Improved Design Method for Nearly Linear-Phase IIR Filters Using Constrained Optimization

Rajeev C. Nongpiur, *Member, IEEE*, Dale J. Shpak, *Senior Member, IEEE*, and Andreas Antoniou, *Life Fellow, IEEE*

Abstract—A new optimization method for the design of nearly linear-phase IIR digital filters that satisfy prescribed specifications is proposed. The group-delay deviation is minimized under the constraint that the passband ripple and stopband attenuation are within the prescribed specifications and either a prescribed or an optimized group delay can be achieved. By representing the filter in terms of a cascade of second-order sections, a non-restrictive stability constraint characterized by a set of linear inequality constraints can be incorporated in the optimization algorithm. An additional feature of the method, which is very useful in certain applications, is that it provides the capability of constraining the maximum gain in transition bands to be below a prescribed level. Experimental results show that filters designed using the proposed method have much lower group-delay deviation for the same passband ripple and stopband attenuation when compared with corresponding filters designed with several state-of-the-art competing methods.

Index Terms—Delay equalization of filters, design of filters by optimization, IIR filter design, nearly linear-phase filters.

I. INTRODUCTION

LINEAR-PHASE filters are important in applications where a flat delay characteristic is necessary. Applications of such filters can be found in the field of audio signal processing where it is important to minimize the relative delay between the frequency components to prevent perceptible audio distortion and in digital communications where nonlinear phase can result in signal spreading thereby causing inter-symbol interference between time-concentrated information symbols. Perfectly linear-phase filters can be easily achieved using FIR filters. However, in most applications a perfectly linear phase response is not required and filters that have approximately linear phase response are quite acceptable. In such cases, IIR filters are more attractive than FIR filters for two main reasons. Firstly, they can satisfy the given filter specifications with a much lower filter order thereby reducing the computational requirement and/or the complexity of hardware and, secondly, they have a much smaller group delay.

The presence of the denominator polynomial in IIR filters makes their design more challenging than that of FIR filters because it results in a highly nonlinear objective function that requires highly sophisticated optimization methods. As IIR filters lack the inherent stability of FIR filters, stability constraints must be incorporated in the design process to ensure that the filter is stable, which means constraining the poles to lie within the unit circle of the z plane.

In applications where the phase response is not important, a fairly large choice of methods is available to the filter designer ranging from closed-form methods based on classical analog filter approximations to numerous optimization methods. In [1], unconstrained algorithms of the quasi-Newton family are used in a least-pth formulation. Filter stability is achieved by means of a well-known stabilization technique whereby poles outside the unit circle are replaced by their reciprocals. MATLAB™ function *iirlpnorm* in [2] implements an unconstrained least-pth quasi-Newton algorithm of the type found in [3]. On the other hand, MATLAB™ function *iirlpnormc* in [2] implements a least-pth Newton method that uses barrier constraints to assure the stability of the resulting filter and also provides for a specified stability margin.

Nearly linear-phase IIR filters can be designed by using the classical *equalizer approach* whereby an IIR filter satisfying prescribed amplitude-response specifications is designed and is then cascaded with an IIR delay equalizer to approximately linearize the phase response [4]. More recent methods typically involve designing IIR filters that simultaneously satisfy both the amplitude- and phase-response constraints, as it results in filters of lower order [5]–[14]. In [8], the frequency-response error is minimized under the constraint that the group-delay deviation is within a prescribed level, while in [5] the design problem is formulated as a cascade of filter sections where each section is represented by a biquadratic transfer function either in the conventional polynomial form or in the polar form and the design problem is solved using a constrained Newton method. An important advantage of this method is the nonrestrictive stability constraint that it uses coupled with the capability of controlling the maximum pole radius. Since the objective function used in [5] is nonconvex, the performance of the algorithm is critically dependent on the initialization point. To overcome difficulties with the nonlinear objective function, iterative methods based on the Steiglitz-McBride (SM) and the Gauss-Newton (GN) algorithms have been employed [15]–[17]. Although convergence cannot be guaranteed in these methods, solutions that are satisfactory for the intended applications are usually obtained. A drawback of the methods in [15]–[17] is the use of restrictive stability constraints that could prevent better designs.

In [18], a minimax design of IIR filters is formulated as a conic quadratic programming problem by approximating each update as a linear approximation step. An advantage

of the method is that it incorporates a nonrestrictive stability constraint using a set of linear inequality constraints; another important advantage is that the final solution is not critically dependent on the initialization point. However, the method has certain drawbacks: it requires the group delay to be specified and, furthermore, it does not provide the option of controlling the relative degree of minimization between the group-delay deviation and the maximum passband ripple.

In this paper, we propose a design method where the group-delay deviation is minimized under the constraints that the passband ripple and stopband attenuation are maintained within the prescribed specifications. The method also provides the option of keeping the group delay fixed at a prescribed value or allowing it to be a free parameter that can be optimized in order to achieve improved filter performance. By designing the filter as a cascade of second-order sections, a necessary and sufficient nonrestrictive stability constraint can be incorporated in the optimization problem as a set of linear inequality constraints. The method eliminates the drawbacks of the method in [18] as it does not require the group delay to be specified, while at the same time it provides the capability of controling the relative quality between the group-delay deviation and the maximum passband ripple. Although the transition region is usually treated as a "don't care" region, in many practical applications excessive gain in this region could be undesirable. Our method provides the capability of controlling the maximum gain in transition bands. Experimental results show that numerous filters designed with the proposed method have much lower group-delay deviation for the same passband ripple and stopband attenuation than corresponding filters designed with several known competing state-of-the-art methods.

The paper is organized as follows. In Section II, we frame the problem as an iterative constrained optimization problem. Then, in Section III we describe a procedure for designing nearly linear-phase IIR filters. In Section IV, performance comparisons between the filters designed using the proposed method and the existing methods are carried out. Conclusions are drawn in Section V.

II. THE OPTIMIZATION PROBLEM

In this section, we frame the design problem at hand as a constrained optimization problem. To this end, we derive formulations for the stability constraints, group-delay deviation, passband ripple, stopband attenuation, and transition-band gain constraints. Then, we incorporate the formulations within the framework of a constrained optimization problem.

We assume that the filter comprises a cascade of second-order sections (SOSs), which can be represented by a product of biquadratic transfer functions of the form

$$H(\mathbf{c}, z) = H_0 \prod_{m=1}^{J} \frac{a_{0m} + a_{1m}z + z^2}{b_{0m} + b_{1m}z + z^2}$$

$$= H_0 \prod_{m=1}^{J} \frac{N_m(z)}{D_m(z)} \tag{1}$$

where

$$\mathbf{c} = [a_{01}\ a_{11}\ b_{01}\ b_{11}\cdots b_{0J}\ b_{1J}\ H_0]^T \tag{2}$$

J is the number of filter sections, $N = 2J$ is the filter order, and H_0 is a positive multiplier constant. An odd-order transfer function can be readily obtained by setting coefficients a_{0m} and b_{0m} to zero in one SOS.

A. Group-Delay Deviation

The group delay corresponding to transfer function $H(\mathbf{c}, z)$ in (1) is given by [4]

$$\tau_h(\mathbf{c}, z) = -\sum_{i=1}^{J} \frac{\alpha_n(\mathbf{c}, z, i)}{\beta_n(\mathbf{c}, z, i)} + \sum_{i=1}^{J} \frac{\alpha_d(\mathbf{c}, z, i)}{\beta_d(\mathbf{c}, z, i)} \tag{3}$$

where

$$\alpha_n(\mathbf{c}, z, i) = 1 - a_{0i}^2 + a_{1i}(1 - a_{0i})\cos\omega \tag{4}$$

$$\beta_n(\mathbf{c}, z, i) = a_{0i}^2 + a_{1i}^2 + 1 + 2a_{0i}(2\cos^2\omega - 1)$$
$$+ 2a_{1i}(a_{0i} + 1)\cos\omega \tag{5}$$

$$\alpha_d(\mathbf{c}, z, i) = 1 - b_{0i}^2 + b_{1i}(1 - b_{0i})\cos\omega \tag{6}$$

$$\beta_d(\mathbf{c}, z, i) = b_{0i}^2 + b_{1i}^2 + 1 + 2b_{0i}(2\cos^2\omega - 1)$$
$$+ 2b_{1i}(b_{0i} + 1)\cos\omega \tag{7}$$

The group-delay deviation at frequency ω is given by

$$e_g(\mathbf{x}, e^{j\omega}) = \tau_h(\mathbf{c}, e^{j\omega}) - \tau \tag{8}$$

where

$$\mathbf{x} = [\mathbf{c}^T \tau]^T \tag{9}$$

and τ is the group delay, which may be an optimization variable. If $\mathbf{x}_k$ is the value of $\mathbf{x}$ at the start of the kth iteration and $\boldsymbol{\delta}$ is the update to $\mathbf{x}_k$, the updated value of the group-delay deviation can be estimated by using the linear approximation

$$e_g(\mathbf{x}_k + \boldsymbol{\delta}, e^{j\omega}) \approx e_g(\mathbf{x}_k, e^{j\omega}) + \nabla e_g(\mathbf{x}_k, e^{j\omega})^T \boldsymbol{\delta} \tag{10}$$

which becomes more accurate as $\|\boldsymbol{\delta}\|_2$ gets smaller.

