ABSTRACT

Recently, it has been observed that the complexity of finite-impulse-response (FIR) filters using the frequency-response masking technique can be significantly reduced by simultaneously optimizing all the subfilters. In order to reduce the filter complexity even further, this paper considers the optimization of a class of FIR filters based on the use of the generalized frequency-response masking approach originally introduced by Lim and Lian. For this class of filters, the two masking filters are constructed by using a common filter part. Again, the simultaneous optimization of all the subfilters is used, resulting in a considerable reduction in the overall filter complexity. An example taken from the literature illustrates the benefits of the proposed approach.

Index Terms—Optimization, narrow-transition-band filter, frequency-response masking approach.

1. INTRODUCTION

One of the most efficient techniques for synthesizing lowpass linear-phase finite-impulse-response (FIR) digital filters with a drastically reduced number of multipliers and adders compared to the conventional direct-form implementation is the frequency-response masking approach [1]–[4]. The price paid for this reduction in the computational complexity is a slight increase in the filter order. A drawback of the original synthesis techniques is that the subfilters have been designed separately. In [5], an approach for simultaneously designing the subfilters has been proposed. It has been shown, by means of an example, that the number of adders and multipliers of the resulting filters are less than 80 percent compared to those obtained by using the original design schemes.

In order to reduce even further the filter complexity, this paper introduces an efficient optimization algorithm for synthesizing FIR filters based on the generalized frequency-response masking approach, where the masking filters are generated using a common filter part, as introduced by Lim and Lian in [6]. The motivation for considering the generalized filters lies in the fact that the amplitude responses for the masking filters after the simultaneous optimization are very similar (see, e.g., Fig. 6 in [5]). An example taken from the literature illustrates the efficiency of the resulting optimized filters.

2. FREQUENCY-RESPONSE MASKING APPROACH

This section reviews shortly two approaches for designing FIR filters based on the use of frequency-response masking approach, namely, the original approach [1]–[4] and the approach based on simultaneously optimizing the subfilters [5]. In addition, the practical filter synthesis is discussed.

Fig. 1 An efficient implementation for a filter synthesized using the frequency-response masking approach. In this implementation, \( G_1(z) \) and \( G_2(z) \) can share their delays if a transposed direct-form implementation (exploiting the coefficient symmetry) is used.

2.1. Original Approach

In the original frequency-response masking approach, the linear-phase FIR filter transfer function is constructed as (see Fig. 1)

\[
H(z) = F(z^L)G_1(z) + [z^{-LN_f/2} - F(z^L)]G_2(z),
\]

(1a)

where

\[
F(z^L) = \sum_{n=0}^{N_F} f(n)z^{-nL},
\]

(1b)

\[
G_1(z) = z^{-M_1} \sum_{n=0}^{N_1} g_1(n)z^{-n}, \quad G_2(z) = z^{-M_2} \sum_{n=0}^{N_2} g_2(n)z^{-n}.
\]

(1c)

Here, the impulse response coefficients \( f(n) \), \( g_1(n) \), and \( g_2(n) \) possess an even symmetry. \( N_F \) is even, whereas both \( N_1 \) and \( N_2 \) are either even or odd. For \( N_1 \geq N_2 \), \( M_1 = 0 \) and \( M_2 = (N_1 - N_2)/2 \), whereas for \( N_1 < N_2 \), \( M_1 = (N_2 - N_1)/2 \) and \( M_2 = 0 \). These selections guarantee that the delays of both of the terms of \( H(z) \) are equal.

The zero-phase frequency response of \( H(z) \) (the phase term \( e^{-jM\omega/2} \) with \( M = LN_F + \max\{N_1, N_2\} \) is omitted) can be expressed as

\[
H(\omega) = H_1(\omega) + H_2(\omega),
\]

(2a)

where

\[
H_1(\omega) = F(\omega)LG_1(\omega), \quad H_2(\omega) = [1 - F(\omega)L]G_2(\omega)
\]

(2b)

with \( F(\omega) \), \( G_1(\omega) \), and \( G_2(\omega) \) being the zero-phase frequency responses of \( F(z) \), \( G_1(z) \), and \( G_2(z) \), respectively.

