An Efficient $\Delta\Sigma$ ADC Architecture for Low Oversampling Ratios

János Márkus, *Student Member, IEEE,* and Gabor C. Temes, *Life Fellow, IEEE*

***Abstract*—As the demand for $\Delta\Sigma$ (Delta-Sigma) analog-to-digital converters (ADCs) with higher bandwidth and higher signal-to-noise ratio (SNR) increases, designers have to look for efficient structures with low oversampling ratio (OSR). The Leslie-Singh or M–0 MASH architecture is often used in such applications. Based on this architecture, a reduced-sample-rate structure was introduced, which needs less chip area and power, but increases the noise floor. This paper describes a modification of the reduced-sample-rate structure which realizes an optimized transfer function, and avoids an SNR loss. In fact, it increases the SNR for high-order modulators. The method can also be applied to one-stage modulators. Simulation results for different MASH ADCs and sensitivity analysis verify the usefulness of the proposed technique.**

***Index Terms*—Leslie–Singh, MASH, modified (rotated) sinc-filter, multistage delta-sigma modulator, optimized noise transfer function, reduced-sample-rate.**

I. Introduction

THE most common $\Delta\Sigma$ ADC structure used for high oversampling ratio (OSR) and high resolution is the high-order $(M = 4 \sim 6)$ one-stage architecture. Their advantages include well-known architectures and design procedures and low sensitivity to the mismatch of its analog components [4, Sec. 5.2].

However, as the demand for $\Delta\Sigma$ analog-to-digital converters (ADCs) with higher bandwidth increases, designers have to contend with a lower OSR, and the classical one-stage high-order structures are no longer efficient. Table I shows the theoretically available SNR for modulators with different orders as a function of the OSR for 1-b quantization [5]. It is clear that at an OSR < 8 the improvement of using high-order 1-b structures becomes insignificant, while it leads to instability and design problems, and also increases the power and area requirements. As an example, for an OSR of 4, the improvement in resolution gained by increasing of the order by 1 is only about 0.6 b. Since in actual high-order modulator design the input needs to be restricted and nonzero poles need to be introduced to stabilize the loop, the achievable SNR is even lower. Hence, 1-b high-order structures are unsuitable for wide-band $\Delta\Sigma$ ADC design. Using multibit feedback, better results can be achieved but as the feedback DAC error limits the achievable performance, dynamic element matching (DEM) [6] or adaptive error compensation techniques must be used [7], [8].

TABLE I
SNR AT DIFFERENT OSRS USING DIFFERENT ORDER ONE-STAGE, ONE-BIT ARCHITECTURES

Order (M)	OSR=2	OSR=4	OSR=8	OSR=16
1	8.5 dB	17.6 dB	26.6 dB	35.6 dB
2	6.8 dB	21.9 dB	36.9 dB	52.0 dB
3	4.4 dB	25.4 dB	46.5 dB	67.6 dB
4	1.5 dB	28.6 dB	55.7 dB	82.8 dB
5	-1.5 dB	31.6 dB	64.7 dB	97.8 dB

A useful alternative is the MASH or multistage $\Delta\Sigma$ converter, in which low-order stages are connected in cascade, each stage converting and cancelling the quantization error of the preceding one, and thus achieving higher SNR [4, Sec. 6.2.1]. In these structures, multibit DACs may be used in the correction stages, which increases the SNR even further, with no need of correcting the DAC errors. The resolution of MASH ADCs is less sensitive to low OSR than that of single-stage high-order modulators [4]. Unfortunately, at low OSR, the mismatch-shaping algorithms commonly used to filter DAC nonlinearity errors out of the band of interest are not efficient anymore. Adaptive DAC error calibration can solve this problem [7], but it requires additional digital circuitry.

Improved performance can be achieved in all these converters by spreading the zeros of the noise transfer function NTF across the band of interest [3]. This technique can increase the SNR by 3.5, 8, 13 and 18 dB for a second-, third-, fourth-, and fifth-order system, respectively. Also, improved decimation filters have been introduced which reduce the aliasing noise by spreading their zeros over the critical stopband frequency range [9].

In this paper, a modified MASH structure is proposed. It is based on the reduced-sample-rate architecture [2], but eliminates the SNR loss inherent in it. The proposed design technique involves the optimization of both the NTF and the decimation filter. It is shown that this process is a generalization of the optimization of a full-speed one-stage architecture [3]. Hence, the results and closed-form relations, derived here for second- and third-order systems, can also be used for optimizing single-stage modulators.

Preliminary results for a second-order system have been disclosed in [10]. Here, the theory is generalized to higher order modulators, and practical realization issues are also discussed.

Manuscript received January 14, 2003; revised August 22, 2003. This work was supported in part by the NSF Center for Design of Analog-Digital Integrated Circuits, by the NSF US-Hungary Grant, under Contract UX026A, and by the Hungarian Fund for Scientific Research under Contract OTKA T 033 053. This paper was recommended by Guest Editors A. Rodríguez-Vázquez, F. Mediero, and O. Feely.

J. Márkus is with the Department of Measurement and Information Systems, Budapest University of Technology and Economics, H-1521 Budapest, Hungary (e-mail: markus@mit.bme.hu).

G. C. Temes is with the Department of Electrical and Computer Science, Oregon State University, Corvallis, OR 97331 USA (e-mail: temes@eecs.oregonstate.edu).

Digital Object Identifier 10.1109/TCSI.2003.821280

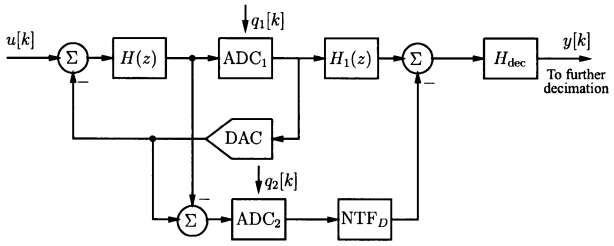


Fig. 1. The generalized Leslie–Singh architecture.

