A Framework for Analysis of High Order Sigma-Delta Modulators

Philip Steiner and Woodward Yang

Division of Applied Sciences

Harvard University

Cambridge, MA

Abstract

In this paper a framework for the analysis of a family of high order interpolative sigma-delta modulators is introduced. It is shown that a large number of high order architectures can be reduced to a *diagonal* form which facilitates the stability analysis of the system. In addition, the diagonal form is a canonical form which illustrates the equivalence of a variety of sigma-delta architectures. Architectural differences are manifested as differences in parameter values in the diagonal form, providing a convenient method of comparison between systems. It is also shown how transformation to the diagonal form may result in a reduction in order of a system. Finally, there is a brief discussion on the importance of diagonal systems as opposed to non-diagonal systems.

Introduction

Sigma-Delta modulation systems are clocked, nonlinear systems which sample an input at signal rates much higher than the Nyquist frequency and feed back a low resolution approximation of this input. This allows for accurate conversion of the signal without relying on high precision components. The basic first order sigma-delta modulator is shown in figure (1).

The nonlinearity of the system is wholly contained within the representation of the quantizer element. There are two main analysis approaches which are used in studying modulators. In the first approach, the quantizer is represented by a thresholding function, in which case a detailed nonlinear analysis is necessary. The advantage of this approach is that no linearizing approximations are made and thus results more accurately reflect the true behavior of the system. Much research using this approach has focussed on the first and second order modulators, including that of Hein and Zakhor [1],[2], Pinault [3], Feely [4], Schreier et. al. [5],[6], Wang [7], and the authors [8].

The second approach is to model the quantizer as a source of additive noise in the system. This linearizes the model of the system and allows for the use of linear systems analysis. The linear method is used by most designers and the results are checked by extensive simulation. The advantage of this method over the nonlinear approach is that the linear analysis provides an estimate of both stability and performance while the nonlinear analysis has typically focussed only on stability. The linear approach has generated a large number of results concerning the shaping of the quantization noise, as well as the stability and performance of systems of all orders. A sample of some of these can be found in [9], [10],[11],[12].

As performance demands on modulators have rapidly increased, so have the implementation and use of high order interpolative modulators. The advantage of high (greater than 2) order modulators is their greater performance at equivalent oversampling ratios. The main problem with them is that they are very frequently unstable. One method of stabilizing high order systems was proposed in [12] and involves the addition of linear feedback elements. However, extensive computer simulation is still required in order to verify stability.

The nonlinear analysis can accurately predict system behavior for any order, yet while several nonlinear analyses of second order systems have been presented, higher order systems have been analytically intractable.

In this paper we propose a way of simplifying the nonlinear analysis of a family of higher order interpolative modulators. This method involves the transformation of state space into a co-ordinate system where the state variables interact *only* through the quantizer function.

Review of Linear Systems

A thorough treatment of linear systems is found in [14]. This section is merely intended to introduce notation and to provide a summary of some important results.

Consider a general, first order, discrete time linear system with input $\vec{u}(n)$, state variables $\vec{x}(n)$, and output $\vec{y}(n)$. The dynamical equations describing the time evolution of the system can be written as

$$\vec{x}(n+1) = \mathbf{A}\vec{x}(n) + \mathbf{B}\vec{u}(n) \tag{1}$$

$$\vec{y}(n) = \mathbf{C}\vec{x}(n) + \mathbf{D}\vec{u}(n) \tag{2}$$

where $\mathbf{A}$ is the *transition matrix* of the system, and the matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are all assumed to be constant (i.e. a time-invariant system). One method of solution is to iteratively solve equations (1) and (2), giving

$$\vec{x}(n) = \mathbf{A}^n\vec{x}(0) + \sum_{k=o}^{n-1} \mathbf{A}^{n-k-1}\mathbf{B}\vec{u}(k) \tag{3}$$

$$\vec{y}(n) = \mathbf{C}\mathbf{A}^n\vec{x}(0) + (\sum_{k=0}^{n-1} \mathbf{C}\mathbf{A}^{n-k-1}\mathbf{B}\vec{u}(k)) + \mathbf{D}\vec{u}(n) \tag{4}$$

In many cases the transition matrix, $\mathbf{A}$, is similar to a diagonal matrix. If this is true then there exists a matrix $\mathbf{P}$ such that [1]

$$\mathbf{P} = \begin{pmatrix} \vdots & \vdots & & \vdots \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_N \\ \vdots & \vdots & & \vdots \end{pmatrix} \tag{5}$$

where $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_N$ are the eigenvectors of $\mathbf{A}$. A transformation of coordinates

$$\vec{x}' = \mathbf{P^{-1}}\vec{x} \tag{6}$$

[1]the matrix $\mathbf{P}$ is not unique in transforming $\mathbf{A}$ to a diagonal form. Any matrix which is a permutation of the columns of $\mathbf{P}$ multiplied by a scalar, will also result in a diagonal form. The permutation of the column vectors rearranges the order of the eigenvalues along the diagonal while scalar multiplication scales all of the eigenvalues.

results in equation (1) becoming

$$\begin{aligned}
\vec{x}\prime(n+1) &= (\mathbf{P^{-1}AP})\vec{x}\prime(n) + (\mathbf{P^{-1}B})\vec{u}(n) \\
&= \mathcal{A}\vec{x}\prime(n) + \mathcal{B}\vec{u}(n) \qquad (7)
\end{aligned}$$

where $\mathcal{A}$ is a diagonal matrix with elements $\lambda_1, \lambda_2, \ldots, \lambda_N$, the eigenvalues of $\mathbf{A}$, along the diagonal. The state variables are now decoupled and iteratively solving for equation (7) results in

$$x_i(n) = \lambda_i^n x_i(0) + \sum_{k=0}^{n-1} \lambda_i^{n-k-1} \mathcal{B}_{ij} u_j(k) \qquad (8)$$

In the case where $\mathbf{A}$ is not similar to a diagonal matrix, the transformation described by equations (5) and (6) can still be applied to equation (1). In this case it is the *generalized* eigenvectors, $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_N$, which are used in the transformation matrix $\mathbf{P}$. After the transformation, the matrix $\mathcal{A}$ is in a non-diagonal, Jordan canonical form. In this case the state variables are only partially decoupled and equation (8) no longer applies.