If ω_{pl} and ω_{ph} are the lower and upper edges of the passband, the L_p-norm of the passband group-delay deviation for the kth iteration is given by

$$\mathbf{E}_p^{(gd)}(k) = \left[\int_{\omega_{pl}}^{\omega_{ph}} |e_g(\mathbf{x}_{k+1}, e^{j\omega})|^p d\omega\right]^{\frac{1}{p}}$$

$$\approx \kappa_g \left[\sum_{i=1}^{N} |e_g(\mathbf{x}_{k+1}, e^{j\omega_i})|^p\right]^{\frac{1}{p}}, \ \omega_i \in \Psi_p$$

$$\approx \left[\sum_{i=1}^{N} |\kappa_g e_g(\mathbf{x}_k, e^{j\omega_i})\right.$$
$$\left. + \kappa_g \nabla e_g(\mathbf{x}_k, e^{j\omega_i})^T \boldsymbol{\delta}|^p\right]^{\frac{1}{p}} \tag{11}$$

where $\Psi_p \in [\omega_{pl}, \omega_{ph}]$ is the set of passband frequency sample points and κ_g is a constant. Expressing (11) in matrix form, we get

$$\mathbf{E}_p^{(gd)}(k) \approx \|\mathbf{C}_k\boldsymbol{\delta} + \mathbf{d}_k\|_p \tag{12}$$

where

$$
\mathbf{C}_k = \begin{bmatrix} \kappa_g \nabla e_g(\mathbf{x}_k, e^{j\omega_1})^T \\ \vdots \\ \kappa_g \nabla e_g(\mathbf{x}_k, e^{j\omega_{N_p}})^T \end{bmatrix} \tag{13}
$$

$$
\mathbf{d}_k = [d_1 \ d_2 \ \cdots \ d_{N_p}]^T, \tag{14}
$$

$$
d_i = \kappa_g e_g(\mathbf{x}_k, e^{j\omega_i}), \quad \omega_i \in \Psi_p. \tag{15}
$$

The right-hand side of (12) is the L_p-norm of an affine function of $\boldsymbol{\delta}$ and, therefore, it is convex with respect to $\boldsymbol{\delta}$ [1].

The quality of the group-delay characteristic of the filter can be measured by using the normalized maximum variation of the filter group delay, τ_h, over the passband as a percentage, i.e., [3]

$$
Q_\tau = \frac{100(\tau_{\max} - \tau_{\min})}{2\tau_{avg}} \tag{16}
$$

where

$$
\tau_{avg} = \frac{\tau_{\max} + \tau_{\min}}{2} \tag{17}
$$

$$
\tau_{\max} = \max_{\omega \in \Psi_p} \tau_h(e^{j\omega}) \tag{18}
$$

$$
\tau_{\min} = \min_{\omega \in \Psi_p} \tau_h(e^{j\omega}). \tag{19}
$$

Hence,

$$
Q_\tau = \frac{100(\tau_{\max} - \tau_{\min})}{(\tau_{\max} + \tau_{\min})} \tag{20}
$$

Q_τ will be referred to as the *maximum group-delay deviation* hereafter.

B. Passband Error

If $H_d(\omega)$ is the desired frequency response of the filter and $\mathbf{c}_k$ is the value of vector $\mathbf{c}$ at the start of the kth iteration, a passband error function at frequency ω can be defined as

$$
e_h(\mathbf{c}_k, e^{j\omega}) = |H(\mathbf{c}_k, e^{j\omega})|^2 - |H_d(\omega)|^2. \tag{21}
$$

Without loss of generality, we can assume that the desired amplitude response is unity in the passband. Therefore, the passband error function becomes

$$
e_h^{(pb)}(\mathbf{c}_k, e^{j\omega}) = |H(\mathbf{c}_k, e^{j\omega})|^2 - 1, \quad \omega \in \Psi_p. \tag{22}
$$

Using the same approach as in Section II-A, the L_p-norm of the passband error function, $\mathbf{E}_p^{(pb)}(k)$, in matrix form can be approximated as

$$
\mathbf{E}_p^{(pb)}(k) \approx \|\mathbf{D}_k^{(pb)}\boldsymbol{\delta} + \mathbf{f}_k^{(pb)}\|_p \tag{23}
$$

where

$$
\mathbf{D}_k^{(pb)} = \begin{bmatrix} \kappa_{pb} \nabla e_h^{(pb)}(\mathbf{c}_k, e^{j\omega_1})^T & 0 \\ \vdots & \vdots \\ \kappa_{pb} \nabla e_h^{(pb)}(\mathbf{c}_k, e^{j\omega_{N_p}})^T & 0 \end{bmatrix}, \quad \omega_i \in \Psi_p \tag{24}
$$

$$
\mathbf{f}_k^{(pb)} = [f_1^{(pb)} \ f_2^{(pb)} \ \cdots \ f_{N_p}^{(pb)}]^T, \tag{25}
$$

$$
f_i^{(pb)} = \kappa_{pb} e_h^{(pb)}(\mathbf{c}_k, e^{j\omega_i}), \tag{26}
$$

$$
\boldsymbol{\delta} = [\boldsymbol{\delta}_c^T \ \delta_\tau]^T. \tag{27}
$$

In the above equations, $\boldsymbol{\delta}_c$ is the vector update for $\mathbf{c}_k$, δ_τ is the scalar update for τ, and κ_{pb} is a constant. The elements of the last column of $\mathbf{D}_k^{(pb)}$ in (24) are all zeros since (23) is independent of τ.

C. Amplitude Response in the Stopband and Transition Band

The frequency response update for the filter at the kth iteration is given by

$$
H(\mathbf{c}_k + \boldsymbol{\delta}_c, e^{j\omega}) \approx H(\mathbf{c}_k, e^{j\omega}) + \nabla H(\mathbf{c}_k, e^{j\omega})^T \boldsymbol{\delta}_c. \tag{28}
$$

By using the same approach as in Section II-B, the L_p-norm of the frequency response in the stopband and transition band can be approximated as

$$
\mathbf{E}_p^{(sb)}(k) \approx \|\mathbf{D}_k^{(sb)}\boldsymbol{\delta} + \mathbf{f}_k^{(sb)}\|_p \tag{29}
$$

$$
\mathbf{E}_p^{(tb)}(k) \approx \|\mathbf{D}_k^{(tb)}\boldsymbol{\delta} + \mathbf{f}_k^{(tb)}\|_p \tag{30}
$$

where

$$
\mathbf{D}_k^{(sb)} = \begin{bmatrix} \kappa_{sb} \nabla H(\mathbf{c}_k, e^{j\omega_1})^T & 0 \\ \vdots & \vdots \\ \kappa_{sb} \nabla H(\mathbf{c}_k, e^{j\omega_{N_s}})^T & 0 \end{bmatrix}, \quad \omega_i \in \Psi_s \tag{31}
$$

$$
\mathbf{D}_k^{(tb)} = \begin{bmatrix} \kappa_{tb} \nabla H(\mathbf{c}_k, e^{j\omega_1})^T & 0 \\ \vdots & \vdots \\ \kappa_{tb} \nabla H(\mathbf{c}_k, e^{j\omega_{N_t}})^T & 0 \end{bmatrix}, \quad \omega_i \in \Psi_t \tag{32}
$$

$$
\mathbf{f}_k^{(sb)} = [f_1^{(sb)} \ f_2^{(sb)} \ \cdots \ f_{N_s}^{(sb)}]^T, \quad \omega_i \in \Psi_s \tag{33}
$$

$$
\mathbf{f}_k^{(tb)} = [f_1^{(tb)} \ f_2^{(tb)} \ \cdots \ f_{N_t}^{(tb)}]^T, \quad \omega_i \in \Psi_t \tag{34}
$$

$$
f_i^{(sb)} = \kappa_{sb} H(\mathbf{c}_k, e^{j\omega_i}) \tag{35}
$$

$$
f_i^{(tb)} = \kappa_{tb} H(\mathbf{c}_k, e^{j\omega_i}). \tag{36}
$$

In the above equations, Ψ_s and Ψ_t are the sets of frequency sample points in the stopband and transition band, respectively, and κ_{sb} and κ_{tb} are constants. Note that Ψ_p, Ψ_s, and Ψ_t may extend over two or more passbands, stopbands, or transition bands, respectively, in the case of multiband filters.

D. Filter Stability

To ensure that the filter is stable, the poles of the transfer function must lie within the unit circle [3]. If $\epsilon_s > 0$ is a stability margin defined as $\epsilon_s = 1 - r_{\max}^{(p)}$ where $r_{\max}^{(p)}$ is the maximum pole radius allowed, the coefficients of the denominator should satisfy the stability conditions given by [5]

$$
b_{0m} \leq 1 - \gamma \tag{37}
$$

$$
b_{1m} - b_{0m} \leq 1 - \gamma \tag{38}
$$

$$
-b_{1m} - b_{0m} \leq 1 - \gamma \tag{39}
$$

where [18]

$$
\gamma = 1 - (1 - \epsilon_s)^2. \tag{40}
$$

Incorporating stability margin ϵ_s would ensure that roundoff errors would not cause filter instability, particularly in fixed-point implementations.

