Roundoff Noise Reduction in State-Space Digital Filters Using High-Order Error Feedback and Realization

Takao Hinamoto and Akimitsu Doi
Hiroshima Institute of Technology
Hiroshima 731-5193, Japan
Email: hinamoto@ieee.org   doi@cc.it-hiroshima.ac.jp

Wu-Sheng Lu
University of Victoria
Victoria, BC, Canada V8W 3P6
Email: wslu@ece.uvic.ca

Abstract— **This paper is concerned with the reduction of roundoff noise effects subject to l_2-scaling constraints for state-space digital filters by means of high-order error feedback and realization. First, the roundoff noise gain is mimimized subject to l_2-scaling constraints by choosing the coordinate transformation matrix appropriately. Optimal high-order error-feedback matrices are then determined so as to minimize the roundoff noise gain in the state-space digital filter transformed by the resulting coordinate transformation matrix. Finally a numerical example is presented to illustrate the utility of the proposed technique.**

I. INTRODUCTION

In the implementation of IIR digital filters in fixed-point arithmetic, the problem of reducing the effects of roundoff noise at the filter output is an important subject of study. It is known that error feedback (EF) is a useful tool for the reduction of finite-word-length (FWL) effects in IIR digital filters. Many EF techniques have been reported in the past for IIR digital filters [1]-[10]. The roundoff noise can also be reduced by introducing a delta operator to IIR digital filters [11]-[13] or by applying a new structure based on the concept of polynomial operators for digital filter implementation [14]. Another useful approach is to construct the state-space filter structure such that the roundoff noise gain is minimized by applying a linear transformation to state-space coordinates subject to l_2-scaling constraints [15]-[18]. As a natural extension of the aforementioned methods, efforts have been made to develop new methods that combine EF and realization, for achieving better performance [19]-[20]. Separately-optimized analytical algorithms have been proposed for state-space digital filters [19]. In [19], jointly-optimized iterative algorithms have also been considered for filters with a general or scalar EF matrix. In [20], a jointly-optimized iterative algorithm has been developed for state-space digital filters with a general, diagonal, or scalar EF matrix by applying a quasi-Newton method.

Unlike the techniques in [19]-[20], in this paper we consider state-space digital filters that utilize high-order EF and realization and propose an algorithm that separately optimizes the EF and realization which minimize roundoff noise subject to l_2-scaling constraints. To this end, an appropriate coordinate transformation matrix is first identified by minimizing the roundoff noise gain. Then high-order EF matrices are determined so as to minimize the round-off noise gain in the state-space digital filter constructed by the above coodinate transformation matrix where the high-order EF matrices are diagonal or scalar matrices. Finally, a numerical example is presented to demonstrate its performance.

II. PROBLEM FORMULATION

Consider a stable, controllable and observable state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_n$ described by

$$
\begin{aligned}
\boldsymbol{x}(k+1) &= \boldsymbol{A}\boldsymbol{x}(k) + \boldsymbol{b}u(k) \\
y(k) &= \boldsymbol{c}\boldsymbol{x}(k) + du(k)
\end{aligned}
\tag{1}
$$

where $\boldsymbol{x}(k)$ is an $n \times 1$ state-variable vector, $u(k)$ is a scalar input, $y(k)$ is a scalar output, and $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$ and d are real constant matrices of appropriate dimensions.

