OPTIMIZATION OF FIR FILTERS USING THE FREQUENCY-RESPONSE MASKING APPROACH

Tapio Saramäki *and* *Håkan Johansson*

Signal Processing Laboratory, P. O. Box 553
Tampere University of Technology
FIN-33101 Tampere, Finland
e-mail: ts@cs.tut.fi

Department of Electrical Engineering
Linköping University
SE-58183 Linköping, Sweden
e-mail: hakanj@isy.liu.se

ABSTRACT

A very efficient technique for drastically reducing the number of multipliers and adders in implementing linear-phase finite impulse response filters is to use the frequency-response masking approach. A drawback in the synthesis techniques proposed up to now is that the subfilters in the overall implementation are designed separately. In order to further reduce the arithmetic complexity of these filters, a two-step optimization technique for simultaneously optimizing the subfilter is introduced. In the first step, a good suboptimal solution is found using a simple iterative algorithm. In the second step, this solution is then used as a start-up solution for further optimization carried out by using the second algorithm of Dutta and Vidyasagar. An example taken from the literature is included illustrating that the number of adders and multipliers for the resulting filters are less than 80 percent compared with the earlier ones.

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 to the significant reduction in the computational complexity is a slight increase in the filter order.

A drawback of the existing synthesis techniques is that the subfilters have been designed separately. This paper introduces a two-step approach for simultaneously designing the subfilters. In the first step, a simple iterative design scheme is used to generate a start-up solution for further optimization. In the second step, this is accomplished by using the second algorithm of Dutta and Vidyasagar [5]. It is 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 filters obtained by using the existing design schemes.

2. FREQUENCY-RESPONSE MASKING APPROACH

This section reviews how to use the frequency-response masking approach for synthesizing linear-phase FIR filters.

2.1. Filter Structure and Frequency Response

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

$$H(z) = F(z^L)G_1(z) + [z^{-LN_F/2} - F(z^L)]G_2(z), \qquad (1a)$$

This work was supported by the Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program (2000–2005)).

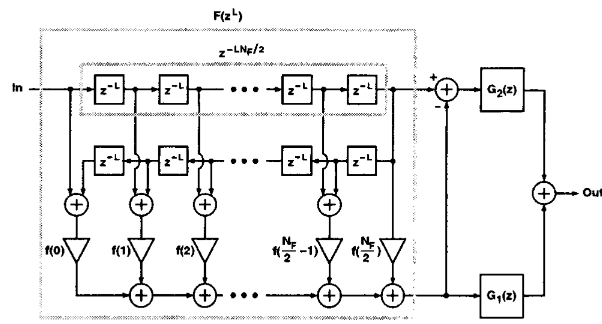

Fig. 1 *An efficient implementation for a filter synthesized using the frequency-response masking approach. In this implementation, $G_1(z)$ and $G_2(z)$ can share their delays if a transposed direct-form implementation (exploiting the coefficient symmetry) is used.*

where

$$F(z^L) = \sum_{n=0}^{N_F} f(n)z^{-nL}, \qquad (1b)$$

$$G_1(z) = z^{-M_1}\sum_{n=0}^{N_1} g_1(n)z^{-n}, \;\; G_2(z) = z^{-M_2}\sum_{n=0}^{N_2} g_2(n)z^{-n}. \qquad (1c)$$

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

$$H(\omega) = H_1(\omega) + H_2(\omega), \qquad (2a)$$

where

$$H_1(\omega) = F(L\omega)G_1(\omega), \quad H_2(\omega) = [1 - F(L\omega)]G_2(\omega) \quad (2b)$$

with

$$F(\omega) = f(N_F/2) + 2\sum_{n=1}^{N_F/2} f(N_F/2 - n)\cos n\omega \qquad (2c)$$

