Design of Very Low-Sensitivity and Low-Noise Recursive Filters Using a Cascade of Low-Order Lattice Wave Digital Filters

Juha Yli-Kaakinen and Tapio Saramäki

***Abstract*—Among the best structures for implementing recursive digital filters are lattice wave digital (LWD) filters (parallel connections of two all-pass filters). They are characterized by many attractive properties, such as a reasonably low coefficient sensitivity, a low roundoff noise level, and the absence of parasitic oscillations. The main drawback is that if the stopband attenuation is very high, then many bits are required for the coefficient representations. In order to get around this problem, a structure consisting of a cascade of LWD filters is introduced in this paper. The main advantage of the proposed structure, compared with the direct LWD filter, is that the poles of the new structure are further away from the unit circle. Consequently, the number of bits required for both the data and coefficient representations are significantly reduced. The price paid for these reductions is a slight increase in the overall filter order. By properly selecting the number of LWD filters and their orders and optimizing them, their coefficients are implementable by using a few powers of two. Filters of this kind are very attractive in very large-scale integration (VLSI) implementations, where a general multiplier is very costly.**

***Index Terms*— All-pass filters, finite precision, lattice wave digital filters, multiplierless design, parallel connections of all-pass filters, recursive filters, VLSI implementations.**

I. Introduction

During the past three decades, a large number of digital filter structures have been developed. The main concern has been the performance of the digital filter in finite wordlength implementation on one hand and the computational complexity of the implementation on the other. It has turned out that very useful digital filter structures for various applications can be constructed by using all-pass subfilters as building blocks. Traditionally, all-pass filters have been used as phase equalizers [1]. Also conventional digital filters (e.g., classical odd-order low-pass and high-pass transfer functions) can be realized as a parallel connection of two all-pass filters. A well-known class of such filters are the lattice wave digital (LWD) filters [2]–[4], which are related to certain analog prototype networks. Direct z-domain techniques have also been advanced for designing filters of this kind and certain extended filter types [5]–[7]. All-pass subfilters are also the basic building blocks of recursive half-band [4], [5], [8]–[12] and Nth-band filters [10], [11], [13]–[15], which have been found to be very efficient in sampling-rate conversion applications. In addition, all the filter types mentioned above can also be designed to have an approximately linear phase response in the passband [11], [14], [16]–[23].

The different types of filters composed of all-pass subfilters share some advantageous properties, such as a reasonably low coefficient sensitivity and a low roundoff noise level. In addition, all these filter types can be realized by using first- and second-order all-pass sections as basic building blocks. The resulting filter structures are highly modular, which makes them suitable for signal processor and VLSI implementations [24].

When considering the parallel connection of two all-pass filters, the coefficient sensitivity is very low in the passband if the all-pass filter structures are such that their transfer functions remain allpass in spite of coefficient quantization. However, the stopband sensitivity is not as good. In most cases, it has turned out that the required coefficient wordlength is proportional to the required stopband attenuation [14]. Therefore, the coefficient wordlength requirements can be reduced if the filter is realized using subfilters with lower stopband attenuations, e.g., in cascade or, more generally, as a tapped cascaded interconnection of identical subfilters [25].

This paper introduces an approach to designing infinite impulse response (IIR) filters using a cascade of different LWD filters. The main advantage of this approach is that the poles of the cascaded lattice filters are further away from the unit circle compared with the direct LWD filters. This means that the number of data bits and the number of bits required for the coefficient representations can be significantly reduced. By properly determining the number of filter sections to be cascaded, as well as their orders, all the coefficient values can be optimized to be representable as two or three powers of two. This makes the proposed filter structure very attractive for VLSI implementations, where a general multiplier element is very costly.

The outline of this paper is as follows. Section II introduces the proposed class of recursive filters consisting of a cascade of LWD filters. In Section III, the optimization problem is stated for designing these filters to meet the given amplitude criteria with very simple coefficient representation forms. The target is to first minimize the number of powers of two required for representing all the coefficients, and then to minimize the number of fractional bits. Section IV shows how to arrive at the desired solution. The first step is to determine the number of LWD filters to be cascaded and their orders

Manuscript received October 18, 1998; revised March 9, 1999. This research was supported by the Academy of Finland. This paper was recommended by Guest Editors F. Maloberti, P. Diniz, and K. Jenkins.

The authors are with the Signal Processing Laboratory, Tampere University of Technology, P.O. Box 557, FIN-33101 Tampere, Finland.

Publisher Item Identifier S 1057-7130(99)05644-X.

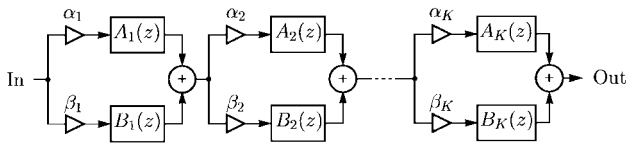


Fig. 1. Proposed recursive filter structure. The $A_k(z)$'s and $B_k(z)$'s are stable all-pass filters.

to exceed the amplitude criteria to provide some tolerance for the coefficient quantization. The second algorithm of Dutta and Vidyasagar [26] turns out to be very efficient to optimize, in the minimax sense, the desired filter with infinite-precision coefficients. The second step of Section IV involves finding the filter with simple coefficient representation forms. For this purpose, a very straightforward quantization scheme is introduced. Finally, in Section V, several examples are included, illustrating the efficiency of the quantization scheme of Section IV and the superiority of the proposed filters over direct LWD filters in finite-wordlength implementations.

II. Proposed Filter Class

Let the transfer function of a recursive digital filter be given by

$$H(z) = \prod_{k=1}^{K} H_k(z) \tag{1a}$$

where

$$H_k(z) = \alpha_k A_k(z) + \beta_k B_k(z). \tag{1b}$$

Here, the $A_k(z)$'s and $B_k(z)$'s for $k = 1, 2, \cdots, K$ are stable all-pass filters of orders M_k and N_k, respectively. An implementation of the above transfer function is depicted in Fig. 1. This contribution concentrates on synthesizing low-pass filters. In this case, $M_k = N_k - 1$ or $M_k = N_k + 1$, so that $M_k + N_k$ (the overall order of $H_k(z)$ is odd). If the $A_k(z)$'s and $B_k(z)$'s are implemented as a cascade of first- and second-order wave digital all-pass structures and M_k and N_k are assumed to be odd and even, respectively, then the $A_k(z)$'s and $B_k(z)$'s are expressible in terms of the adaptor coefficients as follows (see, e.g., [4] or [24]):

$$A_k(z) = \frac{-\gamma_0^{(k)} + z^{-1}}{1 - \gamma_0^{(k)} z^{-1}} \cdot \prod_{l=1}^{m_k} \frac{-\gamma_{2l-1}^{(k)} + \gamma_{2l}^{(k)}(\gamma_{2l-1}^{(k)} - 1)z^{-1} + z^{-2}}{1 + \gamma_{2l}^{(k)}(\gamma_{2l-1}^{(k)} - 1)z^{-1} - \gamma_{2l-1}^{(k)} z^{-2}} \tag{2a}$$