The efficiency of \( H(z) \) as given by Eq. (1) lies in the fact that the pair of complementary filter transfer functions \( F(z^L) \) and

---

The sum of their zero-phase frequency responses, as given by \( F(\omega)L \) and \( 1 - F(\omega)L \), is unity.

---

This work was supported by the Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program (2000–2005)).
z^{−LN_F/2} − F(z^L) is obtained from the prototype complementary transfer functions \( F(z) = \sum_{n=0}^{N_F} f(n)z^{-n} \) and \( z^{-LN_F/2} − F(z) \) by replacing \( z \) by \( z^{-L} \). These substitutions make \( F(I\Omega) \) and \( 1 − F(I\Omega) \) periodic with periodicity equal to \( 2\pi/L \) without increasing the number of multipliers, as shown in Fig. 2. Only the number of delay elements required in the implementation, or equivalently, the overall filter orders are increased by a factor of \( L \), resulting in FIR filters with a sparse impulse response with only every \( L \)-th impulse-response value being nonzero. Most importantly, if \( F(z) \) and \( z^{-LN_F/2} − F(z) \) form a lowpass-highpass filter pair with edges at \( \theta \) and \( \phi \) as shown in Fig. 2(a), then \( F(z^L) \) and \( z^{-LN_F/2} − F(z^L) \) provide several transition bands of width \( (\phi − \theta)/L \). Since the order of a linear-phase FIR filter is roughly inversely proportional to the transition bandwidth, this means that for a conventional FIR filter without zero-valued impulse-response values, the number of multipliers is approximately \( LT \) times larger to generate one of these transition bands.

For a lowpass overall transfer function \( H(z) \), one of the transition bands provided by \( F(z^L) \) or \( z^{-LN_F/2} − F(z^L) \) can be used as that of the overall filter. In the first case, denoted by Case A, the edges are given by (see Fig. 3)

\[
\omega_p = \frac{(2l\pi + \theta)}{L}, \quad \omega_s = \frac{(2l\pi + \phi)}{L},
\]

where \( l \) is a fixed integer, and in the second case, referred to as Case B, by (see Fig. 4)

\[
\omega_p = \frac{(2l\pi - \phi)}{L}, \quad \omega_s = \frac{(2l\pi - \theta)}{L}.
\]

Because of the periodic response of \( F(z^L) \) or \( z^{-LN_F/2} − F(z^L) \) they cannot be used alone. For generating the desired response, the masking filters \( G_1(z) \) and \( G_2(z) \) are used as shown in Figs. 3 and 4 for Cases A and B, respectively. In the original approach, the overall filter with maximum allowable passband and stopband ripples equal to \( \delta_p \) on \( [0, \omega_p] \) and \( \delta_s \) on \( [\omega_s, \pi] \) is carried out as follows:

Step 1: Design \( G_k(z) \) for \( k = 1, 2 \) such that their zero-phase frequency responses approximate unity in their passbands with tolerance less than or equal to 0.96\( \delta_p \) and zero in their stopbands with tolerance less than or equal to 0.96\( \delta_s \).