By taking the quantizations performed before matrix-vector multiplication into account, an finite-word-length (FWL) implementation of the state-space model in (1) with error feed-forward and high-order EF can be expressed as

$$
\begin{aligned}
\tilde{\boldsymbol{x}}(k+1) &= \boldsymbol{A}\boldsymbol{Q}[\tilde{\boldsymbol{x}}(k)] + \boldsymbol{b}u(k) + \sum_{i=1}^{N} \boldsymbol{D}_i e(k-i+1) \\
\tilde{y}(k) &= \boldsymbol{c}\boldsymbol{Q}[\tilde{\boldsymbol{x}}(k)] + du(k) + \boldsymbol{h}e(k)
\end{aligned}
\tag{2}
$$

where $\boldsymbol{h}$ and $\boldsymbol{D}_i$ for $i = 1, 2, \cdots, n$ are referred to as a $1 \times n$ *error-forward vector* and an $n \times n$ *EF matrix*, respectively, and

$$
\boldsymbol{e}(k) = \tilde{\boldsymbol{x}}(k) - \boldsymbol{Q}[\tilde{\boldsymbol{x}}(k)].
$$

In model (2), all coefficient matrices $\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}$, and d are assumed to have an exact fractional B_c bit representation. The FWL state-variable vector $\tilde{\boldsymbol{x}}(k)$ and output $\tilde{y}(k)$ all have a B bit fractional representation, while input $u(k)$ is a $(B-B_c)$-bit fraction. The quantizer $\boldsymbol{Q}[\cdot]$ in (2) rounds the B bit fraction $\tilde{\boldsymbol{x}}(k)$ to $(B-B_c)$ bits after completing the multiplications and additions, where the sign bit is not counted. It is assumed that the roundoff error vector $\boldsymbol{e}(k)$ can be modeled as a zero-mean

noise process with covariance $\sigma^2 \boldsymbol{I}_n$. Subtracting (2) from (1) yields

$$\Delta \boldsymbol{x}(k+1) = \boldsymbol{A}\Delta \boldsymbol{x}(k) + \boldsymbol{A}e(k) - \sum_{i=1}^{N} \boldsymbol{D}_i e(k-i+1)$$

$$\Delta y(k) = \boldsymbol{c}\Delta \boldsymbol{x}(k) + (\boldsymbol{c}-\boldsymbol{h})e(k) \tag{3}$$

where $\Delta \boldsymbol{x}(k) = \boldsymbol{x}(k) - \tilde{\boldsymbol{x}}(k)$ and $\Delta y(k) = y(k) - \tilde{y}(k)$. By taking the z-transform on both sides of (3) and setting $\Delta \boldsymbol{x}(0) = \boldsymbol{0}$, we have

$$\Delta Y(z) = \boldsymbol{H}_e(z)\boldsymbol{E}(z)$$

$$\boldsymbol{H}_e(z) = \boldsymbol{c}(z\boldsymbol{I}_n - \boldsymbol{A})^{-1}\left(\boldsymbol{A} - \sum_{i=1}^{N} \boldsymbol{D}_i z^{-i+1}\right) \tag{4}$$
$$+ \boldsymbol{c} - \boldsymbol{h}$$

where $\Delta Y(z)$ and $\boldsymbol{E}(z)$ represent the z-transforms of $\Delta y(k)$ and $e(k)$, respectively. $\boldsymbol{H}_e(z)$ in (4) is written as

$$\boldsymbol{H}_e(z) = \sum_{k=1}^{\infty} \boldsymbol{c}\left(\boldsymbol{A}^k - \sum_{i=1}^{N} \boldsymbol{A}^{k-i}\boldsymbol{D}_i\right)z^{-k} + \boldsymbol{c} - \boldsymbol{h} \tag{5}$$

where $\boldsymbol{A}^i = \boldsymbol{0}$ for $i < 0$. The normalized noise gain $J_{e1}(\boldsymbol{h}, \boldsymbol{D}) = \sigma_{out}^2/\sigma^2$ with $\boldsymbol{D} = [\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_N]$ is then defined by

$$J_{e1}(\boldsymbol{h}, \boldsymbol{D}) = \mathrm{tr}\left[\frac{1}{2\pi j}\oint_{|z|=1}\boldsymbol{H}_e^*(z)\boldsymbol{H}_e(z)\frac{dz}{z}\right] \tag{6}$$

