decibels) in terms of the maximum passband variation and the minimum stopband attenuation, which are given by

$$A_p = 20 \log_{10} \left(\frac{1 + \delta_p}{1 - \delta_p} \right) \text{dB} \tag{4.2a}$$

and

$$A_s = -20 \log_{10} (\delta_s) \text{ dB}, \tag{4.2b}$$

respectively. Note that both these quantities are positive. Another commonly encountered passband specification is the peak deviation from unity expressed logarithmically, that is, $A_p = 20 \log_{10} (\delta_p)$; A_s is then specified as $A_s = 20 \log_{10} (\delta_s)$. These quantities are negative.

Above, the passband and stopband edges, ω_p and ω_s, have been given in terms of the angular frequency ω. If the sampling frequency of the filter is f_s, then ω is related to the real frequency f through the equation

$$\omega = 2\pi f/f_s. \tag{4.3}$$

For instance, if the sampling frequency is 20 kHz and the passband and stopband edges of a lowpass filter are 4 and 5 kHz, then the band edges in terms of the angular frequency become $\omega_p = 0.4\pi$ and $\omega_s = 0.5\pi$. The third alternative is to specify the edges in terms of the *normalized frequency* defined by

$$f_{\text{norm}} = f/f_s. \tag{4.4}$$

In the above example, the normalized passband and stopband edges are 0.2 and 0.25.

In some applications, it is necessary to preserve the shape of the input signal. This goal is achieved if the phase response of the filter is approximately linear in the passband region $[0, \omega_p]$; that is, $\arg H(e^{j\omega})$ approximates on $[0, \omega_p]$ the linear curve

$$\phi(\omega) = -\tau_0\omega + \tau_1, \tag{4.5}$$

where τ_0 and τ_1 can be freely chosen. Instead of the phase response, the criteria for the phase are usually given in terms of the *group delay* response

$$\tau_g(\omega) = -\frac{d \arg H(e^{j\omega})}{d\omega} \tag{4.6a}$$

or the *phase delay* response

$$\tau_p(\omega) = -\frac{\arg H(e^{j\omega})}{\omega}. \tag{4.6b}$$

These responses have simpler representation forms than the phase response and are often easier to interpret. For instance, the value of the phase delay at a specified frequency point $\omega = \omega_0$ gives directly the delay caused by the filter for a sinusoidal signal of frequency ω_0. If the input signal is periodic or approximately periodic, as an electrocardiogram signal, then the phase delay is required to approximate in the passband a constant τ_0 with the given tolerance δ_d as shown in Figure 4-2. Since the delay of all passband signal components is approximately equal, the signal shape is preserved. If the signal is not periodic, then instead of the phase delay, the group delay may be used. Note that for a constant phase delay, τ_1 is forced to be zero in Eq. (4.5),[1] whereas in the case of a constant group delay, τ_1 can take any value.

[1]In the bandpass or highpass case, the shape of a periodic signal is preserved if τ_1 is a multiple of 2π and τ_0 is a constant.

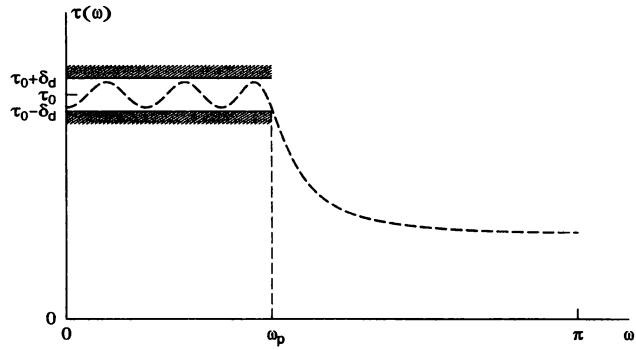


FIGURE 4-2 Tolerance limits for approximation of a constant group or phase delay response.

In the most general case, there are several passbands and stopbands, the desired magnitude response $D(\omega)$ is arbitrary in the passbands, and the allowable approximation error depends on ω in each band. In this case, the specifications can be stated as (see Figure 4-3)

$$D_p(\omega) - e_p(\omega) \leq |H(e^{j\omega})| \leq D_p(\omega) + e_p(\omega) \quad \text{for } \omega \in X_p, \tag{4.7a}$$

$$|H(e^{j\omega})| \leq e_s(\omega) \quad \text{for } \omega \in X_s, \tag{4.7b}$$

where X_p and X_s denote the passband and stopband regions of the filter, respectively. $e_p(\omega)$ is the permissible deviation from the desired passband response $D_p(\omega)$, whereas $e_s(\omega)$ is the allowable deviation from zero in the stopband region. These general specifications can be used, for example, in cases where the input signal of the filter is distorted before or after filtering and it is desired to equalize the am-

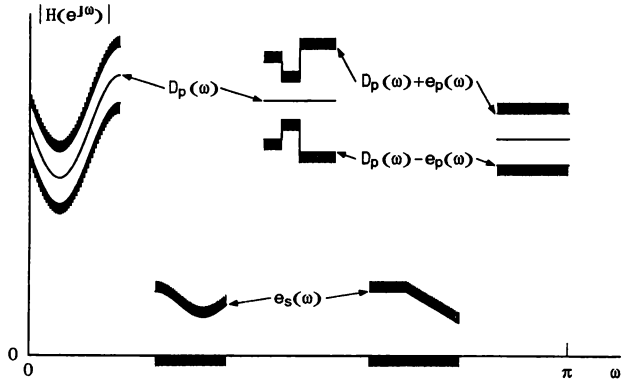


FIGURE 4-3 General specifications for the magnitude response.

plitude distortion. An example of an amplitude distortion occurring after filtering is the one caused by a zero-order hold when reconstructing an analog signal from a discrete-time signal.

The specifications of Eq. (4.7) can be written alternatively in terms of the passband and stopband weighting functions $W_p(\omega)$ and $W_s(\omega)$ as

$$-\delta_p \leq W_p(\omega)[|H(e^{j\omega})| - D_p(\omega)] \leq \delta_p \quad \text{for } \omega \in X_p, \tag{4.8a}$$

$$W_s(\omega)\,|H(e^{j\omega})| \leq \delta_s \quad \text{for } \omega \in X_s. \tag{4.8b}$$

$e_p(\omega)$ is related to δ_p and $W_p(\omega)$ through the equation

$$e_p(\omega) = \delta_p / W_p(\omega), \tag{4.9a}$$

whereas $e_s(\omega)$ is related to δ_s and $W_s(\omega)$ through

$$e_s(\omega) = \delta_s / W_s(\omega). \tag{4.9b}$$

The specifications of Eq. (4.8) can be combined to give the following form, which is useful in many filter design techniques:

$$|E(\omega)| \leq \bar{\epsilon} \quad \text{for } \omega \in X = X_p \cup X_s, \tag{4.10a}$$

where

$$E(\omega) = W(\omega)\,[|H(e^{j\omega})| - D(\omega)] \tag{4.10b}$$

with

$$\bar{\epsilon} = \delta_p, \tag{4.10c}$$

$$D(\omega) = \begin{cases} D_p(\omega) & \text{for } \omega \in X_p \\ 0 & \text{for } \omega \in X_s, \end{cases} \tag{4.10d}$$

and

$$W(\omega) = \begin{cases} W_p(\omega) & \text{for } \omega \in X_p \\ \dfrac{\delta_p}{\delta_s} W_s(\omega) & \text{for } \omega \in X_s. \end{cases} \tag{4.10e}$$

$D(\omega)$ and $W(\omega)$ are called the *desired function* and the *weighting function*, respectively, and $E(\omega)$ is the *weighted error function*. If the maximum absolute value of this function is less than or equal to $\bar{\epsilon}$ on X, then $|H(e^{j\omega})|$ is guaranteed to meet the given criteria.

For instance, in the bandpass case, the specifications are usually stated as

$$1 - \delta_p \leq |H(e^{j\omega})| \leq 1 + \delta_p \quad \text{for } \omega \in [\omega_{p1}, \omega_{p2}], \tag{4.11a}$$

$$|H(e^{j\omega})| \leq \delta_s \quad \text{for } \omega \in [0, \omega_{s1}] \cup [\omega_{s2}, \pi]. \tag{4.11b}$$

These criteria can be written in the above form using

$$X = [0, \omega_{s1}] \cup [\omega_{p1}, \omega_{p2}] \cup [\omega_{s2}, \pi], \tag{4.12a}$$

$$D(\omega) = \begin{cases} 1 & \text{for } \omega \in [\omega_{p1}, \omega_{p2}] \\ 0 & \text{for } \omega \in [0, \omega_{s1}] \cup [\omega_{s2}, \pi], \end{cases} \tag{4.12b}$$

$$W(\omega) = \begin{cases} 1 & \text{for } \omega \in [\omega_{p1}, \omega_{p2}] \\ \delta_p/\delta_s & \text{for } \omega \in [0, \omega_{s1}] \cup [\omega_{s2}, \pi], \end{cases} \tag{4.12c}$$

and

$$\bar{\epsilon} = \delta_p. \tag{4.12d}$$

In a similar manner, the general specifications for the group or phase delay can be stated in terms of the weighted error function as

$$|E_\tau(\omega)| \leq \bar{\epsilon}_\tau \quad \text{for } \omega \in X_p, \tag{4.13a}$$

where

$$E_\tau(\omega) = W_\tau(\omega)[\tau(\omega) - D_\tau(\omega) - \tau_0]. \tag{4.13b}$$

If it is desired to equalize the delay distortion caused by an elliptic filter, then the criteria for an allpass delay equalizer can be written in the above form by selecting $D_\tau(\omega) = -\tau_e(\omega)$, where $\tau_e(\omega)$ is the delay response of the elliptic filter. It should be noted that the actual value of τ_0 in Eq. (4.13b) is not fixed but is an adjustable parameter.

In some cases, it is desired to optimize the frequency-domain behavior of a filter subject to the given time-domain conditions. For instance, in the case of Nyquist or Lth band filters, every Lth impulse-response value is restricted to be zero except for the central value. Furthermore, in some applications, the overshoot of the step response of a digital filter, optimized only in the frequency domain, is too large. In this case, the filter has to be reoptimized with constraints on the ripple of the step response.

4-1-3 Approximation Criteria

Three different error measures are normally used in designing digital filters.

4-1-3-1 Minimax Error Designs. Some applications require that the transfer function coefficients be optimized to minimize the maximum error between the approximating response and the given desired response. The solution minimizing this error function is called a *minimax* or *Chebyshev approximation*. In the case of the weighted error function $E(\omega)$ as given by Eq. (4.10b), the quantity to be minimized is the peak absolute value of $E(\omega)$ on X, that is, the quantity

$$\epsilon = \max_{\omega \in X} |E(\omega)|. \tag{4.14}$$

If the maximum allowable value of ϵ is specified, then the approximation problem is to first find the minimum order of a filter required to meet the given criteria and then optimize the coefficients of a minimum-order transfer function to minimize ϵ. Examples of minimax solutions are elliptic (Cauer) IIR filters and equiripple linear-phase FIR filters.

4-1-3-2 Least-Squared Error Designs. In some cases, instead of the minimax norm, the L_p norm is used. Here, it is desired to minimize the function[2]

$$E_p = \int_X [W(\omega)[|H(e^{j\omega})| - D(\omega)]]^p \, d\omega, \tag{4.15}$$

where p is a positive even integer. It can be shown that as $p \mapsto \infty$, the solution minimizing the above quantity approaches the minimax solution. This fact is exploited in some IIR filter design methods. For FIR filters, L_p error designs are of little practical use since there are efficient algorithms directly available for designing in the minimax sense FIR filters with arbitrary specifications. The exception is the L_2 error or *least-squared error* designs, which can be found very effectively. In this case, the quantity to be minimized is

$$E_2 = \int_X [W(\omega)[|H(e^{j\omega})| - D(\omega)]]^2 \, d\omega. \tag{4.16}$$

4-1-3-3 Maximally Flat Approximations. In the third approach, the approximating response is obtained based on a Taylor series approximation to the given desired response at a certain frequency point and the solution is called a *maximally flat approximation*. In some cases, such as in designing maximally flat (Butterworth) IIR filters, there are two points, one in the passband and one in the stopband, where a Taylor series approximation is applied.

Most of the methods developed for designing digital filters use one of the above approximation criteria. In some synthesis techniques, a combination of these cri-

[2]In the literature, the weighting function $W(\omega)$ is not usually raised to the power of p. If it is desired that the solution minimizing E_p approach the minimax solution as $p \mapsto \infty$, then E_p has to be formed as given by Eq. (4.15).

teria is used. For instance, in the case of Chebyshev IIR filters, a Chebyshev approximation is used in the passband and a maximally flat approximation is used in the stopband.

There exist also several simple filter design techniques that do not use directly the above criteria at all. A typical example of such methods is the design of FIR filters using windows, where the Fourier series of an ideal filter is first truncated and then smoothed using a window function.

4-2 WHY FIR FILTERS?

In many digital signal processing applications, FIR filters are preferred over their IIR counterparts. The main advantages of the FIR filter designs over their IIR equivalents (see Chapter 5 for a review of IIR filter design methods) are the following:

1. FIR filters with exactly linear phase can easily be designed.
2. There exist computationally efficient realizations for implementing FIR filters. These include both nonrecursive and recursive realizations.
3. FIR filters realized nonrecursively are inherently stable and free of limit cycle oscillations when implemented on a finite-wordlength digital system.
4. Excellent design methods are available for various kinds of FIR filters with *arbitrary* specifications.[3]
5. The output noise due to multiplication roundoff errors in an FIR filter is usually very low and the sensitivity to variations in the filter coefficients is also low.

The main disadvantage of conventional FIR filter designs is that they require, especially in applications demanding narrow transition bands, considerably more arithmetic operations and hardware components, such as multipliers, adders, and delay elements than do comparable IIR filters. As the transitions bandwidth of an FIR filter is made narrower, the filter order, and correspondingly the arithmetic complexity, increases inversely proportionally to this width. This makes the implementation of narrow transition band FIR filters very costly. The cost of implementation of an FIR filter can, however, be reduced by using multiplier-efficient realizations, fast convolution algorithms (see Chapter 8), and multirate filtering (see Chapter 14).

This chapter reviews some commonly used methods for designing FIR filters along with several computationally efficient realization methods. The outline of the remaining part of this chapter is as follows. Section 4-3 reviews the properties of linear-phase FIR filters. Sections 4-4 through 4-7 are devoted to very fast design

[3]The design of arbitrary magnitude IIR filters is usually time-consuming and the convergence to the optimum solution is not always guaranteed.

methods. Section 4-4 considers the design of FIR filters based on windowing, Section 4-5 outlines the design of filters that are optimum in the least-mean-square sense and Section 4-6 treats the design of maximally flat filters. Section 4-7 gives some simple analytic design techniques. Section 4-8 is devoted to the design of FIR linear-phase filters that are optimum in the minimax sense. Section 4-9 shows how the design of conventional minimum-phase filters can be accomplished by using a linear-phase filter as a starting point. In Section 4-10, the design of filters having some additional constraints on the frequency- or time-domain response is considered. Finally, Sections 4-11 and 4-12 are devoted to the design of computationally efficient FIR filters. The filters of these two sections are constructed using subfilters as building blocks. In Section 4-12 these subfilters are identical, whereas in Section 4-11 they are different and are obtained from a conventional transfer function by replacing each unit delay element by multiple delays.

4-3 CHARACTERISTICS OF LINEAR-PHASE FIR FILTERS

Some properties of linear-phase FIR filters are reviewed in this section, such as the conditions for linear phase and the zero locations of these filters as well as different representation forms for the frequency response.

4-3-1 Conditions for Linear Phase

Let $\{h[n]\}$ be the impulse response of a causal FIR filter of length $N + 1$. The transfer function of this filter is

$$H(z) = \sum_{n=0}^{N} h[n]z^{-n}. \tag{4.17}$$

The corresponding frequency response is given by

$$H(e^{j\omega}) = \sum_{n=0}^{N} h[n]e^{-jn\omega}. \tag{4.18}$$

In the above, N is the order of the filter.

For many practical FIR filters, exact linearity of phase is often desired. This goal is achieved if the frequency response of the filter is expressible in the form

$$H(e^{j\omega}) = \overline{H}(\omega)e^{j\phi(\omega)}, \tag{4.19a}$$

where

$$\phi(\omega) = \alpha\omega + \beta \tag{4.19b}$$

and $\overline{H}(\omega)$ is a real and even function of ω. The magnitude and the phase of the above function are, respectively,

$$|H(e^{j\omega})| = |\overline{H}(\omega)| \tag{4.20a}$$

and

$$\arg H(e^{j\omega}) = \begin{cases} \alpha\omega + \beta & \text{for } \overline{H}(\omega) \geq 0 \\ \alpha\omega + \beta - \pi & \text{for } \overline{H}(\omega) < 0. \end{cases} \tag{4.20b}$$

$\overline{H}(\omega)$ is called the *zero-phase frequency response*[4] to distinguish it from the magnitude response $|H(e^{j\omega})|$. To simplify the notation, let $H(\omega)$ represent the zero-phase frequency response. It should always be clear from the context whether H is a function of z, $e^{j\omega}$, or ω, that is, whether the transfer function, the frequency response, or the zero-phase frequency response is considered. The relationships between $H(\omega)$ and $|H(e^{j\omega})|$, and between $\phi(\omega)$ and $\arg H(e^{j\omega})$, are shown in Figure 4-4. Note that the zero-phase frequency response of the filter may take both positive and negative values, whereas the magnitude response is strictly nonnegative.

There are the following four types yielding the phase linearity:

Type I: N is even and $h[N - n] = h[n]$ for all n.

Type II: N is odd and $h[N - n] = h[n]$ for all n.

Type III: N is even and $h[N - n] = -h[n]$ for all n.

Type IV: N is odd and $h[N - n] = -h[n]$ for all n.

[4]Some authors call $\overline{H}(\omega)$ the amplitude response of the filter.

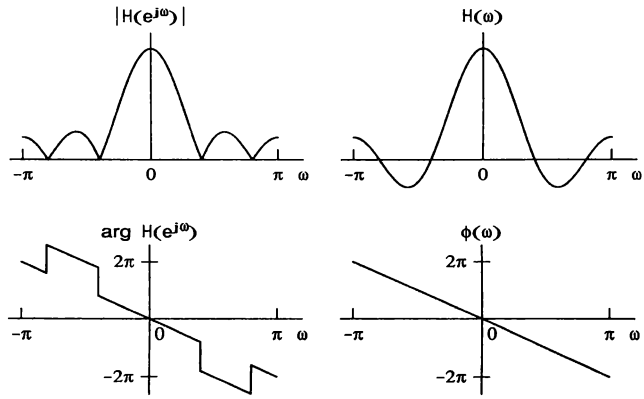


FIGURE 4-4 Relations between the magnitude response $|H(e^{j\omega})|$ and the zero-phase frequency response $H(\omega)$, and between $\arg H(e^{j\omega})$ and $\phi(\omega)$ for an example linear-phase FIR filter.

In each of these four cases, the transfer function is expressible as

$$H(z) = F(z)G(z), \tag{4.21a}$$

where

$$F(z) = \begin{cases} 1 & \text{for Type I} \\ [1 + z^{-1}]/2 & \text{for Type II} \\ [1 - z^{-2}]/2 & \text{for Type III} \\ [1 - z^{-1}]/2 & \text{for Type IV} \end{cases} \tag{4.21b}$$

and

$$G(z) = \sum_{n=0}^{2M} g[n]z^{-n} \tag{4.21c}$$

with

$$g[2M - n] = g[n] \quad \text{for all } n \tag{4.21d}$$

and

$$M = \begin{cases} N/2 & \text{for Type I} \\ (N - 1)/2 & \text{for Type II} \\ (N - 2)/2 & \text{for Type III} \\ (N - 1)/2 & \text{for Type IV.} \end{cases} \tag{4.21e}$$

Hence $H(z)$ can be expressed as a cascade of a fixed term $F(z)$ and a common adjustable term $G(z)$, which itself is a Type I transfer function. The relations between $h[n]$ and $g[n]$ are given in Table 4-1 for the four types. Figure 4-5 shows example impulse responses. In each case, the center of the symmetry is $K = N/2$. For Types I and III, K is an integer and there is an impulse-response sample exactly at this point, whereas for Types II and IV, K is not an integer and it lies between two impulse-response samples. Note that for Type III, the symmetry forces $h[N/2]$ to be equal to zero.

Because of the symmetry property of Eq. (4.21d), $G(z)$ can be expressed as

$$\begin{aligned} G(z) = z^{-M}\,[&g[M] + g[M - 1](z + z^{-1}) \\ &+ g[M - 2](z^2 + z^{-2}) + \cdots + g[0](z^M + z^{-M})]. \end{aligned} \tag{4.22}$$

TABLE 4-1 Relations Between the Coefficients $h[n]$ and $g[n]$

Coefficient	Type I	Type II	Type III	Type IV
$h[0]$	$g[0]$	$\frac{g[0]}{2}$	$\frac{g[0]}{2}$	$\frac{g[0]}{2}$
$h[1]$	$g[1]$	$\frac{g[1] + g[0]}{2}$	$\frac{g[1]}{2}$	$\frac{g[1] - g[0]}{2}$
$h[n]$	$g[n]$	$\frac{g[n] + g[n-1]}{2}$	$\frac{g[n] - g[n-2]}{2}$	$\frac{g[n] - g[n-1]}{2}$
$h[N-1]$	$g[N-1]$	$\frac{g[N-1] + g[N-2]}{2}$	$\frac{-g[N-3]}{2}$	$\frac{g[N-1] - g[N-2]}{2}$
$h[N]$	$g[N]$	$\frac{g[N-1]}{2}$	$-\frac{g[N-2]}{2}$	$-\frac{g[N-1]}{2}$

By substituting $z = e^{j\omega}$ in the above equation, the frequency response of $G(z)$ becomes

$$G(e^{j\omega}) = e^{-jM\omega}[g[M] + g[M-1](2\cos\omega) + g[M-2](2\cos 2\omega) + \cdots + g[0](2\cos M\omega)]. \quad (4.23)$$

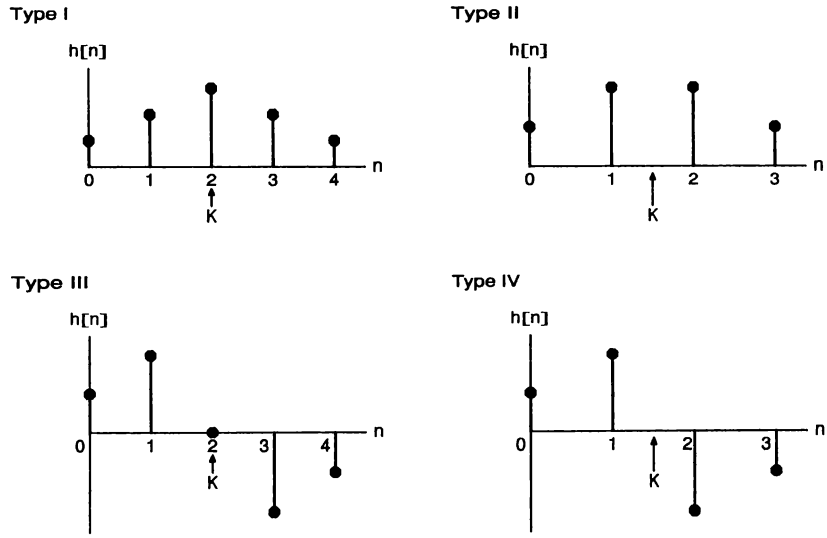


FIGURE 4-5 Example impulse responses for the four different linear-phase types.

Similarly, the frequency response of $F(z)$ can be written, after some manipulations, in the form

$$F(e^{j\omega}) = \begin{cases} 1 & \text{for Type I} \\ e^{-j\omega/2} \cos(\omega/2) & \text{for Type II} \\ e^{-j(\omega-\pi/2)} \sin\omega & \text{for Type III} \\ e^{-j(\omega/2-\pi/2)} \sin(\omega/2) & \text{for Type IV.} \end{cases} \tag{4.24}$$

By combining the above results, the zero-phase frequency response can be expressed as

$$H(\omega) = F(\omega)G(\omega), \tag{4.25}$$

where

$$F(\omega) = \begin{cases} 1 & \text{for Type I} \\ \cos(\omega/2) & \text{for Type II} \\ \sin\omega & \text{for Type III} \\ \sin(\omega/2) & \text{for Type IV,} \end{cases} \tag{4.26}$$

$$G(\omega) = \sum_{n=0}^{M} a[n] \cos n\omega, \tag{4.27a}$$

$$a[n] = \begin{cases} g[M] & \text{for } n = 0 \\ 2g[M-n] & \text{for } n \neq 0, \end{cases} \tag{4.27b}$$

and M is given by Eq. (4.21e). The phase term becomes

$$\phi(\omega) = \begin{cases} -N\omega/2 & \text{for Types I and II} \\ -N\omega/2 + \pi/2 & \text{for Types III and IV.} \end{cases} \tag{4.28}$$

In Section 4-3-3, some other useful representation forms of $H(\omega)$ are given. Figure 4-6 gives example zero-phase frequency responses for the four types. For Type I, $H(\omega)$ is even about $\omega = 0$ and $\omega = \pi$ and the periodicity is 2π. For Type II, the fixed term $F(\omega) = \cos(\omega/2)$ generates a zero for $H(\omega)$ at $\omega = \pi$, making it odd about this point. The periodicity is 4π. Similarly, for Type IV, the fixed term generates a zero at $\omega = 0$. The resulting $H(\omega)$ is odd about $\omega = 0$ and the periodicity is 4π. For Type III, the fixed term gives a zero at both $\omega = 0$ and $\omega = \pi$, making $H(\omega)$ odd about these points. The periodicity is 2π.

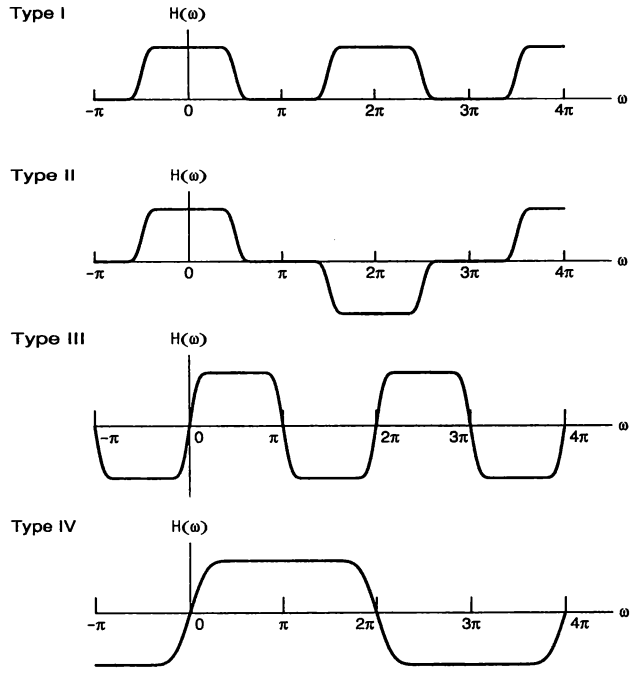


FIGURE 4-6 Example zero-phase frequency responses for the four different linear-phase types.

This chapter concentrates mainly on Types I and II. These filter types are used for conventional filtering applications because, in these cases, the delay caused for sinusoidal signals, $-\phi(\omega)/\omega = N/2$, is independent of the frequency ω. Filters belonging to the remaining two cases have an additional 90° phase shift and they are most suitable for realizing such filters as differentiators and Hilbert transformers (see Chapter 13). In these two cases, the delay caused for sinusoidal signals depends on the frequency. However, the group delay, $-d\phi(\omega)/d\omega$, is a constant (equal to $N/2$ for all linear-phase types).

The above linear-phase filters are also characterized by the property that only $M + 1$ multipliers are needed in the actual implementation because of the symmetry in the filter coefficients. Figure 4-7 gives such an implementation for a Type I filter of order $N = 2M$.

4-3-2 Zero Locations of Linear-Phase FIR Filters

All the poles of an FIR filter lie at the origin. From Eq. (4.21b), it is seen that the fixed term $F(z)$ generates for Type II designs one zero at $z = -1$, for Type III designs one zero at both $z = 1$ and $z = -1$, and for Type IV designs one zero at

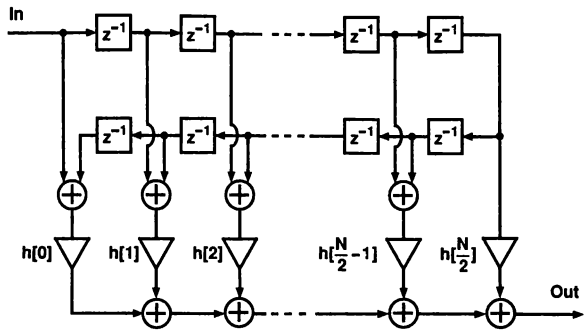


FIGURE 4-7 Implementation for a Type I filter exploiting the symmetry in the filter coefficients.

$z = 1$. What remains is to examine where the zeros of the common adjustable Type I part $G(z)$ are located. From the symmetry condition of Eq. (4.21d), it follows that

$$G(z^{-1}) = z^{2M}G(z). \tag{4.29}$$

This means that $G(z)$ and $G(z^{-1})$ have identical zeros. The zeros of $G(z)$ thus occur in mirror-image pairs. For $G(z)$ with real coefficients, the zeros are either real or occur in complex conjugate pairs. These conditions imply that $G(z)$ can be factored in the form

$$G(z) = g[0]G_1(z)G_2(z)G_3(z), \tag{4.30a}$$

where

$$G_1(z) = \prod_{i=1}^{N_1} \left(1 - \left[2\left(r_i + \frac{1}{r_i}\right)\cos\theta_i\right]z^{-1} + \left[r_i^2 + \frac{1}{r_i^2} + 4\cos^2\theta_i\right]z^{-2} - \left[2\left(r_i + \frac{1}{r_i}\right)\cos\theta_i\right]z^{-3} + z^{-4}\right), \tag{4.30b}$$

$$G_2(z) = \prod_{i=1}^{N_2} (1 - [2\cos\bar{\theta}_i]z^{-1} + z^{-2}), \tag{4.30c}$$

and

$$G_3(z) = \prod_{i=1}^{N_3} \left(1 - \left[\bar{r}_i + \frac{1}{\bar{r}_i}\right]z^{-1} + z^{-2}\right). \tag{4.30d}$$

Here, $4N_1 + 2(N_2 + N_3) = 2M$ and

1. $G_1(z)$ contains the zeros occurring in quadruplets, that is, in complex conjugate and mirror-image pairs off the unit circle at $z = r_i e^{\pm j\theta_i}$, $(1/r_i)e^{\pm j\theta_i}$ for $i = 1, 2, \cdots, N_1$.
2. $G_2(z)$ contains the zeros occurring in complex conjugate pairs on the unit circle at $z = e^{\pm j\theta_i}$ for $i = 1, 2, \cdots, N_2$.
3. $G_3(z)$ contains the zeros occurring in reciprocal pairs on the real axis at $z = \bar{r}_i$, $1/\bar{r}_i$ for $i = 1, 2, \cdots, N_3$.

If $G(z)$ possesses a zero at $z = 1$ or at $z = -1$, then it follows from the symmetry of $G(z)$ and the fact that $G(z)$ is of even order that the number of zeros at this point must be even. If $G(z)$ happens to have k zero-pairs at $z = 1$ (at $z = -1$), then these pairs can be included in $G_3(z)$ by using k terms with $\bar{r}_i = 1$ ($\bar{r}_i = -1$).

For Types II, III, and IV, the locations of the zeros outside the points $z = 1$ and $z = -1$ are similar. The main difference between the four cases is in the number of zeros at $z = 1$ and $z = -1$.

1. Type I designs have either an even number or no zeros at $z = 1$ and at $z = -1$.
2. Type II designs have either an even number or no zeros at $z = 1$, and an odd number of zeros at $z = -1$.
3. Type III designs have an odd number of zeros at $z = 1$ and at $z = -1$.
4. Type IV designs have an odd number of zeros at $z = 1$, and either an even number or no zeros $z = -1$.

Figure 4-8 gives the amplitude response, the impulse response, and the zero locations of a typical Type I filter.

4-3-3 Different Representation Forms for Zero-Phase Frequency Responses

Equations (4.25)–(4.27) give one representation form for the zero-phase frequency response in each of the four different linear-phase cases. This form is used in designing filters in the minimax sense (Section 4-8). Another useful representation form is obtained by expressing $H(\omega)$ directly in terms of the impulse-response coefficients $h[n]$ as follows [OP89; RA75a]:

$$H(\omega) = \begin{cases} h[N/2] + \displaystyle\sum_{n=1}^{N/2} h[N/2 - n][2\cos n\omega] & \text{for Type I} \\ \displaystyle\sum_{n=0}^{(N-1)/2} h[(N-1)/2 - n][2\cos[(n + 1/2)\omega]] & \text{for Type II} \\ \displaystyle\sum_{n=0}^{N/2-1} h[N/2 - 1 - n][2\sin[(n+1)\omega]] & \text{for Type III} \\ \displaystyle\sum_{n=0}^{(N-1)/2} h[(N-1)/2 - n][2\sin[(n + 1/2)\omega]] & \text{for Type IV.} \end{cases} \tag{4.31}$$

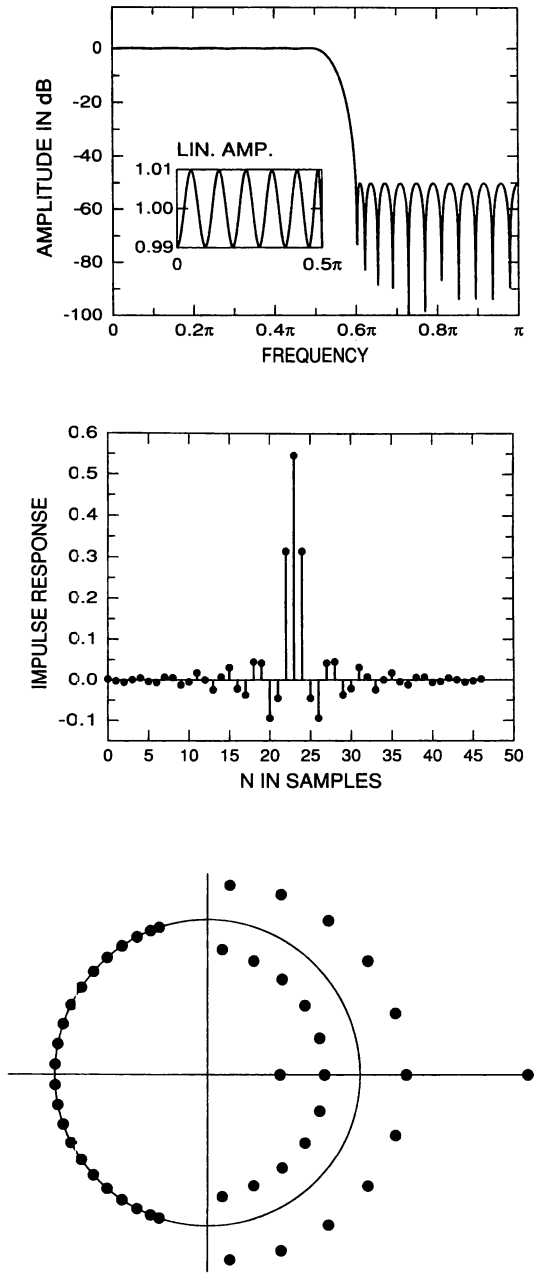


FIGURE 4-8 Amplitude response, impulse response, and zero locations for a typical Type I filter of order $N = 46$.

For later use, we rewrite the above equation in the form

$$H(\omega) = \sum_{n=0}^{M} b[n] \text{ trig } (\omega, n), \tag{4.32a}$$

where

$$\text{trig } (\omega, n) = \begin{cases} 1 & \text{for Type I; } n = 0 \\ 2 \cos n\omega & \text{for Type I; } n > 0 \\ 2 \cos [(n + 1/2)\omega] & \text{for Type II} \\ 2 \sin [(n + 1)\omega] & \text{for Type III} \\ 2 \sin [(n + 1/2)\omega] & \text{for Type IV,} \end{cases} \tag{4.32b}$$

$$b[n] = \begin{cases} h[N/2 - n] & \text{for Type I} \\ h[(N - 1)/2 - n] & \text{for Types II and IV} \\ h[N/2 - 1 - n] & \text{for Type III,} \end{cases} \tag{4.32c}$$

and M is related to N through Eq. (4.21e). This representation form is used in Section 4-5 for designing filters in the least-mean-square sense and in Section 4-10 for designing filters based on linear programming.

A very useful representation form for the common adjustable response part $G(\omega)$ as given by Eq. (4.27) or for the overall response $H(\omega)$ for Type I [$H(\omega) = G(\omega)$] can be derived based on the identity

$$\cos n\omega = T_n(\cos \omega), \tag{4.33}$$

where $T_n(x) = \cos (n \cos^{-1} x)$ is the nth degree Chebyshev polynomial. These polynomials can be generated using the following recursion formulas:

$$T_0(x) = 1, \tag{4.34a}$$

$$T_1(x) = x, \tag{4.34b}$$

$$T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x). \tag{4.34c}$$

Using these equivalences, $\cos n\omega$ can be expressed as an nth degree polynomial in $\cos \omega$ and $G(\omega)$ as an Mth degree polynomial in $\cos \omega$:

$$G(\omega) = \sum_{n=0}^{M} \alpha[n] \cos^n \omega. \tag{4.35}$$

This shows that the zero-phase frequency response of a Type I filter of order $2M$ can be determined as an Mth degree polynomial in $\cos \omega$. This fact is exploited in several synthesis techniques.

4-4 FIR FILTER DESIGN BY WINDOWING

The most straightforward approach to designing FIR filters is to determine the infinite-duration impulse response by expanding the frequency response of an ideal filter in a Fourier series and then to truncate and smooth this response using a window function. The main advantage of this design technique is that the impulse-response coefficients can be obtained in closed form and can be determined very fast even using a calculator. The main drawback is that the passband and stopband ripples of the resulting filter are restricted to be approximately equal.