Sigma Delta Modulators

Dynamical Equations

figure (2) shows an example of a N'th order, interpolative modulator. The outputs of the N integrators are selected as the N independent state variables. While this is not the only state space representation possible, it is well suited to stability analysis, as stability will depend on the bounds of the integrator outputs. With this choice of states, the dynamical equations of the system in figure (2) become

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{pmatrix}_{n+1} = \begin{bmatrix} 1 & 0 & & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & & \cdots & 1 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{pmatrix}_n + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \\ b_N \end{pmatrix} u(n)$$

$$
-\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \\ c_N \end{pmatrix} \Theta[d_k x_k(n)] \tag{9}
$$

and the output is given by

$$
y(n) = \Theta[d_k x_k(n)] \tag{10}
$$

where $d_k x_k(n)$ is an implicit summation over $k = 1 \dots N$, and the quantizing function Θ is defined as

$$
\Theta[z] = \left\{ \begin{array}{ll} +1 & \text{if } z \geq 0 \\ -1 & \text{if } z < 0 \end{array} \right\} \tag{11}
$$

Adopting the notation from the previous section with $\mathbf{B} \rightarrow \vec{b}$, $\vec{u} \rightarrow u$, equation (9) is written more compactly as

$$
\vec{x}(n+1) = \mathbf{A}\vec{x}(n) + \vec{b}u(n) - \vec{c}\,\Theta[\vec{d} \cdot \vec{x}(n)] \tag{12}
$$

Any interpolative modulator of any order can be described by equations (12) and (10). The transition matrix, $\mathbf{A}$, contains all the information concerning the interconnectivity of states, including linear feedback and feed-forward paths. The vector $\vec{b}$ describes the paths for insertion of input into the system, and $\vec{c}$ describes the feedback paths from the quantizer. Therefore, all interpolative system architectures can be represented by equation (12).

Equation (12) is a nonlinear equation, yet the entire nonlinearity is contained within the quantizer function. The first two terms are exactly the same as in equation (1), describing a linear system.

Pseudo-Linear Systems Analysis

It has already been noted that aside from a nonlinear term describing the quantizer, the dynamical equations for a sigma-delta modulator are linear. figure

(3) shows a block diagram of a N'th order system, which is divided into the linear part, $\mathcal{L}$, and the nonlinear quantizer, described in equation (11). The quantizer output can take on only one of two values, $\{-1, 1\}$. The quantizer function thus serves to divide state space into two regions: $\Psi_1 \in \sum d_j x_j \geq 0$ and $\Psi_2 \in \sum d_j x_j < 0$. These two regions are divided by the hyperplane, $\sum d_j x_j = 0$. This is seen in figure (4). In Ψ_1 the dynamical equations (12) and (10) become

$$\vec{x}(n+1) = \mathbf{A}\vec{x}(n) + \vec{b}u(n) - \vec{c} \tag{13}$$

$$y(n) = +1 \tag{14}$$

while in Ψ_2 the dynamical equations are

$$\vec{x}(n+1) = \mathbf{A}\vec{x}(n) + \vec{b}u(n) + \vec{c} \tag{15}$$

$$y(n) = -1 \tag{16}$$

Equations (13) and (15) are linear equations (with the exception of the additive constant $\pm\vec{c}$). The modulator thus behaves linearly while it remains in Ψ_1, obeying equation (13), until a transition is made into Ψ_2 across the $\sum d_j x_j$ hyperplane. The modulator then resumes its linear behavior only it now follows equation (15), until another transition is made back into Ψ_1 and the cycle begins again. Therefore, the behavior of the modulator can be characterized by considering each transition as an initial condition for either equation (13) or equation (15). This is illustrated in figure (5). The system starts in some state $\vec{x}_+(0)$ (arbitrarily chosen to be in Ψ_1). The modulator obeys equation (13) while in Ψ_1 which from equation (3) gives

$$\vec{x}(k) = \mathbf{A}^k \vec{x}_+(0) + \sum_{i=0}^{k-1} \mathbf{A}^{k-i-1}(\vec{b}u(k) - \vec{c}) \tag{17}$$

This holds true only while $\vec{x}(k) \in \Psi_1$. At some time, $k = k^*$, the system will make a transition across the hyperplane into Ψ_2 i.e, $\vec{x}(k^*) \in \Psi_2$ and $\vec{x}(k^*-1) \in \Psi_1$. $\vec{x}(k^*)$ can now be relabelled as $\vec{x}_-(0)$ and the above procedure is repeated with the exception that equation (15) is followed in place of equation (13) so that

$$\vec{x}(k) = \mathbf{A}^k \vec{x}_-(0) + \sum_{i=0}^{k-1} \mathbf{A}^{k-i-1}(\vec{b}u(k) + \vec{c}) \tag{18}$$

Therefore equations (13), (15), and the two sequences $\{\vec{x}_+(0), \vec{x}_+(1), \ldots\}$ and $\{\vec{x}_-(0), \vec{x}_-(1), \ldots\}$ of transition points completely characterize the dynamics of the modulator. Stability of the modulator, for example, can be determined from the stability of the linear systems (13) and (15), and from the convergence and boundedness of the sequences of transition points.

There has been a lot of work based on this type of pseudo-linear analysis (with minor variations) for second order systems ([1],[2],[3],[4],[5],[8]). The results achieved for the second order system have yet to be extended to higher order modulators.

The pseudo-linear method described above was applied to a one-bit quantizer, as described by equation (11), but can readily be extended to multi-bit quantizers. The extra bits of the quantizer cause state space to be divided into more regions with a separate set of linear equations for each region. For example, a 2-bit quantizer (with 4 quantization levels) would give rise to four regions, $\Psi_1, \Psi_2, \Psi_3, \Psi_4$, each with its own distinct set of linear equations. In addition, instead of a single dividing hyperplane a 2-bit quantizer would create three parallel dividing hyperplanes. The pseudo-linear analysis would then require four linear equations and four sequences of transition points, but the procedure would remain essentially the same.

Non-idealities in modulators, such as finite op-amp gain, slew rate, and noise, fall into two categories; linear and nonlinear. Linear non-idealities can be dealt with by the addition of suitable terms into equations (13) and (15). As these terms are linear, no alteration in the method of analysis is required. Gross, nonlinear non-idealities can affect the method of analysis, but small nonlinearities can be addressed using a perturbation analysis.

