

- [8] Y. Sun, and J. K. Fidler, "Design of current-mode multiple output OTA and capacitor filters," *Int. J. Electron.*, vol. 81, no. 1, pp. 95–99, Jan. 1996.
- [9] M. Bhusan and R. W. Newcomb, "Grounding of capacitors in integrated circuits," *Electron. Lett.*, vol. 3, no. 4, pp. 148–149, Apr. 1967.
- [10] K. Pal and R. Singh, "Inductorless current conveyor allpass filter using grounded capacitors," *Electron. Lett.*, vol. 18, no. 1, p. 47, Jan. 1982.

## Sigma-Delta ADC with Reduced Sample Rate Multibit Quantizer

Wei Qin, Bo Hu, and Xieting Ling

**Abstract**—Based on the well-known Leslie–Singh architecture [1], a new cascaded sigma-delta analog-to-digital conversion (ADC) architecture is proposed. It incorporates a multibit quantizer whose sample rate can be significantly lower than the full oversampling speed of the sigma-delta modulator. Simulation results and comparison with other architectures are given. The architecture can be a good choice to extend the use of sigma-delta ADC to high bandwidth applications.

**Index Terms**—Analog-to-digital converters, oversampling converters, sigma-delta modulation.

### I. INTRODUCTION

Sigma-delta analog-to-digital conversion (ADC) has been enjoying increasing popularity over the past 15 years. It exploits the high speed of the modern fine-line CMOS process to obtain high-resolution ADC, without suffering from its inaccuracy.

However, due to its oversampling nature, the conversion bandwidth of sigma-delta ADC is severely limited. This limitation can be partly alleviated by using high-order architecture. It is believed that single-loop sigma-delta modulators with order higher than two are subject to instability and some empirical methods have been introduced to solve this problem [2], [3]. Cascaded multistage architecture, or MASH architecture [4], has also been proposed. It combines several low-order modulators to achieve a high-order noise-shaping function. Another means to reduce the oversampling ratio is to use a multibit quantizer [5]. In this case, highly linear multibit digital-to-analog conversion (DAC) is required, which would complicate the design greatly. In an architecture proposed by Leslie and Singh [1], no multibit DAC is required. This architecture can be viewed as another form of cascaded multistage sigma-delta ADC.

In this paper, based on the Leslie–Singh architecture and successive decimating strategy, a sigma-delta ADC with reduced sample rate multibit quantizer and no multibit DAC is proposed. The following parts of the paper are organized in this way: Section II describes the new architecture in detail, Section III analyzes nonidealities effects, Section IV gives a design example together with simulation results, and Section V makes comparative study between the new architecture and the existing ones.

Manuscript received July 8, 1998; revised February 12, 1999. This paper was recommended by Associate Editor H. Tanimoto.

The authors are with ASIC and System State Key Laboratory, Fudan University, Shanghai 200433, China.

Publisher Item Identifier S 1057-7130(99)04881-8.

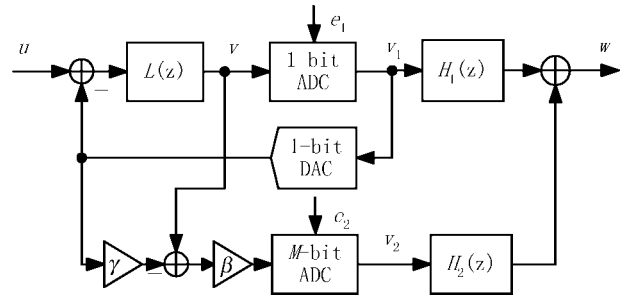


Fig. 1. The generalized Leslie–Singh architecture.

### II. THE NEW ARCHITECTURE

First, let us review the Leslie–Singh architecture [1] as well as the cascaded multistage sigma-delta modulator. Fig. 1 shows a generalized version of a single-loop Leslie–Singh architecture [5].

