The 1999 Midwest Symposium on Circuits and Systems, New Mexico State University, Aug. 1999

Vector Switching Generalizes D/A Noise Shaping

Jeffrey O. Coleman
Dan P. Scholnik

jeffc@alum.mit.edu
scholnik@nrl.navy.mil

Naval Research Laboratory
Radar Division
Washington, DC

Abstract— **The most promising approach to obtaining the extraordinary in-band precision required of the internal DAC in multibit delta-sigma converters is dynamic element matching, in which errors due to DAC-hardware mismatch are spectrally shaped to avoid the signal band. Here we propose an architecture based on a vector switching signal and offer a mathematical framework for the analysis and design of such systems.**

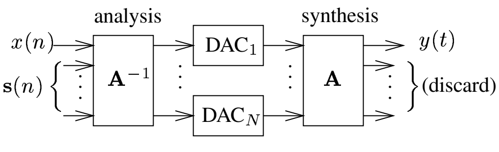

Fig. 1. The generalized noise-shaping DAC.

I. Introduction

Multibit $\Delta\Sigma$ conversion (both D/A and A/D) requires a DAC with exceptional precision in the band of interest. To avoid the expense and limitations of precise circuit calibration much research has been put into shaping the resulting circuit errors so they lie outside the signal band. This is generally accomplished by introducing redundancy into the DAC, implementing it as a summed bank of one-bit subconverters. The spectral properties of the errors can then be controlled by properly selecting which elements to enable for a given input at a given time, a technique known generally as dynamic element matching (DEM) [1].

Several different methods have been proposed for effecting the shaping of element mismatch errors in multibit DACs via dynamic element matching. Carley [2] proposed randomly choosing which elements to enable, whitening the mismatch noise spectrum. Leung and Sutarja [3] proposed two hardware-efficient techniques, clocked averaging and individual level averaging, that outperform random selection. Baird and Fiez [4] introduced a technique they called "data weighted averaging" (DWA), in which the elements to enable are chosen by a sliding window, producing first-order error shaping. Cini, *et. al.* [5] compare several DEM techniques and propose a modified version of DWA they call "double-index averaging". Schreier [6] and Galton [7] presented systems that allowed for DAC noise shaping of arbitrary order, with the DAC structures themselves incorporating $\Delta\Sigma$-like loops.

The present system generalizes on that of Galton [7], who presented an efficient tree-structured architecture for implementing an N-element DAC controlled by a single data line and $N-1$ shaped switching signals. The vector formulation used here is in the spirit of Schreier [6], who proposed calculating the DAC element values directly by quantizing a vector shaping loop. We propose a general vector-switching architecture for multibit DACs which allows an arbitrary number of output levels and any desired shaping-loop response (subject to stability constraints). The analysis here implicitly encompasses subconverter pulse-shape and pulse-timing errors as well as output-level errors. We demonstrate the approach with some simulation results.

This work was partially supported by the AMRFS program (ONR 31) of the Office of Naval Research. The remainder was supported by the Office of Naval Research through its block funding of the Naval Research Laboratory.

II. A System and its Analysis

A. I/O Behavior of the Generalized System.

Using the mixed discrete-/continuous-time convolution

$$(\mathbf{P}\star\mathbf{x})(t) \triangleq T\sum_k \mathbf{P}(t-kT)\mathbf{x}(k)$$

we model a DAC as a system that transforms an input sequence $\mathbf{x}(n)$ with (possibly vector) elements from some finite (scalar or vector) input set to a real (possibly vector) output function $(\mathbf{P}\star\mathbf{x})(t)+\mathbf{e}(t)$ through convolution of the input with (possibly matrix) pulse $\mathbf{P}(t)$ and addition of (possibly vector) error $\mathbf{e}(t)$. For example, the *unit DAC element* of Galton and others has a scalar input and output, a constant error ϵ_l, a zero-order-hold pulse of width T and area $\epsilon_h-\epsilon_l$, and a $\{0,1\}$ input set.

