APPLICATION OF DFT FILTER BANKS AND COSINE MODULATED FILTER BANKS IN FILTERING*

Yuan-Pei Lin and P. P. Vaidyanathan

Department of Electrical Engineering, California Institute of Technology

I. Introduction

The M channel maximally decimated filter bank shown in Fig. 1.1 has been studied extensively, (see references in [1]). When the system in Fig. 1.1 is alias free, it is an LTI system with transfer function $T(z)$, as indicated in Fig. 1.1. $T(z)$ will be called the overall response.

A maximally decimated filter bank is well-known for its application in subband coding. Another application of maximally decimated filter banks is in block filtering, [2]. Convolution through block filtering has the advantages that parallelism is increased and data is processed at a lower rate. However, the computational complexity is comparable to that in direct convolution. In [3], filter banks are used to map long convolutions into smaller ones in the subbands. Computations are then performed in parallel at a lower rate.

More recently [4], another type of filter bank convolver has been developed. In this scheme the convolution is performed in the subbands and for a fixed rate the result of convolution is more accurate than direct convolution. This type of filter bank convolver also enjoys the advantages of block filtering, parallelism and lower working rate. Nevertheless, like block filtering, there is no computational saving.

In this paper, we introduce the new under-decimated system. A filter bank is said to be under-decimated if the number of channels is more than the decimation ratio in the subbands. Two types of low-complexity filter banks can be used for the new system, the DFT filter bank and cosine modulated filter bank [5].

Fig. 1.2 shows the setup of the under-decimated system; it has $2M$ channels but is decimated only by M. In both DFT filter bank case and cosine modulated filter bank case, the system is approximately alias free and overall response is equivalent to a tunable multilevel filter. Properties of the DFT filter banks and the cosine modulated filter banks can be exploited to simultaneously achieve parallelism, computational saving and lower working rate. Furthermore, in both filter banks the implementation cost of the analysis bank is comparable to that of one prototype filter plus some low complexity matrices. The individual analysis and synthesis filters have complex coefficients in the DFT filter bank but have real coefficients in the cosine modulated filter bank.

*The work described in the paper was carried out by Jet Propulsion Laboratory and California Institute of Technology under a contract with National Aeronautics and Space Administration.

1.1 Notation and definition

1. Boldfaced lower case letters are used to represent vectors and boldfaced upper case letters are used to represent matrices. The notations $\mathbf{A}^T$, $\mathbf{A}^*$ and $\mathbf{A}^\dagger$ represent the transpose, conjugate and transpose-conjugate of $\mathbf{A}$, respectively. We define the 'tilde' notation as follows: $\widetilde{\mathbf{A}}(z) = \mathbf{A}^\dagger(1/z^*)$.
2. A filter $H(z)$ with impulse response $h(n)$ is called a Nyquist(N) filter if one of the following two applies. (1) $\sum_{k=0}^{N-1} H(e^{j(\omega-2\pi k/N)}) = c$ for some constant c. (2) $h(n)_{\downarrow N}$ is nonzero only when $n = 0$. The notation $h(n)_{\downarrow N}$ denotes the N fold decimated version of $h(n)$.

II. DFT Filter Banks

The system in Fig. 1.2 is called a DFT filter bank if the analysis filters are shifted versions of a prototype $P(z)$ on the unit circle. Similarly for the synthesis bank. For simplicity, let the synthesis prototype $Q(z) = \widetilde{P}(z)$. As we will only consider FIR filters, some delays can be added to make the individual filters causal. The analysis filters and synthesis filters have the following form.

$$H_k(z) = a_k P(zW^k), F_k(z) = \widetilde{P}(zW^k), W = e^{\frac{j\pi}{M}}. \quad (2.1)$$

On the unit circle $H_k(z)$ or $F_k(z)$ is just a shift of $P(z)$ or $\widetilde{P}(z)$ by $k\pi/M$ except a scalar. Fig 2.1 shows the support of the synthesis filters. The analysis filters are scaled and time-reversed versions of the synthesis filters exept some delays.