The stability conditions in (37)–(39) for the kth iteration can be expressed in matrix form as

$$\mathbf{B}\boldsymbol{\delta} < \mathbf{b}^{(k)} \qquad (41)$$

where

$$\mathbf{B} = \begin{bmatrix} \mathcal{B} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0}\ \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathcal{B} & \mathbf{0}\ \mathbf{0} \end{bmatrix} \qquad (42)$$

$$\mathcal{B} = \begin{bmatrix} 1 & 0 \\ -1 & 1 \\ -1 & -1 \end{bmatrix} \qquad (43)$$

$$\mathbf{b}^{(k)} = \begin{bmatrix} \mathbf{b}_1^{(k)} \\ \vdots \\ \mathbf{b}_J^{(k)} \end{bmatrix} \qquad (44)$$

$$\mathbf{b}_m^{(k)} = \begin{bmatrix} 1 - \epsilon_s - b_{0m}^{(k)} \\ 1 - \epsilon_s - b_{1m}^{(k)} + b_{0m}^{(k)} \\ 1 - \epsilon_s + b_{1m}^{(k)} + b_{0m}^{(k)} \end{bmatrix}. \qquad (45)$$

E. Optimization Problem

The optimization is carried out by minimizing the group-delay deviation under the constraints that the passband error, stopband attenuation, and transition band attenuation are within prescribed levels. This can be done by solving the optimization problem

$$\begin{aligned} \text{minimize} \quad & \|e_g(\mathbf{x}, e^{j\omega})\|_p \\ \text{subject to:} \quad & \text{passband error function} < \Gamma_{pb} \\ & \text{stopband gain } < \Gamma_{sb} \\ & \text{transition band gain} < \Gamma_{tb} \\ & \text{stability margin} = \epsilon_s \end{aligned} \qquad (46)$$

where Γ_{pb}, Γ_{sb}, and Γ_{tb} are the maximum prescribed levels for the passband error, stopband gain, and transition band gain, respectively.

Using (12), (23), (29), (30), and (41) the problem to be solved in the kth iteration can be expressed as

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{C}_k\boldsymbol{\delta} + \mathbf{d}_k\|_p \\ \text{subject to:} \quad & \|\mathbf{D}_k^{(pb)}\boldsymbol{\delta} + \mathbf{f}_k^{(pb)}\|_p < \Gamma_{pb} \\ & \|\mathbf{D}_k^{(sb)}\boldsymbol{\delta} + \mathbf{f}_k^{(sb)}\|_p < \Gamma_{sb} \\ & \|\mathbf{D}_k^{(tb)}\boldsymbol{\delta} + \mathbf{f}_k^{(tb)}\|_p < \Gamma_{tb} \\ & \|\boldsymbol{\delta}\|_2 < \Gamma_{small} \\ & \mathbf{B}\boldsymbol{\delta} < \mathbf{b}^{(k)} \end{aligned} \qquad (47)$$

where $\boldsymbol{\delta}$ is the optimization variable. The optimum value of $\boldsymbol{\delta}$ is then used to update the optimizing parameters for the next iteration. The bound on the norm of the update vector assures the validity of the linear approximation and at the same time eliminates the need for a line search step that is required in most optimization algorithms [18].

The value of p in the above optimization problem can be any positive integer. The most significant values for p are 2 and ∞.

In the first case, the L_2-norm would be minimized, which would result in a least-squares solution, whereas, in the second case, the L_∞-norm would be minimized, which would result in a minimax solution. In a least-squares solution, the squares of the passband and stoband errors would be minimized whereas in a minimax solution the maxima of the absolute values of the passband and stopband errors would be minimized. In this paper, we explore the use of the L_∞-norm.

The optimization problem in (47) can be solved by starting with an initialization filter that satisfies at least some of the specifications, if at all possible, in order to reduce the amount of computation required. The initialization filter may satisfy only the stopband and transition band constraints but not the passband constraint. In order to handle such scenarios, we relax the constraints for the passband and the norm of the parameter update by adding a variable δ_{rlx} which is also minimized while its value is constrained to be positive; when $\delta_{rlx} = 0$, the original constraints are restored. The relaxation of the norm of the parameter update is to facilitate a greater change between iterations so that all constraints are quickly satisfied and δ_{rlx} is reduced to a small value. With this modification, the problem in (47) for the case $p = \infty$ becomes

$$\begin{aligned} \text{minimize} \quad & \|\mathbf{C}_k\boldsymbol{\delta} + \mathbf{d}_k\|_\infty + W\delta_{rlx} \\ \text{subject to:} \quad & \|\mathbf{D}_k^{(pb)}\boldsymbol{\delta} + \mathbf{f}_k^{(pb)}\|_\infty < \Gamma_{pb} + \delta_{rlx} \\ & \|\mathbf{D}_k^{(sb)}\boldsymbol{\delta} + \mathbf{f}_k^{(sb)}\|_\infty < \Gamma_{sb} \\ & \|\mathbf{D}_k^{(tb)}\boldsymbol{\delta} + \mathbf{f}_k^{(tb)}\|_\infty < \Gamma_{tb} \\ & \|\boldsymbol{\delta}\|_2 < \Gamma_{small} + \delta_{rlx} \\ & \delta_{rlx} > 0 \\ & \mathbf{B}\boldsymbol{\delta} < \mathbf{b}^{(k)} \end{aligned} \qquad (48)$$

where $\boldsymbol{\delta}$ and δ_{rlx} are optimization variables, and $W > 0$ is a weighing factor for the relaxation variable. It should be noted that if the initialization filter is feasible, the use of the relaxation parameter δ_{rlx} is not necessary although in some cases it has been found to speed up the convergence. For the situation where the numerator and denominator orders are required to be different, the appropriate numerator or denominator coefficients in some of the SOSs can be set to zero by initializing them with zero and correspondingly constraining their updated values to zero. That is, if $x_k(i)$ corresponds to a coefficient in an SOS that is required to be zero in every iteration k, we intialize $x_0(i) = 0$ and correspondingly include the additional constraint on the corresponding updated value, i.e., $\delta(i) = 0$.

The optimization problem in (48) can be easily expressed as a *second-order cone programmming* (SOCP) problem as in [18] and solved using efficient SOCP solvers such as the one available in the SeDuMi optimization toolbox [19] for MATLAB™. The various L_∞-norms in (48) are efficiently evaluated by the SOCP solver.

III. DESIGN PROCEDURE

Two general strategies for the design of digital filters have been developed to deal with design problems where the group delay is not specified or with problems where a prescribed group delay is required. In the former case, the group delay can be used as an additional independent variable that can be optimized in

order to bring about additional improvements to the filter being designed.

A. Optimized Group Delay

When the group delay is assumed to be an independent variable, it is important that the initialization filter be chosen to be close to the desired optimal filter in order to assure fast convergence. To this end, a good first step would be to design the lowest-order IIR filter that would satisfy only the amplitude-response specifications. An *elliptic* filter would be the most suitable choice since it gives the lowest-order IIR filter for any given amplitude-response specifications because of the optimality of the elliptic approximation. Such a filter can be obtained by using the design method described in Chap. 12 of [3].

To reduce the group-delay deviation of the filter in the passband, a number of additional general biquadratic SOSs are included depending on the degree of linearity required in the phase response. To achieve fast convergence, the additional SOSs are initialized as allpass sections. The poles and zeros of the additional SOSs are initially distributed in the passband sector of the z plane, namely, the sector bounded by the passband edge frequencies. Under these circumstances, the transfer function assumes the form

$$
\begin{aligned}
H_{init}(z) &= H_{ellip}(z) \\
&\cdot G_0 \prod_{k=1}^{M} \frac{(z - r_k^{-1} e^{j\omega_k})(z - r_k^{-1} e^{-j\omega_k})}{(z - r_k e^{j\omega_k})(z - r_k e^{-j\omega_k})}, \\
&\omega_k \in \Psi_p
\end{aligned}
\tag{49}
$$

where $H_{ellip}(z)$ is the transfer function of the elliptic filter, G_0 is a normalizing gain factor, M is the number of additional allpass SOSs, and $0 < r_k < 1$. An initialization group delay that was found to work well is the average of the passband filter-equalizer combination, which can be estimated as

$$
\tau_{init} = \frac{-1}{\omega_{ph} - \omega_{pl}} \int_{\omega_{pl}}^{\omega_{ph}} \frac{d}{d\omega} \left\{ \arg \left[H_{init}(e^{j\omega}) \right] \right\} d\omega.
\tag{50}
$$

Although using the initializaton group delay in (50) always results in solutions that are comparable or better than those achieved with existing methods, we have found some instances where starting with the maximum or minimum passband group delay can yield a better solution. Therefore, to maximize the possibility for obtaining the best solution, the optimization can also be initialized with the minimum and maximum passband group-delays given by

$$
\tau_{init}^{(max)} = \max_{\omega} \left(-\frac{d}{d\omega} \left\{ \arg \left[H_{init}(e^{j\omega}) \right] \right\} \right),
$$
$$
\omega \in \Psi_p
\tag{51}
$$
$$
\tau_{init}^{(min)} = \min_{\omega} \left(-\frac{d}{d\omega} \left\{ \arg \left[H_{init}(e^{j\omega}) \right] \right\} \right),
$$
$$
\omega \in \Psi_p
\tag{52}
$$

and then selecting the best solution from among the three results. However, if there is a need to reduce the amount of computation required, the filter can be designed by using only the initialization group delay in (50).