The paper is organized as follows. Section II discusses the existing multistage architectures, while Section III describes the optimization for a second-order system in detail. Section IV discusses the extension of the optimization process for third and higher order systems, and Section V verifies the theoretical results with simulations. Finally, Section VI contains the conclusions.

II. The Reduced-Sample-Rate Architecture

A. The Leslie–Singh Architecture

A popular multistage architecture was introduced by Leslie and Singh [1]. Their topology combines multibit quantization with single-bit feedback. It can be regarded as an $M-0$ MASH structure, where the second stage (the multibit ADC) processes the quantization error (Fig. 1) of the first one. In the ideal case (with no mismatch errors between the analog and digital noise transfer functions), the output of this modulator contains only the input signal $(U(z))$ and the quantization noise of the second stage $(Q_2(z))$

$$Y(z) = H_1(z)H_{\rm dec}(z)U(z) + {\rm NTF}_D(z)H_{\rm dec}(z)Q_2(z) \quad (1)$$

where $H_1(z)$ is the digital replica of the signal transfer function $({\rm STF})$ of the second stage (which is only a delay if the second stage is a pipeline converter), ${\rm NTF}_D(z)$ is the digital replica of the first-stage noise transfer function $({\rm NTF}(z))$, and $H_{\rm dec}(z)$ is the transfer function of the decimation filter.

The main advantage of the architecture is that it uses only single-bit feedback, and hence avoids the matching problems inherent in a multibit feedback DAC, but still provides the higher resolution and increased SNR of a modulator with multibit quantizer.

The structure requires in the second stage a multibit analog-to-digital converter $({\rm ADC}_2)$ operating at the oversampling clock frequency. For high clock rates, this makes its implementation difficult. This problem has been addressed [2] by introducing the reduced-sample-rate architecture shown in Fig. 2. Its operation will be discussed next.

B. The Reduced-Sample-Rate Architecture [2]

Often, the noise transfer function of the first stage of the MASH structure is simply an Mth-order difference function, i.e.,

$${\rm NTF} = \left(1 - z^{-1}\right)^M \quad (2)$$

where M is the order of the modulator.

Using a multistage decimating strategy, the first stage of the decimation filter is usually the sinc filter

$$H_{\rm dec}(z) = \frac{1}{N^{M'}} \frac{(1 - z^{-N})^{M'}}{(1 - z^{-1})^{M'}} \quad (3)$$

where N is the decimation factor of the first stage and $M' \geq M$ is the order of the decimator [11], [12].

The reduced-sample-rate architecture is based on the observation that, if $M' = M$ is chosen, then the ${\rm NTF}_D$ and the denominator of the decimation filter cancel, so (1) becomes

$$Y(z) = H_1(z)H_{\rm dec}(z)U(z) + {\rm num}\left[H{\rm dec}(z)\right]Q_2(z) \quad (4)$$

where ${\rm num}\left[H_{\rm dec}(z)\right]$ denotes the numerator of the transfer function $H_{\rm dec}(z)$. As this numerator contains only z^{-N} as a variable, one can use a decimated clock frequency for its implementation. Thus, after a minor rearrangement, a structure can be obtained in which the second stage containing the multibit quantizer can operate at a reduced clock rate (Fig. 2).

It should be noted that the reduced-sample-rate architecture cannot realize modulators with nonzero poles stabilizing the loop of the first stage. However, as discussed in Section I (cf. Table I), using modulator orders higher than third is inefficient in a low OSR design. For this reason, this paper focuses on second- and third-order systems for which it is usually unnecessary to shift the poles of the noise transfer function.

Another drawback of the reduced-sample-rate architecture is that it uses a decimation filter with an order only $M' = M$, and hence it causes the output noise power to increase N times, resulting in a noise-floor increase in the output of the modulator [2], [12]. This can also be seen from Fig. 2: with every doubling of N, the white quantization noise of the second quantizer doubles due to aliasing.

In Sections III–V, a design method is introduced which restores this SNR loss.

III. Optimization of the Second-Order Structure

In the reduced-sample-rate (RSR) system discussed in Section II, both the NTF and $H_{\rm dec}(z)$ had only multiple (M-tuple) zeros and poles. In the RSR architecture, ${\rm num}\left[H_{\rm dec}(z)\right]$ replaces the NTF. Hence, the in-band noise can be reduced by relocating the zeros of $H_{\rm dec}(z)$ from $z = 1$, spreading them optimally across the signal band. This technique has been described earlier in a different context for single-stage modulators in [3]. As shown below, the minimization problem solved here can be regarded as a generalization of the one described in [3].

Once the zeros of the decimation filter have been altered, the poles of $H_{\rm dec}(z)$ need also be moved to keep its magnitude constant in the passband. In addition, since the reduced-sample-rate structure is based on the cancellation of the ${\rm NTF}_D$, the NTF of the first stage also needs to be adjusted.

In the following, the proposed design technique for accomplishing this will be discussed in the context of a second-order modulator.

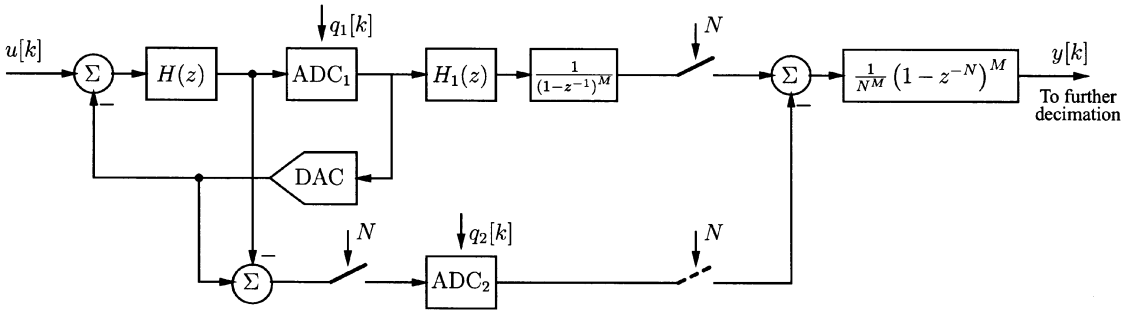


Fig. 2. The reduced-sample-rate architecture. M is the order of the modulator and the decimation filter, and N is the decimating ratio.