and

$$B_k(z) = \prod_{l=1}^{n_k} \frac{-\hat{\gamma}_{2l-1}^{(k)} + \hat{\gamma}_{2l}^{(k)}(\hat{\gamma}_{2l-1}^{(k)} - 1)z^{-1} + z^{-2}}{1 + \hat{\gamma}_{2l}^{(k)}(\hat{\gamma}_{2l-1}^{(k)} - 1)z^{-1} - \hat{\gamma}_{2l-1}^{(k)} z^{-2}} \tag{2b}$$

where

$$m_k = (M_k - 1)/2, \quad n_k = N_k/2. \tag{2c}$$

If $A_k(z)$ possesses a real pole at $z = r_0^{(k)}$ and m_k complex-conjugate pole pairs at $z = r_l^{(k)} \exp(\pm j\theta_l^{(k)})$ for $l = 1, 2, \cdots, m_k$, and $B_k(z)$ possesses n_k complex-conjugate pole pairs at $z = \hat{r}_l^{(k)} \exp(\pm j\hat{\theta}_l^{(k)})$ for $l = 1, 2, \cdots, n_k$, then

$$\gamma_0^{(k)} = r_0^{(k)} \tag{3a}$$

$$\gamma_{2l-1}^{(k)} = -(r_l^{(k)})^2, \quad \text{for } l = 1, 2, \cdots, m_k \tag{3b}$$

$$\gamma_{2l}^{(k)} = \frac{2r_l^{(k)} \cos(\theta_l^{(k)})}{1 + (r_l^{(k)})^2}, \quad \text{for } l = 1, 2, \cdots, m_k \tag{3c}$$

$$\hat{\gamma}_{2l-1}^{(k)} = -(\hat{r}_l^{(k)})^2, \quad \text{for } l = 1, 2, \cdots, n_k \tag{3d}$$

and

$$\hat{\gamma}_{2l}^{(k)} = \frac{2\hat{r}_l^{(k)} \cos(\hat{\theta}_l^{(k)})}{1 + (\hat{r}_l^{(k)})^2}, \quad \text{for } l = 1, 2, \cdots, n_k. \tag{3e}$$

III. Optimization Problem

Before stating the optimization problem, we denote the transfer function of the filter by $H(\Phi, z)$, where Φ is the adjustable parameter vector

$$\begin{aligned}\Phi = [&\alpha_1, \beta_1, r_0^{(1)}, \cdots, r_{m_1}^{(1)}, \theta_1^{(1)}, \cdots, \theta_{m_1}^{(1)}, \hat{r}_1^{(1)}, \cdots, \hat{r}_{n_1}^{(1)} \\ &\hat{\theta}_1^{(1)}, \cdots, \hat{\theta}_{n_1}^{(1)}, \cdots, \alpha_K, \beta_K, r_0^{(K)}, \cdots, r_{m_K}^{(K)} \\ &\theta_1^{(K)}, \cdots, \theta_{m_K}^{(K)}, \hat{r}_1^{(K)}, \cdots, \hat{r}_{n_K}^{(K)}, \hat{\theta}_1^{(K)}, \cdots, \hat{\theta}_{n_K}^{(K)}].\end{aligned} \tag{4}$$

The amplitude specifications for the filter are stated as follows:

$$1 - \delta_p \le |H(\Phi, e^{j\omega})| \le 1, \quad \text{for } \omega \in [0, \omega_p] \tag{5a}$$

$$|H(\Phi, e^{j\omega})| \le \delta_s, \quad \text{for } \omega \in [\omega_s, \pi]. \tag{5b}$$

Alternatively, these criteria are expressible as

$$|E(\Phi, \omega)| \le 1, \quad \text{for } \omega \in [0, \omega_p] \cup [\omega_s, \pi] \tag{6a}$$

$$E(\Phi, \omega) \le 0, \quad \text{for } \omega \in [0, \omega_p] \tag{6b}$$

where

$$E(\Phi, \omega) = W(\omega)[|H(\Phi, e^{j\omega})| - D(\omega)] \tag{6c}$$

with

$$D(\omega) = \begin{cases} 1, & \omega \in [0, \omega_p] \\ 0, & \omega \in [\omega_s, \pi] \end{cases} \tag{6d}$$

and

$$W(\omega) = \begin{cases} 1/\delta_p, & \omega \in [0, \omega_p] \\ 1/\delta_s, & \omega \in [\omega_s, \pi]. \end{cases} \tag{6e}$$

This work concentrates on the coefficient quantization in fixed-point arithmetic. In many implementations, it is attractive to carry out the multiplication of a data sample by a filter coefficient value using a sequence of shifts and adds. For such a purpose, it is desirable to express the coefficient values in the form

$$\sum_{r=1}^{R} a_r 2^{-P_r} \tag{7}$$

where each of the a_r's is either 1 or -1 and the P_r's are positive integers in the increasing order. The target is to find all the coefficient values included in Φ, as given by (4), in such a way that: 1) R, the number of powers of two, is made as small as possible and 2) P_R, the number of fractional bits, is made as small as possible.

The optimization problem under considerations is the following.

Optimization Problem: Find K, the number of subfilters, the M_k's, and N_k's, as well as the adjustable parameter vector Φ, as given by (4), in such a way that we have the following.

1) $H(\Phi, z)$ meets the criteria given by (5) or (6).
2) The coefficients included in Φ are quantized to achieve the above-mentioned target for their representations.

IV. Filter Optimization

The solution to the stated optimization problem can be found in two steps. In the first step, a filter with infinite-precision coefficients is determined in such a way that it exceeds the given amplitude criteria to provide some tolerance for the coefficient quantization. The second step involves finding a filter meeting the given criteria with simple coefficient representation forms.

A. Optimization of Infinite-Precision Filters

The problem is to find the adjustable parameter vector Φ to minimize on $[0, \omega_p] \cup [\omega_s, \pi]$ the peak absolute value of $E(\Phi, \omega)$, as given by (6c)–(6e), subject to the condition of (6b). To solve this problem, we discretize the passband and stopband regions into the frequency points $\omega_i \in [0, \omega_p], i = 1, 2, \cdots, L_p$ and $\omega_i \in [\omega_s, \pi], i = L_p+1, L_p+2, \cdots, L_p+L_s$. The resulting discrete minimax problem is to find Φ to minimize

$$\epsilon = \max_{1 \le i \le L_p + L_s} \{|E(\Phi, \omega_i)|\} \tag{8a}$$

subject to

$$E(\Phi, \omega_i) \le 0, \quad i = 1, 2, \cdots, L_p. \tag{8b}$$

In order to meet the criteria of (5) or (6), K as well as the M_k's and N_k's for $k = 1, 2, \cdots, K$, have to be selected such that the minimized ϵ becomes less than or equal to unity.