The required filter can be designed by using the following algorithm:

Step 1: Obtain the transfer function of the required elliptic filter, $H_{ellip}(z)$, that satisfies the required amplitude response specifications; e.g., by using the D-Filter software package [20].

Step 2: Set the number of additional general biquadratic SOSs to M and select r_k and ω_k to construct the transfer function $H_{init}(z)$ as in (49); from $H_{init}(z)$, compute the initialization group delay τ_{init} using (50). This can be easily done by using D-Filter [20].

Step 3a: Using $H_{init}(z)$ and τ_{init} for initialization, solve the optimization problem in (48).

Step 3b (optional): Solve the optimization problem in (48) using $\{H_{init}(z), \tau_{init}^{(max)}\}$ and $\{H_{init}(z), \tau_{init}^{(min)}\}$ for initialization and then select the solution that has the smallest value of Q_τ in *Steps 3a* and *3b*.

Step 4: Using (20), compute the maximum group-delay deviation of the filter, Q_τ, obtained in *Step 3*. If Q_τ is less than the prescribed value, the filter specifications are satisfied and the algorithm is terminated; otherwise, set $M = M + 1$ and go to *Step 2*.

The optional step, *Step 3b*, can be carried out if the amount of computation required is not a critical factor, in order to increase the possibility for obtaining a better solution.

B. Prescribed Group Delay

When a prescribed group delay is required, the initialization procedure described in Section III-A is not appropriate. A more appropriate initialization scheme would be to use the balance model truncation (BMT) method described in [21], [22]. The main steps of the BMT involve converting a high-order FIR filter into a state-space balanced model, then reducing the model order, and finally converting the lower-order model to a reduced-order IIR filter.

To ensure that the IIR filter obtained with the BMT method has a group delay that is close to the prescribed value, the initialization linear-phase FIR filter is designed to have a group delay, τ_{pr}, that is close to the prescribed value. This can be done by selecting the filter length as

$$
L_{fir} = 2\lceil \tau_{pr} \rceil + 1
\tag{53}
$$

where $\lceil \cdot \rceil$ is the ceiling operator. The transfer function of the IIR filter can be expressed as

$$
H_{init}(z) = \hat{G}_0 \frac{\prod_i (z - \check{z}_i)}{D_{bmt}(z)} = \frac{\sum_i \check{b}_i z^{m-i}}{D_{bmt}(z)}
\tag{54}
$$

where the normalizing gain factor $\hat{G}_0$ is chosen to ensure that the average passband gain of the filter is unity.

Sometimes, the filter obtained with the BMT method may have one or more zeros that are located far away from the origin. Such zeros can slow down the optimization algorithm. Experimental results have shown that faster convergence can be achieved by moving any zeros with a radius greater than a prescribed maximum $r_{\max}^{(z)}$ to the origin of the z plane by letting

$$\check{z}_i = \begin{cases} 0 & \text{if } |\bar{z}_i| > r_{\max}^{(z)} \\ \bar{z}_i & \text{otherwise.} \end{cases} \quad (55)$$

With the group delay fixed, the problem in (48) can be simplified to

$$
\begin{aligned}
\text{minimize} \quad & \|\mathbf{C}_k \mathbf{I}_{ce} \boldsymbol{\delta}_c + \mathbf{d}_k\|_\infty + W \delta_{rlx} \\
\text{subject to:} \quad & \|\mathbf{D}_k^{(pb)} \mathbf{I}_{ce} \boldsymbol{\delta}_c + \mathbf{f}_k^{(pb)}\|_\infty < \Gamma_{pb} + \delta_{rlx} \\
& \|\mathbf{D}_k^{(sb)} \mathbf{I}_{ce} \boldsymbol{\delta}_c + \mathbf{f}_k^{(sb)}\|_\infty < \Gamma_{sb} \\
& \|\mathbf{D}_k^{(tb)} \mathbf{I}_{ce} \boldsymbol{\delta}_c + \mathbf{f}_k^{(tb)}\|_\infty < \Gamma_{tb} \\
& \|\boldsymbol{\delta}_c\|_2 < \Gamma_{small} + \delta_{rlx} \\
& \delta_{rlx} > 0 \\
& \mathbf{B} \mathbf{I}_{ce} \boldsymbol{\delta}_c < \mathbf{b}^{(k)} \quad (56)
\end{aligned}
$$

where $\boldsymbol{\delta}_c$ is the optimization variable and $\mathbf{I}_{ce}$ is given by

$$\mathbf{I}_{ce} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (57)$$

The required filter can then be designed by using the following procedure:

Step 1: Design a linear-phase FIR filter of length given by (53) with the prescribed passband- and stopband-edge frequencies, by using D-Filter or the MATLAB™ function *fir1* or some other way.

Step 2: If the total number of SOSs is M_{tot}, the IIR filter order is $2M_{tot}$. Using the BMT method, transform the FIR filter obtained in *Step 1* to an IIR filter of order $2M_{tot}$.

Step 3: Form the transfer function in (54).

Step 4: Using $\tilde{H}_{init}(z)$ in (54) for initialization, solve the optimization problem in (56) for the prescribed group delay of τ_{pr}.

Step 5: Using (20) compute the group-delay deviation, Q_τ, of the filter obtained in *Step 4*. If Q_τ is less than the maximum prescribed value, the filter specifications are satisfied and the algorithm is terminated; otherwise, set $M_{tot} = M_{tot} + 1$ and go to *Step 2*.

C. Practical Considerations

The computational effort required to complete a design is directly proportional to the number of frequency sample points used in sets Ψ_p, Ψ_s, and Ψ_t. To reduce the number of sample points and at the same time prevent spikes in the error function, the *nonuniform variable sampling* (NVS) technique described in [23] (see also Chapter 16 of [3]) was used for the choice of sample points. When compared with the standard uniform

sampling method, the NVS technique not only reduces the computational effort by around an order of magnitude but it also reduces the passband ripple and improves the convergence of our algorithms.

The weight W for the relaxation parameter, δ_{rlx}, in (48) should not be too small as this can make the optimization algorithm unstable and prevent it from converging; at the same time, it should also not be too large as this can slow down the convergence process. Typical values of W that have been found to work well range between 500 to 5000.

The computational efficiency of the algorithm is also dependent on two other aspects: the efficiency of the SOCP solver and the number of iterations required to achieve convergence. For the former, most SOCP solvers, including the one in the SeDuMi toolbox for MATLAB™, are polynomial-time type algorithms which can be easily run on inexpensive computers such as a laptop. For the latter, the number of iterations to achieve convergence is determined by how close the initialization point is to the final solution. Sometimes, the solution keeps improving with each iteration, but beyond a certain point the degree of improvement is too small to be of practical significance and the optimization can be terminated. Sometimes, the objective function may at some point show very small improvement, or even increase for several iterations before rapidly decreasing again. To ensure that the optimization is not prematurely terminated, the values of the objective function for the L_o most recent iterations are compared with the minimum value achieved before and if each of these values is larger than the minimum value, the optimization is terminated. A value of 40 for L_o was found to work well.

IV. EXPERIMENTAL RESULTS

In order to compare our method with other state-of-the-art competing methods, we have designed and tested many nearly linear-phase IIR filters satisfying a diverse range of specifications. Ten examples are included in this paper.

Parameters Γ_{small} and W in (48) were set to 0.01 and 1000, respectively, while $r_{\max}^{(z)}$ in (55) was set to 2.5. The M poles of the allpass part of the transfer function in (49) were initialized with $r_k = 0.8$ and ω_k uniformly distributed in the passband of the required filter. The default maximum pole radius was set to 0.98 except in Example 10 where it was set to a value slightly higher than the maximum pole radius of the elliptic filter, namely, 0.991.

The frequency sets Ψ_p, Ψ_s, and Ψ_t were determined by using the NVS technique [23] with a dense set of 2000 uniformly-spaced virtual frequencies and 68 actual sample frequencies for each passband and stopband, respectively. Six of the actual sample frequencies located near the passband edge in Ψ_p and stopband edge in Ψ_s were fixed with a separation of 7.8×10^{-4} radians between them. For each transition band, 500 virtual sample frequencies and 18 actual sample frequencies were used.

In some of the design examples, namely, in Examples 1, 2, and 3, the passband gain of the competing filter used for comparison was constrained to have a maximum value of 1. To achieve consistency between our filters and the competing filters, we normalized the multiplier constant in some of the competing

designs to achieve an average passband gain of unity. This was done by adjusting the multiplier constant so that

$$\bar{H}_0 = \frac{2H_0}{g_{\max} + g_{\min}} = \frac{2H_0}{1 + g_{\min}} \qquad (58)$$

where $g_{\max}$ and $g_{\min}$ are the maximum and minimum passband gains, respectively.

