A SYSTEMATIC ALGORITHM FOR THE DESIGN OF MULTIPLIERLESS LATTICE WAVE DIGITAL FILTERS

Juha Yli-Kaakinen and Tapio Saramäki

Institute of Signal Processing
Tampere University of Technology
P.O. Box 553, FIN-33101 Tampere, Finland
email: {ylikaaki, ts}@cs.tut.fi

ABSTRACT

This paper describes an efficient algorithm for designing multiplierless lattice wave digital filters (parallel connections of two all-pass filters) with short coefficient wordlength. The coefficient optimization is performed in two steps. First, a nonlinear optimization algorithm is used for determining a parameter space of the infinite-precision coefficients including the feasible space where the filter meets the given criteria. The second step involves finding the filter parameters in this space so that the resulting filter meets the given criteria with the simplest coefficient representation forms. Comparisons with some other existing quantization schemes show that the proposed algorithm gives better finite-precision solutions in all examples taken from the literature.

Index Terms—Coefficient quantization, multiplierless design, lattice wave digital filters, parallel connection of all-pass filters.

1 INTRODUCTION

WHEN using a custom or a semi-custom integrated circuit or a programmable logic device for practically implementing a digital filter, the silicon area, the computational complexity, the power consumption, and the maximal achievable sampling rate is highly dependent on the coefficient wordlength. Therefore, the wordlength should be as short as possible but still sufficient to satisfy the filter specifications. In addition, in highly customized very large-scale integration (VLSI) implementations, the general multiplier element is very costly. Therefore, it is beneficial to carry out the multiplication of a data sample by each filter coefficient value using a sequence of shifts and adds or subtracts. The shifts are often hardwired and, therefore, essentially free. Thus, only a few adders or subtracters are required for implementing each coefficient. Such an implementation is usually called "multiplierless".

In order to generate multiplierless filter implementations, it is very essential that a digital filter is realized using a low-sensitivity structure being very insensitive to variations in the filter coefficients. The importance of such a structure is that if the effect of the coefficient value deviation from the ideal value is small, then short coefficient wordlengths can be used with only slightly violating the infinite-precision filter specifications, resulting in a faster, smaller, and less expensive hardware [1].

One of the best structures for implementing recursive digital filters are the lattice wave digital (LWD) filters [2]–[4] that are related to certain analog prototype networks. An LWD filter consists of parallel connection of two all-pass filters. This filter class is characterized by many attractive properties, such as a reasonably low coefficient sensitivity, a low roundoff noise level, and the absence of parasitic oscillations. Moreover, the number of multipliers required in the implementation is directly the filter order, unlike in some other implementation forms, such as in the canonic direct-form realizations requiring approximately twice the number of multipliers. In addition, these all-pass subfilters can be realized by using first- and second-order sections as basic building blocks. The resulting filter structures are highly modular, thereby making them suitable for VLSI implementations [5], [6].

This paper describes an efficient algorithm for designing LWD filters with short coefficient wordlength. This algorithm is based on the following observation: Finding the smallest and largest values for both the radius and the angle of all the complex-conjugate pole pairs and the smallest and largest values for the radius of a possible real pole so that the given criteria are still met enables one to find a parameter space including the feasible space where the filter specifications are satisfied. After determining this larger space, all what is needed is to check whether in this space there exist the desired discrete values for the coefficient representations. This strategy is general but particularly efficient for filters implemented as a parallel connection of two all-pass filters due to the fact that for these filters only the denominator coefficients of the all-pass sections have to be quantized. Several examples taken from the literature are included illustrating that in all the examples the proposed quantization scheme results in a better finite-precision solution than other existing quantization techniques.

2 LATTICE WAVE DIGITAL FILTERS

The transfer function for the LWD filter can be expressed as

$$H(z) = \frac{1}{2}[A_1(z) + A_2(z)], \tag{1}$$

where $A_1(z)$ and $A_2(z)$ are real stable all-pass filters of orders M and N, respectively. This contribution concentrates on designing low-pass filters. In this case, $M = N-1$ or $M = N+1$ so that $M+N$, the overall order of $H(z)$, is odd.