Let  $u$  be the input signal normalized by the 1-bit DAC's quantization step. The DAC's output is scaled by  $\gamma$  and subtracted from the loop filter output  $v$ . The difference is then passed to the multibit quantizer in the second stage with gain  $\beta$ . With the linear model of the 1-bit quantizer, i.e.,  $v_1 = \alpha v + e_1$ , where  $\alpha$  is the gain of the 1-bit quantizer and can be determined by the unity loop gain assumption [6], the signal transfer function  $G$  and noise transfer function  $H$  of the first stage in Fig. 1 are

$$G = \frac{\alpha L}{1 + \alpha L} \quad (1)$$

$$H = \frac{1}{1 + \alpha L}. \quad (2)$$

Then the digital output  $v_1$  and  $v_2$  can be viewed as

$$v_1 = Gu + He_1 \quad (3)$$

$$v_2 = \beta \left( \frac{v_1 - e_1}{\alpha} - \gamma \cdot v_1 \right) + e_2. \quad (4)$$

If the digital filters  $H_1$  and  $H_2$  are chosen to be

$$H_1 = 1 - H(1 - \alpha \cdot \gamma) \quad (5)$$

$$H_2 = \frac{\alpha}{\beta} H \quad (6)$$

then the output  $w$  will become

$$w = H_1 v_1 + H_2 v_2 = Gu + \frac{\alpha}{\beta} H e_2. \quad (7)$$

In (7), the coarse quantization noise  $e_1$  of the first stage can be totally cancelled in an ideal case. Usually,  $H$  here is a differential function, i.e.,  $H = (1 - z^{-1})^k$ , then the available dynamic range (dB) is

$$DR = 6.02 \cdot [M + (k + 0.5) \cdot \log_2 \text{OSR}] + 20 \cdot \log \left[ \frac{\beta \sqrt{2k + 1}}{\alpha \pi^k} \right] + 1.76 \quad (8)$$

where  $k$  is the order of loop filter  $L(z)$  and oversampling ratio (OSR) represents the oversampling ratio. The cascaded multistage sigma-delta modulator, or MASH architecture, which appeared earlier [4], is based on the same noise cancellation principle. The only difference is that another sigma-delta modulator is used instead of the multibit quantizer.

Evidently, the Leslie–Singh structure requires a multibit quantizer operating at the full oversampling speed, which makes its implementation difficult. Only very fast ADC architectures, such as flash