We now show that with proper design of $P(z)$, this DFT filter bank is approximately alias free and the overall response is equivalent to a tunable multilevel filter. Moreover, the overall response can be a real-coefficient linear-phase filter as desired. Efficient implementation of the DFT filter bank will also be discussed.

2.1 Suppression of aliasing error

Consider the under-decimated system in Fig. 1.2, a $2M$ channel filter bank decimated by M. The suppression of

0-7803-2440-4/94/ $4.00

aliasing error due to downsampling in the subbands can be explained pictorially. Take the first subband as an example. Because of decimation followed by expansion, there will be $M-1$ image copies of $H_0(z)$, as shown in Fig. 2.2. If $P(z)$ has stopband edge less than π/M, these image copies will be suppressed to the level of the stopband attenuation of $P(z)$. The reasoning of aliasing suppression in the other subbands follows.

We now present the mathematical counterpart of the above discussion. The output, $\widehat{X}(z)$, is related to the input, $X(z)$, by

$$\widehat{X}(z) = \sum_{i=0}^{M-1} A_i(z)X(zW^{2i}). \tag{2.2}$$

The alias transfer function, $A_i(z)$, is defined as

$$A_i(z) = \frac{1}{M}\sum_{k=0}^{2M-1} H_k(zW^{2i})F_k(z). \tag{2.3}$$

The under-decimated system is alias free if $A_i(z) = 0$, for $i = 1, 2, \ldots, M-1$.

With analysis filters and synthesis filters chosen as in (2.1), the alias transfer function is

$$A_i(z) = \frac{1}{M}\sum_{k=0}^{2M-1} a_k P(zW^{2i+k})\widetilde{P}(zW^k), \tag{2.4}$$

Assume

$$P(zW^{2i})\widetilde{P}(z) \approx 0, \quad i = 1, \ldots, M-1. \tag{2.5}$$

This assumption is reasonable if $P(z)$ has stopband edge less than π/M and large enough stopband attenuation. In this case $A_i(z) \approx 0$ and the DFT filter bank is almost alias free. Notice that the degree of alias suppression improves with the stopband attenuation of $P(z)$.

2.2 The overall response of the DFT filter bank

For a $2M$ channel system decimated by M as shown in Fig. 1.2, the overall response $T(z)$ [1] is

$$T(z) = \frac{1}{M}\sum_{k=0}^{2M-1} H_k(z)F_k(z). \tag{2.6}$$

Substitute the expression of $H_k(z)$ and $F_k(z)$ in (2.1), then

$$T(e^{j\omega}) = \frac{1}{M}\sum_{k=0}^{2M-1} a_k |P(e^{j(\omega - k\pi/M)})|^2. \tag{2.7}$$

When $|P(e^{j\omega})|^2$ is a Nyquist($2M$) filter, it can be shown that the addition of $a_k|P(e^{j(\omega-k\pi/M)})|^2$ in (2.7) will not result in any bumps or dips in the magnitude response of $T(z)$. The definition of a Nyquist filter is given in Sec. I. Detailed explanation can be found in [1]. With (2.7), we can plot a typical magnitude response of $T(z)$ as in Fig. 2.3, which shows that the overall response is equivalent to a multilevel filter. Since the value of a_k can be chosen freely, $T(z)$ is actually a tunable multilevel filter.

Furthermore, if $P(z)$ has real coefficients and we choose $a_k = a_{2M-k}, k = 1, 2, \ldots, M$, it can be verified that the resulting $T(z)$ is a linear-phase real-coefficient filter.