The above problem can be solved in a straightforward manner by using the second algorithm proposed by Dutta and Vidyasagar in [26], as is shown in Appendix A. For this nonlinear optimization algorithm, the convergence to the global optimum cannot be assured. Hence, a good guess for the initial filter has an extensive effect on the convergence of the algorithm to the optimal solution. It has turned out that, in many cases, it is beneficial to select all the α_k's and β_k's to be equal to $1/2$, as for the conventional LWD filters. Furthermore, it is in most cases advantageous to select all the $A_k(z)$'s and the $B_k(z)$'s to be of the same order, respectively. Only in cases where K, the number stages, is large and the required passband ripple is relatively low, it is beneficial to give other values for α_k and β_k only in one section. For the case where $\alpha_k = \beta_k = 1/2$ for $k = 1, 2, \cdots, K$, the starting point filter for further optimization can be determined by using several identical copies of the same subfilter. The passband and stopband ripples for this subfilter should be approximately equal to δ_p/K and $\sqrt[K]{\delta_s}$, respectively.[1] For the case where for one section α_k and β_k have values different from $1/2$, a good initial filter for further optimization can be achieved by using the procedure described in Appendix B.

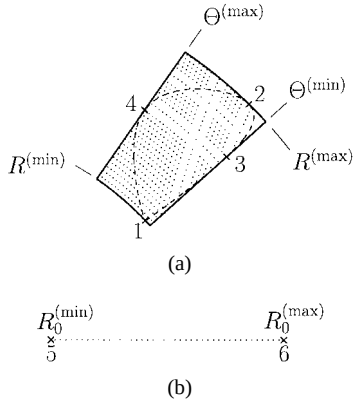


Fig. 2. Typical search spaces for the poles when three powers of two with seven fractional bits ($R = 3$ and $P_R = 7$) are used for the adaptor coefficients. (a) Upper half-plane pole for the complex-conjugate pole pair. (b) Real pole.

B. Optimization of Finite-Precision Filters

It has turned out that a very straightforward quantization scheme for the filter coefficients is obtained as follows in the case where all the α_k's and β_k's are equal to $1/2$. For each complex-conjugate pole pair, the largest and smallest values for both the radius and angle are determined in such a way that by reoptimizing the remaining pole parameter and the locations of the remaining poles, the overall criteria as given by (5) or (6) can still be met. For each real pole, the smallest and largest values for the radius are found in a similar manner.

The above procedure gives for the upper-half-plane pole for each complex-conjugate pole pair $r_l^{(k)} \exp(\pm j\theta_l^{(k)})$ for $l = 1, 2, \cdots, m_k$ and for $k = 1, 2, \cdots, K$ and $\hat{r}_l^{(k)} \exp(\pm j\hat{\theta}_l^{(k)})$ for $l = 1, 2, \cdots, n_k$ and for $k = 1, 2, \cdots, K$ the region $R\exp(j\Theta)$ where $R^{(\min)} \le R \le R^{(\max)}$ and $\Theta^{(\min)} \le \Theta \le \Theta^{(\max)}$, as illustrated in Fig. 2(a). The crosses numbered by 1, 2, 3, and 4 correspond, respectively, to the points where the smallest radius $R^{(\min)}$, the largest radius $R^{(\max)}$, the smallest angle $\Theta^{(\min)}$, and the largest angle $\Theta^{(\max)}$ are reached. Inside this region, there is the feasible region given by the dashed line in Fig. 2(a), where the pole can be located such that by relocating the remaining poles, the given overall criteria are still met by using infinite-precision arithmetic. For each real pole $r_0^{(k)}$ for $k = 1, 2, \cdots, K$, there is the corresponding region $R_0^{(\min)} \le R \le R_0^{(\max)}$ that is simultaneously the feasible region. In Fig. 2(b), the crosses numbered by 5 and 6 indicate $R_0^{(\min)}$ and $R_0^{(\max)}$, respectively.

[1] There is clearly a tradeoff between the number of subfilters, K, and the order of the subfilter; the higher is the value of K, the lower is the order of the subfilter. However, since the subfilter order is restricted to be an odd integer, there are only few practical combinations for the subfilter order and K. It is not necessary for the subfilter being an odd order elliptic filter to exactly meet the ripple requirements. This is due to the fact that further optimization makes the subfilters different and simultaneously improves the overall filter performance.

The next step is to find in the above regions those pole locations which are achievable by implementing the adaptor coefficients in the form of (7) with the given R and the given largest value for P_R. The dots in Fig. 2 indicate these pole locations. Note that the distributions are very irregular due to the desired representation form. For the complex-conjugate pole pairs, the larger region is used since it can be found by applying the algorithm to be described later only four times. All what is still needed is to check whether there exists a combination of the discrete pole positions with which the given overall criteria are met. More details on how to effectively find the desired finite-precision filter have been described in [27] and [28].

The above-mentioned infinite-precision regions can be determined conveniently by using the second algorithm of Dutta and Vidyasagar [26], as shown in Appendix A. In this case, there are $\Sigma_{k=1}^{K}\ (M_k + N_k)$ problems of the form: find Φ to minimize ψ subject to

$$|E(\Phi,\omega_i)| - 1 \le 0, \qquad i = 1,2,\cdots,L_p + L_s \tag{9a}$$

$$E(\Phi,\omega_i) \le 0, \qquad i = 1,2,\cdots,L_p. \tag{9b}$$

For these problems, ψ is $r_l^{(k)}, 1 - r_l^{(k)}$ for $l = 0,1,\cdots,m_k$, $k = 1,2,\cdots,K$; $\theta_l^{(k)}$, $\pi - \theta_l^{(k)}$ for $l = 1,2,\cdots,m_k$, $k = 1,2,\cdots,K$; and $\hat{r}_l^{(k)}$, $1 - \hat{r}_l^{(k)}$, $\hat{\theta}_l^{(k)}$, $\pi - \hat{\theta}_l^{(k)}$ for $l = 1,2,\cdots,n_k$, $k = 1,2,\cdots,K$, respectively.[2]

If for one section α_k and β_k are not equal to $1/2$, then, in addition to the poles or equivalently the adaptor coefficients, these parameters are included in the above quantization scheme.

The proposed quantization scheme provides significant advantages over those based on the use of simulated annealing or genetic algorithms. First of all, it is always guaranteed that the optimum solution can be found provided that it exists. Second, the computational workload to arrive at the optimum discrete-valued solution is in most cases significantly smaller than in the two above-mentioned algorithms.

V. Numerical Examples

This section shows, by means of examples, the efficiency and flexibility of the quantization scheme described in the previous section as well as the superiority of the proposed filters over direct LWD filters in finite-wordlength implementations. More examples can be found in [27].

A. Example 1

It is desired to design a filter with the passband and stopband edges at $\omega_p = 0.05\pi$ and at $\omega_s = 0.1\pi$, respectively. The maximum allowable passband ripple and the required stopband attenuation are 0.5 dB ($\delta_p = 0.0559$) and 100 dB ($\delta_s = 10^{-5}$), respectively.

[2] In these problems, the optimization is performed, using special arrangements, in such a manner that the above-mentioned infinite-precision regions are not allowed to completely overlap, thus reducing the computational complexity of the overall quantization scheme.