The vector notation better suits the proposed composite DAC system in Fig. 1. First, the internal bank of N scalar D/A subconverters can together be considered a single vector DAC with some diagonal matrix pulse $\mathbf{P}(t)$, some vector $\mathbf{e}(t)$, and a Cartesian-product input set. Second, the entire system is then a DAC: Using $\hat{\mathbf{u}}_1$ as the first standard unit vector and suppressing time dependence for notational brevity, the system output in terms of composite input vector $\mathbf{w}^{\mathrm{T}} \triangleq (x,\, \mathbf{s}^{\mathrm{T}})$ is $y=\hat{\mathbf{u}}_1^{\mathrm{T}}\mathbf{A}\mathbf{P}\star\mathbf{A}^{-1}\mathbf{w}+\hat{\mathbf{u}}_1^{\mathrm{T}}\mathbf{A}\mathbf{e}=\hat{\mathbf{u}}_1^{\mathrm{T}}\mathbf{A}\mathbf{P}\mathbf{A}^{-1}\star\mathbf{w}+\hat{\mathbf{u}}_1^{\mathrm{T}}\mathbf{A}\mathbf{e}$.

Suppose synthesis matrix $\mathbf{A}$ is "orthogonal" in the sense that $\mathbf{A}^T\mathbf{A}=N\mathbf{I}$, so that analysis matrix $\mathbf{A}^{-1}=\frac{1}{N}\mathbf{A}^T$, and for convenience write $\mathbf{A}^T$ with its first column in partitioned off as $\mathbf{A}^T=\begin{pmatrix}\mathbf{r}, & \mathbf{R}\end{pmatrix}$. Define scalar pulse $p(t)$ to be the system's unit-sample response with $\mathbf{s}=0$, so that $p(t)=\mathbf{r}^T\mathbf{P}(t)\mathbf{A}^{-1}\hat{\mathbf{u}}_1=\mathbf{r}^T\mathbf{P}(t)\mathbf{r}/N$. Then defining $\mathbf{Q}(t)\triangleq\mathbf{P}(t)-p(t)\mathbf{I}$ will result in $\mathbf{r}^T\mathbf{Q}(t)\mathbf{r}=0$, and the decomposition $\mathbf{P}(t)=p(t)\mathbf{I}+\mathbf{Q}(t)$ permits us to write $y=\mathbf{r}^T\mathbf{P}\mathbf{A}^{-1}\star\mathbf{w}+\mathbf{r}^T\mathbf{e}=p\star x+\frac{1}{N}\mathbf{r}^T\mathbf{Q}\mathbf{R}\star\mathbf{s}+\mathbf{r}^T\mathbf{e}$. For convenience, let

$\mathbf{q}^{\mathrm{T}}(t) \stackrel{\Delta}{=} \mathbf{r}^T\mathbf{Q}(t)\mathbf{R}/N$, so that $y = p \star x + \frac{1}{N}\mathbf{r}^T\mathbf{Q}\mathbf{R} \star \mathbf{s} + \mathbf{r}^T\mathbf{e}$ can be written concisely as $y = p \star x + \mathbf{q}^{\mathrm{T}} \star \mathbf{s} + \mathbf{r}^T\mathbf{e}$, showing that the system is a DAC for input $x(n)$ with pulse $p(t)$ and output error $(\mathbf{q}^{\mathrm{T}} \star \mathbf{s})(t) + \mathbf{r}^{\mathrm{T}}\mathbf{e}(t)$ provided the inputs to the subconverter DACs have acceptable values.

In the typical design $\mathbf{r}$ is a vector of ones. If there is no crosstalk between the subconverters, $\mathbf{P}(t)$ will be diagonal, $p(t)$ will be the average of its elements, and $\mathbf{Q}(t)$ will be a diagonal matrix of the deviations of the subconverter pulses from that average. If there is crosstalk $\mathbf{P}(t)$ will include (presumably small) off-diagonal elements, changing the picture slightly. Either way, $\mathbf{q}(t)$ completely captures the hardware errors to do with mismatched DAC levels (different pulse amplitudes along the diagonal of $\mathbf{P}$), timing error (different time shifts of those pulses), frequency response (different pulse shapes), and crosstalk (off-diagonal elements). The ideal case is represented by $\mathbf{q}(t) = 0$, and any nonzero value becomes a vector impulse response driven by switching process $\mathbf{s}(n)$.

The error term $\mathbf{e}(t)$ models the errors internal to each DAC subconverter. In the common case where the subconverters are unit DAC elements this represents just a constant amplitude offset. If the subconverters themselves are composite DACs, then $\mathbf{e}(t)$ is recursively defined in terms of the mismatch and offset errors of the subconverters.