If $A_1(z)$ and $A_2(z)$ are implemented as a cascade of first- and second-order wave digital all-pass structures and M and N are assumed to be an odd and even integer, respectively, then $A_1(z)$ and $A_2(z)$ are expressible in terms of the adaptor coefficients as follows (see, e.g., [4]):

$$A_1(z) = \frac{-\gamma_0 + z^{-1}}{1 - \gamma_0 z^{-1}} \prod_{l=1}^{m} \frac{-\gamma_{2l-1} + \gamma_{2l}(\gamma_{2l-1} - 1)z^{-1} + z^{-2}}{1 + \gamma_{2l}(\gamma_{2l-1} - 1)z^{-1} - \gamma_{2l-1}z^{-2}} \tag{2a}$$

and

$$A_2(z) = \prod_{l=m+1}^{m+n} \frac{-\gamma_{2l-1} + \gamma_{2l}(\gamma_{2l-1} - 1)z^{-1} + z^{-2}}{1 + \gamma_{2l}(\gamma_{2l-1} - 1)z^{-1} - \gamma_{2l-1}z^{-2}}, \tag{2b}$$

where $m = (M-1)/2$ and $n = N/2$.

If $A_1(z)$ possesses a real pole at $z = r_0$ and m complex-conjugate pole pairs at $z = r_l \exp(\pm j\theta_l)$ for $l = 1, 2, \ldots, m$ and $A_2(z)$ possesses n complex-conjugate pole pairs at $z = r_l \exp(\pm j\theta_l)$ for $l = m+1, m+2, \ldots, m+n$, then

$$\gamma_0 = r_0, \quad \text{whereas} \quad \gamma_{2l-1} = -r_l^2 \quad \text{and} \quad \gamma_{2l} = \frac{2r_l \cos\theta}{1 + r_l^2}, \tag{3}$$

for $l = 1, 2, \ldots, m+n$.

This work was supported by the Academy of Finland, project No. 44876 (Finnish Centre of Excellence Program (2000–2005)). Juha Yli-Kaakinen was also financed by a postdoctoral research grant from the Academy of Finland, project No. 75492.

3 STATEMENT OF THE PROBLEM

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

$$\Phi = [r_0, r_1, \ldots, r_{m+n}, \theta_1, \theta_2, \ldots, \theta_{m+n}]. \tag{4}$$

Given the passband and stopband edges ω_p and ω_s, respectively, as well as the passband and stopband ripples δ_p and δ_s, respectively, the magnitude 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 can be expressed as[1]

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

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

where

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

with

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

The stability of the resulting filter is guaranteed if the poles of the allpass sections $A_1(z)$ and $A_2(z)$ lie inside the unit circle, that is, it is required that

$$|r_0| < 1 \quad \text{and} \quad r_l < 1 \quad \text{for } l = 1, 2, \ldots, m+n. \tag{7}$$

This contribution concentrates on 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 or subtracts. For such a purpose, it is desired to express the coefficient values in the form

$$\textstyle\sum_{r=1}^{R} a_r 2^{-P_r}, \tag{8}$$

where each a_r is either 1 or -1 and the P_r's are nonnegative integers in the increasing order. In this case, the goal is to find all the coefficient values so that: 1) R, the number of powers-of-two terms, is made as small as possible, 2) P_R, the maximum number of shifts, is made as small as possible. For this purpose, it is attractive to use the canonic-signed-digit (CSD) representation. This representation is characterized by the fact that no two consecutive digits a_r are both nonzero, that is, for the minimal R, $a_r a_{r+1} = 0$ for $r = 1, 2, \ldots, R-1$. The number of adders and subtracters required to realize a CSD coefficient is one less than the number of nonzero digits in this coefficient representation form.

An estimate for the implementation cost of the filter is the number of adders and subtracters required to implement all the adaptor coefficients, that is, the implementation cost is given by

$$\textstyle\sum_{l=0}^{2(m+n)} \sigma_l, \tag{9}$$

where the σ_l's are the number of adders and subtracters required to implement the γ_l's.

The optimization problem under consideration is the following:

Optimization Problem: Given ω_p, ω_s, δ_p, and δ_s, find M and N, and the parameter vector Φ in such a manner that, first, the criteria of Eq. (5) [or Eq. (6)] and Eq. (7) are met after quantizing the adaptor coefficient values corresponding to the parameters included in Φ to achieve the above-mentioned form for their representations and, then, the implementation cost, as given by Eq. (9), is minimized.