Substituting (5) into (6) yields

$$J_{e1}(\boldsymbol{h}, \boldsymbol{D})$$
$$= \mathrm{tr}\left[\boldsymbol{A}^T\boldsymbol{W}_o\boldsymbol{A} - \sum_{i=1}^{N}\left\{(\boldsymbol{A}^T)^i\boldsymbol{W}_o\boldsymbol{D}_i + \boldsymbol{D}_i^T\boldsymbol{W}_o\boldsymbol{A}^i\right\}\right.$$
$$+ \sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{D}_i^T\left\{(\boldsymbol{A}^T)^{j-i}\boldsymbol{W}_o + \boldsymbol{W}_o\boldsymbol{A}^{i-j}\right\}\boldsymbol{D}_j$$
$$\left.- \sum_{i=1}^{N}\boldsymbol{D}_i^T\boldsymbol{W}_o\boldsymbol{D}_i\right] + (\boldsymbol{c}-\boldsymbol{h})(\boldsymbol{c}-\boldsymbol{h})^T \tag{7}$$

where $\boldsymbol{W}_o$ is the observability Gramian of the filter that can be obtained by solving the Lyapunov equation

$$\boldsymbol{W}_o = \boldsymbol{A}^T\boldsymbol{W}_o\boldsymbol{A} + \boldsymbol{c}^T\boldsymbol{c}.$$

Assuming that EF matrcies $\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_N$ are diagonal, the normalized noise gain in (7) can be written as

$$J_{e1}(\boldsymbol{h}, \boldsymbol{D}) = \mathrm{tr}\left[\boldsymbol{A}^T\boldsymbol{W}_o\boldsymbol{A} - 2\sum_{i=1}^{N}\boldsymbol{W}_o\boldsymbol{A}^i\boldsymbol{D}_i\right.$$
$$\left.+ \sum_{i=1}^{N}\sum_{j=1}^{N}\boldsymbol{W}_o\boldsymbol{A}^{|i-j|}\boldsymbol{D}_i\boldsymbol{D}_j\right] \tag{8}$$
$$+ (\boldsymbol{c}-\boldsymbol{h})(\boldsymbol{c}-\boldsymbol{h})^T.$$

It is noted that the l_2-scaling constraints on the state variables involves the controllability Gramian $\boldsymbol{K}_c$ of the filter computed by solving the Lyapunov equation

$$\boldsymbol{K}_c = \boldsymbol{A}\boldsymbol{K}_c\boldsymbol{A}^T + \boldsymbol{b}\boldsymbol{b}^T.$$

A different yet equivalent state-space description of (1), $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_n$, can be obtained via a coordinate transformation $\overline{\boldsymbol{x}}(k) = \boldsymbol{T}^{-1}\boldsymbol{x}(k)$ with

$$\overline{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}, \quad \overline{\boldsymbol{b}} = \boldsymbol{T}^{-1}\boldsymbol{b}, \quad \overline{\boldsymbol{c}} = \boldsymbol{c}\boldsymbol{T}. \tag{9}$$

Accordingly, the controllability and observability Gramians for the new state-space model $(\overline{\boldsymbol{A}}, \overline{\boldsymbol{b}}, \overline{\boldsymbol{c}}, d)_n$ can be expressed as

$$\overline{\boldsymbol{K}}_c = \boldsymbol{T}^{-1}\boldsymbol{K}_c\boldsymbol{T}^{-T} \quad \text{and} \quad \overline{\boldsymbol{W}}_o = \boldsymbol{T}^T\boldsymbol{W}_o\boldsymbol{T}, \tag{10}$$

respectively. The l_2-scaling constraints are imposed on the state-variable vector $\overline{\boldsymbol{x}}(k)$ so that

$$(\overline{\boldsymbol{K}}_c)_{ii} = (\boldsymbol{T}^{-1}\boldsymbol{K}_c\boldsymbol{T}^{-T})_{ii} \text{ for } i = 1, 2, \cdots, n. \tag{11}$$