B. *Spectral Analysis.*

A Wiener-filter development reveals that for a DAC with error function $\epsilon(t)$ the spectral error density that should be minimized in the signal band (where $S_x(fT) > 0$) is $S_\epsilon(f) - T^{-1}|S_{\epsilon x}(f)|^2/S_x(fT)$, where cross-spectral density $S_{\epsilon x}(f)$ is the Fourier transform of continuous-/discrete-time crosscorrelation $R_{\epsilon x}(\tau) \stackrel{\Delta}{=} E\left[\epsilon(nT+\tau)x^*(n)\right]$, assumed independent of n (for joint wide-sense stationarity). Here $\epsilon(t) = (\mathbf{q}^{\mathrm{T}} \star \mathbf{s})(t) + \mathbf{r}\,\mathbf{e}(t)$, so this density becomes

$$\begin{aligned}
&\mathbf{Q}^{\mathrm{T}}(f)T\left(S_{\mathbf{s}}(fT) - \tfrac{S_{\mathbf{s}x}(fT)S^H_{\mathbf{s}x}(fT)}{S_x(fT)}\right)\mathbf{Q}^*(f)\\
&+ 2\mathrm{Re}\left\{\mathbf{Q}^{\mathrm{T}}(f)\left(S^H_{\mathbf{es}}(f) - \tfrac{S_{\mathbf{s}x}(fT)S^H_{\mathbf{e}x}(f)}{S_x(fT)}\right)\mathbf{r}^H\right\}\\
&+ \mathbf{r}\left(S_{\mathbf{e}}(f) - \tfrac{S_{\mathbf{e}x}(f)S^H_{\mathbf{e}x}(f)}{TS_x(fT)}\right)\mathbf{r}^H.
\end{aligned}$$

If $\mathbf{e}(t)$ is a DC process the second and third terms vanish except at $f = 0$, and if DC errors are unimportant in the application we can take $\epsilon(t) = (\mathbf{q}^{\mathrm{T}} \star \mathbf{s})(t)$.

The interpretation yields the design goals. We want this component of vector process $\mathbf{s}(n)$ along a direction defined by vector $\mathbf{q}(t)$ to have as little power as possible in the signal band. Vector $\mathbf{q}(t)$ is completely determined by hardware errors and is completely unpredictable in vector direction, so this is effectively a requirement that $\mathbf{s}(n)$ have very little signal-band power in any vector direction. Curiously, this analysis suggests that it is best if whatever residual power remains correlates with $x(n)$ so that it will appear as part of the signal.

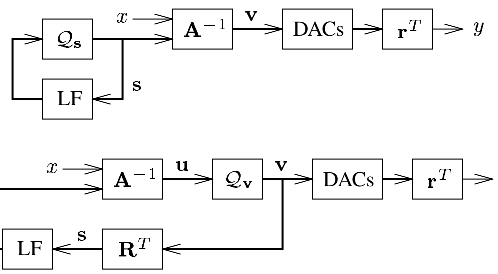

Fig. 2. Quantization of $\mathbf{s}$ in the upper system is replaced with the equivalent but more-practical quantization of $\mathbf{v}$ in the lower.

C. *Constraining the Subconverter DAC Inputs.*

The last section showed that vector switching signal $\mathbf{s}(n)$ must have certain spectral properties. Here we examine the other requirement on $\mathbf{s}(n)$, that it be constrained so as to place the inputs to the subconverter DACs in their allowable input ranges.

The upper diagram in Fig. 2 shows the system of Fig. 1 again but in streamlined notation and with the explicit inclusion of an inputless $\Delta\Sigma$-style loop for generating switching vector $\mathbf{s}$. The loop-filter LF determines the spectrum of the switching vector and hence the DAC output errors. It becomes the job of quantizer $\mathcal{Q}_{\mathbf{s}}$ to quantize and limit the loop-filter output in such a way that $\mathbf{v}$ satisfies the input requirements of the DACs. Since $\mathbf{R}^T\mathbf{v} = \mathbf{s}$, however, the loop-filter input can alternatively be generated from $\mathbf{v}$ as in the lower diagram, which includes a more-important change: the quantization is done at the DAC inputs. The advantage is simplicity; quantizer $\mathcal{Q}_{\mathbf{v}}$ can directly place the necessary limitations on $\mathbf{v}$ without the complication of an intervening matrix transformation. However, $\mathcal{Q}_{\mathbf{v}}$ must now operate under the constraint $\mathbf{r}^T\mathbf{v} = x$ to ensure that the relation between x and $\mathbf{v}$ is not disturbed by the quantization.