[1]These specifications are typical of most recursive filters built using allpass filters as building blocks. In these cases, the filter structure constrains the maximum of the magnitude response to be unity.

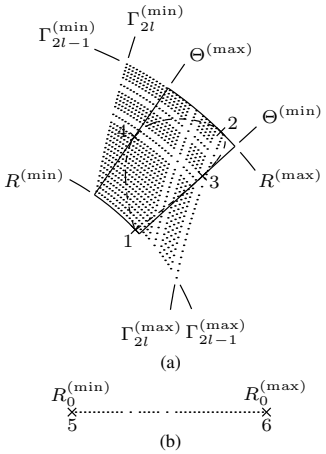

Fig. 1 Typical search spaces for the poles when three powers-of-two terms with seven fractional bits ($R = 3$ and $P_R = 7$) are used for the adaptor coefficient representations. (a) Upper-half-plane pole for the complex-conjugate pole pair. (b) Real pole.

4 FILTER OPTIMIZATION

The solution to the stated optimization problem can be found in the following two steps. In the first step, the smallest and largest values are determined for each adjustable parameter by reoptimizing the remaining unknowns in the parameter vector in such a manner that the given specifications are met. This enables one to find the parameter space of the infinite-precision coefficients including the feasible space where the filter meets the specifications. The second step involves finding the filter parameters in this space so that the resulting filter meets the given criteria with the simplest coefficient representation forms.

4.1 Optimization of Infinite-Precision Filters

It has turned out that a very straightforward quantization scheme for the filter coefficients is obtained as follows: For each complex-conjugate pole pair, the smallest and largest values for both the radius and the angle are determined so that by reoptimizing the locations of the remaining poles the given overall criteria, as given by Eq. (5) [or Eq. (6)] and Eq. (7) can still be met. For the 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 of each complex-conjugate pole pair $r_l \exp(\pm j\theta_l)$ for $l = 1, 2, \ldots, m+n$ 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. 1(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. 1(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 the real pole r_0, there exists the corresponding region $R_0^{(\min)} \le R \le R_0^{(\max)}$ that is simultaneously the feasible region. In Fig. 1(b) the crosses numbered by 5 and 6 indicate $R_0^{(\min)}$ and $R_0^{(\max)}$, respectively.

For the complex-conjugate pole pairs, the larger region is used since it can be found by applying only four times the algorithm to be described next. In order to find the above-mentioned regions for all the poles, there are $2 + 4(m+n)$ problems of the following form: Find the adjustable parameter vector Φ to minimize ψ subject to the conditions of Eq. (6) and Eq. (7). For these problems, ψ is r_0 and $1 - r_0$ for the real pole, whereas for the complex-conjugate pole pairs, ψ is selected to be r_l, $1 - r_l$, θ_l, and $\pi - \theta_l$ for $l = 1, 2, \ldots, m+n$. In order to prevent the poles from changing their ordering, e.g., to inhibit the outermost pole pair from becoming the second outermost pole pair when minimizing its radius, an additional constraint of

Table 1 Optimized parameter vectors $\Phi^{(k)}$ for $k = 1, 2, \ldots, 14$ in the example of Subsection 5.1.

	$\Phi^{(1)}$	$\Phi^{(2)}$	$\Phi^{(3)}$	$\Phi^{(4)}$	$\Phi^{(5)}$	$\Phi^{(6)}$	$\Phi^{(7)}$	$\Phi^{(8)}$	$\Phi^{(9)}$	$\Phi^{(10)}$	$\Phi^{(11)}$	$\Phi^{(12)}$	$\Phi^{(13)}$	$\Phi^{(14)}$
ψ	r_0	$1-r_0$	r_1	r_2	r_3	$1-r_1$	$1-r_2$	$1-r_3$	θ_1	θ_2	θ_3	$\pi-\theta_1$	$\pi-\theta_2$	$\pi-\theta_3$
r_0	0.2441	0.5771	0.4419	0.2441	0.4773	0.5332	0.5528	0.5332	0.2441	0.4594	0.5332	0.4951	0.4422	0.5036
r_1	0.7823	0.8325	0.7616	0.7823	0.8008	0.8560	0.8440	0.8560	0.7823	0.8162	0.8560	0.8259	0.7616	0.8526
r_2	0.4529	0.6718	0.5734	0.4529	0.6234	0.6741	0.6750	0.6741	0.4529	0.5975	0.6741	0.6508	0.5736	0.6595
r_3	0.9450	0.9520	0.9397	0.9450	0.9364	0.9618	0.9555	0.9618	0.9450	0.9497	0.9618	0.9382	0.9397	0.9615
θ_1	0.3611π	0.3620π	0.3816π	0.3611π	0.3935π	0.3707π	0.3653π	0.3707π	0.3611π	0.3615π	0.3707π	0.3997π	0.3816π	0.3951π
θ_2	0.2429π	0.2382π	0.3076π	0.2429π	0.2983π	0.2632π	0.2511π	0.2632π	0.2429π	0.2358π	0.2632π	0.2904π	0.3076π	0.2843π
θ_3	0.4079π	0.4064π	0.4099π	0.4079π	0.4163π	0.4052π	0.4059π	0.4052π	0.4079π	0.4069π	0.4052π	0.4215π	0.4099π	0.4307π