In the experiments, the optimization was first carried out by allowing the group delay to be variable without imposing a constraint on the maximum gain in transition bands. In some designs, the filter obtained had an anomaly in a transition band whereby the gain increased above unity. This may not be a problem in many applications but if the anomaly is undesirable it can be eliminated by imposing a constraint on the maximum gain for the transition band with the anomaly. In such situations, a second design was carried out to ensure that the gain in transition bands is always equal to or less than unity.

Our method can be used to design filters with optimized or fixed group delay. However, some of the competing methods do not provide for an optimized group delay. In such examples, we have designed two filters, one with fixed and the other with optimized group delay for the sake of comparison.

To deal with the above possibilities, some of the following four design variants have been carried out for the examples considered:

1) *Design A-1* The group delay is optimized and no constraint is applied on the transition-band gain.
2) *Design A-2* The group delay is optimized and a constraint is applied on the transition-band gain with $\Gamma_{tb} = 1$.
3) *Design B-1* The group delay is fixed and no constraint is applied on the transition-band gain.
4) *Design B-2* The group delay is fixed and a constraint is applied on the transition-band gain with $\Gamma_{tb} = 1$.
5) *Design B-3* The group delay is fixed and Γ_{tb} is adjusted so as to ensure that the transition band gain is close to that of the competing filter.

Examples 1 to 3 involve a lowpass, a highpass, and a bandpass filter and the designs obtained are compared with designs based on the classical equalizer approach described in [4]. Examples 4, 5, 6, 8, and 9 concern different lowpass filters and Example 7 concerns a highpass filter. The designs obtained with the proposed method are compared with corresponding designs obtained by using the methods in [5], [8], [12]–[14], [17], [18], and [11], respectively. Example 10 is concerned with a selective IIR bandpass-filter design which is compared with an optimal FIR filter satisfying the same specifications.

It should be noted that the optional design step, *Step 3b*, in Section III-A for the design of filters with optimized group delay was included in all of the design examples except Example 10.

A. Examples 1, 2, and 3

The required design specifications for the filters of Examples 1, 2, and 3 are given in Table I, Table III, and Table V, the results obtained are summarized in Table II, Table IV, and Table VI and the amplitude responses and delay characteristics are plotted in Figs. 1–3, respectively. The poles and zeros of our designs for Examples 1 to 3 and also 4 to 9 are given in [24]. In designs where a transition-band anomaly occurred both designs A-1 and A-2 were carried out as seen in the tables.
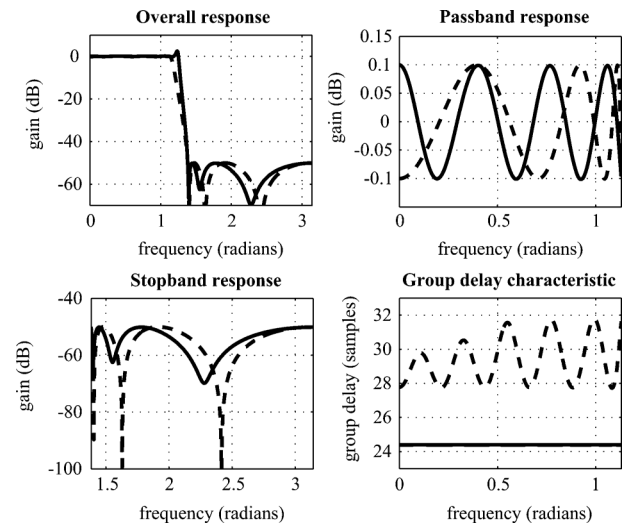

Fig. 1. Overall, passband, and stopband amplitude responses and group-delay characteristic for Design A-1 of the proposed method (solid curves) and the method in [4] (dashed curves) for Example 1.
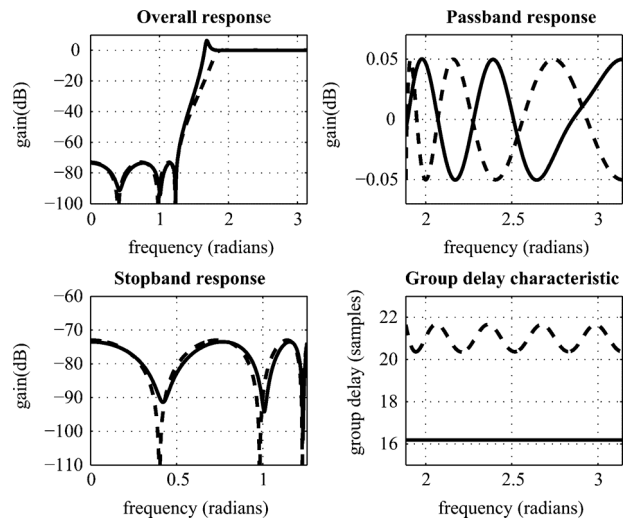
Fig. 2. Overall, passband, and stopband amplitude responses and group-delay characteristic for Design A-1 of the proposed method (solid curves) and the method in [4] (dashed curves) for Example 2.

The competing filters for Examples 1 and 2 correspond to the first and second examples in [4], while that of Example 3 is obtained from p. 761 of [3].

As can be seen in Table II, Table IV, and Table VI and Figs. 1–3, the IIR filters designed using the proposed method have much smaller maximum group-delay deviation for practically the same passband ripple and minimum stopband attenuation as the designs obtained with the classical equalizer method.

B. Examples 4, 5, 6, 7, and 8

The competing filters for Examples 4, 5, and 6 correspond to the fourth example in [5], the first example in [18], and the first example in [17], respectively. For Example 7 there are two competing filters, namely, the second example in [12] and the fourth example in [13], while for Example 8 the two competing filters were taken from the third example in [14].

Fig. 3. Overall, stopband, and passband amplitude responses and group-delay characteristic for Design A-1 of the proposed method (solid curves) and the method in [3] (dashed curves) for Example 3.
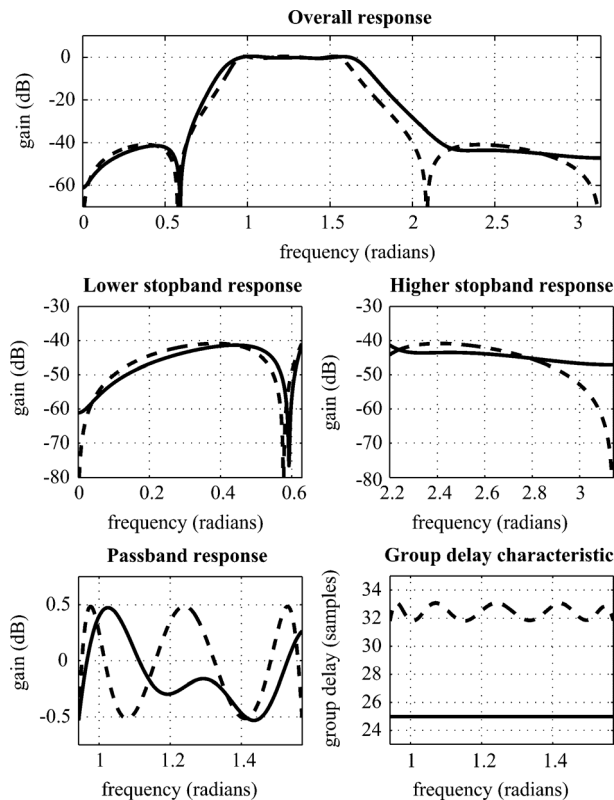
TABLE I
LOWPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 1

Parameters	Values
Maximum passband ripple, dB	0.2
Minimum stopband attenuation, dB	50
Passband edge, rad/s	0.36π
Stopband edge, rad/s	0.44π
Maximum pole radius	0.98

TABLE II
DESIGN RESULTS FOR EXAMPLE 1 (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [4]
Elliptic filter order	NA	NA	6
Equalizer filter order	NA	NA	10
Total filter order	16	16	16
Max PB ripple, dB	0.2	0.2	0.2
Min SB atten., dB	50.14	50.14	50.14
Max TB gain, dB	2.5	0	-0.2
τ_{avg}	24.38	24.45	29.75
Q_τ	0.00796	0.0132	6.82

PB: passband; TB:transition band; SB: stopband

For Examples 5, 7, and 8, the group delay was prescribed by the competing methods and, therefore, both designs A and B were carried out.