A simple constrained vector quantization now applies. We wish to choose an allowable $\mathbf{v}$ so that $\|\mathbf{v} - \mathbf{u}\|$ is minimized under the constraint $\mathbf{r}^T\mathbf{v} = x$. This nearest-neighbor decision rule is especially simple if one extra condition is met. Suppose that under the $\mathbf{r}^T\mathbf{v} = x$ constraint $\|\mathbf{v}\|$ is independent of the choice of $\mathbf{v}$. In that case $\|\mathbf{v} - \mathbf{u}\|^2 = \|\mathbf{v}\|^2 - 2\mathbf{u}^T\mathbf{v} + \|\mathbf{u}\|^2$ is affected only in the cross term by the choice of $\mathbf{v}$. Consequently, we can minimize $\|\mathbf{v} - \mathbf{u}\|$ by choosing $\mathbf{v}$ to maximize $\mathbf{u}^T\mathbf{v}$ under the constraint $\mathbf{r}^T\mathbf{v} = x$.

In the most common case this condition on $\|\mathbf{v}\|$ is met naturally and the decision rule becomes quite simple. Suppose each DAC element is binary with larger and smaller output values of β_ℓ and β_s, and suppose that $\mathbf{r}$ is a vector of ones. Then the constraint $\mathbf{r}^T\mathbf{v} = x$ fixes the numbers of β_ℓ and β_s elements in $\mathbf{v}$, in turn fixing $\|\mathbf{v}\|^2$. Suppose then that in order to meet the constraint, $\mathbf{v}$ must contain n_ℓ and n_s elements respectively equal to β_ℓ and β_s. Then $\mathbf{u}^T\mathbf{v}$ will be maximized by positioning these elements in $\mathbf{v}$ in alignment with the n_ℓ and n_s largest and smallest elements of $\mathbf{u}$. The nearest-neighbor rule is reduced to the identification of the n_ℓ largest elements of $\mathbf{u}$.

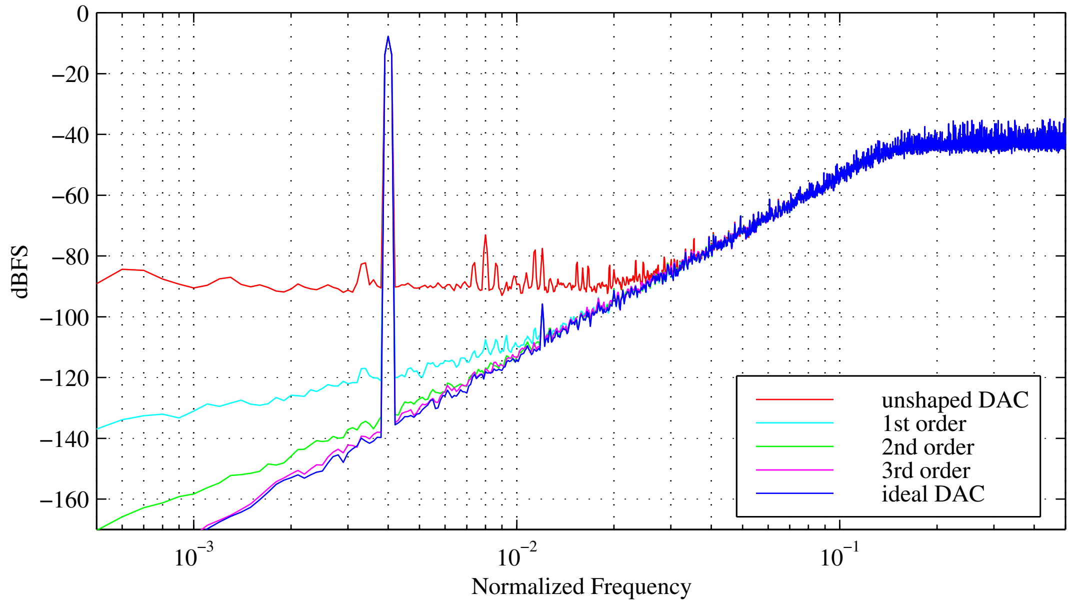

Fig. 3. Power spectra for the example, demonstrating the effects of element mismatch errors and the improvement possible by shaping the errors.

