# A NEW CLASS OF VERY LOW SENSITIVITY CASCADE-FORM DIGITAL-FILTERS BASED ON "PASSIVE" SECOND ORDER SINGLE-INPUT SINGLE-OUTPUT BUILDING BLOCKS*

P. P. Vaidyanathan
Department of Electrical Engineering
California Institute of Technology
Pasadena, CA 91125

S. K. Mitra
Dept. of Electrical & Computer Eng.
University of California
Santa Barbara, CA 93106

ABSTRACT

A new type of cascade form structure for digital filtering is proposed, with each building block being a second order section, that satisfies certain passivity properties. This passivity is essentially a "structure-induced" boundedness on the transfer function magnitude, and leads to low passband sensitivity. In addition, the cascade nature ensures low stopband sensitivity, as zeros on the unit circle continue to remain on the unit circle in spite of the quantization. The structure itself is independent of the pole locations and therefore meets a wide range of filtering applications.

## 1. INTRODUCTION

A number of contributions in the area of digital filtering have drawn attention to the importance of structures that require very few number of bits per multiplier coefficient [1], [2], [3]. In a recent work [4], a general z-domain theory and synthesis procedure are advanced, for low sensitivity digital filter implementations, the structures being based on interconnections of the "lossless digital two-pair." In [4], for a given "bounded real" transfer function G(z), a structure is obtained as an interconnection of losses digital two-pairs, in the form shown in Fig. 1. The losslessness property of the two-pairs is maintained in spite of the perturbations of multipliers internal to the two-pairs, and therefore, the boundedness of $G(z)$ is "structurally" forced. This is the key factor behind the low sensitivity properties of the structures in [4]. A number of well-known structures, such as the wave digital filters[3], cascaded lattice structures [5], and the coupled form circuit [6] turn out to be special cases of the structures presented in [4].

The purpose of this paper is to develop low-sensitivity structures that are not based on the lossless two-pair, but obtained as a cascade of single-input single-output second order sections. Each second order section is implemented in a form that makes it "structurally passive," so that each section has low passband sensitivity property. In addition, each section is implemented so that zeros of the transfer function that are on the unit circle continue to remain on the unit circle in spite of the multiplier quantization. As a consequence, the stopband behavior, under quantized conditions, is excellent as well.

In Section 2 we briefly review the relation between structural passivity and low passband sensitivity of a digital filter transfer function. In Sections 3 and 4 we develop a second order structure for transfer functions with zeros on the unit circle, that satisfies the "structural passivity" requirements in a particular manner. In Section 5 we present some simulation examples that demonstrate the low sensitivity property of filters realized as a cascade of these second order building blocks.

## 2. PRELIMINARIES

Let us consider a digital filter transfer function G(z), implemented with a structure that has the multiplier coefficients $m_1, m_2 \ldots m_M$. Assume G(z) to be stable and scaled such that $|G(e^{j\omega})|_{max} = 1$. Let the structure be such that if any of the parameters $m_i$ is perturbed about its nominal value, $|G(e^{j\omega})|$ continues to be bounded above by unity. Thus, at a frequency $\omega = \omega_0$ where $|G(e^{j\omega_0})| = 1$, perturbation of a coefficient $m_i$ can only decrease the magnitude of G. Thus, a plot of $|G(e^{j\omega_0})|$ against $m_i$ is as shown in Fig. 2. This shows that the first-order sensitivity of $|G(e^{j\omega})|$ with respect to each $m_i$ is zero at $\omega = \omega_0$ :

$$\left.\frac{\partial |G(e^{j\omega})|}{\partial m_i}\right|_{\omega = \omega_0} = 0 \tag{1}$$

Thus, "structural boundedness" or "passivity" is the key to low passband sensitivity. The structures developed in [4] which are based on two-input two-output building blocks satisfy the property of Eqn. (1) for each multiplier independently.