A. Optimization of the Decimation Filter

Assuming that the decimation ratio of the first digital filter stage is N, a second-order decimating transfer function with its zeros and poles shifted from $z = 1$ can be written as follows:

$$H_{\rm dec}(z) = \frac{1 - 2(\cos N\alpha)z^{-N} + z^{-2N}}{1 - 2(\cos\alpha)z^{-1} + z^{-2}} \tag{5}$$

where $z = e^{j2\pi f/f_s}$ and α represents the angle of the complex conjugate zeros. If $\alpha = 0$, the expression simplifies to the transfer function of the classic sinc filter. The transfer function in (5) defines the rotated sinc (RS) filter; it was first proposed in [9]. In our system, α should be optimized to minimize the inband noise power in the output of the whole RSR system. Assuming that the quantization noise of the second-stage ADC is white, the optimum value satisfies

$$\begin{aligned}\alpha_{\rm opt} &= \arg\min_{\alpha} \int_0^{f_s/2{\rm OSR}} |{\rm num}(H_{\rm dec})|^2\, df \\ &= \arg\min_{\alpha} \int_0^{f_s/2{\rm OSR}} \left|1 - 2(\cos N\alpha)z^{-N} + z^{-2N}\right|^2 \\ &\quad \times df. \end{aligned} \tag{6}$$

The assumption that the second-stage quantization noise is additive and white is valid under most circumstances, since: 1) the resolution of the quantizer is usually high (5–12 b) and 2) the input to the quantizer is the quantization error of the first stage, which is usually a "busy" signal, especially when dithering is used in the first stage or when the first loop uses multibit quantization.

As stated above, this process is similar to the optimization of the NTF of a single-stage $\Delta\Sigma$ modulator [3]. The difference is that in our case (6) contains one more parameter N, the decimation ratio of the first stage of the decimation filter. This complicates the problem somewhat.

To solve the minimization problem, [3] used an approximating method assuming OSR $\gg$ 1. Here, however, the optimal solution depends on the ratio of OSR$/N$, which is usually small ($\leq$4). In the second-order example shown in Section V, this number equals 2; in the case of a single-stage decimation filter, as used, e.g., in incremental $\Delta\Sigma$ structures [13], it may equal 1.

For this reason, an exact closed-form solution of the optimization is needed. It is proven in the Appendix that the optimal α is given by

$$\alpha_{\rm opt} = \frac{1}{N}\cos^{-1}\left(\frac{\sin\left(\frac{\pi N}{\rm OSR}\right)}{\frac{\pi N}{\rm OSR}}\right). \tag{7}$$

For $N = 1$, this formula gives the exact optimal zero-placement in the NTF of a single-stage $\Delta\Sigma$ modulator. If $N = 1$ and OSR $\gg 1$ then the following approximation can be derived using Taylor series expansion:

$$\begin{aligned}\frac{\alpha_{\rm opt}}{\omega_B} &= \frac{\alpha_{\rm opt}}{\frac{2\pi}{2{\rm OSR}}} = \frac{\rm OSR}{\pi}\cos^{-1}\left(\frac{\sin\left(\frac{\pi}{\rm OSR}\right)}{\frac{\pi}{\rm OSR}}\right)\Bigg|_{{\rm OSR}\gg 1} \\ &\approx \frac{\rm OSR}{\pi}\cos^{-1}\left(1 - \frac{1}{6}\left(\frac{\pi}{\rm OSR}\right)^2 + \frac{1}{120}\left(\frac{\pi}{\rm OSR}\right)^4\right) \\ &\approx \frac{\rm OSR}{\pi}\left(\frac{1}{3}\sqrt{3}\frac{\pi}{\rm OSR} - \frac{1}{270}\sqrt{3}\left(\frac{\pi}{\rm OSR}\right)^3\right) \\ &\approx \frac{1}{\sqrt{3}} \end{aligned} \tag{8}$$

which agrees with the result in [3, Table I].

Comparing the exact result with that given by the approximate one (Fig. 3), it turns out that the error in the approximate result is usually small (it can be seen also from the Taylor-series expansion derived above). For OSR$/N$ = 2, the increase in noise power using the approximate solution is only about 0.1 dB. However, if OSR$/N$ = 1 (as in the case of an incremental converter), the increase is about 0.5 dB for a second-order modulator and about 1 dB for a third-order one.

B. Optimization of the First Stage of the Modulator

Since the NTF of the optimized structure has a pair of complex zeros in place of the usual double real zeros at $z = 1$, the architecture of the first stage has to be modified. To realize the complex conjugate zeros, at least one additional feedback loop has to be introduced in the original structure.