Canonical Representation and Diagonal Modulators

The pseudo-linear analysis described in the last section proposes one method of dealing with the nonlinearity of the quantizer function, without making any assumptions about the shape of the quantization noise. This method is frequently applied to stability analysis of a given modulator ([1],[2],[3]) and usually involves the calculation of a potential function of the form

$$f(\vec{x}) = constant \tag{19}$$

or of a state space trajectory of the form

$$x_N = g(x_1, x_2, \ldots, x_{N-1}) \tag{20}$$

In the case of a second order order modulator with a 2-dimensional state space it is usually possible to analytically describe a potential function or trajectory. As the order of the modulator is increased, however, an analytical solution for the potential function becomes increasingly intractable.

There exists a class of high order systems which allow for simplification of the above-mentioned complexity problem. The modulator shown in figure (1) and described by equation (9) has a transition matrix, $\mathbf{A}$, which is in Jordan canonical form, i.e. there are no linear transformations which will simplify the form of $\mathbf{A}$. This means that no simplification, with respect to the potential function, is readily apparent. Now consider the modulator in figure (5). This modulator includes generalized linear feedback paths between the integrators (feed-forward paths have been left out but can be included for further generalization). The dynamical state space equation for this modulator is

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{pmatrix}_{n+1} = \begin{bmatrix} 1 & z_{12} & & \cdots & z_{1N} \\ 1 & 1 & z_{23} & \cdots & z_{2N} \\ 0 & 1 & 1 & \cdots & z_{3N} \\ \vdots & & & \ddots & \vdots \\ 0 & & \cdots & 1 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \\ x_N \end{pmatrix}_{n} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ \\ b_N \end{pmatrix} u(n)$$
$$- \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ \\ c_N \end{pmatrix} \Theta[d_k x_k(n)] \tag{21}$$

Inclusion of the linear feedback paths puts the transition matrix, $\mathbf{A}$, into a non-canonical form. Thus, $\mathbf{A}$ can be isospectrally rotated to its equivalent Jordan canonical form. In some cases the Jordan canonical form of $\mathbf{A}$ is a diagonal matrix. These *diagonal* modulators , i.e. modulators whose transition matrix is similar to a diagonal matrix, make up an important class of sigma delta systems because they provide the greatest simplification of the potential function, and because most implementations fall into this class. The remainder of this paper deals with diagonal modulators, although some of the results may apply to non-diagonal, canonical systems. A brief discussion of the non-diagonal case if found in the last section.

The transformation described by equations (5),(6), and (7) can now be applied to the system (21) resulting in

$$\begin{aligned} \vec{x}\prime(n+1) &= (\mathbf{P^{-1}AP})\vec{x}\prime(n) + (\mathbf{P^{-1}}\vec{b})u(n) - (\mathbf{P^{-1}}\vec{c})\Theta[\vec{d}\mathbf{PP^{-1}}\vec{x}(n)] \\ &= \mathcal{A}\vec{x}\prime(n) + \vec{b}\prime u(n) - \vec{c}\prime\Theta[\vec{d}\prime \cdot \vec{x}\prime(n)] \end{aligned} \tag{22}$$

where the matrix $\mathcal{A}$ is a diagonal matrix with the eigenvalues of $\mathbf{A}$ as its elements.

figure (6) shows a diagram of this diagonal modulator. The diagonal form of $\mathcal{A}$ has two important implications.

First, all the state variables in this new coordinate system are decoupled and interact only through the quantizer function. The dynamical equation for any state variable $x_{k'}$ is given by

$$x_{k'}(n+1) = \lambda_k x_{k'}(n) + b_{k'}u(n) - c_{k'}\Theta[\vec{d'} \cdot \vec{x'}(n)] \tag{23}$$

in which the other state variables, $x_{j'} j \neq k$, appear only in the quantizer function. Since the pseudo-linear analysis transforms this quantizer function into a division of state space, the interaction between states is embedded in the hyperplane which divides Ψ_1 from Ψ_2. This greatly simplifies the stability analysis of the system, reducing the complexity of the potential function.

Second, a large number of matrices of the form of $\mathbf{A}$ in equation (21) are similar to diagonal matrices. Therefore, a large number of system architectures with different feed-forward and feedback paths can be represented in the diagonal form, with different architectures in the interpolative form reflected as differences in parameter values in the diagonal form. In addition, as it has been found that linear feedback paths are usually required to stabilize higher order modulators ([12]), a large number of implementations fall within the diagonal classification.

The simplification that the diagonal representation brings to the analysis of higher order systems can be accompanied by several complications. Recall that $\mathbf{A}$ is a matrix whose elements are all real. Thus, the characteristic equation of $\mathbf{A}$ is a N'th order polynomial with real coefficients.

$$\begin{aligned} det(\mathbf{A} - \lambda\mathbf{I}) &= 0 \\ &= \sum_{i=0}^{N} a_i \lambda^i \quad a_i \in \Re \end{aligned} \tag{24}$$

The roots of this equation are either real or pairs of complex conjugates. The real roots result in equations, in the canonical representation, of the form

$$x_{k'}(n+1) = \lambda_k x_{k'}(n) + b_{k'}u(n) - c_{k'}\Theta[\vec{d'} \cdot \vec{x'}(n)] \tag{25}$$

where $\lambda_k, b_{k'}, c_{k'}$, and $x_{k'}(0)$ are all real so that $x_{k'}$ only takes on real values. The complex conjugate roots result in paired equations. If λ_j is a complex root and λ_{j+1} is its conjugate then

$$\begin{aligned} x_{j'}(n+1) &= \lambda_j x_{j'}(n) + b_{j'}u(n) - c_{j'}\Theta[\vec{d'} \cdot \vec{x'}(n)] \\ x_{j+1'}(n+1) &= \lambda_{j+1} x_{j+1'}(n) + b_{j+1'}u(n) - c_{j+1'}\Theta[\vec{d'} \cdot \vec{x'}(n)] \end{aligned} \tag{26}$$

where $\lambda_{j+1} = \lambda_j^*, b_{j+1'} = (b_{j'})^*, c_{j+1} = (c_{j'})^*, x_{j+1}(0) = (x_{j'}(0))^*$ and thus $x_{j+1}(m) = (x_{j'}(m))^*$ for all m.

Therefore in exchange for decoupling of the state variables, some of the states in the canonical representation may take on complex values. The added complexity of the latter, however, is far outweighed by the simplification provided by the former. It is also important to note that these coordinate transformations affect only the linear part of the modulator. In figure (2), it is only the internal representation of the linear block $\mathcal{L}$ which is changed. The input, $u(n)$, as well as the input to the quantizer and the output from the quantizer remain unaffected. Therefore even with complex state variables, the argument of Θ is always real and the pseudo-linear analysis can still be applied.