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

0-7803-6685-9/01/$10.00©2001 IEEE

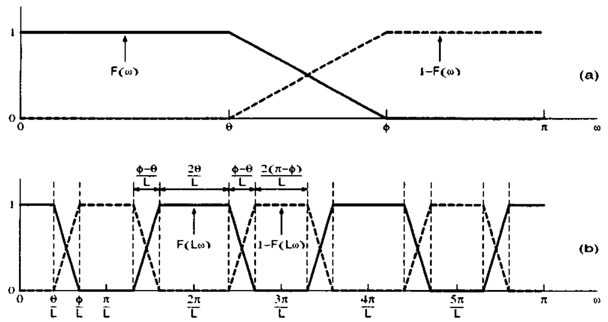

Fig. 2 *Generation of a complementary periodic filter pair by starting with a lowpass-highpass complementary pair. (a) Prototype filter responses $F(\omega)$ and $1 - F(\omega)$. (b) Periodic responses $F(L\omega)$ and $1 - F(L\omega)$ for $L = 6$.*

2.2. Efficiency of the Use of the Transfer Function $F(z^L)$

The efficiency of $H(z)$ as given by Eq. (1) lies in the fact that of the pair of transfer functions $F(z^L)$ and $z^{-LN_F/2} - F(z^L)$ can be generated from the pair of *prototype* transfer functions $F(z) = \sum_{n=0}^{N_F} f(n)z^{-n}$ and $z^{-N_F/2} - F(z)$ by replacing z^{-1} by z^{-L}. This increases the filter orders to LN_F, but since only every Lth impulse response value is nonzero, the number of adders and multipliers remain the same. The above prototype pair forms a *complementary* filter pair since their zero-phase frequency responses, $F(\omega)$ and $1 - F(\omega)$ with $F(\omega)$ given by Eq. (2c), add up to unity. Figure 2(a) illustrates the relations between these responses in the case of a lowpass-highpass filter pair with edges at θ and ϕ.

As illustrated in Fig. 2(b), the substitution $z^{-L} \mapsto z^{-1}$ preserves the complementary property resulting in the *periodic* responses $F(L\omega)$ and $1 - F(L\omega)$, which are frequency-axis compressed versions of the prototype responses such that the interval $[0,\ L\pi]$ is shrunk onto $[0,\ \pi]$. Since the periodicity of the prototype responses is 2π, the periodicity of the resulting responses is $2\pi/L$ and they contain several passband and stopband regions in the interval $[0,\ \pi]$.

For a lowpass filter $H(z)$, 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)

$$\omega_p = (2l\pi + \theta)/L,\ \ \omega_s = (2l\pi + \phi)/L, \qquad (3)$$

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

$$\omega_p = (2l\pi - \phi)/L,\ \ \omega_s = (2l\pi - \theta)/L. \qquad (4)$$

The widths of these transition bands are $(\phi - \theta)/L$, which is only $1/L$-th of that of the prototype filters implying that the arithmetic complexity of the periodic transfer functions to provide one of the transition bands is only $1/L$-th of that of a conventional nonperiodic filter[1].

2.3. Use of the Masking Filters $G_1(z)$ and $G_2(z)$ in Earlier Synthesis Schemes

The role of the *masking filters* $G_1(z)$ and $G_2(z)$ is two-fold. First, they are designed in such a manner that in the passband of the overall filter, the subresponses $H_1(\omega)$ and $H_2(\omega)$ as given by Eq. (2b) approximate $F(L\omega)$ and $1 - F(L\omega)$, respectively. Hence, $H(\omega) = H_1(\omega) + H_2(\omega)$ approximates unity, as is desired. In the stopband, these filters attenuate the extra unwanted passband and transition band regions of $F(L\omega)$ and $1 - F(L\omega)$. These goals are achieved in Cases A and B by selecting the passband and stopband edges of the two lowpass masking filter to be located as shown in Figs. 3 and 4, respectively.