Fig. 4 shows a recently proposed structure with an NTF = $\left(1 - z^{-1}\right)^2 = 1 - 2z^{-1} + z^{-2}$[14] and two modified structures realizing NTF = $1 - (2 - g)z^{-1} + z^{-2}$. In the modified configurations, $g = 2 - 2\cos\alpha$. Both structures can implement the

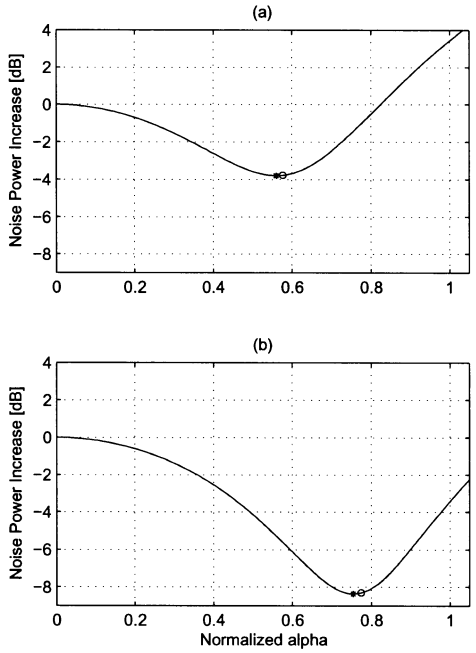


Fig. 3. The maximum SNR improvement as a function of the normalized $\alpha_n = \alpha/\omega_B$ at an OSR $= 4$ and $N = 2$. ∘ is the optimum point found by the approximation method, $*$ is the optimum point of the exact method. (a) Second-order system (maximum improvement 3.8 dB). (b) Third-order system (maximum improvement 8.3 dB).

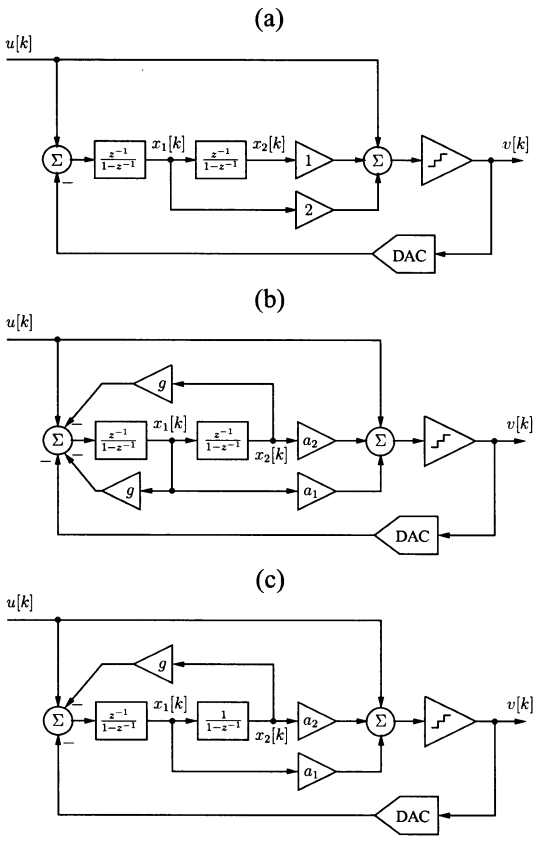


Fig. 4. Realization of the first stage. (a) Modulator with NTF $= (1-z^{-1})^2 = 1-2z^{-1}+z^{-2}$. (b) Integrator based (CIFF) modulator with NTF $= 1-(2-g)z^{-1}+z^{-2}$, $a_1 = 2-g$, $a_2 = 1-g$. (c) Resonator based (CRFF) modulator with NTF $= 1-(2-g)z^{-1}+z^{-2}$, $a_1 = 1$, $a_2 = 1-g$.

NTF $= 1-2(\cos\alpha)z^{-1}+z^{-2}$ transfer function required for the optimized RSR structure.

The circuit of Fig. 4(b) uses two delaying integrators, similar to that shown in Fig. 4(a); however, this structure needs two additional feedback branches, which makes its implementation somewhat complicated. By contrast, the structure of Fig. 4(c) needs only one internal feedback path and uses simpler coefficients, but the whole outer loop contains only one delay, which may cause difficulties in its circuit-level implementation, especially at a high clock rate. In a wideband switched-capacitor circuit realization, having multiple feedback may be less of a challenge than too few delay blocks. In general, the best configuration depends on the specifications of the converter.

The structures of Fig. 4(b) and (c) share the advantages of the original circuit of Fig. 4(a): 1) due to the single DAC feedback path and the feedforward path carrying the input signal u, the STF of the system is exactly 1, and hence there is no input signal processed by the integrators, which reduces distortion and 2) a delayed version of the quantization noise of the first stage appears directly at the output of the second integrator, and therefore there is no need for a subtraction (as used in Fig. 2) to feed the noise to the second stage [14].

Assuming a unit-valued power spectral density for the second-stage quantization noise q_2, the total noise power in the output of the modulator is given by $\int_0^{f_s/(2\mathrm{OSR})} |\mathrm{num}(H_{\mathrm{dec}})|^2\, df$. Evaluation of this integral for the original and optimized cases, with $\alpha = 0$ and $\alpha = \alpha_{\mathrm{opt}}$, respectively, shows that the optimized system achieves for e.g., OSR$/N = 2$ a 3.8 dB better noise suppression than the original RSR one [Fig. 3(a)]. This more than compensates for the 3-dB SNR loss introduced by the RSR structure when $N = 2$. Note that the optimized second-order modulator cannot eliminate the added noise totally for higher decimation ratios (i.e., when $N > 2$), but it works well up to an OSR $= 8$. For third-order modulators, the SNR is improved more significantly as discussed below.

The NTF of the structure introduced here has a maximum gain of $2+2\cos\alpha \le 4$ at $f = f_s/2$. According to Lee's rule [15](also verified by simulations), for typical values of α, and using a one-bit quantizer, such NTF leads to an unstable modulator. To achieve stability, one can increase the resolution of the quantizer in the loop—this was done in the simulations of Section V. Another stabilizing technique is to shift poles in the NTF so as to reduce the maximum out-of-band gain. However, since the design of the reduced-sample-rate architecture is based on the cancellation of the complete NTF$_D$ by the denominator of $H_{\mathrm{dec}}(z)$, this method cannot be used in this case.

IV. Higher Order Structures

As discussed in Section I, using higher order structures with low OSR is not very efficient (cf. Table I). However, in some applications, one still needs to use third-order or higher NTFs. Section IV-A extends the previously given solution for third-order systems, while Section IV-B discusses the general higher order case.