In this paper we develop second order single-input single-output structures such that the property of Eqn. (1) holds in a restricted manner, but is still sufficient to satisfy low sensitivity requirement. Specifically, we construct a function $f(m_1, \ldots m_M)$ of multipliers so that, if each $m_i$ is quantized such that $f(m_1 \ldots m_N)$ always decreases from its ideal value, then Eqn. (1) holds. Thus, with specific quantization schemes for multiplier parameters, the second

*Work supported in part by NSF ENG 82-18310 and in part by Caltech funds.



CH1945-5/84/0000-0111 $1.00 © 1984 IEEE

order sections satisfy low-sensitivity property.

The term "br (bounded real) function" is frequently used in this paper. A digital filter transfer function G(z) is called "bounded real" if a) G(z) is real for real z, b) G(z) is stable, and c) $|G(e^{j\omega})| \le 1$ for all $\omega$.

## 3. DERIVATION OF THE STRUCTURE

We wish to design a stable biquadratic transfer function of the following form:

$$G(z) = \frac{n_0+2n_1z^{-1}+n_0z^{-2}}{1+2d_1z^{-1}+d_2z^{-2}} \tag{2}$$

with a transmission zero on the unit circle of the z-plane at a frequency $\omega_1$ given by $\cos^{-1}(-n_1/n_0)$. The constant $n_0$ is such that $|G(e^{j\omega})|$ has a maximum value equal to unity at a frequency $\omega_2$. Clearly, G(z) is br and therefore the following function:

$$\frac{F(z)}{r} = \frac{1+G(z)}{1-G(z)} \tag{3}$$

is discrete-time positive real (pr) for any positive r. If we perform the bilinear transformation:

$$z^{-1} = \frac{1-s}{1+s} \tag{4}$$

and obtain the following function Z(s):

$$Z(s) = F(z)\Big|_{z=\frac{1+s}{1-s}}$$

then Z(s) is a conventional second order positive real function. In order to satisfy the conditions imposed on G(z), we should design the pr function Z(s) in such a way that it satisfies the following conditions:

$$Z(j\Omega_1) = r \quad \text{i.e., a purely real quantity} \tag{5}$$

$$Z(j\Omega_2) = j\,X(j\Omega_2) \quad \text{i.e., a purely imaginary quantity} \tag{6}$$

Here $\Omega_1 = \tan\frac{\omega_1}{2}$, and $\Omega_2 = \tan\frac{\omega_2}{2}$. Let us now consider the following continuous-time biquadratic function:

$$Z(s) = \frac{a_0s^2+a_1s+1}{b_0s^2+b_1s+1} \tag{7}$$

It is well known that Z(s) is pr if the coefficients $a_0, a_1, b_0$, and $b_1$ are positive and, in addition, if the following inequality is satisfied:

$$(a_0+b_0-a_1b_1)^2 \le 4\,a_0b_0 \tag{8}$$

Now setting $\text{Im}[Z(j\Omega_1)] = 0$ we get:

$$a_1(1-b_0\Omega_1^2)\Omega_1 - b_1(1-a_0\Omega_1^2)\Omega_1 = 0$$

which gives the nontrivial solution:

$$\Omega_1^2 = \frac{a_1-b_1}{a_1b_0-a_0b_1} \tag{9}$$

Next, setting $\text{Re}\,Z(j\Omega_2) = 0$ we get

$$a_0b_0\Omega_2^4 + \Omega_2^2[a_1b_1-a_0-b_0] + 1 = 0 \tag{10}$$

and in order to have a real solution $\Omega_2^2$, we should have:

$$(a_1b_1-a_0-b_0)^2 - 4a_0b_0 \ge 0 \tag{11}$$

which, along with the condition of Eqn. (8), implies

$$a_1b_1-a_0-b_0 = \pm 2\sqrt{a_0b_0} \tag{12}$$

Eqn. (10) then gives the solution,

$$\Omega_2^2 = -\frac{a_1b_1-a_0-b_0}{2a_0b_0} \tag{13}$$

which dictates the choice of sign in Eqn. (12), leading to

$$\Omega_2^2 = \frac{1}{\sqrt{a_0b_0}} \tag{14}$$

In summary, a pr function of the form (7), with coefficients satisfying the condition

$$a_1b_1-a_0-b_0 = -2\sqrt{a_0b_0}$$

is purely real at $s = j\Omega_1$, where $\Omega_1^2$ is given by Eqn. (9) and purely imaginary at $s = j\Omega_2$ where $\Omega_2^2$ is given by Eqn. (14).