TABLE I
Optimized Finite-Precision Adaptor Coefficients for the Direct LWD Filter in Example 1

$A(z)$	$B(z)$
$\gamma_0 = \ \ 974 \cdot 2^{-10}$	
$\gamma_1 = -940 \cdot 2^{-10}$	$\widehat{\gamma}_1 = -934 \cdot 2^{-10}$
$\gamma_2 = \ \ 1014 \cdot 2^{-10}$	$\widehat{\gamma}_2 = \ \ 1020 \cdot 2^{-10}$
$\gamma_3 = -1007 \cdot 2^{-10}$	$\widehat{\gamma}_3 = -964 \cdot 2^{-10}$
$\gamma_4 = \ \ 1009 \cdot 2^{-10}$	$\widehat{\gamma}_4 = \ \ 1011 \cdot 2^{-10}$

TABLE II
Optimized Finite-Precision Adaptor Coefficients for the Cascade of Two LWD Filters in Example 1

$A(z)$	$B(z)$
$\gamma_0^{(1)} = \ 1 - 2^{-3} - 2^{-7}$	
$\gamma_1^{(1)} = -1 + 2^{-4} + 2^{-8}$	$\widehat{\gamma}_1^{(1)} = -1 + 2^{-2} - 2^{-4}$
$\gamma_2^{(1)} = \ 1 - 2^{-6} - 2^{-8}$	$\widehat{\gamma}_2^{(1)} = \ 1 - 2^{-6}$
$\gamma_0^{(2)} = \ 1 - 2^{-3} + 2^{-5}$	
$\gamma_1^{(2)} = -1 + 2^{-5} + 2^{-7}$	$\widehat{\gamma}_1^{(2)} = -1 + 2^{-3}$
$\gamma_2^{(2)} = \ 1 - 2^{-6} - 2^{-8}$	$\widehat{\gamma}_2^{(2)} = \ 1 - 2^{-6} + 2^{-8}$

The minimum order of a direct LWD filter to meet the given amplitude criteria is seven.[3] However, this filter just meets the given criteria. Therefore, to allow some tolerance for the coefficient quantization, the filter order has to be increased to nine. Using the quantization scheme described above, the given criteria are met by a filter with $K = 1$ and $\alpha_1 = \beta_1 = 1/2$. The optimized discrete-valued adaptor coefficients are given in Table I. In this case, 10 fractional bits[4] are needed for the adaptor coefficients. Among the solutions satisfying the given amplitude specifications with ten fractional bits, the one with the smallest ϵ, as given by (8a), has been selected.

For $K = 2$, the given criteria are met by $\alpha_k = \beta_k = 1/2$, $M_k = 3$, and $N_k = 2$ for $k = 1,2$. Table II gives the optimized finite-precision adaptor coefficients. In this case, all the coefficients can be represented as two or three powers of two, and eight fractional bits are needed. A total of only eight adders[5] are required to implement all the filter coefficients. Among the solutions satisfying the given amplitude specifications with the smallest number of adders, the one with the smallest ϵ, as given by (8a), has been selected.

For $K = 4$, the given criteria are met by $\alpha_k = \beta_k = 1/2$, $M_k = 1$, and $N_k = 2$ for $k = 1,2,3,4$. The optimized adaptor coefficients are given in Table III. As for $K = 2$, eight fractional bits are required. In this case, nine adders are

[3] It is well known that the odd order elliptic low-pass filter is the most selective filter being implementable as a parallel connection of two all-pass filters (see, e.g., [4]).

[4] The filter specifications cannot be satisfied using four powers of two for the adaptor coefficients. Therefore, the coefficients are represented as fixed-point binary numbers as $-a_0 + \Sigma_{r=1}^{R}\ a_r 2^{-r}$, where a_r for $r = 0,1,\cdots,R$ is either 0 or 1. Here, R is the number of fractional bits.

[5] When the adaptors shown in Fig. 9 in [4] are used, the actual multiplier to be implemented is always positive and less than or equal to half. Therefore, in this case, the number of adders required for implementing the adaptor coefficients becomes smaller.

TABLE III
OPTIMIZED FINITE-PRECISION ADAPTOR COEFFICIENTS FOR THE CASCADE OF FOUR LWD FILTERS IN EXAMPLE 1

$A(z)$	$B(z)$
$\gamma_0^{(1,2)} = 1 - 2^{-2} + 2^{-5}$	$\widehat{\gamma}_1^{(1,2)} = -1 + 2^{-3}$ $\widehat{\gamma}_2^{(1,2)} = 1 - 2^{-5} + 2^{-7}$
$\gamma_0^{(3)} = 1 - 2^{-3}$	$\widehat{\gamma}_1^{(3)} = -1 + 2^{-3} - 2^{-6}$ $\widehat{\gamma}_2^{(3)} = 1 - 2^{-6} - 2^{-8}$
$\gamma_0^{(4)} = 1 - 2^{-3} - 2^{-5}$	$\widehat{\gamma}_1^{(4)} = -1 + 2^{-3} - 2^{-7}$ $\widehat{\gamma}_2^{(4)} = 1 - 2^{-6} - 2^{-8}$

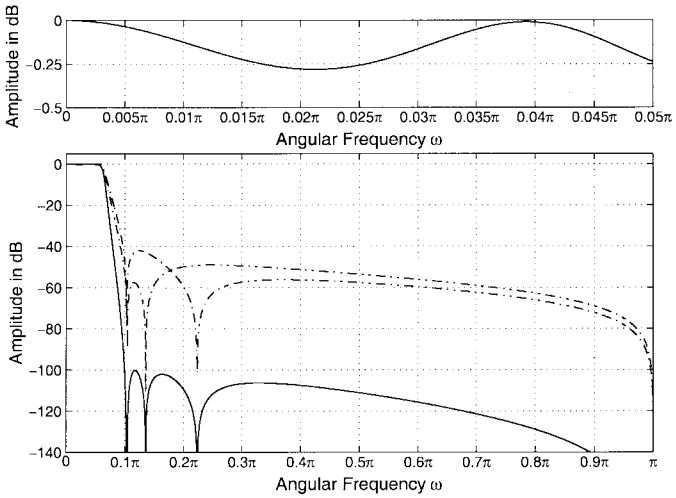


Fig. 3. Some amplitude responses for the cascade of two optimized finite-precision LWD filters in Example 1. The solid and dot-dashed lines show the responses for the overall filter and the subfilters, respectively.

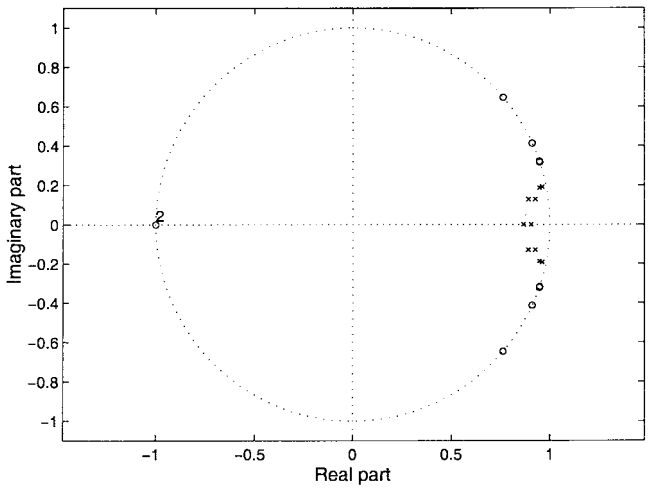


Fig. 4. Pole–zero plot for the cascade of two optimized finite-precision LWD filters in Example 1.

required to implement all the filter coefficients. Note that two sections are identical. The solution has been selected as for the $K = 2$ case.