4-4-1 Design Process

FIR filter design based on windowing generally begins by specifying the ideal zero-phase frequency response $H_{id}(\omega)$. Ideal Type I responses in the lowpass, highpass, bandpass, and bandstop cases are shown in Figure 4-9. Since $H_{id}(\omega)$ is even about $\omega = 0$ and periodic in ω with period 2π, it can be expanded in a Fourier series as follows:

$$H_{id}(\omega) = h_{id}^{(0)}[0] + 2\sum_{n=1}^{\infty} h_{id}^{(0)}[n] \cos n\omega, \tag{4.36a}$$

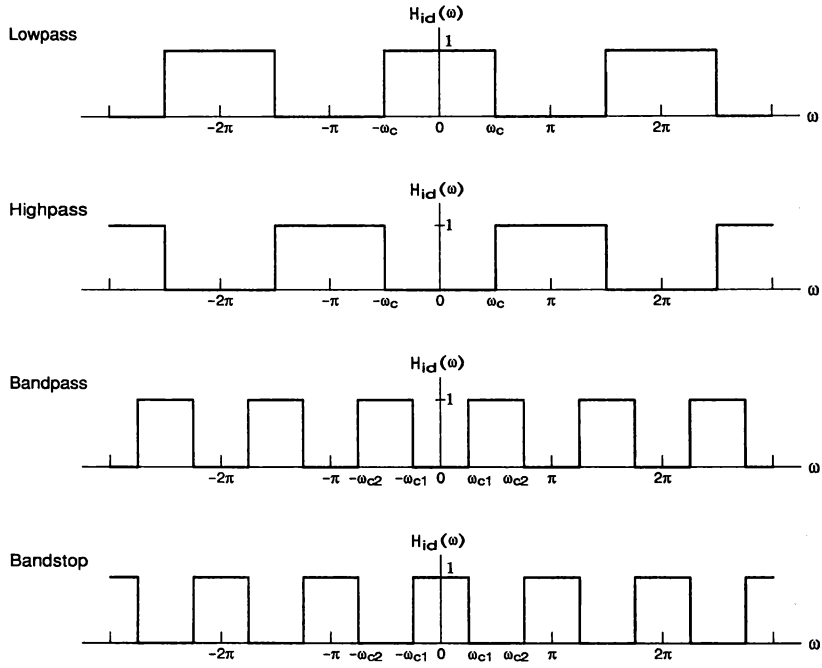


FIGURE 4-9 Zero-phase frequency responses of ideal Type I filters.

where

$$h_{\text{id}}^{(0)}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} H_{\text{id}}(\omega)\cos(n\omega)\,d\omega, \qquad 0 \le n \le \infty. \tag{4.36b}$$

The corresponding impulse response is of infinite duration and even about $n = 0$; that is, $h_{\text{id}}^{(0)}[-n] = h_{\text{id}}^{(0)}[n]$ (see Figure 4-10). For instance, in the lowpass case

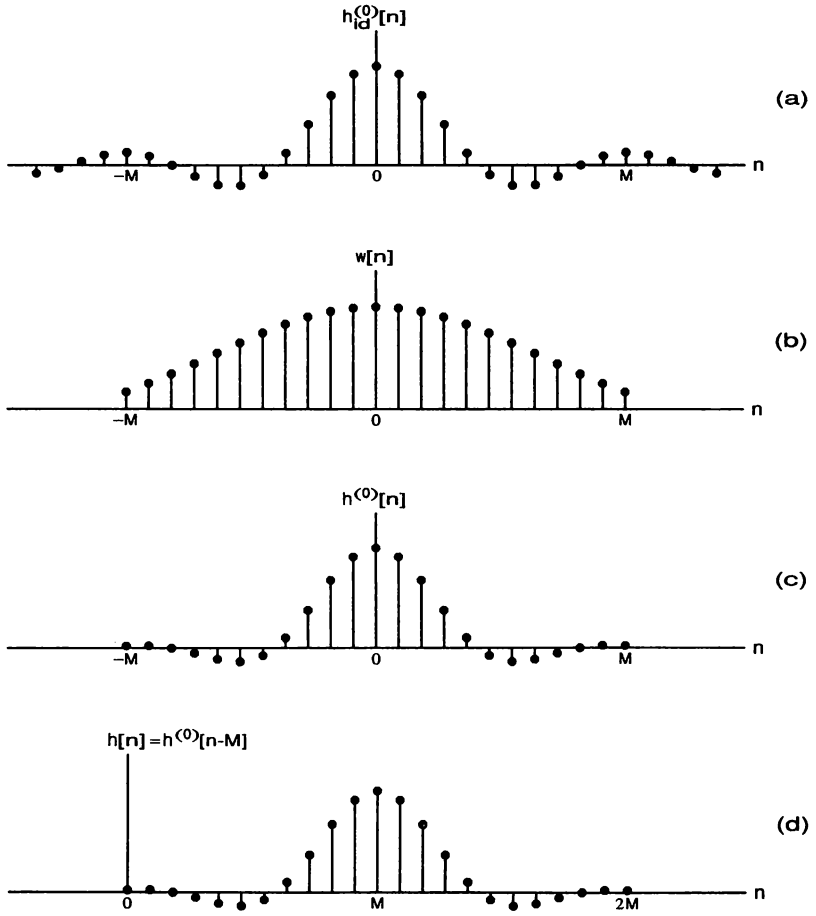


FIGURE 4-10 Impulse responses involved in designing an FIR filter by windowing. (a) Infinite-duration impulse response for an ideal zero-phase filter. (b) Impulse response for a window function. (c) Response obtained by truncating and smoothing the ideal response by the window function. (d) Response for the corresponding causal filter.

with cutoff edge ω_c, the coefficients are

$$h_{\text{id}}^{(0)}[n] = \frac{\omega_c}{\pi}\left(\frac{\sin(\omega_c n)}{\omega_c n}\right) = \begin{cases} \omega_c/\pi, & n = 0 \\ \sin(\omega_c n)/(\pi n), & |n| > 0. \end{cases} \tag{4.37}$$

The coefficients for highpass, bandpass, and bandstop filters are given in Table 4-2.

An approximating finite-duration impulse response is then generated by truncating and smoothing the above response according to

$$h^{(0)}[n] = \mathrm{w}[n]h_{\text{id}}^{(0)}[n], \tag{4.38}$$

where $\mathrm{w}[n]$ is a *window function* that is nonzero for $-M \le n \le M$ and zero otherwise (see Figure 4-10). Finally, the coefficients of the corresponding causal realizable FIR filter of order $2M$ are obtained by shifting the location of the central impulse-response coefficient from $n = 0$ to $n = M$, giving

$$h[n] = h^{(0)}[n - M], \qquad 0 \le n \le 2M. \tag{4.39}$$

TABLE 4-2 Coefficients of Ideal Zero-Phase Type I Filters

Type	Coefficients
Lowpass filter with edge	$h_{\text{id}}^{(0)}[0] = \dfrac{\omega_c}{\pi}$
angle ω_c	$h_{\text{id}}^{(0)}[n] = \dfrac{\sin(\omega_c n)}{\pi n}, \quad \lvert n\rvert > 0$
Highpass filter with edge	$h_{\text{id}}^{(0)}[0] = 1 - \dfrac{\omega_c}{\pi}$
angle ω_c	$h_{\text{id}}^{(0)}[n] = -\dfrac{\sin(\omega_c n)}{\pi n}, \quad \lvert n\rvert > 0$
Bandpass filter with edge	$h_{\text{id}}^{(0)}[0] = \dfrac{(\omega_{c2} - \omega_{c1})}{\pi}$
angles ω_{c1} and ω_{c2}	$h_{\text{id}}^{(0)}[n] = \dfrac{1}{\pi n}[\sin(\omega_{c2} n) - \sin(\omega_{c1} n)], \quad \lvert n\rvert > 0$
Bandstop filter with edge	$h_{\text{id}}^{(0)}[0] = 1 - \dfrac{(\omega_{c2} - \omega_{c1})}{\pi}$
angles ω_{c1} and ω_{c2}	$h_{\text{id}}^{(0)}[n] = \dfrac{1}{\pi n}[\sin(\omega_{c1} n) - \sin(\omega_{c2} n)], \quad \lvert n\rvert > 0$

For Type III filters, $H_{id}(\omega)$ (e.g., the ideal response of a Hilbert transformer or differentiator) is odd about $\omega = 0$ (cf. Figure 4-6) and the Fourier series contains sine terms, instead of cosine terms, and $h_{id}^{(0)}[0]$ is absent. In this case, $h_{id}^{(0)}[-n] = -h_{id}^{(0)}[n]$ and the terms in the series are $2h_{id}^{(0)}[-n] \sin n\omega$, where $h_{id}^{(0)}[-n]$ can be determined from Eq. (4.36b) by replacing cos $(n\omega)$ by sin $(n\omega)$.

For Type II, $H_{id}(\omega)$ is odd about $\omega = \pi$ and the periodicity is 4π (see Figure 4-11(a)). The design of a Type II filter of odd-order N to approximate this response can be performed by first applying the above process with $M = N$ to the response $H_{id}(2\omega)$, which is a Type I response [PA87] (see Figure 4-11(b)). This gives an FIR filter of order $2N$. Since $H_{id}(2\omega)$ is odd about $\omega = \pi/2$, $h_{id}^{(0)}[n] = 0$ for $n = 0, \pm 2, \pm 4, \cdots$. Correspondingly, $h[N \pm 2r] = 0$ for $r = 0, 1, \cdots, (N - 1)/2$. The Type II filter whose zero-phase frequency response approximates the ideal response of Figure 4-11(a) is then obtained by discarding these zero-valued impulse-response samples, resulting in a filter of order N. The design of Type IV filters can be converted into the design of Type III filters in the same manner [PA87].

4-4-2 Direct Truncation of an Ideal Impulse Response

The simplest window is the *rectangular window* for which

$$w[n] = \begin{cases} 1, & -M \le n \le M \\ 0, & \text{otherwise.} \end{cases} \tag{4.40}$$

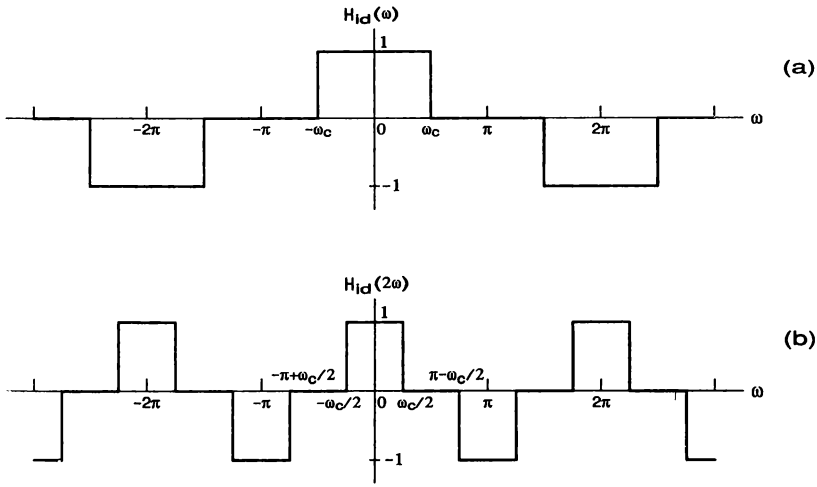


FIGURE 4-11 (a) Ideal Type II lowpass response $H_{id}(\omega)$. (b) Response $H_{id}(2\omega)$, which can be approximated by Type I filters.

The use of this window corresponds to a direct truncation of the infinite-duration impulse response and leads to a solution exhibiting large ripples before and after the discontinuity of the ideal frequency response. This is the well-known *Gibbs phenomenon*. As an example, Figure 4-12 gives the resulting responses $H(\omega)$ for Type I lowpass filters with $\omega_c = 0.4\pi$ for $M = 10$ and $M = 30$. The corresponding filter orders ($N = 2M$) are 20 and 60, respectively. As seen from this figure, the transition bandwidth of $H(\omega)$ becomes narrower when M is increased, but the maximum ripples in the passband and stopband regions remain about the same. In both cases, the first stopband extremum has the value of -0.09 (21-dB attenuation) and the last passband extremum has the value of 1.09.

The Gibbs phenomenon can be explained by the fact that $H(\omega)$ is related to the ideal response $H_{id}(\omega)$ and the frequency response of the window function

$$\Psi(\omega) = \sum_{n=-M}^{M} w[n] e^{-jn\omega} = w[0] + 2 \sum_{n=1}^{M} w[n] \cos n\omega \tag{4.41}$$

through

$$H(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_{id}(\theta)\, \Psi(\omega - \theta)\, d\theta. \tag{4.42}$$

For the rectangular window,

$$\Psi(\omega) = \sum_{n=-M}^{M} e^{-jn\omega} = \frac{\sin((2M+1)\omega/2)}{\sin(\omega/2)}. \tag{4.43}$$

This response is depicted in Figure 4-13 for $M = 10$ and $M = 30$. As seen from this figure, $\Psi(\omega)$ appears as a gradually decaying sinusoid starting at a middle

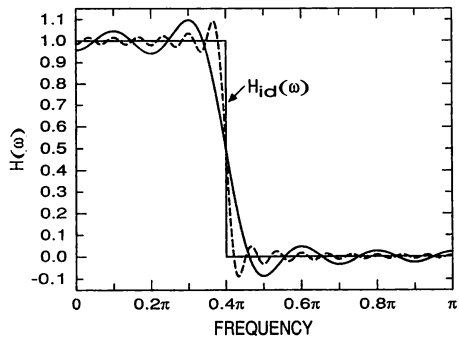


FIGURE 4-12 Responses of Type I lowpass filters designed by truncating the impulse response of an ideal filter with $\omega_c = 0.4\pi$. The solid and dashed lines give the responses for $M = 10$ and $M = 30$, respectively.

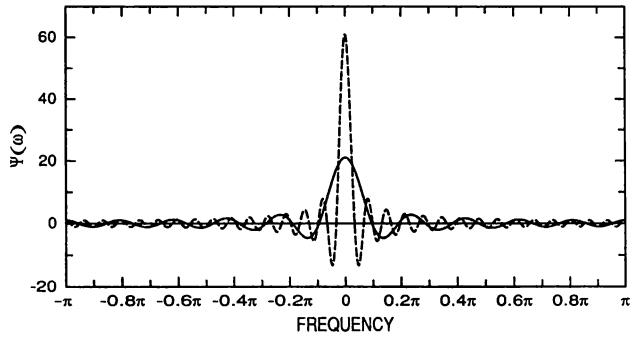


FIGURE 4-13 Frequency responses for the rectangular window for $M = 10$ (solid line) and $M = 30$ (dashed line).

lobe, called the *mainlobe*, whose width is twice that of the *sidelobes* being situated in the intervals between the zeros.

According to Eq. (4.42), the value of $H(\omega)$ at any frequency point ω is obtained in the lowpass case with cutoff edge ω_c by integrating $\Psi(\omega - \theta)$ with respect to θ over the interval $[-\omega_c, \omega_c]$. This is illustrated in Figure 4-14. For $\omega = \pi$, only small ripples of $\Psi(\omega - \theta)$ are inside this interval, resulting in a small value of

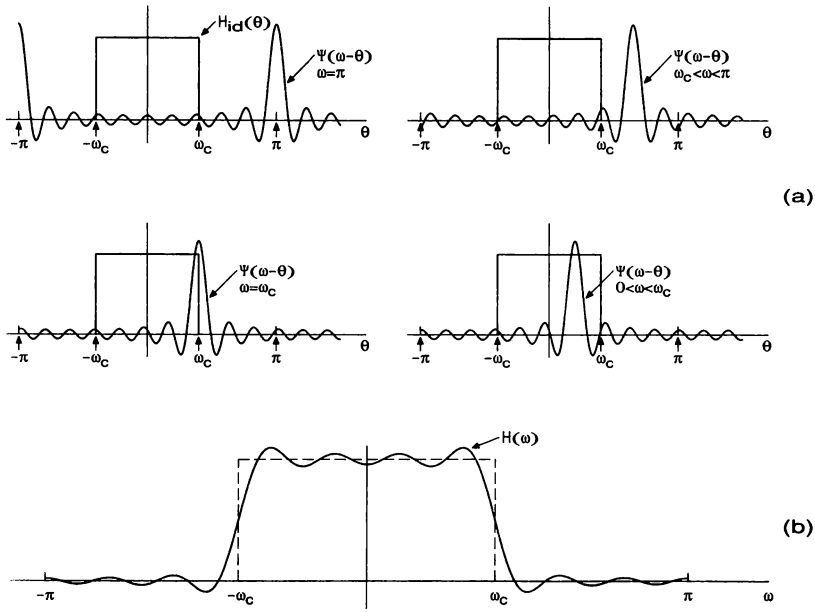


FIGURE 4-14 Explanation of the Gibbs phenomenon. (a) Convolution process. (b) Response for the resulting filter.

$H(\omega)$ at $\omega = \pi$. As ω is made smaller, larger ripples of $\Psi(\omega - \theta)$ are entering into the interval, resulting in larger values in $H(\omega)$ for $\omega < \pi$. The ripples are due to the fact that the area under every second sidelobe of $\Psi(\omega)$ is of opposite sign. For $\omega = \omega_c$, half of the mainlobe is inside the interval $[-\omega_c, \omega_c]$. Since the integral of $\Psi(\omega)$ over the interval $[-\pi, \pi]$ is one and most of the energy is concentrated in the mainlobe, the value of $H(\omega)$ at $\omega = \omega_c$ is approximately $\frac{1}{2}$. When ω is further decreased, the whole mainlobe enters the interval and the area in this interval is approximately one, resulting in the passband response of $H(\omega)$. The ripples around one are due to the fact that the sidelobes of $\Psi(\omega - \theta)$, which are of different heights, go inside the interval $[-\omega_c, \omega_c]$ and leave it as ω varies.

As M is increased, the widths of the mainlobe and the sidelobes decrease. However, the area under each lobe remains the same since at the same time the heights of the lobes increase (see Figure 4-13). This means that as M is increased, the oscillations of the resulting filter response occur more rapidly but do not decrease (see Figure 4-12).

4-4-3 Fixed Window Functions

The Gibbs phenomenon can be reduced by using a less abrupt truncation of the Fourier series. This is achieved by using a window function that tapers smoothly towards zero at both ends. Some of the well-known fixed window functions $w[n]$ [BL58; HA78; HA87; KA63, KA66; RA75a] are summarized in Table 4-3 along with their frequency responses $\Psi(\omega)$.[5] For these fixed window functions, the only adjustable parameter is M, half the filter order. The plots of the frequency responses are given in Figure 4-15 for the last four windows in Table 4-3 for $M = 128$. Also, the responses of the filters resulting when using these windows for $\omega_c = 0.4\pi$ are shown in this figure.

Figure 4-16 depicts, in the lowpass case, a typical relation between $H(\omega)$ and $\Psi(\omega)$, which is given in terms of $\theta - \omega_c$ in order to center the response at the cutoff edge. Note the close similarity to the case where $\Psi(\omega)$, $H_{id}(\omega)$, and $H(\omega)$ correspond to the impulse response, the step excitation, and the response of a continuous-time filter, respectively. As seen from the figure, $H(\omega)$ satisfies approximately $H(\omega_c + \omega) + H(\omega_c - \omega) = 1$ in the vicinity of the cutoff edge ω_c. This means that $H(\omega_c) \approx \frac{1}{2}$. Furthermore, the maximum passband deviation from unity and the maximum stopband deviation from zero are about the same, and the peak passband overshoot $(1 + \delta)$ and the peak negative stopband undershoot $(-\delta)$ occur at the same distance from the discontinuity point ω_c. The distance between these two overshoot points is for most windows approximately equal to the mainlobe width Δ_M. The criteria met by $H(\omega)$ can be given by

$$1 - \delta \leq H(\omega) \leq 1 + \delta \quad \text{for } \omega \in [0, \omega_p], \tag{4.44a}$$

$$-\delta \leq H(\omega) \leq \delta \quad \text{for } \omega \in [\omega_s, \pi], \tag{4.44b}$$

[5]The definitions of the window functions of Table 4-3 differ slightly in the literature.

TABLE 4-3 Some Commonly Used Fixed Windows for FIR Filter Design

Window Type	Window Function, $w[n]$, $-M \le n \le M$	Frequency Response, $\Psi(\omega)$
Rectangular	1	$\Psi_R(\omega) \equiv \sin[(2M+1)\omega/2]/\sin(\omega/2)$
Bartlett	$1 - \dfrac{\lvert n\rvert}{M+1}$	$\dfrac{1}{M+1}[\sin[(M+1)\omega/2]/\sin(\omega/2)]^2$
Hann	$\frac{1}{2}\left[1 + \cos\left[\dfrac{2\pi n}{2M+1}\right]\right]$	$0.5\Psi_R(\omega) + 0.25\Psi_R\left(\omega - \dfrac{2\pi}{2M+1}\right) + 0.25\Psi_R\left(\omega + \dfrac{2\pi}{2M+1}\right)$
Hamming	$0.54 + 0.46\cos\left[\dfrac{2\pi n}{2M+1}\right]$	$0.54\Psi_R(\omega) + 0.23\Psi_R\left(\omega - \dfrac{2\pi}{2M+1}\right) + 0.23\Psi_R\left(\omega + \dfrac{2\pi}{2M+1}\right)$
Blackman	$0.42 + 0.5\cos\left[\dfrac{2\pi n}{2M+1}\right] + 0.08\cos\left[\dfrac{4\pi n}{2M+1}\right]$	$0.42\Psi_R(\omega) + 0.25\Psi_R\left(\omega - \dfrac{2\pi}{2M+1}\right) + 0.25\Psi_R\left(\omega + \dfrac{2\pi}{2M+1}\right) + 0.04\Psi_R\left(\omega - \dfrac{4\pi}{2M+1}\right) + 0.04\Psi_R\left(\omega + \dfrac{4\pi}{2M+1}\right)$

where $\omega_p(\omega_s)$ is defined to be the highest frequency where $H(\omega) \ge 1 - \delta$ (the lowest frequency where $H(\omega) \le \delta$). The width of the transition band, $\Delta\omega = \omega_s - \omega_p$, is thus less than the mainlobe width Δ_M. This means that for a good window function, the mainlobe width has to be as narrow as possible. On the other hand, for a small ripple value δ, it is required that the area under the sidelobes of $\Psi(\omega)$ be as small as possible. These two requirements contradict each other.

For the fixed window functions given in Table 4-3, $H(\omega)$ is characterized by the facts that δ is approximately a constant, regardless of the values of M and ω_c, and the transition bandwidth is inversely proportional to the filter order $N = 2M$; that is,

$$\Delta\omega = \omega_s - \omega_p \approx \gamma/(2M), \tag{4.45}$$

where γ is also approximately a constant. Some properties of these window functions are summarized in Table 4-4. It gives, for each window, the mainlobe width

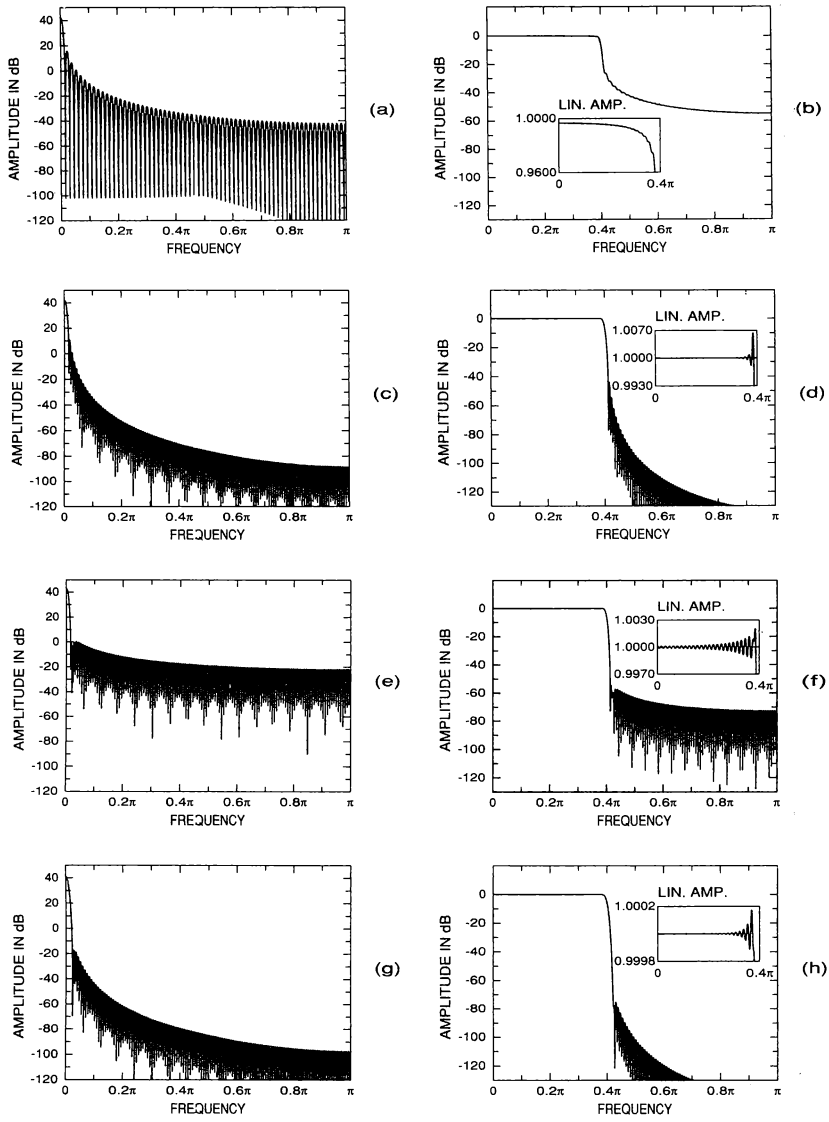


FIGURE 4-15 Frequency responses of the window functions and the resulting filters for the last four windows in Table 4-3 for $M = 128$ and $\omega_c = 0.4\pi$. (a,b) Bartlett window. (c,d) Hann window. (e,f) Hamming window. (g,h) Blackman window.

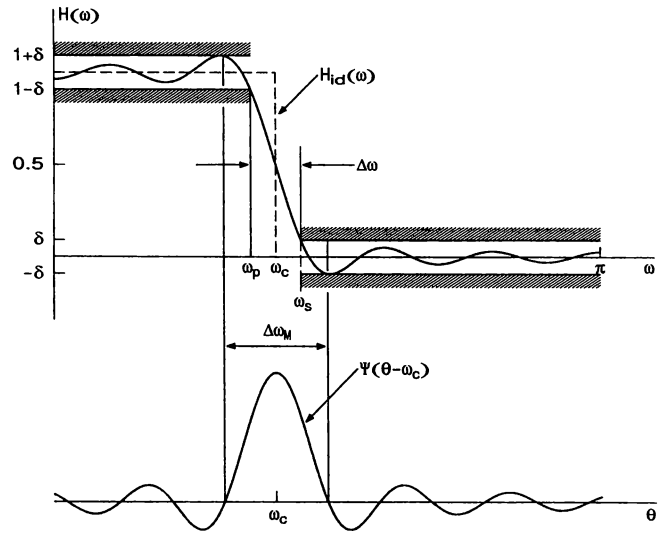


FIGURE 4-16 Typical relations between the frequency response of the window function and the resulting filter response in the lowpass case with cutoff edge ω_c.

Δ_M and the maximum sidelobe ripple in decibels in the case where $\Psi(\omega)$ is scaled to be unity at $\omega = 0$. Furthermore, the minimum stopband attenuation

$$A_s = -20 \log_{10} \delta \tag{4.46}$$

and the dependence of the transition bandwidth on $2M$ are given.[6] The only exception is the Bartlett window for which the filter response has no zeros on the

[6]These values have been determined for the case $\omega_c = 0.4\pi$ and $M = 128$.

TABLE 4-4 Properties of Some Commonly Used Fixed Windows

Window Type	Mainlobe Width Δ_M	Sidelobe Ripple	A_s	$\Delta\omega = \omega_s - \omega_p$
Rectangular	$\frac{4\pi}{2M+1}$	−13.3 dB	20.9 dB	$1.84\pi/(2M)$
Bartlett	$\frac{4\pi}{M+1}$	−26.5 dB	See text	See text
Hann	$\frac{8\pi}{2M+1}$	−31.5 dB	43.9 dB	$6.22\pi/(2M)$
Hamming	$\frac{8\pi}{2M+1}$	−42.7 dB	54.5 dB	$6.64\pi/(2M)$
Blackman	$\frac{12\pi}{2M+1}$	−58.1 dB	75.3 dB	$11.13\pi/(2M)$

unit circle (see Figure 4-15) and it is difficult to locate the stopband edge. The above window functions suffer from the drawback that A_s cannot be varied. Only ω_p and ω_s can be adjusted by properly selecting ω_c and M. Since ω_c is in the center of the transition band, it is selected to be $\omega_c = (\omega_p + \omega_s)/2$. Then M is determined from Eq. (4.45), where the specific value of γ can be found in Table 4-4 for each window type.

4-4-4 Adjustable Window Functions

The above problem can be overcome by using window functions having an additional parameter with which A_s can be varied. There exist three approaches to obtaining good adjustable windows. The first alternative is to minimize the energy in the sidelobes of the frequency response of the window function $w[n]$, whereas the second one is to minimize the peak sidelobe ripple. Both the Kaiser window [KA66, KA74] and the Saramäki window [SA89a, SA91a] provide an approximately optimum solution to the first problem, whereas the Dolph–Chebyshev window [HE68] is the solution to the second problem. The third alternative is to properly combine the first two approaches [SA91a].

The Kaiser window [KA66, KA74] is given by

$$w[n] = \begin{cases} I_0\left[\alpha\sqrt{1-\left(\dfrac{n}{M}\right)^2}\right] \Big/ I_0(\alpha), & -M \le n \le M \\ 0, \quad \text{otherwise}, \end{cases} \tag{4.47}$$

where α is the adjustable parameter and $I_0(x)$ is the modified zeroth-order Bessel function of the first kind, which has the simple power series expansion

$$I_0(x) = 1 + \sum_{r=1}^{\infty}\left[\frac{(x/2)^r}{r!}\right]^2. \tag{4.48}$$

For most practical applications, about 20 terms in the above summation are sufficient to arrive at reasonably accurate values of $w[n]$.

For the Saramäki window [SA89a], the frequency response of the unscaled window function ($\overline{w}[0]$ is not equal to unity) can be expressed in the forms

$$\begin{aligned} \overline{\Psi}(\omega) &= \sum_{n=-M}^{M} \overline{w}[n]e^{-jn\omega} = 1 + \sum_{k=1}^{M} 2T_k[\gamma\cos\omega + (\gamma - 1)] \\ &= \frac{\sin\left[\dfrac{2M+1}{2}\cos^{-1}\{\gamma\cos\omega + (\gamma-1)\}\right]}{\sin\left[\frac{1}{2}\cos^{-1}\{\gamma\cos\omega + (\gamma-1)\}\right]}, \end{aligned} \tag{4.49a}$$

where $T_k[x]$ is the kth degree Chebyshev polynomial and

$$\gamma = \left(1 + \cos\frac{2\pi}{2M+1}\right) \Big/ \left(1 + \cos\frac{2\beta\pi}{2M+1}\right). \tag{4.49b}$$

Here, β is the adjustable parameter, which has been selected such that the mainlobe width is $4\beta\pi/(2M+1)$. This is β times that of the rectangular window. In the special case $\beta = 1$, $\overline{\Psi}(\omega)$ is the frequency response of the rectangular window.

The desired normalized window function (w[0] = 1) is

$$\mathrm{w}[n] = \begin{cases} \overline{\mathrm{w}}[n]/\overline{\mathrm{w}}[0], & -M \le n \le M \\ 0, & \text{otherwise} \end{cases} \tag{4.50}$$

and the corresponding frequency response is $\Psi(\omega) = \overline{\Psi}(\omega)/\overline{\mathrm{w}}[0]$.

The unscaled coefficients $\overline{\mathrm{w}}[n]$ can be expressed as

$$\overline{\mathrm{w}}[n] = v_0[n] + 2\sum_{k=1}^{M} v_k[n], \tag{4.51}$$

where the $v_k[n]$'s can be calculated using the following recursion formulas:

$$v_0[n] = \begin{cases} 1, & n = 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.52a}$$

$$v_1[n] = \begin{cases} \gamma - 1, & n = 0 \\ \gamma/2, & |n| = 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.52b}$$

$$v_k[n] = \begin{cases} 2(\gamma - 1)v_{k-1}[n] - v_{k-2}[n] \\ \quad +\gamma(v_{k-1}[n-1] + v_{k-1}[n+1]), & -k \le n \le k \\ 0, & \text{otherwise.} \end{cases} \tag{4.52c}$$

For the Dolph–Chebyshev window [HE68], the frequency response of the unscaled window function can be expressed as

$$\overline{\Psi}(\omega) = T_M[\gamma \cos\omega + (\gamma - 1)], \tag{4.53a}$$

where

$$\gamma = \left(1 + \cos\frac{\pi}{2M}\right) \Big/ \left(1 + \cos\frac{2\beta\pi}{2M+1}\right), \tag{4.53b}$$

and the unscaled coefficients are

$$\overline{w}[n] = v_M[n], \tag{4.54}$$

where $v_M[n]$ can be determined using the recursion relations of Eq. (4.52). Here, β has been selected as for the Saramäki window to make the mainlobe width β times that of the rectangular window.

The transitional windows introduced by Saramäki [SA91a] combine the properties of the Dolph–Chebyshev and Saramäki windows. For the mainlobe width being β times that of the rectangular window, the unscaled frequency response of this window is given by

$$\overline{\Psi}(\omega) = \sum_{n=-M}^{M} \overline{w}(n) e^{-jn\omega} = \prod_{k=1}^{M} (\cos \omega - \cos \omega_k), \tag{4.55a}$$

where

$$\omega_k = \rho \omega_k^{(1)} + (1 - \rho)\omega_k^{(2)} \tag{4.55b}$$

with

$$\omega_k^{(1)} = 2 \cos^{-1}\left[\frac{\cos [\beta\pi/(2M + 1)]}{\cos [\pi/(2M + 1)]} \cos\left(\frac{k\pi}{2M + 1}\right)\right] \tag{4.55c}$$

and

$$\omega_k^{(2)} = 2 \cos^{-1}\left[\frac{\cos [\beta\pi/(2M + 1)]}{\cos [\pi/(4M)]} \cos\left(\frac{(2k - 1)\pi}{4M}\right)\right]. \tag{4.55d}$$

Here, $\omega_k^{(1)}$ and $\omega_k^{(2)}$ for $k = 1, 2, \cdots, M$ are the zero locations of the Saramäki and the Dolph–Chebyshev windows, respectively. For $\rho = 1$ and $\rho = 0$, $\overline{\Psi}(\omega)$ is the unscaled response for the Saramäki and the Dolph–Chebyshev window, respectively. For this transitional window, $0 < \rho < 1$ is an adjustable parameter in addition to β. Accurate values for the unscaled window coefficients $\overline{w}[n]$ are obtained from [PA87]

$$\overline{w}[n] = \frac{1}{2M + 1}\left[\overline{\Psi}(0) + 2 \sum_{k=1}^{M} \overline{\Psi}\left(\frac{2\pi k}{2M + 1}\right) \cos\left(\frac{2\pi nk}{2M + 1}\right)\right]. \tag{4.56}$$

Alternatively, the coefficients can be determined by evaluating $\overline{\Psi}(\omega)$ at 2^I ($>2M + 1$) equally spaced frequencies and using the inverse fast Fourier transform (see Chapter 8). With a slightly increased amount of calculation, this window gives a higher attenuation for the resulting filter than the other adjustable windows considered above.

The advantages of the above adjustable windows compared to the fixed windows are their near optimality and flexibility. Given ω_p, ω_s, and the minimum stopband attenuation A_s of the filter, the adjustable parameter (α for the Kaiser window and β for the Saramäki, Dolph–Chebyshev, and transitional windows) can

be determined to give the desired value for A_s, whereas M can be determined to give the desired value for the transition bandwidth $\Delta\omega = \omega_s - \omega_p$ of the filter. Experimentally obtained estimation formulas for the adjustable parameter and M are given in Table 4-5 for each window type [KA66, KA74; SA89a, SA91a]. For the transitional window, these equations are for

$$\rho = \begin{cases} 0.4, & A_s \leq 50 \\ 0.5, & 50 < A_s \leq 75 \\ 0.6, & 75 < A_s, \end{cases} \tag{4.57}$$

which has turned out to be a good selection in most cases [SA91a]. Like for fixed window functions, the cutoff edge of the ideal filter is selected to be $\omega_c = (\omega_p + \omega_s)/2$ to center the transition band of the resulting filter at this point.

Because of the characteristics of the Dolph–Chebyshev and transitional windows, the estimation formulas developed for these windows are not as accurate as those for the Kaiser and Saramäki windows.

An informative way to compare the performances of adjustable windows is to design several classes of filters with various values of the adjustable parameter for fixed values of M and ω_c. Based on the resulting filter frequency responses, a plot of the stopband attenuation as a function of the parameter $D = 2M(\omega_s - \omega_p)$ can be generated (D, instead of $\omega_s - \omega_p$, is used to make the plot almost independent of M). Figure 4-17 gives such plots for the above-mentioned adjustable windows for $\omega_c = 0.4\pi$ and $M = 128$. For the Kaiser window and the Saramäki window, the difference in the plots is very small. For comparison purposes, a corresponding plot is also included for filters for which the passband and stopband ripples $\delta_p = \delta_s$ have been minimized in the minimax sense for the given value of D. This plot thus gives an upper limit for the stopband attenuation attainable using window functions. For the Kaiser window and the Saramäki window, the resulting attenuation is 5–7 dB less than this upper limit. The stopband attenuation obtained by the Dolph–Chebyshev window is 1–4 dB worse than that of the Kaiser or Saramäki window. For the transitional window, the improvement is typically 2–4 dB over the Kaiser and Saramäki windows and the resulting attenuation approaches the upper limit.

Example 4.1. It is desired to design with each adjustable window considered above a filter with an 80-dB stopband attenuation for $M = 128$ and $\omega_c = 0.4\pi$. Using the formulas given in Table 4-5, the values for the adjustable parameters for the Kaiser, the Saramäki, the Dolph–Chebyshev, and the transitional windows become $\alpha = 7.857$, $\beta = 2.702$, $\beta = 2.770$, and $\beta = 2.587$, respectively. The resulting attenuations are 79.68, 80.17, 79.29, and 80.75 dB. Figure 4-18 gives the frequency responses of both the window functions and the resulting filters in the case of an exactly 80-dB attenuation.[7] The transition bandwidths for these filters are

[7]The value of the adjustable parameter giving exactly the desired attenuation A_s can be obtained in two steps. First, the actual filter attenuation, denoted by A_r, is determined for the estimated value of the adjustable parameter. Then this parameter is reestimated by using, instead of A_s, $A_s - (A_r - A_s)$ in the estimation formula.

TABLE 4-5 Estimation Formulas for the Adjustable Parameter and M for Adjustable Windows to Give the Desired Attenuation A_s and Transition Bandwidth $\omega_s - \omega_p$

Window Type	Adjustable Parameter	M
Kaiser	$\alpha = \begin{cases} 0.1102(A_s - 8.7), & A_s > 50 \\ 0.5842(A_s - 21)^{0.4} & \\ \quad +0.07886(A_s - 21), & 21 < A_s < 50 \\ 0, & A_s < 21 \end{cases}$	$M = \dfrac{A_s - 7.95}{14.36(\omega_s - \omega_p)/\pi}$
Saramäki	$\beta = \begin{cases} 0.000121(A_s - 21)^2 & \\ \quad +0.0224(A_s - 21) + 1, & A_s \le 65 \\ 0.033A_s + 0.062, & 65 < A_s \le 110 \\ 0.0345A_s - 0.097, & A_s > 110 \end{cases}$	$M = \dfrac{A_s - 8.15}{14.36(\omega_s - \omega_p)/\pi}$
Dolph–Chebyshev	$\beta = \begin{cases} 0.0000769(A_s)^2 & \\ \quad +0.0248A_s + 0.330, & A_s \le 60 \\ 0.0000104(A_s)^2 & \\ \quad +0.0328A_s + 0.079, & A_s > 60 \end{cases}$	$M = \dfrac{1.028A_s - 8.4}{14.36(\omega_s - \omega_p)/\pi}$
Transitional	$\beta = \begin{cases} 0.000154(A_s)^2 & \\ \quad +0.0153A_s + 0.465, & A_s \le 60 \\ 0.0000204(A_s)^2 & \\ \quad +0.0303A_s + 0.032, & A_s > 60 \end{cases}$	$M = \dfrac{0.00036(A_s)^2 + 0.951A_s - 9.4}{14.36(\omega_s - \omega_p)/\pi}$

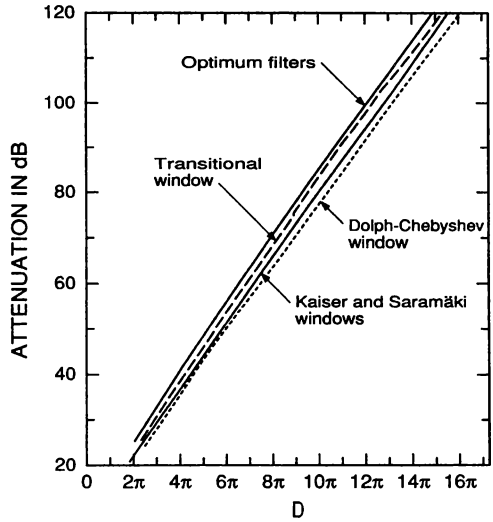


FIGURE 4-17 Plots of the minimum stopband attenuation of the resulting filter versus $D = 2M(\omega_s - \omega_p)$ for adjustable windows. $M = 128$ and $\omega_c = 0.4\pi$. For comparison purposes, a corresponding plot for filters for which $\delta_p = \delta_s$ has been minimized in the minimax sense is included.

0.0393π, 0.0390π, 0.0406π, and 0.0373π. From Figure 4-18, it is seen that the responses for the Kaiser window and the Saramäki window as well as the responses of the resulting filters are practically the same. It is interesting to observe from this figure that the filter response for the Dolph–Chebyshev window is flatter than the corresponding responses for the Kaiser and Saramäki windows, whereas the filter response for the transitional window is between those of the Saramäki and Dolph–Chebyshev windows. Also note that the ripples of the sidelobes are of the same height for the Dolph–Chebyshev window. In Section 4-7, this property is utilized in designing filters having an equiripple behavior in the stopband.