If we now convert the pr function Z(s) into a discrete time br function G(z) by applying the following transformation

$$G(z) = \frac{Z(s)-r}{Z(s)+r}\Bigg|_{s=\frac{1-z^{-1}}{1+z^{-1}}} \tag{15}$$

then we have a discrete time br function of the following form:

$$G(z) = \frac{a_0-c_0+1-r+2z^{-1}[(1-r)-(a_0-c_0)]+z^{-2}[a_0-c_0+1-r]}{a_0+c_0+2a_1+1+r+2z^{-1}[1+r-a_0-c_0]+z^{-2}[a_0+c_0+1+r-2a_1]} \tag{16}$$

which has a transmission zero at the frequency $\omega_1$ and attains a peak magnitude of unity at the frequency $\omega_2$. In Eqn. (16),

$$c_0 \triangleq rb_0 > 0$$

Now consider the implementation of Eqn. (16). Under ideal (infinite precision) conditions, the multiplier coefficients $a_0$, $c_0$, $a_1$, and r are such that Eqn. (12) holds. If we now wish to quantize the multiplier coefficients, we can do so in such a way that Eqn. (8) (if not Eqn. (12)) continues to hold. This ensures that, at the frequency $\Omega_2$, the real part of $Z(j\Omega_2)$ becomes positive rather than negative. In other words, $|G(e^{j\omega_2})|$ decreases from unity, satisfying the zero-sensitivity conditions of Section 2. Finally, note that the quantization of $a_0$, $c_0$, $a_1$, and r does not affect their signs and so the other requirement for br property is also satisfied. It should be noticed that Eqn. (16) involves a constant term in the denominator that is different from unity. In an actual implementation, this can be taken care of by a multiplier of value $m = 1/(a_0+c_0+2a_1+1+r)$ and this multiplier typically requires greater precision than the other multiplier coefficients.

## 4. DERIVATION OF THE VALUES OF MULTIPLIER COEFFICIENTS

Given a transfer function of the form of Eqn. (2), let us first rewrite it in the form

$$G(z) = \frac{kn_0 + 2kn_1z^{-1} + kn_0z^{-2}}{k + 2kd_1z^{-1} + kd_2z^{-2}} \qquad (17)$$

where k should be determined so that the decomposition of multiplier coefficients as shown in Eqn. (16) can be accomplished. Now Eqn. (16) can be written as

$$G(z) = \frac{\alpha + 2\gamma z^{-1} + \alpha z^{-2}}{\beta + 2a_1 + 2\delta z^{-1} + (\beta-2a_1)z^{-2}} \qquad (18)$$

where

$$\alpha = a_0-c_0+1-r \ , \ \beta = a_0+c_0+1+r \ ,$$

$$\gamma = 1-r-(a_0-c_0) \ , \ \delta = 1+r-a_0-c_0 \ .$$

Note that $\alpha+\beta+\gamma+\delta = 4$. From Eqns. (17) and (18) we have

$$\alpha = kn_0 \ , \ \beta = \frac{k(1+d_2)}{2} \ , \ \gamma = kn_1 \ , \ \delta = kd_1 \qquad (19)$$

Thus, the multiplier coefficients to be implemented are given by

$$a_0=1-\frac{\gamma+\delta}{2}, \ c_0=1-\frac{\alpha+\delta}{2}, \ r=1-\frac{\alpha+\gamma}{2}, \ a_1=\frac{k(1-d_2)}{4} \qquad (20)$$

Given any br transfer function, we can thus find a set of positive numbers $a_0$, $c_0$, $a_1$, r.