Step 2: Design \( F(I\Omega) \) such that the overall response \( H(\omega) \) approximates unity with tolerance \( \delta_p \) on (see Figs. 3 and 4)

\[
\Omega^{(p)}_p = \begin{cases} 
(2l\pi - \theta)/L, & \text{for Case A} \\
(2l\pi - \phi)/L & \text{for Case B}
\end{cases}
\]

\[
\Omega^{(s)}_s = \begin{cases} 
(2l\pi + \phi)/L, & \text{for Case A} \\
(2l\pi + \theta)/L & \text{for Case B}
\end{cases}
\]

2.2. Design Scheme Based on Simultaneously Optimizing the Subfilters

In [5], an approach for simultaneously optimizing the subfilters has been proposed. In the overall algorithm, the goal is to design \( F(z^L) \) to provide the desired overall filter performance on \( [\Omega_{p1}, \Omega_{p2}] \cap [\Omega_{s1}, \Omega_{s2}] \), where \( \Omega_{p1} = 2l\pi/L \), \( \Omega_{p2} = (2l\pi + \theta)/L \), \( \Omega_{s1} = (2l\pi + \phi)/L \), and \( \Omega_{s2} = (2l\pi + \theta)/L \) for Case A designs, whereas \( \Omega_{p1} = (2l - 1)\pi/L \), \( \Omega_{p2} = (2l\pi - \phi)/L \), \( \Omega_{s1} = (2l\pi - \theta)/L \), and \( \Omega_{s2} = 2l\pi/L \) for Case B designs (see Figs. 3 and 4). The roles of \( G_1(z) \) and \( G_2(z) \) are to give the overall filter performance in the remaining parts of the passband and stopband. This can be done in two steps. First, an iterative algorithm is used to alternately designing \( F(z^L) \) and the masking filters \( G_1(z) \) and \( G_2(z) \) to take care of their frequency-response-shaping responsibilities until the difference between successive overall solutions is within the given tolerance limit. Then, nonlinear optimization techniques, like the Dutta-Vidyasagar algorithm [7] or sequential quadratic programming [8], are applied to improve the filter performance using the result of the first step as a start-up solution.
2.3. Practical Filter Synthesis

In practice, \( \omega_p \) and \( \omega_s \) are given and \( l, L, \theta, \) and \( \phi \) must be determined to give a solution with a significantly reduced arithmetic complexity. For a given value of \( L, \) either Case A or Case B can be used provided that \( L \) is not too large [1]–[3]. Case A is applicable if \( l, \theta, \) and \( \phi \) are determined as \(^2\)

\[
L = \left\lfloor \frac{\omega_p}{(2\pi)} \right\rfloor, \quad \theta = \frac{2\pi}{L} - \frac{L\omega_p}{2}, \quad \phi = \frac{L\omega_p}{2} - 2\pi \tag{5a}
\]

and the resulting \( \theta \) and \( \phi \) satisfy \( 0 \leq \theta < \phi \leq \pi. \) Similarly, Case B can be used if \( l, \theta, \) and \( \phi \) are determined as \(^3\)

\[
L = \left\lfloor \frac{\omega_s}{(2\pi)} \right\rfloor, \quad \theta = 2\pi - \frac{L\omega_s}{2}, \quad \phi = 2\pi - \frac{L\omega_s}{2} \tag{5b}
\]

and the resulting \( \theta \) and \( \phi \) satisfy \( 0 \leq \theta < \phi \leq \pi. \) 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 number of multipliers, that is, \( N_F/2 + 1 + \left[ N_1/2 + 1 \right] + \left[ N_2/2 + 1 \right] \) if the coefficient symmetries are exploited. For both of the above approaches, a good estimate for \( N_F \) is the minimum order for the zero-phase frequency response of \( F(z) \) to stay within \( 1 \pm \delta_p (\pm \delta_s) \) on \([0, \theta) \cup [\phi, \pi]\). For the original approach, good estimates for \( N_1 \) and \( N_3 \) are the minimum orders for the zero-phase frequency responses of \( G_k(z) \) for \( k = 1, 2 \) to stay within the same limits in their passbands and stopbands. For the filters designed using the simultaneous optimization, good estimates for \( N_1 \) and \( N_2 \) are 60 percent of those for the earlier designs. Based on these observation, the values of \( L \) giving the lowest complexities for the above approaches can be found in the near vicinity of \( L_{opt} = 1/\sqrt{2(\omega_s - \omega_p)/\pi} \) and \( L_{opt} = 1/\sqrt{1.6(\omega_s - \omega_p)/\pi} \), respectively. The best results are usually obtained for those values of \( L \) where \( N_1 \) and \( N_2 \) are nearly equal.