In Example 6, we considered the design of a lowpass filter following a strategy reported in [17] whereby the filter is designed as a cascade combination of an IIR filter and an FIR filter. For the comparison, we considered two filter structures for the proposed method: the first is similar to the structure used in [17] and comprises a 4th-order IIR filter in cascade with an

TABLE III
HIGHPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 2

Parameters	Values
Maximum passband ripple, dB	0.1
Minimum stopband attenuation, dB	73
Passband edge, rad/s	0.6π
Stopband edge, rad/s	0.4π
Maximum pole radius	0.98

TABLE IV
DESIGN RESULTS FOR EXAMPLE 2 (HIGHPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [4]
Elliptic filter order	NA	NA	6
Equalizer filter order	NA	NA	8
Total filter order	14	14	14
Max PB ripple, dB	0.1	0.1	0.1
Min SB atten., dB	73.5	73.5	73.1
Max TB gain, dB	6.32	0.002	-0.103
τ_{avg}	16.19	18.02	21
Q_τ	0.00104	0.00905	3.13

PB: passband; TB:transition band; SB: stopband

TABLE V
BANDPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 3

Parameters	Values
Maximum passband ripple, dB	1.0
Minimum stopband attenuation, dB	41
Low stopband edge, rad/s	0.2π
Low passband edge, rad/s	0.3π
High passband edge, rad/s	0.5π
High stopband edge, rad/s	0.7π
Maximum pole radius	0.98

TABLE VI
DESIGN RESULTS FOR EXAMPLE 3 (BANDPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [3]
Elliptic filter order	NA	NA	6
Equalizer filter order	NA	NA	8
Total filter order	14	14	14
Max PB ripple, dB	1	1	1
Min SB atten., dB	41.4	41.37	41.37
Max TB gain, dB	0.33	0	-1
τ_{avg}	24.93	25.54	32.44
Q_τ	0.000461	0.00126	1.96

PB: passband; TB:transition band; SB: stopband

11th-order FIR filter; the second is a 10th-order IIR filter which is equivalent to the first structure in terms of the number of multipliers.

The prescribed specifications for the filters in Examples 4 to 8 are given in Table VII, Table IX, Table XII, Table XV, and Table XVIII, respectively. The results obtained are summarized in Table VIII, Table X, Table XI, Table XIII, Table XIV, Table XVI, Table XVII, Table XIX, and Table XX and the frequency responses obtained are plotted in Figs. 4, 5, and Fig. 6. As can be seen in these tables and figures, the IIR filters designed using the proposed method have much smaller maximum group-delay deviation for practically the same passband ripple and minimum stopband attenuation as the designs obtained with the competing methods in [5], [12], [13], [17], [18], and [14] except that our method also yields increased minimum stopband attenuation relative to that achieved with the method in [17].
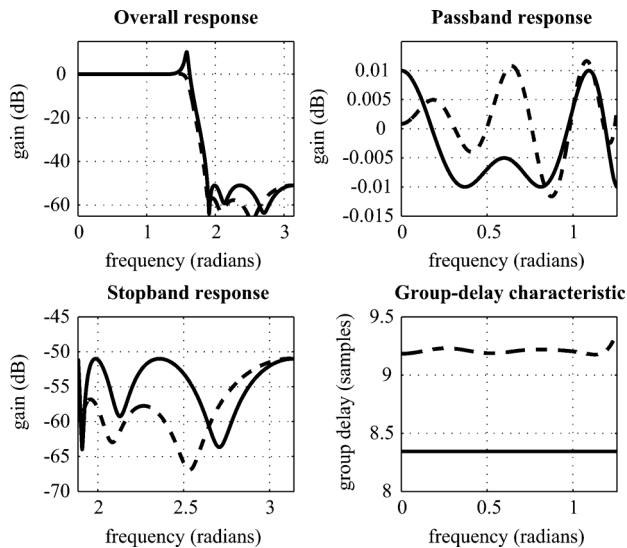
Fig. 4. Overall, passband, and stopband amplitude responses and group-delay characteristic for Design A-1 of the proposed method (solid curves) and the method in [5] (dashed curves) for Example 4.

Fig. 6. Overall, passband, and stopband amplitude responses and group-delay characteristic for Design A-1 of the all-IIR filter structure of the proposed method (solid curves) and the method in [17] (dashed curves) for Example 6.
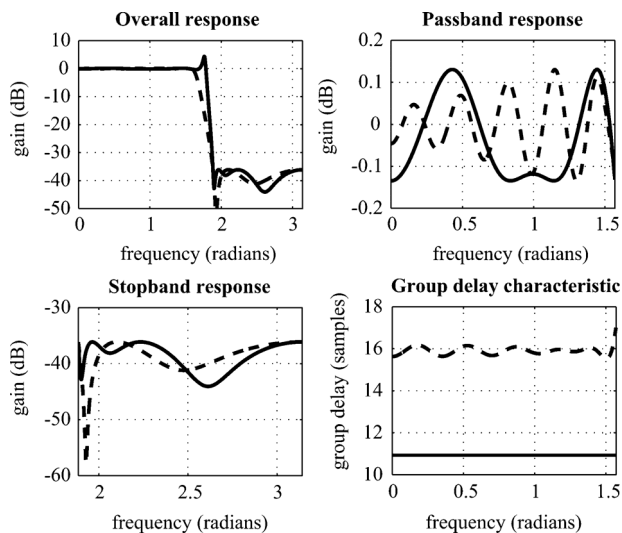

Fig. 5. Overall, passband, and stopband amplitude responses and group-delay characteristic for Design A-1 of the proposed method (solid curves) and the method in [18] (dashed curves) for Example 5.

C. Example 9

We also carried out comparisons with lowpass filter 1-iii in the first example of [8]. Our designs A-1, A-2, and B-1 have slightly better amplitude-response specifications and, in addition, offer reduced values of Q_τ relative to that achieved in [8], namely, 0.00043, 0.00089, 0.086 for designs A-1, A-2, and B-1, respectively, versus 0.13 for the design in [8]. We then compared the same designs with filter MMPE(2) in the first example of [11] and again found that our designs offer slightly better amplitude-response specifications and much lower values of Q_τ relative to that for the design in [11] which was 0.11. The poles and zeros and additional results for these filters are given in [24]. It should be noted that no filter coefficients are provided in [11] but from the design results reported in [11] we were able to estimate all the per-
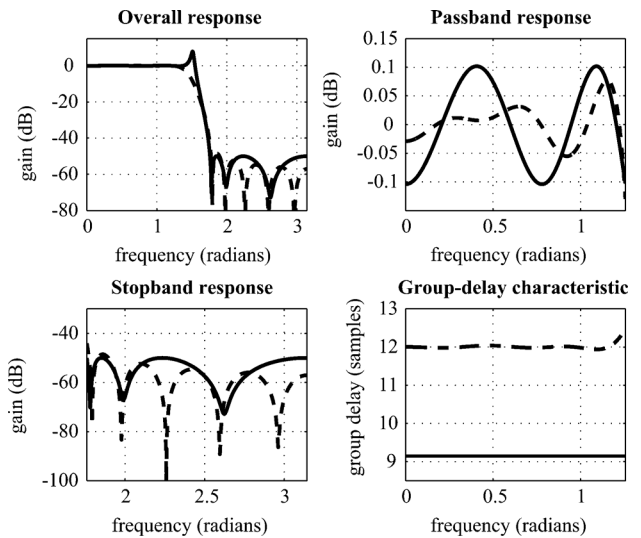
TABLE VII
LOWPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 4

Parameters	Values
Maximum passband ripple, dB	0.025
Minimum stopband attenuation, dB	50
Passband edge, rad/s	0.4π
Stopband edge, rad/s	0.6π
Maximum pole radius	0.98

TABLE VIII
DESIGN RESULTS FOR EXAMPLE 4 (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [5]
Total filter order	10	10	10
Max PB ripple, dB	0.022	0.022	0.023
Min SB atten., dB	51	51	50.9
Max TB gain, dB	10.18	0.002	0.17
τ_{avg}	8.35	8.59	9.28
Q_τ	0.000472	0.20	1.07

PB: passband; TB:transition band; SB: stopband

TABLE IX
LOWPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 5

Parameters	Values
Maximum passband ripple, dB	0.266
Minimum stopband attenuation, dB	37
Passband edge, rad/s	0.5π
Stopband edge, rad/s	0.6π
Maximum pole radius	0.98
Group delay (samples)	15.9

TABLE X
DESIGN RESULTS FOR EXAMPLE 5 FOR VARIABLE
GROUP DELAY (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [18]
Total filter order	12	12	12
Max PB ripple, dB	0.265	0.265	0.266
Min SB atten., dB	36.145	36.146	36.146
Max TB gain, dB	4.37	0	-0.136
τ_{avg}	10.92	11.40	16.26
Q_τ	0.00449	0.0188	4.54

PB: passband; TB:transition band; SB: stopband

formance parameters except for the maximum transition-band gain.