4-5 DESIGN OF FIR FILTERS IN THE LEAST-MEAN-SQUARE SENSE

The second straightforward approach for designing FIR filters is based on the use of the least-squared approximation [FA74; KA63, KA66; KE72; LI83c; PA87; TU70; VA87]. In this case, the problem is to find the filter coefficients to minimize

$$E_2 = \int_X [W(\omega)[H(\omega) - D(\omega)]]^2 \, d\omega, \tag{4.58}$$

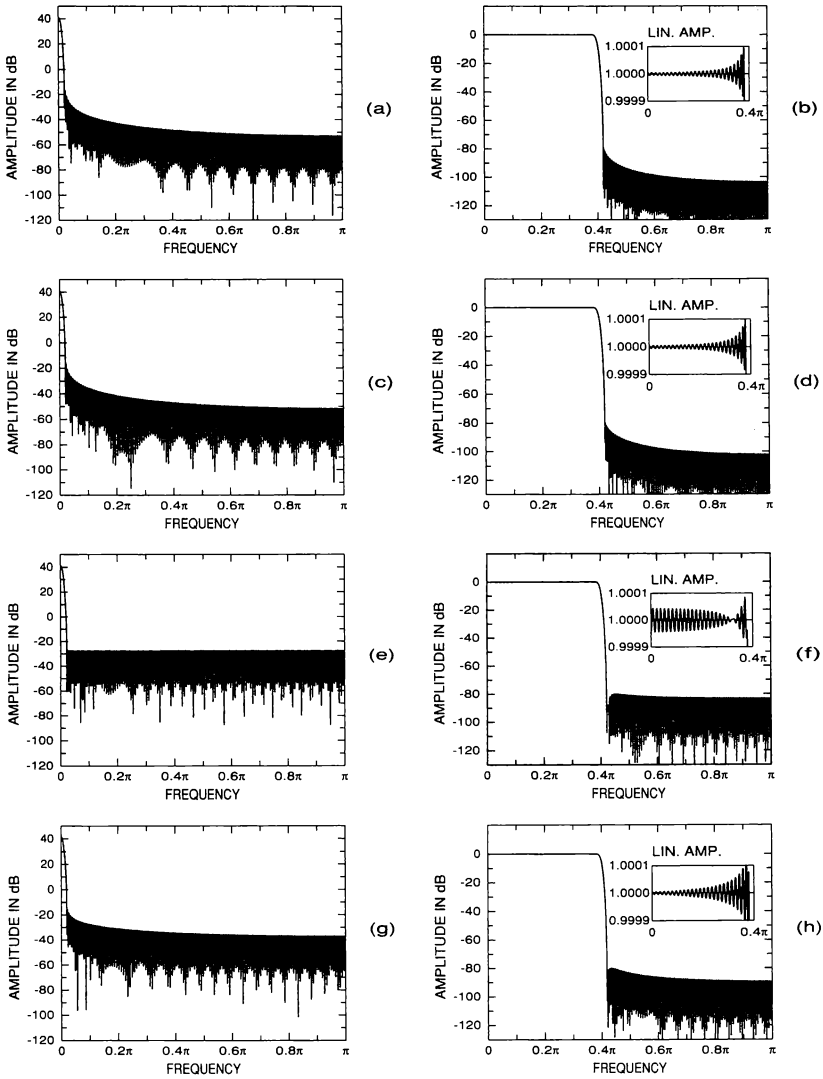


FIGURE 4-18 Frequency responses of the window functions and the resulting filters for adjustable windows giving an 80-dB attenuation for the filter when $M = 128$ and $\omega_c = 0.4\pi$. (a,b) Kaiser window. (c,d) Saramäki window. (e,f) Dolph–Chebyshev window. (g,h) Transitional window.

where X contains the passband and stopband regions, $D(\omega)$ is a desired response, and $W(\omega)$ is a positive weighting function. If $D(\omega)$ and $W(\omega)$ are sampled at a very dense grid of frequencies $\omega_1, \omega_2, \cdots, \omega_K$ on X, then minimization of the above equation may be achieved by minimizing

$$E_2 = \sum_{k=1}^{K} [W(\omega_k)[H(\omega_k) - D(\omega_k)]]^2. \tag{4.59}$$

As shown in Section 4-3-3, $H(\omega)$ can be expressed in the four different linear-phase cases in the form (see Eq. (4.32))

$$H(\omega) = \sum_{n=0}^{M} b[n] \text{ trig } (\omega, n). \tag{4.60}$$

By substituting this for $H(\omega_k)$ in Eq. (4.59) and transferring $W(\omega_k)$ inside the parentheses we obtain

$$E_2 = \sum_{k=1}^{K} \left[W(\omega_k) \sum_{n=0}^{M} b[n] \text{ trig } (\omega_k, n) - W(\omega_k)D(\omega_k) \right]^2. \tag{4.61}$$

This equation can be written in the following quadratic form

$$E_2 = \mathbf{e}^T\mathbf{e}, \tag{4.62a}$$

where

$$\mathbf{e} = \mathbf{Xb} - \mathbf{d} \tag{4.62b}$$

with

$$\mathbf{X} = \begin{bmatrix} W(\omega_1) \text{ trig } (\omega_1, 0) & W(\omega_1) \text{ trig } (\omega_1, 1) & \cdots & W(\omega_1) \text{ trig } (\omega_1, M) \\ W(\omega_2) \text{ trig } (\omega_2, 0) & W(\omega_2) \text{ trig } (\omega_2, 1) & \cdots & W(\omega_2) \text{ trig } (\omega_2, M) \\ \vdots & \vdots & & \vdots \\ W(\omega_K) \text{ trig } (\omega_K, 0) & W(\omega_K) \text{ trig } (\omega_K, 1) & \cdots & W(\omega_K) \text{ trig } (\omega_K, M) \end{bmatrix}, \tag{4.62c}$$

$$\mathbf{b} = [b[0], b[1], \ldots, b[M]]^T, \tag{4.62d}$$

and

$$\mathbf{d} = [W(\omega_1)D(\omega_1), W(\omega_2)D(\omega_2), \ldots, W(\omega_K)D(\omega_K)]^T. \tag{4.62e}$$

Here, $\mathbf{e}$ is a K length vector with the kth element being $W(\omega_k)[H(\omega_k) - D(\omega_k)]$. The optimum solution of minimizing E_2 is given by [LI83c; PA87]

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{d} \tag{4.63}$$

and it satisfies the "normal equations" [PA87]

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{d}. \tag{4.64}$$

It should be noted that if K is much larger than M, then Eq. (4.63) should not be solved directly because it becomes ill conditioned. In this case, the direct solution will probably have large errors. Parks and Burrus [PA87] recommend the use of the software package LINPACK [DO79], which has a special program for solving the above problem.

In the case where both $W(\omega)$ and $D(\omega)$ are piecewise-constant functions, a significantly simpler procedure for finding the optimum solution can be generated [LI83c].

Example 4.2. Consider the design of a Type I lowpass filter of order $N = 46$ ($M = 23$) for $\omega_p = 0.5\pi$ and $\omega_s = 0.6\pi$. $D(\omega) = 1$ on $[0, \omega_p]$ and $D(\omega) = 0$ on $[\omega_s, \pi]$. Figure 4-19 shows the resulting responses in two cases. In both cases, $W(\omega) = 1$ on $[0, \omega_p]$, whereas $W(\omega)$ is unity on $[\omega_s, \pi]$ in the first case and 10 in the second case. The effect of the stopband weighting is clearly seen from the figure. It is also seen that the maximum deviations between the actual and the desired responses are much larger near the passband and stopband edges. This is characteristic of the least-squared-error designs. If the maximum deviations are desired to be minimized, then it is preferred to design the filter in the minimax sense (see Section 4-8). Compare Figure 4-19 to Figure 4-8, which gives a re-

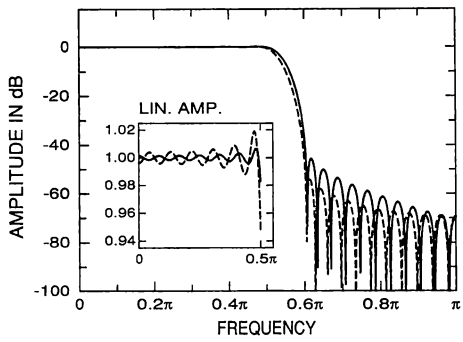


FIGURE 4-19 Amplitude responses for least-squared-error FIR filters of order 46. The solid and dashed lines give the responses for the filters with stopband weighting of 1 and 10, respectively.

sponse for an FIR filter optimized in the minimax sense. The filter orders in these two figures are the same.

4-6 MAXIMALLY FLAT FIR FILTERS

The third straightforward approach for designing FIR filters is to use filters with maximally flat response around $\omega = 0$ and $\omega = \pi$ [HA77; HE71b; KA77b, KA79; VA84]. The advantages of these filters are that the design is extremely simple and they are useful in applications where the signal is desired to be preserved with very small error near the zero frequency. If the maximum deviation from the desired response is required to be minimized, then it is preferred to use filters designed in the minimax sense. These minimax designs meet the given criteria with a significantly reduced filter order.

Consider a Type I filter with transfer function

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n}, \qquad h[2M - n] = h[n]. \tag{4.65}$$

For maximally flat designs, it is advantageous to express $H(\omega)$ as an Mth degree polynomial in cos ω as follows (cf. Section 4-3-3):

$$H(\omega) = \sum_{n=0}^{M} \alpha[n] \cos^n \omega. \tag{4.66}$$

This $H(\omega)$ is determined in such a way that it has $2K$ zeros at $\omega = \pi$ and $H(\omega) - 1$ has $2L = 2(M - K + 1)$ zeros at $\omega = 0$. M is thus related to L and K through $M = K + L - 1$. The above conditions are satisfied if $H(\omega)$ can be written simultaneously in the forms

$$H(\omega) = \left[\frac{1 + \cos \omega}{2}\right]^K \sum_{n=0}^{L-1} d[n] \left[\frac{1 - \cos \omega}{2}\right]^n$$

$$= \cos^{2K}(\omega/2) \sum_{n=0}^{L-1} d[n] \sin^{2n}(\omega/2) \tag{4.67a}$$

and

$$H(\omega) = 1 - \left[\frac{1 - \cos \omega}{2}\right]^L \sum_{n=0}^{K-1} \bar{d}[n] \left[\frac{1 + \cos \omega}{2}\right]^n$$

$$= 1 - \sin^{2L}(\omega/2) \sum_{n=0}^{K-1} \bar{d}[n] \cos^{2n}(\omega/2). \tag{4.67b}$$

The coefficients $d[n]$ and $\bar{d}[n]$ giving the desired solution are

$$d[n] = \frac{(K-1+n)!}{(K-1)!n!}, \qquad \bar{d}[n] = \frac{(L-1+n)!}{(L-1)!n!}. \tag{4.68}$$

The resulting $H(\omega)$ is characterized by the facts that it achieves the value one at $\omega = 0$ and its first $2L-1$ derivatives are zero at this point, whereas it achieves the value zero at $\omega = \pi$ with its first $2K-1$ derivatives being zero at this point. The primary unknowns of the above filters are K and L. Given the filter specifications, the problem is thus to determine these integers such that the criteria are satisfied.

Kaiser [KA79] has stated the specifications for maximally flat filters as shown in Figure 4-20. Here, β is the center of the transition band and γ is the width of the transition band, which is defined as the region where the response varies from 0.95 (passband edge angle) to 0.05 (stopband edge angle). For meaningful specifications, γ has to satisfy $0 \leq \gamma \leq \min(2\beta, 2\pi - 2\beta)$. In the design procedure proposed by Kaiser, the lower estimate for $M = K + L - 1$ (half the filter order) is given by

$$M_{\text{lower}} = (\pi/\gamma)^2. \tag{4.69}$$

Then ρ is determined by

$$\rho = (1 + \cos\beta)/2. \tag{4.70}$$

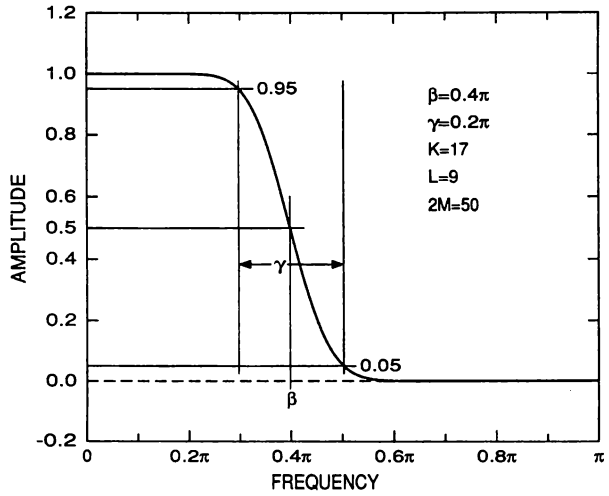


FIGURE 4-20 Specifications for a maximally flat lowpass filter and response for a filter of order 50 meeting the criteria $\beta = 0.4\pi$ and $\gamma = 0.2\pi$.

The next step is to determine

$$K_p = \langle \rho M_p \rangle, \tag{4.71}$$

where $\langle x \rangle$ stands for the nearest integer of x, for the values of M_p in the range $M_{\text{lower}} \leq M_p \leq 2M_{\text{lower}}$. Finally, the values of the integers K_p and M_p for which the ratio K_p/M_p is closest to ρ are selected. The corresponding values of K, L, and M are then $K = K_p$, $L = M_p - K_p$, and $M = M_p - 1$, respectively. With these selections of K and L, the given value of β can be achieved accurately.

Example 4.3. Consider the specifications $\beta = 0.4\pi$ and $\gamma = 0.2\pi$. The above procedure results in $K = 17$ and $L = 9$. The order of the filter is thus $2(K + L - 1) = 50$. The amplitude response of this filter is depicted in Figure 4-20.

The transfer function having the zero-phase frequency response as given by Eq. (4.67a) or (4.67b) can be implemented using the conventional direct-form structure shown in Figure 4-7.[8] Alternatively, the transfer function can be written in the forms

$$H(z) = \left(\frac{1 + z^{-1}}{2}\right)^{2K} \sum_{n=0}^{L-1} (-1)^n d[n] z^{-(L-1-n)} \left(\frac{1 - z^{-1}}{2}\right)^{2n} \tag{4.72a}$$

and

$$H(z) = z^{-M} - (-1)^L \left(\frac{1 - z^{-1}}{2}\right)^{2L} \sum_{n=0}^{K-1} \bar{d}[n] z^{-(K-1-n)} \left(\frac{1 + z^{-1}}{2}\right)^{2n}. \tag{4.72b}$$

The advantage of realizing the transfer function in the above forms lies in the fact that these implementation forms have significantly fewer multipliers than the direct-form structure [VA84].

4-7 SOME SIMPLE FIR FILTER DESIGNS

There are two special cases where the optimum solution in the minimax sense can be obtained analytically [HE73]. The first analytically solvable case is the one where the zero-phase frequency response is monotonically decaying in the passband region and exhibits an equiripple behavior in the stopband region $[\omega_s, \pi]$. An

[8] In this case, it is preferred to calculate the filter coefficients by evaluating $H(\omega)$ at 2^l ($>2M + 1$) equally spaced frequencies and using the inverse discrete Fourier transform (see Chapter 8). This guarantees that the resulting coefficient values are accurate enough. This is the procedure used by Kaiser [KA79].

equiripple behavior on $[\omega_s, \pi]$ can be achieved by mapping the Mth degree Chebyshev polynomial $T_M(x)$ to the ω-plane such that the region $[-1, 1]$, where $T_M(x)$ oscillates within the limits ± 1, is mapped to the region $[\omega_s, \pi]$ (see Figure 4-21). The desired transformation mapping $x = 1$ to $\omega = \omega_s$ and $x = -1$ to $\omega = \pi$ is

$$x = \gamma \cos \omega + (\gamma - 1), \qquad \gamma = \frac{2}{1 + \cos \omega_s}, \tag{4.73}$$

resulting in the following zero-phase frequency response of a Type I filter of order $2M$:

$$\overline{H}(\omega) = T_M[(2 \cos \omega + 1 - \cos \omega_s)/(1 + \cos \omega_s)]. \tag{4.74}$$

The response taking the value $1 + \delta_p$ at $\omega = 0$ is then

$$H(\omega) = (1 + \delta_p)\overline{H}(\omega)/\overline{H}(0). \tag{4.75}$$

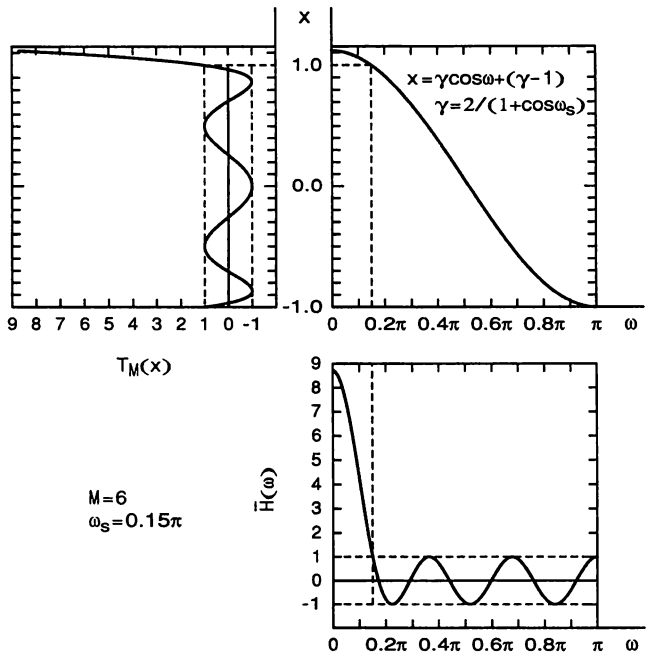


FIGURE 4-21 Generation of a zero-phase frequency response oscillating within the limits ± 1 in the stopband $[\omega_s, \pi]$ based on mapping the Mth degree Chebyshev polynomial $T_M(x)$ to the ω-plane.

This function oscillates on $[\omega_s, \pi]$ within the limits $\pm\delta_s$, where

$$\delta_s = (1 + \delta_p)/\overline{H}(0). \tag{4.76}$$

Based on the properties of Chebyshev polynomials, it can be shown that the value of M (half the filter order) giving the specified stopband ripple δ_s is [HE73]

$$M = \frac{\cosh^{-1}[(1 + \delta_p)/\delta_s]}{\cosh^{-1}[(3 - \cos\omega_s)/(1 + \cos\omega_s)]}. \tag{4.77}$$

Example 4.4. Figure 4-22 gives responses with $\omega_s = 0.1\pi$ and $\delta_p = 0.1$ for $M = 15$ and $M = 30$. The corresponding filter orders are 30 and 60, respectively. The disadvantage of these designs is that all their zeros lie on the unit circle and the passband region, where the response decays from $1 + \delta_p$ to $1 - \delta_p$, is narrow and cannot be controlled.

The response that is equiripple in the passband $[0, \omega_p]$ oscillating within the limits $1 \pm \delta_p$ and monotonically decaying in the region $[\omega_p, \pi]$ can be derived in the same manner. This solution is given by [HE73]

$$H(\omega) = 1 - \delta_p T_M[(-2\cos\omega + 1 + \cos\omega_p)/(1 - \cos\omega_p)]. \tag{4.78}$$

If it is desired that $H(\pi) = -\delta_s$, then δ_p can be determined from

$$\delta_p = (1 + \delta_s)/T_M[(3 + \cos\omega_p)/(1 - \cos\omega_p)]. \tag{4.79}$$

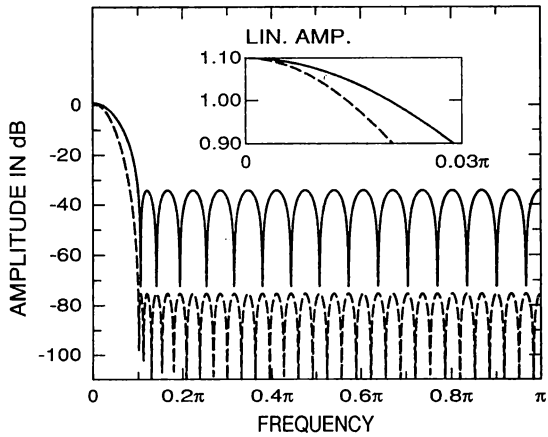


FIGURE 4-22 Responses for filters having an equiripple stopband behavior and a monotonically decaying passband response for $M = 15$ (solid line) and $M = 30$ (dashed line). $\omega_s = 0.1\pi$ and $\delta_p = 0.1$.

The minimum value of M required to meet the given ripple requirements can be determined from Eq. (4.77) by interchanging δ_p and δ_s and by replacing ω_s by $\pi - \omega_p$.

4-8 DESIGN OF FIR FILTERS IN THE MINIMAX SENSE

One of the main advantages of FIR filters over their IIR counterparts is that there exists an efficient algorithm for optimizing in the minimax sense arbitrary-magnitude FIR filters. For IIR filters, the design of arbitrary-magnitude filters is usually time-consuming and the convergence to the best solution is not always guaranteed. This section introduces this algorithm for designing linear-phase FIR filters and shows its flexibility by means of several examples. The resulting filters are optimal in the sense that they meet the given arbitrary specifications with the minimum filter order. This section considers also some properties of these optimum linear-phase FIR filters.

4-8-1 Remez Multiple Exchange Algorithm

The most efficient algorithm for designing optimum magnitude FIR filters with arbitrary specifications is the Remez multiple exchange algorithm [CH66; RI64]. The most frequently used method for implementing this algorithm is the one originally advanced by Parks and McClellan [PA72a]. Further improvements to the implementation of the Remez algorithm have been proposed by McClellan, Parks, and Rabiner [MC73a, MC73b, MC79; PA72b; RA75b]. As a result of this work, a program for designing arbitrary-magnitude FIR filters has been reported in McClellan et al. [MC73b, MC79]. This program is directly applicable to obtaining optimal designs for most types of FIR filters like lowpass, highpass, bandpass, and bandstop filters, Hilbert transformers, and digital differentiators. Also, filters having several passbands and stopbands can be designed directly. The amount of computation required for designing optimum filters can be significantly reduced by using techniques proposed by Antoniou [AN82, AN83] and Bonzanigo [BO82]. This section concentrates on the original FIR filter design program of McClellan, Parks, and Rabiner [MC73b, MC79]. This method is referred to later as the MPR algorithm.

4-8-1-1 Characterization of the Optimum Solution. The Remez multiple exchange algorithm is the most powerful algorithm for finding the coefficients $a[n]$ of the function

$$G(\omega) = \sum_{n=0}^{M} a[n] \cos n\omega \tag{4.80}$$

to minimize on a closed subset X of $[0, \pi]$ the peak absolute value of the following weighted error function:

$$E(\omega) = \overline{W}(\omega)[G(\omega) - \overline{D}(\omega)], \tag{4.81}$$

that is, the quantity

$$\epsilon = \max_{\omega \in X} |E(\omega)|. \tag{4.82}$$

It is required that $\overline{D}(\omega)$ be continuous on X and $\overline{W}(\omega) > 0$. This algorithm can be used in all four linear-phase cases based on the fact that the zero-phase frequency response $H(\omega)$ of a filter of order N can be expressed, according to the discussion of Section 4-3-1, in the form

$$H(\omega) = F(\omega)G(\omega), \tag{4.83}$$

where $G(\omega)$ is as given by Eq. (4.80) and

$$F(\omega) = \begin{cases} 1 & \text{for Type I} \\ \cos(\omega/2) & \text{for Type II} \\ \sin\omega & \text{for Type III} \\ \sin(\omega/2) & \text{for Type IV,} \end{cases} \qquad M = \begin{cases} N/2 & \text{for Type I} \\ (N-1)/2 & \text{for Type II} \\ (N-2)/2 & \text{for Type III} \\ (N-1)/2 & \text{for Type IV.} \end{cases} \tag{4.84}$$

If the desired function for $H(\omega)$ on X is $D(\omega)$ and the weighting function is $W(\omega)$, then the error function can be written in the form of Eq. (4.81) as follows:

$$\begin{aligned} E(\omega) &= W(\omega)[H(\omega) - D(\omega)] = W(\omega)[F(\omega)G(\omega) - D(\omega)] \\ &= W(\omega)F(\omega)[G(\omega) - D(\omega)/F(\omega)] = \overline{W}(\omega)[G(\omega) - \overline{D}(\omega)], \end{aligned} \tag{4.85a}$$

where

$$\overline{W}(\omega) = F(\omega)W(\omega), \qquad \overline{D}(\omega) = D(\omega)/F(\omega). \tag{4.85b}$$

The Remez multiple exchange algorithm can be constructed on the basis of the following characterization theorem [CH66; RI64].

Characterization Theorem. Let $G(\omega)$ be of the form of Eq. (4.80). Then $G(\omega)$ is the best unique solution minimizing ϵ as given by Eq. (4.82) if and only if there exist at least $M + 2$ points $\omega_1, \omega_2, \ldots, \omega_{M+2}$ in X such that

$$\omega_1 < \omega_2 < \cdots < \omega_{M+1} < \omega_{M+2}$$

$$E(\omega_{i+1}) = -E(\omega_i), \qquad i = 1, 2, \ldots, M+1$$

$$|E(\omega_i)| = \epsilon, \qquad i = 1, 2, \ldots, M+2.$$

In other words, the optimum solution is characterized by the fact that the weighted error function $E(\omega)$ alternatingly achieves the values $\pm\epsilon$, with ϵ being the peak

absolute value of the weighted error, at least at $M + 2$ consecutive points in X. Figure 4-23 gives the response of a typical optimum Type I lowpass filter of order $N = 12$ and the corresponding error function. In this case, $H(\omega) \equiv G(\omega)$, $X = [0, \omega_p] \cup [\omega_s, \pi]$, $\overline{D}(\omega) = D(\omega) = 1$ and $\overline{W}(\omega) = W(\omega) = 1$ for $\omega \in [0, \omega_p]$, whereas $D(\omega) = 0$ and $W(\omega) = 2$ for $\omega \in [\omega_s, \pi]$. Note that the above weighting function makes the stopband ripple of the optimum filter to be half of that of the passband ripple. In this case, $M = N/2 = 6$ so that $G(\omega)$ contains seven unknowns $a[0], a[1], \ldots, a[6]$. The number of extremal points is $M + 2 = 8$ so that there is one more extremal frequency than there are unknowns, as required by the characterization theorem. According to the theorem, it is thus easy to check whether a given solution is the optimum one. If the relative weighting between the stopband and passband errors is k and there exists a solution $H(\omega)$ that alternatingly goes through the values $1 \pm \epsilon$ in the passband and through the values $\pm\epsilon/k$ in the stopband, and the overall number of these extrema is at least $M + 2$, then this solution is, according to the characterization theorem, the best unique solution. In the lowpass case, both ω_p and ω_s are always extremal points, and $H(\omega_p) = 1 - \epsilon$ and $H(\omega_s) = \epsilon/k$ so that $E(\omega_p) = -\epsilon$ and $E(\omega_s) = \epsilon$.

4-8-1-2 The McClellan-Parks-Rabiner (MPR) Algorithm. Given a set of $M + 2$ points on X, denoted by $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M+2}\}$, the unknown coefficients $a[0], a[1], \ldots, a[M]$ and ϵ can be determined such that $E(\omega)$ satisfies

$$E(\omega_k) = \overline{W}(\omega_k)[G(\omega_k) - \overline{D}(\omega_k)] = (-1)^k\epsilon, \qquad k = 1, 2, \ldots, M + 2. \tag{4.86}$$

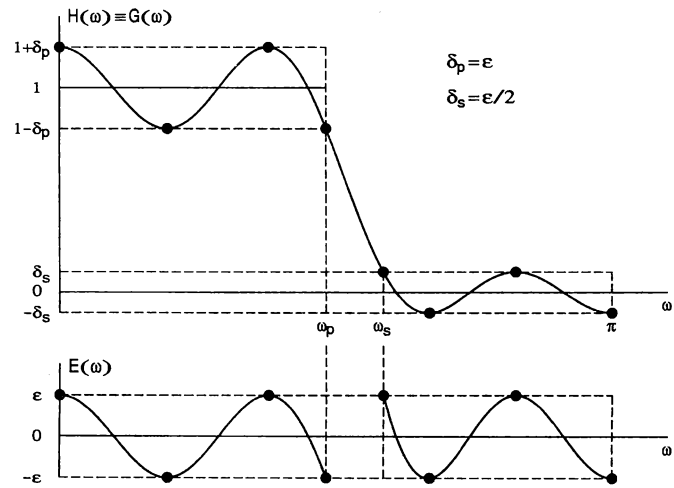


FIGURE 4-23 Zero-phase frequency response and error function for an optimum Type I linear-phase lowpass filter of order $2M = 12$. The stopband weighting is two times that of the passband.

This can be achieved by solving for the unknowns the following system of $M + 2$ linear equations:

$$\sum_{n=0}^{M} a[n] \cos n\omega_k - (-1)^k \epsilon / \overline{W}(\omega_k) = \overline{D}(\omega_k), \qquad k = 1, 2, \ldots, M + 2. \tag{4.87}$$

The resulting $E(\omega)$ goes alternatingly through the values $\pm \epsilon$ at the points ω_k. If X consists of the above set of $M + 2$ points, that is, $X = \Omega$, then $|\epsilon|$ is the peak absolute value of $E(\omega)$ on X and the conditions of the above characterization theorem are satisfied.[9] The Remez exchange algorithm makes use of this fact. The problem is simply to find a set Ω on X in such a way that the optimum solution on Ω is simultaneously the optimum solution on the overall set X. This is achieved if the value of $|\epsilon|$ is simultaneously the peak absolute value of $E(\omega)$ on the overall set X. The Remez algorithm iteratively finds the desired set of $M + 2$ extremal points using the following steps:

1. Select an initial set of $M + 2$ extremal points $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M+2}\}$ in X.
2. Solve the system of $M + 2$ linear equations given by Eq. (4.87) for the unknowns $a[0], a[1], \ldots, a[M]$ and ϵ.
3. Find on X, $M + 2$ extremal points of the resulting $E(\omega)$, where $|E(\omega)| \geq |\epsilon|$. If there are more than $M + 2$ extremal points, retain $M + 2$ extrema such that the largest absolute values are included with the condition that the sign of the error function $E(\omega)$ alternates at the selected points. Store the abscissas of the extrema into $\overline{\Omega} = \{\overline{\omega}_1, \overline{\omega}_2, \ldots, \overline{\omega}_{M+2}\}$.
4. If $|\omega_k - \overline{\omega}_k| \leq \alpha$ for $k = 1, 2, \ldots, M + 2$ (α is a small number), then go to the next step. Otherwise, set $\Omega = \overline{\Omega}$ and go to Step 2.
5. Calculate the filter coefficients and plot the frequency response.

The above algorithm starts by selecting $M + 2$ initial extremal points ω_k. These points can be selected, for example, to lie equidistantly on X. Then the coefficients of $G(\omega)$ and ϵ are solved for at Step 2 such that $E(\omega)$ satisfies

$$E(\omega_k) = (-1)^k \epsilon, \qquad k = 1, 2, \ldots, M + 2; \tag{4.88}$$

that is, it alternately achieves the values $\pm\epsilon$ at the points $\omega_1, \omega_2, \ldots, \omega_{M+2}$, as required by the characterization theorem. However, all the points ω_k are not the true external points of the resulting $E(\omega)$. This is illustrated in Figure 4-24, which gives $E(\omega)$ after performing Step 2 for the first time. As seen from this figure, most of the initial extremal points are not the true extremal points and the maximum absolute value of $E(\omega)$ is higher than $|\epsilon|$ around these points. Thus the $M + 2$ new extremal points $\overline{\omega}_k$ of $E(\omega)$ are located next and these extremal points replace the ω_k's. After that, the coefficients of $G(\omega)$ and ϵ are redetermined such that Eq.

[9]When solving Eq. (4.87) for the unknowns, the resulting ϵ is either positive or negative.

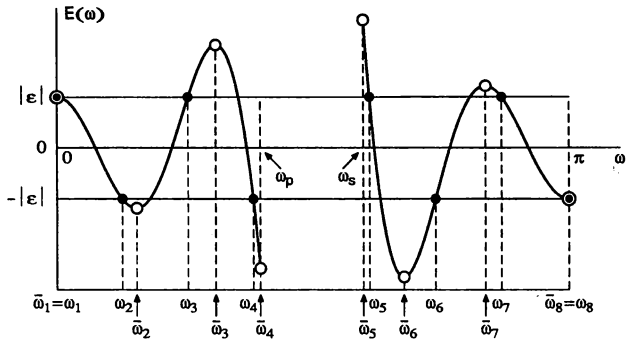


FIGURE 4-24 Error function obtained by forcing it to alternatingly go through the values $\pm\epsilon$ at the selected extremal points ω_k, $k = 1, 2, \ldots, M + 2$. $M = 6$ and $X = [0, \omega_p] \cup [\omega_s, \pi]$. The $\bar{\omega}_k$'s are the true extremal points of the error function.

(4.88) is satisfied at the new extremal points. The process is repeated until the ω_k's become the true extremal points of $E(\omega)$. In the above process, the absolute value of ϵ increases in each iteration loop.

The set of linear equations at Step 2 can be solved conveniently by first calculating ϵ analytically as

$$\epsilon = \frac{\sum_{k=1}^{M+2} b_k \bar{D}(\omega_k)}{\sum_{k=1}^{M+2} b_k (-1)^{k+1} / \bar{W}(\omega_k)}, \tag{4.89a}$$

where

$$b_k = \prod_{\substack{i=1 \\ i \neq k}}^{M+2} \frac{1}{(\cos \omega_k - \cos \omega_i)}. \tag{4.89b}$$

After calculating ϵ, it is known that $G(\omega)$ achieves the value

$$C_k = \bar{D}(\omega_k) + (-1)^k \epsilon / \bar{W}(\omega_k) \tag{4.90}$$

at the kth extremal point. To get around the numerical sensitivity problems, the Lagrange interpolation formula in the barycentric form is used to express $G(\omega)$ as

$$G(\omega) = \frac{\sum_{k=1}^{M+1} \left(\frac{\beta_k}{\cos \omega - \cos \omega_k} \right) C_k}{\sum_{k=1}^{M+1} \left(\frac{\beta_k}{\cos \omega - \cos \omega_k} \right)}, \tag{4.91a}$$

where

$$\beta_k = \prod_{\substack{i=1 \\ i \neq k}}^{M+1} \frac{1}{(\cos \omega_k - \cos \omega_i)}. \tag{4.91b}$$

Note that after solving ϵ, $M + 1$ points, instead of $M + 2$ points, are required to uniquely determine $G(\omega)$. In the MPR algorithm [MC73b, MC79] $G(\omega)$ is expressed in the above form. This is because the actual coefficient values $a[n]$ are not needed in intermediate calculations. After the convergence of the algorithm, the $a[n]$'s are determined by evaluating $G(\omega)$ at $2M + 1$ equally spaced frequency points and then applying the inverse discrete Fourier transform (see Chapter 8). From the $a[n]$'s, the filter coefficients $h[n]$ can then be determined according to the discussion of Section 4-3-1.

In the practical implementation of the MPR algorithm, the extrema of $E(\omega)$ at Step 3 are located by evaluating $E(\omega)$ over a dense set of frequencies spanning the approximation region X. As a rule of thumb, a good selection of the number of grid points is $16M$. Typically, four to eight iterations of the above algorithm are required to arrive at the optimum solution in lowpass cases. In designing filters having several passband and stopband regions, the number of iterations is typically two or three times that required for designing lowpass filters.

4-8-2 Properties of the Optimum Filters

Before illustrating the use of the above algorithm in practical filter design problems, some properties of optimum filters are reviewed. In the lowpass case, the filter design parameters are the passband edge ω_p, the stopband edge ω_s, the passband ripple δ_p, and the stopband ripple δ_s. The remaining parameter to be determined is the minimum filter order N required to meet the given criteria. If N is prescribed, then the ripple ratio

$$k = \delta_p/\delta_s, \tag{4.92}$$

instead of δ_p and δ_s, is usually specified. In the latter case, the optimum solution is obtained by using the following desired response and weighting function in the MPR algorithm:

$$D(\omega) = \begin{cases} 1 & \text{for } \omega \in [0, \omega_p] \\ 0 & \text{for } \omega \in [\omega_s, \pi], \end{cases} \tag{4.93a}$$

$$W(\omega) = \begin{cases} 1 & \text{for } \omega \in [0, \omega_p] \\ k & \text{for } \omega \in [\omega_s, \pi]. \end{cases} \tag{4.93b}$$

In this case, $X = [0, \omega_p] \cup [\omega_s, \pi]$.

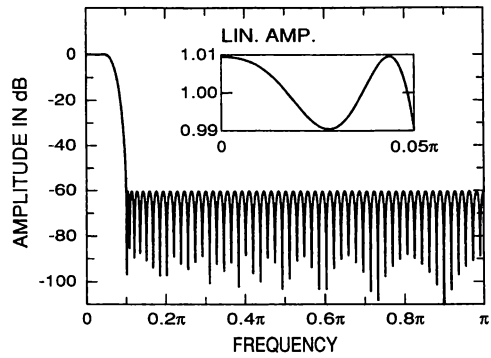


FIGURE 4-25 Amplitude response for an optimum Type I filter of order $N = 2M = 108$.

Example 4.5. Figure 4-25 gives the optimized response of a Type I filter of order $N = 108$ ($M = 54$) for $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, and $k = 10$. The resulting ripples are given by $\delta_p = 0.00955$ and $\delta_s = 0.000955$. In this case, $N = 108$ is the minimum filter order to meet the ripple requirements $\delta_p \leq 0.01$ and $\delta_s \leq 0.001$.

Except for the case of the Chebyshev solutions considered in Section 4-7, there exist no analytic relations between the lowpass filter parameters N, ω_p, ω_s, δ_p, and δ_s. However, rather accurate estimates, based on empirical data, have been reported by Herrmann et al. [HE73], Kaiser [KA74], and Rabiner [RA73c] for the minimum filter order N. Kaiser has proposed the particularly simple formula [KA74]

$$N \approx \frac{-20 \log_{10}(\sqrt{\delta_p \delta_s}) - 13}{14.6[(\omega_s - \omega_p)/(2\pi)]} \tag{4.94}$$

for predicting the filter order N. A somewhat more accurate formula due to Herrmann et al. [HE73] is

$$N \approx \frac{D_\infty(\delta_p, \delta_s) - F(\delta_p, \delta_s)[(\omega_s - \omega_p)/(2\pi)]^2}{(\omega_s - \omega_p)/(2\pi)}, \tag{4.95a}$$

where

$$D_\infty(\delta_p, \delta_s) = [a_1(\log_{10} \delta_p)^2 + a_2 \log_{10} \delta_p + a_3] \log_{10} \delta_s - [a_4(\log_{10} \delta_p)^2 + a_5 \log_{10} \delta_p + a_6] \tag{4.95b}$$

and

$$F(\delta_p, \delta_s) = b_1 + b_2[\log_{10} \delta_p - \log_{10} \delta_s] \tag{4.95c}$$

with

$$a_1 = 0.005309, \quad a_2 = 0.07114, \quad a_3 = -0.4761, \tag{4.95d}$$

$$a_4 = 0.00266, \quad a_5 = 0.5941, \quad a_6 = 0.4278, \tag{4.95e}$$

$$b_1 = 11.01217, \quad b_2 = 0.51244. \tag{4.95f}$$

This formula has been developed for $\delta_s < \delta_p$. If $\delta_s > \delta_p$, then the estimate is obtained by interchanging δ_p and δ_s in the formula.

If the ripples of the filter are rather small, then both formulas give approximately the same result. However, when the ripple values are large, the latter formula gives a better estimate. For this formula, the estimation error is typically less than 2%. From the above formulas, it is seen that the required filter order is roughly inversely proportional to the transition bandwidth.

In the case of the specifications of Example 4.5, both of the above formulas give $N = 101$. For the optimized filter of order 101, the ripples are given by $\delta_p = 0.0157$ and $\delta_s = 0.00157$, showing that the filter order has to be increased. When determining the actual minimum filter order, it must be taken into consideration that sometimes a filter of order $N - 1$ has lower ripple values than a filter of order N. For instance, for the case $\omega_p = 0.6856\pi$, $\omega_s = 0.83246\pi$, and $k = 1$, the Type I filter of order $N = 10$ achieves $\delta_p = \delta_s = 0.1282$, whereas the Type II filter of order $N = 9$ achieves $\delta_p = \delta_s = 0.1$ [RA73b]. Based on this, it is advantageous to determine separately the minimum orders for both Type I filters (N is even) and Type II filters (N is odd), and then to select the lower order. For the specifications of Example 4.5, the minimum orders of Type I and Type II filters to meet the ripple requirements of $\delta_p = 0.01$ and $\delta_s = 0.001$ are 108 and 109, respectively, so that $N = 108$ is the minimum order.

An informative way to study the various types of optimum lowpass filter solutions is to plot the transition bandwidth

$$\Delta\omega = \omega_s - \omega_p \tag{4.96}$$

of the filter versus ω_p for fixed values of N, δ_p, and δ_s [PA73; RA73a, RA73b]. Figure 4-26 gives such plots for Type I optimum filters with $N = 14$ ($M = 7$), $N = 16$ ($M = 8$), and $N = 18$ ($M = 9$) for $\delta_p = \delta_s = 0.1$.[10] As seen from this figure, all three curves alternate between sharp minima and flat-topped maxima. We consider in greater detail the filters corresponding to the six points, denoted by the letters A, B, C, D, E, and F, in the curve for $N = 16$ ($M = 8$). The responses of these filters are given in Figure 4-27. Filters C and F correspond to the points where the local minimum of $\Delta\omega$ occurs with respect to ω_p. These are special *extraripple* or *maximal ripple* solutions whose error function exhibits $M + 3 = 11$

[10]In constructing these curves, $k = 1$ has been used in the MPR algorithm. For each value of ω_p, the minimum transition bandwidth to meet the given ripple requirements has been determined by decreasing the stopband edge ω_s until the filter just meets the ripple requirements.