3. DESIGN SCHEME FOR GENERALIZED FREQUENCY-RESPONSE MASKING APPROACH

Due to the fact that the amplitude responses for the masking filters \( G_1(z) \) and \( G_2(z) \) after the simultaneous optimization are very similar (see, e.g., Fig. 6 in [5]), it is beneficial to construct the masking filters using a common filter part \( G_3(z) \). In this case, the overall transfer function is given by (see Fig. 5)

\[
H(z) = \left[ F(z^L)G_1(z) + [z^{-LN_F/2} - F(z^L)]G_2(z) \right] G_3(z), \tag{6a}
\]

where \( F(z^L) \) and \( G_k(z) \) for \( k = 1, 2 \) are given by Eq. (1) and

\[
G_3(z) = \sum_{n=0}^{N_3} g_3(n) z^{-n} \tag{6b}
\]

For this generalized frequency-response masking approach, it has turned out that the solutions giving the lowest complexities are usually obtained when one of the subfilters \( G_1(z) \) or \( G_2(z) \) in Fig. 5 is selected to be a pure delay, implying that the order of remaining one is two times this delay. Based on this observation, the design algorithm for generating the periodic filter \( F(z^L) \) as well as the subfilters \( G_k(z) \) for \( k = 1, 2, 3 \) is carried out as follows:

Step 1: Design \( F(z^L) \), and \( G_k(z) \) for \( k = 1, 2 \) using the design scheme proposed in [5] and briefly described in Subsection 2.2.

---

\(^2\) \( \lceil x \rceil \) stands for the largest integer that is smaller than or equal to \( x \).

\(^3\) \( \lfloor x \rfloor \) stands for the smallest integer that is larger than or equal to \( x \).

---

4. NUMERICAL EXAMPLES

This section illustrates, by means of an example, the efficiency of the filters resulting when applying the proposed technique compared to those obtained using the earlier design schemes.

Consider the specifications [3]–[5]: \( \omega_p = 0.4\pi, \omega_s = 0.402\pi, \delta_p = 0.01, \) and \( \delta_s = 0.001. \) For the optimum conventional direct-form FIR filter design, the minimum order to meet the given criteria is 2541, requiring 2541 adders and 1271 multipliers when the coefficient symmetry is exploited.

For the original designs, \( L = 16 \) minimizes the number of multipliers required in the implementation [3], [4]. For \( L = 16, \) the overall filter is a Case A design with \( l = 3, \theta = 0.4\pi, \) and \( \phi = 0.432\pi. \) The minimum orders for \( G_1(z), G_2(z), \) and \( F(z) \) to meet the given specifications are \( N_1 = 70, N_2 = 98, \) and \( N_F = 162, \) respectively. The overall number of multipliers and adders for this design are 168 and 330, respectively, that are 13% of those required by an equivalent conventional direct-form design (1271 and 2541). The overall filter order is 2690 that is only 6% higher than that of the direct-form design (2541).

For \( L = 16, \) the best\(^1\) solution resulting when using the simultaneous synthesis scheme described in [5] is obtained by \( N_1 = 47, N_2 = 57, \) and \( N_F = 160. \) For this filter, the number of multipliers and adders are 134 and 264, respectively, that are approximately

\(^1\)According to the discussions of Subsection 2.1, this corresponds to the case where \( N_1, \) the order of \( G_1(z), \) is zero and the additional delay term is needed to make the delays of \( G_1(z) \) and \( G_2(z) \) equal.