Based on the analysis, the problem at hand is to design the EF diagonal matrices $\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_N$ and the optimal coordinate transformation matrix $\boldsymbol{T}$ separately minimizing

$$J_{e2}(\boldsymbol{T}, \boldsymbol{D}) = \mathrm{tr}\left[\overline{\boldsymbol{A}}^T\overline{\boldsymbol{W}}_o\overline{\boldsymbol{A}} - 2\sum_{i=1}^{N}\overline{\boldsymbol{W}}_o\overline{\boldsymbol{A}}^i\boldsymbol{D}_i\right.$$
$$\left.+ \sum_{i=1}^{N}\sum_{j=1}^{N}\overline{\boldsymbol{W}}_o\overline{\boldsymbol{A}}^{|i-j|}\boldsymbol{D}_i\boldsymbol{D}_j\right] \tag{12}$$

subject to l_2-scaling constraints in (11) where the error feed-forward vector $\boldsymbol{h}$ is assumed to be chosen as $\boldsymbol{h} = \overline{\boldsymbol{c}}$.

III. ANALYTIC METHOD FOR SEPARATE OPTIMIZATION OF (12) SUBJECT TO (11)

The separate minimization is carried out in two major steps. First, we fix the EF matrices to $\boldsymbol{D}_i = \boldsymbol{0}$ for $i = 1, 2, \cdots, N$ in (12) so that the objective function is reduced to $J_{e2}(\boldsymbol{T}, \boldsymbol{0}) = \mathrm{tr}[\overline{\boldsymbol{A}}^T\overline{\boldsymbol{W}}_o\overline{\boldsymbol{A}}]$ which is minimized with respect to matrix $\boldsymbol{T}$ subject to the l_2-scaling constraints in (11). Second, with $\boldsymbol{T}$ optimized in the first step, (12) is minimized under the fixed $\boldsymbol{T}$ with respect to matrices $\boldsymbol{D}_1, \boldsymbol{D}_2, \cdots, \boldsymbol{D}_N$. To perform the first step, we define the Lagrange function

$$J_o(\boldsymbol{P}, \lambda) = \mathrm{tr}[\boldsymbol{A}^T\boldsymbol{W}_o\boldsymbol{A}\boldsymbol{P}] - \lambda\left(\mathrm{tr}[\boldsymbol{K}_c\boldsymbol{P}^{-1} - n\right) \tag{13}$$

where $\boldsymbol{P} = \boldsymbol{T}\boldsymbol{T}^T$. The optimal coordinate transformation matrix $\boldsymbol{T}$ can be determined by solving the equations

$$\frac{\partial J_o(\boldsymbol{P}, \lambda)}{\partial \boldsymbol{P}} = \boldsymbol{0}, \quad \frac{\partial J_o(\boldsymbol{P}, \lambda)}{\partial \lambda} = 0 \tag{14}$$

which is led to [19]

$$\boldsymbol{T} = \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{n}\theta_i\right)^{\frac{1}{2}}\boldsymbol{V}^{-\frac{1}{2}}\left[\boldsymbol{V}^{\frac{1}{2}}\boldsymbol{K}_c\boldsymbol{V}^{\frac{1}{2}}\right]^{\frac{1}{4}}\boldsymbol{U} \tag{15}$$

where

$$\boldsymbol{V} = \boldsymbol{A}^T\boldsymbol{W}_o\boldsymbol{A}$$

$\theta_1^2, \theta_2^2, \cdots, \theta_n^2$ are the eigenvalues of matrix $\boldsymbol{K}_c \boldsymbol{W}_c$ and $\boldsymbol{U}$ is an arbitrary $n \times n$ orthogonal matrix.