the following form:

$$r_1 \le r_{m+1} \le r_2 \le r_{m+2} \le \cdots \le r_m \le r_{m+n} \tag{10}$$

is required.

To solve these problems, the passband and stopband regions are discretized into the frequency points $\omega_i \in [0,\, \omega_p]$, $i = 1, 2, \ldots, L_p$ and $\omega_i \in [\omega_s,\, \pi]$, $i = L_p + 1, L_p + 2, \ldots, L_p + L_s$. The resulting discrete minimization problem is to find Φ to minimize ψ subject to

$$E(\Phi,\, \omega_i) - 1 \le 0 \quad \text{for } i = 1, 2, \ldots, L_p + L_s \tag{11a}$$

$$E(\Phi,\, \omega_i) \le 0 \quad \text{for } i = 1, 2, \ldots, L_p \tag{11b}$$

and the constraints of Eqs. (7) and (10).

The above-mentioned problems can be solved conveniently by using the second algorithm of Dutta and Vidyasagar [7] or the function `fmincon` from the optimization toolbox provided by MathWorks, Inc. [8]. For more details, see [9].

Based on these smallest and largest values, the smallest and largest values for the adaptor coefficients can be determined as follows:

$$\begin{aligned}
\gamma_0^{(\min)} &= -r_0^{(\max)}, & \gamma_0^{(\max)} &= -r_0^{(\min)}, \\
\gamma_{2l-1}^{(\min)} &= -\big(r_l^{(\max)}\big)^2, & \gamma_{2l-1}^{(\max)} &= -\big(r_l^{(\min)}\big)^2, \\
\gamma_{2l}^{(\min)} &= \frac{2r_l^{(\min)} \cos\theta_l^{(\max)}}{1 + \big(r_l^{(\min)}\big)^2}, \text{ and } & \gamma_{2l}^{(\max)} &= \frac{2r_l^{(\min)} \cos\theta_l^{(\min)}}{1 + \big(r_l^{(\max)}\big)^2}
\end{aligned} \tag{12}$$

for $l = 1, 2, \ldots, m + n$. The smallest and largest values for the fixed-wordlength adaptor coefficients corresponding to the innermost complex-conjugate pole-pair are also depicted in Fig. 1.

4.2 Optimization of Finite-Precision Filters

It has been experimentally proved that the parameter space defined above forms a space including the feasible space where the filter specifications are satisfied. After finding this larger space, all what is needed is to check whether in this space there exist combinations of the discrete pole positions with which the given overall criteria are met.

This search can be done in a straightforward manner by first finding the sets of CSD numbers Γ_l for $l = 0, 1, \ldots, M+N-1$ between the smallest and largest values of each adaptor coefficient, i.e., for $l = 0, 1, \ldots, M + N - 1$

$$\left\{ \Gamma_l \in \mathbb{CSD}_{(R,P_R)} \;\middle|\; \gamma_l^{(\min)} \le \Gamma_l \le \gamma_l^{(\max)} \right\}, \tag{13}$$

where $\mathbb{CSD}_{(R,P_R)}$ denotes the space of CSD numbers for R, the given maximum number of powers-of-two terms and P_R, the maximum number of fractional bits [cf. Eq. (8)]. The magnitude response is then evaluated for each combination of the Γ_l's to check whether the filter meets the given specifications.