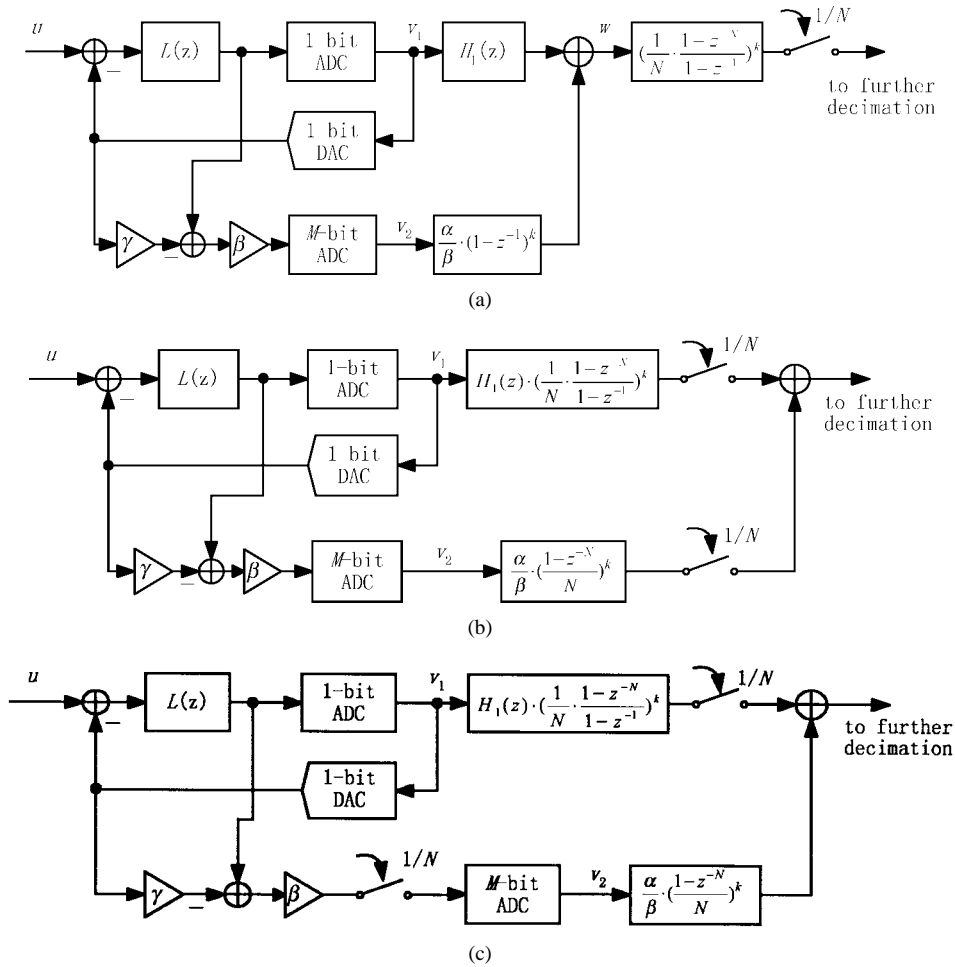


Fig. 2. Evolution to the reduced sample rate architecture.

and pipeline ADC, are eligible for even moderate-speed application. Moreover, the output multibit digital stream at full oversampling ratio adds an extra burden to the successive decimation circuit. All of these make the Leslie–Singh architecture less favorable than the MASH architecture.

However, if  $H$  is a differential function, there will be some similarity between transfer function  $H_2$  and the sinc decimation filter [7], which is commonly used for sigma–delta output. And, for the Leslie–Singh architecture, this similarity can be exploited to reduce the sample rate of the multibit ADC. When  $H$  is a  $k$ th-order differential function, from (6),  $H_2$  should be

$$H_2 = \frac{\alpha}{\beta} (1 - z^{-1})^k. \quad (9)$$

If we  $N$ -fold decimate the output word stream with a  $k$ th-order sinc filter, as is shown in Fig. 2(a),  $H_2$  can be cancelled with the denominator of the sinc filter. The resulting filter in Fig. 2(b) is simply a differential function of  $z^N$ , and its position can be exchanged with that of the  $N$ -fold down-sampler. Moreover, the multibit ADC can also be moved behind the down-sampler, thus  $N$  times reducing its sample rate. Clearly, the evolution process in Fig. 2 is strict and reversible, so all characteristics of the reduced sample rate architecture are identical to those of a Leslie–Singh sigma–delta ADC  $N$ -fold decimated by a  $k$ th-order sinc filter. The 1-bit quantizer in Fig. 2(c) still operates at full OSR, but its implementation is easy since it is simply a comparator.

It was shown by Candy [7] that the most economical initial decimation stage for a  $k$ th-order modulator is a  $(k + 1)$ th-order sinc

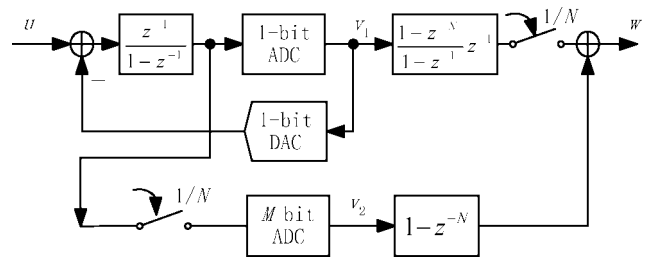


Fig. 3. The FRC architecture.