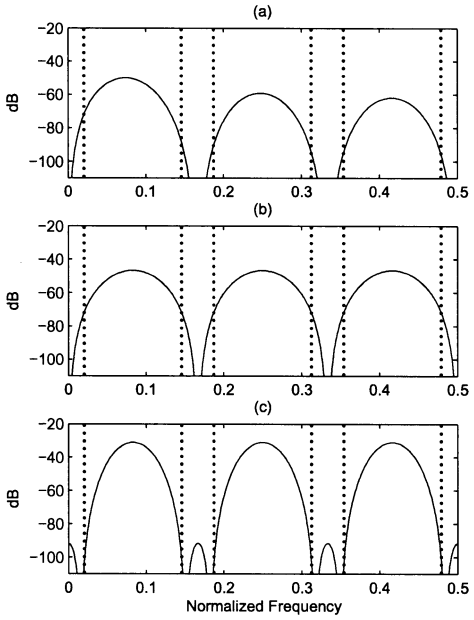


Fig. 5. The NTF of different modulators (input: $Q_2(z)$, output: after the first decimation stage, $M = 3$, OSR $= 24$, $N = 6$). The dotted lines show the boundary of the signal- and aliasing bands. (a) Full-speed structure with a decimation filter order $M' = M + 1 = 4$. (b) RSR structure ($M' = M = 3$). (c) Optimized RSR structure ($M' = M = 3$, optimized zero-placement).

A. Optimization of the NTF of a Third-Order Modulator

In the third-order case, if the NTF of the original system is in the form given in (2), the optimization problem becomes

$$\alpha_{\text{opt}} = \arg\min_{\alpha} \int_0^{f_s/2\text{OSR}} \left|(1 - z^{-N}) \times (1 - 2(\cos N\alpha)z^{-N} + z^{-2N})\right|^2 df \quad (9)$$

where $z = e^{j2\pi f/f_s}$. As shown in the Appendix, the solution of this problem can also be found in closed form, giving the optimized α as

$$\alpha_{\text{opt}} = \frac{1}{N}\cos^{-1}\frac{2\frac{\pi N}{\text{OSR}} - 4\sin\frac{\pi N}{\text{OSR}} + \sin 2\frac{\pi N}{\text{OSR}}}{4\left(-\frac{\pi N}{\text{OSR}} + \sin\frac{\pi N}{\text{OSR}}\right)}. \quad (10)$$

Again, it can be shown using Taylor series expansion (see the Appendix) that the solution converges to $\sqrt{3/5}$, the solution given by Schreier in [3], assuming $\text{OSR}/N \gg 1$.

To verify the results discussed above, Fig. 3 shows the in-band noise power in decibels relative to the nonoptimized one in a second- and third-order system. In the third-order case, one can achieve more than 8 dB improvement before decimation. Hence, the optimized structure can give a better SNR than even the original full-speed architecture, for first-stage decimation ratios up to $N = 6$.

To illustrate the effect of the optimization, Fig. 5 shows the total NTF from the second-stage quantizer $Q_2(z)$ to the output of the first decimation stage. The frequency is normalized to the original sampling rate, and OSR $= 24$, $N = 6$ are used. Notice that due to the fact that $M' = M$ is used in the RSR architecture, the peaks of the NTF do not decrease with the frequency [Fig. 5(b)] as for the original structure [Fig. 5(a)], and the signal band values are similar. However, using the optimization described above, the suppression of the noise is significantly improved in the signal band and in the aliasing bands [Fig. 5(c)].

B. Higher Order Systems

For the design of fourth- and higher order systems, which requires at least a two-dimensional minimization, the general closed-form solution becomes very complicated. Assuming $\text{OSR}/N \gg 1$, the formulation becomes much simpler, allowing closed-form solutions up to fifth-order modulators [3] and numerical solutions for higher orders.

A general "cook-book" design procedure for the optimized system consists of the following steps.

- Find M, the order of the first stage, and the resolution of the second stage.
- Optimize the coefficients of the NTF: use the exact formulas for $M < 4$, or the approximate method for higher orders.
- Stabilize the loop by using multibit feedback if necessary (and use digital calibration of the DAC errors as described, e.g., in [7]).
- Design the first stage of the modulator for the given NTF (e.g., with the proposed second-order solution or with the help of [16]).

V. Design Examples and Practical Considerations

A. An Ideal Second-Order System

To verify the proposed technique, first an ideal second-order two-stage converter with an OSR of 4 has been designed and simulated using MATLAB. The system is illustrated in Fig. 6. Although the architecture of Fig. 4(b) contains two delays in the forward path of the analog modulator, in this example the architecture of Fig. 4(c) with a 4-b quantizer is used, as this contains less coefficients and thus it is less sensitive to mismatch problems. The second stage of the example contains a reduced-speed 10-b quantizer. Thanks to the RSR configuration, the second-stage DAC can be operated at one half of the oversampling clock frequency.

The output spectra after the first decimation stage are shown in Fig. 7. The RSR structure (Fig. 7(b), $\text{SNR} = 98.61$ dB) has 3 dB lower SNR than the original full-speed structure (Fig. 7(a), $\text{SNR} = 101.62$ dB). Using the proposed method (Fig. 7(c), $\text{SNR} = 101.71$ dB), the SNR is actually slightly improved compared to the original structure, while still allowing the half-speed operation of the bulk of the analog circuitry!

B. Nonideal Third-Order System

As the MASH structure is based on the cancellation of the first-stage quantization noise, mismatch errors can limit the performance of the whole system. The main error sources are found in the analog integrators, namely the finite op-amp gain and the capacitor mismatch [17], [18].