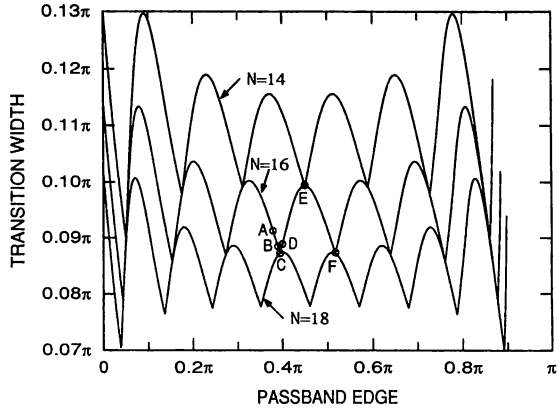


FIGURE 4-26 Transition bandwidth ($\omega_s - \omega_p$) as a function of the passband edge ω_p for $\delta_p = \delta_s = 0.1$ for filters with $N = 14$ ($M = 7$), $N = 16$ ($M = 8$), and $N = 18$ ($M = 9$).

extrema with equal amplitude. This is one more than that required by the characterization theorem. Furthermore, it follows from this theorem that these extraripple solutions are also the optimum solutions for $\overline{M} = M + 1$, or equivalently for $N = 18$. This is because the number of extrema is $\overline{M} + 2$ for these filters. The explanation to this is that the first and last impulse-response coefficients $h[n]$ of the filter with higher order become exactly zero when the filter with lower order has the extraripple solution. This gives $a[\overline{M}] = a[9] = 0$ for the filter with $N = 18$ so that the responses of the two filters coincide.

When ω_p is made smaller, the resulting filter has $M + 2$ equal amplitude extrema, as well as one smaller amplitude extremum at $\omega = 0$ (Filter B). When ω_p is further decreased, the extra extremum disappears (Filter A). On the other hand, if ω_p is made larger, the resulting filter (Filter D) has one smaller ripple at $\omega = \pi$. Also, this ripple disappears when ω_p is further increased. Filter E in Figure 4-27 corresponds to the case where the filter with $N = 14$ has the same solution (extraripple solution for the filter with $N = 14$).

Hence for Type I filters, there are three kinds of optimum solutions: solutions having $M + 2$ equal amplitude extrema, special solutions having $M + 3$ equal amplitude extrema, and solutions having, in addition to $M + 2$ equal amplitude extrema, one smaller extremum. For Type II filters, the properties are quite similar [RA73b]. The basic difference is that the Type II filters have an odd order ($N = 2M + 1$) and they have a fixed zero at $z = -1$ ($\omega = \pi$).

4-8-3 Some Useful Properties of Optimum Type I Filters

Consider a Type I transfer function of the form

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n}, \qquad h[2M - n] = h[n]. \tag{4.97}$$

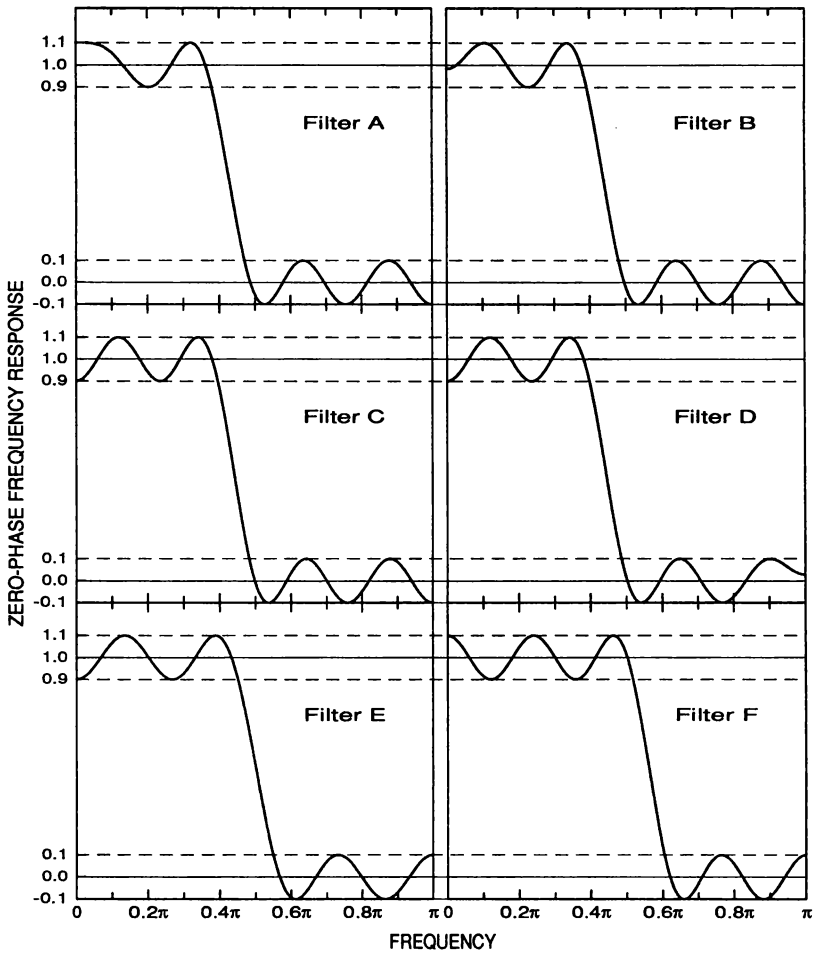


FIGURE 4-27 Responses for six of the filters of Figure 4-26.

The corresponding zero-phase frequency response is given by

$$H(\omega) = h[M] + \sum_{n=1}^{M} 2h[M-n] \cos n\omega. \tag{4.98}$$

On the basis of $H(z)$, we can construct three Type I filters having the following transfer functions:

$$G(z) = \begin{cases} z^{-M} - H(z) & \text{for Case A} \\ (-1)^M H(-z) & \text{for Case B} \\ z^{-M} - (-1)^M H(-z) & \text{for Case C.} \end{cases} \tag{4.99}$$

The zero-phase frequency responses of these three filters can be written as

$$G(\omega) = \begin{cases} 1 - H(\omega) & \text{for Case A} \\ H(\pi - \omega) & \text{for Case B} \\ 1 - H(\pi - \omega) & \text{for Case C.} \end{cases} \tag{4.100}$$

In Case A, the impulse-response coefficients of $G(z)$ are related to the coefficients of $H(z)$ via $g[M] = 1 - h[M]$ and $g[n] = -h[n]$ for $n = 0, 1, \ldots, M - 1$. By substituting these values into

$$G(\omega) = g[M] + \sum_{n=1}^{M} 2g[M - n] \cos n\omega, \tag{4.101}$$

we end up with $G(\omega)$ shown in Eq. (4.100). In Case B, the coefficients $g[n]$ are related to the $h[n]$'s via $g[M - n] = h[M - n]$ for n even and $g[M - n] = -h[M - n]$ for n odd. Using the facts that for n even $\cos n\omega = \cos n(\pi - \omega)$ and for n odd $-\cos n\omega = \cos n(\pi - \omega)$, we can write $G(\omega)$ in the above form. The fact that $G(\omega)$ is expressible in Case C as shown in Eq. (4.100) follows directly from the properties of the Case A and Case B filters.

In Case A, the filter pair $H(z)$ and $G(z)$ is called a *complementary filter pair* since the sum of their zero-phase frequency responses is unity; that is,

$$H(\omega) + G(\omega) = 1. \tag{4.102}$$

This means that if $H(z)$ is a lowpass design with $H(\omega)$ oscillating within the limits $1 \pm \delta_p$ on $[0, \omega_p]$ and within the limits $\pm\delta_s$ on $[\omega_s, \pi]$, then $G(z)$ is a highpass filter with $G(\omega)$ oscillating within $\pm\delta_p$ on $[0, \omega_p]$ and within $1 \pm \delta_s$ on $[\omega_s, \pi]$ (see Figures 4-28(a) and 4-28(b)). An implementation of $G(z)$ is shown in Figure 4-29. The delay term z^{-M} can be shared with $H(z)$ in this implementation. Hence at the expense of one additional adder, a complementary filter pair can be implemented.

If $H(\omega)$ is as shown in Figure 4-28(a), then the Case B filter is a highpass design with $G(\omega)$ oscillating within $\pm\delta_s$ on $[0, \pi - \omega_s]$ and within $1 \pm \delta_p$ on $[\pi - \omega_p, \pi]$ (see Figure 4-28(c)). $G(\omega)$ for the Case C filter, in turn, varies within $1 \pm \delta_s$ on $[0, \pi - \omega_s]$ and within $\pm\delta_p$ on $[\pi - \omega_p, \pi]$ (see Figure 4-28(d)). An implementation of the Case C filter is depicted in Figure 4-29. This implementation is very important in many cases as it allows us to implement a wideband filter $G(z)$ using a delay term and a transfer function that is obtained from a narrowband filter $H(z)$ by simply changing the sign of every second coefficient value. This is because there are computationally efficient implementations for narrowband filters, as will be seen in Section 4-11.

4-8-4 The Use of the MPR Algorithm

The MPR algorithm is very flexible for solving many kinds of approximation problems in the minimax sense. The user specifies first the filter type and the order of

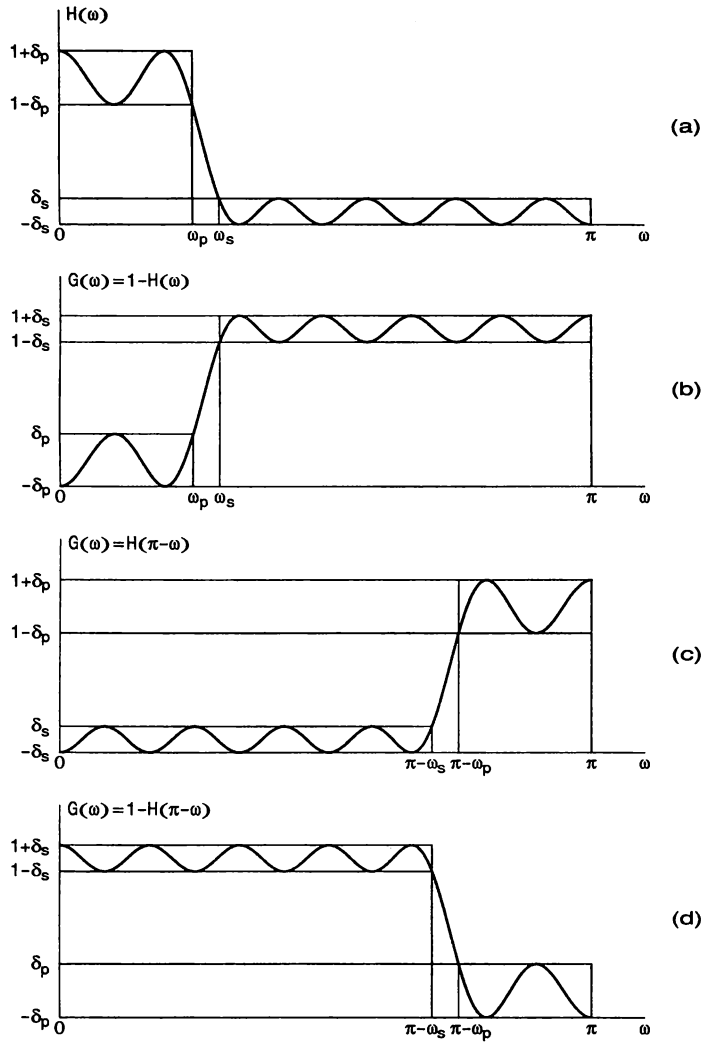
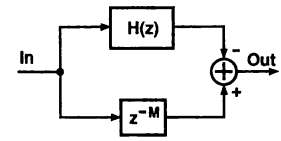
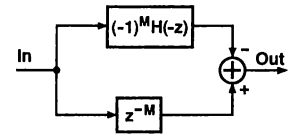


FIGURE 4-28 Responses for the prototype filter and three filters developed on the basis of this prototype filter. (a) Prototype lowpass filter. (b) Case A. (c) Case B. (d) Case C.

the filter. The different filter types are conventional frequency-selective filters having multiple passbands and stopbands, differentiators, and Hilbert transformers. The conventional filters are Type I and Type II filters, whereas in the last two cases, the filters are Type III and Type IV filters considered in greater detail in Chapter 13. A FORTRAN program implementing the MPR algorithm can be found in McClellan et al. [MC73b, MC79]. In this program, instead of the filter order N, the length of the impulse response, $N + 1$, is used. After giving the filter type



Case A



Case C

FIGURE 4-29 Implementations for Case A and Case C filters.

and the length, the program automatically checks whether the filter is a Type I, II, III, or IV design. It also changes the desired function and the weighting function correspondingly and concentrates on determining the common adjustable filter part.

The user also supplies an integer and the program selects the number of grid points to be $M + 1$ times this integer. A good selection for this integer is 16. When designing multiband filters, the user specifies the edges of the bands as well as the desired value and the weighting for each band. If other than piecewise-constant desired functions and weighting functions are desired to be used, the program has subroutines EFF and WATE, which can be used in these cases. It should be noted that the basic frequency variable f of the program is related to the angular frequency ω via

$$f = \omega/(2\pi). \tag{4.103}$$

For instance, if a desired edge angle is 0.4π, then the edge for the program is 0.2. Some examples are now given to illustrate the flexibility of the MPR algorithm.

Example 4.6. It is desired to design a lowpass filter with passband and stopband regions $[0, 0.3\pi]$ and $[0.4\pi, \pi]$, respectively. The passband ripple δ_p is restricted to be at most 0.002 on $[0, 0.15\pi]$ and at most 0.01 in the remaining region $[0.15\pi, 0.3\pi]$. The stopband ripple δ_s is at most 0.0001 (80-dB attenuation) on $[0.4\pi, 0.6\pi]$ and at most 0.001 (60-dB attenuation) on $[0.6\pi, \pi]$. Furthermore, the overall filter is implemented in the form

$$H_{\text{ove}}(z) = H_{\text{fix}}(z)H(z), \tag{4.104}$$

where the fixed term $H_{\text{fix}}(z)$ has zero pairs on the unit circle at the angular frequencies $\omega = \pm 0.4\pi, \pm 0.45\pi, \pm 0.5\pi, \pm 0.55\pi, \pm 0.6\pi, \pm 0.65\pi$. The desired overall filter can be obtained by designing $H(z)$ using the following desired and

weighting functions:[11]

$$D(\omega) = \begin{cases} 1/|H_{\text{fix}}(\omega)|, & \omega \in [0,\ 0.3\pi] \\ 0, & \omega \in [0.4\pi,\ \pi], \end{cases} \tag{4.105a}$$

$$W(\omega) = \begin{cases} 5|H_{\text{fix}}(\omega)|, & \omega \in [0,\ 0.15\pi] \\ |H_{\text{fix}}(\omega)|, & \omega \in [0.15\pi,\ 0.3\pi] \\ 100|H_{\text{fix}}(\omega)|, & \omega \in [0.4\pi,\ 0.6\pi] \\ 10|H_{\text{fix}}(\omega)|, & \omega \in [0.6\pi,\ \pi]. \end{cases} \tag{4.105b}$$

The given criteria are met when the peak absolute value of the corresponding error function becomes smaller than or equal to 0.01. The minimum order of $H(z)$ to meet the criteria is 54. The amplitude response of the resulting overall design is depicted in Figure 4-30.

Example 4.7. This example illustrates the use of the MPR algorithm for designing FIR filters with a very flat passband and equiripple stopband. These filters have been proposed by Vaidyanathan [VA85] and their transfer function is of the form

$$H(z) = z^{-M} - (-1)^L[(1 - z^{-1})/2]^{2L}\overline{H}(z), \tag{4.106}$$

[11]In general, if the overall filter is of the form $H_{\text{ove}}(z) = H_{\text{fix}}(z)H(z)$ and the desired and weighting functions for $H_{\text{ove}}(z)$ are $D(\omega)$ and $W(\omega)$, respectively, then the given criteria are met by designing $H(z)$ using the desired function $D(\omega)/H_{\text{fix}}(\omega)$ and the weighting function $H_{\text{fix}}(\omega)W(\omega)$. This follows from the fact that the weighted error function can be written as $E(\omega) = W(\omega)[H_{\text{ove}}(\omega) - D(\omega)] = W(\omega)H_{\text{fix}}(\omega)[H(\omega) - D(\omega)/H_{\text{fix}}(\omega)]$. If $H_{\text{fix}}(\omega)$ is zero at some points in the approximation interval, then the new weighting function becomes zero at these points. This problem can be avoided by disregarding these grid points when using the MPR algorithm. The absolute values of $H_{\text{fix}}(\omega)$ are used in Eq. (4.105) to make the weighting function positive.

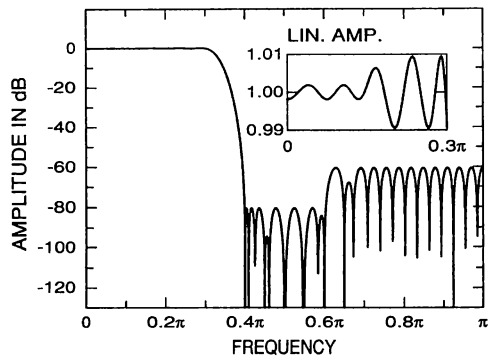


FIGURE 4-30 Amplitude response for a filter having some fixed zero pairs on the unit circle and unequal passband and stopband weightings.

where $M = L + K$ with K being half the order of $\overline{H}(z)$. Here, $\overline{H}(z)$ of even order $2K$ is designed using the following desired and weighting functions:

$$D(\omega) = \begin{cases} 0, & \omega \in [0, \omega_p] \\ 1/[\sin(\omega/2)]^{2L}, & \omega \in [\omega_s, \pi], \end{cases} \tag{4.107a}$$

$$W(\omega) = \begin{cases} [\sin(\omega_p/2)]^{2L}, & \omega \in [0, \omega_p] \\ (\delta_p/\delta_s)\,[\sin(\omega/2)]^{2L}, & \omega \in [\omega_s, \pi], \end{cases} \tag{4.107b}$$

where $[\sin(\omega/2)]^{2L}$ is the zero-phase frequency response of $(-1)^L[(1 - z^{-1})/2]^{2L}$. Figure 4-31 gives the resulting overall response

$$H(\omega) = 1 - [\sin(\omega/2)]^{2L}\overline{H}(\omega) \tag{4.108}$$

for the case $\omega_p = 0.6\pi$, $\omega_s = 0.7\pi$, $\delta_p \leq 0.016$, $\delta_p/\delta_s = 5$, and $2L = 16$ [VA85]. The minimum even order of $\overline{H}(z)$ to meet the criteria is $2K = 46$. In the above, the desired and weighting functions have been selected such that $[\sin(\omega/2)]^{2L}\overline{H}(\omega)$ achieves the peak absolute value of the corresponding error function ($\epsilon = 0.0144$) at $\omega = \omega_p$ and oscillates within the limits $1 \pm \epsilon/5$ on $[\omega_s, \pi]$. Correspondingly, $H(\omega)$ achieves the value $1 - \epsilon$ at $\omega = \omega_p$ and oscillates within the limits $\pm\epsilon/5$ in the stopband. On $[0, \omega_p]$, $[\sin(\omega_p/2)]^{2L}\overline{H}(\omega)$ varies within the limits $\pm\epsilon$. Because of the term $[\sin(\omega/2)]^{2L}$, $[\sin(\omega/2)]^{2L}\overline{H}(\omega)$ approximates very accurately zero and $H(\omega)$ approximates very accurately unity in the beginning of the interval $[0, \omega_p]$. Note that the fixed term has $2L$ zeros at $z = 1$ (at $\omega = 0$).

Example 4.8. Let the bandpass filter specifications be $\omega_{s1} = 0.2\pi$, $\omega_{p1} = 0.25\pi$, $\omega_{p2} = 0.6\pi$, $\omega_{s2} = 0.7\pi$, $\delta_{s1} = 0.001$, and $\delta_p = \delta_{s2} = 0.01$. In this case, the desired function is unity in the passband and zero in the stopbands, whereas the

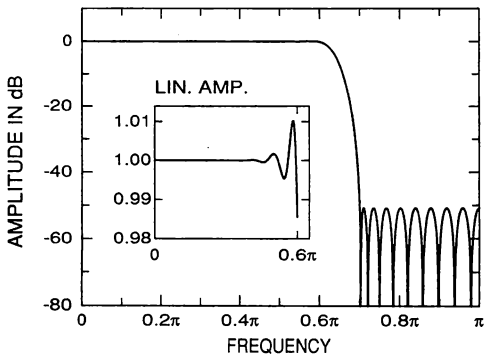


FIGURE 4-31 Amplitude response for a filter with a very flat passband and equiripple stopband.

weighting function is 10 in the first stopband and 1 in both the passband and the second stopband. The minimum order to meet the above specifications is 102. Figure 4-32(a) gives the response of the optimum filter. This response is optimal according to the characterization theorem even though it has an unacceptable transition band peak of 15 dB. This is possible because the approximation is restricted to the passband and stopband regions only and the transition bands are considered as don't-care bands. For designs with a single transition band there are no unacceptable transition band ripples. However, for filters having more than one transition band, this phenomenon of large transition band peaks occurs when the widths of the transition bands are different [RA74]; the larger the difference, the greater the problem.

The transition band peak can easily be attenuated by including the transition bands in the overall approximation interval and requiring that the response stays within the limits $-\delta_{s1}$ and $1 + \delta_p$ in the first transition band and within the limits $-\delta_{s2}$ and $1 + \delta_p$ in the second transition band. This can be done by selecting the desired function to be $\frac{1}{2}(1 + \delta_p - \delta_{s1})$ and $\frac{1}{2}(1 + \delta_p - \delta_{s2})$ in the first and second

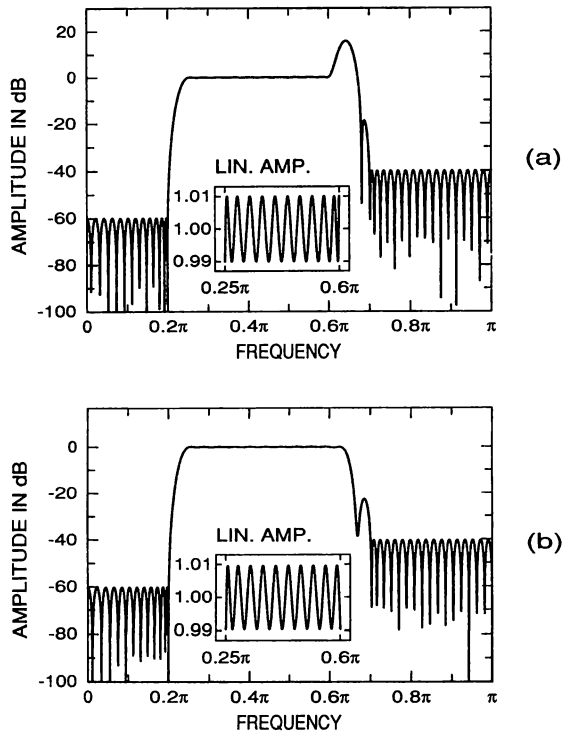


FIGURE 4-32 Amplitude responses for bandpass filters. (a) Filter designed without transition band constraints. (b) Filter designed with transition band constraints.

transition bands, respectively. If the weighting in the passband is unity, then the weighting functions in the transition bands are selected to be $\delta_p/[\frac{1}{2}(1 + \delta_p + \delta_{s1})]$ and $\delta_p/[\frac{1}{2}(1 + \delta_p + \delta_{s2})]$, respectively. These selections guarantee that if the passband ripple of the resulting filter is less than or equal to the specified δ_p, then the response stays within the desired limits in the transition bands.[12] When including the transition bands in the approximation problem, the filter order has to be increased only by one (to 103) to meet the resulting specifications. Figure 4-32(b) gives the response of this filter. For other techniques for attenuating undesired transition band ripples, see Rabiner et al. [RA74].

4-9 DESIGN OF MINIMUM-PHASE FIR FILTERS

The attractive property of Type I and Type II linear-phase FIR filters is that their delay is a constant and thus they cause no phase distortion to the signal. The delay is equal to half the filter order. This means that the delay becomes very long for high-order filters required in cases demanding a narrow transition band. In some applications, such a long delay is not tolerable. In those cases, a smaller group delay can be achieved in the passband region by using minimum-phase FIR filters. There exist also applications where linear phase is not required and the symmetry in the coefficients of linear-phase FIR filters cannot be exploited. In those cases, nonlinear-phase filters meet the same amplitude criteria with a reduced number of multipliers and delay elements. If the passband of the filter is very wide, then a saving by almost a factor of 2 can be achieved in the filter order [GO81; LE75]. This section outlines the design of nonlinear-phase filters based on the design scheme of Herrmann and Schüssler [HE70], which can be used for synthesizing filters with unweighted stopband response. For more general techniques, see Boite and Leich [BO81], Kamp and Wellekens [KA83b], and Grenez [GR83].

Consider a nonlinear-phase FIR filter with transfer function

$$H(z) = \sum_{n=0}^{M} h[n]z^{-n}. \tag{4.109}$$

The zeros of the transfer function

$$\hat{H}(z) = z^{-M}H(z^{-1}) = \sum_{n=0}^{M} h[M - n]z^{-n} \tag{4.110}$$

are reciprocal to those of $H(z)$. This implies that the function

$$G(z) = H(z)\hat{H}(z) = z^{-M}H(z)H(z^{-1}) \tag{4.111}$$

[12]The lower and upper edges for the first transition band are selected as $\omega_{s1} + \alpha$ and $\omega_{p1} - \alpha$, respectively, where α is a small number. Similarly, for the second transition band, the edges are $\omega_{p2} + \alpha$ and $\omega_{s2} - \alpha$. These selections prevent the desired function from becoming discontinuous at the edges.

is the transfer function of a Type I linear-phase filter of order $2M$. Since $G(z)$ must be factorizable into the terms $H(z)$ and $\hat{H}(z)$, its zeros on the unit circle have to be double. From the above equation, it follows also that the magnitude-squared function of $H(z)$ can be expressed as

$$|H(e^{j\omega})|^2 = G(\omega). \tag{4.112}$$

Since $G(z)$ possesses double zeros on the unit circle, $G(\omega)$ has double zeros on $[0, \pi]$, making it nonnegative on $[0, \pi]$. These facts show that the design of a nonlinear-phase FIR filter of order M can be accomplished in terms of a Type I linear-phase filter of order $2M$ having double zeros on the unit circle.

Based on this, Herrmann and Schüssler [HE70] have proposed the following simple design procedure:

1. Design a Type I linear-phase FIR filter transfer function $\overline{G}(z)$ of order $2M$ using the MPR algorithm such that $\overline{G}(\omega)$ oscillates within the limits $1 \pm \bar{\delta}_p$ in the passband $[0, \omega_p]$ and within the limits $\pm\bar{\delta}_s$ in the stopband $[\omega_s, \pi]$ (see Figure 4-33(a)). This $\overline{G}(z)$ has single zeros on the unit circle.
2. Form $G(z) = \bar{\delta}_s z^{-M} + \overline{G}(z)$. The resulting $G(\omega) = \bar{\delta}_s + \overline{G}(\omega)$ is nonnegative on $[\omega_s, \pi]$, oscillating within zero and $2\bar{\delta}_s$ (see Figure 4-33(b)). On $[0, \omega_p]$, $G(\omega)$ oscillates within the limits $1 + \bar{\delta}_s \pm \bar{\delta}_p$. $G(\omega)$ has double zeros at those frequency points where $\overline{G}(\omega)$ takes the stopband minimum value of $-\bar{\delta}_s$. Correspondingly, $G(z)$ has double zeros on the unit circle at those frequencies.
3. Perform the factorization of $G(z) = H(z)z^{-M}H(z^{-1})$ such that $H(z)$ contains the zeros inside the unit circle and one each of the double zeros on the unit circle. Scale $H(z)$ such that the passband average of the resulting filter $AH(z)$ is equal to unity (see Figure 4-33(c)).

The desired scaling constant at Step 3 is

$$A = \frac{2}{\sqrt{1 + \bar{\delta}_p + \bar{\delta}_s} + \sqrt{1 - \bar{\delta}_p + \bar{\delta}_s}}. \tag{4.113}$$

If it is required that the magnitude response of the scaled filter $AH(z)$ approximates unity in the passband with tolerance δ_p and zero in the stopband with tolerance δ_s, then the passband and stopband ripples of the linear-phase filter at Step 1 must satisfy

$$\bar{\delta}_p \le \frac{2\delta_p}{1 + (\delta_p)^2 - (\delta_s)^2/2}, \qquad \bar{\delta}_s \le \frac{(\delta_s)^2/2}{1 + (\delta_p)^2 - (\delta_s)^2/2}. \tag{4.114}$$

The most difficult part in the above procedure is the factorization of $G(z)$ into the terms $H(z)$ and $z^{-M}H(z^{-1})$. The direct approach is simply to pick up the zeros of $G(z)$. However, if the order of $G(z)$ is high, conventional root-finding proce-

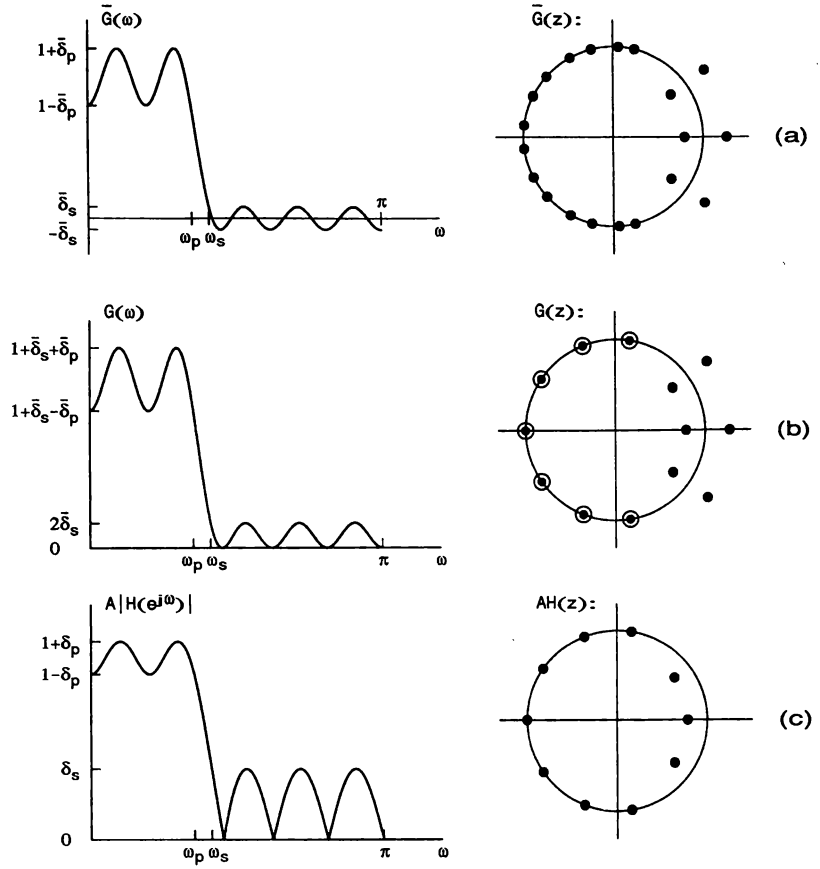


FIGURE 4-33 Steps for designing an equiripple minimum-phase FIR filter of order 10 with the aid of a linear-phase FIR filter of order 20.

dures cannot be used for locating the zeros. Another approach is to perform the factorization without finding the roots of $G(z)$. Such techniques have been proposed independently by Boite and Leich [BO81] and Mian and Naider [MI82a] (see also Boite and Leich [BO84]). For a review of different techniques for performing the factorization, see Schüssler and Steffen [SC88].

The filter obtained by selecting the zeros to lie on or inside the unit circle is called a *minimum-phase* filter. If the zeros outside the unit circle are selected, then the resulting filter is called a *maximum-phase* design.

Example 4.9. Let the specifications be $\omega_p = 0.5\pi$, $\omega_s = 0.6\pi$, $\delta_p = 0.01$, and $\delta_s = 0.00316$ (50-dB attenuation). Using Eq. (4.114), the ripples of $\bar{G}(z)$ at Step 1

become $\bar{\delta}_p \approx 0.02$ and $\bar{\delta}_s \approx 5 \times 10^{-6}$. The minimum even order to meet these criteria is 74 so that the order of the corresponding minimum-phase filter is 37. The common amplitude response of the minimum-phase and maximum-phase filters as well as their group delay responses, zero locations, and impulse responses are given in Figure 4-34. The minimum order of a linear-phase filter to meet the same criteria is 46 (see Figure 4-8) so that the saving in the filter order provided by the minimum-phase filter is 20%. For the minimum- and maximum-phase fil-

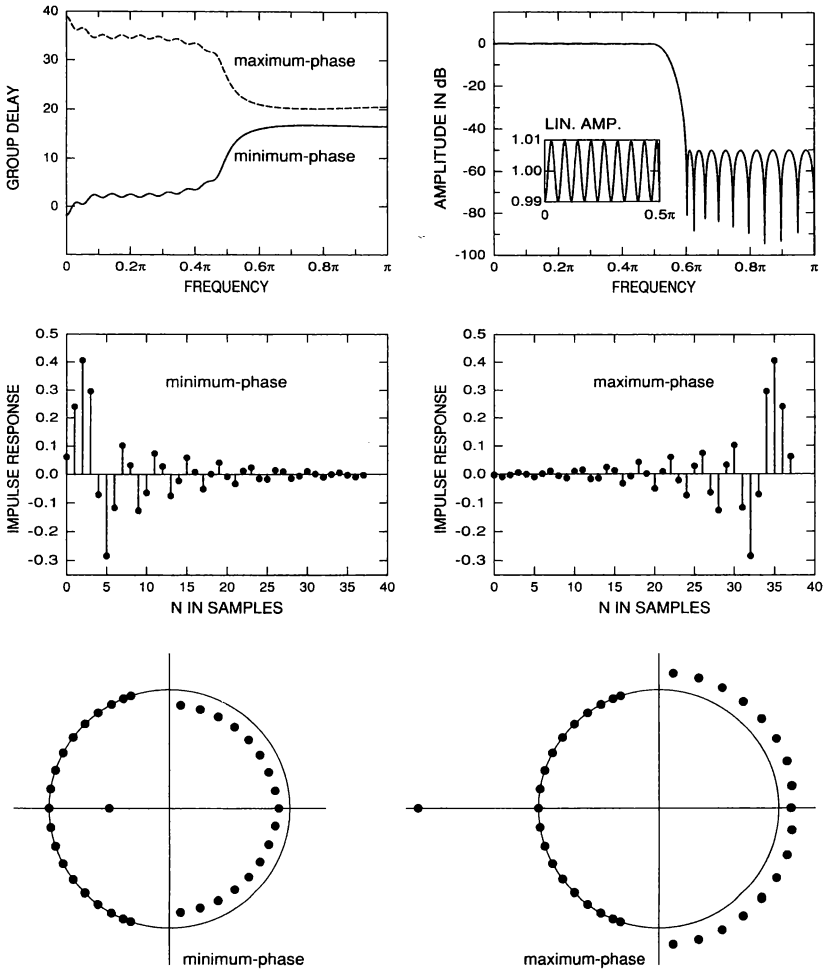


FIGURE 4-34 Group delay responses, amplitude response, zero locations, and impulse responses for minimum- and maximum-phase FIR filters having the same amplitude response.

ters, the phase and the group delay responses have the smallest and largest values, respectively, among the filters having the same amplitude response. It is also interesting to observe from Figure 4-34 that most of the energy of the impulse response of the minimum-phase (maximum-phase) term is concentrated in the beginning (end) of the response.

4-10 DESIGN OF FIR FILTERS WITH CONSTRAINTS IN THE TIME OR FREQUENCY DOMAIN

The previous sections concentrated only on designing FIR filters to meet the given amplitude criteria in some sense. However, there exist applications where there are constraints in the time domain or in the frequency domain. For example, in some applications the transient part of the step response must be constrained to vary within given limits [LI83b; RA72a, RA72b]. Another example is the design of Nyquist filters or Lth band filters with every Lth impulse-response value being zero except for the central value [LI85; MI82b; SA87b, SA88c; VA87]. Furthermore, in some cases, there are flatness constraints in the passband response of the filter [KA83a; ST79].

This section considers techniques for solving the above-mentioned approximation problems. In some cases, the desired solution can be obtained by properly modifying the design methods proposed in the previous sections. In the remaining cases, new techniques are required. Perhaps the most flexible design method for finding the optimum solution to various constrained approximation problems is linear programming [HE71a; KA83a; LI83b; RA72a, RA72b; ST79]. The advantage of this technique is that the convergence to the optimum solution is guaranteed. With linear programming, it is also possible to find the optimum solution to the unconstrained minimax approximation problems considered previously. The disadvantage, however, is that the required computation to arrive at the optimum solution is rather large. Therefore it is preferred to use linear programming only in those cases that cannot be handled with other faster design techniques.

4-10-1 Linear Programming Approach for FIR Filter Design

Linear programming is a very flexible approach for solving many constrained approximation problems in the minimax sense. Mathematically, the linear programming problem [DA63; HA63; LA73] can be stated in the form of the following primal problem: find the unknowns x_k, $k = 1, 2, \ldots, N$, subject to the constraints

$$x_k \geq 0, \qquad k = 1, 2, \ldots, N, \tag{4.115a}$$

$$\sum_{k=1}^{N} \gamma_{lk} x_k = \beta_l, \qquad l = 1, 2, \ldots, M\ (M < N), \tag{4.115b}$$

such that

$$\sigma = \sum_{k=1}^{N} \alpha_k x_k \tag{4.115c}$$

is minimized. In this problem, γ_{lk}, α_k, and β_l are constants. The above problem is mathematically equivalent to the following dual problem: find the unknowns y_l, $l = 1, 2, \ldots, M$, subject to the constraints

$$\sum_{l=1}^{M} \gamma_{lk} y_l \leq \alpha_k, \qquad k = 1, 2, \ldots, N, \tag{4.116a}$$

such that

$$\rho = \sum_{l=1}^{M} \beta_l y_l \tag{4.116b}$$

is maximized.

For digital filter design problems, the dual problem is the most natural form. There exist several well-defined procedures [DA63; HA63; LA73] for arriving at the desired solution within $M + N$ iterations. Lim has introduced an efficient special purpose algorithm for designing FIR filters [LI83b]. This is faster than general purpose algorithms.

Linear programming can be applied in a straightforward manner to those problems where the approximating function is linear; that is, it can be expressed in the form

$$H(\omega) = \sum_{n=0}^{R} b[n]\Phi(\omega, n), \tag{4.117}$$

where the $b[n]$'s are unknowns. According to the discussion of Section 4-3-3, the zero-phase frequency response of a linear-phase FIR filter can be expressed in all four cases in the above form (see Eq. (4.32)). Also, in many other cases, the approximating function can be written in this form. For instance, in the conventional frequency-sampling methods, the filter response is expressible in the above form [RA72a, RA72b].

A general constrained frequency-domain approximation problem, which can be solved using linear programming, can be stated in the following form: find the unknowns $b[n]$ to minimize

$$\delta_1 = \max_{\omega \in X_1} |E(\omega)|, \tag{4.118a}$$

where

$$E(\omega) = W(\omega)[H(\omega) - D(\omega)] \tag{4.118b}$$

subject to

$$\max_{\omega \in X_2} |E(\omega)| \leq \delta_2. \tag{4.118c}$$

Here, X_1 contains a part of the passband and stopband regions and X_2 contains the remaining part. For instance, by selecting X_1 and X_2 to be the stopband and passband regions of the filter, respectively, the stopband variation can be minimized for the given maximum allowable passband variation. Problems of this kind cannot be solved directly using the MPR algorithm considered in Section 4-8.

By sampling $W(\omega)$ and $D(\omega)$ along a dense grid of frequencies $\omega_1^{(1)}, \omega_2^{(1)}, \ldots, \omega_{K_1}^{(1)}$ on X_1, and along a grid of frequencies $\omega_1^{(2)}, \omega_2^{(2)}, \ldots, \omega_{K_2}^{(2)}$ on X_2, the problem can be stated in the form of the dual problem as follows: find $b[0], b[1], \ldots, b[R]$, and δ_1 subject to the constraints

$$\sum_{n=0}^{R} b[n]\Phi(\omega_k^{(1)}, n) - \delta_1 / W(\omega_k^{(1)}) \leq D(\omega_k^{(1)}), \qquad k = 1, 2, \cdots, K_1, \tag{4.119a}$$

$$-\sum_{n=0}^{R} b[n]\Phi(\omega_k^{(1)}, n) - \delta_1 / W(\omega_k^{(1)}) \leq -D(\omega_k^{(1)}), \qquad k = 1, 2, \cdots, K_1, \tag{4.119b}$$

$$\sum_{n=0}^{R} b[n]\Phi(\omega_k^{(2)}, n) \leq D(\omega_k^{(2)}) + \delta_2 / W(\omega_k^{(2)}), \qquad k = 1, 2, \cdots, K_2, \tag{4.119c}$$

$$-\sum_{n=0}^{R} b[n]\Phi(\omega_k^{(2)}, n) \leq -D(\omega_k^{(2)}) + \delta_2 / W(\omega_k^{(2)}), \qquad k = 1, 2, \cdots, K_2, \tag{4.119d}$$

such that

$$\rho = -\delta_1 \tag{4.119e}$$

is maximized.