Suppose that the eigenvalue-eigenvector decomposition of $\left[\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{K}_c \boldsymbol{V}^{\frac{1}{2}} \right]^{\frac{1}{2}}$ can be written as

$$\left[\boldsymbol{V}^{\frac{1}{2}} \boldsymbol{K}_c \boldsymbol{V}^{\frac{1}{2}} \right]^{\frac{1}{2}} = \boldsymbol{R} \, \mathrm{diag}\{\theta_1, \theta_2, \cdots, \theta_n\} \boldsymbol{R}^T \quad (16)$$

where $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}_n$. By numerical manipulations, we can obtain an $n \times n$ orthogonal matrix $\boldsymbol{S}$ such that [17]

$$\boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^{\mathbf{T}} = \begin{bmatrix} 1 & * & \cdots & * \\ * & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ * & \cdots & * & 1 \end{bmatrix} \quad (17)$$

where

$$\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$$
$$\lambda_i = \frac{n\theta_i}{\theta_1 + \theta_2 + \cdots + \theta_n} \quad \text{for} \quad i = 1, 2, \cdots, n.$$

Below we consider two cases of the EF matrices.

Case 1: $\boldsymbol{D}_p$ *is a diagonal matrix for* $p = 1, 2, \cdots, N$

In this case, matrix $\boldsymbol{D}_p$ assumes the form

$$\boldsymbol{D}_p = \mathrm{diag}\{d_{p1}, d_{p2}, \cdots, d_{pn}\} \quad \text{for} \quad p = 1, 2, \ldots, N. \quad (18)$$

It follows that

$$\frac{\partial J_{e2}(\boldsymbol{T}, \boldsymbol{D})}{\partial d_{pl}} = -2(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^i)_{ll} + 2\sum_{k=1}^{N} d_{kl}(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^{|p-k|})_{ll}$$
$$= 0 \quad \text{for} \quad l = 1, 2, \cdots, n. \quad (19)$$

As a result, matrix $\boldsymbol{D}_p$ can be derived from

$$\begin{bmatrix} d_{1l} \\ d_{2l} \\ \vdots \\ d_{Nl} \end{bmatrix} = \begin{bmatrix} p_{0l} & p_{1l} & \cdots & p_{(N-1)l} \\ p_{1l} & p_{0l} & \cdots & p_{(N-2)l} \\ \vdots & \vdots & \ddots & \vdots \\ p_{(N-1)l} & p_{(N-2)l} & \cdots & p_{0l} \end{bmatrix}^{-1} \begin{bmatrix} p_{1l} \\ p_{2l} \\ \vdots \\ p_{Nl} \end{bmatrix} \quad (20)$$

where $p_{pl} = \left(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^p \right)_{ll}$.

Case 2: $\boldsymbol{D}_p$ *is a scalar matrix for* $p = 1, 2, \cdots, N$

In this case, the EF matrices assume the form $\boldsymbol{D}_p = \alpha_p \boldsymbol{I}_n$ with a scalar α_p. It follows that

$$\frac{\partial J_{e2}(\boldsymbol{T}, \boldsymbol{D})}{\partial \alpha_p} = -2\,\mathrm{tr}\left(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^p \right) + 2\sum_{k=1}^{N} \alpha_k \, \mathrm{tr}\left(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^{|p-k|} \right)$$
$$= 0 \quad \text{for} \quad p = 1, 2, \cdots, N. \quad (21)$$

Therefore, scalars $\alpha_1, \alpha_2, \cdots, \alpha_N$ can be computed using

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} = \begin{bmatrix} q_0 & q_1 & \cdots & q_{N-1} \\ q_1 & q_0 & \cdots & q_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ q_{N-1} & q_{N-2} & \cdots & q_0 \end{bmatrix}^{-1} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_N \end{bmatrix} \quad (22)$$

where $q_p = \mathrm{tr}\left(\overline{\boldsymbol{W}}_o \overline{\boldsymbol{A}}^p \right)$.