Reduction of Order

Consider the diagonal modulator represented by equation (22) and shown in figure (6). It has already been indicated that the only interaction between the state variables takes place in the quantizer function. It follows that if $d_{k'} = 0$ for some state $x_{k'}$, this state variable does not interact in any way with the other state variables and, by equations (10) and (11), does not contribute in any way to the determination of the output. $x_{k'}$ is not observable and does not affect the state space trajectory of the other state variables in any way. This variable can be ignored, and in fact entirely removed from the dynamical equations without changing the output produced for a given input. This is illustrated in figure (7).

By the above argument, all state variables $x_{k'}$ for which $d_{k'} = 0$ can be dropped from equation (22), reducing the order of system. Assume that $d_{k'} = 0$ for some number, m, of states. These variables are removed and the remaining state variables and parameters are reindexed from 1 to $N - m$. Equation (22) becomes

$$\vec{x}\prime(n+1) = \mathcal{A}\vec{x}\prime(n) + \vec{b}\prime u(n) - \vec{c}\prime\Theta[\vec{d}\prime \cdot \vec{x}\prime(n)] \tag{27}$$

where all the vectors $\vec{x}\prime, \vec{b}\prime, \vec{c}\prime, \vec{d}\prime$ are now $(N-m) \times (1)$. $\mathcal{A}$ is a square matrix of dimension $N-m$

$$\mathcal{A} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{N-m} \end{pmatrix} \tag{28}$$

and it is assumed that the state variables are indexed such that the λ_i are arranged in increasing order along the diagonal.

There is another case in which reduction of order of the system may occur: that of repeated eigenvalues. Assume that λ_j is a repeated eigenvalue of multiplicity l. Since the eigenvalues are arranged in increasing order along the diagonal, eigenvalues of the same value will be adjacent, i.e. $\lambda_j = \lambda_{j+1} = \ldots = \lambda_{j+l}$ $(j+l \leq N-m)$. Now a transformation matrix, $\mathbf{G}$ can be formed, such that $\mathbf{G}$ is block diagonal and given by

$$\mathbf{G} = \begin{pmatrix} \mathbf{I_j} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D_l} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I_{N-m-j-l}} \end{pmatrix} \tag{29}$$

where the $\mathbf{I_i}$ are identity matrices of dimension i and $\mathbf{D_l}$

$$\mathbf{D_l} = \begin{pmatrix} d_j & d_{j+1} & d_{j+2} & \ldots & d_{j+l} \\ d_j & -d_{j+1} & d_{j+2} & \ldots & d_{j+l} \\ d_j & d_{j+1} & -d_{j+2} & \ldots & d_{j+l} \\ \vdots & & & \ddots & \vdots \\ d_j & d_{j+1} & d_{j+2} & \ldots & -d_{j+l} \end{pmatrix} \tag{30}$$

is a nonsingular matrix of rank l. Thus $\mathbf{G}$ is invertible and can be used as a transformation matrix

$$\vec{x}\prime\prime = \mathbf{G}\vec{x}\prime \tag{31}$$

where $\vec{x}\prime$ is the state variable from equation (27). Applying this transformation to the canonical system in equation (27) yields

$$
\begin{aligned}
\mathbf{G}\vec{x}\prime(n+1) &= (\mathbf{G}\mathcal{A}\mathbf{G^{-1}})\mathbf{G}\vec{x}\prime(n) + \mathbf{G}\vec{b}\prime u(n) - \mathbf{G}\vec{c}\prime\,\Theta[\vec{d}\prime\mathbf{G^{-1}}\mathbf{G}\vec{x}\prime(n)] \\
\vec{x}\prime\prime(n+1) &= \mathcal{A}\vec{x}\prime\prime(n) + \vec{b}\prime\prime u(n) - \vec{c}\prime\prime\,\Theta[\vec{d}\prime\prime \cdot \vec{x}\prime\prime(n)] \qquad (32)
\end{aligned}
$$

Notice that the transformation leaves $\mathcal{A}$ unchanged. In addition $x_k\prime\prime = x_k\prime$ for $k < j$ and $k > j + l$, so that the vectors $\vec{b}\prime, \vec{c}\prime$ and $\vec{d}\prime$ are changed only in the $j \ldots j + l$ elements. Of specific interest is $\vec{d}\prime\prime$ which is given by

$$
\vec{d}\prime\prime = \left[\begin{array}{cccccccccccc} d_1\prime & \ldots & d_{j-1}\prime & 1 & 0 & 0 & \ldots & 0 & d_{j+l+1} & \ldots & d_{N-m}\prime \end{array} \right] \qquad (33)
$$

Since $\mathcal{A}$ remains unchanged the system is still in canonical form and the state variables are still decoupled. By the argument presented at the beginning of this section, any state variable, $x_k\prime$, for which $d_k\prime = 0$ can be dropped from equation (32). Equation (33) shows that after the transformation $d_{j+1}\prime\prime = d_{j+2}\prime\prime = \ldots = d_{j+l}\prime\prime = 0$ and $d_j\prime\prime = 1$. Therefore the variables $x_{j+1}\prime\prime, \ldots, x_{j+1}\prime\prime$ can be dropped from equation (32), and the order of the system has been reduced by $l - 1$.

For each repeated eigenvalue of multiplicity l, the order of the system is reduced by $l - 1$. Therefore an interpolative system of order N can be transformed into a diagonal system of order $N - m - \sum_i (l_i - 1)$, where the sum is over all distinct eigenvalues of $\mathbf{A}$. Clearly, the lower order system retains the same input/output relationship as the original system, and thus precisely the same performance.

An Example of a Diagonal Modulator

Consider the modulator in figure (8), taken from [13]. This is a fifth order interpolative modulator with state space equations

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}_{(n+1)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & \alpha_1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & \alpha_2 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}_{(n)} + \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} u(n)$$

$$- \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \Theta[\begin{pmatrix} d_1 & d_2 & d_3 & d_4 & d_5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}] \tag{34}$$