[1]Recall that the order of a linear-phase FIR filter is roughly inversely proportional to the transition bandwidth. Note that the orders of both the periodic filters and the corresponding nonperiodic filters are approximately the same, but the conventional filter does not contain zero-valued impulse-response coefficients.

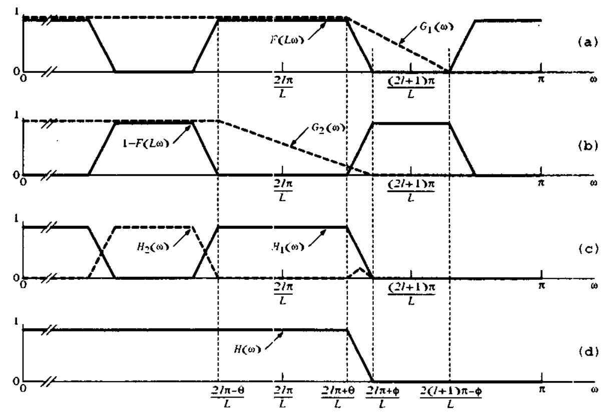

Fig. 3 *Case A design of a lowpass filter using the frequency-response masking technique.*

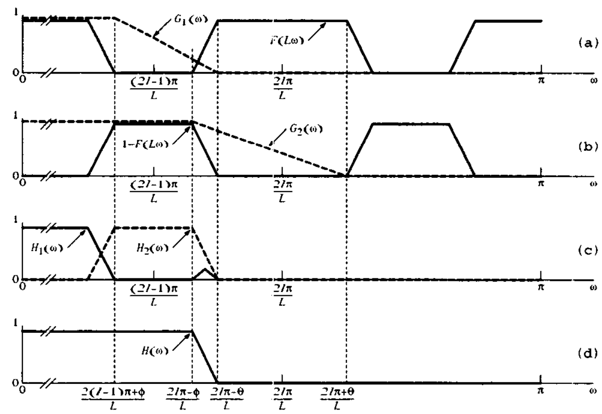

Fig. 4 *Case B design of a lowpass filter using the frequency-response masking technique.*

2.4. Existing Filter Design Technique

Based on the observations made in [1], the existing design of $H(z)$ 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:

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 $0.9\delta_p$ and zero in their stopbands with tolerance less than or equal to $0.9\delta_s$.

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

$$\Omega_p^{(F)} = \begin{cases} [(2l\pi - \theta)/L,\ (2l\pi + \theta)/L] & \text{for Case A} \\ [[2(l-1)\pi + \phi]/L,\ (2l\pi - \phi)/L] & \text{for Case B} \end{cases}$$

and approximates zero with tolerance δ_s on

$$\Omega_s^{(F)} = \begin{cases} [(2l\pi + \phi)/L,\ [2(l+1)\pi - \phi]/L] & \text{for Case A} \\ [(2l\pi - \theta)/L,\ (2l\pi + \theta)/L] & \text{for Case B.} \end{cases}$$

3. PROPOSED TWO-STEP DESIGN SCHEME

This section describes the proposed two-step algorithm for designing filters with a reduced arithmetic complexity. How to select the filter orders and the design parameters l, L, θ and ϕ will be considered in the next section.

3.1. Algorithm For Finding an Initial Filter

Assuming that the filter orders and other design parameters have been predetermined, an initial solution can be found effectively using the following procedure:

Step 1: Set $r = 1$, $\epsilon_G^{(r)} = \epsilon_F^{(r)} = 0$. Determine the parameters of $F^{(r)}(z)$ to minimize

$$\max_{\omega \in [0,\ \theta] \cup [\phi,\ \pi]} |W(\omega)[F^{(r)}(\omega) - D(\omega)]|.$$

Here, the first band is the passband, where $W(\omega)$ and $D(\omega)$ are equal to unity. The second band is the stopband, where $W(\omega)$ is equal to δ_p/δ_s and $D(\omega)$ is equal to zero. In the sequel, the same desired and weighting functions are used.