III. Example

The preceding discussion is illustrated through a simple example. Consider a three-element, four-output-level DAC, with 1% rms element level errors chosen from a Gaussian distribution. Other than the level errors, the elements are considered ideal; no pulse shape or offset errors is simulated. Let $\mathbf{r} = (1\ 1\ 1)^T$, so that the final output is just the sum of the three subconverters. The remaining rows of $\mathbf{A}$ are chosen arbitrarily to satisfy the orthogonality requirement. The input to the DAC is the output of a third-order $\Delta\Sigma$ modulator driven by a sinusoid. We consider the cases of first-, second-, and third-order shaping of the element mismatch errors. The actual transfer functions were chosen experimentally, and in general the assertion [8] that transfer functions that result in stable one-bit $\Delta\Sigma$ modulators will also be stable for element shaping appears to hold.

Figure 3 shows the various power spectra for this example, normalized to a full-scale sinusoidal input. The ideal curve shows the power spectrum of the $\Delta\Sigma$ modulator output, with no DAC-introduced errors. The unshaped curve represents the DAC output with a static assignment of input codes to element outputs according to a thermometer encoding. The middle three curves show the results of increasing the order of the element mismatch-shaping. The characteristic 20, 40, and 60 dB/decade slopes are apparent in the lowpass region of the plot. The third-order shaped curve is indistinguishable from that of the ideal DAC.

IV. Future Work

Although it is clear that noise-shaped dynamic element matching is a viable solution to linearize multibit DACs, open questions remain. One is how to optimally select the combination of $\Delta\Sigma$ shaping and DAC element mismatch shaping. The example presented suggests that using the same-order loop for both can yield essentially ideal results. The structural and functional similarity of the (scalar) $\Delta\Sigma$ modulator and the (vector) mismatch-shaping DAC suggest that it might be useful to consider a single unifying structure that directly modulates a real input onto one-bit DAC elements. Such a vector $\Delta\Sigma$ modulator might further simplify the analysis and lead to more efficient architectures.

References

[1] R. J. van de Plassche, "Dynamic element matching for high-accuracy monolithic D/A converters," *IEEE Journal of Solid-State Circuits*, vol. 11, pp. 795–800, Dec. 1976.

[2] L. R. Carley, "A noise-shaping coder topology for 15+ bit converters," *IEEE Journal of Solid-State Circuits*, vol. 24, no. 2, pp. 267–273, Apr. 1989.

[3] B. H. Leung and S. Sutaria, "Multibit $\Sigma - \Delta$ A/D convertor incorporating a novel class of dynamic element matching techniques," *IEEE Trans. Circuits and Systems II*, vol. 39, pp. 35–51, Jan. 1992.

[4] R. T. Baird and T. S. Fiez, "Linearity enhancement of multi-bit $\Delta\Sigma$ A/D and D/A converters using data weighted averaging," *IEEE Trans. Circuits and Systems II*, vol. 42, pp. 753–762, Dec. 1995.

[5] D. Cini, C. Samori, and A. L. Lacaita, "Double-index averaging: A novel technique for dynamic element matching in Σ-Δ A/D convertors," *IEEE Trans. Circuits and Systems II*, vol. 46, no. 4, pp. 353–358, Apr. 1999.

[6] R. Schreier and B. Zhang, "Noise-shaped multibit D/A convertor employing unit elements," *Electronics Letters*, vol. 31, no. 20, pp. 1712–1713, Sept. 1995.

[7] I. Galton, "Spectral shaping of circuit errors in digital-to-analog convertors," *IEEE Trans. Circuits and Systems II*, vol. 44, no. 10, pp. 808–817, Oct. 1997.

[8] T. Shui, R. Schreier, and F. Hudson, "Modified mismatch-shaping for continuous-time delta-sigma modulators," in *Proc. IEEE Custom Integrated Circuits Conf.*, 1998, pp. 225–228.