\(^2\)The measure of goodness is the overall number of multipliers. If there exist several solutions requiring the same minimum number of multipliers, then, first, the solution with the minimum value of \( N_F \) is selected and, second, the one having a lower value for the maximum of \( N_1 \) and \( N_2 \) is selected. In this case, the overall filter order, as given by \( LN_F + \max\{N_1, N_2\}, \) is minimized.
80% of those of the original design. The overall filter order reduces to 2617.

For these filters, the overall number of multipliers is minimized by \( L = 21 \). This filter is a Case A design with \( l = 4 \), \( \theta = 0.4\pi \), and \( \phi = 0.4/2\pi \). The best solution is obtained by \( N_1 = 55 \), \( N_2 = 77 \), and \( N_3 = 122 \). This filter requires 129 multipliers and 254 adders that are approximately 77% of those of the original best design for \( L = 16 \).

For the generalized frequency-response masking approach with simultaneously optimizing the subfilters, the overall number of multipliers is minimized also by \( L = 21 \). The best solution is obtained by \( N_1 = 0 \), \( G(z) = z^{-N_1/2} = z^{-23} \), \( N_2 = 46 \), \( N_3 = 55 \), and \( N_P = 122 \). This filter requires 114 multipliers and 223 adders that are approximately 88% of those obtained by simultaneously optimizing the masking filters without the common part. Some of the characteristics for the filters designed using various algorithms are summarized in Table 1. Here, \( N_M \) denotes the number of multipliers required to implement the overall filter.

For the best design with \( L = 21 \), Figs. 6 and 7 show the responses for \( F(L\omega) \) and \( 1 - F(L\omega) \) as well as \( G_2(\omega) \) and \( G_3(\omega) \), respectively. The zero-plots for the masking filters are shown in Fig. 8. The overall response \( H(\omega) \) is depicted in Fig. 9. Many interesting observation can be made from these figures. For instance, \( F(L\omega) [1 - F(L\omega)] \) varies approximately between \(-0.5 \) and \( 2.5 \) \([\pm 1.5 \) and \( 1.5] \).

### REFERENCES


---

If the peak scaling and two’s complement arithmetic are desired to be used and \( G_1(z) \) and \( G_2(z) \) share the delays by using the transposed direct-form approach, then exploiting the coefficient symmetry in Fig. 1, there exist two alternatives for the scaling. In the first alternative, the overall input is divided by a constant \( \beta \) being the maximum value of \( F(L\omega) \) and the coefficients of \( G_1(z) \) and \( G_2(z) \) are multiplied by \( \beta \). In the second alternative, the coefficients \( f(n) \) in Fig. 1 as well as the output of the delay line \( z^{-LNP/2} \) are divided by \( \beta \).

---

Table 1 Summary of Filter Designs in the Example under Consideration

<table>
<thead>
<tr>
<th>Method</th>
<th>( L )</th>
<th>( N_F )</th>
<th>( N_1 )</th>
<th>( N_2 )</th>
<th>( N_3 )</th>
<th>( N_M )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional direct-form filter</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>1271</td>
</tr>
<tr>
<td>Original [3], [4]</td>
<td>16</td>
<td>162</td>
<td>70</td>
<td>98</td>
<td>–</td>
<td>168</td>
</tr>
<tr>
<td>Proposed</td>
<td>21</td>
<td>122</td>
<td>0</td>
<td>46</td>
<td>55</td>
<td>114</td>
</tr>
</tbody>
</table>

---

**Fig. 6** Responses for \( F(L\omega) \) (solid line) and \( 1 - F(L\omega) \) (dot-dashed line) for the best proposed filter for \( L = 21 \).

**Fig. 7** Responses for \( G_2(\omega) \) (solid line) and \( G_3(\omega) \) (dashed line). \( G_1(z) = 1 \) since \( G_1(z) = z^{-23} \).

**Fig. 8** Zero-plots for the masking filters. Circles and crosses denote the roots of the \( G_2(z) \) and \( G_3(z) \), respectively.

**Fig. 9** Response for the best proposed overall filter for \( L = 21 \).