Note that in the dual problem the constraints are formed in such a way that a linear combination of the unknowns is less than or equal to a constant. In the above problem, δ_2 is a constant and δ_1 is an unknown. This explains the difference between Eqs. (4.119a) and (4.119b) and Eqs. (4.119c) and (4.119d). The above equations have been constructed such that, after finding the optimum solution, $-\delta_1 \leq E(\omega_k^{(1)}) \leq \delta_1$ and $-\delta_2 \leq E(\omega_k^{(2)}) \leq \delta_2$ at the selected grid points. Note also that in the dual problem a linear combination of unknowns is maximized and maximizing $-\delta_1$ implies minimizing δ_1.

It is easy to include in the dual problem various constraints that are expressible in the form of Eq. (4.116a). For instance, it is straightforward to add constraints

of the form

$$\left.\frac{d^l H(\omega)}{d^l \omega}\right|_{\omega=\omega_k} = \sum_{n=0}^{R} b[n] \left.\frac{d^l \Phi(\omega, n)}{d^l \omega}\right|_{\omega=\omega_k} \leq 0 \qquad (4.120a)$$

or

$$\left.\frac{d^l H(\omega)}{d^l \omega}\right|_{\omega=\omega_k} = \sum_{n=0}^{R} b[n] \left.\frac{d^l \Phi(\omega, n)}{d^l \omega}\right|_{\omega=\omega_k} \geq 0, \qquad (4.120b)$$

where l is an integer and ω_k is a grid point. Here, the constraint expressed by Eq. (4.120a) is directly in the desired form. The constraint of Eq. (4.120b) can be written in the form of Eq. (4.116a) by multiplying the left-hand side by -1 and replacing $\geq$ by $\leq$. By adding a constraint of the form of Eq. (4.120a) with $l = 1$ at each grid point in the passband region, the passband response of the filter can be forced to be monotonically decreasing. Steiglitz has presented a FORTRAN code for designing filters of this kind [ST79]. Furthermore, the first L derivatives of $H(\omega)$ can be forced to be zero at $\omega = \omega_k$ by simultaneously using the constraints of Eqs. (4.120a) and (4.120b) for $l = 1, 2, \ldots, L$.

In addition, if it is desired that $H(\omega)$ achieve exactly the value A at $\omega = \omega_k$, this condition can be included by using the following two constraints:

$$\sum_{n=0}^{R} b[n]\Phi(\omega_k, n) \leq A, \qquad -\sum_{n=0}^{R} b[n]\Phi(\omega_k, n) \leq -A. \qquad (4.121)$$

Example 4.10. Consider the design of a Type I linear-phase filter of order 70 having $[0, 0.3\pi]$ and $[0.4\pi, \pi]$ as the passband and stopband regions, respectively. To illustrate the flexibility of linear programming, several constraints are included. First, the filter has fixed zero pairs at the angular frequencies $\pm 0.4\pi$, $\pm 0.45\pi$, $\pm 0.5\pi$, $\pm 0.55\pi$, $\pm 0.6\pi$, and $\pm 0.65\pi$, and $H(\omega)$ achieves the value of unity at $\omega = 0$ with its first four derivatives being zero at this point. Second, the maximum deviation from unity on $[0, 0.15\pi]$ is 0.002 and the maximum deviation from zero on $[0.4\pi, 0.6\pi]$ is 0.0001, whereas the response is desired to be optimized in the remaining regions with weighting of unity on $[0.15\pi, 0.3\pi]$ and 10 on $[0.6\pi, \pi]$. The last part of this problem can be expressed in the form of Eq. (4.119) using $X_1 = [0.15\pi, 0.3\pi] \cup [0.6\pi, \pi]$ and $X_2 = [0, 0.15\pi] \cup [0.4\pi, 0.6\pi]$. $D(\omega)$ is 1 on $[0, 0.3\pi]$ and 0 on $[0.4\pi, \pi]$. $W(\omega)$ is 1 on $[0, 0.3\pi]$, 20 on $[0.4\pi, 0.6\pi]$, and 10 on $[0.6\pi, \pi]$, whereas $\delta_2 = 0.002$. To include the first part, Eq. (4.121) is used with $A = 0$ at the frequency points where the filter has fixed zeros and with $A = 1$ at the zero frequency. Equations (4.120a) and (4.120b) are used with $l = 1, 2, 3, 4$ at the zero frequency. The optimized filter response is shown in Figure 4-35. The resulting ripple values on $[0.15\pi, 0.3\pi]$ and $[0.6\pi, \pi]$ are 0.00637 and 0.000637, respectively.

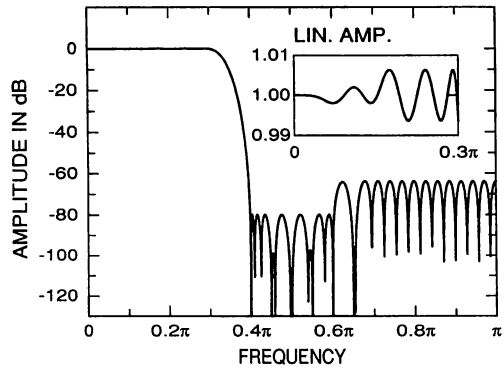


FIGURE 4-35 Amplitude response for an optimized filter of order 70 having several constraints in the frequency domain.

It is also straightforward to include time-domain constraints in the approximation problem. For instance, some of the unknowns $b[n]$, $n \in S$ can be fixed and the remaining ones can be optimized. In this case, the desired solution can be found by using the following approximating function,

$$H(\omega) = \sum_{\substack{n=0 \\ n \notin S}}^{R} b[n]\Phi(\omega, n), \tag{4.122}$$

and by including the effect of the fixed terms in the desired function by changing it to be

$$\overline{D}(\omega) = D(\omega) - \sum_{n \in S} b[n]\Phi(\omega, n). \tag{4.123}$$

Example 4.11. This example shows how linear programming can be used for designing filters with constraints on the step response, which is related to the impulse-response coefficients $h[n]$ through

$$g[n] = \sum_{m=0}^{n} h[m]. \tag{4.124}$$

As an example, Figures 4-36(a) and 4-36(b) give the amplitude and step responses for a filter of order 46 optimized without any constraints in the time domain. For this filter, the passband and stopband ripples are related via $\delta_p = \sqrt{10}\delta_s$, $\omega_p = 0.5\pi$, and $\omega_s = 0.6\pi$. The maximum undershoot of the step response occurring at $n = 21$ is -0.0921 and $g[n] = 0.9903$ for $n \geq 46$. It is desired that $g[n] = 1$ for $n \geq 46$ and

$$-\delta_{\text{step}} \leq g[n] \leq \delta_{\text{step}} \quad \text{for } 0 \leq n \leq K, \tag{4.125}$$

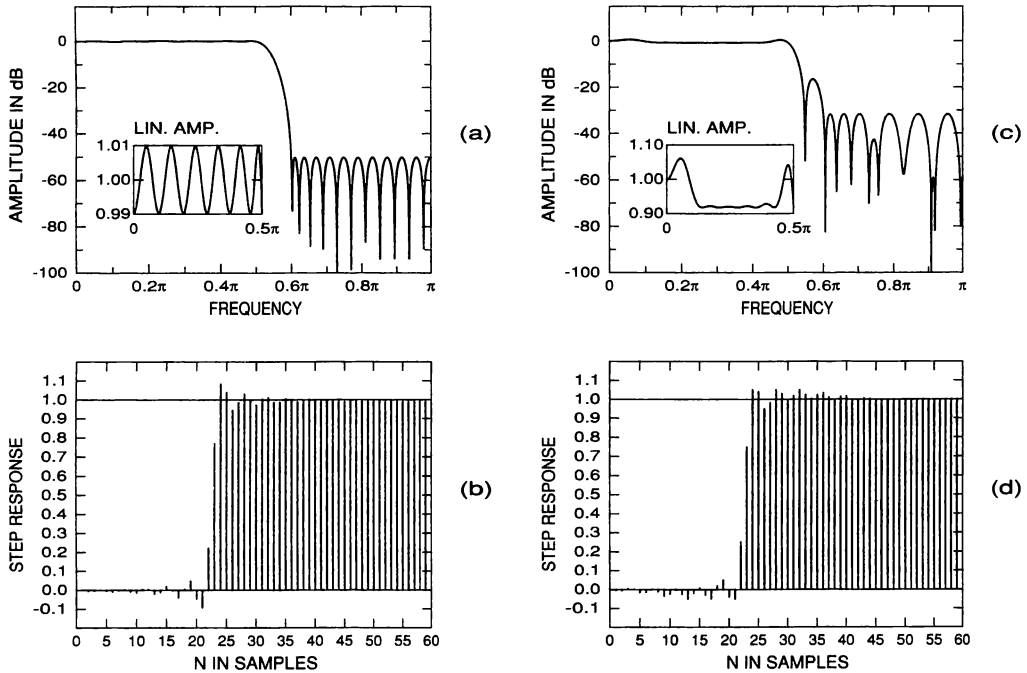


FIGURE 4-36 Amplitude and step responses for filters of order 46. (a, b) Optimum equiripple filter. (c, d) Filter optimized with constraints on the step response.

where $K = 21$ and $\delta_{\text{step}} = 0.05$. The first condition can be satisfied by requiring that $H(0) = 1$. The second constraint is linear in the $h[n]$'s and can thus easily be included in the dual problem. Because of this condition, it is advantageous to express $H(\omega)$ directly in terms of the $h[n]$'s as ($M = N/2$ for Type I designs)

$$H(\omega) = h[M] + \sum_{n=1}^{M} h[M - n](2 \cos n\omega), \tag{4.126}$$

so that $\Phi(\omega, M) = 1$ and $\Phi(\omega, M - n) = 2 \cos n\omega$ for $n > 0$. The amplitude and step responses for the filter optimized with the above constraints are shown in Figures 4-36(c) and 4-36(d), respectively. It is seen that these time-domain conditions increase significantly the ripples of the amplitude response. For other examples for designing filters having constraints on the step response, see Lim [LI83b].

4-10-2 Design of Lth Band (Nyquist) Filters

Consider again a Type I linear-phase FIR filter with transfer function

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n}, \qquad h[2M - n] = h[n]. \tag{4.127}$$

This filter is defined to be an Lth band filter if its coefficients satisfy[13] (see Figure 4-37)

$$h[M] = 1/L, \tag{4.128a}$$

$$h[M + rL] = 0 \quad \text{for } r = \pm 1, \pm 2, \cdots, \pm \lfloor M/L \rfloor . \tag{4.128b}$$

These filters, also called Nyquist filters, play an important role in designing digital transmission systems [LI85; MI82b; SA87b, SA88c; VA87] and filter banks [VA89]. They can also be used as efficient decimators and interpolators since every Lth impulse-response coefficient is zero except for the central coefficient. An important subclass of these filters are half-band filters, which are considered in greater detail in Section 4-10-3.

It can be shown [MI82b] that the time-domain conditions of Eq. (4.128) imply some limitations on the frequency response of the filter. First, the passband edge (in the lowpass case) is restricted to be less than π/L and the stopband edge to be larger than π/L. Usually, the edges are given in terms of an excess bandwidth factor ρ as follows (see Figure 4-37):

$$\omega_p = (1 - \rho)\pi/L, \qquad \omega_s = (1 + \rho)\pi/L. \tag{4.129}$$

Second, if the maximum deviation of $H(\omega)$ from zero on $[\omega_s, \pi]$ is δ_s, then the maximum deviation of $H(\omega)$ from unity on $[0, \omega_p]$ is in the worst case $\delta_p =$

[13] $\lfloor x \rfloor$ stands for integer part of x.

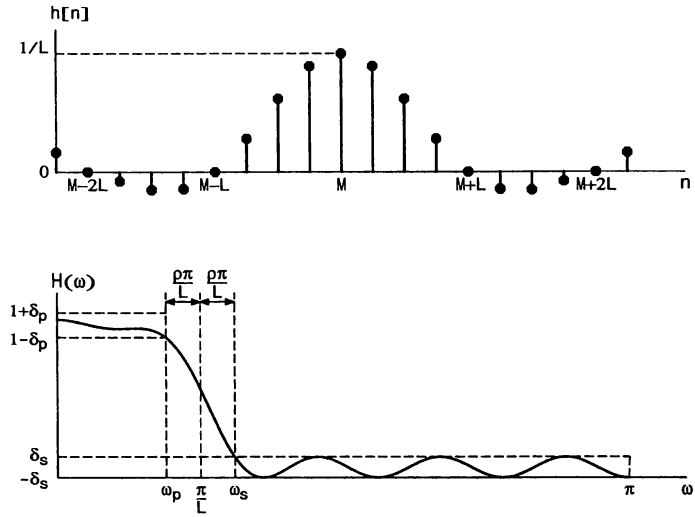


FIGURE 4-37 Typical impulse response and zero-phase frequency response for an FIR Nyquist filter.

$(L-1)\delta_s$. Usually, δ_p is much smaller than this upper limit. Since δ_p is guaranteed to be relatively small for a small value of δ_s, the filter synthesis can concentrate on shaping the stopband response.

The stopband response can be optimized either in the minimax sense or in the least-mean-square sense. In the case of the minimax criterion, the problem is to find the coefficients of $H(z)$ such that the time-domain conditions of Eq. (4.128) are satisfied and

$$\delta_s = \max_{\omega \in [\omega_s, \pi]} |W(\omega)H(\omega)| \tag{4.130}$$

is minimized, where $W(\omega)$ is a positive weighting function. In the case of the least-mean-square criterion, the quantity to be minimized is

$$E_2 = \int_{\omega_s}^{\pi} [W(\omega)H(\omega)]^2 \, d\omega. \tag{4.131}$$

In some applications, it is desired to factorize $H(z)$ into the minimum-phase and maximum-phase terms. In this case, an additional constraint that $H(\omega)$ be non-negative is required. This subsection concentrates on minimax designs. For least-squared-error filters, see Vaidyanathan and Nguyen [VA87] and Nguyen et al. [NG88].

In order to find $H(z)$ minimizing δ_s as given by Eq. (4.130) and simultaneously

meeting the time-domain conditions of Eq. (4.128), it is split into two parts [SA87b] as follows:

$$H(z) = H_p(z)H_s(z) = \sum_{n=0}^{2K} h_p[n]z^{-n} \sum_{n=0}^{2(M-K)} h_s[n]z^{-n}, \qquad (4.132a)$$

where

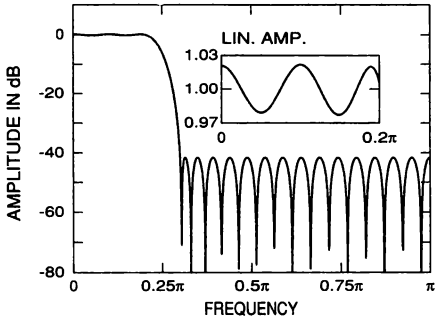
$$K = \lfloor M/L \rfloor . \qquad (4.132b)$$

Here, both $H_p(z)$ and $H_s(z)$ are Type I linear-phase filters. $H_p(z)$ has its zeros off the unit circle and is determined such that the time-domain conditions of Eq. (4.128) are satisfied, whereas $H_s(z)$ has its zeros on the unit circle and is used for providing the desired stopband response. For any $H_s(z)$, $H_p(z)$ can be determined such that the overall filter $H(z) = H_p(z)H_s(z)$ satisfies the time-domain conditions. This leads to a system of $2\lfloor M/L \rfloor + 1$ linear equations in the $2\lfloor M/L \rfloor + 1$ coefficients $h_p[n]$ of $H_p(z)$. Utilizing the fact that the coefficients of $H_p(z)$ as well as the time-domain conditions are symmetric, a system of $\lfloor M/L \rfloor + 1$ equations needs to be solved. The remaining problem is to find $H_s(z)$ to give the minimum value of δ_s. The algorithm for iteratively determining the desired $H_s(z)$ consists of the following steps [SA87b]:

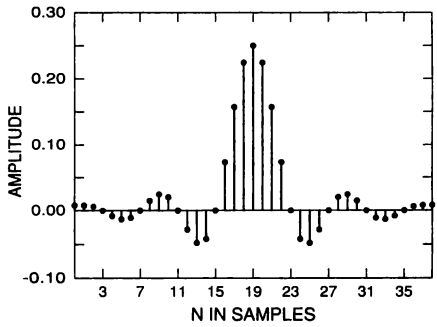
1. Set $H_p(\omega) \equiv 1$ and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{M-K+1}\} = \{0, 0, \ldots, 0\}$.
2. Find $H_s(\omega)$ such that $H_s(0) = 1$ and $W(\omega)H_p(\omega)H_s(\omega)$ alternatingly achieves at least at $M - K + 1$ consecutive points on $[\omega_s, \pi]$ the extremum values $\pm\delta_s$. Store the extremal points into $\overline{\Omega} = \{\overline{\omega}_1, \overline{\omega}_2, \ldots, \overline{\omega}_{M-K+1}\}$.
3. Determine $H_p(z)$ such that the time-domain conditions of Eq. (4.128) are satisfied.
4. If $|\omega_k - \overline{\omega}_k| \leq \alpha$ for $k = 1, 2, \ldots, M - K + 1$ (α is a small number), then stop. Otherwise set $\Omega = \overline{\Omega}$ and go to Step 2.

The desired $H_s(\omega)$ at Step 2 can be found using the MPR algorithm. The desired function is zero on $[\omega_s, \pi]$ and the weighting function is $W(\omega)H_p(\omega)$. $H_s(\omega)$ can be forced to take the value unity at $\omega = 0$ by selecting a very narrow passband region $[0, \epsilon]$, setting $D(\omega) \equiv 1$, and by using a large weighting function in this region. If a very narrow passband region is used, then the MPR algorithm selects automatically only one grid point ($\omega = 0$) in this region. Typically, three to five iterations of the above algorithm are needed to arrive at the desired solution. Another approach for designing Lth band filters is to use linear programming [LI85; SA88c]. However, linear programming requires significantly more computation time than the above algorithm.

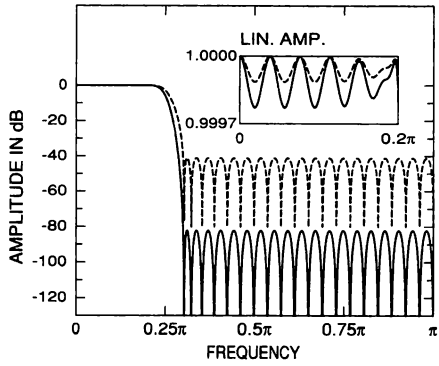
Example 4.12. The specifications are $L = 4$ and $\rho = 0.2$ ($\omega_p = 0.2\pi$ and $\omega_s = 0.3\pi$), and $\delta_s = 0.01$ (40-dB attenuation). The amplitude and impulse responses for an optimized filter of order 38 are shown in Figures 4-38(a) and 4-38(b), respectively.



(a)



(b)



(c)

FIGURE 4-38 Fourth-band filters with $\rho = 0.2$ ($\omega_p = 0.2\pi$ and $\omega_s = 0.3\pi$). (a, b) Amplitude and impulse responses for a minimax linear-phase design of order 38. (c) Amplitude responses for the overall factorizable minimax design (solid line) of order 102 and for the minimum-phase term (dashed line) of order 51.

If it is desired that $H(z)$ be factorizable into the minimum- and maximum-phase terms, then the subfilter $H_s(z)$ is rewritten in the form

$$H_s(z) = [\overline{H}_s(z)]^2, \qquad \overline{H}_s(z) = \sum_{n=0}^{M-K} \overline{h}_s[n]z^{-n}. \tag{4.133}$$

Here, $\overline{H}_s(z)$ is either a Type I linear-phase filter ($M - K$ is even) or a Type II filter ($M - K$ is odd). The resulting overall zero-phase frequency response is given by

$$H(\omega) = H_p(\omega)[\overline{H}_s(\omega)]^2. \tag{4.134}$$

Since the zeros of $H_p(z)$ are off the unit circle, $H(\omega)$ is nonnegative, as is desired. In this case, the minimization of the stopband ripple can be performed by slightly modifying the above algorithm. The basic difference is that now $\overline{H}_s(\omega)$ is determined at Step 2 such that $\overline{H}_s(0) = 1$ and $(\sqrt{W(\omega)H_p(\omega)})\overline{H}_s(\omega)$ oscillates within the limits $\pm\overline{\delta}_s$ on $[\omega_s, \pi]$ with $\lfloor (M - K)/2 \rfloor + 1$ extremal frequencies. Correspondingly, $W(\omega)H_p(\omega)[\overline{H}_s(\omega)]^2$ oscillates within the limits 0 and $\delta_s = (\overline{\delta}_s)^2$ on $[\omega_s, \pi]$. The advantage of this approach is that both the minimum- and maximum-phase terms of $H(z)$ contain $\overline{H}_s(z)$ and only $H_p(z)$ must be factored in order to get the overall minimum-phase and maximum-phase designs.

Example 4.13. The specifications for the minimum-phase and maximum-phase filters are those of Example 4.12. The required stopband ripple for $H(\omega)$ is $(\overline{\delta}_s)^2 = 0.0001$. Figure 4-38(c) gives the amplitude responses for an optimized overall filter of order 102 (solid line) and for the minimum-phase (or maximum-phase) term of order 51 (dashed line).

4-10-3 Design of Half-Band Filters

A very important subclass of Lth band filters in many applications are half-band filters ($L = 2$). For these filters,

$$h[M] = \tfrac{1}{2}, \tag{4.135a}$$

$$h[M + 2r] = 0 \quad \text{for } r = \pm 1, \pm 2, \ldots, \pm \lfloor M/2 \rfloor . \tag{4.135b}$$

A filter satisfying these conditions can be generated in two steps by starting with a Type II (M is odd) transfer function

$$G(z) = \sum_{n=0}^{M} g[n]z^{-n}, \qquad g[M - n] = g[n]. \tag{4.136}$$

In the first step, zero-valued impulse-response samples are inserted between the $g[n]$'s (see Figures 4-39(a) and 4-39(b)), giving the following Type I transfer func-

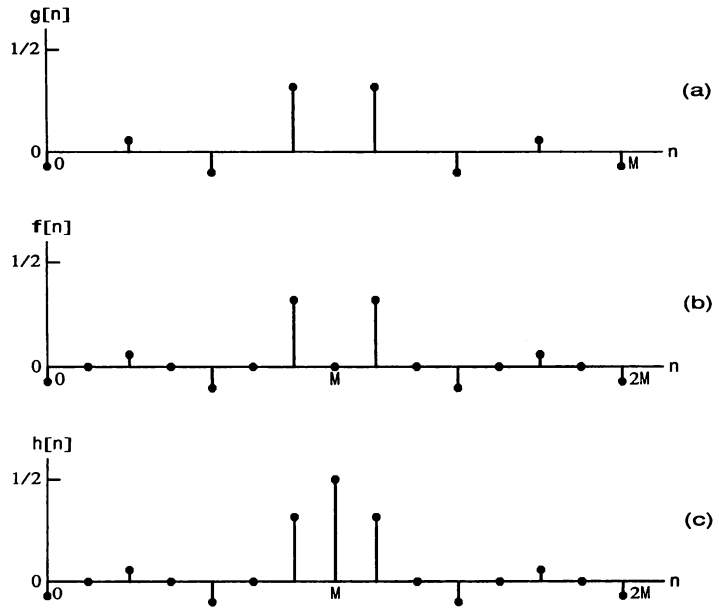


FIGURE 4-39 Generation of the impulse response of a half-band filter of order $2M$ with M odd. (a) Impulse response $g[n]$ of a Type II filter of order M. (b) Impulse response $f[n]$ resulting when inserting zero-valued samples between the $g[n]$'s. (c) Desired impulse response $h[n]$ of a half-band filter obtained by replacing $f[M] = 0$ by $h[M] = \frac{1}{2}$.

tion of order $2M$:

$$F(z) = \sum_{n=0}^{2M} f[n]z^{-n} = G(z^2) = \sum_{n=0}^{M} g[n]z^{-2n}. \tag{4.137}$$

The second step is then to replace the zero-valued impulse-response sample at $n = M$ by $\frac{1}{2}$ (see Figure 4-39(c)), resulting in the desired transfer function

$$H(z) = \sum_{n=0}^{2M} h[n]z^{-n} = \tfrac{1}{2}z^{-M} + F(z) = \tfrac{1}{2}z^{-M} + \sum_{n=0}^{M} g[n]z^{-2n}. \tag{4.138}$$

This gives $h[M] = \frac{1}{2}$, $h[n] = g[n/2]$ for n even, and $h[n] = 0$ for n odd and $n \neq M$, as is desired.

The zero-phase frequency responses of $H(z)$, $F(z)$, and $G(z)$ are related through

$$H(\omega) = \tfrac{1}{2} + F(\omega) = \tfrac{1}{2} + G(2\omega). \tag{4.139}$$

Based on these relations, the design of a lowpass half-band filter with passband edge at ω_p and passband ripple of δ can be accomplished by determining $G(z)$ such that $G(\omega)$ oscillates within $\frac{1}{2} \pm \delta$ on $[0, 2\omega_p]$ (see Figure 4-40(a)). Since $G(z)$ is a Type II transfer function, it has one fixed zero at $z = -1$ ($\omega = \pi$). $G(z)$ can be designed directly with the aid of the MPR algorithm using only one band $[0, 2\omega_p]$, $D(\omega) = \frac{1}{2}$, and $W(\omega) = 1$. Since $G(z)$ has a single zero at $z = -1$, $G(\omega)$ is odd about $\omega = \pi$. Hence $G(2\pi - \omega) = -G(\omega)$ and $G(\omega)$ oscillates within $-\frac{1}{2} \pm \delta$ on $[2\pi - 2\omega_p, 2\pi]$. The corresponding $F(\omega) = G(2\omega)$ stays within $\frac{1}{2} \pm \delta$ on $[0, \omega_p]$ and within $-\frac{1}{2} \pm \delta$ on $[\pi - \omega_p, \pi]$ (see Figure 4-40(b)). Finally, $H(\omega)$ approximates unity on $[0, \omega_p]$ with tolerance δ and zero on $[\pi - \omega_p, \pi]$ with the same tolerance δ (see Figure 4-40(c)).

For the resulting $H(\omega)$, the passband and stopband ripples are thus the same and the passband and stopband edges are related through $\omega_s = \pi - \omega_p$. In general,

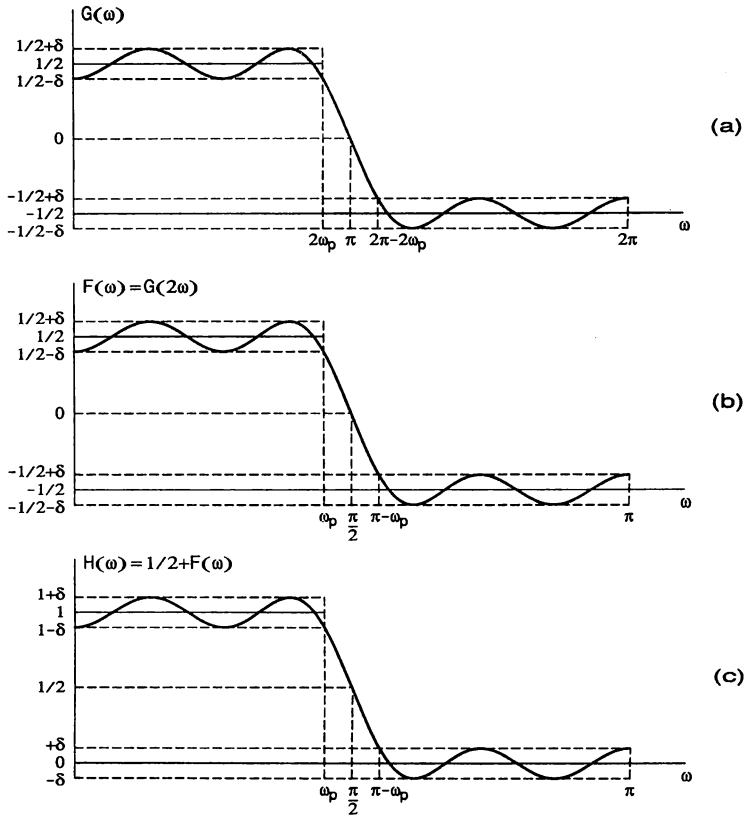


FIGURE 4-40 Design of a half-band lowpass filter.

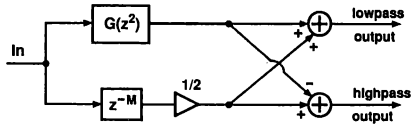


FIGURE 4-41 Implementation of a complementary lowpass–highpass half-band filter pair.

$H(\omega)$ satisfies

$$H(\omega) + H(\pi - \omega) = 1. \tag{4.140}$$

This makes $H(\omega)$ symmetric about the point $\omega = \pi/2$ such that the sum of the values of $H(\omega)$ at $\omega = \overline{\omega} < \pi/2$ and at $\omega = \pi - \overline{\omega} > \pi/2$ is equal to unity (see Figure 4-40(c)).

Figure 4-41 gives an implementation for the half-band filter as a parallel connection of $G(z^2)$ and $\frac{1}{2}z^{-M}$. This implementation is very attractive because in this case the complementary highpass output having the zero-phase frequency response $1 - H(\omega)$ is obtained directly by subtracting $G(z^2)$ from $\frac{1}{2}z^{-M}$. Note that the delay term z^{-M} can be shared with $G(z^2)$. The number of nonzero coefficient values in $G(z^2)$ is $M + 1$. By exploiting the symmetry in these coefficients, only $(M + 1)/2$ multipliers (M is odd) are needed to implement a lowpass–highpass filter pair of order $2M$. Figure 4-42 gives responses for a complementary half-band filter pair.

4-11 DESIGN OF FIR FILTERS USING PERIODIC SUBFILTERS AS BASIC BUILDING BLOCKS

One approach to reduce the cost of implementation of an FIR filter is to construct the overall filter using subfilters whose transfer function is of the form $F(z^L)$. There exist several design techniques [FA81; JI84; LI86; NE84, NE87; RA88; SA88a, SA88b, SA90], some of which are reviewed in this section.

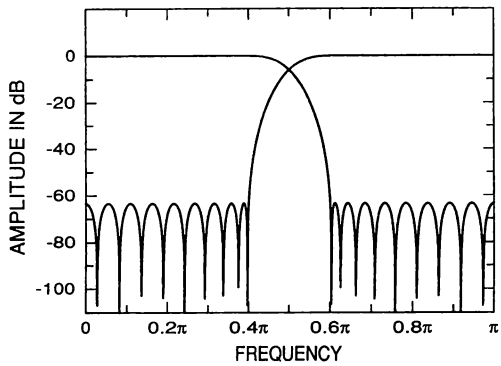


FIGURE 4-42 Responses for a complementary half-band filter pair of order 34 for $\omega_p = 0.4\pi$. The implementation of this filter pair requires only nine multipliers.

4-11-1 Periodic Filters

Consider a linear-phase transfer function of the form

$$A(z) = F(z^L) = \sum_{n=0}^{N_F} f[n]z^{-nL}, \qquad f[N_F - n] = f[n]. \tag{4.141}$$

This transfer function is obtainable from a conventional transfer function

$$F(z) = \sum_{n=0}^{N_F} f[n]z^{-n} \tag{4.142}$$

by replacing z^{-1} by z^{-L}, that is, by substituting for each unit delay L unit delays. Figure 4-43 gives for $A(z)$ an implementation that exploits the coefficient symmetry. Note that there is a multiplier only after every Lth delay term. The order of $A(z)$ is LN_F and its zero-phase frequency response is

$$A(\omega) = F(L\omega). \tag{4.143}$$

$A(\omega)$ is thus a frequency-axis compressed version of $F(\omega)$ such that the interval $[0, L\pi]$ is shrunk onto $[0, \pi]$. Figure 4-44 gives the resulting $A(\omega)$ in the case where $F(\omega)$ is a lowpass design with passband and stopband edges at θ and ϕ. Since the periodicity of $F(\omega)$ is 2π, the periodicity of $A(\omega)$ is $2\pi/L$ and it contains several passband and stopband regions in the interval $[0, \pi]$. It should be noted that this applies when N_F is even.[14]

[14]If N_F is odd, then $A(\omega)$ changes sign at $\omega = \pi/L, 3\pi/L, 5\pi/L, \ldots$ so that $A(\omega)$ approximates alternatingly unity and minus unity in the consecutive passband regions. This is because $F(z)$ is in this case a Type II design and $F(\omega)$ is odd about the points $\omega = \pi, 3\pi, 5\pi, \ldots$ (see Figure 4-6).

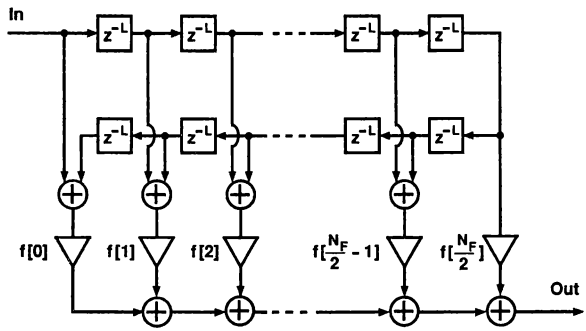


FIGURE 4-43 An implementation for a linear-phase transfer function $F(z^L)$ when N_F is even.

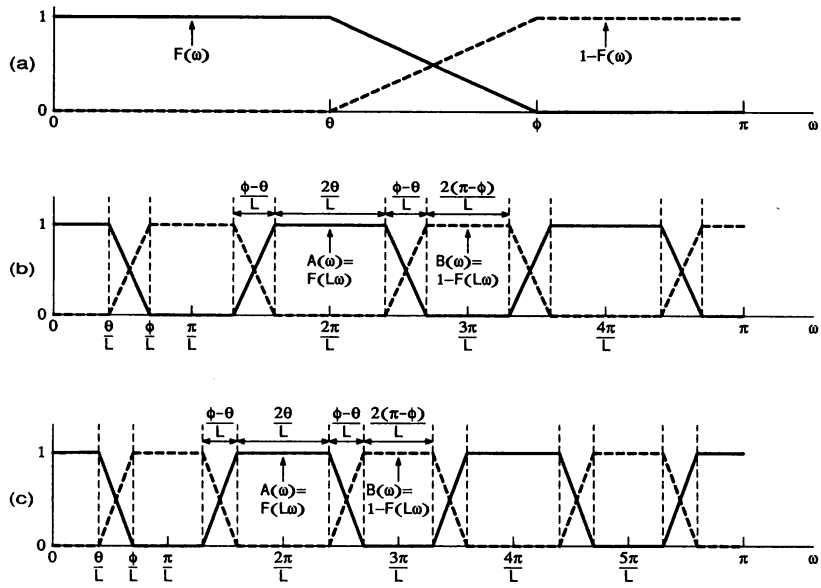


FIGURE 4-44 Periodic responses $A(\omega) = F(L\omega)$ and $B(\omega) = 1 - F(L\omega)$ when $F(\omega)$ is a lowpass design. (a) Responses $F(\omega)$ and $1 - F(\omega)$. (b) Responses $F(L\omega)$ and $1 - F(L\omega)$ for $L = 5$. (c) Responses $F(L\omega)$ and $1 - F(L\omega)$ for $L = 6$.

When N_F is even, the complementary transfer function of $A(z)$ is (cf. Section 4-8-3)

$$B(z) = z^{-LN_F/2} - A(z) = z^{-LN_F/2} - F(z^L). \tag{4.144}$$

The zero-phase frequency response of the corresponding filter is

$$B(\omega) = 1 - A(\omega) = 1 - F(L\omega) \tag{4.145}$$

and its passband regions are the stopband regions of $A(\omega)$ and vice versa (see Figure 4-44).

As seen from Figure 4-44, the periodic transfer functions $F(z^L)$ and $z^{-LN_F/2} - F(z^L)$ provide several transition bands of width $(\phi - \theta)/L$, which can be used as a transition band for a lowpass filter. The attractive property of these filters is that the number of nonzero impulse-response values to provide one of these transition bands is only $(1/L)$th of that of a conventional nonperiodic filter. This follows from the facts that the required FIR filter order is roughly inversely proportional to the transition bandwidth (cf. Section 4-8-2) and the transition bandwidth of the prototype filter $F(z)$, which determines the number of nonzero impulse-response coefficients, is $\phi - \theta$. This is L times wider. Note that the orders of the periodic

filters and that of the nonperiodic filter are approximately the same, but the nonperiodic filter has no zero-valued impulse-response samples.

Because of periodic responses, $F(z^L)$ or $z^{-LN_F/2} - F(z^L)$ cannot be used alone for synthesizing a lowpass filter. The desired result can be achieved by properly combining these filters with conventional nonperiodic filters.

4-11-2 Frequency-Response Masking Approach

A very elegant approach to exploiting the attractive properties of $A(z)$ and $B(z)$ has been proposed by Lim [LI86]. In this approach, the overall transfer function is constructed as

$$H(z) = F(z^L)G_1(z) + [z^{-LN_F/2} - F(z^L)]G_2(z). \tag{4.146}$$

An implementation[15] of this transfer function is shown in Figure 4-45. The zero-phase frequency response of this filter can be written as

$$H(\omega) = F(L\omega)G_1(\omega) + [1 - F(L\omega)]G_2(\omega) \tag{4.147}$$

provided that the delays of $G_1(z)$ and $G_2(z)$ are equal.[16]

For a lowpass design, the transition band of $H(\omega)$ can be selected to be one of the transition bands provided by $F(z^L)$ or $z^{-LN_F/2} - F(z^L)$. In the first case, referred to as Case A, the edges of $H(\omega)$ are selected to be (see Figure 4-46)

$$\omega_p = (2l\pi + \theta)/L, \qquad \omega_s = (2l\pi + \phi)/L, \tag{4.148}$$

where l is a fixed integer, and in the second case, referred to as Case B,

$$\omega_p = (2l\pi - \phi)/L, \qquad \omega_s = (2l\pi - \theta)/L. \tag{4.149}$$

[15]Note that the delay term $z^{-LN_F/2}$ can be shared with $F(z^L)$. Also, $G_1(z)$ and $G_2(z)$ can share the same delay elements if they are implemented using the transposed direct-form structure (exploiting the coefficient symmetry).

[16]This means that the orders of both $G_1(z)$ and $G_2(z)$, denoted by N_{G1} and N_{G2}, must be either even or odd and if N_{G1} and N_{G2} are not equal, then, in order to equalize the delays, the delay term $z^{-(N_{G1}-N_{G2})/2}$ ($z^{-(N_{G2}-N_{G1})/2}$) must be added to $G_2(z)$ ($G_1(z)$) if $N_{G1} - N_{G2}$ is positive (negative).

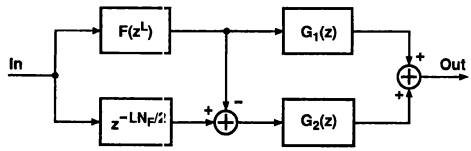


FIGURE 4-45 The structure of a filter synthesized using the frequency-response masking technique.

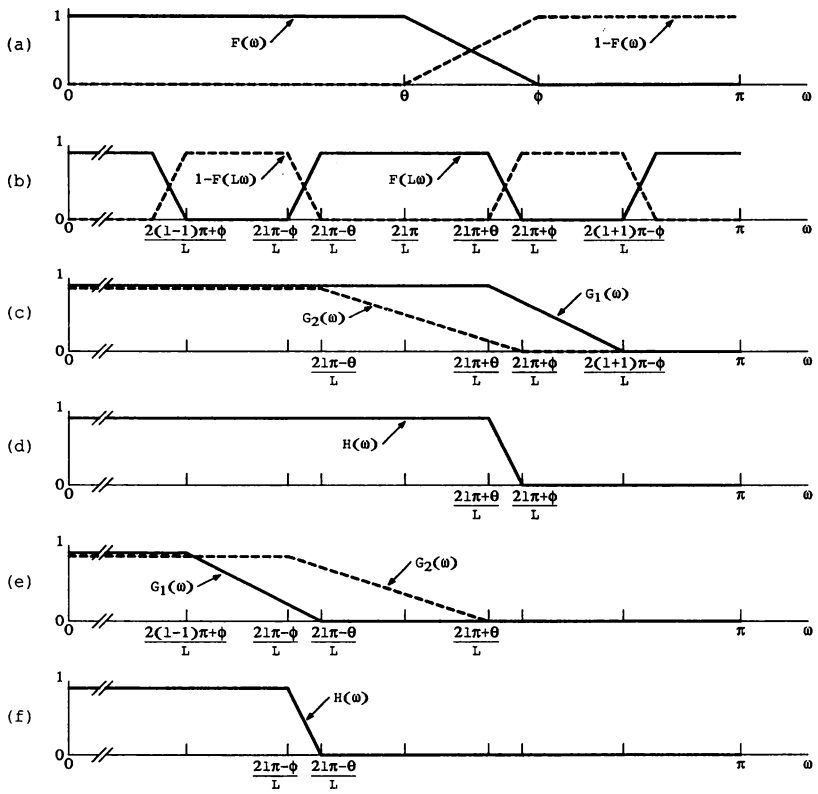


FIGURE 4-46 Design of lowpass filters using the frequency-response masking technique.