In Figure 1, the dots indicate the allowable locations for the poles when three powers-of-two terms with seven fractional bits are used for the adaptor coefficient representations. Note that the distributions are highly irregular for a few powers-of-two terms due to the desired coefficient representation form. However, as can be seen from this figure, there are, particularly for the innermost complex-conjugate pole, regions where the angle of the pole corresponding to quantized values of γ_{2l-1} and γ_{2l} is smaller than $\Theta^{(\min)}$ or larger than $\Theta^{(\max)}$. For this reason, it is advisable to check if the angle of the discrete pole is in prespecified region in order to avoid the vain evaluation of the corresponding magnitude response.

It should be pointed out that for a certain wordlength, there are typically several solutions which will meet the magnitude specifications. Therefore, it is advisable to find first all the solutions and then to choose among them the one with the best attenuation characteristics or the minimum number of adders required to implement all the multiplier coefficients for the given wordlength.

5 NUMERICAL EXAMPLES

The purpose of this section is twofold. First, the performance of the proposed quantization scheme is illustrated by means of an example. Second, comparisons with some other existing quantization schemes show that the proposed algorithm gives better finite-precision solutions in all examples taken from the literature.

5.1 Illustrative Example

This example is included to illustrate the performance of the proposed overall synthesis scheme. It is desired to design a low-pass filter with the passband and stopband edges at $\omega_p = 0.4\pi$ and at $\omega_p = 0.5\pi$, respectively. The maximum allowable passband ripple and the required stopband attenuation are 0.2 dB ($\delta_p = 0.0228$) and 60 dB ($\delta_s = 10^{-3}$), respectively. The minimum odd-order of an elliptic filter to meet the given amplitude criteria is seven.[2]

The optimized parameter vectors $\Phi^{(k)}$ for $k = 1, 2, \ldots, 14$ after the infinite-precision optimization of Subsection 4.1 are shown in Table 1. In this table, $\Phi^{(1)}$ and $\Phi^{(2)}$ are the optimized parameter vectors for $\psi = r_0$ or $\psi = 1 - r_0$, respectively, $\Phi^{(k)}$ for $k = 3, 4, \ldots, 8$ are the optimized solutions for $\psi = r_l$ or $\psi = 1 - r_l$ for $l = 1, 2, 3$, respectively, and $\Phi^{(k)}$ for $k = 9, 10, \ldots, 14$ are the optimized solutions for $\psi = \theta_l$ or $\psi = \pi - \theta_l$ for $l = 1, 2, 3$, respectively. The corresponding smallest and largest values for the adaptor coefficients $\gamma_l^{(\min)}$ and $\gamma_l^{(\max)}$ for $l = 0, 1, \ldots, 6$, derived using Eq. (12), are shown in Table 2. The overall CPU-time required for solving all these infinite-precision optimization problems is approximately 51 seconds when using a MATLAB code running on a 500 MHz AlphaServer DS20 with $L_p = L_s = 100$.

For this filter, all the coefficient values can be represented as two or three powers-of-two terms, that is, R, the maximum number of powers-of-two terms, is three, whereas seven fractional bit ($P_R = 7$) are required to meet the magnitude specifications.[3] The permissible discrete coefficient values between $\gamma_l^{(\min)}$ and $\gamma_l^{(\max)}$ for $l = 0, 1, \ldots, 6$ are also shown in Table 2. In this case, the number of discrete values between the smallest and largest values of the adaptor coefficients for the selected CSD coefficient representation form are 38, 28, 27, 9, 11, 4, and 10, respectively, that is, the overall number of coefficient value combinations is approximately $114 \cdot 10^6$.

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

[3] In this case, six fractional bits is the shortest wordlength for which there exist discrete values between all the smallest and largest values of the adaptor coefficients. However, for this coefficient wordlength, there is no solution satisfying the specifications.