The transition matrix in equation (34) is similar to a diagonal matrix and has eigenvalues $1, 1 \pm \sqrt{\alpha_1}, 1 \pm \sqrt{\alpha_2}$. The transformation matrix, $\mathbf{P}$ of equation (5) is given by

$$\mathbf{P} = \begin{pmatrix} \alpha_1\alpha_2 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{\alpha_1}(\alpha_1 - \alpha_2) & -\sqrt{\alpha_1}(\alpha_1 - \alpha_2) & 0 & 0 \\ -\alpha_2 & \alpha_1 - \alpha_2 & \alpha_1 - \alpha_2 & 0 & 0 \\ 0 & \sqrt{\alpha_1} & -\sqrt{\alpha_1} & \sqrt{\alpha_2} & -\sqrt{\alpha_2} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{35}$$

which has inverse

$$\mathbf{P^{-1}} = \begin{pmatrix} \frac{1}{\alpha_1\alpha_2} & 0 & 0 & 0 & 0 \\ \frac{1}{2\alpha_1(\alpha_1-\alpha_2)} & \frac{1}{2\sqrt{\alpha_1}(\alpha_1-\alpha_2)} & \frac{1}{2(\alpha_1-\alpha_2)} & 0 & 0 \\ \frac{1}{2\alpha_1(\alpha_1-\alpha_2)} & \frac{-1}{2\sqrt{\alpha_1}(\alpha_1-\alpha_2)} & \frac{1}{2(\alpha_1-\alpha_2)} & 0 & 0 \\ \frac{-1}{2\alpha_2(\alpha_1-\alpha_2)} & \frac{-1}{2\sqrt{\alpha_2}(\alpha_1-\alpha_2)} & \frac{-1}{2(\alpha_1-\alpha_2)} & \frac{1}{2\sqrt{\alpha_2}} & \frac{1}{2} \\ \frac{-1}{2\alpha_2(\alpha_1-\alpha_2)} & \frac{1}{2\sqrt{\alpha_2}(\alpha_1-\alpha_2)} & \frac{-1}{2(\alpha_1-\alpha_2)} & \frac{-1}{2\sqrt{\alpha_2}} & \frac{1}{2} \end{pmatrix} \tag{36}$$

Applying the transformation of equation (6) to equation (34) yields

$$\begin{pmatrix} x_1\prime \\ x_2\prime \\ x_3\prime \\ x_4\prime \\ x_5\prime \end{pmatrix}_{(n+1)} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1+\sqrt{\alpha_1} & 0 & 0 & 0 \\ 0 & 0 & 1-\sqrt{\alpha_1} & 0 & 0 \\ 0 & 0 & 0 & 1+\sqrt{\alpha_2} & 0 \\ 0 & 0 & 0 & 0 & 1-\sqrt{\alpha_2} \end{pmatrix} \begin{pmatrix} x_1\prime \\ x_2\prime \\ x_3\prime \\ x_4\prime \\ x_5\prime \end{pmatrix}_n$$

$$+ \begin{pmatrix} b_1\prime \\ b_2\prime \\ b_3\prime \\ b_4\prime \\ b_5\prime \end{pmatrix} u(n) - \begin{pmatrix} c_1\prime \\ c_2\prime \\ c_3\prime \\ c_4\prime \\ c_5\prime \end{pmatrix} \Theta[\begin{pmatrix} d_1\prime & d_2\prime & d_3\prime & d_4\prime & d_5\prime \end{pmatrix} \begin{pmatrix} x_1\prime \\ x_2\prime \\ x_3\prime \\ x_4\prime \\ x_5\prime \end{pmatrix}]$$

where

$$\vec{b\prime} = \vec{c\prime} = \begin{pmatrix} \frac{1}{\alpha_1\alpha_2} \\ \frac{1}{2\alpha_1(\alpha_1-\alpha_2)} \\ \frac{1}{2\alpha_1(\alpha_1-\alpha_2)} \\ \frac{-1}{2\alpha_2(\alpha_1-\alpha_2)} \\ \frac{1}{2\alpha_2(\alpha_1-\alpha_2)} \end{pmatrix}$$

$$\vec{d\prime} = \begin{pmatrix} \alpha_1\alpha_2 d_1 - \alpha_2 d_3 + d_5 \\ \sqrt{\alpha_1}(\alpha_1-\alpha_2)d_2 + (\alpha_1-\alpha_2)d_3 + \sqrt{\alpha_1}d_4 + d_5 \\ \sqrt{\alpha_1}(\alpha_2-\alpha_1)d_2 + (\alpha_1-\alpha_2)d_3 - \sqrt{\alpha_1}d_4 + d_5 \\ \sqrt{\alpha_2}d_4 + d_5 \\ -\sqrt{\alpha_2}d_4 + d_5 \end{pmatrix} \tag{37}$$

If any of the elements of $\vec{d\prime}$ are zero then those state variables can be eliminated from the equation reducing the order of the system. Now consider the special case $\alpha_1 = \alpha_2 = \alpha$. In this case the transformation matrix $\mathbf{P}$, given in equation (35), becomes singular because $\vec{v}_2 = \vec{v}_3, \vec{v}_4 = \vec{v}_5$ in equation (5). The transition matrix, $\mathbf{A}$, is no longer similar to a diagonal matrix and generalized eigenvectors must be used, giving

$$\mathbf{P} = \begin{pmatrix} \alpha^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\alpha & 0 & 2\alpha \\ -\alpha & 0 & 2\sqrt{\alpha} & 0 & -2\sqrt{\alpha} \\ 0 & \sqrt{\alpha} & 1+\sqrt{\alpha} & -\sqrt{\alpha} & 1-\sqrt{\alpha} \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{38}$$

which leads to

$$\mathcal{A} = \mathbf{P^{-1}AP} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1+\sqrt{\alpha} & 1 & 0 & 0 \\ 0 & 0 & 1+\sqrt{\alpha} & 0 & 0 \\ 0 & 0 & 0 & 1-\sqrt{\alpha} & 1 \\ 0 & 0 & 0 & 0 & 1-\sqrt{\alpha} \end{pmatrix} \tag{39}$$

This non-diagonal canonical form of the transition matrix requires that the relation $\alpha_1 = \alpha_2$ hold to infinite precision. If $\alpha_1 \approx \alpha_2$ only for some finite (although possibly very high) precision, the transition matrix is similar to a diagonal matrix. This illustrates the discussion of the following section.

Non-Diagonal Canonical Modulators

Non-diagonal canonical modulators (NDC modulators), as opposed to the diagonal modulators, must be analysed on a system-by-system basis. Some NDC modulators, such as the one of figure (1) and equation (9), provide no easy way to simplify the form of the potential function. Other systems, such as the one described in equation (34) with $\alpha_1 \equiv \alpha_2 \equiv \alpha$ can be simplified as is shown in figure (9). After the transformation, the fifth order modulator of equation (34) becomes one first order system and two second order systems, in parallel, interacting only through the quantizer function.