For Case A designs, $H(\omega)$ approximates unity on $[0, \omega_p]$ and zero on $[\omega_s, \pi]$ if the two masking filters $G_1(z)$ and $G_2(z)$ are lowpass filters with the following passband and stopband edges (see Figures 4-46(c) and 4-46(d)):

$$\omega_p^{(G_1)} = \omega_p = \{2l\pi + \theta\}/L, \qquad \omega_s^{(G_1)} = \{2(l+1)\pi - \phi\}/L, \tag{4.150a}$$

$$\omega_p^{(G_2)} = \{2l\pi - \theta\}/L, \qquad \omega_s^{(G_2)} = \omega_s = \{2l\pi + \phi\}/L. \tag{4.150b}$$

On $[0, \omega_p^{(G_2)}]$, $G_1(\omega) \approx 1$ and $G_2(\omega) \approx 1$ so that $H(\omega) \approx F(L\omega) + [1 - F(L\omega)] = 1$, as is desired. On $[\omega_p^{(G_2)}, \omega_p]$, $G_1(\omega) \approx 1$, $F(L\omega) \approx 1$, and $1 - F(L\omega) \approx 0$, giving $H(\omega) \approx F(L\omega)G_1(\omega) \approx 1$ regardless of the behavior of $G_2(\omega)$ in this region. On $[\omega_s^{(G_1)}, \pi]$, $G_1(\omega) \approx 0$ and $G_2(\omega) \approx 0$ so that $H(\omega) \approx 0$. Since $F(L\omega) \approx 0$ on $[\omega_s, \omega_s^{(G_1)}]$, the stopband region of $G_1(z)$ can start at $\omega = \omega_s^{(G_1)}$, instead of $\omega = \omega_s$.

For Case B designs, the required edges of the two masking filters $G_1(z)$ and

$G_2(z)$ are (see Figures 4-46(e) and 4-46(f))

$$\omega_p^{(G_1)} = \{2(l-1)\pi + \phi\}/L, \qquad \omega_s^{(G_1)} = \omega_s = \{2l\pi - \theta\}/L, \tag{4.151a}$$

$$\omega_p^{(G_2)} = \omega_p = \{2l\pi - \phi\}/L, \qquad \omega_s^{(G_2)} = \{2l\pi + \theta\}/L. \tag{4.151b}$$

In Lim [LI86], the effects of the ripples of the subresponses $G_1(\omega)$, $G_2(\omega)$, and $F(L\omega)$ on the ripples of the overall response $H(\omega)$ have been studied carefully. Based on these observations, the design of the overall filter with passband and stopband ripples of δ_p and δ_s can be accomplished for both Case A and Case B in the following two steps:

1. Design $G_k(z)$ for $k = 1, 2$ using the MPR algorithm such that $G_k(\omega)$ approximates unity on $[0, \omega_p^{(G_k)}]$ with tolerance $0.85\delta_p \ldots 0.9\delta_p$ and zero on $[\omega_s^{(G_k)}, \pi]$ with tolerance $0.85\delta_s \ldots 0.9\delta_s$.[17]
2. Design $F(L\omega)$ such that the overall response $H(\omega)$ as given by Eq. (4.147) approximates unity on

$$\Omega_p^{(F)} = \begin{cases} [\omega_p^{(G_2)}, \omega_p] = [\{2l\pi - \theta\}/L, \{2l\pi + \theta\}/L] \\ \qquad \text{for Case A} \\ [\omega_p^{(G_1)}, \omega_p] = [\{2(l-1)\pi + \phi\}/L, \{2l\pi - \phi\}/L] \\ \qquad \text{for Case B} \end{cases} \tag{4.152a}$$

with tolerance δ_p and zero on

$$\Omega_s^{(F)} = \begin{cases} [\omega_s, \omega_s^{(G_1)}] = [\{2l\pi + \phi\}/L, \{2(l+1)\pi - \phi\}/L] \\ \qquad \text{for Case A} \\ [\omega_s, \omega_s^{(G_2)}] = [\{2l\pi - \theta\}/L, \{2l\pi + \theta\}/L] \\ \qquad \text{for Case B} \end{cases} \tag{4.152b}$$

with tolerance δ_s.

The design of $F(L\omega)$ can be performed using linear programming [LI86]. Another, computationally more efficient, alternative is to use the MPR algorithm. It can be shown that $F(L\omega)$ meets the given criteria if the maximum absolute value of the error function given in Table 4-6 is on $[0, \theta] \cup [\phi, \pi]$ less than or equal to unity. Even though this error function looks very complicated, it is straightforward to use the subroutines EFF and WATE in the MPR algorithm for optimally designing $F(z)$.

[17]To reduce the order of $G_1(z)$, a smaller weighting can be used in the MPR algorithm on those regions of $G_1(z)$ where $F(L\omega)$ has one of its stopbands. As a rule of thumb, for the regions in the passband (stopband) of $G_1(z)$, the weighting can be selected to be one-tenth of the original passband (stopband) weighting. Similarly, the order of $G_2(z)$ can be reduced by using a smaller weighting on those regions where $F(L\omega)$ has one of its passbands.

TABLE 4-6 Error Function for Designing $F(z)$ in the Frequency-Response Masking Approach

$$E(\omega) = W_F(\omega)[F(\omega) - D_F(\omega)]$$

where

$$D_F(\omega) = [u(\omega) + l(\omega)]/2, \qquad W_F(\omega) = 2/[u(\omega) - l(\omega)]$$

with

$$u(\omega) = \min\left(\frac{D(\omega) - G_2[h_1(\omega)]}{G_1[h_1(\omega)] - G_2[h_1(\omega)]} + e_1(\omega), \frac{D(\omega) - G_2[h_2(\omega)]}{G_1[h_2(\omega)] - G_2[h_2(\omega)]} + e_2(\omega)\right)$$

$$l(\omega) = \max\left(\frac{D(\omega) - G_2[h_1(\omega)]}{G_1[h_1(\omega)] - G_2[h_1(\omega)]} - e_1(\omega), \frac{D(\omega) - G_2[h_2(\omega)]}{G_1[h_2(\omega)] - G_2[h_2(\omega)]} - e_2(\omega)\right)$$

$$e_1(\omega) = \frac{\delta(\omega)}{|G_1[h_1(\omega)] - G_2[h_1(\omega)]|}, \qquad e_2(\omega) = \frac{\delta(\omega)}{|G_1[h_2(\omega)] - G_2[h_2(\omega)]|}$$

and

$$D(\omega) = \begin{cases} 1 & \text{for } \omega \in [0, \theta] \\ 0 & \text{for } \omega \in [\phi, \pi], \end{cases} \qquad \delta(\omega) = \begin{cases} \delta_p & \text{for } \omega \in [0, \theta] \\ \delta_s & \text{for } \omega \in [\phi, \pi] \end{cases}$$

$$h_1(\omega) = (2l\pi + \omega)/L, \qquad h_2(\omega) = \begin{cases} (2l\pi - \omega)/L & \text{for } \omega \in [0, \theta] \\ (2(l+1)\pi - \omega)/L & \text{for } \omega \in [\phi, \pi] \end{cases}$$

for Case A and

$$D(\omega) = \begin{cases} 0 & \text{for } \omega \in [0, \theta] \\ 1 & \text{for } \omega \in [\phi, \pi], \end{cases} \qquad \delta(\omega) = \begin{cases} \delta_s & \text{for } \omega \in [0, \theta] \\ \delta_p & \text{for } \omega \in [\phi, \pi] \end{cases}$$

$$h_1(\omega) = (2l\pi - \omega)/L, \qquad h_2(\omega) = \begin{cases} (2l\pi + \omega)/L & \text{for } \omega \in [0, \theta] \\ (2(l-1)\pi + \omega)/L & \text{for } \omega \in [\phi, \pi] \end{cases}$$

for Case B

In practice, ω_p and ω_s are given and l, L, θ, and ϕ must be determined. To ensure that Eq. (4.148) yields a desired solution with $0 \le \theta < \phi \le \pi$, it is required that (see Figures 4-44 and 4-46)

$$\frac{2l\pi}{L} \le \omega_p, \qquad \omega_s \le \frac{(2l+1)\pi}{L} \tag{4.153}$$

for some positive integer l. In this case,

$$l = \lfloor L\omega_p/(2\pi) \rfloor, \qquad \theta = L\omega_p - 2l\pi, \qquad \phi = L\omega_s - 2l\pi. \tag{4.154}$$

Similarly, to ensure that Eq. (4.149) yields a desired solution with $0 \le \theta < \phi \le \pi$, it is required that

$$\frac{(2l-1)\pi}{L} \le \omega_p, \qquad \omega_s \le \frac{2l\pi}{L} \tag{4.155}$$

for some positive integer l. In this case,[18]

$$l = \lceil L\omega_s/(2\pi) \rceil, \qquad \theta = 2l\pi - L\omega_s, \qquad \phi = 2l\pi - L\omega_p. \tag{4.156}$$

For any set of ω_p, ω_s, and L, either Eq. (4.154) or (4.156) (not both) will yield the desired θ and ϕ, provided that L is not too large. If $\theta = 0$ or $\phi = \pi$, then the resulting specifications for $F(\omega)$ are meaningless and the corresponding value of L cannot be used.

The remaining problem is to determine L to minimize the filter complexity. The following example illustrates how this problem can be solved.

Example 4.14. The filter specifications are [JI84; SA88b] $\omega_p = 0.4\pi$, $\omega_s = 0.402\pi$, $\delta_p = 0.01$, and $\delta_s = 0.0001$. The minimum order of an optimum conventional direct-form design to meet the given criteria, estimated by Eq. (4.95), is 3138, requiring 1570 multipliers. Table 4-7 gives, for the admissible values of L in the range $3 \le L \le 30$, l, θ, and ϕ as well as whether the overall filter is a Case A or Case B design. It shows also estimated orders for $F(z)$, $G_1(z)$, and $G_2(z)$, denoted by N_F, N_{G1}, and N_{G2}, as well as the overall number of multipliers,[19] $N_F/2 + 1 + \lfloor (N_{G1} + 2)/2 \rfloor + \lfloor (N_{G2} + 2)/2 \rfloor$. These orders have been estimated using Eq. (4.95) with the passband and stopband ripples being the specified ones.[20] N_F is approximately only $(1/L)$th of the order of an equivalent direct-form design and decreases with increasing L. The widths of the transition bands of $G_1(\omega)$ and $G_2(\omega)$ are $[2\pi - \theta - \phi]/L$ and $[\theta + \phi]/L$, respectively, so that their sum is $2\pi/L$ and decreases with increasing L. The overall number of multipliers is usually the smallest at those values of L for which these widths and, correspondingly, N_{G1} and N_{G2} are of the same order. This happens when $\theta + \phi$ is approximately equal to π. If this is true for several values of L, then the best result is obtained by increasing L until the decrease in the number of multipliers in $F(z)$ is less than the increase in the overall number of multipliers in $G_1(z)$ and $G_2(z)$. The estimated filter orders are usually so close to the actual minimum orders

[18] $\lceil x \rceil$ stands for the smallest integer which is larger than or equal to x.

[19] When the symmetry in the filter coefficients is exploited, a Type I linear-phase filter of order N (even) requires $N/2 + 1$ multipliers, whereas a Type II filter (N odd) requires $(N + 1)/2$ multipliers. N_F is forced, according to the discussion of Section 4-11-1, to be even in order to get the desired solution.

[20] For N_F, the nearest even value greater than or equal to the estimated value has been used. For Case A designs, the ripples in the passband $[0, \theta]$ and in the stopband $[\phi, \pi]$ are δ_p and δ_s, respectively, whereas for Case B designs they are interchanged. Note that for the estimation formula of Eq. (4.95), δ_p and δ_s are the larger and smaller ripple values, respectively. If the estimated value of N_{G1} is even (odd), then N_{G2} has also been selected to be even (odd).

TABLE 4-7 Estimation of L Minimizing the Number of Multipliers in the Frequency-Response Masking Approach

L	Case	l	θ	ϕ	N_F	N_{G1}	N_{G2}	Number of Multipliers
3	B	1	0.794π	0.8π	1046	46	10	554
4	B	1	0.392π	0.4π	786	19	31	420
6	A	1	0.4π	0.412π	524	31	47	303
7	A	1	0.8π	0.814π	448	113	27	296
8	B	2	0.784π	0.8π	392	120	32	275
9	B	2	0.382π	0.4π	350	46	72	237
11	A	2	0.4π	0.422π	286	58	84	217
12	A	2	0.8π	0.824π	262	200	46	257
13	B	3	0.774π	0.8π	242	191	51	244
14	B	3	0.372π	0.4π	224	71	113	206
16	A	3	0.4π	0.432π	196	86	120	204
17	A	3	0.8π	0.834π	184	291	65	272
18	B	4	0.764π	0.8π	174	259	73	255
19	B	4	0.362π	0.4π	166	96	156	212
21	A	4	0.4π	0.442π	150	113	157	212
22	A	4	0.8π	0.844π	142	388	84	310
23	B	5	0.754π	0.8π	136	324	92	279
24	B	5	0.352π	0.4π	130	120	200	228
26	A	5	0.4π	0.452π	120	142	192	230
27	A	5	0.8π	0.854π	116	490	102	357
28	B	6	0.744π	0.8π	112	385	113	307
29	B	6	0.342π	0.4π	108	144	246	252

that they can be used for determining the value of L minimizing the filter complexity.

With the estimated filter orders, $L = 16$ gives the best result. The actual filter orders are $N_F = 198$, $N_{G1} = 83$, and $N_{G2} = 123$. The responses of the subfilters as well as that of the overall design are given in Figure 4-47. The overall number of multipliers and adders for this design are 204 and 406, respectively, which are 13% of those required by an equivalent conventional direct-form design (1570 and 3138). The overall filter order is 3291, which is only 5% higher than that of the direct-form design (3138).

The complexity of the filter can be reduced further by using a two-level frequency-response masking. For examples, see Lim [LI86] and Saramäki and Fam [SA88b].

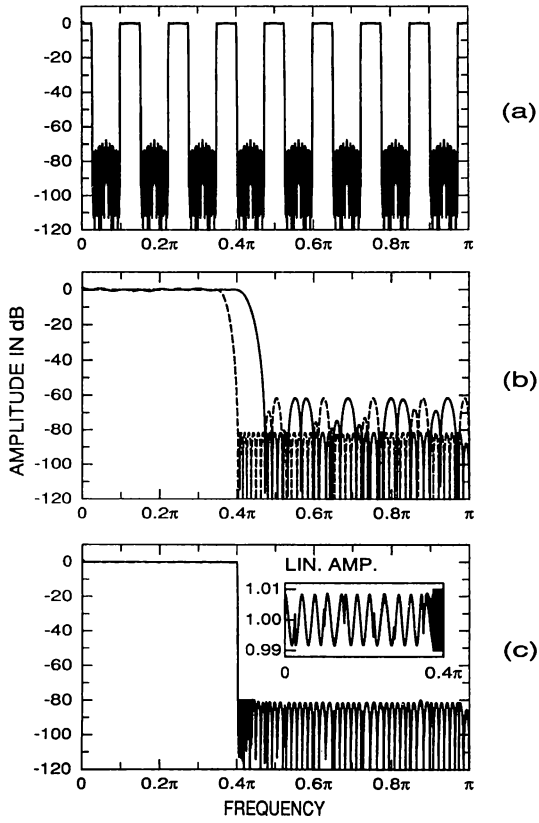


FIGURE 4-47 Amplitude responses for a lowpass filter synthesized using the frequency-response masking approach. (a) Periodic response $F(L\omega)$. (b) Subresponses $G_1(\omega)$ (solid line and $G_2(\omega)$ (dashed line) (c) Overall filter.

4-11-3 Design of Narrowband Lowpass Filters

When ω_s is less than $\pi/2$, then the overall filter can be synthesized in the following simplified form [JI84; NE84; SA88a]:

$$H(z) = F(z^L)G(z). \tag{4.157}$$

The zero-phase frequency response of this filter is given by

$$H(\omega) = F(L\omega)G(\omega). \tag{4.158}$$

If the passband and stopband edges of $F(z)$ are θ and ϕ, then the edges of the first transition band of $F(L\omega)$ are (see Figure 4-48)

$$\omega_p = \theta/L, \qquad \omega_s = \phi/L. \tag{4.159}$$

$F(L\omega)$ does not provide the desired attenuation in the regions where it has extra unwanted passbands and transition bands, that is, in the region

$$\Omega_s(L, \omega_s) = \bigcup_{k=1}^{\lfloor L/2 \rfloor} \left[k\frac{2\pi}{L} - \omega_s, \min\left(k\frac{2\pi}{L} + \omega_s, \pi \right) \right]. \tag{4.160}$$

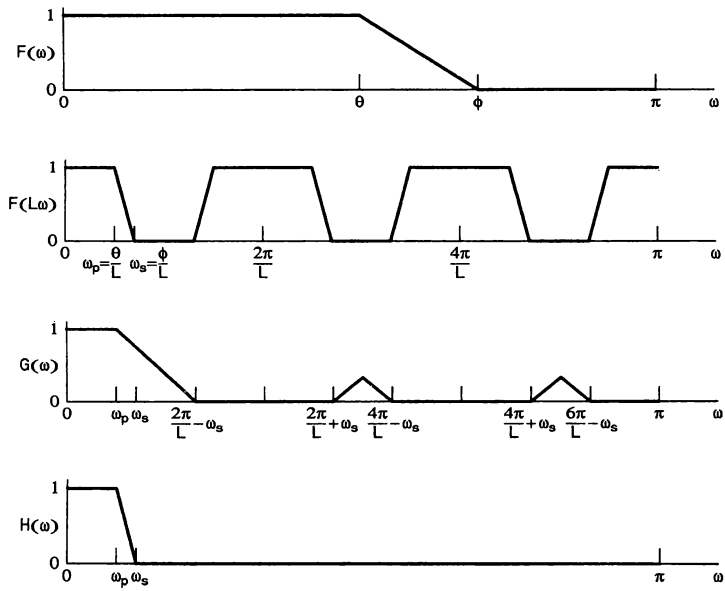


FIGURE 4-48 Design of a narrowband lowpass filter using a cascade of a periodic and a nonperiodic filter.

Therefore the role of the nonperiodic filter $G(z)$ is to provide enough attenuation in this region.

The simplest way to determine $F(z)$ and $G(z)$ such that $H(z)$ is a lowpass design with edges at ω_p and ω_s and ripples of δ_p and δ_s is to design these subfilters using the MPR algorithm to satisfy

$$1 - \delta_p^{(F)} \le F(\omega) \le 1 + \delta_p^{(F)} \quad \text{for } \omega \in [0, L\omega_p], \tag{4.161a}$$

$$-\delta_s \le F(\omega) \le \delta_s \quad \text{for } \omega \in [L\omega_s, \pi], \tag{4.161b}$$

$$1 - \delta_p^{(G)} \le G(\omega) \le 1 + \delta_p^{(G)} \quad \text{for } \omega \in [0, \omega_p], \tag{4.161c}$$

$$-\delta_s \le G(\omega) \le \delta_s \quad \text{for } \omega \in \Omega_s(L, \omega_s), \tag{4.161d}$$

where

$$\delta_p^{(G)} + \delta_p^{(F)} = \delta_p. \tag{4.161e}$$

The ripples $\delta_p^{(F)}$ and $\delta_p^{(G)}$ can be selected, for example, to be half the overall ripple δ_p. In the above specifications, both $F(z^L)$ and $G(z)$ have $[0, \omega_p]$ as a passband region.

Another alternative, resulting in a considerably reduced order of $G(z)$, is to design simultaneously $F(\omega)$ to meet

$$1 - \delta_p \le F(\omega)G(\omega/L) \le 1 + \delta_p \quad \text{for } \omega \in [0, L\omega_p], \tag{4.162a}$$

$$-\delta_s \le F(\omega)G(\omega/L) \le \delta_s \quad \text{for } \omega \in [L\omega_s, \pi], \tag{4.162b}$$

and $G(\omega)$ to meet

$$G(0) = 1, \tag{4.163c}$$

$$-\delta_s \le F(L\omega)G(\omega) \le \delta_s \quad \text{for } \omega \in \Omega_s(L, \omega_s). \tag{4.163d}$$

In this case, $G(z)$ has all its zeros on the unit circle and concentrates on providing for the overall filter the desired attenuation on $\Omega_s(L, \omega_s)$ (see Figure 4-49). $F(L\omega)$ equalizes the passband distortion caused by $G(\omega)$ and provides the required attenuation in its stopband regions.

The desired overall response can be found by designing iteratively $F(z)$ to meet Eq. (4.162) and $G(z)$ to meet Eq. (4.163) until the difference between successive overall solutions is within the given tolerance limits. The algorithm consists of the following steps [SA88a]:

1. Set $F(\omega) \equiv 1$.
2. Determine $G(\omega)$ using the MPR algorithm to minimize on $[0, \epsilon] \cup \Omega_s(L, \omega_s)$ the peak absolute value of $E_G(\omega) = W_G(\omega)[G(\omega) - D_G(\omega)]$, where

$$D_G(\omega) = \begin{cases} 1 & \text{for } \omega \in [0, \epsilon] \\ 0 & \text{for } \omega \in \Omega_s(L, \omega_s), \end{cases} \qquad W_G(\omega) = \begin{cases} \alpha & \text{for } \omega \in [0, \epsilon] \\ F(L\omega) & \text{for } \omega \in \Omega_s(L, \omega_s). \end{cases} \tag{4.164}$$

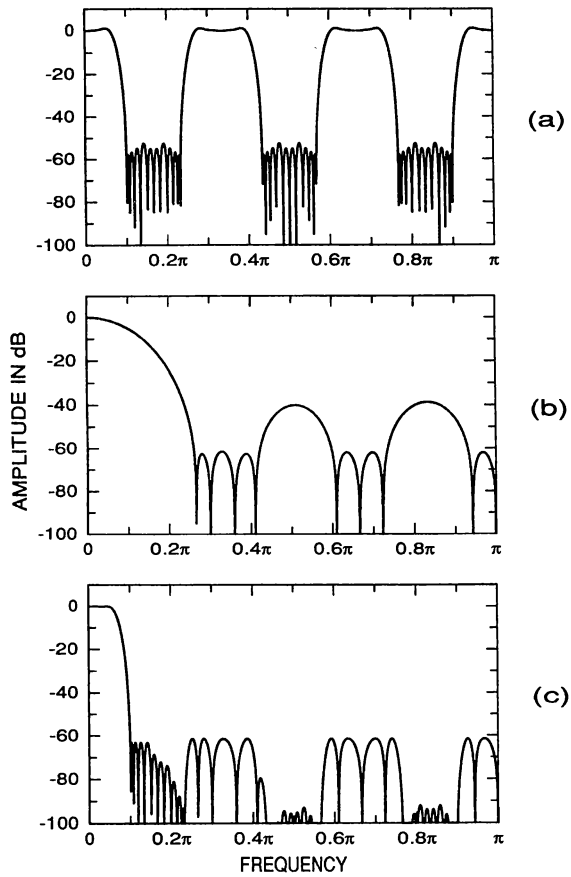


FIGURE 4-49 Amplitude responses for an optimized filter of the form $H(z) = F(z^L)G(z)$ with $L = 6$. The specifications are $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, $\delta_p = 0.01$, and $\delta_s = 0.001$. (a) $F(z^L)$ of order 17 in z^L. (b) $G(z)$ of order 17. (c) Overall filter.

By selecting ϵ to be a very small number and α to be a very large number, the MPR algorithm uses only one grid point ($\omega = 0$) on $[0, \epsilon]$ and forces $G(\omega)$ to take the value of unity at $\omega = 0$.

3. Determine $F(\omega)$ using the MPR algorithm to minimize on $[0, L\omega_p] \cup [L\omega_s, \pi]$ the peak absolute value of $E_F(\omega) = W_F(\omega)[F(\omega) - D_F(\omega)]$, where

$$D_F(\omega) = \begin{cases} 1/G(\omega/L) & \text{for } \omega \in [0, L\omega_p] \\ 0 & \text{for } \omega \in [L\omega_s, \pi], \end{cases} \tag{4.165a}$$

$$W_F(\omega) = \begin{cases} G(\omega/L) & \text{for } \omega \in [0, L\omega_p] \\ \delta_p G(\omega/L)/\delta_s & \text{for } \omega \in [L\omega_s, \pi]. \end{cases} \tag{4.165b}$$

4. Repeat Steps 2 and 3 until the difference between successive solutions is within the given tolerance limits.

Typically, three to five iterations of the above algorithm are required to arrive at the desired solution.

Given the filter specifications, the remaining problem is to optimize L and the orders of $G(z)$ and $F(z^L)$ to minimize the overall number of multipliers. For the order of $F(z^L)$ in z^L, a good estimate is

$$N_F = N/L, \tag{4.166}$$

where N is the minimum order of a conventional nonperiodic FIR filter to meet the given overall criteria. For the order of $G(z)$, a good estimate has been found to be

$$N_G = \cosh^{-1}\left(\frac{1}{\delta_s}\right)\left[\frac{1}{\cosh^{-1} X\left(\omega_p, \frac{2\pi}{L} - \frac{\omega_p + 2\omega_s}{3}\right)} + \frac{L/2}{\cosh^{-1} X\left(\frac{L\omega_p}{2}, \pi - \frac{L(\omega_p + 2\omega_s)}{6}\right)}\right], \tag{4.167a}$$

where

$$X(\omega_1, \omega_2) = (2\cos\omega_1 - \cos\omega_2 + 1)/(1 + \cos\omega_2). \tag{4.167b}$$

L has to be selected such that the stopband edge of $F(z)$, $\phi = L\omega_s$, is less than π. This means that L must be less than π/ω_s. After estimating the required orders for $G(z)$ and $F(z)$, the remaining problem is to decrease or increase the orders to find the actual minimum orders. Since the frequency-response-shaping responsibilities are very well shared with the subfilters, the minimum orders can be found rather independently. First, the minimum order of $F(z)$ can be determined for the estimated order of $G(z)$, and then the minimum order of $G(z)$ is determined. Again, an example is used to illustrate how the best value of L can be found.

Example 4.15. The specifications are $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, $\delta_p = 0.01$, and $\delta_s = 0.001$. Table 4-8 gives for the admissible values of L ($2 \le L \le 9$) both the estimated and actual minimum orders of $F(z)$ and $G(z)$ as well as the edges of $F(z)$. As seen from the table, the estimated orders are very close to the actual ones, showing that the best value of L can easily be determined based on the above

TABLE 4-8 Estimated and Actual Minimum Filter Orders for $2 \le L \le 9$ for a Narrowband Filter Synthesized as a Cascade of a Periodic and Nonperiodic Filter

			Estimated		Actual		Number of
L	θ	ϕ	N_F	N_G	N_F	N_G	Multipliers
2	0.1π	0.2π	54	3	54	3	30
3	0.15π	0.3π	36	6	35	6	22
4	0.2π	0.4π	27	9	26	9	19
5	0.25π	0.5π	22	13	21	14	19
6	0.3π	0.6π	18	17	17	17	18
7	0.35π	0.7π	15	22	14	22	20
8	0.4π	0.8π	13	27	12	28	22
9	0.45π	0.9π	12	34	10	36	25

estimation formulas. The minimum of the total number of multipliers is obtained by increasing L until the decrease in the number of multipliers of $F(z^L)$ becomes smaller than the increase in the number of multipliers of $G(z)$. The amplitude responses for the subfilters and the overall design are shown in Figure 4-49 for the best value of L, $L = 6$. The orders of both $F(z^L)$ (in z^L) and $G(z)$ are 17. This design requires 18 multipliers and 34 adders. The minimum order of a conventional direct-form design is 108, requiring 55 multipliers and 108 adders. The price paid for these reductions in the number of arithmetic operations is a slight increase in the overall filter order (from 108 to 119).

Example 4.16. Further savings in the number of arithmetic operations can be achieved by implementing $G(z)$ using special structures [KI88; SA88a, SA89b]. A particularly efficient implementation is provided by a transfer function of the form [SA89b]

$$G(z) = \prod_{r=1}^{M} T_r(z), \tag{4.168a}$$

where

$$T_r(z) = 2^{-Pr} \sum_{k=0}^{K_r-1} z^{-k} = 2^{-Pr} \frac{1 - z^{-Kr}}{1 - z^{-1}} \tag{4.168b}$$

and 2^{-Pr}, with P_r integer-valued, is a scaling constant. An efficient implementation of $T_r(z)$ is depicted in Figure 4-50.[21] The implementation of the above $G(z)$ re-

[21] If modulo arithmetic (e.g., 1's or 2's complement arithmetic) and the worst-case scaling (corresponds to peak scaling in this case) are used, the output of $T_r(z)$ implemented as shown in Figure 4-50 is correct even though internal overflows may occur. For details, see Saramäki et al. [SA88a]. This implementation is very attractive because, in this case, the system does not need initial resetting and the effect of temporary miscalculations vanishes automatically from the output in a finite time.

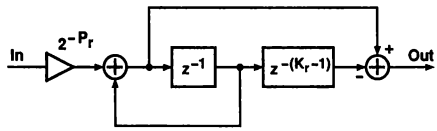


FIGURE 4-50 Efficient implementation for $T_r(z) = 2^{-Pr}(1 - z^{-Kr})/(1 - z^{-1})$.

quires no general multipliers. The zero-phase frequency response of $T_r(z)$ is

$$T_r(\omega) = 2^{-Pr} \sin(K_r\omega/2)/\sin(\omega/2) \tag{4.169}$$

and it provides $K_r - 1$ zeros on the unit circle, located at $\omega = \pm k2\pi/K_r$ for $k = 1, 2, \ldots, \lfloor (K_r - 1)/2 \rfloor$ and at $\omega = \pi$ for K_r even. The design of $G(z)$ involves determining M, the number of $T_r(z)$'s, and the K_r's in such a way that the resulting $G(\omega)$ provides enough attenuation on $\Omega_s(L, \omega_s)$ as given by Eq. (4.160). The criteria of the previous example are met by $L = 8$, $F(z^L)$ of order 9 in z^L, and $G(z)$ consisting of four $T_r(z)$'s with $K_1 = 18$, $K_2 = 16$, $K_3 = 12$, and $K_4 = 11$. This filter requires only 5 general multipliers, 17 adders, and 129 delay elements. The amplitude responses of the subfilters $G(z)$ and $F(z^L)$ are given in Figure 4-51(a), whereas Figure 4-51(b) gives the overall response.[22]

4-11-4 Design of Wideband Lowpass Filters

The design of a wideband filter can be accomplished based on the fact (see Section 4-8-3) that if $\overline{H}(z)$ of even order $2M$ is a lowpass design with the following edges and ripples,

$$\overline{\omega}_p = \pi - \omega_s, \quad \overline{\omega}_s = \pi - \omega_p, \quad \overline{\delta}_p = \delta_s, \quad \overline{\delta}_s = \delta_p, \tag{4.170}$$

then

$$H(z) = z^{-M} - (-1)^M \overline{H}(-z) \tag{4.171}$$

is a lowpass filter having the passband and stopband edges of ω_p and ω_s and the passband and stopband ripples of δ_p and δ_s. Hence, if ω_p and ω_s of the desired filter are larger than $\pi/2$, then $\overline{\omega}_p$ and $\overline{\omega}_s$ of $\overline{H}(z)$ are less than $\pi/2$. This enables us to design $\overline{H}(z)$ in the form

$$\overline{H}(z) = F(z^L)G(z) \tag{4.172}$$

[22]The K_r's are larger than necessary to meet the criteria for $G(z)$ and they provide some zeros on $[\omega_s, 2\pi/L - \omega_s] = [0.1\pi, 0.15\pi]$, making the requirements of $F(z^L)$ less stringent in this region (see Figure 4-51). It can be shown that in order to guarantee that the overall filter meets the given criteria in this case, $F(z)$ has to be designed such that $G(\omega/L)$ is replaced by max $[|G(\omega/L)|, |G((2\pi - \omega)/L)|]$ in Eq. (4.162b) [SA89b].

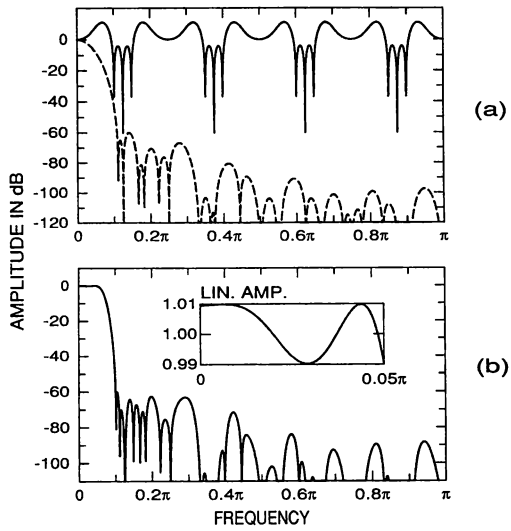


FIGURE 4-51 Amplitude responses for an optimized filter of the form $H(z) = F(z^L)G(z)$, where $G(z)$ is a cascade of four $T_r(z)$'s and $L = 8$. The filter specifications are those of Figure 4-49. (a) $F(z^L)$ (solid line) and $G(z)$ (dashed line). (b) Overall filter.

using the techniques of the previous subsection. The resulting overall transfer function is then

$$H(z) = z^{-M} - (-1)^M F((-z)^L)G(-z), \tag{4.173}$$

where M is half the order of $F(z^L)G(z)$. An implementation of this transfer function is shown in Figure 4-52.[23] To avoid half-sample delays, the order has to be even.

Example 4.17. Let the wideband filter specifications be $\omega_p = 0.9\pi$, $\omega_s = 0.95\pi$, $\delta_p = 0.001$, and $\delta_s = 0.01$. From Eq. (4.170), the specifications of $\overline{H}(z)$ become $\overline{\omega}_p = 0.05\pi$, $\overline{\omega}_s = 0.1\pi$, $\overline{\delta}_p = 0.01$, and $\overline{\delta}_s = 0.001$. These are the narrowband

[23]The delay term z^{-M} can be shared with $\overline{F}(z^L)$.

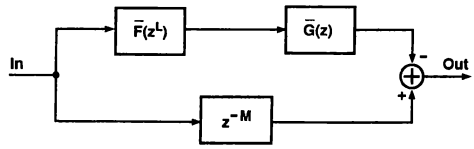


FIGURE 4-52 Implemetation for a wideband filter of the form $H(z) = z^{-M} - (-1)^M F((-z)^L)G(-z)$. $\overline{F}(z^L) \equiv (-1)^M F((-z)^L)$ and $\overline{G}(z) \equiv G(-z)$.

specifications of Example 4.15. Hence the desired wideband design is obtained by using the subfilters $F(z^L)$ and $G(z)$ of Figure 4-49. However, the overall order of the filter of Figure 4-49 is odd (119) and the resulting delay contains a half-sample delay. Therefore, in order to achieve the desired solution with even order, the order of $G(z)$ has to be increased by one. Figure 4-53 gives the amplitude response of the resulting filter. This design requires 19 multipliers, 36 adders, and 120 delay elements, whereas the corresponding numbers for a conventional direct-form equivalent of order 108 are 55, 108, and 108, respectively.

4-11-5 Generalized Designs

The Jing–Fam approach [JI84] is based on iteratively using the facts that a narrowband filter can be implemented effectively as $H(z) = F(z^L)G(z)$ and a wideband filter in the form of Eq. (4.173). If $\omega_s < \pi/2$, then the overall transfer function is first expressed as

$$H(z) \equiv \overline{H}_1(z) = G_1(z)F_1(z^{L_1}) \tag{4.174}$$

and the simultaneous criteria for $G_1(z)$ and $F_1(z)$ are stated according to Eq. (4.161). By denoting by $\delta_p^{(1)}$ the passband ripple of $G_1(z)$, the passband and stopband ripples of $F_1(z)$ are $\delta_p - \delta_p^{(1)}$ and δ_s, respectively. If L_1 is selected such that the passband and stopband edges of $F_1(z)$, $L_1\omega_p$ and $L_1\omega_s$, become larger than $\pi/2$, then $F_1(z)$ can be implemented in terms of a narrowband filter

$$\overline{H}_2(z) = G_2(z)F_2(z^{L_2}) \tag{4.175}$$

in the form

$$F_1(z) = z^{-M_1} - (-1)^{M_1}\overline{H}_2(-z). \tag{4.176}$$

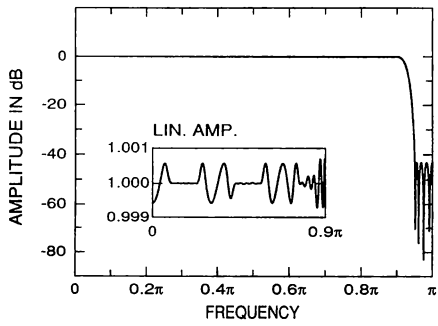


FIGURE 4-53 Amplitude response for a wideband filter implemented as shown in Figure 4-52.

The required passband and stopband ripples of $\overline{H}_2(z)$ are δ_s and $\delta_p - \delta_p^{(1)}$, respectively, whereas the passband and stopband edges are less than $\pi/2$ and given by

$$\omega_p^{(2)} = \pi - L_1\omega_s, \qquad \omega_s^{(2)} = \pi - L_1\omega_p. \tag{4.177}$$

The specifications for $\overline{H}_2(z)$ are thus similar to those of $\overline{H}_1(z)$ and the process can be repeated. Continuing in this manner and designing the last stage in the form $\overline{H}_R(z) = G_R(z)$ results in the following overall transfer function [JI84; SA88b]:

$$H(z) \equiv \overline{H}_1(z) = G_1(z)F_1(z^{L_1}), \tag{4.178a}$$

where

$$F_1(z) = z^{-M_1} - (-1)^{M_1}\overline{H}_2(-z), \qquad \overline{H}_2(z) = G_2(z)F_2(z^{L_2}), \tag{4.178b}$$

$$F_2(z) = z^{-M_2} - (-1)^{M_2}\overline{H}_3(-z), \qquad \overline{H}_3(z) = G_3(z)F_3(z^{L_3}), \tag{4.178c}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$F_{R-2}(z) = z^{-M_{R-2}} - (-1)^{M_{R-2}}\overline{H}_{R-1}(-z), \qquad \overline{H}_{R-1}(z) = G_{R-1}(z)F_{R-1}(z^{L_{R-1}}), \tag{4.178d}$$

$$F_{R-1}(z) = z^{-M_{R-1}} - (-1)^{M_{R-1}}\overline{H}_R(-z), \qquad \overline{H}_R(z) = G_R(z), \tag{4.178e}$$

with M_r for $r = 1, 2, \ldots, R-1$ being half the order of $\overline{H}_{r+1}(z)$.

The above equations give $H(z)$ in an implicit form. The corresponding explicit form is given later. Here, the $G_r(z)$'s for $r = 1, 2, \ldots, R$ are the filters to be designed. The overall criteria are met if the $G_r(z)$'s satisfy (cf. Eq. (4.161))

$$1 - \delta_p^{(r)} \le G_r(\omega) \le 1 + \delta_p^{(r)} \quad \text{for } \omega \in [0, \omega_p^{(r)}], \tag{4.179a}$$

$$-\delta_s^{(r)} \le G_r(\omega) \le \delta_s^{(r)} \quad \text{for } \omega \in \Omega_s^{(r)}, \tag{4.179b}$$

where

$$\Omega_s^{(r)} = \begin{cases} \bigcup_{k=1}^{\lfloor L_r/2 \rfloor} \left[k\dfrac{2\pi}{L_r} - \omega_s^{(r)}, \min\left(k\dfrac{2\pi}{L_r} + \omega_s^{(r)}, \pi \right) \right] & \text{for } r < R \\ \left[\omega_s^{(R)}, \pi \right] & \text{for } r = R. \end{cases} \tag{4.179c}$$

The $\omega_p^{(r)}$'s and $\omega_s^{(r)}$'s for $r = 1, 2, \ldots, R$ are the edges of the $\overline{H}_r(z)$'s and can be determined iteratively as

where $\omega_p^{(1)} = \omega_p$ and $\omega_s^{(1)} = \omega_s$, and the $\delta_s^{(r)}$'s as

$$\delta_s^{(r)} = \begin{cases} \delta_p - \sum\limits_{\substack{k=1 \\ k \text{ odd}}}^{r-1} \delta_p^{(k)} & \text{for } r \text{ even} \\ \delta_s - \sum\limits_{\substack{k=2 \\ k \text{ even}}}^{r-1} \delta_p^{(k)} & \text{for } r \text{ odd.} \end{cases} \tag{4.179e}$$

In order for the overall filter to meet the given ripple requirements, $\delta_s^{(R)}$ and the $\delta_p^{(r)}$'s have to satisfy for R even

$$\sum_{\substack{k=2 \\ k \text{ even}}}^{R} \delta_p^{(k)} = \delta_s, \qquad \delta_s^{(R)} + \sum_{\substack{k=1 \\ k \text{ odd}}}^{R-1} \delta_p^{(k)} = \delta_p \tag{4.180a}$$

or for R odd

$$\sum_{\substack{k=1 \\ k \text{ odd}}}^{R} \delta_p^{(k)} = \delta_p, \qquad \delta_s^{(R)} + \sum_{\substack{k=2 \\ k \text{ even}}}^{R-1} \delta_p^{(k)} = \delta_s. \tag{4.180b}$$

If $2\pi/L_r - \omega_s^{(r)} < \pi/2$ for $r < R$ or $\omega_s^{(R)} < \pi/2$, then the number of multipliers in $G_r(z)$ can be reduced by designing it, using the techniques of Section 4-11-3, in the form

$$G_r(z) = G_r^{(1)}(z^{K_r}) G_r^{(2)}(z). \tag{4.181}$$

After some manipulations, $H(z)$ as given by Eqs. (4.178) and (4.181) can be rewritten in the explicit form shown in Table 4-9. If $G_r(z)$ is a single-stage design, then $G_r^{(1)}(z^{K_r}) \equiv 1$ and $H_r^{(1)}(z) \equiv 1$. It can be shown that in order to obtain the desired overall solution, the order of $G_r(z)$ for $r \geq 2$, denoted by N_r in Table 4-9, has to be even. The realization of $H(z)$ is given Figure 4-54,[24] where

$$d_r = \overline{M}_r - \overline{M}_{r+1}, \quad r = 2, 3, \ldots, R-1, \quad d_R = \overline{M}_R. \tag{4.182}$$

If the edges ω_p and ω_s of $H(z)$ are larger than $\pi/2$, then we set $H(z) \equiv F_1(z)$. In this case, $\delta_p^{(1)} \equiv 0$, $L_1 \equiv 1$, and $G_1(z)$, $\omega_p^{(1)}$, and $\omega_s^{(1)}$ are absent. Furthermore, $\omega_p^{(2)} = \pi - \omega_s$ and $\omega_s^{(2)} = \pi - \omega_p$, and $H_1(z)$ is absent in Figure 4-54 and in Table 4-9.