Fig. 3 shows for the $K = 2$ design the amplitude responses of both sections, as well as that of the overall filter. In addition, the passband details of the amplitude response are shown for the overall filter. The pole–zero plot for the overall design

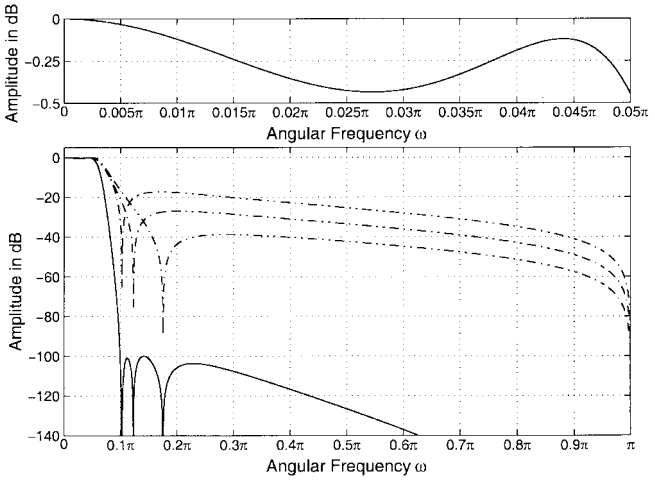


Fig. 5. Some amplitude responses for the cascade of four optimized finite-precision LWD filters in Example 1. The solid and dot-dashed lines show the responses for the overall filter and the subfilters, respectively. Two subfilters are identical (the dot-dashed line with the lowest attenuation).

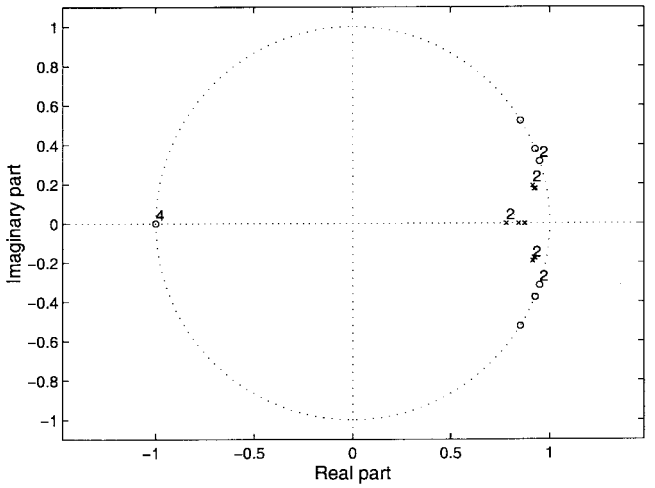


Fig. 6. Pole–zero plot for the cascade of four optimized finite-precision LWD filters in Example 1.

is depicted in Fig. 4. Similar characteristics for $K = 4$ are shown in Figs. 5 and 6, respectively.

The above cascades of two and four low-order LWD filter sections are very attractive for VLSI implementations, since no general multipliers are needed. The price paid for this is a slight increase in the overall filter order compared to the direct LWD filter. For $K = 2$, the order increases from nine to ten and for $K = 4$, from nine to twelve.

Another advantage of the proposed filters compared to the direct LWD filter is the fact that the radius of the outermost complex-conjugate pole pair is significantly smaller. For $K = 1, K = 2$, and $K = 4$, these values are 0.991 66, 0.980 28, and 0.947 74, respectively. When using the adaptors shown in Fig. 9 in [4], the output noise gains are 42.9, 34.3, and 30.8 dB for $K = 1, K = 2$, and $K = 4$, respectively. This shows that for $K = 2$ and $K = 4$, roughly one and two fewer bits are required for the data representation to arrive at approximately the same output noise level as with the corresponding direct LWD filter. It should be pointed out that lower output noise

TABLE IV
OPTIMIZED FINITE-PRECISION ADAPTOR COEFFICIENTS FOR THE DIRECT LWD FILTER IN EXAMPLE 2

$A(z)$	$B(z)$
$\gamma_0 = 1 - 2^{-3} + 2^{-6}$	
$\gamma_1 = -1 + 2^{-3} + 2^{-6} + 2^{-9}$	$\widehat{\gamma}_1 = -1 + 2^{-2} - 2^{-4} + 2^{-9}$
$\gamma_2 = 1 - 2^{-5}$	$\widehat{\gamma}_2 = 1 - 2^{-6} + 2^{-9}$
$\gamma_3 = -1 + 2^{-5} - 2^{-7} - 2^{-9}$	$\widehat{\gamma}_3 = -1 + 2^{-4} + 2^{-6}$
$\gamma_4 = 1 - 2^{-4} - 2^{-8}$	$\widehat{\gamma}_4 = 1 - 2^{-4} + 2^{-6} - 2^{-8}$

TABLE V
OPTIMIZED FINITE-PRECISION ADAPTOR COEFFICIENTS FOR THE CASCADE OF FOUR LWD FILTERS IN EXAMPLE 2

$A(z)$	$B(z)$
$\gamma_0^{(1,2)} = 2^{-1} + 2^{-3}$	$\widehat{\gamma}_1^{(1,2)} = -1 + 2^{-2} - 2^{-5}$
	$\widehat{\gamma}_2^{(1,2)} = 1 - 2^{-3} + 2^{-5}$
$\gamma_0^{(3)} = 2^{-1} + 2^{-3} + 2^{-5}$	$\widehat{\gamma}_1^{(3)} = -1 + 2^{-2}$
	$\widehat{\gamma}_2^{(3)} = 1 - 2^{-3} + 2^{-5}$
$\gamma_0^{(4)} = 1 - 2^{-2} + 2^{-5}$	$\widehat{\gamma}_1^{(4)} = -1 + 2^{-2} - 2^{-4}$
	$\widehat{\gamma}_2^{(4)} = 1 - 2^{-4}$

values can be achieved by using other adaptor structures (see, e.g., [24]).

B. Example 2

The criteria are the same as in Example 1 except that the passband and stopband edges are now doubled, that is, $\omega_p = 0.1\pi$ and $\omega_s = 0.2\pi$. This example is included to emphasize the fact that for the specifications with larger edge values, significantly fewer fractional bits are required for the proposed cascade-form filters.

For $K = 1$, the criteria are met by a filter with $\alpha_1 = \beta_1 = 1/2, M_1 = 5$ and $N_1 = 4$, as in Example 1. Table IV gives the optimized discrete-valued adaptor coefficients. In this case, four powers of two with nine fractional bits are required for the direct LWD filter.