2.3 Implementation of the DFT filter banks

There exist efficient implementations for the DFT filter bank. To see this, express $P(z)$ as

$$P(z) = \sum_{n=0}^{2M-1} E_n(z^{2M})z^{-n}, \tag{2.8}$$

where $E_n(z)$ is the nth type 1 polyphase component of $P(z)$ [1]. The analysis filters can be rewritten as

$$H_k(z) = a_k \sum_{n=0}^{2M-1} E_n(z^{2M})W^{-kn}z^{-n}, 0 \le k < 2M. \tag{2.9}$$

Let

$$\mathbf{h}(z) = [H_0(z)\ H_1(z) \ldots\ H_{2M-1}(z)]^T. \tag{2.10}$$

The vector $\mathbf{h}(z)$ takes the form

$$\mathbf{h}(z) = \mathbf{A}\mathbf{W}^*\mathbf{E}(z)\mathbf{e}(z), \tag{2.11}$$

where $\mathbf{e}(z) = [1\ \ z^{-1} \ldots z^{-(2M-1)}]^T$, $\mathbf{A}$ and $\mathbf{E}$ are diagonal matrices with $\mathbf{A}_{kk} = a_k$ and $[\mathbf{E}(z)]_{kk} = E_k(z^{2M})$. The DFT matrix $\mathbf{W}$ has entries $\mathbf{W}_{kn} = W^{kn}$, $0 \le k, n < 2M$. Eq. (2.11) renders a picture of polyphase implementation for the analysis bank, Fig. 2.4. The complexity is that of the prototype filter $P(z)$ plus a DFT matrix and some scalars. Similarly for the synthesis bank. Notice that all the computations involved in the filter bank are performed after the M-fold decimators; parallelism, lower rate and lower complexity are achieved at the same time.

III. Cosine Modulated Filter Banks

In the DFT filter bank, the analysis and synthesis filters have complex coefficients. The new under-decimated cosine modulated filter bank comes into play if real coefficients are desired.

A filter bank, Fig. 1.1 or Fig. 1.2, is said to be cosine modulated if all analysis and synthesis filters are generated by cosine or sine modulation of one or two prototype filters. In this section we introduce the new under-decimated cosine modulated filter bank. As in Fig. 1.2, the new system has $2M$ channels but is decimated by M. Individual analysis and synthesis filters have real coefficients. Aliasing is suppressed to the level of the stopband

attenuation of the prototypes and the overall response is a linear-phase tunable multilevel filter. Furthermore, there exists efficient implementation of this cosine modulated filter bank. The complexity of the analysis bank is that of the prototype filter plus two DCT matrices (Appendix A) and some scalars. The complexity of an $M \times M$ DCT matrix is of order only $M \log(M)$ [6]. Complexity for the synthesis bank is similar.

3.1 Construction of the new system from the consideration of alias cancellation and suppression

In the cosine modulated filter bank, all analysis and synthesis filters have real coefficients. Each filter has positive and negative spectral occupancy as opposed to single-sided spectral support in the DFT filter bank. This incurs a problem that we do not have in the DFT filter bank. More details and a proposed solution of this new problem are given below.

Let $P(z)$ be the prototype of the analysis filters. As in DFT filter bank, we take the synthesis prototype $Q(z) = \widetilde{P}(z)$ for simplicity. Let $P_k(z) = P(zW^{k+0.5}), k = 0, 1, \ldots, 2M-1$. We can combine $P_k(z)$ and $P_{2M-1-k}(z)$, to get real-coefficient filters. For instance,

$$\begin{aligned} F_k(z) &= \widetilde{P}_k(z) + \widetilde{P}_{2M-1-k}(z), \\ H_k(z) &= a_k \widetilde{F}_k(z), \end{aligned} \quad k = 0, 1, \ldots, M-1. \quad (3.1a)$$

Fig. 3.1 shows the spectral supports of the synthesis filters. The analysis filters are scaled and time-reversed versions of corresponding synthesis filters except some delays.