Step 2: Set $r = r + 1$. Determine the parameters of $G_k^{(r)}(z)$ for $k = 1, 2$ to minimize

$$\epsilon_G^{(r)} = \max_{\omega \in [0,\ \Omega_{p1}] \cup [\Omega_{s2},\ \pi]} |W(\omega)[H_G^{(r)}(\omega) - D(\omega)]|,$$

where

$$H_G^{(r)}(\omega) = F^{(r-1)}(L\omega)G_1^{(r)}(\omega) + [1 - F^{(r-1)}(L\omega)]G_2^{(r)}(\omega),$$

$\Omega_{p1} = 2l\pi/L$ and $\Omega_{s2} = (2l + 1)\pi/L$ for Case A designs, and $\Omega_{p1} = (2l - 1)\pi/L$ and $\Omega_{s2} = 2l\pi/L$ for Case B designs (see Figs. 3 and 4).

Step 3: Determine the parameters of $F^{(r)}(z)$ to minimize

$$\epsilon_F^{(r)} = \max_{\omega \in [\Omega_{p1},\ \Omega_{p2}] \cup [\Omega_{s1},\ \Omega_{s2}]} |W(\omega)[H_F^{(r)}(\omega) - D(\omega)]|,$$

where

$$H_F^{(r)}(\omega) = F^{(r)}(L\omega)G_1^{(r)}(\omega) + [1 - F^{(r)}(L\omega)]G_2^{(r)}(\omega),$$

$\Omega_{p2} = (2l\pi + \theta)/L$ and $\Omega_{s1} = (2l\pi + \phi)/L$ for Case A designs, and $\Omega_{p2} = (2l\pi - \phi)/L$ and $\Omega_{s1} = (2l\pi - \theta)/L$ for Case B designs (see Figs. 3 and 4).

Step 4: If $|\epsilon_G^{(r)} - \epsilon_G^{(r-1)}| \leq \Delta$ and $|\epsilon_F^{(r)} - \epsilon_F^{(r-1)}| \leq \Delta$, where Δ is a prescribed tolerance, then stop. Otherwise, go to Step 2.

Steps 1 and 3 can be accomplished very fast by using the Remez algorithm [6], whereas Step 2 can be implemented using linear programming. The basic idea in the above algorithm is to share the frequency-response-shaping responsibilities in such a manner that $G_1(z)$ and $G_2(z)$ concentrate mainly on generating the desired response on $[0,\ \Omega_{p1}]$ and $[\Omega_{s2},\ \pi]$, whereas $F(z^L)$ concentrates on the regions $[\Omega_{p1},\ \Omega_{p2}]$ and $[\Omega_{s1},\ \Omega_{s2}]$ (see Figs. 3 and 4).

3.2. Further Optimization

The solution obtained using the above algorithm can be further improved by using it as a start-up solution for the second algorithm of Dutta and Vidyasagar [5]. For details on how to apply this algorithm see, e.g., [7].

4. PRACTICAL FILTER SYNTHESIS

In practice, ω_p and ω_s are given and l, L, θ, and ϕ 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, θ, and ϕ are determined as [2]

$$l = \lfloor L\omega_p/(2\pi) \rfloor,\ \theta = L\omega_p - 2l\pi,\ \phi = L\omega_s - 2l\pi \qquad (5a)$$

and the resulting θ and ϕ satisfy $0 \leq \theta < \phi \leq \pi$. Similarly, Case B can be used if l, θ, and ϕ are determined as [3]

$$l = \lceil L\omega_s/(2\pi) \rceil,\ \theta = 2l\pi - L\omega_s,\ \phi = 2l\pi - L\omega_p \qquad (5b)$$