IV. A NUMERICAL EXAMPLE

As a numerical example, we consider a state-space digital filter $(\boldsymbol{A}, \boldsymbol{b}, \boldsymbol{c}, d)_3$ specified by

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.453770 & -1.556160 & 1.974860 \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} 0 & 0 & 0.2420961 \end{bmatrix}^T$$

$$\boldsymbol{c} = \begin{bmatrix} 0.095706 & 0.095086 & 0.327556 \end{bmatrix}$$

$$d = 0.015940.$$

The controllability and observability Gramians $\boldsymbol{K}_c$ and $\boldsymbol{W}_o$ of the above filter were computed as

$$\boldsymbol{K}_c = \begin{bmatrix} 1.000000 & 0.872501 & 0.562822 \\ 0.872501 & 1.000000 & 0.872501 \\ 0.562822 & 0.872501 & 1.000000 \end{bmatrix}$$

$$\boldsymbol{W}_o = \begin{bmatrix} 0.820742 & -2.035323 & 1.628159 \\ -2.035323 & 5.307270 & -4.264912 \\ 1.628159 & -4.264912 & 3.941488 \end{bmatrix}.$$

The noise gain of the filter with no error feedforward and no EF was then computed from (8) as

$$J_{e1}(\boldsymbol{0}, \boldsymbol{0}) = 10.0695.$$

The optimal coordinate transformation matrix was constructed using (15) as

$$\boldsymbol{T} = \begin{bmatrix} -1.066701 & 1.326232 & 0.257799 \\ -0.321892 & 1.155334 & -0.070508 \\ 0.252063 & 0.827689 & -0.002870 \end{bmatrix}.$$

With $\boldsymbol{T}$ found and fixed, we now consider two cases: one employs a single EF matrix $\boldsymbol{D}_1$ (i.e. $N = 1$) and the other uses two EF matrices $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ (i.e. $N = 2$). In each case, EF matrix assumes the form of either diagonal or scalar matrices. The EF matrices were optimized using the method presented in Section 3, and the results in terms of noise gain $J_{e2}(\boldsymbol{T}, \boldsymbol{D})$ are summarized in Table I, where " 3-Bit Quantization" refers to value of $J_{e2}(\boldsymbol{T}, \boldsymbol{D})$ where the fraction part of each component of $\boldsymbol{D}$ was rounded to 3 bits in a power-of-two representation. From the table, it is observed that (i) Diagonal EF matrices offer improved performance relative to their scalar counterparts; (ii) employing more EF matrices helps reduce the noise gain in a significant way, and (iii) compared with their infinite-precision counterparts, the use of quantized EF matrices reduces implementation complexity with only slightly degraded performance.

TABLE I
SEPARATE OPTIMIZATION

N	$\boldsymbol{D}$	Infinite Precision	3-Bit Quantization
1	Diagonal	0.1209	0.1252
	Scalar	0.1325	0.1368
2	Diagonal	0.0612	0.0659
	Scalar	0.0992	0.1038

V. CONCLUSION

The separate optimization problem of high-order EF and realization for roundoff noise minimization subject to l_2-scaling constraints in state-space digital filters has been investigated. After choosing the optimal coordinate transformation matrix under the l_2-scaling constraints, high-order EF matrices are determined so as to minimize the roundoff noise gain in the optimal realization where high-order EF matrices are diagonal or scalar matrices. A numerical example has been presented to demonstrate the validity and effectiveness of the proposed technique. A technique for solving the joint optimization problem of high-order EF and realization for roundoff noise minimization in state-space digital filters has been reported in [21].

REFERENCES

[1] H. A. Spang, III and P. M. Shultheiss, "Reduction of quantizing noise by use of feedback," *IRE Trans. Commun. Syst.*, vol. CS-10, pp. 373-380, Dec. 1962.

[2] T. Thong and B. Liu, "Error spectrum shaping in narrowband recursive digital filters," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 25, pp. 200-203, Apr. 1977.

[3] T. L. Chang and S. A. White, "An error cancellation digital filter structure and its distributed-arithmetic implementation," *IEEE Trans. Circuits Syst.*, vol. 28, pp. 339-342, Apr. 1981.