Aliasing error created in the subbands is now very different from that in the DFT filter bank because each filter consists of two shifts of $P(z)$ or $\widetilde{P}(z)$. To illustrate this situation, consider the kth subband. Due to decimation followed by expansion, $P_k(z)$ has $M-1$ image copies and $P_{2M-1-k}(z)$ also has $M-1$ image copies. Since every filter has real coefficients, inspecting the overlapping between images of $P_k(z)$ and $\widetilde{P}_{2M-1-k}(z)$ is sufficient. Referring to Fig. 3.2, of the $M-1$ image copies of $P_k(z)$, $M-2$ of them fall into the stopband of $\widetilde{P}_{2M-1-k}(z)$ provided that $P(z)$ has stopband edge less than π/M. If $P(z)$ has large enough stopband attenuation, these images will be suppressed by $\widetilde{P}_{2M-1-k}(z)$. However, one of the image copies of $P_k(z)$ will overlap with the spectral support of $\widetilde{P}_{2M-1-k}(z)$. In particular, one image copy of $P_k(z)$ will overlap with $\widetilde{P}_{2M-1-k}(z)$ from the right when k is even and from the left when k is odd. Fig. 3.2 shows the case when k is even. This type of aliasing error can not be suppressed in the synthesis bank.

Our solution to this problem is to introduce a second subsystem that has similar aliasing error to cancel the existing one. For this particular purpose, the filters of the second subsystem are required to have a stacking similar to that of the first subsystem. Let the second subsystem have analysis filters $H_k(z)$ and synthesis filters $F_k(z)$, $k = M, M+1, \ldots, 2M-1$. We have found that with the following choice, the negative of the alias component from the first subsystem will occur in the second subsystem.

$$\begin{aligned} F_k(z) &= j\widetilde{P}_k(z) - j\widetilde{P}_{2M-1-k}(z), \\ H_k(z) &= a_k \widetilde{F}_k(z), \end{aligned} \quad M \le k < 2M. \quad (3.1b)$$

Specifically, the $k+M$ subband creates the negative of the aliasing error from the kth subband. Through this construction the spectral supports of filters in the second subsystem resemble that of filters in the first subsystem.

We now verify that indeed the alias transfer function $A_i(z) \approx 0$. With filters constructed as in (3.1) and the expression of alias transfer functions in (2.3), we have

$$A_i(z) = \frac{2}{M} \sum_{k=0}^{M-1} a_k (P_{k+2i}\widetilde{P}_k + P_{2M-1-k+2i}\widetilde{P}_{2M-1-k}).$$

If $P(z)$ satisfy (2.5), then $A_i(z) \approx 0$.

3.2 The overall response $T(z)$

Using (3.1) and (2.6), the overall response is

$$T(e^{j\omega}) = \frac{2}{M} \sum_{k=0}^{2M-1} a_k |P_k(e^{j\omega})|^2. \quad (3.2)$$

The above expression for the overall response is similar to that in the DFT filter bank, (2.7). If $|P(e^{j\omega})|^2$ is a Nyquist($2M$) filter, $T(z)$ is a tunable multilevel filter as in the DFT filter bank. Also from (3.2), we know the overall response has linear phase.

Summarizing, we have shown that with filters constructed in (3.1), the system in Fig. 1.2 is a cosine modulated filter bank with approximate alias cancellation and the overall response is a linear-phase tunable multilevel filter. As shown in Appendix A, complexity of the analysis bank is that of the prototype filter $P(z)$ plus two DCT matrices and some scalars.

Alternative stacking of filter responses

In addition to the configuration of filters shown in Fig. 3.1, a different stacking can also be applied. Fig. 3.3 shows this alternative. The spectral supports of the second set of synthesis filters are different from the spectral supports of the first set of synthesis filters. Again the analysis filters are scaled and time-reversed versions of synthesis filters except some delays. In this case, the scheme of alias cancellation stills works and the argument that the overall response is a tunable multilevel filter continues to hold after minor adjustments.

Remark. In the discussion of DFT and cosine modulated filter banks, we have chosen the synthesis prototype $Q(z) = \widetilde{P}(z)$ for simplicity. It can be shown that both

systems still function like tunable multilevel filters with approximate alias suppression if (1) $P(zW^2)Q(z) \approx 0$ and (2) $P(z)Q(z)$ is Nyquist($2M$).

IV. Design techniques

For the design of the prototypes, we have found two simple methods.