For $K = 4$, the given criteria are met by $\alpha_k = \beta_k = 1/2$, $M_k = 1$, and $N_k = 2$, for $k = 1, 2, 3, 4$, as in Example 1. The optimized adaptor coefficients are given in Table V. In this case, only five fractional bits are needed compared to eight bits required by the corresponding Example 1 filter. Again, two sections are identical.

C. Example 3

The band edges are the same as in Example 2, that is, $\omega_p = 0.1\pi$ and $\omega_s = 0.2\pi$. The maximum allowable passband ripple and the required stopband attenuation are 0.02 dB ($\delta_p = 1.1513 \cdot 10^{-3}$) and 80 dB ($\delta_s = 10^{-4}$), respectively.[6]

[6]In the case where all the α_k's and β_k's are equal to $1/2$, the filter structure constrains the maximum of the amplitude response to be unity and this response oscillates in the passband within 1 and $1 - \delta_p$. However, in the case where for one section α_k and β_k are not equal to $1/2$, the amplitude response can be allowed to oscillate in the passband within $1 \pm \delta_p$. Therefore, if the peak-to-peak passband ripple in decibels is A_p, then for the first case, δ_p is determined from $A_p = -20\log_{10}(1 - \delta_p)$. For the latter case, in turn, δ_p is determined from $A_p = 20\log_{10}[(1 + \delta_p)/(1 - \delta_p)]$.

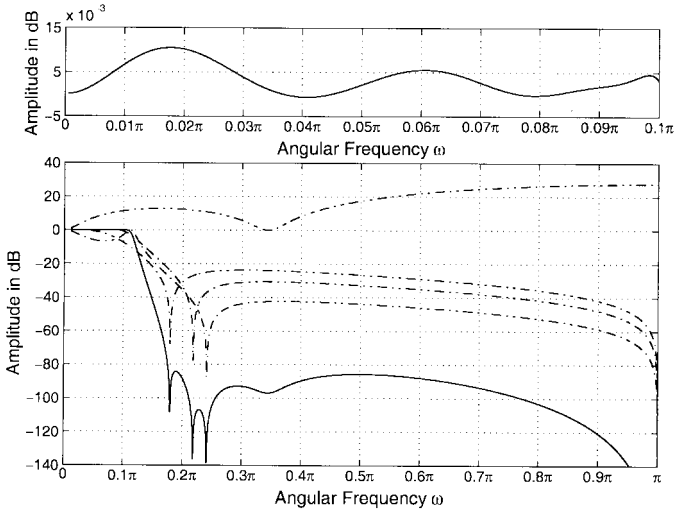


Fig. 7. Some amplitude responses for the cascade of four optimized finite-precision LWD filters in Example 3. The solid and dot-dashed lines show the responses for the overall filter and the subfilters, respectively. The uppermost dot-dashed line correspond to the section where α_k and β_k are not equal to $1/2$.

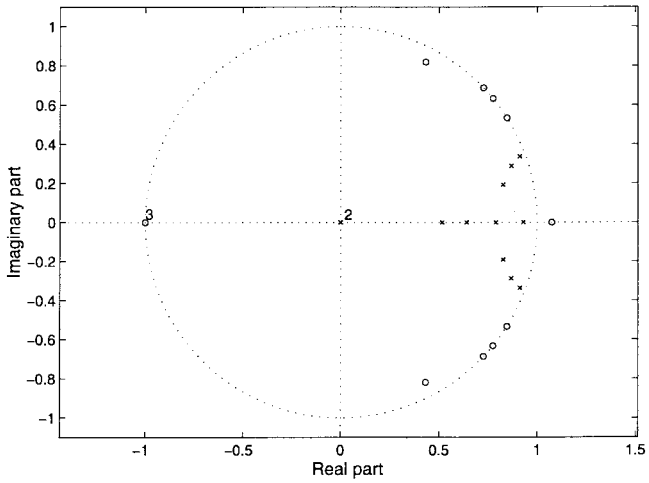


Fig. 8. Pole–zero plot for the cascade of four optimized finite-precision LWD filters in Example 3.

This example illustrates the usefulness of the filters where for one section α_k and β_k are not equal to $1/2$.

For $K = 4$, the given criteria are met by $\alpha_k = \beta_k = 1/2$ for $k = 1, 2, 3, \alpha_4 = 1495 \cdot 2^{-7}$, and $\beta_4 = -1623 \cdot 2^{-7}$, while $M_k = 1$ and $N_k = 2$ for $k = 1, 2, 3, 4$. In this case, seven fractional bits are required for the adaptor coefficients. In the case where all the α_k's and β_k's are equal to $1/2$, six sections with $M_k = 1$ and $N_k = 2$ for $k = 1, 2, \cdots, 6$ are required.

The amplitude responses of all the sections, as well as that of the overall filter, are shown in Fig. 7. The pole–zero plot for the overall filter is depicted in Fig. 8. For this design, the second-order all-pass filter has been forced to be z^{-2} for the section for which α_k and β_k are not equal to $1/2$. Therefore, the overall filter has a double pole at the origin.

VI. CONCLUSION

A structure consisting of a cascade of LWD filters has been introduced for designing recursive digital filters requiring a

high stopband attenuation. The main advantage of the proposed filters compared to the direct LWD filters is that the poles of the proposed structure are further away from the unit circle. Therefore, these filters have a lower coefficient sensitivity to coefficient quantization effects than the corresponding direct LWD filters. Also, their noise variance due to the multiplication roundoff errors is significantly smaller.

APPENDIX A

This appendix shows how the second algorithm of Dutta and Vidyasagar [26] can be applied to solving the constrained nonlinear optimization problems stated in Section IV.

In order to apply the second algorithm of Dutta and Vidyasagar to solving the problem stated in Section IV-A, we restate it in the following form: find the parameter vector Φ of the filter to minimize

$$\epsilon = \max_{1\le i\le I_1} \{f_i(\Phi)\} \tag{A1}$$

subject to

$$g_i(\Phi) \le 0, \qquad i = 1, 2, \cdots, I_2 \tag{A2}$$

where $I_1 = L_p + L_s$ and for $i = 1, 2, \cdots, I_1$

$$f_i(\Phi) = |W(\omega_i)[|H(\Phi, e^{j\omega_i})| - D(\omega_i)]| \tag{A3}$$

whereas $I_2 = L_p$ and for $i = 1, 2, \cdots, I_2$

$$g_i(\Phi) = W(\omega_i)[|H(\Phi, e^{j\omega_i})| - D(\omega_i)]. \tag{A4}$$

The main idea in the algorithm is to gradually find ϕ and Φ to minimize the following function:

$$P(\Phi, \phi) = \sum_{i|f_i(\Phi)>\phi} [f_i(\Phi) - \phi]^2 + \sum_{i|g_i(\Phi)>0} w_i[g_i(\Phi)]^2. \tag{A5}$$

In (A5), the first summation contains only those $f_i(\Phi)$'s for $i = 1, 2, \cdots, I_1$ that are larger than ϕ. Similarly, the second summation contains only those $g_i(\Phi)$'s for $i = 1, 2, \cdots, I_2$ that are larger than zero. The w_i's are the weights given by the user. Usually they are selected to be equal. Their values have some effect on the convergence rate of the algorithm. If ϕ is very large, then Φ can be found to make $P(\Phi, \phi)$ zero or practically zero. On the other hand, if ϕ is too small, then $P(\Phi, \phi)$ cannot be made zero. The key idea is to find the minimum of ϕ, for which there exists Φ such that $P(\Phi, \phi)$ becomes zero or practically zero. In this case $\epsilon \approx \phi$.