filter, with which the increase of baseband quantization noise is below 0.25 dB if the stream is decimated to above four times the Nyquist rate. Candy also pointed out that if decimated with a  $k$ th-order sinc filter, the noise power would increase by  $N$  times, where  $N$  is the decimation ratio. This means that by reducing  $N$  times the sample rate of the multibit quantizer in Fig. 2, we would have to pay  $0.5 \log_2 N$  bits of resolution. Then the dynamic range in dB of Fig. 2 will be

$$DR = 6.02 \cdot [M + (k + 0.5) \cdot \log_2 \text{OSR} - 0.5 \cdot \log_2 N] + 20 \cdot \log \left[ \frac{\beta \sqrt{2k+1}}{\alpha \pi^k} \right] + 1.76. \quad (10)$$

$N$ , or the speed reduction rate of the multibit quantizer, can be selected among the factors of OSR according to the resolution and speed requirements of specific applications. If  $N$  is chosen to be small so that the multibit quantizer operates at an oversampling ratio above four, the sinc filter can still be used in successive decimation,

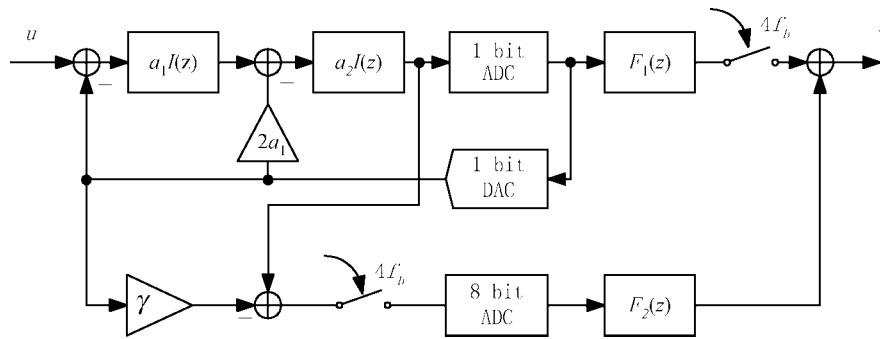


Fig. 4. A second-order design example.

because after  $k$ th-order sinc decimation in Fig. 2, the noise shape remains unchanged, although its power increases [7].

The 0.5-bit/octave cost is the same as the well-known resolution-speed tradeoff of conventional Nyquist rate ADC, and is acceptable in many cases. For instance, with a second-order modulator, a 6-bit multibit quantizer and an OSR of 64, 16 bits of resolution can be achieved under the Leslie-Singh architecture. However, under the architecture shown in Fig. 2(c), the same resolution can be obtained with an 8-bit quantizer operating at only four times the Nyquist rate. The 2-bit cost is worth paying, considering the significant reduction of speed by 16 times. In both cases, 2 bits of resolution loss due to the gain optimization of the integrators to prevent their output from overloading is considered.

#### A. FRC Architecture

Fig. 3 shows a special first-order case, in which  $\alpha = \beta = 1$ ,  $\gamma = 0$ ,  $N = \text{OSR}$  and  $H_1 = z^{-1}$ ,  $H_2 = 1 - z^{-1}$ .

The decimation filter for the 1-bit stream is simply a delayed first-order sinc filter and can be viewed as an accumulate-and-dump counter. The multibit output stream is first-order differentiated and summed with the counter's output. This special first-order structure is the same as a recently proposed feedforward residue compensation (FRC) structure [8], which is intended to extend the resolution of Nyquist rate ADC.

However, higher order FRC architectures are different from our architecture and have practical problems in implementation [8]. This is because of the fact that FRC is more like a conventional ADC than a sigma-delta ADC. While totally based on the sigma-delta modulation and the decimation technique, the architecture proposed here is relatively straightforward and flexible.

#### B. Design Considerations

As is mentioned above, tradeoff between resolution and speed exists and should be considered carefully. Besides, the antialias effect should also be taken into account when we choose the value of  $N$ . The attenuation of the out-of-band signal of a  $k$ th-order sinc filter can be evaluated by [7]

$$\Phi = \left( \frac{\sin(\pi \cdot (1/N - 1/(2 \cdot \text{OSR})))}{\sin(\pi/(2 \cdot \text{OSR}))} \right)^k. \quad (11)$$

The higher the order  $k$ , or the smaller the decimation fold  $N$ , the better the antialias attenuation will be. Since  $k$  is usually no larger than two for cascaded architectures, the antialias effect may be an important factor in determining the first-stage decimation fold  $N$ , and hence, the sample rate of the multibit ADC.