TABLE XI
DESIGN RESULTS FOR EXAMPLE 5 FOR FIXED GROUP DELAY (LOWPASS FILTER)

Parameters	Design B-1	Method in [18]
Total filter order	12	12
Max PB ripple, dB	0.263	0.266
Min SB atten., dB	36.146	36.146
Max TB gain, dB	-0.023	-0.136
τ_{avg}	15.98	16.26
Q_τ	2.69	4.54

PB: passband; TB:transition band; SB: stopband

TABLE XII
LOWPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 6

Parameters	Values
Maximum passband ripple, dB	0.25
Minimum stopband attenuation, dB	44
Passband edge, rad/s	0.4π
Stopband edge, rad/s	0.56π
Maximum pole radius	0.98

TABLE XIII
DESIGN RESULTS FOR EXAMPLE 6 WHERE THE PROPOSED FILTER IS AN ALL-IIR FILTER STRUCTURE (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [17]
IIR filter order	10	10	4
FIR filter order	0	0	11
Max PB ripple, dB	0.207	0.207	0.207
Min SB atten., dB	50	50	44.1
Max TB gain, dB	7.89	0	-0.133
τ_{avg}	9.15	9.79	12.2
Q_τ	0.00130	0.159	2.25

PB: passband; TB:transition band; SB: stopband

TABLE XIV
DESIGN RESULTS FOR EXAMPLE 6 WHERE THE PROPOSED FILTER IS AN IIR-FIR FILTER COMBINATION (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [17]
IIR filter order	4	4	4
FIR filter order	11	11	11
Max PB ripple, dB	0.204	0.205	0.207
Min SB atten., dB	50	50	44.1
Max TB gain, dB	8.77	0	-0.133
τ_{avg}	8.72	9.35	12.2
Q_τ	0.0149	0.382	2.25

PB: passband; TB:transition band; SB: stopband

TABLE XV
HIGHPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 7

Parameters	Values
Maximum passband ripple, dB	0.720
Minimum stopband attenuation, dB	27
Passband edge, rad/s	0.525π
Stopband edge, rad/s	0.475π
Maximum pole radius	0.98
Group delay (samples)	12

D. Example 10

The prescribed specifications for the high-selectivity bandpass filter are given in Table XXI. For this filter, we obtained type A-1 and A-2 IIR designs and also a corresponding optimal

TABLE XVI
DESIGN RESULTS FOR EXAMPLE 7 FOR VARIABLE GROUP DELAY (HIGHPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [12]
Filter order	14	14	14
Max PB ripple, dB	0.491	0.492	0.720
Min SB atten., dB	31	30.98	27.36
Max TB gain, dB	7.5	0.021	0.155
τ_{avg}	11.95	12.7	13.74
Q_τ	0.00515	1.71	16.4
Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [13]
Filter order	14	14	14
Max PB ripple, dB	0.491	0.492	0.492
Min SB atten., dB	31	30.98	30.7
Max TB gain, dB	7.5	0.021	3.86
τ_{avg}	11.95	12.7	12.79
Q_τ	0.00515	1.71	9

PB: passband; TB:transition band; SB: stopband

TABLE XVII
DESIGN RESULTS FOR EXAMPLE 7 FOR FIXED GROUP DELAY (HIGHPASS FILTER)

Parameters	Design B-1	Design B-2 $\Gamma_{tb} = 1$	Method in [12]
Filter order	14	14	14
Max PB ripple, dB	0.491	0.491	0.720
Min SB atten., dB	31	31	27.36
Max TB gain, dB	7.08	0.022	0.155
τ_{avg}	12	12	13.74
Q_τ	0.018	4.8	16.4
Parameters	Design B-1	Design B-3	Method in [13]
Filter order	14	14	14
Max PB ripple, dB	0.491	0.491	0.492
Min SB atten., dB	31	31	30.7
Max TB gain, dB	7.08	3.79	3.86
τ_{avg}	12	12	12.79
Q_τ	0.018	0.028	9

PB: passband; TB:transition band; SB: stopband

TABLE XVIII
LOWPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 8

Parameters	Values
Maximum passband ripple, dB	0.1
Minimum stopband attenuation, dB	44
Passband edge, rad/s	0.5π
Stopband edge, rad/s	0.55π
Maximum pole radius	0.98
Group delay (samples)	15

TABLE XIX
DESIGN RESULTS FOR EXAMPLE 8 FOR VARIABLE GROUP DELAY (LOWPASS FILTER)

Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [14] 1st Design
Filter order	18	18	18
Max PB ripple, dB	0.0904	0.09	0.0995
Min SB atten., dB	45.5	45.5	44.77
Max TB gain, dB	7	0	11.25
τ_{avg}	19.6	20.69	14.45
Q_τ	0.204	3.35	5.65
Parameters	Design A-1	Design A-2 $\Gamma_{tb} = 1$	Method in [14] 2nd Design
Filter order	18	18	18
Max PB ripple, dB	0.0904	0.09	0.0907
Min SB atten., dB	45.5	45.5	45.36
Max TB gain, dB	7	0	21.90
τ_{avg}	19.6	20.69	14.46
Q_τ	0.204	3.35	5.09

PB: passband; TB:transition band; SB: stopband

TABLE XX
DESIGN RESULTS FOR EXAMPLE 8 FOR FIXED GROUP DELAY (LOWPASS FILTER)

Parameters	Design B-1	Method in [14] 1st Design	Method in [14] 2nd Design
Filter order	18	18	18
Max PB ripple, dB	0.09	0.0995	0.0907
Min SB atten., dB	45.5	44.77	45.36
Max TB gain, dB	7.43	11.25	21.90
τ_{avg}	15	14.45	14.46
Q_τ	3.18	5.65	5.09

PB: passband; TB:transition band; SB: stopband

TABLE XXI
BANDPASS DIGITAL FILTER SPECIFICATIONS FOR EXAMPLE 10

Parameters	Values
Maximum passband ripple, dB	1.0
Minimum stopband attenuation, dB	68
Low stopband edge, rad/s	1
Low passband edge, rad/s	1.1
High passband edge, rad/s	1.9
High stopband edge, rad/s	2
Maximum pole radius	0.991

TABLE XXII
COMPARISON BETWEEN DESIGN A-1 OF THE IIR FILTERS AND AN
EQUIVALENT FIR FILTER (EXAMPLE 10)

Parameters	Elliptic Filter	IIR Filter 1	IIR Filter 2	IIR Filter 3	IIR Filter 4	FIR Filter
Total filter order	14	20	22	24	26	142
Max PB ripple, dB	1	0.98	0.98	0.98	0.78	0.98
Min SB atten., dB	68.5	68	68	68	68.3	68
Max TB gain, dB	-1	11.73	10.71	11.36	10.64	-0.5
τ_{avg}	63.33	29.5	34.1	38.57	41.72	71
Q_τ	83.77	19.85	6.95	1.89	0.58	0
No. of multiplications	29	41	45	49	53	72
No. of additions	28	40	44	48	52	142
No of delays	14	20	22	24	26	142

PB: passband; TB:transition band; SB: stopband

TABLE XXIII
COMPARISON BETWEEN DESIGN A-2 OF THE IIR FILTERS AND AN
EQUIVALENT FIR FILTER (EXAMPLE 10)

Parameters	Elliptic Filter	IIR Filter 1	IIR Filter 2	IIR Filter 3	IIR Filter 4	FIR Filter
Total filter order	14	22	24	26	28	142
Max PB ripple, dB	1	0.98	0.98	0.98	0.98	0.98
Min SB atten., dB	68.5	68	68	68	68	68
Max TB gain, dB	-1	0.01	0.01	0.02	0.01	-0.5
τ_{avg}	63.33	33.66	38.59	44.75	49.27	71
Q_τ	83.77	17.38	9.6	3.6	0.49	0
No. of multiplications	29	45	49	53	57	72
No. of additions	28	44	48	52	56	142
No of delays	14	22	24	26	28	142

PB: passband; TB:transition band; SB: stopband

FIR design that satisfies the same specifications. The optimization was carried out for various IIR-filter orders by varying the number of additional biquadratic sections. The FIR filter was designed using the *weighted-Chebyshev* method described in Chapter 15 of [3]. The results obtained are presented in Tables XXII and XXIII . In these tables, we also include the numbers of multiplications, additions, and unit delays per sampling period required by each design. These numerical values correspond to a metric of the computational effort required in a software implementation or the complexity of the hardware in a hardware implementation since multiplications, additions, and unit delays translate directly into multipliers, adders, and unit-delay elements, respectively. We have assumed a cascade realization of SOSs both for the IIR and FIR filters. For the IIR filters we have assumed a direct canonic realization which would require a total of $4J + 1$ multiplications, $4J$ additions,

and $2J$ unit delays in general where J is the number of sections [3]. In the case of the FIR filter, $J + 1$ multiplications, $2J$ additions, and $2J$ unit delays would be required in view of the symmetry property of the transfer function coefficients. From Tables XXII and XXIII , we observe a clear trade-off between filter complexity and group delay versus maximum group-delay deviation. Evidently, an IIR design would offer significant reduction in the number of arithmetic operations and group delay of the filter but at the cost of a higher maximum group-delay deviation. For most applications, a perfectly linear-phase response is not required and a value of Q_τ in the range of 1 to 10, depending on the application, would be entirely acceptable. In such applications, an IIR design that is much more economical than an FIR design would be possible. Note that the complexity of the IIR design is reduced and its efficiency is increased as the allowable value of Q_τ is increased. This constitutes a most valuable trade-off that would allow a filter designer to design the most economical filter that would satisfy the required specifications for the intended application.

The above design examples have shown that the proposed design method yields filters that satisfy arbitrary prescribed amplitude-response specifications with the lowest maximum group-delay deviation compared to those achieved with the competing design methods considered. The solution obtained sometimes depends on the initialization filter and group delay used, which implies that global convergence cannot be guaranteed. In effect, the proposed method gives quality suboptimal designs that may sometimes be globally optimal.

V. CONCLUSION

A method for the design of nearly linear-phase IIR digital filters that satisfy prescribed specifications has been described. In the proposed method, the group-delay deviation is minimized under the constraints that the passband ripple and minimum stopband attenuation meet the specifications and either a prescribed or an optimized group delay can be achieved. By designing the filter as a cascade of second-order sections, a nonrestrictive stability constraint characterized by a set of linear inequality constraints can be incorporated in the optimization algorithm. An additional feature of the method, which is very useful in certain applications, is the inherent capability of constraining the maximum gain in transition bands to be below a prescribed value. This facilitates the elimination of transition-band anomalies which sometimes occur in filters designed by optimization.