Shortcut design: The constraints for the prototypes can be easily satisfied through the following two steps. (1) Design a Nyquist($2M$) $P(z)$ with stopband edge less than π/M. (2) Design $Q(z)$ such that the passband of $Q(z)$ covers the passband and transition bands of $P(z)$ but $P(zW^2)$ falls into the stopband of $Q(z)$, Fig. 4.1. As $Q(z)$ is so designed and $P(z)$ is Nyquist($2M$), $P(z)Q(z)$ will be very close to a Nyquist($2M$) filter and $P(zW^2)Q(z) \approx 0$. Notice that this design technique involves no optimization at all.

Kaiser window design: In this approach, the synthesis prototype $Q(z) = \widetilde{P}(z)$. The analysis prototype $P(z)$ is designed through Kaiser window. The impulse response of $P(z)$ is $p(n) = w(n)h_i(n)$, where $w(n)$ is a Kaiser window and $h_i(n) = \sin(\omega_c n)/\pi n$ is the impulse response of the ideal filter with cutoff frequency ω_c, [1]. After we choose the stopband attenuation and the width of the transition band, the length of the window can be estimated by a formula developed by Kaiser. In this case, the window is completely determined. The cutoff frequency ω_c is the only parameter left to be optimized such that $P(z)\widetilde{P}(z)$ is close to a Nyquist($2M$) filter. Using the second definition for a Nyquist filter, we can choose a simple objective function, $\phi_{Kaiser} = \max_{n,n\neq 0} |p(n)*p(-n)|_{\downarrow 2M}$. We can adjust the parameter ω_c to find the best $P_0(z)$ which yields the smallest ϕ_{Kaiser}. Experiments show that ϕ_{Kaiser} is a convex function of ω_c. Very good design can be obtained as will be demonstrated in Example 4.1.

Example 4.1 *Tunable multilevel filter.* A 20 channel cosine modulated under-decimated filter bank is used in this example. In this case $M = 10$. The prototypes are designed through Kaiser window approach. The analysis prototype $P(z)$ is linear-phase with order $N = 110$, stopband attenuation 75 dB, passband edge $\omega_p = 0.04\pi$ and stopband edge $\omega_s = 0.098\pi$. The synthesis prototype $Q(z)$ is the time reversed version of $P(z)$, $Q(z) = z^{-N}\widetilde{P}(z)$. Fig. 4.2(a) show the magnitude response of $P(z)$.

After designing the $P(z)$, we can tune a_k to obtain the desired overall response, $T(z)$. For instance, we set $a_0 = a_1 = 1$, $a_2 = a_3 = a_4 = 0$, $a_5 = a_6 = a_7 = 0.7$ and $a_8 = a_9 = a_{10} = 0.3$. The magnitude response of the resulting $T(z)$ is plotted in Fig. 4.2(b). Since $T(z)$ has linear phase, we do not show the phase response. Fig. 4.2(c) shows the corresponding dB plot of Fig. 4.2(b).

Appendix A.

Express $P(z)$ in terms of type 1 polyphase components as in (2.8). The analysis filters are

$$\begin{aligned} H_k(z) &= 2\sum_{n=0}^{2M-1} a_k E_n(-z^{2M}) \cos(\tfrac{\pi}{M}(k+0.5)n), \\ H_{k+M}(z) &= 2\sum_{n=0}^{2M-1} a_k E_n(-z^{2M}) \sin(\tfrac{\pi}{M}(k+0.5)n), \\ & k = 0, 1, \ldots, M-1. \end{aligned}$$

Let $\mathbf{h}(z)$ be as in (2.10), then

$$\mathbf{h}(z) = 2\mathbf{A}\underbrace{\begin{pmatrix} \mathbf{C} & -\mathbf{\Lambda S} \\ \mathbf{S} & \mathbf{\Lambda C} \end{pmatrix}}_{\mathbf{T}} \begin{pmatrix} \mathbf{E}_0(z^M) & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_1(z^M) \end{pmatrix} \mathbf{e}(z),$$