Gain  $\beta$  and  $\gamma$  in Figs. 1 and 2 are incorporated to maximize the input range. The smaller the value of  $\beta$ , the less likely that the input of the multibit quantizer will saturate, but, according to (10), the

lower the resolution will be. So another tradeoff exists. A positive  $\gamma$  helps to limit the maximum input level of the multibit quantizer and has no effect on the resolution. The value of  $\beta$  and  $\gamma$  can be optimized through repetitive simulation [6].

### III. NONIDEALITIES

Like the Leslie-Singh architecture and the MASH architecture, the proposed architecture is susceptible to circuit nonidealities because of its cascaded nature [5]. Mainly two kinds of nonidealities will affect its performance: gain mismatch and finite gain of operational amplifier. For the sake of brevity, we only focus on second-order modulators here.

#### A. Gain Mismatch

The leaky error caused by gain mismatch in Fig. 1 can be calculated as

$$\varepsilon_g = (\delta_\beta + \delta_\alpha) \cdot (1 - z^{-1})^2 \cdot e_1 + [\delta_\alpha - \delta_\beta + \alpha \cdot \gamma \cdot (\delta_\beta + \delta_\gamma)] \cdot (1 - z^{-1})^4 \cdot e_1 \quad (12)$$

where  $\delta_\alpha$ ,  $\delta_\beta$ , and  $\delta_\gamma$  are fractional errors of the 1-bit quantizer's gain  $\alpha$  and the gains  $\beta$  and  $\gamma$ . The second term in (12) contains double-shaped leaky quantization noise, which contributes most at high frequency but has little effect on baseband, so normally only the first term in (12) is considered [6]. However, since the out-of-band quantization noise was not suppressed completely when decimated with a  $k$ th-order sinc filter in Fig. 2, variation of high-frequency noise will affect the performance of our novel architecture. So, the second term in (12) should also be considered here. Consequently, the product of  $\alpha$  and  $\gamma$  should be kept no larger than unity, in order to minimize the mismatch effects of gain  $\beta$  and  $\gamma$ .

#### B. Finite Gain of Operational Amplifier

Finite gain of an operational amplifier will cause a gain error and a pole error in a switched capacitor (SC) integrator. Both errors are inversely proportional to the gain [5]. Integrator gain error can be included in  $\delta_\alpha$  of (12). Pole error may cause problems more serious than the mismatch error. For second-order modulators, the leaky error due to the pole errors  $\delta_{\theta 1}$  and  $\delta_{\theta 2}$  of the two integrators is approximately [5]

$$\varepsilon_\theta = (\delta_{\theta 1} + \delta_{\theta 2}) \cdot (1 - z^{-1}) \cdot e_1. \quad (13)$$

The noise is only first-order shaped and will probably have larger power than (12). For a first-order modulator, this leaky noise is even unshaped [5]. Thereby, if high resolution is desired, the gain-squaring technique of SC circuits [9] should be used to reduce the pole error.

It is worth noting that if the first-order shaped noise in (13) is uncorrelated with the fine quantization noise  $e_2$  in (7), its power will

TABLE I  
COMPARISON WITH OTHER ARCHITECTURES

Modulator Type	Advantages of Proposed Structure	Disadvantages of Proposed Structure
High-Order Single Loop	guaranteed stability simple loop filter design high SNR for low OSR	sensitive to circuit non-idealities, multi-bit quantizer required
Multi-bit Single Loop	no multi-bit DAC required low output rate of multi-bit stream simple circuit design	sensitive to circuit non-idealities
MASH	high SNR for low OSR low output rate of multi-bit stream	multi-bit quantizer required
High-Speed MASH	no multi-bit DAC required low output rate of multi-bit stream	high resolution multi-bit quantizer required
Leslie-Singh	low sample rate of multi-bit quantizer low output rate of multi-bit stream	0.5bit/octave loss of resolution