Experimental results have shown that the nearly linear-phase IIR filters designed using the proposed method have a much lower maximum group-delay deviation for the same passband ripple and minimum stopband attenuation when compared with several filters designed with state-of-the-art competing methods. It has also been demonstrated that nearly linear-phase IIR filters offer some substantial advantages when compared with their exact linear-phase FIR counterparts such as lower group delay and filter complexity without compromising the required amplitude-response specifications.

REFERENCES

[1] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms And Engineering Applications*. New York: Springer, 2007.

[2] D. J. Shpak, *MATLAB™ functions iirlpnorm and iirlpnormc* in the *MATLAB™ Filter Design Toolbox*. Natick, MA, USA: The Math-Works, 2002.

[3] A. Antoniou, *Digital Signal Processing: Signals, Systems, and Filters*. New York: McGraw-Hill, 2005.

[4] C. Charalambous and A. Antoniou, "Equalisation of recursive digital filters," *Proc. Inst. Elect. Eng.*, vol. 127, no. 5, pp. 219–225, Oct. 1980.

[5] D. Guindon, D. J. Shpak, and A. Antoniou, "Design methodology for nearly linear-phase recursive digital filters by constrained optimization," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 7, pp. 1719–1731, Jul. 2010.

[6] W.-S. Lu, "Design of recursive digital filters with prescribed stability margin: A parameterization approach," *IEEE Trans. Circuits Syst. I: Fund. Theory Appl.*, vol. 45, pp. 1289–1298, Sep. 1998.

[7] W.-S. Lu, "An argument-principle based stability criterion and application to the design of IIR digital filters," in *Proc. IEEE Intern. Symp. Circuits Syst.*, 2006, pp. 4431–4434.

[8] J. L. Sullivan and J. W. Adams, "PCLS IIR digital filters with simultaneous frequency response magnitude and group delay specifications," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 2853–2861, Nov. 1998.

[9] A. Jiang and H. K. Kwan, "IIR digital filter design with new stability constraint based on argument principle," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 56, pp. 583–593, Mar. 2009.

[10] X. Lai and Z. Lin, "Minimax design of IIR digital filters using a sequential constrained least-squares method," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3901–3906, Jul. 2010.

[11] X. Lai and Z. Lin, "Minimax phase error design of IIR digital filters with prescribed magnitude and phase responses," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 980–991, Feb. 2012.

[12] A. Jiang and H. K. Kwan, "Minimax design of IIR digital filters using iterative SOCP," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 6, pp. 1326–1337, Jun. 2010.

[13] A. Jiang and H. K. Kwan, "Minimax design of IIR digital filters using SDP relaxation technique," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, pp. 378–390, Feb. 2010.

[14] X. Lai, Z. Lin, and H. K. Kwan, "A sequential minimization procedure for minimax design of IIR filters based on second-order factor updates," *IEEE Trans. Circuits Syst.-II: Exp. Briefs*, vol. 58, no. 1, pp. 51–55, Jan. 2011.

[15] W.-S. Lu and A. Antoniou, "Design of digital filters and filter banks by optimization: A state of the art review," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Tampere, Finland, Sep. 2000, pp. 351–354.

[16] B. Dumitrescu and R. Niemist, "Multistage IIR filter design using convex stability domains defined by positive realness," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 962–974, Apr. 2004.

[17] M. C. Lang, "Least-squares design of IIR filters with prescribed magnitude and phase responses and a pole radius constraint," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3109–3121, Nov. 2000.

[18] W.-S. Lu and T. Hinamoto, "Optimal design of IIR digital filters with robust stability using conic-quadratic-programming updates," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1581–1592, Jun. 2003.

[19] J. F. Sturm, "Using SeDuMi1.02, a MATLAB™ toolbox for optimization over symmetric cones," *Optimiz. Methods Softw.*, vol. 11-12, pp. 625–653, 1999.

[20] [Online]. Available: http://www.ece.uvic.ca/~dsp/

[21] I. Kale, J. Gryka, G. D. Cain, and B. Beliczynski, "FIR filter order reduction: Balanced model truncation and Hankel-norm optimal approximation," *IEEE Proc. Vis. Image Signal Process.*, vol. 141, pp. 168–174, Jun. 1994.

[22] L. Prenobo and L. M. Silverman, "Model reduction via balanced state space representation," *IEEE Trans. Autom. Control*, vol. AC-27, no. 4, pp. 382–387, Apr. 1982.

[23] A. Antoniou, "Improved minimax optimization algorithms and their application in the design of recursive digital filters," *IEE Proc.*, vol. 138, no. 6, pp. 724–730, Dec. 1991.

[24] [Online]. Available: http://www.ece.uvic.ca/~andreas/JournalPapers/Poles_and_Zeros_and_Additional_Results.pdf

Rajeev C. Nongpiur (S'01–A'05–M'12) received the B.Tech. degree in electronics and communications engineering from the Indian Institute of Technology, Kharagpur, in 1998 and the Ph.D. degree from the University of Victoria, British Columbia, Canada, in 2005.

From 1998 to 2000, he worked as a systems engineer at Wipro Technologies, from 2004 to 2008 as a Research Scientist at QNX Software Systems, and from 2008 to 2010 as Senior DSP Engineer with Unication Co., Ltd., Vancouver, Canada. He is currently serving as Research Associate in the Department of Electrical and Computer Engineering, University of Victoria, British Columbia, Canada. His research interests are in the areas of signal processing for digital communications, speech, audio, and biomedical applications. He is the author of more than 15 patents in the area of audio signal processing.

Dr. Nongpiur is a member of the IEEE Circuits and Systems and Signal Processing Societies.

Dale J. Shpak (S'79–M'86–SM'09) received the B.Sc. (electrical engineering) degree from the University of Calgary, Canada, in 1980. From 1980 to 1982, he was an engineer for the City of Calgary Electric System while he received the M.Eng. degree in electronics also from the University of Calgary. In 1989, he received the Ph.D. degree from the University of Victoria, Canada.

In 1988, he joined the Department of Engineering, Royal Roads Military College, as a Professor and returned to industry when it closed in 1995. Between 1982 and 1987, he performed research on computer systems, microelectronics, and DSP algorithms and implementation. Since 1989, he has held a faculty position with the University of Victoria in addition to his other professional activities. As an Adjunct Professor of Electrical and Computer Engineering, he receives ongoing NSERC funding for research programs with his graduate students. He joined the Department of Computer Science, Camosun College, in 1999. He has instructed more than 30 different courses including object-oriented programming, computer networks, digital circuit design, digital filters, materials science, software engineering, and real-time and concurrent systems. He held several positions in industry where he developed algorithms, software, circuits, networking systems, and embedded systems. Since 1984, he has served as a consultant and develops software and embedded systems for products including audio processing, wireless sensing and control, and remote sensing. He is a principal developer of award-winning products, including the Filter Design Toolbox for MATLAB. His principal research interests are in the areas of signal processing for communications and audio, design and implementation of embedded systems, and digital filter design. He also develops free software for music education.

Dr. Shpak is a Member of the Association of Professional Engineers of the Province of British Columbia.
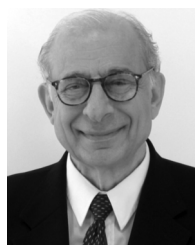
Andreas Antoniou (M'69–SM'79–F'82–LF'04) received the B.Sc.(Eng.) and Ph.D. degrees in electrical engineering from the University of London, U.K., in 1963 and 1966, respectively.

He taught at Concordia University from 1970 to 1983, was the founding Chair of the Department of Electrical and Computer Engineering, University of Victoria, B.C., Canada, from 1983 to 1990, and is currently Professor Emeritus. His teaching and research interests are in the area of digital signal processing. He is the author of *Digital Signal Processing: Signals, Systems, and Filters* (New York: McGraw-Hill, 2005), and coauthor with Wu-Sheng Lu of *Practical Optimization: Algorithms and Engineering Applications* (New York: Springer, 20070.

Dr. Antoniou is a Fellow of the IET. He served first as an Associate Editor and after, as Editor–in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1983 to 1987, as a Distinguished Lecturer of the IEEE Signal Processing and the Circuits and Systems Societies during 2003–2004 and 2006–2007, respectively, and as General Chair of the 2004 International Symposium on Circuits and Systems. He was awarded the CAS Golden Jubilee Medal by the IEEE Circuits and Systems Society, the B.C. Science Council Chairman's Award for Career Achievement for 2000, the Doctor Honoris Causa degree by the National Technical University, Athens, Greece, in 2002, the 2005 IEEE Circuits and Systems Society Technical Achievement Award, the 2008 IEEE Canada Outstanding Engineering Educator Silver Medal, the 2009 IEEE Circuits and Systems Society Education Award, the 2011 Craigdarroch Gold Medal for Career Achievement and the 2011 Legacy Award for Research both from the University of Victoria.