where $\mathbf{\Lambda}$, $\mathbf{A}$, $\mathbf{E}_0(z^M)$ and $\mathbf{E}_1(z^M)$ are diagonal matrices with $\mathbf{\Lambda}_{kk} = (-1)^k, \mathbf{A}_{kk} = \mathbf{A}_{k+M,k+M} = a_k, [\mathbf{E}_0(z^M)]_{kk} = E_k(-z^{2M})$, and $[\mathbf{E}_1(z^M)]_{kk} = E_{k+M}(-z^{2M})$ for $k = 0, 1, \ldots, M-1$. The matrices $\mathbf{C}$ and $\mathbf{S}$ are of dimension $M \times M$ with

$$\begin{aligned} [\mathbf{C}]_{kn} &= \cos(\tfrac{\pi}{M}(k+0.5)n) \\ [\mathbf{S}]_{kn} &= \sin(\tfrac{\pi}{M}(k+0.5)n), \end{aligned} \quad 0 \le k, n, < M.$$

It can shown that $\mathbf{C}$ and $\mathbf{S}$ are related by $\mathbf{S} = \mathbf{\Lambda C J}$, where $\mathbf{J}$ is an $M \times M$ matrix with nonzero entries $\mathbf{J}_{kn} = 1$, when $k + n = M$, $0 \le k, n < M$. Using this property, we have $\mathbf{T} = \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda C} \end{pmatrix} \begin{pmatrix} \mathbf{I} & -\mathbf{J} \\ \mathbf{J} & \mathbf{I} \end{pmatrix}$.

With all these informations, we can draw the implementation of the cosine modulated filter bank in Fig. A.1. The matrix $\mathbf{J}$ merely reorders the inputs and requires no computation. From the block diagram, we observe that the complexity of the analysis bank is that of the analysis prototype filter plus two DCT matrices and some scalars.

References

[1] P. P. Vaidyanathan, *Multirate systems and filter banks*, Englewood Cliffs, Prentice Hall, 1993.

[2] C. S. Burrus. "Block implementation of digital filters," *IEEE Trans. on Circuit theory*, pp. 697-701, Nov. 1971.

[3] M. Vetterli, "Running FIR and IIR filtering using multirate filter banks," *IEEE Trans. on ASSP*, pp. 730-738, May, 1988.

[4] P. P. Vaidyanathan, "Orthonormal and Biorthonormal Filter Banks as Convolvers, and Convolutional Coding Gain," *IEEE Tran. on SP*, pp. 2110-2130, June 1993.

[5] H. J. Nussbaumer, "Pseudo QMF filter bank," *IBM Tech. disclosure Bulletin*, vol. 24, pp. 3081-3087, Nov. 1981.

[6] P. Yip, and K. R. Rao, "Fast discrete transforms," in *Handbook of digital signal processing*, edited by D. F. Elliott, Academic Press, San Diego, CA, 1987.

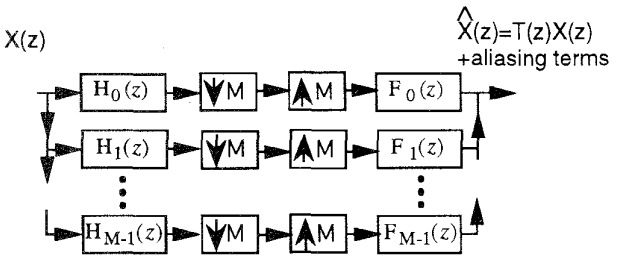


Fig. 1.1. M channel maximally decimated filter bank.

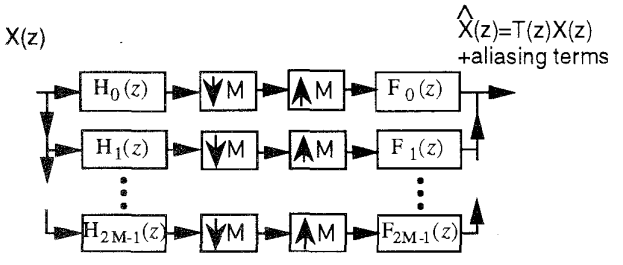


Fig. 1.2. 2M channel under-decimated filter bank.