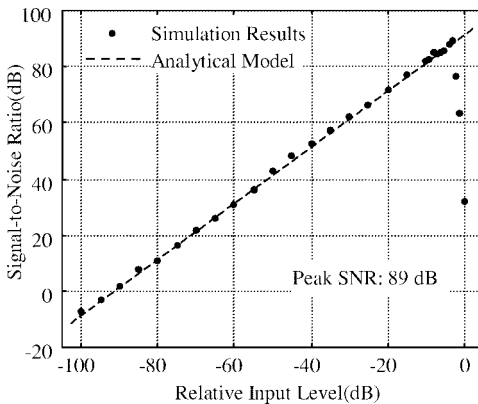


Fig. 5. SNR.

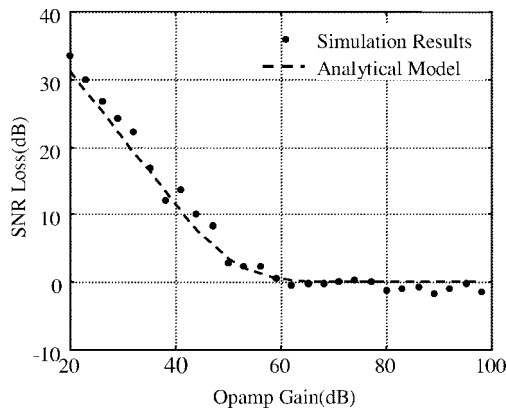


Fig. 6. Effect of finite opamp gain.

not increase when decimated with a second-order sinc filter. Hence, the power of  $\varepsilon_\theta$  is allowed to be  $N$  times as large as  $\varepsilon_2$ , which means that a proposed design with an  $M$ -bit quantizer is as tolerant to the pole errors of the integrators as a Leslie-Singh architecture with a quantizer of  $M - 0.5 \cdot \log_2 N$  bits.

IV. SIMULATION RESULTS

Refer to the second-order modulator in Fig. 4, in which  $I(z) = z^{-1}/(1 - z^{-1})$  and  $a_1, a_2, \gamma,$  and  $\beta$  are chosen to be 0.4, 0.625, 0.25, and 1, respectively. The integrator output is limited within 1.25 times of the quantization step. OSR is 32 and the 8-bit ADC operates at four times the Nyquist rate, i.e.,  $N = 8$ .

According to the unity loop gain assumption [6]

$$\alpha = \frac{1}{\alpha_1 \alpha_2} = 4. \tag{14}$$

Then from (5) and (6), filters  $H_1$  and  $H_2$  are

$$H_1 = 1 \tag{15}$$

$$H_2 = 4(1 - z^{-1})^2. \tag{16}$$

According to Fig. 2(c), in Fig. 4

$$F_1 = \frac{1}{64} \cdot \left( \frac{1 - z^{-8}}{1 - z^{-1}} \right)^2 \tag{17}$$

$$F_2 = \frac{1}{16} \cdot (1 - z^{-N})^2. \tag{18}$$

Simulation results with Simulink [10] are shown in Figs. 5 and 6. Fig. 5 shows different signal-to-noise ratio (SNR) under different input levels. 0-dB input is defined to be a sinusoidal, whose peak-to-peak amplitude equals the quantization step. The simulated results match well with those calculated with (10) when the input is below -3 dB. The peak SNR is 89 dB, so 14 bits of resolution is achieved.

Fig. 6 shows the loss of the SNR under different opamp gains. The simulation results also match well with those predicted by (12) and (13), and prove our assertion that the leaky noise in (13) can be larger than the fine quantization noise. An opamp gain of 60 dB is enough in this case.