A nonideal third-order ADC using the cascade-of-resonators feed-forward (CRFF) architecture [4, Sec.5.6.4], was simulated to verify the effectiveness of the proposed method in the presence of nonideal effects. The analyzed structure used a

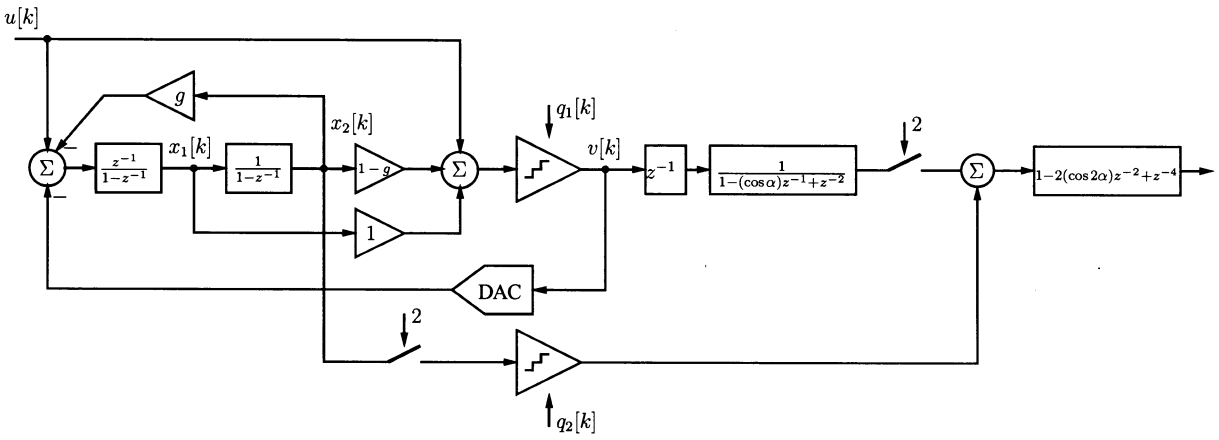


Fig. 6. The simulated ideal second-order system (OSR $= 4$, $N = 2$, $\alpha = 0.44$, $g = 0.19$, $1 - g = 0.81$). The resolution of the first and second quantizer is 4 and 10 b, respectively.

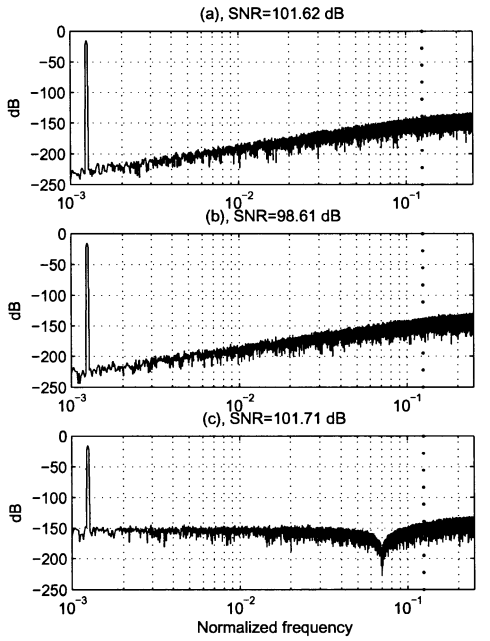


Fig. 7. The output spectra of an ideal second-order modulator. The dotted line shows the boundary of the signal band. (a) Full-speed second stage, third-order decimator, SNR $= 101.62$ dB. (b) RSR architecture, second-order decimation, SNR $= 98.61$ dB. (c) Proposed architecture, second-order decimation, SNR $= 101.71$ dB.

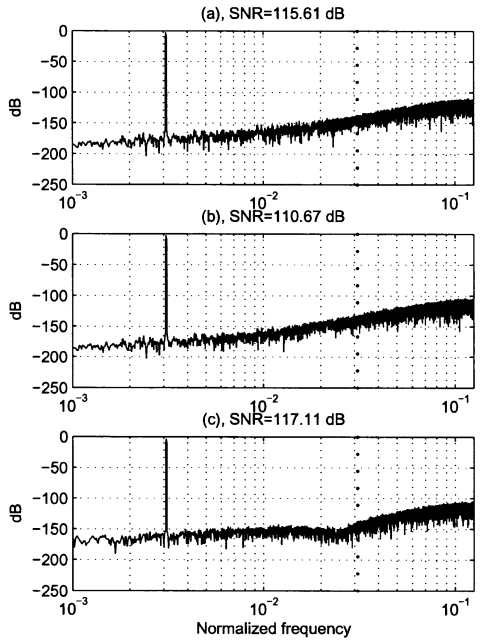


Fig. 8. The output spectra of a nonideal third-order modulator. The dotted line shows the boundary of the signal band. (a) Full-speed second stage. (b) RSR architecture. (c) Proposed architecture.

third-order 5-b first stage and a 5-b pipeline second stage. An OSR of 16 and a decimation ratio $N = 4$ were assumed. A 0.1% mismatch error was used, and a 40-dB op-amp dc gain was simulated with correlated double sampling [19]. The simulated system incorporated data-weighted averaging [20] in the first-stage DAC, which was also assumed to be calibrated to 12-b linearity with one of the available correction methods [7], [21], [22].

Fig. 8 compares the simulated spectra after the first decimation stage for systems with full-speed second stage, reduced-rate suboptimal second stage [2], and the proposed scheme. Although the nonidealities limit the SNR performance, the proposed architecture still delivers the best performance.

C. First-Stage Design Issues

Since the proposed architecture requires a modification of the first stage to spread the zeros of its NTF across the band of interest, the resulting added complexity and sensitivity are of great interest to the designer. Hence, these issues are briefly discussed next for a second-order system.