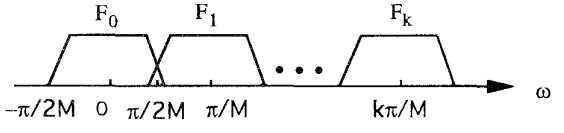


Fig. 2.1. Spectral support of synthesis filters.

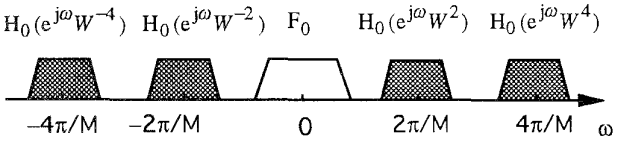


Fig. 2.2. Image copies of $H_0(z)$ due to decimation followed by expansion and the spectral support of $F_0(z)$.

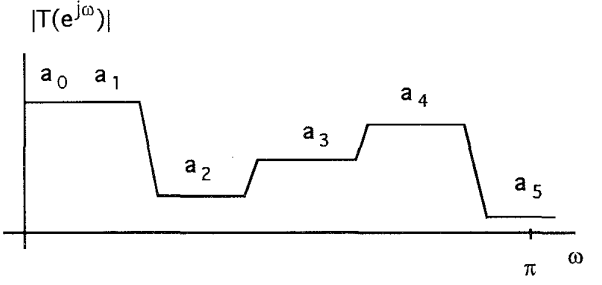


Fig. 2.3. A typical magnitude response of T(z), a multilevel filter.

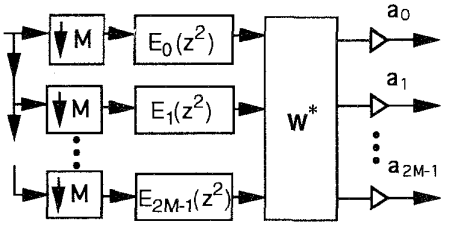


Fig. 2.4. Efficient implementation of the analysis bank of the 2M channel DFT filter bank. The DFT matrix, **W** is of dimension 2M by 2M.

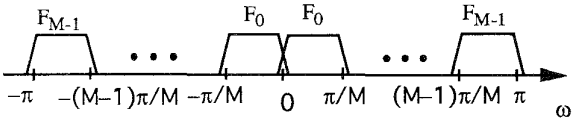


Fig. 3.1. Spectral support of the synthesis filters.

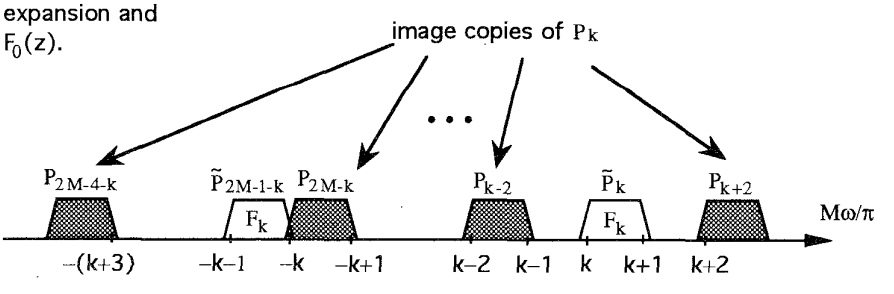


Fig. 3.2. Image copies of $P_k(z)$ due to decimation followed by expansion and the spectral support of $F_k(z)$ when k is even.

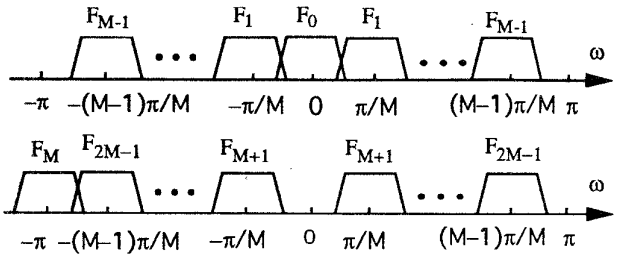


Fig. 3.3. Spectral support of synthesis filters for a different stacking.