The remaining problem is to select R, the L_r's, the K_r's, and the ripple values such that the filter complexity is minimized. The following example illustrates this.

[24]The use of the extra delay terms can be avoided by using the transposed structure. In this case, the delay terms can be shared with $H_R^{(1)}(z^{K_R L_R})$ or, if this filter stage is not present, with $H_R^{(2)}(z^{L_R})$. This is because the overall order of this filter stage is usually larger than the sum of the d_r's.

TABLE 4-9 Explicit Form for the Transfer Function in the Jing–Fam Approach

$$H(z) = H_1(z^{\bar{L}_1})[I_2 z^{-\bar{M}_2} + H_2(z^{\bar{L}_2})[I_3 z^{-\bar{M}_3} + H_3(z^{\bar{L}_3})[\cdots$$
$$[I_{R-1} z^{-\bar{M}_{R-1}} + H_{R-1}(z^{\bar{L}_{R-1}})[I_R z^{-\bar{M}_R} + H_R(z^{\bar{L}_R})]] \cdots]]],$$

where

$$H_r(z^{\bar{L}_r}) = H_r^{(1)}(z^{K_r\bar{L}_r}) H_r^{(2)}(z^{\bar{L}_r})$$
$$H_r^{(1)}(z) = G_r^{(1)}(J_r^{(1)} z), \qquad H_r^{(2)}(z) = S_r G_r^{(2)}(J_r^{(2)} z)$$
$$S_1 = 1, \qquad S_r = -(-1)^{\bar{M}_r/\bar{L}_r}, \qquad r = 2, 3, \cdots, R$$
$$J_1^{(2)} = 1, \qquad J_2^{(2)} = -1, \qquad J_r^{(2)} = -[J_{r-1}^{(2)}]^{L_{r-1}}, \qquad r = 3, 4, \cdots, R$$
$$J_r^{(1)} = [J_r^{(2)}]^{K_r}$$
$$\bar{L}_1 = 1, \qquad \bar{L}_r = \prod_{k=1}^{r-1} L_k, \qquad r = 2, 3, \cdots, R$$
$$\bar{M}_R = \tfrac{1}{2}\bar{L}_R N_R, \qquad \bar{M}_{R-r} = \bar{M}_{R-r+1} + \tfrac{1}{2}\bar{L}_{R-r} N_{R-r}, \qquad r = 1, 2, \cdots, R-2$$
$$I_2 = 1, \qquad I_r = [J_{r-1}^{(2)}]^{\bar{M}_r/\bar{L}_{r-1}}, \qquad r = 3, 4, \cdots, R$$
$$N_r = K_r N_r^{(1)} + N_r^{(2)}$$

$N_r^{(1)}$ and $N_r^{(2)}$ are the orders of $G_r^{(1)}(z)$ and $G_r^{(2)}(z)$, respectively.

Example 4.18. Consider the specifications of Example 4.14, that is, $\omega_p = 0.4\pi$, $\omega_s = 0.402\pi$, $\delta_p = 0.01$, and $\delta_s = 0.0001$. In this case, the only alternative is to select $L_1 = 2$. The resulting passband and stopband regions for $G_1(z)$ are

$$\Omega_p^{(1)} = [0, 0.4\pi], \qquad \Omega_s^{(1)} = [0.598\pi, \pi].$$

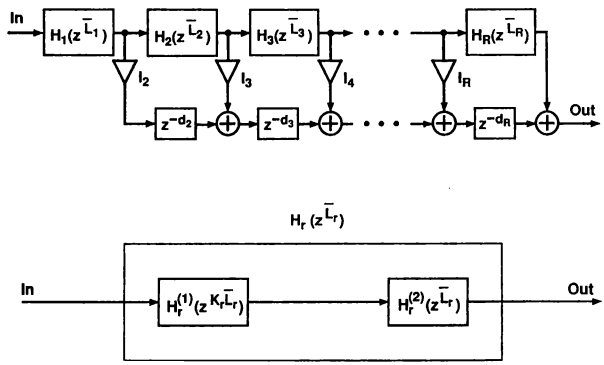


FIGURE 4-54 The structure of a filter synthesized using the Jing–Fam approach.

For $\overline{H}_2(z)$, the edges become $\omega_p^{(2)} = \pi - L_1\omega_s = 0.196\pi$ and $\omega_s^{(2)} = \pi - L_1\omega_p = 0.2\pi$. For L_2, there are two alternatives to make the edges of $\overline{H}_3(z)$, $\omega_p^{(3)} = \pi - L_2\omega_s^{(2)}$ and $\omega_s^{(3)} = \pi - L_2\omega_p^{(2)}$, less than $\pi/2$. These are $L_2 = 3$ and $L_2 = 4$. We select $L_2 = 4$, giving for $G_2(z)$ the following passband and stopband regions:

$$\Omega_p^{(2)} = [0, 0.196\pi], \qquad \Omega_s^{(2)} = [0.3\pi, 0.7\pi] \cup [0.8\pi, \pi].$$

The edges of $\overline{H}_3(z)$ take the values shown in Table 4-10. By selecting $R = 5$, $L_3 = 3$, and $L_4 = 2$, the edges of $\overline{H}_4(z)$ and $\overline{H}_5(z) \equiv G_5(z)$ become as shown in Table 4-10. The passband and stopband regions for $G_3(z)$, $G_4(z)$, $G_5(z)$ are

$$\Omega_p^{(3)} = [0, 0.2\pi], \qquad \Omega_s^{(3)} = [0.4507\pi, 0.8827\pi],$$

$$\Omega_p^{(4)} = [0, 0.352\pi], \quad \Omega_s^{(4)} = [0.6\pi, \pi],$$

$$\Omega_p^{(5)} = [0, 0.2\pi], \qquad \Omega_s^{(5)} = [0.296\pi, \pi].$$

For $R = 5$, it is required that (see Eq. (4.180)) $\delta_p^{(1)} + \delta_p^{(3)} + \delta_p^{(5)} = \delta_p$ and $\delta_p^{(2)} + \delta_p^{(4)} + \delta_s^{(5)} = \delta_s$. The simplest way is to select the ripple values in these summations to be equal. In this case, the required ripples for the $G_r(z)$'s become as shown in Table 4-10. Since the stopband edges of the first and fourth subfilter are larger than $\pi/2$, they are single-stage filters and can be designed using the MPR algorithm. The remaining subfilters can be synthesized, using the techniques of Section 4-11-3, to be two-stage filters. Table 4-10 gives the parameters describing the overall filter, whereas Figure 4-55 gives the resulting overall response. This design

TABLE 4-10 Data for a Filter Designed Using the Jing–Fam Approach

	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$
$\omega_p^{(r)}$	0.4π	0.196π	0.2π	0.352π	0.2π
$\omega_s^{(r)}$	0.402π	0.2π	0.216π	0.4π	0.296π
$\delta_p^{(r)}$	$\frac{1}{3} \times 10^{-2}$	$\frac{1}{3} \times 10^{-4}$	$\frac{1}{3} \times 10^{-2}$	$\frac{1}{3} \times 10^{-4}$	$\frac{1}{3} \times 10^{-2}$
$\delta_s^{(r)}$	10^{-4}	$\frac{2}{3} \times 10^{-2}$	$\frac{2}{3} \times 10^{-4}$	$\frac{1}{3} \times 10^{-2}$	$\frac{1}{3} \times 10^{-4}$
L_r	2	4	3	2	—
K_r	—	3	2	—	3
$N_r^{(1)}$	—	26	13	—	25
$N_r^{(2)}$	38	10	10	32	19
N_r	38	88	36	32	94
L_r	1	2	8	24	48
$J_r^{(1)}$	—	-1	1	—	-1
$J_r^{(2)}$	1	-1	-1	1	-1
M_r	—	2872	2784	2640	2256
I_r	—	1	1	1	1
S_r	1	-1	-1	-1	1
d_r	–	88	144	384	2256

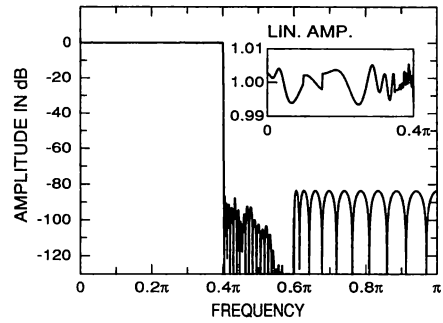


FIGURE 4-55 Amplitude response for a filter synthesized using the Jing–Fam approach.

requires only 93 multipliers and 177 adders, which are less than 6% of those required by an optimum conventional filter (1570 and 3138). The price paid for these reductions is an increased overall filter order (from 3138 to 5782). The overall filter order as well as the number of multipliers can be decreased by selecting the ripple values in a more optimum manner [JI84; SA88b]. Also, in order to arrive at the best solution, different choices of R as well as all alternatives to select the L_r's such that the $\omega_s^{(r)}$'s become smaller than $\pi/2$ are worth going through.

The above Jing–Fam synthesis technique can be used directly for almost all lowpass filter specifications. The only exceptions are filters with $\omega_p < \pi/2$ and $\omega_s > \pi/2$. One alternative to design these filters is to shift the passband and stopband edges of the filter by a factor of $\frac{3}{2}$ using a decimation by a factor of $\frac{3}{2}$ at the input of the filter and an interpolation by the same factor at the output of the filter [JI84].

When comparing the Jing–Fam approach and the frequency-response masking approach of Lim with each other, the Jing–Fam approach normally gives filters with fewer multipliers at the expense of a larger overall filter order [SA88b]. One attractive feature of the Jing–Fam approach is that it can be combined with multirate filtering to reduce the filter complexity even further [RA90] (see Chapter 14).

4-11-6 Design of Other Types of Filters

This section has concentrated on designing lowpass filters. However, the reviewed synthesis techniques can also be applied in a straightforward manner to designing other types of filters.

The design of a highpass filter $H(z)$ with edges at ω_s and ω_p and ripples of δ_s and δ_p can be performed, according to the discussion of Section 4-8-3, by first designing a lowpass filter $\overline{H}(z)$ with edges at $\overline{\omega}_p = \pi - \omega_p$ and $\overline{\omega}_s = \pi - \omega_s$ and ripples of $\overline{\delta}_p = \delta_p$ and $\overline{\delta}_s = \delta_s$. The desired highpass filter is then $H(z) = (-1)^M \overline{H}(-z)$ with M being half the filter order.

Figure 4-56 illustrates how a narrowband bandpass filter can be designed in the form $H(z) = F(z^L)G(z)$. In this case, $L = 5$, the passband and stopband edges are $\omega_{p1}, \omega_{p2} = 0.7\pi \pm 0.04\pi$ and $\omega_{s1}, \omega_{s2} = 0.7\pi \pm 0.06\pi$, respectively, and $\delta_p = \delta_s = 0.001$. $F(\omega)$ has been determined to be a bandpass design in such a way that $F(L\omega)$ takes care of the overall response in the interval $[0.6\pi, 0.8\pi]$, whereas $G(\omega)$ provides the desired attenuation on the extra passbands and transition bands of $F(L\omega)$. The details for designing bandpass filters of this type can be found in Saramäki et al. [SA88a]. $F(z^L)$ is of order 68 in z^L and $G(z)$ is of order 32. The overall design thus requires 52 multipliers and 100 adders, whereas the minimum order of an equivalent conventional direct-form design is 336, requiring 169 multipliers and 336 adders. The complementary bandstop filter can be implemented

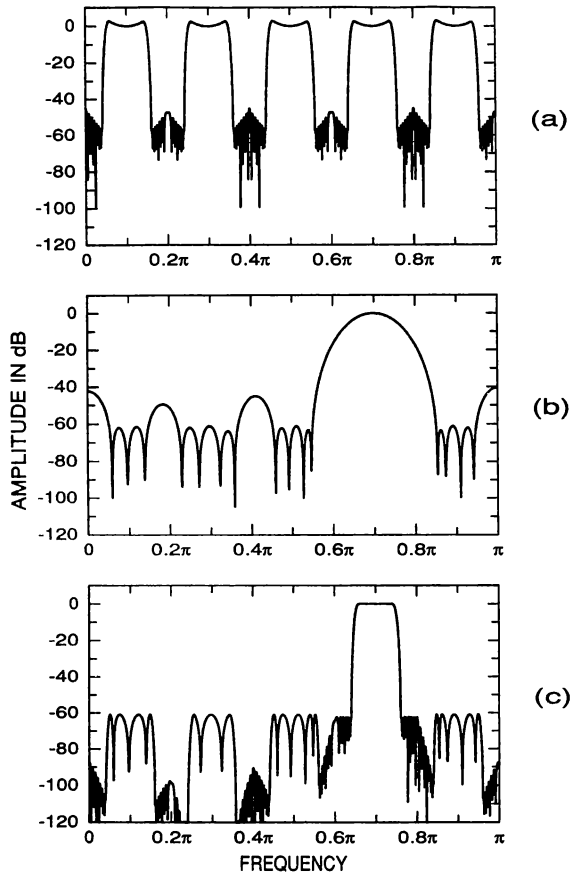


FIGURE 4-56 Amplitude responses for an optimized bandpass filter of the form $H(z) = F(z^L)G(z)$ with $L = 5$. (a) $F(z^L)$ of order 68 in z^L. (b) $G(z)$ of order 32. (c) Overall filter.

directly in the form $z^{-M} - H(z) = z^{-M} - F(z^L)G(z)$, where the delay z^{-M} can be shared with $F(z^L)$.

Another straightforward technique for designing bandpass filters has been proposed by Neuvo, Rajan, and Mitra [NE87; RA88]. This technique is based on the fact that if $\overline{H}(z)$ is a Type I or Type II linear-phase transfer function of order N, then

$$H(z) = e^{jN\omega_0/2}\overline{H}(ze^{j\omega_0}) + e^{-jN\omega_0/2}\overline{H}(ze^{-j\omega_0}) \tag{4.183}$$

is also a linear-phase transfer function of the same type and the same order. The impulse responses of these two filters are related via $h[n] = 2\bar{h}[n] \cos[(n - N/2)\omega_0]$ and the zero-phase frequency responses via (see Figure 4-57)

$$H(\omega) = \overline{H}(\omega + \omega_0) + \overline{H}(\omega - \omega_0). \tag{4.184}$$

Therefore, if $\overline{H}(z)$ is a lowpass design with edges at $\overline{\omega}_p$ and $\overline{\omega}_s$, then $H(z)$ is a bandpass design (see Figure 4-57) with passband edges at $\omega_0 \pm \overline{\omega}_p$ and stopband edges at $\omega_0 \pm \overline{\omega}_s$. The ripples of $H(z)$ are in the worst case $\delta_s = 2\bar{\delta}_s$ and $\delta_p = \bar{\delta}_p$

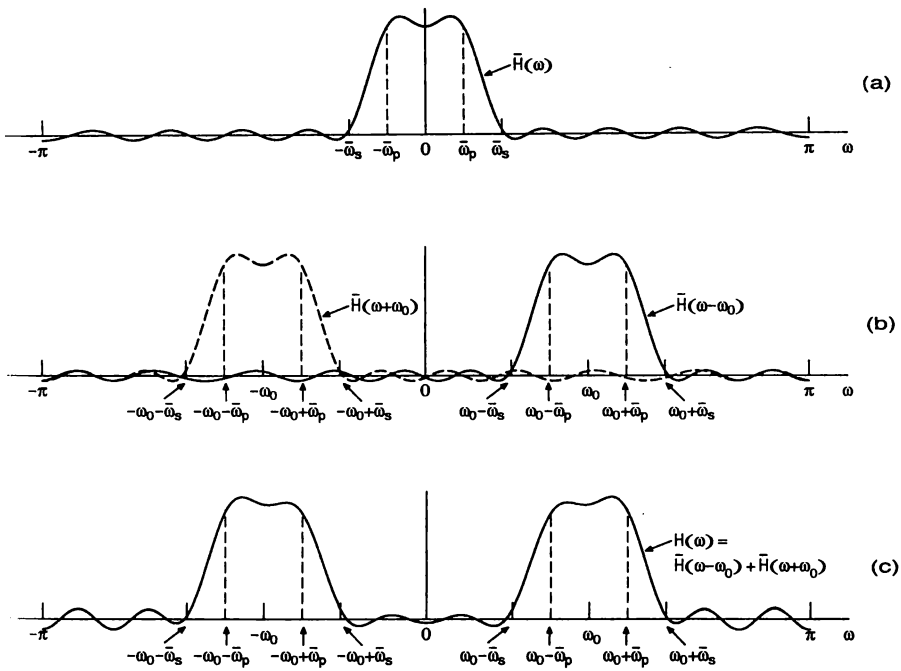


FIGURE 4-57 Design of a linear-phase bandpass filter from a lowpass prototype by modulation. (a) Response $\overline{H}(\omega)$ of the prototype with transfer function $\overline{H}(z)$. (b) Responses $\overline{H}(\omega - \omega_0)$ and $\overline{H}(\omega + \omega_0)$. (c) Response of $H(z) = e^{jN\omega_0/2}\overline{H}(ze^{j\omega_0}) + e^{-jN\omega_0/2}\overline{H}(ze^{-j\omega_0})$.

$+ \bar{\delta}_s$, where $\bar{\delta}_p$ and $\bar{\delta}_s$ are the ripples of $\overline{H}(z)$. In Neuvo et al. [NE87], an efficient structure has been derived for the overall filter in the case where the prototype lowpass filter is of the form $\overline{H}(z) = F(z^L)G(z)$, whereas in Rajan et al. [RA88] a corresponding structure has been given for the case where $\overline{H}(z)$ is designed using the frequency-response masking approach described in Section 4-11-2.

The third approach to exploit the techniques of this section is to design the bandpass filter as a cascade of a lowpass filter $H_{LP}(z)$ and a highpass filter $H_{HP}(z)$, that is, in the form

$$H(z) = H_{LP}(z)H_{HP}(z). \tag{4.185}$$

Here, $H_{LP}(z)$ is designed to provide the second transition band $[\omega_{p2}, \omega_{s2}]$ and $H_{HP}(z)$ to provide the first transition band $[\omega_{s1}, \omega_{p1}]$. The resulting ripple in the first (second) stopband is approximately equal to the stopband ripple of the highpass (lowpass) filter, whereas the passband ripple is in the worst case equal to the sum of the passband ripples of the subfilters.

4-12 DESIGN OF FIR FILTERS USING IDENTICAL SUBFILTERS AS BASIC BUILDING BLOCKS

Another approach to reduce the cost of implementation of an FIR filter is to design it by interconnecting a number of identical subfilters with the aid of a few additional adders and multipliers. Such an approach has been suggested originally by Kaiser and Hamming [KA77a] and improved by Nakamura and Mitra [NA82]. This section concentrates on the most general approach proposed by Saramäki [SA87a]. The main advantage of using identical copies of the same filter lies in the fact that with this approach it is relatively easy to synthesize selective FIR filters without general multipliers [SA91b].

4-12-1 Filter Structures and Conditions for the Subfilter and Tap Coefficients

Figure 4-58 gives two general structures for implementing a linear-phase FIR filter as a tapped cascaded interconnection of N identical subfilters (for other alternatives, see Saramäki [SA87a]). The subfilter has a Type I linear-phase transfer function

$$F_M(z) = \sum_{n=0}^{2M} f[n]z^{-n}, \quad f[2M - n] = f[n]. \tag{4.186}$$

The subscript M is used to emphasize that the delay of the subfilter is M. The frequency response of the structure of Figure 4-58(a) is [SA87a]

$$H(e^{j\omega}) = e^{-jNM\omega} H(\omega), \tag{4.187a}$$

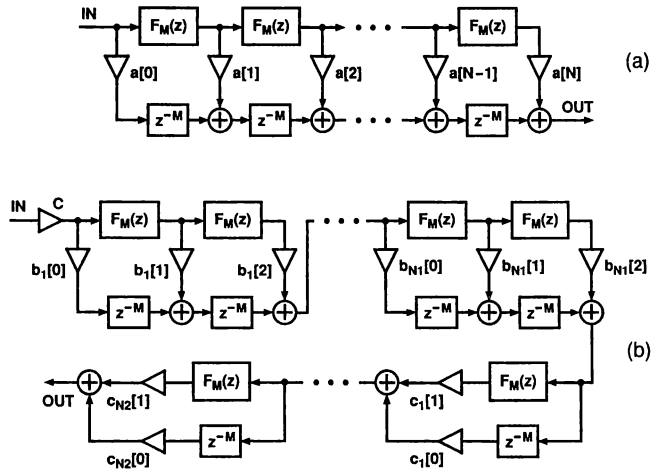


FIGURE 4-58 Two general structures for implementing a linear-phase FIR filter as a tapped cascaded interconnection of N identical subfilters of even order $2M$.

where

$$H(\omega) = \sum_{n=0}^{N} a[n][F_M(\omega)]^n \tag{4.187b}$$

with

$$F_M(\omega) = f[M] + 2\sum_{n=1}^{M} f[M-n]\cos n\omega. \tag{4.187c}$$

The additional tap coefficients $a[n]$ and the subfilter $F_M(z)$ can be determined such that $H(\omega)$ meets

$$1 - \delta_p \le H(\omega) \le 1 + \delta_p \quad \text{for } \omega \in X_p, \tag{4.188a}$$

$$-\delta_s \le H(\omega) \le \delta_s \quad \text{for } \omega \in X_s, \tag{4.188b}$$

where the passband and stopband regions, X_p and X_s, respectively, may consist of several bands. Based on the fact that $H(\omega)$ can be obtained from the polynomial

$$P(x) = \sum_{n=0}^{N} a[n]x^n \tag{4.189}$$

using the substitution (see Eq. (4.187b))

$$x = F_M(\omega), \tag{4.190}$$

the general simultaneous conditions for the $a[n]$'s and $F_M(z)$ can be stated as

$$1 - \delta_p \le P(x) \le 1 + \delta_p \quad \text{for } x_{p1} \le x \le x_{p2}, \tag{4.191a}$$

$$-\delta_s \le P(x) \le \delta_s \quad \text{for } x_{s1} \le x \le x_{s2}, \tag{4.191b}$$

$$x_{p1} \le F_M(\omega) \le x_{p2} \quad \text{for } \omega \in X_p, \tag{4.192a}$$

$$x_{s1} \le F_M(\omega) \le x_{s2} \quad \text{for } \omega \in X_s. \tag{4.192b}$$

Figures 4-59, 4-60, and 4-61 exemplify these relations in three different cases to be considered in more details in this section. As seen from these figures, the substitution $x = F_M(\omega)$ can be regarded as a transformation that maps the passband region $x_{p1} \le x \le x_{p2}$ of $P(x)$ (the stopband region $x_{s1} \le x \le x_{s2}$) onto the passband region X_p (stopband region X_s) of $H(\omega)$. Hence the amplitude values are preserved and only the argument axis is changed. Alternatively, $P(x)$ can be interpreted as an *amplitude change function* [KA77a], which tells that if the subfilter response $F_M(\omega)$ achieves the value x_0, then the overall response $H(\omega)$ achieves the value $P(x_0)$ without regard of the frequency. The passband and stopband regions

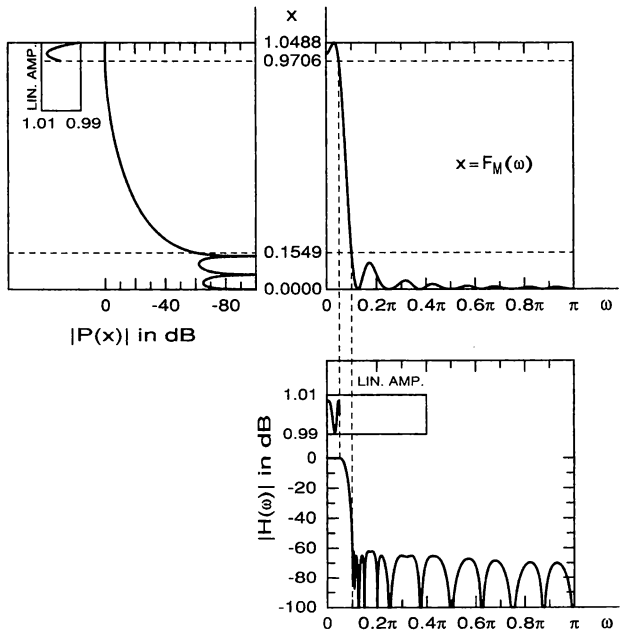


FIGURE 4-59 Design of a composite filter using four prescribed subfilters to meet the lowpass criteria: $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, $\delta_p = 0.01$, and $\delta_s = 0.001$. In this case, $x_{s1} = 0$, $x_{s2} = 0.1549$, $x_{p1} = 0.9706$, and $x_{p2} = 1.0488$ are determined by the subfilter response $F_M(\omega)$.

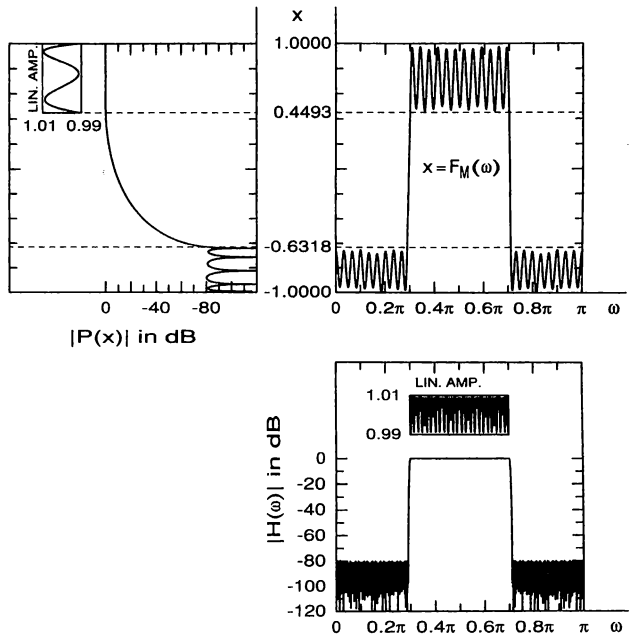


FIGURE 4-60 Design of a composite filter using eight subfilters to meet the bandpass criteria: ω_{p1}, $\omega_{p2} = 0.5\pi \pm 0.2\pi$, ω_{s1}, $\omega_{s2} = 0.5\pi \pm 0.21\pi$, $\delta_p = 0.01$, and $\delta_s = 0.0001$. Case A simultaneous specifications are used for $F_M(\omega)$ and $P(x)$ with $x_{s1} = -1$, $x_{p2} = 1$, $x_{s2} = -0.6318$, and $x_{p1} = 0.4493$.

of $F_M(\omega)$ and $H(\omega)$ are thus the same and all that happens is that the multiple use of the same subfilter reduces the large passband and stopband variations in $F_M(\omega)$ to small variations in $H(\omega)$.

The following two problems[25] can be solved in a straightforward manner:

Problem I Given N, the number of subfilters, optimize the $a[n]$'s (or, equivalently, $P(x)$) and $F_M(z)$ to meet the given criteria with the minimum subfilter order $2M$.

Problem II Given $F_M(z)$, optimize the $a[n]$'s to meet the given criteria with the minimum value of N.

For Problem II, the parameters x_{p1}, x_{p2}, x_{s1}, and x_{s2} are fixed and determined by $F_M(z)$ (see Figure 4-59). For Problem I, these parameters are adjustable and their

[25] In addition to these problems, $F_M(z)$ and $P(x)$ can be optimized to minimize N, the number of subfilters, for the given subfilters order $2M$. Also, the subfilter order can be minimized for the given N and the given values of the $a[n]$'s, like in the Kaiser–Hamming approach [KA77a].

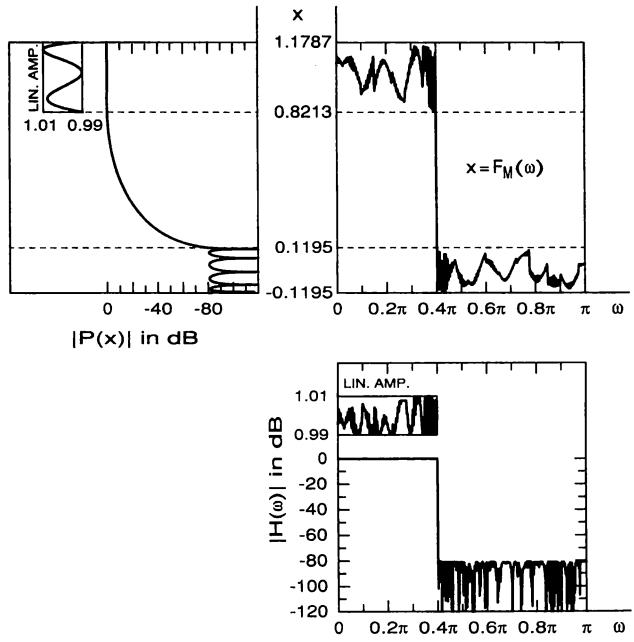


FIGURE 4-61 Design of a composite filter using eight subfilters to meet the lowpass criteria: $\omega_p = 0.4\pi$, $\omega_s = 0.402\pi$, $\delta_p = 0.01$, and $\delta_s = 0.0001$. Case B simultaneous specifications are used for $F_M(\omega)$ and $P(x)$ with x_{p1}, $x_{p2} = 1 \pm 0.1787$ and x_{s1}, $x_{s2} = \pm 0.1195$.

number can be reduced, without loss of generality, from four to two in the following two useful ways:

Case A: $x_{s1} = -1$, $x_{p2} = 1$; x_{s2} and x_{p1} are adjustable.

Case B: $x_{s1} = -\bar{\delta}_s$, $x_{s2} = \bar{\delta}_s$, $x_{p1} = 1 - \bar{\delta}_p$, $x_{p2} = 1 + \bar{\delta}_p$; $\bar{\delta}_p$ and $\bar{\delta}_s$ are adjustable.

Case A is beneficial when the subfilter is a conventional direct-form design, as in this case the subfilter is automatically peak scaled with the maximum and minimum values of $F_M(\omega)$ being $+1$ and -1, respectively (see Figure 4-60). In Case B, the subfilter criteria are conventional with the maximum passband deviation from unity (maximum stopband deviation from zero) being $\bar{\delta}_p(\bar{\delta}_s)$ (see Figure 4-61).

The additional tap coefficients in the structure of Figure 4-58(b) can be obtained by factoring $P(x)$ as given by Eq. (4.189) into the second-order and first-order terms as

$$P(x) = C \prod_{k=1}^{N_1} [b_k[2]x^2 + b_k[1]x + b_k[0]] \prod_{k=1}^{N_2} [c_k[1]x + c_k[0]], \quad (4.193)$$

where $2N_1 + N_2 = N$. The advantages of this structure compared to that of Figure 4-58(a) are that the extra delays z^{-M} can be shared with the subfilters $F_M(z)$ and its sensitivity to variations in the tap coefficients is lower [SA87a].

4-12-2 Filter Optimization

For the above problems, the design of $P(x)$ can be accomplished conveniently with the aid of an FIR filter using the substitution

$$x = \alpha \cos \Omega + \beta \tag{4.194}$$

in $P(x)$, yielding

$$G(\Omega) = P(\alpha \cos \Omega + \beta) = \sum_{n=0}^{N} g[n] \cos^n \Omega, \tag{4.195a}$$

where

$$g[n] = \sum_{r=n}^{N} a[r] \binom{r}{n} \alpha^n \beta^{r-n}. \tag{4.195b}$$

Being expressible as an Nth degree polynomial in $\cos \Omega$, $G(\Omega)$ is the zero-phase frequency response of a Type I linear-phase FIR filter of order $2N$ (see Section 4-3-3 for details) and can be designed using standard FIR filter design algorithms.[26] By selecting

$$\alpha = (x_{p2} - x_{s1})/2, \qquad \beta = (x_{p2} + x_{s1})/2, \tag{4.196}$$

the x-plane regions $[x_{p1}, x_{p2}]$ and $[x_{s1}, x_{s2}]$ are mapped, respectively, onto the Ω-plane regions $[0, \Omega_p]$ and $[\Omega_s, \pi]$, where (see Figure 4-62)

$$\Omega_p = \cos^{-1}\left[\frac{2x_{p1} - x_{p2} - x_{s1}}{x_{p2} - x_{s1}}\right], \qquad \Omega_s = \cos^{-1}\left[\frac{2x_{s2} - x_{p2} - x_{s1}}{x_{p2} - x_{s1}}\right], \tag{4.197}$$

and the conditions for $P(x)$ can be expressed in terms of $G(\Omega)$ as

$$1 - \delta_p \le G(\Omega) \le 1 + \delta_p \quad \text{for } 0 \le \Omega \le \Omega_p, \tag{4.198a}$$

$$-\delta_s \le G(\Omega) \le \delta_s \quad \text{for } \Omega_s \le \Omega \le \pi. \tag{4.198b}$$

[26]These algorithms give the impulse-response coefficients $\bar{g}[n]$ of the corresponding filter. $G(\Omega)$ can be expressed as $G(\Omega) = \bar{g}[N] + 2\Sigma_{n=1}^{N} \bar{g}[N - n] \cos n\Omega$, which can be rewritten in the form of Eq. (4.195a) using the identity $\cos n\Omega = T_n(\cos \Omega)$, where $T_n(x)$ is the nth degree Chebyshev polynomial.

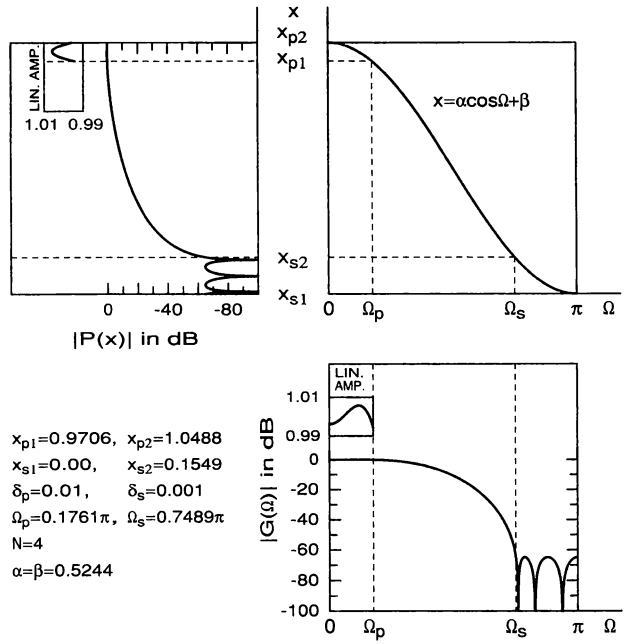


FIGURE 4-62 Design of the polynomial $P(x)$ with the aid of an FIR filter response $G(\Omega)$ for the given x_{s1}, x_{s2}, x_{p1}, and x_{p2} and the given δ_p and δ_s.

$G(\Omega)$ meeting these conventional lowpass specifications can then be converted back into the polynomial $P(x)$ using the substitution

$$\cos \Omega = [x - \beta]/\alpha. \tag{4.199}$$

The resulting tap coefficients $a[n]$ can be determined from the $g[n]$'s according to Eq. (4.195b).

Example 4.19. This example illustrates how multiplier-free filters can be designed by first determining a computationally efficient subfilter with higher ripple values than the required ones and then using the additional tap coefficients to reduce these ripples to the desired level (Problem II). Consider again the specifications: $\omega_p = 0.05\pi$, $\omega_s = 0.1\pi$, $\delta_p = 0.01$, and $\delta_s = 0.001$. For narrowband cases of this kind, a particularly efficient subfilter transfer function is of the form [SA87a]

$$F_M(z) = \left[2^{-P}\frac{1 - z^{-K}}{1 - z^{-1}}\right]^2 [cz^{-K/2} + d(1 + z^{-K})], \tag{4.200}$$

where 2^{-P}, with P integer-valued, is a scaling multiplier and $M = 3K/2 - 1$. An efficient implementation of this transfer function is depicted in Figure 4-63. By

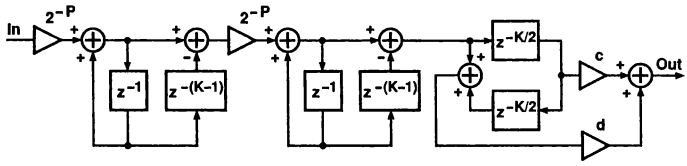


FIGURE 4-63 An implementation of the proposed subfilter.

selecting $K = 16$, $P = 4$, $c = 2$, and $d = -2^{-1}$, the resulting subfilter requires no general multipliers and $F_M(\omega)$ varies within $x_{p1} = 0.9706$ and $x_{p2} = 1.0488$ on $[0, \omega_p]$ and within $x_{s1} = 0$ and $x_{s2} = 0.1549$ on $[\omega_s, \pi]$ (see Figure 4-59). Using Eq. (4.197), the edges of $G(\Omega)$ become $\Omega_p = 0.1761\pi$ and $\Omega_s = 0.7489\pi$ (see Figure 4-62). The minimum even order $2N$ to meet the resulting criteria is 8 so that the required number of subfilters is $N = 4$. When the corresponding polynomial $P(x)$ is factored in the form of Eq. (4.193), only first-order sections are present. By fixing $c_k[1] = 1$ for $k = 1, 2, 3, 4$, the remaining coefficients take the infinite-precision values shown in Table 4-11. The given criteria are still met when these coefficient values are quantized to the easily implementable values shown also in Table 4-11. The resulting composite filter requires no general multiplications, making it very useful for hardware or VLSI implementation. The responses of Figure 4-59 are for this overall design.

Problem I can be solved by finding Ω_p and Ω_s for $G(\Omega)$ of the given even order $2N$ in such a way that it meets the criteria of Eq. (4.198) and the corresponding subfilter criteria become as mild as possible so that they can be met by the minimum even order $2M$. For any $G(\Omega)$, the corresponding polynomial $P(x)$ is obtained using the substitution of Eq. (4.199), where

$$\alpha = 1, \qquad \beta = 0 \tag{4.201}$$

TABLE 4-11 Tap Coefficients for the Filter of Example 4.19

Infinite-Precision	Coefficients
$c_1[1] = 1$	$c_1[0] = -0.009995$
$c_2[1] = 1$	$c_2[0] = -0.075844$
$c_3[1] = 1$	$c_3[0] = -0.144123$
$c_4[1] = 1$	$c_4[0] = -1.323373$
$C = -3.967595$	
Quantized	**Coefficients**
$c_1[1] = 1$	$c_1[0] = 0$
$c_2[1] = 1$	$c_2[0] = -2^{-4}$
$c_3[1] = 1$	$c_3[0] = -2^{-3} - 2^{-6}$
$c_4[1] = 1$	$c_4[0] = -2^{0} - 2^{-2} - 2^{-4}$
$C = -2^2$	

for Case A and

$$\alpha = \frac{2}{2 + \cos \Omega_p - \cos \Omega_s}, \qquad \beta = \frac{1 - \cos \Omega_s}{2 + \cos \Omega_p - \cos \Omega_s} \tag{4.202}$$

for Case B. In Case A, the resulting passband and stopband regions of $P(x)$ are $[x_{p1}, 1]$ and $[-1, x_{s2}]$, where

$$x_{p1} = \cos \Omega_p, \qquad x_{s2} = \cos \Omega_s, \tag{4.203}$$

whereas, in Case B, the corresponding regions are $[1 - \bar{\delta}_p, 1 + \bar{\delta}_p]$ and $[-\bar{\delta}_s, \bar{\delta}_s]$, where

$$\bar{\delta}_p = \frac{1 - \cos \Omega_p}{2 + \cos \Omega_p - \cos \Omega_s}, \qquad \bar{\delta}_s = \frac{1 + \cos \Omega_s}{2 + \cos \Omega_p - \cos \Omega_s}. \tag{4.204}$$

Figure 4-64 exemplifies these relations. Note that $P(x)$ for Case B can be obtained from the Case A polynomial by simply replacing x by $[x - \beta]/\alpha$, where α and β are given by Eq. (4.202).