As Fig. 4 shows, compared to the original first stage, the modified one [Fig. 4(c)] requires a new feedback branch (g) and the modification of the gain of a feed-forward path from 1 to $1 - g$. (Similarly, for a third-order stage, one new and two modified branches are needed.) The new NTF can then be found readily in the parametric form

$$\mathrm{NTF}_{\mathrm{par}} = \frac{1 + (-2 + g)z^{-1} + z^{-2}}{1 + (a_1 + a_2 - 2 + g)z^{-1} + (1 - a_1)z^{-2}}. \quad (11)$$

To match this with the required NTF

$$\mathrm{NTF}_{\mathrm{req}} = 1 - 2(\cos\alpha)z^{-1} + z^{-2} \tag{12}$$

the parameter values $a_1 = 1$, $a_2 = 1 - g$, and $g = 2 - 2\cos\alpha$ must be chosen.

If g is inaccurate, there will be two effects. First, the NTF zero will shift along the unit circle. This will be unimportant since (as Fig. 3 illustrates) the SNR is insensitive to the exact location of this zero. Secondly, as both the zero and the pole of first stage's analog NTF will shift slightly away from the ideal one, the mismatch between the analog and the digital NTF will cause a leakage error. It can be shown analytically that the leakage error at low frequencies is in the same order as the mismatch error (e.g., 0.1% mismatch error would cause about 60 dB suppression of the first-stage quantization error). Using multibit first-stage or mismatch compensation algorithms, this effect can be eliminated.

The value of g is determined by N and the OSR. The higher the OSR, the smaller g will be. For OSR $= 4$ and $N = 2$, the result is $g = 0.19$. For increased N and OSR, $\alpha \sim 1/N$ and $g \sim \alpha^2 \sim 1/N^2$, so $g \ll 1$ will result. This makes the implementation less accurate. However, the proposed design method is aimed specifically at low-OSR ADCs, and for such systems (as discussed above) the effect on the SNR is quite small.

In summary, neither the complexity nor the sensitivity of the analog first stage are significantly increased by the new design method.

VI. Conclusion

The paper showed that combining the use of the RSR architecture introduced in [2] with an optimization of the noise transfer function (NTF) and of the transfer function of the decimation filter, one can make up for the SNR loss introduced by the RSR structure. In higher order cases, the SNR can actually increase over that of a full-sample-rate nonoptimized modulator.

In addition, it was shown that the optimization problem solved here is a generalization of the optimal zero placement for the NTF of a one-stage modulator, described earlier by Schreier [3]. Closed-form equations for the optimal zero placement have been derived for second- and third-order modulators and compared with the previously reported approximations.

The price paid for the improved NTF is a more complex design process for the first stage of the MASH structure and for both stages of the decimation filter. However, if the decimation ratio of the first stage of the decimation filter is low (which is the case for a high-bandwidth, low-OSR converter), then the added coefficients in the first stage are of the same order as the other coefficients. In addition, efficient techniques to realize the decimation filter with the optimized coefficients were introduced in [23] and remain applicable in the modified system described here.

The sensitivity of the modified RSR to circuit nonidealities was also examined and it is close to that of the original structure described in [2].

Appendix I
Derivation of the Optimal α for a Second-Order System

To minimize the in-band noise power, the following optimization problem needs to be solved:

$$\begin{aligned}\alpha_{\mathrm{opt}} &= \arg\min_{\alpha} \int_0^{f_s/2\mathrm{OSR}} \left|\mathrm{num}(H_{\mathrm{dec}}(z))\right|^2 df \\ &= \arg\min_{\alpha} \int_0^{f_s/2\mathrm{OSR}} \left|1 - 2(\cos N\alpha)z^{-N} + z^{-2N}\right|^2 \\ &\quad \times df. \end{aligned} \tag{13}$$

Let $\vartheta = 2\pi f/f_s$ and $z = e^{j\vartheta}$. The square of the absolute value of $H(z)$ can be written as

$$\left|1 - 2(\cos N\alpha)z^{-N} + z^{-2N}\right|^2 = (\cos N\vartheta - \cos N\alpha)^2. \tag{14}$$

To find the minimum, the first derivative of $|H(z)|^2$ given in (13) with respect to α needs to be equated to zero. Introducing the new variable ϑ and changing the limits of the integral gives

$$\frac{d}{d\alpha}\int_0^{\pi/\mathrm{OSR}} (\cos N\vartheta - \cos N\alpha)^2\, d\vartheta = 0 \tag{15}$$

$$\int_0^{\pi/\mathrm{OSR}} 2\,(\cos N\vartheta - \cos N\alpha)\,(\sin N\alpha) N d\vartheta = 0 \tag{16}$$

$$2N(\sin N\alpha)\int_0^{\pi/\mathrm{OSR}} (\cos N\alpha - \cos N\vartheta)\, d\vartheta = 0 \tag{17}$$

$$2(\sin N\alpha)\left(\sin\frac{\pi N}{\mathrm{OSR}} - (\cos N\alpha)\frac{\pi N}{\mathrm{OSR}}\right) = 0. \tag{18}$$

The above equation gives two solutions for α. However, $\sin(N\alpha) = 0$, i.e., $\alpha = k2\pi/N,\ k \in \mathbb{Z}$ gives a local maximum due to the symmetric behavior of the function. Thus, the desired minimum occurs for

$$\sin\frac{\pi N}{\mathrm{OSR}} - (\cos N\alpha)\frac{\pi N}{\mathrm{OSR}} = 0 \tag{19}$$

i.e., for

$$\alpha = \frac{1}{N}\cos^{-1}\frac{\sin\frac{\pi N}{\mathrm{OSR}}}{\frac{\pi N}{\mathrm{OSR}}}. \tag{20}$$

Appendix II
Derivation of the Optimal α for a Third-Order System

In the third-order case, the function to be minimized is

$$\begin{aligned}\alpha_{opt} = \arg\min_{\alpha} \int_0^{f_s/2\mathrm{OSR}} &\left|(1 - z^{-N})\right. \\ &\left.\times\,(1 - 2(\cos N\alpha)z^{-N} + z^{-2N})\right|^2 df. \end{aligned} \tag{21}$$