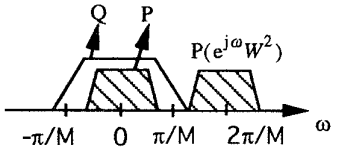


Fig. 4.1. The passband of Q(z) covers the passband and transition band of P(z) but $P(e^{j\omega}W^2)$ falls into the stopband of Q(z).

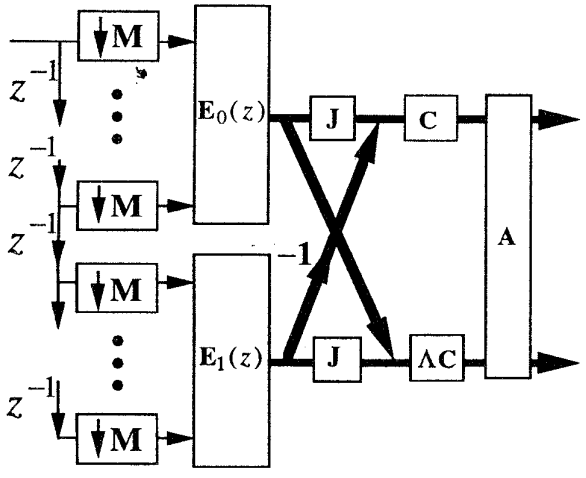
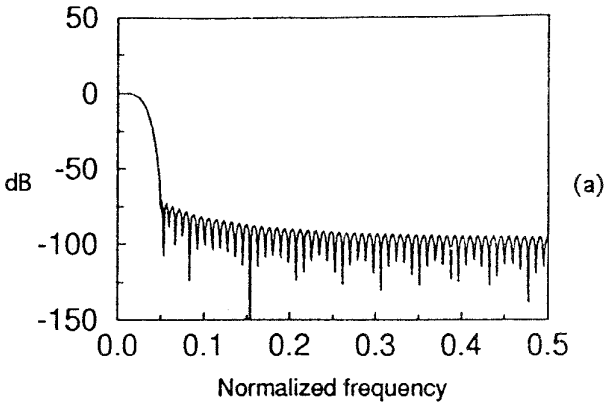
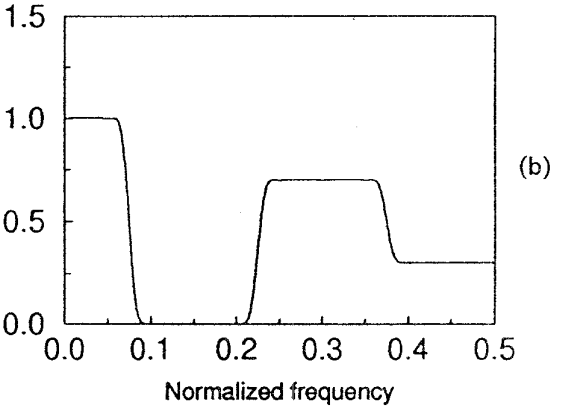


Fig. A.1. Efficient implementation of the analysis bank of the under-decimated cosine modulated 2M channel filter bank.





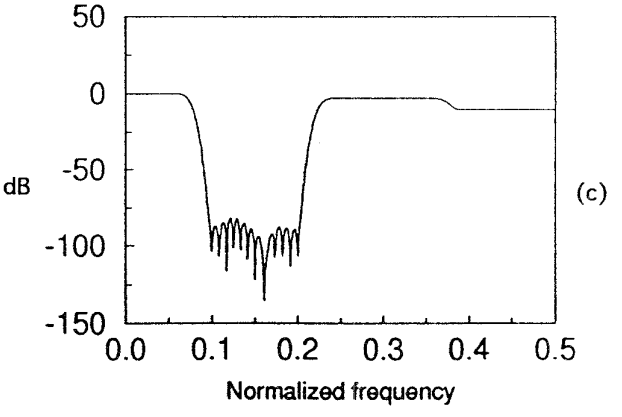


Fig. 4.2. Example 4.1. (a) The magnitude response of the prototype, P(z). (b) The magnitude response of the overall response T(z). (c) The magnitude response of the overall response T(z) in dB plot.