The algorithm is carried out in the following steps.

Step 1: Set $B_{\text{low}} = 0, B_{\text{high}} = 10^4$, $\phi_1 = B_{\text{low}}$, and $k = 1$.

Step 2: Find $\hat{\Phi}_k$ to minimize $P(\Phi, \phi_k)$.

Step 3: Evaluate

$$M_{\text{low}} = \phi_k + \sqrt{P(\hat{\Phi}_k, \phi_k)/n} \tag{A6}$$

where n is the number of the $f_i(\hat{\Phi}_k)$'s satisfying $f_i(\hat{\Phi}_k) > \phi_k$ and

$$M_{\text{high}} = \phi_k + \frac{P(\hat{\Phi}_k, \phi_k)}{\sum_{i|f_i(\hat{\Phi}_k)>\phi_k} [f_i(\hat{\Phi}_k) - \phi_k]}. \tag{A7}$$

Step 4: If $M_{\text{high}} \le B_{\text{high}}$, then set $\phi_{k+1} = M_{\text{high}}$. Otherwise, set $\phi_{k+1} = M_{\text{low}}$. Also, set $\phi_0 = \phi_{k+1} - \phi_k$.

Step 5: Set $B_{\text{low}} = M_{\text{low}}$ and $S = P(\hat{\Phi}_k, \phi_k)$.

Step 6: Set $k = k + 1$.

Step 7: Find $\hat{\Phi}_k$ to minimize $P(\Phi, \phi_k)$.

Step 8: If $(B_{\text{high}} - B_{\text{low}})/B_{\text{high}} \le \epsilon_1$ or $\phi_0/\phi_k \le \epsilon_1$, then stop. Otherwise, go to the next step.

Step 9: If $P(\hat{\Phi}_k, \phi_k) > \epsilon_2$, then go to Step 3. Otherwise, if $S \le \epsilon_3$, then stop. If none is true, then set $B_{\text{high}} = \phi_k$, $S = 0$, $\phi_k = B_{\text{low}}$, and go to Step 7.

In the above algorithm, we have used $\epsilon_1 = \epsilon_2 = \epsilon_3 = 10^{-14}$. A very crucial issue to arrive at least at a local optimum is to perform optimization at Steps 2 and 7 effectively. We have used the Fletcher–Powell algorithm [29]. When applying the Fletcher–Powell algorithm the partial derivatives of the objective function with respect to the unknowns are needed. Another very crucial issue is to find good starting point values for the elements of the adjustable vector Φ. The effectiveness of the above algorithm lies in the fact that at Steps 2 and 7 it exploits a criterion closely resembling the one used in the least-mean-square optimization. This guarantees that the objective function is well behaved.

In the case of the optimization problem of Section IV-B, we restate it as: find ψ and the parameter vector Φ of the filter to minimize

$$\epsilon = \psi \tag{A8}$$

subject to

$$g_i(\Phi) \le 0, \qquad i = 1, 2, \cdots, I \tag{A9}$$

where $I = 2L_p + L_s$ and for $i = 1, 2, \cdots, L_p + L_s$

$$g_i(\Phi) = |W(\omega_i)[|H(\Phi, e^{j\omega_i})| - D(\omega_i)]| - 1 \tag{A10}$$

and for $i = L_p + L_{s+1}, L_p + L_{s+2} \cdots, 2L_p + L_s$

$$g_i(\Phi) = W(\omega_{i-L_p-L_s})[|H(\Phi, e^{j\omega_{i-L_p-L_s}})| - D(\omega_{i-L_p-L_s})]. \tag{A11}$$

In this case, the objective function under consideration is given by

$$P(\Phi, \psi, \phi) = [\psi - \phi]^2 + \sum_{i|g_i(\Phi)>0} w_i[g_i(\Phi)]^2. \tag{A12}$$

The target is to find Φ, ψ, and ϕ in such a way that the above function becomes zero with the minimum value of $\phi = \psi$. The algorithm can be carried out as described above. The main difference is that now a single parameter ψ being related to one of the elements of the adjustable parameter vector Φ is desired to be minimized subject to the given constraints on Φ.

APPENDIX B

This appendix shows how a good initial filter can be generated for the case where $\alpha_k = \beta_k$ for $k = 1, 2, \cdots, K-1$, $\alpha_K \ne 1/2$, $\beta_K \ne 1/2$, and all the sections are of the same order. The details can be found in [27]. The desired overall filter can be found using the following procedure:

Step 1: Determine the value of k_0 in the following transfer function

$$F(z) = [(1 + k_0 - \delta_p) - k_0 z^{-1}][(1 + z^{-1})/2]^{K-1} \quad \text{(B1)}$$

such that the maximum of $|F(e^{j\omega})|$ achieves the value of unity. Select $\alpha_K = 1 + k_0 - \delta_p$ and $\beta_K = -k_0$.

Step 2: Find the largest frequency point, denoted by Ω_p, such that $|F(e^{j\omega})| \leq 1 - \delta_p$ for $0 \leq \omega \leq \Omega_p$ and the smallest frequency point, denoted by Ω_s, such that $|F(e^{j\omega})| \leq \delta_s$ for $\Omega_s \leq \omega \leq \pi$.

Step 3: The amplitude response of the overall filter stays within 1 and $1 - \delta_p$ in the passband region $[0, \omega_p]$ and is less than or equal to δ_s in the stopband region $[\omega_s, \pi]$ by selecting the $A_k(z)$'s and $B_k(z)$'s as follows. First, all the subfilters are made identical, that is, $A_k(z) = A(z)$ and $B_k(z) = B(z)$ for $k = 1, 2, \cdots, K$. Second, it is required that the amplitude response of $[A(z)+B(z)]/2$ stays in the same passband within the limits 1 and $1 - \hat{\delta}_p$, where

$$\hat{\delta}_p = 1 - \cos(\Omega_p/2) \quad \text{(B2)}$$

and in the same stopband within the limits zero and $\hat{\delta}_s$, where

$$\hat{\delta}_s = \cos(\Omega_s/2). \quad \text{(B3)}$$

$[A(z) + B(z)]/2$ at Step 3 can be found by simply designing a minimum odd order elliptic filter that roughly[7] meets the above specifications and is implementable as a parallel connection of two all-pass filters.

ACKNOWLEDGMENT

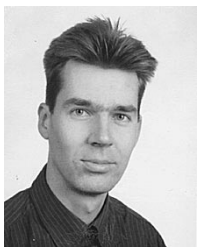
The authors wish to thank the anonymous reviewers for their constructive comments and suggestions. They also thank Prof. O. Vainio and Dr. H. Johansson for valuable discussions.