### Sensitivity Properties of the Resulting Structure and Relation to Other Structures

As mentioned earlier, an implementation in which $a_0$, $c_0$, $a_1$, and r are the multiplier coefficients has low sensitivity with respect to these multipliers. This conclusion does not assume any particular pole locations in the z-plane. In addition, from Eqn. (16) we can verify the following results:

$$\left.\frac{\partial G}{\partial a_0}\right|_{z=1} = 0, \ \left.\frac{\partial G}{\partial a_1}\right|_{z=\pm 1} = 0, \ \left.\frac{\partial G}{\partial c_0}\right|_{z=1} = 0, \ \left.\frac{\partial G}{\partial r}\right|_{z=-1} = 0 \qquad (21a,b,c,d)$$

Thus, we have zero sensitivity with respect to the multipliers $a_0$, $a_1$, and $c_0$, at $\omega = 0$, and this is significant for poles close to the point $z = 1$. Similarly for poles close to $z = -1$, Eqns. (21b) and (21d) are significant.

A number of methods are well-known for decomposing the denominator coefficients of a digital biquad, in order to achieve low sensitivity [1], [2]. In [2] are presented some of these techniques, which are extensions of the technique due to Agarwal and Burrus [1]. Basically the denominator is written in one of the following two forms:

$$D(z) = 1 - (2+\alpha_1)z^{-1} + (1+\alpha_1-\alpha_2)z^{-2} \qquad (22)$$

or

$$D(z) = 1 - (2+\alpha_1+\alpha_2)z^{-1} + (1+\alpha_1)z^{-2} \qquad (23)$$

The all-pole transfer function $H(z) = 1/D(z)$ has the magnitude $1/|\alpha_2|$ in Eqns. (22) and (23). In either case, it is independent of $\alpha_1$. Thus, the structure is such that $H(z)$ is bounded by the constant $1/|\alpha_2|$ regardless of $\alpha_1$. This is a simple special case of structural boundedness in a restricted sense (i.e., around $z = 1$). This leads to zero sensitivity with respect to $\alpha_1$ at the point $\omega = 0$, as pointed out in [2].

The structure represented in Eqn. (16) is a more general decomposition technique in which the numerator and deonminator are coupled. The decomposition itself is independent of the pole locations.

## 5. SIMULATION EXAMPLES

A tenth order bandpass elliptic filter with passband attenuation $\leq 0{\cdot}2$dB and passband edges at $0{\cdot}1\pi$ and $0{\cdot}2\pi$ was implemented, with each second-order section being of the form of Eqn. (16). The resulting structure (the br-cascade) was simulated with finite precision for the multiplier coefficients and compared to similar simulation of the conventional cascade form (direct form-cascade). Figure 3 shows the passbands of the ideal and actual implementations for both the conventional cascade form and the new structure. The improved performance in the passband, of the br cascade, is clearly in evidence. Figure 4 shows the entire frequency response, which clearly demonstrates that the stopband of the br-cascade structures is at least as good as that of the conventional cascade-form.

## 6. CONCLUDING REMARKS

The implementation of Eqn. (16) requires 5 multipliers, namely $a_0$, $c_0$, $a_1$, r and the quantity $1/(a_0+c_0+2a_1+1+r)$, whereas a scaled direct form second order section requires only 4 multipliers (if zeros are on the unit circle). However, the lower sensitivity of the br-cascade compared to the direct-form cascade is significant enough to make it more efficient in an overall sense. For example, consider an 8-bit microcomputer implementation where two precisions are possible, viz., 8 bits or 16 bits. Assuming that the direct form cascade requires 11 bits whereas the br-cascade requires only 8 bits per multiplier for a given performance level, it is clearly more efficient to implement five single precision multipliers as against four double precision multipliers.

The roundoff noise properties of the new structure remain to be studied. It will also be of interest to investigate whether the passivity property of the second order sections helps to suppress limit cycles.