V. COMPARISON

The proposed architecture provides an alternative choice for engineers. Its advantages as well as disadvantages, when compared with other architectures, are listed in Table I. High-speed MASH architecture in Table I refers to a modified MASH architecture whose second stage is a multibit sigma-delta modulator [11].

Generally speaking, the proposed architecture has good low OSR performance and is fit for higher bandwidth applications. Combination use of this architecture with others is also possible. For instance, multibit modulator or MASH architecture can precede the reduced sample rate multibit quantizer.

VI. CONCLUSION

A new cascaded sigma-delta ADC architecture is proposed, which includes a sigma-delta modulator and a reduced sample rate multibit quantizer. The architecture is flexible and provides one more choice for ADC applications. Resolution-speed tradeoff and antialias requirement are important factors in determining the sample rate of the multibit quantizer. This architecture may be used to reduce the oversampling ratio of sigma-delta ADC and extend its use into high-bandwidth applications.

## REFERENCES

- [1] T. C. Leslie and B. Singh, "An improved sigma-delta modulator architecture," in *Proc. IEEE Int. Symp. Circuits and Systems '85*, p. 372.
- [2] W. L. Lee and G. C. Sodini, "A topology for higher order interpolative coders," in *Proc. IEEE Int. Symp. Circuits and Systems '87*, p. 459.
- [3] R. W. Adams, P. F. Ferguson, A. Ganesan, S. Vincelette, A. Volpe, and R. Libert, "Theory and practical implementation of a fifth-order sigma-delta A/D converter," *J. Audio Eng. Soc.*, vol. 39, pp. 515–528, July 1991.
- [4] Y. Matsuya, K. Uchimura, A. Iwata, T. Kobayashi, M. Ishikawa, and T. Yoshitome, "A 16-bit oversampling A-to-D conversion technology using triple-integration noise shaping," *IEEE J. Solid-State Circuits*, vol. 22, pp. 921–929, Dec. 1987.
- [5] S. R. Norsworthy, R. Schreier, and G. C. Temes, *Delta-Sigma Data Converters—Theory, Design and Simulation*. New York: IEEE Press, 1997.
- [6] L. A. Williams, III and B. A. Wooley, "Third-order cascaded sigma-delta modulators," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 489–497, May 1991.
- [7] J. C. Candy, "Decimation for sigma delta modulation," *IEEE Trans. Commun.*, vol. COM-34, pp. 72–76, Jan. 1986.
- [8] R. Harjani and T. A. Lee, "FRC: A method for extending the resolution of Nyquist rate converters using oversampling," *IEEE Trans. Circuits and Systems II*, vol. 45, pp. 482–494, Apr. 1998.
- [9] K. Huang, F. Maloberti, and G. Temes, "Switched-capacitor integrators with low finite-gain sensitivity," *Electron. Lett.*, vol. 21, pp. 1156–1157, Nov. 1985.
- [10] *The Simulink Manual*. The Mathworks, Inc., Natick, MA, 1997.
- [11] B. P. Brandt and B. A. Wooley, "A 50-MHz multibit sigma-delta modulator for 12-b 2-MHz A/D conversion," *IEEE J. Solid-State Circuits*, vol. 26, pp. 1746–1756, Dec. 1991.

## Efficient Architectures to Recover the Regularized Least Squares Solution

R. Sundaram

**Abstract**—Several practical applications are concerned with the identification of the least squares (LS) solution. The objective is to attain this solution accurately and efficiently while conserving resources. The computational and storage requirements to determine the LS solution by any iterative procedure become prohibitively large as the problem dimensions grow. This brief presents some architectures based on thresholded binary networks which recover regularized LS solutions by partitioning such networks and adopting a switching operation between active and inactive partitions to optimize the objective function. Also, an iterative method based on steepest descent is briefly discussed and implemented. It yields reliable estimates of the regularized LS solution, while providing savings in computation and storage.