Table 2 Smallest and largest values for the infinite-precision and finite-precision adaptor coefficients in the example of Subsection 5.1.

l	$\gamma_l^{(\min)}$	$\gamma_l^{(\max)}$	$\Gamma_l \cdot 2^7$	$\gamma_l^{(\text{opt})}$
0	0.244 130	0.577 150	{32–42, 44, 46–50, 52, 54–73}	$60 \cdot 2^{-7}$
1	−0.732 802	−0.580 039	−{92, 88, 84, 82–78, 76}	$-82 \cdot 2^{-7}$
2	0.304 435	0.413 204	{39–42, 44, 46–50, 52}	$44 \cdot 2^{-7}$
3	−0.455 581	−0.205 127	−{58–54, 52, 50–46, 44, 42–27}	$-48 \cdot 2^{-7}$
4	0.490 484	0.678 424	{55–74, 76, 78–82, 84}	$69 \cdot 2^{-7}$
5	−0.924 996	−0.876 865	−{118, 116, 114, 113}	$-114 \cdot 2^{-7}$
6	0.215 922	0.293 096	{28–37}	$34 \cdot 2^{-7}$

A total of only eleven adders and/or subtracters are required to implement all the multipliers for this coefficient representation form.[4] In this case, there is only one solution meeting the magnitude specifications for the given coefficient representation form. The optimized adaptor coefficient values, denoted by the $\gamma_l^{(\text{opt})}$'s, are also shown in Table 2. The CPU time required when using a Fortran 95 program on a 500 MHz AlphaServer DS20 to arrive at this solution with $L_p = L_s = 40$ was 13 min.

5.2 Comparisons with Other Quantization Algorithms Resulting in Multiplierless Overall Filters

This subsection compares the performance of the proposed quantization algorithm for designing multiplierless LWD filters in four low-pass examples taken from the literature. In all examples, the criteria are stated in terms of the passband edge ω_p, the stopband edge ω_s, A_p, the passband ripple in decibels, that is, $A_p = -20\log_{10}(1-\delta_p)$, and the minimum stopband attenuation $A_s = -20\log_{10}\delta_s$. Also, R, the maximum allowed number of powers-of-two terms for each adaptor coefficient, and P_R, the number of fractional bits, are given for both the reference designs and filters resulting when using the proposed optimization algorithm.

Example 1: Consider the specifications [10] $\omega_p = 0.27\pi$, $\omega_s = 0.4\pi$, $A_p = 0.2$ dB, and $A_s = 30$ dB. The minimum odd-order of an elliptic filter to meet the specifications is five. For the finite-precision filter optimized in [10], $P_R = 7$ and $R = 4$ are used for the coefficient representations to meet the given criteria. For this finite-precision filter, six adders and/or subtracters are required to implement all the adaptor coefficients. For the finite-precision filter optimized using the proposed algorithm, only three adders and/or subtracters are needed. For this design, $R = 2$ and $P_R = 4$.

Example 2: Consider the half-band filter specifications [10] $\omega_p = 0.44\pi$ and $A_s = 46$ dB. Due to the properties of half-band IIR filters, $\omega_s = \pi - \omega_p = 0.56\pi$ and $\delta_p \approx \delta_s^2/2$, giving $A_p = 1.1 \cdot 10^{-4}$ dB (see, e.g., [11]). For the optimized finite-precision filter of order nine in [10], $P_R = 8$, $R = 4$, and the number of adders and/or subtracters needed to implement all the coefficients is eight. The proposed algorithm results in the optimized filter with $P_R = 8$, $R = 3$, and requiring only five adders and/or subtracters to implement all the adaptor coefficients.

Example 3: In [12], [13], it was desired to design a fifth-order low-pass filter having a 0.125-dB passband ripple on $[0, 0.375\pi]$ while the required stopband attenuation is 14 dB on $[0.5\pi, 0.575\pi]$ and 32 dB on $[0.575\pi, \pi]$. For the finite-precision design optimized in [12], [13], $P_R = 5$, $R = 3$, and four adders and/or subtracters are required to meet the specifications. For the filter resulting when using the proposed algorithm, only two adders and/or subtracters are needed for the same $P_R = 5$ when $R = 2$.

Example 4: Consider the specifications [14] $\omega_p = 0.4125\pi$, $\omega_s = 0.575\pi$, $A_p = 0.045$ dB, and $A_s = 44$ dB. It was claimed in [14], that the number of adders required to implement all the adaptor coefficients is six when $P_R = 7$ and $R = 3$. However, by evaluating the magnitude response using the given adaptor coefficient values, it can be observed that the resulting passband ripple is approximately 0.063 dB. Hence, the quantized filter does not meet the specifications. For the filter resulting when using the proposed algorithm, ten adders and/or subtracters with $P_R = 7$ are required to meet the specifications for $R = 4$.