References

[1] R.C. Agarwal and C.S. Burrus, "New Recursive Digital Filter Structures Having Very Low Sensitivity and Roundoff Noise," IEEE Trans. on Circuits and Systems, pp. 921-926, Dec. 1975.
[2] S. Nishimura, K. Hirano, and R.N. Pal, "A New Class of Very Low Sensitivity and Low Roundoff Noise Recursive Digital Filter Structures," IEEE Trans. on Circuits and Systems, pp. 1152-1158, Dec. 1981.
[3] A. Fettweis, "Digital Filter Structures Related to Classical Filter Networks," in Digital Signal Processing II, Edited by Digital Signal Processing Committee, IEEE ASSP Society, pp. 475-485, IEEE Press, 1975.
[4] P.P. Vaidyanathan and S.K. Mitra, "A General Theory and Synthesis Procedure for Low Sensitivity Digital Filter Structures," Tech. Report #82-09, Department of Electrical and Computer Engineering, Univ. of Calif., Santa Barbara, 1982.
[5] A. H. Gray and J. D. Markel, "A Normalized Digital Filter Structure," IEEE Trans. on Acoustics, Speech, and Signal Processing, pp.268-277, June 1975.
[6] A.V. Oppenheim and R.W. Schafer, "Digital Signal Processing," Prentice Hall, Inc., 1975.

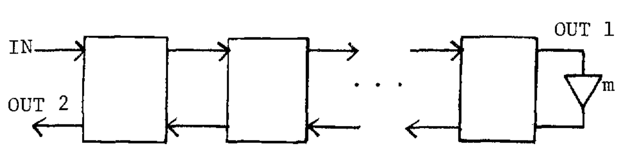


Fig. 1. The lbr-cascade realization

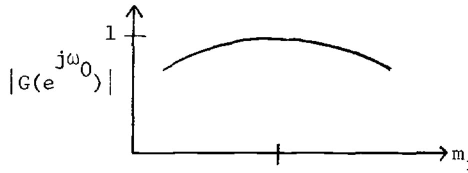


Fig. 2. The zero-sensitivity property

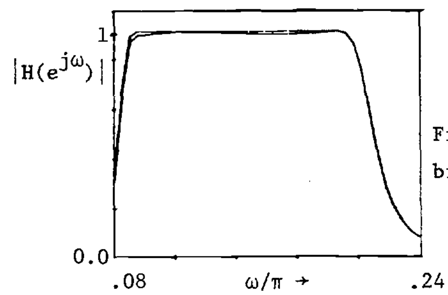


Fig. 3(a) br-cascade

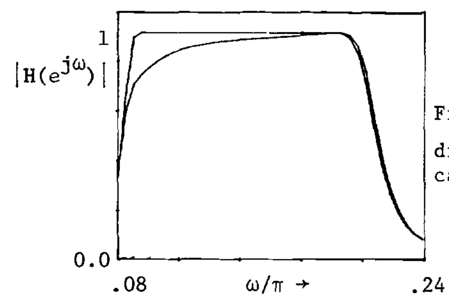


Fig. 3(b) direct-form cascade

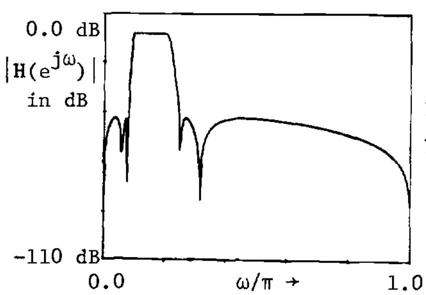


Fig. 4(a) br-cascade

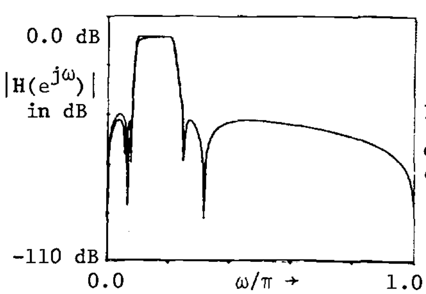


Fig. 4(b) direct-form cascade