### I. INTRODUCTION

The linear shift invariant (LSI) degradation model is well established and widely used to recover image data from corrupted samples [1], [3]. The two most dominant forms of degradation affecting images are blurring and noise. Inverse filters, which merely seek to undo the blurring, have limited usefulness. They accentuate the noise corruption at higher frequencies since blurring emphasizes

Manuscript received December 1, 1997; revised February 19, 1999. This paper was recommended by Associate Editor J. B. Dias.

The author was with Information Systems Inc., West Lafayette, IN 47906 USA. He is now at 2189 Bedell Rd., #9, Grand Island, NY 14027 USA.

Publisher Item Identifier S 1057-7130(99)04873-9.

the low-frequency content. Regularized restoration [5] is achieved by noise-sensitive filters at the cost of a systematic bias in the solution. Iterative restoration procedures [2]–[4] permit the inclusion of external or *a priori* constraints and termination of convergence prior to the attainment of the limiting solution. Thresholded binary networks based on the discrete Hopfield model [6]–[11] lead to robust retrieval of the limiting solution. This brief discusses architectures based on partitions of thresholded binary networks. Section II briefly outlines the iterative formulation of the LS problem. Section III presents update strategies on partitions of the network. Section IV identifies the application to image restoration. Section V contains the conclusions.

### II. ITERATIVE FORMULATION OF THE PROBLEM

Let  $y(i, j)$  denote the observed (degraded) image intensity at the location  $(i, j)$  specified on a  $V \times V$  rectangular grid using cartesian coordinates. Then, according to the LSI model

$$y(i, j) = \sum_{m, n \in R_h} h(m, n)x(i - m, j - n) + \nu(i, j). \quad (1)$$

Here,  $R_h$  denotes the region of support of the point spread function (PSF), with  $h(m, n)$  representing the discretized blur,  $x(i, j)$  the unknown (true) intensity at the location  $(i, j)$ , and  $\nu(i, j)$  the additive white noise sample at the location  $(i, j)$ . The lexicographic ordering of the data, PSF, and noise samples (either by row or column) leads to the matrix-vector relation

$$Y = HX + \hat{N} \quad (2)$$

where  $Y$ ,  $X$  and  $\hat{N}$  are now  $V^2 \times 1$  column vectors. The matrix  $H$  contains the blur parameters and is doubly block-circulant of dimension  $V^2 \times V^2$ . The unregularized least squares (LS) solution  $\hat{X} = \hat{X}_{UR}$  corresponds to minimizing  $\phi(\hat{X})$  as given by

$$\phi(\hat{X}) = \frac{1}{2} \|Y - H\hat{X}\|^2. \quad (3)$$

This solution is  $\hat{X}_{UR} = (H^t H)^{-1} H^t Y$ . When regularization is incorporated, the objective function in (3) is modified to

$$\psi(\hat{X}) = \frac{1}{2} \|Y - H\hat{X}\|^2 + \frac{\lambda}{2} \|D\hat{X}\|^2 \quad (4)$$

where  $D$  represents the chosen regularization matrix (also doubly block-circulant) and  $\lambda$  is the chosen regularization constant. This leads to the regularized LS solution  $\hat{X} = \hat{X}_R$  given by  $\hat{X}_R = (H^t H + \lambda D^t D)^{-1} H^t Y$ . It is easy to relate  $\hat{X}_{UR}$  to  $\hat{X}_R$  as

$$\hat{X}_R = (W_\lambda)^{-1} (H^t H) \hat{X}_{UR} \quad (5)$$

where  $W_\lambda = H^t H + \lambda D^t D = W_\lambda^t$ . Henceforth,  $W_\lambda$  will be referred to as the interconnect matrix and  $\theta = H^t Y$  as the bias vector. The regularized LS solution  $\hat{X}_R$  is then

$$\hat{X}_R = W_\lambda^{-1} \theta. \quad (6)$$

The objective function in (4) can also be minimized using thresholded binary networks based on the discrete Hopfield model [6]–[10]. The network operates recursively on elements which assume binary states (i.e., either of two levels) based on a thresholding nonlinearity. The updates occur in a sequential or parallel manner as discussed in the next section.