[4] If the adaptors shown in Fig. 9 in [4] are used for implementing the first- and second-order sections, then the total number of adders becomes eight.

Table 3 Comparison with other quantization algorithms.

Example	$M+N$	Reference		Proposed				
		P_R	N_A	P_R	N_A	A_p (dB)	A_s (dB)	CPU
1	5	7	6	4	3	$8.2 \cdot 10^{-3}$	30.6	6 s
2	9	8	7	8	5	$8.5 \cdot 10^{-4}$	47.1	2 s
3	5	5	4	5	2	$4.6 \cdot 10^{-2}$	26.4/34.9	3 s
4[a]	5	7	6	7	10	$4.3 \cdot 10^{-2}$	44.3	1 s

[a] The given criteria are not met (see Example 4).

Summary: The characteristics of the multiplierless filters designed using various algorithm are summarized in Table 3. Here, N_A is the number of adders and/or subtracters required to meet the specifications. CPU denotes the computer time required by the finite-precision optimization with $L_p = L_s = 40$ for a Fortran 95 program running a on 500 MHz AlphaServer DS20. As seen from this table, N_A for the proposed quantization scheme is in all cases considerably smaller than that being achievable by using other existing algorithms. A MATLAB m-file containing the optimized finite-precision coefficient values and for evaluating the corresponding magnitude responses can be downloaded from `http://alpha.cc.tut.fi/~ylikaaki/ISCCSP04LWD/`.

REFERENCES

[1] L. Wanhammar, *DSP Integrated Circuits*. New York: Academic, 1998.

[2] A. Fettweis, H. Levin, and A. Sedlmeyer, "Wave digital lattice filters," *Int. J. Circuit Theory Appl.*, vol. 2, pp. 203–211, June 1974.

[3] A. Fettweis, "Wave digital filters: Theory and practice," *Proc. IEEE*, vol. 74, pp. 270–327, Feb. 1986.

[4] L. Gazsi, "Explicit formulas for lattice wave digital filters," *IEEE Trans. Circuits Syst.*, vol. CAS-32, pp. 68–88, Jan. 1985.

[5] T. Saramäki and T. Ritoniemi, "Optimization of digital filter structures for VLSI implementation," *Automatica*, vol. 34, pp. 111–116, 1993.

[6] L. D. Milić and M. D. Lutovac, "Design of multiplierless elliptic IIR filters with a small quantization error," *IEEE Trans. Signal Processing*, vol. 47, pp. 469–479, Feb. 1999.

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

[8] T. Coleman, M. A. Branch, and A. Grace, *Optimization Toolbox User's Guide*, The MathWorks, Inc., Jan. 1999, Version 2.

[9] T. Saramäki and J. Yli-Kaakinen, "Design of digital filters and filter banks by optimization: Applications," Tampere International Center for Signal Processing, Tech. Rep. No. 15, Apr. 2002. [Online]. Available: http://alpha.cc.tut.fi/~ylikaaki/TICSP02.pdf

[10] M. D. Lutovac and L. Milić, "Design of computationally efficient elliptic IIR filters with a reduced number of shift-and-add operations in multipliers," *IEEE Trans. Signal Processing*, vol. 45, pp. 2422–2430, Oct. 1997.

[11] W. Wegener, "Wave digital directional filters with reduced number of multipliers and adders," *Int. J. Electron. Commun. (AEÜ)*, vol. 33, pp. 239–243, June 1979.

[12] F. Catthoor, J. Vandewalle, and H. De Man, "Simulated-annealing-based optimization of coefficient and data word-lengths in digital filters," *Int. J. Circuit Theory Appl.*, vol. 16, pp. 371–390, 1988.

[13] L. Claesen, F. Catthoor, D. Lanneer, G. Goosens, S. Note, J. V. Meerbergen, and H. De Man, "Automatic synthesis of signal processing benchmark using the CATHEDRAL silicon compilers," in *Proc. IEEE Conf. Custom Integrated Circuits*, May 16–19 1988, pp. 14.7/1–14.7/4.

[14] U. Kaiser, "A CAD system for the design of digital filter algorithms for the reduced-instruction set digital signal processor RISP," in *Proc. IEEE Int. Symp. Circuits Syst.*, Singapore, June 11–14 1991, pp. 45–48.