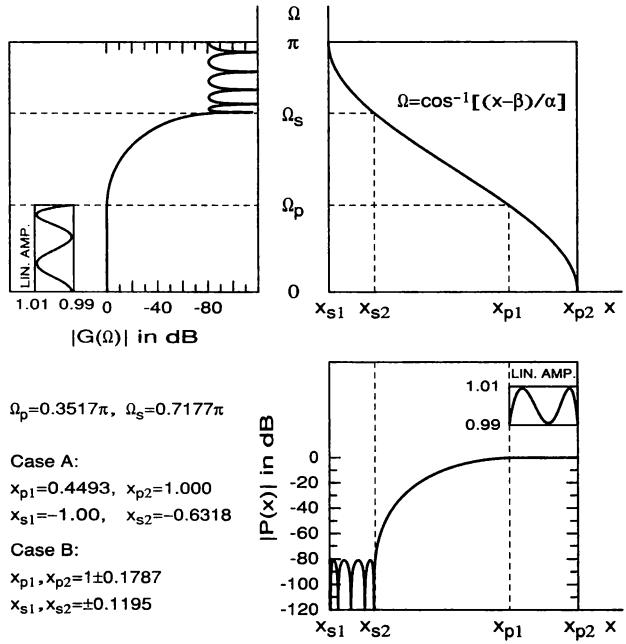


FIGURE 4-64 Relations of the Case A and Case B polynomials $P(x)$ to the best extra-ripple solution $G(\Omega)$ for the given values of δ_p and δ_s. $N = 8$, $\delta_p = 0.009$, and $\delta_s = 0.00009$.

It has turned out that the mildest subfilter criteria are typically obtained at those values of Ω_p and Ω_s for which $G(\Omega)$ has an extraripple solution for the specified values of δ_p and δ_s (see Section 4-8-2). The best extraripple solution is the one for which Ω_p and $\pi - \Omega_s$ are the most equal. As an example, Figure 4-64 gives the best extraripple solution for $N = 8$, $\delta_p = 0.009$, and $\delta_s = 0.00009$ along with the corresponding polynomials $P(x)$ in Cases A and B. Note that the allowable passband and stopband variations for the subfilter are in both cases huge compared to those of the overall design. In Section 4-12-3, the solutions of Figure 4-64 are used as a starting point for synthesizing multiplier-free filters for $\delta_p = 0.01$ and $\delta_s = 0.0001$.

The desired extraripple solutions can be found directly using the algorithm of Hofstetter et al. [HO71]. This algorithm can also be implemented by slightly modifying the MPR algorithm.[27]

Example 4.20. Let the specifications be $\omega_p = 0.4\pi$, $\omega_s = 0.402\pi$, $\delta_p = 0.01$, and $\delta_s = 0.0001$. Table 4-12 gives the minimum subfilter orders $2M$ for various values of N, the number of subfilters, along with the subfilter specifications in

[27]The modified program is available by writing to the author of this chapter.

TABLE 4-12 Data for Filters Synthesized Using Identical Subfilters

Number of Subfilters	Subfilter Order	$\cos \Omega_p$ $\cos \Omega_s$	$\bar{\delta}_p$ $\bar{\delta}_s$	Number of Distinct Coefficients	Overall Filter Order
$N = 1$	3138			1570	3138
$N = 2$	2056	0.98038	0.005000	1032	4112
		−0.94450	0.014142		
$N = 4$	1046	0.79100	0.057687	529	4184
		−0.83208	0.046348		
$N = 6$	692	0.58445	0.125781	354	4152
		−0.71930	0.084963		
$N = 8$	514	0.43774	0.183502	267	4112
		−0.62629	0.121968		
$N = 10$	408	0.33818	0.228955	216	4080
		−0.55241	0.154843		
$N = 15$	268	0.29826	0.266590	151	4020
		−0.33403	0.252999		
$N = 20$	200	0.28324	0.288587	122	4000
		−0.20044	0.321925		
$N = 30$	132	0.17262	0.355611	98	3960
		−0.15404	0.363594		
$N = 40$	98	0.11975	0.391903	91	3920
		−0.12635	0.388965		
$N = 50$	78	0.08953	0.414360	91	3900
		−0.10775	0.406070		

Cases A and B. Also, the number of distinct coefficients,[28] $N + M + 2$, and the overall filter order, $2MN$, are given in the table. The case $N = 1$ corresponds to the conventional direct-form design. It is interesting to observe that the overall filter order for all the cases is approximately 1.3 times (within 1.24–1.33) that of the direct-form design. If all the identical subfilters are implemented separately, then the overall multiplication rate per sample, $N(M + 1) + N + 1$, is higher than that of the direct-form equivalent. However, the structures of Figure 4-58 become advantageous if all the subfilters are implemented using a single subfilter by applying multiplexing. Since the subfilter order can be reduced to any value by increasing the number of subfilters, it has the potential of being realized by a fast short convolution algorithm (see Chapter 8) or implemented using an integrated FIR filter chip.

4-12-3 Design of FIR Filters Without General Multipliers

Using the structure of Figure 4-58(b), it is relatively easy to design high-order filters without general multipliers. Filters of this kind are very attractive in VLSI implementation where a general multiplier is very costly. These filters can be designed in two steps [SA91b]. In the first step, the additional tap coefficients in the structure of Figure 4-58(b) are quantized to values that are simple combinations of powers-of-two. The second step then involves designing the subfilter in such a way that there are no general multipliers. It is relatively easy to get such a subfilter without time-consuming optimization since the ripple values of the subfilter are very large (cf. Table 4-12) and, consequently, large coefficient quantization errors are allowed.[29]

To allow some quantization error for the tap coefficients in the structure of Figure 4-58(b), $G(\Omega)$ of the given order $2N$ is first designed to be the best extra-ripple solution for the passband and stopband ripples of $0.8\delta_p \cdot \cdot \cdot 0.9\delta_p$ and $0.8\delta_s \cdot \cdot \cdot 0.9\delta_s$. This $G(\Omega)$ is then converted to the Case A or Case B polynomial $P(x)$ according to the discussion of Section 4-12-2 and the passband and stopband regions of $P(x)$, $[x_{p1}, x_{p2}]$ and $[x_{s1}, x_{s2}]$, are located. In both Case A and Case B, the resulting $P(x)$ can be factored in the form

$$P(x) = D \prod_{k=1}^{N_1} (x^2 + \alpha_k x + \beta_k) \prod_{k=1}^{N_2} (x + \gamma_k). \tag{4.205}$$

[28]For the structure of Figure 4-58(b), the number of distinct coefficients is $N + M + 2$ if the blocks in the structure are scaled in such a way that the coefficients $b_k[2]$ in the second-order blocks and the coefficients $c_k[1]$ in the first-order blocks become unity. This can be done without loss of generality.

[29]The rule of thumb for direct rounding of FIR filter coefficients is that if the allowed quantization error is made double, one bit is saved. Also, the order of the subfilter is significantly reduced compared to the order of the overall filter. Another rule of thumb for direct rounding is that if there are two filters with the same allowable quantization error and the order of the first filter is one-fourth that of the second filter, then the first filter requires one bit less.

A very straightforward technique to arrive at simple tap coefficients is based on expressing the coefficients of the second- and first-order terms as

$$\alpha_k = b_k[1]/b_k[2], \qquad \beta_k = b_k[0]/b_k[2], \tag{4.206}$$

$$\gamma_k = c_k[0]/c_k[1]. \tag{4.207}$$

The resulting $P(x)$ can be written in the following form corresponding to the structure of Figure 4-58(b):

$$P(x) = C \prod_{k=1}^{N_1} (b_k[2]x^2 + b_k[1]x + b_k[0]) \prod_{k=1}^{N_2} (c_k[1]x + c_k[0]). \tag{4.208}$$

If the coefficients of a second-order term are desired to be quantized to two powers-of-two values, that is, values of the form $\pm 2^{-P_1} \pm 2^{-P_2}$, then a simple technique is to first set $b_k[2]$ to take all possible two powers-of-two values within $\frac{1}{2}$ and 1. Then, for each value of $b_k[2]$, the remaining coefficient values $b_k[1]$ and $b_k[0]$ are determined from Eq. (4.206) and quantized to the closest two powers-of-two values. Finally, those values that provide the closest approximations to α_k and β_k are selected. Quantized values for the $c_k[1]$'s and $c_k[0]$'s can be found in the same manner. The next step is to select C such that the average of $P(x)$ in the passband region $[x_{p1}, x_{p2}]$ is unity and it is checked whether $P(x)$ is within the limits $1 \pm \delta_p$ is the passband region $[x_{p1}, x_{p2}]$ and within the limits $\pm\delta_s$ in the stopband region $[x_{s1}, x_{s2}]$. If not, some of the coefficients require three powers-of-two representations.[30]

What remains is to design a multiplier-free subfilter such that its zero-phase frequency response $F_M(\omega)$ stays within the limits x_{p1} and x_{p2} in the passband region and within the limits x_{s1} and x_{s2} in the stopband region.

Example 4.21. Consider the ripples $\delta_p = 0.01$ and $\delta_s = 0.0001$ and the case where eight subfilters ($N = 8$) are used. Figure 4-64 gives the best extraripple solution of $G(\Omega)$ for the ripple values $0.9\delta_p$ and $0.9\delta_s$ as well as the corresponding polynomial $P(x)$ in both Case A and Case B together with its passband and stopband regions. In both cases, $P(x)$ still meets the given criteria when the additional tap coefficients are quantized, using the above procedure, to the values shown in Table 4-13. In Case A, it is required that $F_M(\omega)$ stays within the limits 0.4493 and 1 in the passband(s) and within the limits -1 and -0.6318 in the stopband(s). In Case B, the required passband and stopband ripples are $\bar{\delta}_p = 0.1787$ and $\bar{\delta}_s = 0.1195$, respectively. In both cases, the allowable passband and stopband variations are thus huge, making the quantization of the subfilter coefficients rather

[30]It has turned out that the second-order terms of $P(x)$ and those first-order terms of $P(x)$ that do not possess zeros in the stopband interval $[x_{s1}, x_{s2}]$ are the most sensitive and require sometimes a higher accuracy.

TABLE 4-13 Quantized Tap Coefficients for the Structure of Figure 4-58(b)

Case A		
$b_1[2] = 2^{-1} + 2^{-6}$	$b_1[1] = -2^0 + 2^{-3} - 2^{-7}$	$b_1[0] = 2^{-1} - 2^{-6} - 2^{-7}$
$c_1[1] = 2^{-1} + 2^{-2}$	$c_1[0] = -2^0 + 2^{-4} + 2^{-5}$	
$c_2[1] = 2^0$	$c_2[0] = 2^0 - 2^{-7}$	
$c_3[1] = 2^0 - 2^{-4}$	$c_3[0] = 2^0 - 2^{-3}$	
$c_4[1] = 2^{-1} + 2^{-3}$	$c_4[0] = 2^{-1} + 2^{-6}$	
$c_5[1] = 2^0 - 2^{-3}$	$c_5[0] = 2^{-1} + 2^{-3}$	
$c_6[1] = 2^0 - 2^{-3}$	$c_6[0] = 2^{-1} + 2^{-4}$	
$C = -2^2 - 2^1 - 2^{-1} - 2^{-5}$		
Case B		
$b_1[2] = 2^0 - 2^{-3} + 2^{-6}$	$b_1[1] = -2^1 + 2^{-4} + 2^{-8}$	$b_1[0] = 2^0 + 2^{-3} - 2^{-8}$
$c_1[1] = 2^{-1} + 2^{-4} + 2^{-7}$	$c_1[0] = -2^{-1} - 2^{-2}$	
$c_2[1] = 2^{-1} + 2^{-6}$	$c_2[0] = 2^{-4} - 2^{-8}$	
$c_3[1] = 2^{-1} + 2^{-3}$	$c_3[0] = 2^{-5} + 2^{-6}$	
$c_4[1] = 2^0$	$c_4[0] = 2^{-7}$	
$c_5[1] = 2^0 - 2^{-5}$	$c_5[0] = -2^{-4}$	
$c_6[1] = 2^0 - 2^{-5}$	$c_6[0] = -2^{-3} + 2^{-6}$	
$C = -2^8 + 2^4 + 2^3 - 2^0$		

trivial. To illustrate this, the design of a bandpass filter for the passband edges of ω_{p1}, $\omega_{p2} = 0.5\pi \pm 0.2\pi$ and stopband edges of ω_{s1}, $\omega_{s2} = 0.5\pi \pm 0.21\pi$ is considered. The minimum even subfilter order to meet the Case A criteria is 112. If the subfilter order is increased to 120, then the given criteria are still met when direct rounding is used to quantize the coefficient values to the closest two powers-of-two values in 8-bit representations. These values are shown in Table 4-14.[31] Note that because of the symmetry of the filter specifications, every second impulse-response value is equal to zero. The responses of Figure 4-60 are for the resulting composite design. The overall filter order is 960, whereas the minimum order of an equivalent conventional direct-form design is 636. The price paid for getting a multiplier-free design is thus a 50% increase in the filter order. If the subfilter order is increased to 136, then with direct rounding we end up with the very simple 6-bit coefficient values of Table 4-15.[32]

Using the frequency-response masking approach described in Section 4-11-2, the lowpass filter criteria for Case B with $\omega_p = 0.4\pi$ and $\omega_s = 0.402\pi$ are met by a subfilter of the form $F_M(z) = F(z^L)G_1(z) + [z^{-N_F L/2} - F(z^L)]G_2(z)$, where L

[31]The infinite-precision filter has been designed such that the desired function is 0.7247 in the passband and -0.8159 in the stopbands and the weighting function is 1 in the passband and 1.55 in the stopbands. With this weighting, the allowable quantization errors in the passband and in the stopbands are about the same.

[32]The infinite-precision filter has been designed using a stopband weighting of 1.7.

TABLE 4-14 Quantized Coefficients[a] for a Bandpass Subfilter of Order 120

$f[0] = 8 \times 2^{-8}$	$f[2] = 20 \times 2^{-8}$	$f[4] = -15 \times 2^{-8}$	$f[6] = 6 \times 2^{-8}$	$f[8] = 3 \times 2^{-8}$
$f[10] = -4 \times 2^{-8}$	$f[12] = -2 \times 2^{-8}$	$f[14] = 5 \times 2^{-8}$	$f[16] = -1 \times 2^{-8}$	$f[18] = -5 \times 2^{-8}$
$f[20] = 4 \times 2^{-8}$	$f[22] = 4 \times 2^{-8}$	$f[24] = -6 \times 2^{-8}$	$f[26] = -1 \times 2^{-8}$	$f[28] = 7 \times 2^{-8}$
$f[30] = -4 \times 2^{-8}$	$f[32] = -6 \times 2^{-8}$	$f[34] = 8 \times 2^{-8}$	$f[36] = 3 \times 2^{-8}$	$f[38] = -12 \times 2^{-8}$
$f[40] = 4 \times 2^{-8}$	$f[42] = 12 \times 2^{-8}$	$f[44] = -12 \times 2^{-8}$	$f[46] = -7 \times 2^{-8}$	$f[48] = 20 \times 2^{-8}$
$f[50] = -4 \times 2^{-8}$	$f[52] = -28 \times 2^{-8}$	$f[54] = 28 \times 2^{-8}$	$f[56] = 34 \times 2^{-8}$	$f[58] = -120 \times 2^{-8}$
$f[60] = -48 \times 2^{-8}$				

[a] $f[n]$ is zero for n odd.

TABLE 4-15 Quantized Coefficients[a] for a Bandpass Subfilter of Order 136

$f[0] = -2 \times 2^{-6}$	$f[2] = 5 \times 2^{-6}$	$f[4] = 0$	$f[6] = 0$	$f[8] = 1 \times 2^{-6}$
$f[10] = 0$	$f[12] = -1 \times 2^{-6}$	$f[14] = 0$	$f[16] = 1 \times 2^{-6}$	$f[18] = -1 \times 2^{-6}$
$f[20] = -1 \times 2^{-6}$	$f[22] = 1 \times 2^{-6}$	$f[24] = 0$	$f[26] = -1 \times 2^{-6}$	$f[28] = 1 \times 2^{-6}$
$f[30] = 1 \times 2^{-6}$	$f[32] = -2 \times 2^{-6}$	$f[34] = 0$	$f[36] = 2 \times 2^{-6}$	$f[38] = -1 \times 2^{-6}$
$f[40] = -2 \times 2^{-6}$	$f[42] = 2 \times 2^{-6}$	$f[44] = 1 \times 2^{-6}$	$f[46] = -3 \times 2^{-6}$	$f[48] = 1 \times 2^{-6}$
$f[50] = 3 \times 2^{-6}$	$f[52] = -3 \times 2^{-6}$	$f[54] = -2 \times 2^{-6}$	$f[56] = 5 \times 2^{-6}$	$f[58] = -1 \times 2^{-6}$
$f[60] = -7 \times 2^{-6}$	$f[62] = 7 \times 2^{-6}$	$f[64] = 8 \times 2^{-6}$	$f[66] = -30 \times 2^{-6}$	$f[68] = -12 \times 2^{-6}$

[a] $f[n]$ is zero for n odd.

TABLE 4-16 Quantized Coefficients for a Subfilter Designed Using the Frequency-Response Masking Approach

$f[0] = 2 \times 2^{-6}$	$f[1] = -4 \times 2^{-6}$	$f[2] = -3 \times 2^{-6}$	$f[3] = -2 \times 2^{-6}$	$f[4] = 1 \times 2^{-6}$
$f[5] = 0$	$f[6] = -1 \times 2^{-6}$	$f[7] = -2 \times 2^{-6}$	$f[8] = 0$	$f[9] = 2 \times 2^{-6}$
$f[10] = 1 \times 2^{-6}$	$f[11] = -2 \times 2^{-6}$	$f[12] = -3 \times 2^{-6}$	$f[13] = 1 \times 2^{-6}$	$f[14] = 4 \times 2^{-6}$
$f[15] = 1 \times 2^{-6}$	$f[16] = -5 \times 2^{-6}$	$f[17] = -6 \times 2^{-6}$	$f[18] = 5 \times 2^{-6}$	$f[19] = 17 \times 2^{-6}$
$f[20] = 28 \times 2^{-6}$				
$g_1[0] = 2 \times 2^{-6}$	$g_1[1] = 3 \times 2^{-6}$	$g_1[2] = -1 \times 2^{-6}$	$g_1[3] = -2 \times 2^{-6}$	$g_1[4] = -1 \times 2^{-6}$
$g_1[5] = 3 \times 2^{-6}$	$g_1[6] = 2 \times 2^{-6}$	$g_1[7] = -4 \times 2^{-6}$	$g_1[8] = -5 \times 2^{-6}$	$g_1[9] = 4 \times 2^{-6}$
$g_1[10] = 20 \times 2^{-6}$	$g_1[11] = 28 \times 2^{-6}$			
$g_2[0] = -3 \times 2^{-6}$	$g_2[1] = -1 \times 2^{-6}$	$g_2[2] = 1 \times 2^{-6}$	$g_2[3] = 2 \times 2^{-6}$	$g_2[4] = 1 \times 2^{-6}$
$g_2[5] = -1 \times 2^{-6}$	$g_2[6] = -2 \times 2^{-6}$	$g_2[7] = 0$	$g_2[8] = 3 \times 2^{-6}$	$g_2[9] = 3 \times 2^{-6}$
$g_2[10] = -1 \times 2^{-6}$	$g_2[11] = -5 \times 2^{-6}$	$g_2[12] = -2 \times 2^{-6}$	$g_2[13] = 7 \times 2^{-6}$	$g_2[14] = 19 \times 2^{-6}$
$g_2[15] = 24 \times 2^{-6}$				

= 16, the order of $F(z)$ is 40, and the orders of $G_1(z)$ and $G_2(z)$ are 22 and 30, respectively. This filter has been slightly overdesigned[33] such that direct rounding can be used to quantize the filter coefficients to the 6-bit values shown in Table 4-16. Only one coefficient ($g_2[14] = 19 \times 2^{-6}$) requires a three powers-of-two representation. Note that no optimization has been used in finding these coefficient values. The overall order is 70% higher than that of a direct-form equivalent (5360 compared to 3138). The responses of Figure 4-61 are for the resulting overall design.

In the above, direct rounding has been used for quantizing the subfilter coefficients. Another technique, leading to better results, is to use mixed integer linear programming [LI83a, LI90].

4-13 SUMMARY

Several techniques have been reviewed for synthesizing linear-phase FIR filters along with efficient realization methods. This chapter started with very fast design methods: designs based on windowing (Section 4-4), least-squared-error designs (Section 4-5), maximally flat designs (Section 4-6), and simple analytic designs (Section 4-7). All these techniques suffer from some drawbacks. For filters designed using windowing, the passband and stopband ripples are approximately equal. Maximally flat filters are useful in applications where the signal should be preserved very accurately near the zero frequency, and least-squared-error filters in applications where white Gaussian noise in the filter stopband is desired to be attenuated as much as possible. However, if the maximum deviation from the given response is of main interest, then filters designed in the minimax sense meet the criteria with a significantly reduced filter order. These filters have been designed in Section 4-8 using the Remez multiple exchange algorithm. This algorithm is the most powerful method for designing arbitrary-magnitude FIR filters in the minimax sense. It can also be used in a straightforward manner for synthesizing nonlinear-phase FIR filters (Section 4-9).

In some applications there are additional constraints in the time domain or in the frequency domain, such as in the case of Nyquist filters or in the case where the transient part of the step response is restricted to vary within the given limits. Filters meeting these additional constraints can be designed in some cases by properly using the Remez algorithm as a subroutine (Section 4-10). However, the most flexible design method for constrained approximation problems of various kinds is linear programming, which has been considered in detail in Section 4-10.

The last two sections of this chapter have been devoted to the design of computationally efficient linear-phase FIR filters. Section 4-11 concentrated on syn-

[33]When designing $G_1(z)$ and $G_2(z)$, a uniform weighting has been used both in the passband and in the stopband. It has turned out that for filters with large passband and stopband ripple values, it is not worth reducing weightings in some parts of the passband and stopband like for filters with small ripple values (see Section 4-11-2). The stopband weighting of 1.6 has been used.

thesizing filters that use as basic building blocks the transfer functions obtained by replacing each unit delay element in a conventional transfer function by multiple delays. By properly combining these transfer functions together, filters with significantly fewer multipliers and adders can be designed. Section 4-12 introduced another approach to reduce the cost of implementation of FIR filters, based on the use of identical subfilters. With this approach, multiplier-free highly selective filters can be designed in a systematic manner. In addition to these design methods, filters with a reduced number of arithmetic operations can be synthesized using structures that generate piecewise-polynomial impulse responses [CH84a, CH84b] or FIR filters that mimic the performance of IIR filters [FA81; SA90]. In these two approaches, very effective implementations are achieved by using feedback loops. Furthermore, multirate filtering (see Chapter 14) and fast convolution algorithms (see Chapter 8) provide efficient implementations for FIR filters.

REFERENCES

[AN82] A. Antoniou, Accelerated procedure for the design of equiripple nonrecursive digital filters. *IEE Proc., Part G: Electron. Circuits Syst.* **129,** 1–10 (1982); and *Erratum: ibid.* p. 107 (1982).

[AN83] A. Antoniou, New improved method for the design of weighted-Chebyshev, nonrecursive, digital filters. *IEEE Trans. Circuits Syst.* **CAS-30,** 740–750 (1983).

[BL58] R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra*. Dover, New York, 1958.

[BO81] R. Boite and H. Leich, A new procedure for the design of high order minimum phase FIR digital or CCD filters. *Signal Process.* **3,** 101–108 (1981).

[BO82] F. Bonzanigo, Some improvements to the design programs for equiripple FIR filters. *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., Paris, France, 1982*, pp. 274–277 (1982).

[BO84] R. Boite and H. Leich, Comments on 'A fast procedure to design equiripple minimum-phase FIR filters.' *IEEE Trans. Circuits Syst.* **CAS-31,** 503–504 (1984).

[CH66] E. W. Cheney, *Introduction to Approximation Theory*. McGraw-Hill, New York, 1966.

[CH84a] S. Chu and C. S. Burrus, Efficient recursive realizations of FIR filters. Part I: The filter structures. *Circuits, Syst. Signal Process.* **3**(1), 3–20 (1984).

[CH84b] S. Chu and C. S. Burrus, Efficient recursive realizations of FIR filters. Part II: Design and applications. *Circuits, Syst. Signal Process.* **3**(1), 21–57 (1984).

[DA63] G. Dantzig, *Linear Programming and Extensions*. Princeton University Press, Princeton, NJ, 1963.

[DO79] J. J. Dongarra, J. R. Bunch, C. B. Moler, and G. W. Stewart, *LINPACK Users' Guide*. SIAM, Philadelphia, PA, 1979.

[FA74] D. C. Farden and L. L. Scharf, Statistical design of nonrecursive digital filters. *IEEE Trans. Acoust., Speech,, Signal Process.* **ASSP-22,** 188–196 (1974).

[FA81] A. T. Fam, MFIR filters: Properties and applications. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-29,** 1128–1136 (1981).

[GO81] E. Goldberg, R. Kurshan, and D. Malah, Design of finite impulse response digital filters with nonlinear phase response. *IEEE-Trans. Acoust., Speech, Signal Process.* **CAS-29,** 1003–1010 (1981).

[GR83] F. Grenez, Design of linear or minimum-phase FIR filters by constrained Chebyshev approximation. *Signal Process.* **5,** 325–332 (1983).

[HA63] G. Hadley, *Linear Programming*. Addison-Wesley, Reading, MA, 1963.

[HA77] R. W. Hamming, *Digital Filters*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[HA78] F. J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform. *Proc. IEEE* **66,** 51–83 (1978).

[HA87] F. J. Harris, Multirate FIR filters for interpolating and desampling. In *Handbook of Digital Signal Processing, Engineering Applications* (D. F. Elliott, ed.), Chapter 3 (see Appendix). Academic Press, San Diego, CA, 1987.

[HE68] H. D. Helms, Nonrecursive digital filters: Design methods for achieving specifications on frequency response. *IEEE Trans. Audio Electroacoust.* **AU-16,** 336–342 (1968).

[HE70] O. Herrmann and H. W. Schüssler, Design of nonrecursive digital filters with minimum phase. *Electron. Lett.* **6,** 329–330 (1970); also reprinted in L. R. Rabiner and C. M. Rader, eds., *Digital Signal Processing*, pp. 185–186. IEEE Press, New York, 1972.

[HE71a] H. D. Helms, Digital filters with equiripple or minimax responses. *IEEE Trans. Audio Electroacoust.* **AU-19,** 87–93 (1971); also reprinted in L. R. Rabiner and C. M. Rader, eds., *Digital Signal Processing*, pp. 131–137. IEEE Press, New York, 1972.

[HE71b] O. Herrmann, On the approximation problem in nonrecursive digital filter design. *IEEE Trans. Circuit Theory* **CT-18,** 411–413 (1971); also reprinted in L. R. Rabiner and C. M. Rader, eds., *Digital Signal Processing*, pp. 202–203. IEEE Press, New York, 1972.

[HE73] O. Herrmann, L. R. Rabiner, and D. S. K. Chan, Practical design rules for optimum finite impulse response lowpass digital filters. *Bell Syst. Tech. J.* **52,** 769–799 (1973).

[HO71] E. Hofstetter, A. Oppenheim, and J. Siegel, A new technique for the design of non-recursive digital filters. *Proc. 5th Annu. Princeton Conf. Inf. Sci. Syst.*, pp. 64–72 (1971); also reprinted in L. R. Rabiner and C. M. Rader, eds., *Digital Signal Processing*, pp. 187–194. IEEE Press, New York, 1972.

[JI84] Z. Jing and A. T. Fam, A new structure for narrow transition band, lowpass digital filter design. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-32,** 362–370 (1984).

[KA63] J. F. Kaiser, Design methods for sampled data filters. *Proc. Allerton Conf. Circuit Syst. Theory, 1st, Monticello, IL, 1963*, pp. 221–236 (1963); also reprinted in L. R. Rabiner and C. M. Rader, eds., *Digital Signal Processing*, pp. 20–34. IEEE Press, New York, 1972.

[KA66] J. F. Kaiser, Digital filters. In *System Analysis by Digital Computer* (F. F. Kuo and J. F. Kaiser, eds.), Chapter 7. Wiley, New York, 1966.

[KA74] J. F. Kaiser, Nonrecursive digital filter design using the I_0-sinh window function. *Proc. IEEE Int. Symp. Circuits Syst., 1974, San Francisco*, pp. 20–23 (1974); also reprinted in Digital Signal Processing Committee and IEEE ASSP, eds., *Selected Papers in Digital Signal Processing, II*, pp. 123–126. IEEE Press, New York, 1975.

[KA77a] J. F. Kaiser and R. W. Hamming, Sharpening the response of a symmetric nonrecursive filter by multiple use of the same filter. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-25,** 415–422 (1977).

[KA77b] J. F. Kaiser and W. A. Reed, Data smoothing using low-pass digital filters. *Rev. Sci. Instrum.* **48,** 1447–1457 (1977).

[KA79] J. F. Kaiser, Design subroutine (MXFLAT) for symmetric FIR low pass digital filters with maximally-flat pass and stop bands. In *Programs for Digital Signal Processing* (Digital Signal Processing Committee and IEEE ASSP, eds.), pp. 5.3-1–5.3-6. IEEE Press, New York, 1979.

[KA83a] J. F. Kaiser and K. Steiglitz, Design of FIR filters with flatness constraints. *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Boston, 1983*, pp. 197–200 (1983).

[KA83b] Y. Kamp and C. J. Wellekens, Optimal design of minimum-phase FIR filters. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-31,** 922–926 (1983).

[KE72] W. C. Kellogg, Time domain design of nonrecursive least mean-square digital filters. *IEEE Trans. Audio Electrosacoust.* **AU-20,** 155–158 (1972).

[KI88] H. Kikuchi, H. Watanabe, and T. Yanagisawa, Interpolated FIR filters using cyclotomic polynomials. *Proc. IEEE Int. Symp. Circuits Syst., Espoo, Finland, 1988*, pp. 2009–2012 (1988).

[LA73] A. Land and S. Powell, *Fortran Codes for Mathematical Programming*. Wiley, New York, 1973.

[LE75] P. Leistner and T. W. Parks, On the design of FIR digital filters with optimum magnitude and minimum phase. *Arch. Elektron. Uebertragungstech.* **29,** 270–274 (1975).

[LI83a] Y. C. Lim and S. R. Parker, FIR filter design over a discrete powers-of-two coefficient space. *IEEE Trans. Acoust. Speech, Signal Process.* **ASSP-31,** 583–591 (1983).

[LI83b] Y. C. Lim, Efficient special purpose linear programming for FIR filter design. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-31,** 963–968 (1983).

[LI83c] Y. C. Lim and S. R. Parker, Discrete coefficient FIR digital filter design based upon an LMS criteria. *IEEE Trans. Circuits Syst.* **CAS-30,** 723–739 (1983).

[LI85] J. K. Liang, R. J. P. de Figueiredo, and F. C. Lu, Design of optimal Nyquist, partial response, *N*th band, and nonuniform tap spacing FIR digital filters using linear programming techniques. *IEEE Trans. Circuits System.* **CAS-32,** 386–392 (1985).

[LI86] Y. C. Lim, Frequency-response masking approach for the synthesis of sharp linear phase digital filters. *IEEE Trans. Circuits System.* **CAS-33,** 357–364 (1986).

[LI90] Y. C. Lim, Design of discrete-coefficient-value linear phase FIR filters with optimum normalized peak ripple magnitude. *IEEE Trans. Circuits Syst.* **CAS-37,** 1480–1486, (1990).

[MC73a] J. H. McClellan and T. W. Parks, A unified approach to the design of optimum FIR linear phase digital filters. *IEEE Trans. Circuit Theory* **CT-20,** 697–701 (1973).

[MC73b] J. H. McClellan, T. W. Parks, and L. R. Rabiner, A computer program for designing optimum FIR linear phase digital filters. *IEEE Trans. Audio Electroacoust.* **AU-21,** 506–526 (1973); also reprinted in Digital Signal Processing Committee and IEEE ASSP, eds., *Selected Papers in Digital Signal Processing, II*, pp. 97–117. IEEE Press, New York, 1975.

[MC79] J. H. McClellan, T. W. Parks, and L. R. Rabiner, FIR linear phase filter design program. In *Programs for Digital Signal Processing*, (Digital Signal Processing committee and IEEE ASSP, eds.), pp. 5.1-1–5.1-13. IEEE Press, New York, 1979.

[MI82a] G. A. Mian and A. P. Naider, A fast procedure to design equiripple minimum-phase FIR filters. *IEEE Trans. Circuits Syst.* **CAS-29,** 327–331 (1982).

[MI82b] F. Mintzer, On half-band, third-band, and Nth-band FIR filters and their design. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-30,** 734–738 (1982).

[NA82] S. Nakamura and S. K. Mitra, Design of FIR digital filters using tapped cascaded FIR subfilters. *Circuits, Syst. Signal Process.* **1**(1), 43–56 (1982).

[NE84] Y. Neuvo, C.-Y. Dong, and S. K. Mitra, Interpolated finite impulse response filters. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-32,** 563–570 (1984).

[NE87] Y. Neuvo, G. Rajan, and S. K. Mitra, Design of narrow-band FIR bandpass digital filters with reduced arithmetic complexity. *IEEE Trans. Circuits Syst.* **CAS-34,** 409–419 (1987).

[NG88] T. Q. Nguyen, T. Saramäki, and P. P. Vaidyanathan, Eigenfilters for the design of special transfer functions with applications in multirate signal processing. *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., New York, 1988*, pp. 1467–1470 (1988).

[OP89] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Chapters 5 and 7. Prentice-Hall, Englewood Cliffs, NJ, 1989.

[PA72a] T. W. Parks and J. H. McClellan, Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circuit Theory* **CT-19,** 189–194 (1972).

[PA72b] T. W. Parks and J. H. McClellan, A program for the design of linear phase finite impulse response digital filters. *IEEE Trans. Audio Electroacoust.* **AU-20,** 195–199 (1972).

[PA73] T. W. Parks, L. R. Rabiner, and J. H. McClellan, On the transition width of finite impulse response digital filters. *IEEE Trans. Audio Electroacoust.* **AU-21,** 1–4 (1973).

[PA87] T. W. Parks and C. S. Burrus, *Digital Filter Design*, Chapters 2 and 3. Wiley, New York, 1987.

[RA72a] L. R. Rabiner, The design of finite impulse response digital filters using linear programming techniques. *Bell Syst. Tech. J.* **51,** 1177–1198 (1972).

[RA72b] L. R. Rabiner, Linear program design of finite impulse response (FIR) digital filters. *IEEE Trans. Audio Electroacoust.* **AU-20,** 280–288 (1972).

[RA73a] L. R. Rabiner and O. Herrmann, The predictability of certain optimum finite impulse response digital filters. *IEEE Trans. Circuit Theory* **CT-20,** 401–408 (1973).

[RA73b] L. R. Rabiner and O. Herrmann, On the design of optimum FIR low-pass filters with even impulse response duration. *IEEE Trans. Audio Electroacoustic* **AU-21,** 329–336 (1973).

[RA73c] L. R. Rabiner, Approximate design relationships for low-pass FIR digital filters. *IEEE Trans. Audio Electroacoust.* **AU-21,** 456–460 (1973); also reprinted in Digital Signal Processing Committee and IEEE ASSP, eds., *Selected Papers in Digital Signal Processing, II*, pp. 118–122. IEEE Press, New York, 1975.

[RA74] L. R. Rabiner, J. F. Kaiser, and R. W. Schafer, Some considerations in the design of multiband finite impulse-response digital filters. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-22,** 462–472 (1974).

[RA75a] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, Chapter 3. Prentice-Hall, Englewood Cliffs, NJ, 1975.

[RA75b] L. R. Rabiner, J. H. McClellan, and T. W. Parks, FIR digital filter design techniques using weighted Chebyshev approximation. *Proc. IEEE* **63,** 595–610 (1975); also reprinted in Digital Signal Processing Committee and IEEE ASSP, eds., *Selected Papers in Digital Signal Processing, II*, pp. 81–96. IEEE Press, New York, 1975.

[RA88] G. Rajan, Y. Neuvo, and S. K. Mitra, On the design of sharp cutoff wideband FIR filters with reduced arithmetic complexity. *IEEE Trans. Circuits Syst.* **CAS-35,** 1447–1454 (1988).

[RA90] T. Ramstad and T. Saramäki, Multistage, multirate FIR filter structures for narrow transition-band filters. *Proc. IEEE Int. Symp. Circuits Syst., New Orleans, Louisiana, 1990*, pp. 2017–2021 (1990).

[RI64] J. R. Rice, *The Approximation of Functions*, Vol. 1. Addison-Wesley, Reading, MA, 1964.

[SA87a] T. Saramäki, Design of FIR filters as a tapped cascaded interconnection of identical subfilters. *IEEE Trans. Circuits Syst.* **CAS-34,** 1011–1029 (1987).

[SA87b] T. Saramäki and Y. Neuvo, A Class of FIR Nyquist (*N*th-Band) filters with zero intersymbol interference. *IEEE Trans. Circuits Syst.* **CAS-34,** 1182–1190 (1987).

[SA88a] T. Saramäki, Y. Neuvo, and S. K. Mitra, Design of computationally efficient interpolated FIR filters. *IEEE Trans. Circuits Syst.* **CAS-35,** 70–88 (1988).

[SA88b] T. Saramäki and A. T. Fam, Subfilter approach for designing efficient FIR filters. *Proc. IEEE Int. Symp. Circuits Syst., Espoo, Finland, 1988*, pp. 2903–2915 (1988).

[SA88c] H. Samueli, On the design of optimal equiripple FIR digital filters for data transmission applications. *IEEE Trans. Circuits Syst.* **CAS-35,** 1542–1546 (1988).

[SA89a] T. Saramäki, A class of window functions with nearly minimum sidelobe energy for designing FIR filters. *Proc. IEEE Int. Symp. Circuits Syst., Portland, Oregon, 1989*, pp. 359–362 (1989).

[SA89b] T. Saramäki and S. K. Mitra, Design of efficient interpolated FIR filters using recursive running sum filters. *Proc. Int. Symp. Networks, Syst. Signal Process., 6th, Zagreb, Yugoslavia*, pp. 20–23 (1989).

[SA90] T. Saramäki and A. T. Fam, Properties and structures of linear-phase FIR filters based on switching and resetting of IIR filters. *Proc. IEEE Int. Symp. Circuits Syst., New Orleans, Louisiana, 1990*, pp. 3271–3274 (1990).

[SA91a] T. Saramäki, Adjustable windows for the design of FIR filters—A tutorial (invited paper). *Proc. Mediter. Electrotech. Conf., 6th, Ljubljana, Yugoslavia, 1991*, pp. 28–33 (1991).

[SA91b] T. Saramäki, A systematic technique for designing highly selective multiplier-free FIR filters. *Proc. IEEE Int. Conf. Circuits Syst., Singapore, 1991*, pp. 484–487 (1991).

[SC88] H. W. Schüssler and P. Steffen, Some advanced topics in filter design. In *Advanced Topics in Signal Processing*. (J. S. Lim and A. V. Oppenheim, eds.), Chapter 8. Prentice-Hall, Englewood Cliffs, NJ, 1988.

[ST79] K. Steiglitz, Optimal design of FIR digital filters with monotone passband response. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-27,** 643–649 (1979).

[TU70] D. W. Tufts and J. T. Francis, Designing digital lowpass filters: Comparison of some methods and criteria. *IEEE Trans. Audio Electroacoust.* **AU-18,** 487–494 (1970).

[VA84] P. P. Vaidyanathan, On maximally-flat linear-phase FIR filters. *IEEE Trans. Circuits Syst.* **CAS-31,** 830–832 (1984).

[VA85] P. P. Vaidyanathan, Optimal design of linear-phase FIR digital filters with very flat passbands and equiripple stopbands. *IEEE Trans. Circuits Syst.* **CAS-32,** 904–917 (1985).

[VA87] P. P. Vaidyanathan and T. Q. Nguyen, Eigenfilters: A new approach to least-squares FIR filter design and applications including Nyquist filters. *IEEE Trans. Circuits Syst.* **CAS-34,** 11–23 (1987).

[VA89] P. P. Vaidyanathan, T. Q. Nguyen, Z. Doganata, and T. Saramäki, Improved technique for design of perfect reconstruction FIR QMF bands with lossless polyphase matrices. *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-37,** 1042–1056 (1989).