REFERENCES

[1] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
[2] R. Nouta, "The Jaumann structure in wave-digital filters," *Int. J. Circuit Theory Applicat.*, vol. 2, pp. 163–174, June 1974.
[3] A. Fettweis, H. Levin, and A. Sedlmeyer, "Wave digital lattice filters," *Int. J. Circuit Theory Applicat.*, vol. 2, pp. 203–211, June 1974.
[4] L. Gazsi, "Explicit formulas for lattice wave digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 68–88, Jan. 1985.
[5] R. Ansari and B. Liu, "A class of low-noise computationally efficient recursive digital filters with applications to sampling rate alterations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 90–97, Feb. 1985.
[6] T. Saramäki, "On the design of digital filters as a sum of two all-pass filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 1191–1193, Nov. 1985.
[7] P. P. Vaidyanathan, S. K. Mitra, and Y. Neuvo, "A new approach to the realization of low-sensitivity IIR digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 350–361, Apr. 1986.
[8] W. Wegener, "Wave digital directional filters with reduced number of multipliers and adders," *Arch. Elektr. Übertrag. Tech.*, vol. 33, pp. 239–243, June 1979.
[9] J. A. Nossek and H.-D. Schwartz, "Wave digital lattice filters with applications in communication systems," in *Proc. 1983 IEEE Int. Symp. Circuits and Systems*, Newport Beach, CA, pp. 845–848.

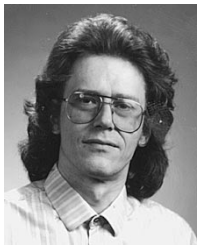
[7] Further optimization makes the identical subfilters different and improves the overall filter performance.

[10] R. A. Valenzuela and A. G. Constantinides, "Digital signal processing schemes for efficient interpolation and decimation," *Proc. Inst. Elect. Eng., Part G*, vol. 130, pp. 225–235, Dec. 1983.
[11] R. Ansari and B. Liu, "Efficient sampling rate alteration using recursive (IIR) digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1366–1373, Dec. 1983.
[12] A. Fettweis, J. A. Nossek, and K. Meerkötter, "Reconstruction of signals after filtering and sampling rate reduction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 893–902, Aug. 1985.
[13] L. Taxen, "Polyphase filter banks using wave digital filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp. 423–428, June 1981.
[14] M. Renfors and T. Saramäki, "Recursive Nth-band digital filters, Part I: Design and properties, Part II: Design of multistage decimators and interpolators," *IEEE Trans. Circuits Syst.*, vol. CAS-34, pp. 24–51, Jan. 1987.
[15] W. Drews and L. Gazsi, "A new design method for polyphase filters using all-pass sections," *IEEE Trans. Circuits Syst.*, vol. CAS-33, pp. 346–348, Mar. 1986.
[16] M. Renfors and T. Saramäki, "A class of approximately linear phase digital filters composed of allpass subfilters," in *Proc. 1986 IEEE Int. Symp. Circuits and Systems*, San Jose, CA, pp. 678–681.
[17] C. W. Kim and R. Ansari, "Approximately linear phase IIR digital filters using allpass sections," in *Proc. 1986 IEEE Int. Symp. Circuits and Systems*, San Jose, CA, May 1986, pp. 661–664.
[18] F. Leeb, "Lattice wave digital filters with simultaneous conditions on amplitude and phase," in *Proc. 1991 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Toronto, ON, Canada, pp. 1645–1648.
[19] J. Földvári-Orosz, T. Henk, and E. Simonyi, "Simultaneous amplitude and phase approximation for lumped and sampled filters," *Int. J. Circuit Theory Applicat.*, vol. 19, pp. 77–100, 1991.
[20] B. Jaworski and T. Saramäki, "Linear phase IIR filters composed of two parallel allpass sections," in *Proc. 1994 IEEE Int. Symp. Circuits and Systems*, London, U.K., pp. 537–540.
[21] A. Jones, S. S. Lawson, and T. Wicks, "Design of cascaded allpass structures with magnitude and delay constraints using simulated annealing and quasi-Newton methods," in *Proc. 1991 IEEE Int. Symp. Circuits Syst.*, Singapore, pp. 2439–2442.
[22] S. S. Lawson and T. Wicks, "Design of efficient digital filters satisfying arbitrary loss and delay specifications," *Proc. Inst. Elect. Eng., Pt. G*, vol. 139, pp. 611–620, Oct. 1992.
[23] K. Surma-aho and T. Saramäki, "A systematic approach for designing approximately linear phase recursive digital filters," *IEEE Trans. Circuits Syst.*, vol. 46, July 1999.
[24] M. Renfors and E. Zigouris, "Signal processor implementation of digital all-pass filters," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 714–729, May 1988.
[25] T. Saramäki and M. Renfors, "A novel approach for the design of IIR filters as a tapped cascaded interconnection of identical allpass filters," in *Proc. 1987 IEEE Int. Symp. Circuits Syst.*, Philadelphia, PA, pp. 629–632.
[26] S. R. K. Dutta and M. Vidyasagar, "New algorithms for constrained minimax optimization," *Math. Programming*, vol. 13, pp. 140–155, 1977.
[27] J. Yli-Kaakinen, "Optimization of recursive digital filters for practical implementation," Dipl. Eng. M.Sc. thesis, Dept. Elect. Eng., Tampere Univ. Tech., Finland, 1998. [Online]. Available HTTP: http://alpha.cc.tut.fi/~y131373/Thesis.html
[28] J. Yli-Kaakinen and T. Saramäki, "An efficient algorithm for the design of lattice wave digital filters with short coefficient wordlength," in *1999 IEEE Int. Symp. Circuits Syst.*, Orlando, FL, 1999, vol. III, pp. 443–448.
[29] R. Fletcher and M. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. 163–168, 1963.



Juha Yli-Kaakinen was born in Heinola, Finland, on March 2, 1970. He received the Dipl. Eng. degree in electrical engineering from Tampere University of Technology, Tampere, Finland, in 1998.

Since 1995, he has been with the Signal Processing Laboratory, Tampere University of Technology. His research interests are in digital signal processing, especially in digital filter optimization for communication systems and VLSI implementations.



Tapio Saramäki was born in Orivesi, Finland, on June 12, 1953. He has received the degrees of Dipl. Eng. (Hons.) and Doctor of Technology (Hons.) in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1978 and 1981, respectively.

Since 1977, he has held various research and teaching positions at Tampere University of Technology, where he is currently a Professor of Signal Processing and a Docent of Telecommunications. He is also a co-founder and a system-level designer of VLSI Solution Oy, Tampere, Finland, specializing in VLSI implementations of sigma–delta modulators and signal-processing algorithms. In 1982, 1985, 1986, 1990, and 1998, he was a Visiting Research Scholar at the University of California at Santa Barbara. He has written more than 150 international journal and conference articles, six international book chapters, and holds three patents. His research interests are in digital signal processing, especially filter and filter bank design, VLSI implementations, and communications applications, as well as approximation and optimization theories. He is a founding member of the Median-Free Group International.

Dr. Saramäki received the 1987 Guillemin-Cauer Award for the best paper of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.