It is possible to make some general statements about when an interpolative modulator has a NDC form. The use of "resonators" (a linear feedback path from the output of an integrator to the input of the integrator directly preceding it), such as is the case in figure (8), frequently results in NDC forms. If a modulator employs only a single resonator, then its canonical form is almost certainly non-diagonal. More generally, if a N'th order interpolative modulator is constructed with less than $(N-1)/2$ resonators *as the only linear feedback paths*, its canonical form is non-diagonal. Even if the modulator has $(N-1)/2$ or more resonators, a NDC form can result if some of the resonator coefficients are exactly equal. Inclusion of linear feedback paths other than resonators virtually guarantees a diagonal canonical form.

The NDC modulators therefore include a very large number of sigma-delta system implementations. A question must thus be asked as to *how important are the diagonal systems*? The answer is that they are very important for two reasons.

First, they are theoretically very interesting systems. Diagonal modulators are the simplest form of higher order modulators and can be analysed analytically. They can therefore provide insight into other higher order systems. In addition, all diagonal systems are encompassed within a single analysis method, as opposed to the NDC forms which require a system by system analysis. The architectures of different diagonal systems can be compared on a parameter value basis, once the different systems are put into diagonal form.

Second, all silicon implementations of sigma-delta modulators are diagonal modulators. This statement would seem to conflict with previous ones, however there is no conflict. It is true that in theory, many implemented systems are NDC systems. This is true in theory only, however. The non-diagonal Jordan form relies on element values of infinite precision. Almost any perturbation away from these values causes $\mathbf{A}$ to be similar to a diagonal matrix ([15]). Consider the system described in equation (34). In order for this matrix have the NDC form of equation (39) it is required that $\alpha_1 = \alpha_2$ to infinite precision, *and* that all of the 1's in the matrix be *exactly* equal to 1 to infinite precision. Conditions like these do not hold in a real system, and thus any implemented system is practically guaranteed to be in diagonal form.

Conclusion

Interpolative sigma delta modulators are linear systems with the addition of a nonlinear quantizer function. The nonlinearity thus introduced by the quantizer can be eliminated by dividing up state space into regions in which the quantizer output is constant. This pseudo-linear analysis is applicable to M-bit quantizers with a division of state space into M regions.

Diagonal modulators are a class of high order interpolative modulators whose transition matrices are similar to a diagonal matrix. In the diagonal form all of the state variables become decoupled, interacting only through the quantizer function. The diagonal form provides 1) a method of comparing a variety of architectures by transforming differences in architecture to differences in parameter values, and 2) a great simplification of the potential function, and thus any stability analysis. In addition, the diagonal form represents the true order of the modulator.

Non-diagonal canonical modulators make up an important class of systems; however, because they require infinite precision of the elements in the transition matrix, most implementations of these systems only approximate the NDC form and are in fact diagonal modulators.

Acknowledgments

This work was supported in part by an NSF Young Investigator Award (MIPS-92-57964), an ARO Young Investigator Award (DAAL03-92-G-0375), and a grant from the Joint Services Electronics Program (N00014-89-J-1023). In addition P. Steiner was supported in part by a fellowship from the Quebec government FCAR program. We also thank Wai Lee and Robert Adams of Analog Devices for their useful suggestions and input, and Bryan Steiner for providing a unique perspective on the problem.

References

[1] S. Hein, A. Zakhor. "Stability and scaling of double loop Sigma Delta modulators",*Proc. ISCAS 1992*, vol. 3, pp. 1312-15

[2] S.Hein, A. Zakhor. Sigma Delta Modulators, Kluwer Academic Publishers, 1993

[3] S.C. Pinault, P.V. Lopresti, "On the behavior of the double-loop sigma-delta modulator", *IEEE Trans. Cir. and Sys. II*, Aug. 1993, vol. 40, no. 8, pp. 467-79

[4] O. Feely, L.O. Chua, "Nonlinear dynamics of a class of analog-to-digital converters", *Int. Jour. of Bifurcation and Chaos in App. Sc. and Eng.*, June 1992, vol. 2, no. 2, pp. 325-40

[5] R. Schreier, W.M. Snelgrove, "Sigma Delta Modulation is a mapping", *Proc. ISCAS 1991*, vol. 5, pp. 2415-18

[6] B. Zhang, M. Goodson, R. Schreier, "Invariant Sets for General Second-Order Low-pass Delta-Sigma Modulators with DC Inputs", *Proc. ISCAS 1994*, vol. 6, pp. 1-4

[7] Hongmo Wang, "A Geometric View of Sigma Delta Modulation", *IEEE Trans. Cir. and Sys.*, June 1992, vol. 39, no. 6, pp. 402-5

[8] P. Steiner, W. Yang, "Stability Analysis of the Second Order $\Sigma - \Delta$ Modulator", *Proc. ISCAS 1994*, vol. 5, pp. 365-8

[9] S.H. Ardlan, J.J. Paulos, "An Analysis of Nonlinear Behavior in Delta-Sigma Modulators", *IEEE Trans. Cir. and Sys.*, June 1987, vol. CAS-34, no. 6, pp. 593-603

[10] J.C. Candy, O.J. Benjamin, "The Structure of Quantization Noise from Sigma-Delta Modulation", *IEEE Trans. Communic.*, Sept. 1981, vol. COM-29, no. 9, pp. 1316-23

[11] R.M. Gray, "Quantization Noise Spectra", *Trans. Info. Th.*, Nov. 1990, vol. 36, no. 6, pp. 1220-40

[12] K.C.H Chao, S. Nadeem, W. Lee, C.G. Sodini, "A Higher Order Topology for Interpolative Modulators for Oversampling A/D Converters", *IEEE Trans. Cir. and Sys.*, Mar. 1990, vol. 37, no. 3, pp. 309-18

[13] Robert Adams, "Design Aspects of High-Order Delta-Sigma A/D Converters", in *ISCAS'94 Circuits and Systems Tutorials*, pp. 235-60, June 1994

[14] R.W. Brockett, Finite dimensional linear systems, Wiley, 1970

[15] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, 1989

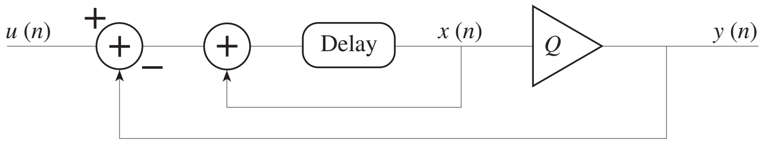

Figure 1: First Order $\Sigma - \Delta$ Converter (Q is quantizer function)