and the resulting θ and ϕ 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, $N_F/2+1+\lfloor(N_1+2)/2\rfloor+\lfloor(N_2+2)/2\rfloor$ if the coefficient symmetries are exploited. For both the earlier and the present designs, 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 $[0,\ \theta]$ ($[\phi,\ \pi]$). For earlier designs, good estimates for N_1 and N_2 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 new technique, good estimates for N_1 and N_2 are 60 percent of those for the earlier designs. Based on this observation, it can be shown that the values of L giving the lowest complexities can be found in the near vicinity of

$$L_{\text{opt}} = 1/\sqrt{1.6(\omega_s - \omega_p)/\pi}\,. \qquad (6)$$

The best results are usually obtained for those values of L where N_1 and N_2 are nearly equal.

5. NUMERICAL EXAMPLES

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

Consider the specifications [3], [4]: $\omega_p = 0.4\pi$, $\omega_s = 0.402\pi$, $\delta_p = 0.01$, 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 earlier 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[4] solution resulting when using the proposed synthesis scheme is obtained by $N_1 = 47$, $N_2 = 57$, and

[2] $\lfloor x \rfloor$ stands for the largest integer that is smaller than or equal to x.

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

[4] 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.

$N_F = 160$. For this filter, the number of multipliers and adders are 134 and 264, respectively, that are approximately 80% of those of the earlier design. The overall filter order reduces to 2617.

For the proposed filters, the overall number of multipliers is minimized by $L = 21$ [the value obtained using Eq. (6)]. This filters is a Case A design with $l = 4$, $\theta = 0.4\pi$, and $\phi = 0.442\pi$. The best solution is obtained by $N_1 = 55$, $N_2 = 77$, and $N_F = 122$. This filter requires 129 multipliers and 254 adders that are approximately 77% of those of the earlier best design for $L = 16$.

For the best design with $L = 21$, Figs. 5 and 6 show the responses $F(L\omega)$ and $1 - F(L\omega)$ and $G_1(\omega)$ and $G_2(\omega)$, respectively. The responses $H_1(\omega) = F(L\omega)G_1(\omega)$ and $H_2(\omega) = [1 - F(L\omega)]G_2(\omega)$ are shown in Fig. 7, whereas the overall response $H(\omega) = H_1(\omega) + H_2(\omega)$ is depicted in Fig. 8. Two interesting observations can be made from these figures. First, $F(L\omega)$ and $F(L\omega)G_1(\omega)$ in the passband region [$1 - F(L\omega)$ and $[1 - F(L\omega)]G_2(\omega)$ in the passband region] varies approximately between -2 and 4 [-3 and 3].[5] Second, the responses $G_1(\omega)$ and $G_2(\omega)$ are very similar.

6. CONCLUSION

The arithmetic complexity of FIR filters using the frequency-response masking approach has been reduced by simultaneously optimizing the subfilters. Future work is devoted to characterizing in more details the behavior of the resulting filters as well as to applying a similar technique for the multistage frequency-response masking approach.

REFERENCES

[1] Y. C. Lim, "Frequency-response masking approach for the synthesis of sharp linear phase digital filters," *IEEE Trans. Circuits and Syst.*, vol. CAS-33, pp. 357–364, Apr. 1986.

[2] Y. C. Lim and Y. Lian, "The optimum design of one- and two-dimensional FIR filters using the frequency response masking technique," *IEEE Trans. Circuits and Syst. II*, vol. CAS-33, pp. 357–364, Feb. 1993.

[3] T. Saramäki, "Design of computationally efficient FIR filters using periodic subfilters as building blocks" in *The Circuits and Filters Handbook*, edited by W.-K. Chen, CRC Press, Inc., 1995, pp. 2578–2601.

[4] T. Saramäki and Y. C. Lim, "Use of the Remez algorithm for designing FIR filters utilizing the frequency-response masking approach," in *Proc. IEEE Int. Symp. Circuits Syst.*, Orlando, FL, May 1999, pp. 449–455.