[4] D. C. Munson and D. Liu, "Narrowband recursive filters with error spectrum shaping," *IEEE Trans. Circuits Syst.*, vol. 28, pp. 160-163, Feb. 1981.

[5] W. E. Higgins and D. C. Munson, "Noise reduction strategies for digital filters: Error spectrum shaping versus the optimal linear state-space formulation," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 30, pp. 963-973, Dec. 1982.

[6] M. Renfors, "Roundoff noise in error-feedback state-space filters," *Proc. Int. Conf. Acoustics, Speech, Signal Processing* (ICASSP'83), pp. 619-622, Apr. 1983.

[7] W. E. Higgins and D. C. Munson, "Optimal and suboptimal error-spectrum shaping for cascade-form digital filters," *IEEE Trans. Circuits Syst.*, vol. 31, pp. 429-437, May 1984.

[8] T. I. Laakso and I. O. Hartimo, "Noise reduction in recursive digital filters using high-order error feedback," *IEEE Trans. Signal Processing*, vol. 40, pp. 1096-1107, May 1992.

[9] P. P. Vaidyanathan, "On error-spectrum shaping in state-space digital filters," *IEEE Trans. Circuits Syst.*, vol. 32, pp. 88-92, Jan. 1985.

[10] D. Williamson, "Roundoff noise minimization and pole-zero sensitivity in fixed-point digital filters using residue feedback," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1210-1220, Oct. 1986.

[11] G. Li and M. Gevers, "Roundoff noise minimization using delta-operator realizations," *IEEE Trans. Signal Processing*, vol. 41, pp. 629-637, Feb. 1993.

[12] D. Williamson, "Delay replacement in direct form structures", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 453-460, Apr. 1988.

[13] M. M. Ekanayake and K. Premaratne, "Two-dimensional delta-operator formulated discrete-time systems: Analysis and synthesis of minimum roundoff noise realizations," *Proc. IEEE Int. Symp. Circuits Syst.* (ISCAS'96), vol. 2, pp. 213-216, May 1996.

[14] G. Li and Z. Zhao, "On the generalized DFIIt structure and its state-space realization in digital filter implementation," *IEEE Trans. Circuits Syst. I*, vol. 51, pp. 769-778, Apr. 2004.

[15] S. Y. Hwang, "Roundoff noise in state-space digital filtering: A general analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 256-262, June 1976.

[16] C. T. Mullis and R. A. Roberts, "Synthesis of minimum roundoff noise fixed-point digital filters," *IEEE Trans. Circuits Syst.*, vol. 23, pp. 551-562, Sept. 1976.

[17] S. Y. Hwang, "Minimum uncorrelated unit noise in state-space digital filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, pp. 273-281, Aug. 1977.

[18] L. B. Jackson, A. G. Lindgren and Y. Kim, "Optimal synthesis of second-order state-space structures for digital filters," *IEEE Trans. Circuits Syst.* , vol. 26, pp. 149-153, Mar. 1979.

[19] T. Hinamoto, H. Ohnishi and W.-S. Lu, "Roundoff noise minimization of state-space digital filters using separate and joint error feedback/coordinate transformation," *IEEE Trans. Circuits Syst. I*, vol. 50, pp. 23-33, Jan. 2003.

[20] W.-S. Lu and T. Hinamoto, "Jointly optimized error-feedback and realization for roundoff noise minimization in state-space digital filters," *IEEE Trans. Signal Processing*, vol. 53, pp. 2135-2145, June 2005.

[21] T. Hinamoto, A. Doi and W.-S. Lu, "Jointly optimized high-order error feedback and realization for roundoff noise minimization in state-space digital filters", *Proc. 54th 2011 IEEE Int. Midwest Symp. Circuits Syst.*, Seoul, Korea, CD-ROM, Aug. 2011.