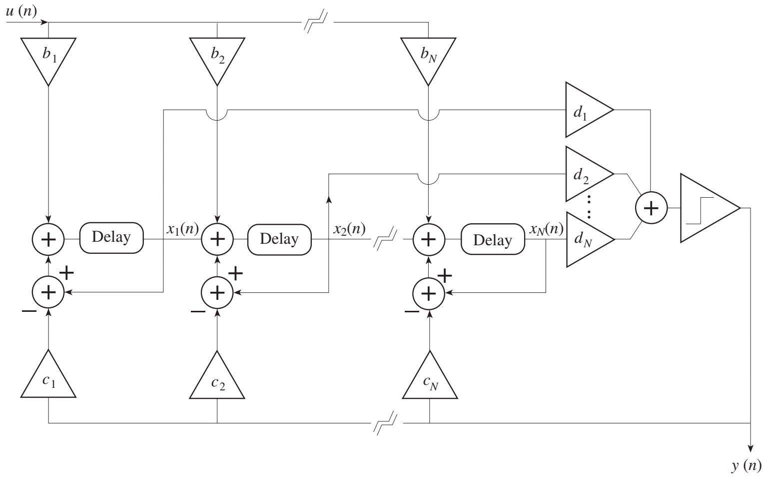

Figure 2: N'th order interpolative modulator with no linear feedback paths

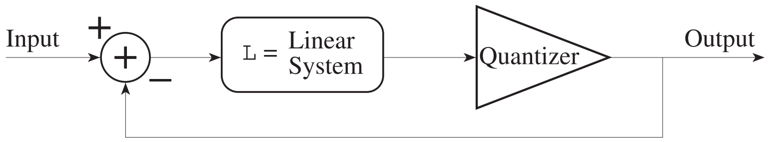

Figure 3: Block diagram of $\Sigma - \Delta$ systems

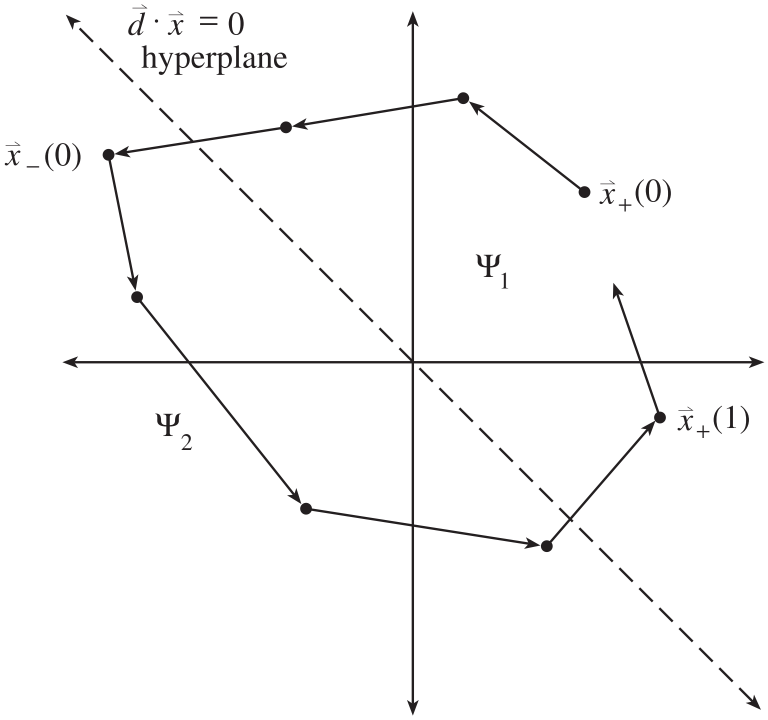

Figure 4: N-Dimensional state space divided into 2 regions by the quantizer function.