[5] S. R. K. Dutta and M. Vidyasagar, "New algorithms for constrained minimax optimization," *Math. Programming*, vol 13., pp. 140–155, 1977.

[6] 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.*, vol. AU-21, pp. 506–526, Dec. 1973.

[7] J. Yli-Kaakinen and T. Saramäki, "Design of very low-sensitivity and low-noise recursive digital filters using a cascade of low-order lattice wave digital filters," *IEEE Trans.Circuits Syst. II*, vol. 46, pp. 906–914, July 1999.

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

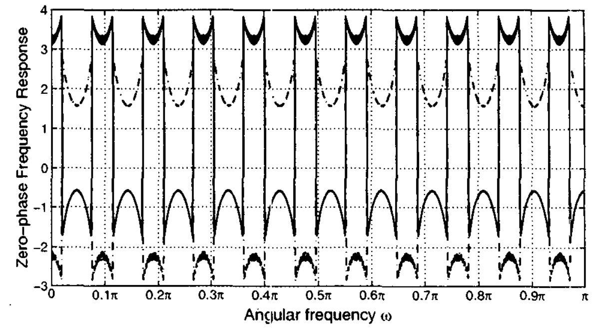

Fig. 5 *Responses for $F(L\omega)$ (solid line) and $1 - F(L\omega)$ (dot-dashed line) for the best proposed filter for $L = 21$.*

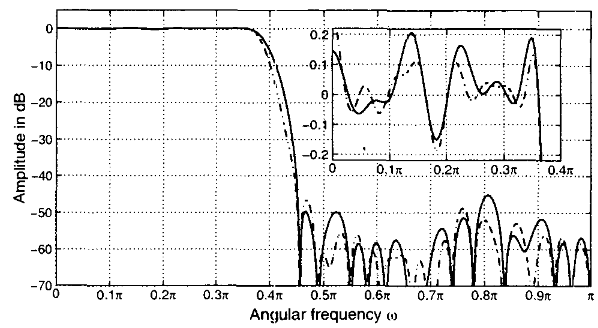

Fig. 6 *Responses for $G_1(\omega)$ (solid line) and $G_2(\omega)$ (dot-dashed line) for the best proposed filter for $L = 21$.*

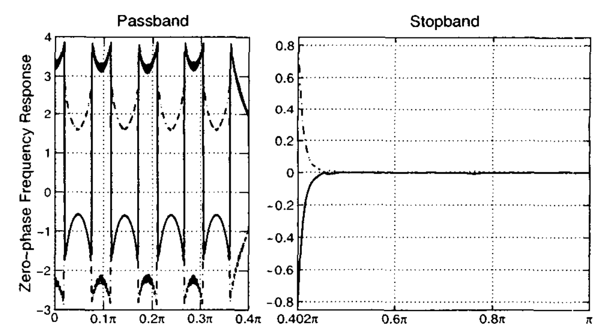

Fig. 7 *Responses for $F(L\omega)G_1(\omega)$ (solid line) and $[1 - F(L\omega)]G_2(\omega)$ (dot-dashed line) for the best proposed filter for $L = 21$.*

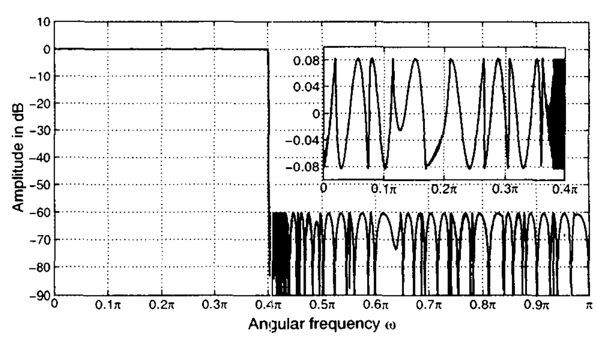

Fig. 8 *Response for the best proposed overall filter for $L = 21$.*