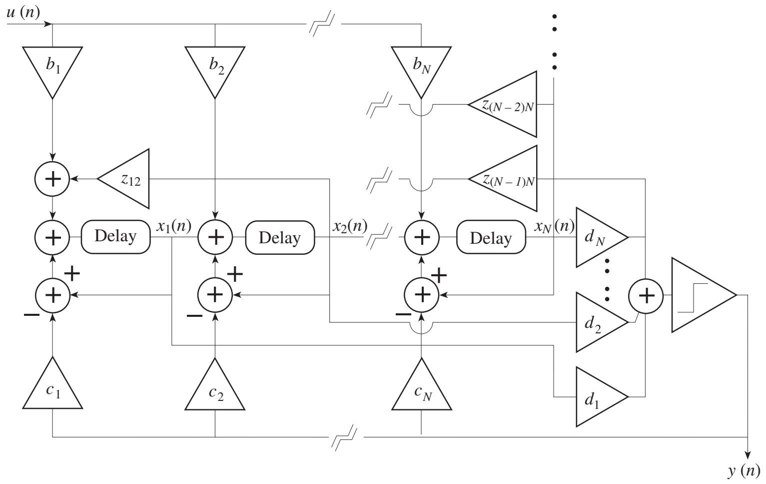

Figure 5: A general N'th order modulator in non-canonical form

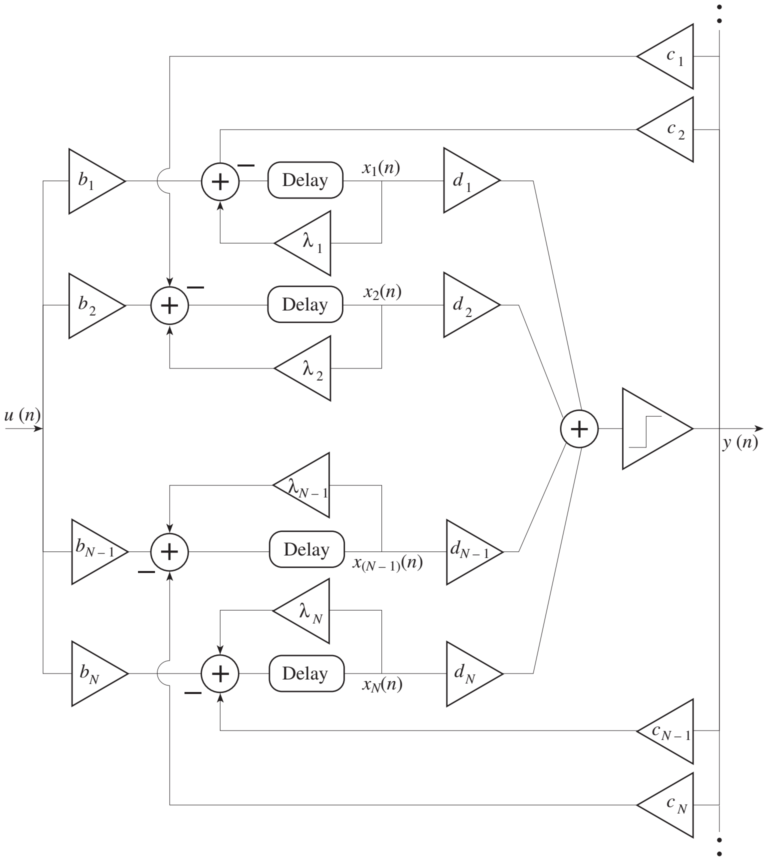

Figure 6: N'th order diagonal modulator

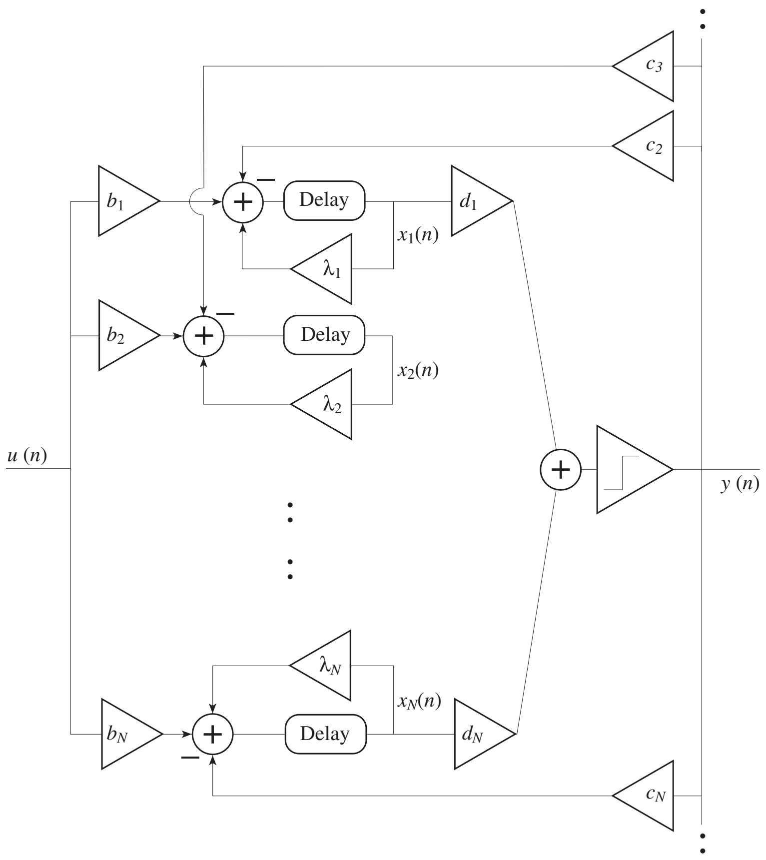

Figure 7: N'th order diagonal modulator with $d_2 = 0$. Notice that x_2 affects neither $y(k)$ nor any of the other states, x_j. x_2 can thus be ignored, and dropped from the dynamical equations.

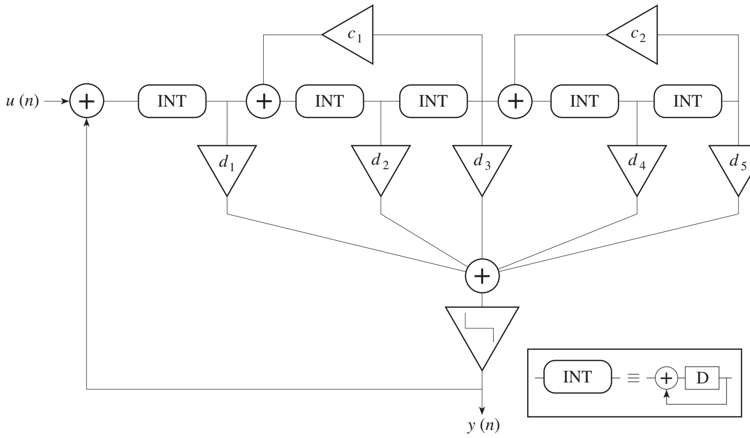

Figure 8: A 5th order modulator with two resonators.

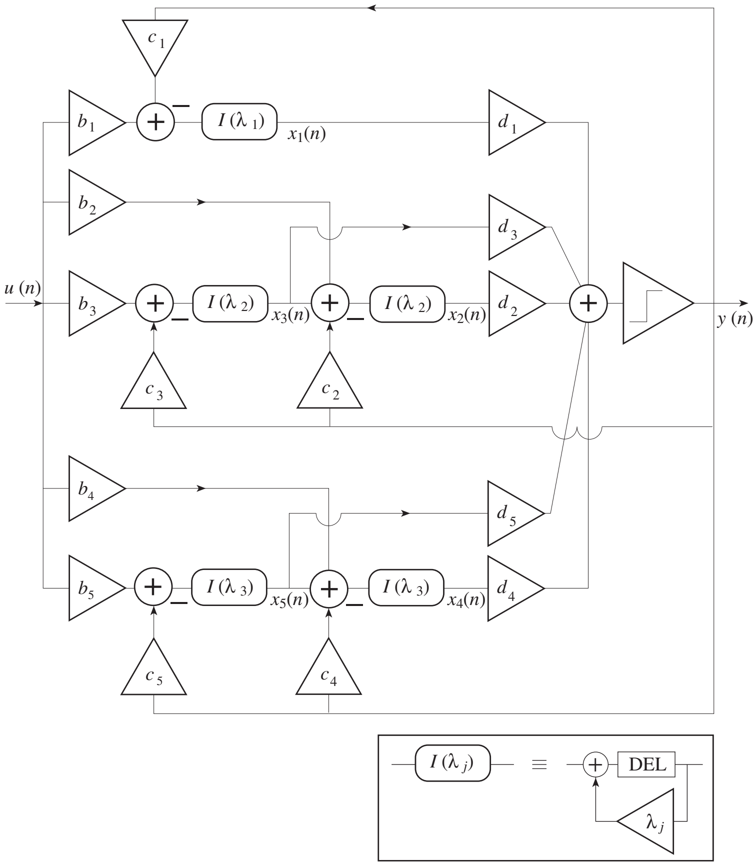

Figure 9: A 5th order NDC modulator where five cascaded integrators have become one first order and two second order modulators in parallel.