As in the previous discussion, substituting $\vartheta = 2\pi f/f_s$, $z = e^{j\vartheta}$ gives

$$\frac{d}{d\alpha}\int_0^{\pi/\mathrm{OSR}} \left(4\left(\sin\frac{N\vartheta}{2}\right)(\cos N\vartheta) - (\cos N\alpha)\right)^2 d\vartheta = 0 \tag{22}$$

$$N(\sin N\alpha)\int_0^{\pi/\mathrm{OSR}} \left(\sin\frac{N\vartheta}{2}\right)^2 (\cos N\vartheta - \cos N\alpha)\, d\vartheta = 0. \tag{23}$$

Since $\sin N\alpha = 0$ represents a local maximum, we can simplify and rearrange (23) to obtain

$$\int_0^{\pi/\mathrm{OSR}} \left(\sin\frac{N\vartheta}{2}\right)^2 (\cos N\vartheta)\, d\vartheta = (\cos N\alpha)\int_0^{\pi/\mathrm{OSR}} \left(\sin\frac{N\vartheta}{2}\right)^2 d\vartheta \tag{24}$$

$$\cos N\alpha = \frac{\int_0^{\pi/\mathrm{OSR}} 2\left(\sin\frac{N\vartheta}{2}\right)^2 (\cos N\vartheta)\, d\vartheta}{\int_0^{\pi/\mathrm{OSR}} 2\left(\sin\frac{N\vartheta}{2}\right)^2 d\vartheta}. \tag{25}$$

Using $2(\sin N\vartheta/2)^2 = 1 - \cos N\vartheta$ and $2(\cos N\vartheta)^2 = 1 + \cos 2N\vartheta$,

$$\cos N\alpha = \frac{2\frac{\pi N}{\mathrm{OSR}} - 4\sin\frac{\pi N}{\mathrm{OSR}} + \sin 2\frac{\pi N}{\mathrm{OSR}}}{4\left(-\frac{\pi N}{\mathrm{OSR}} + \sin\frac{\pi N}{\mathrm{OSR}}\right)} \tag{26}$$

results. Hence, the optimum occurs for

$$\alpha = \frac{1}{N}\cos^{-1}\frac{2\frac{\pi N}{\mathrm{OSR}} - 4\sin\frac{\pi N}{\mathrm{OSR}} + \sin 2\frac{\pi N}{\mathrm{OSR}}}{4\left(-\frac{\pi N}{\mathrm{OSR}} + \sin\frac{\pi N}{\mathrm{OSR}}\right)}. \tag{27}$$

Appendix III
Proof of the Approximate Solution for a Third-Order System

Assuming $N = 1$ and $\mathrm{OSR} \gg 1$, we can use Taylor-series expansion around $x = \pi N/\mathrm{OSR} = 0$. This gives

$$\frac{\alpha_{\mathrm{opt}}}{\omega_B} = \frac{\alpha_{opt}}{\frac{2\pi}{2\mathrm{OSR}}} \tag{28}$$

$$= \frac{1}{x}\cos^{-1}\left.\frac{2x - 4\sin x + \sin 2x}{-4x + 4\sin x}\right|_{x\ll 1} \tag{29}$$

$$\approx \frac{1}{x}\cos^{-1}\left(1 - \frac{3}{10}x^2\right) \approx \frac{1}{x}\sqrt{\frac{3}{5}}x = \sqrt{\frac{3}{5}}. \tag{30}$$

Acknowledgment

The authors are grateful to Prof. I. Kollár for his useful comments and support and J. Silva for the useful discussions and some of his models.

References

[1] T. C. Leslie and B. Singh, "An improved sigma-delta modulator architecture," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'90)*, vol. 1, New Orleans, LA, May 1–3, 1990, pp. 372–375.

[2] W. Qin, B. Hu, and X. Ling, "Sigma-delta ADC with reduced sample rate multibit quantizer," *IEEE Trans. Circuits Syst. II*, vol. 46, pp. 824–828, June 1999.

[3] R. Schreier, "An empirical study of high-order single-bit delta-sigma modulators," *IEEE Trans. Circuits Syst. II*, vol. 40, pp. 461–466, Aug. 1993.

[4] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta-Sigma Data Converters*. Piscataway, NJ: IEEE Press, 1997.

[5] B. P. Brandt, "Oversampled analog-to-digital conversion," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1991.

[6] Y. Geerts, M. S. J. Steyaert, and W. Sansen, "A high performance multibit $\Delta\Sigma$ CMOS ADC," *IEEE J. Solid-State Circuits*, vol. 35, pp. 1829–40, Dec. 2000.

[7] X. Wang, U. Moon, G. C. Temes, and P. Kiss, "Digital correlation technique for the estimation and correction of DAC errors in multibit MASH $\Delta\Sigma$ ADC's," in *Proc. 4th Int. Conf. Advanced A/D and D/A Conversion Techniques and Their Applications; 7th Eur. Workshop on ADC Modeling and Testing (ADDA-EWADC 2002)*, Prague, Czech Republic, June 26–28, 2002, pp. 39–42.

[8] X. Wang, U. Moon, M. Liu, and G. C. Temes, "Digital correlation technique for the estimation and correction of DAC errors in multibit MASH delta-sigma ADC's," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2002)*, vol. IV, Scottsdale, AZ, May 26–29, 2002, pp. 691–694.

[9] L. Lo Presti, "Efficient modified-sinc filters for sigma-delta A/D converters," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 1204–1213, Nov. 2000.

[10] J. Márkus and G. C. Temes, "An efficient delta-sigma noise-shaping architecture for wideband applications," in *Proc. 4th Int. Conf. Advanced A/D and D/A Conversion Techniques and their Applications; 7th Eur. Workshop on ADC Modeling and Testing (ADDA-EWADC 2002)*, Prague, Czech Republic, June 26–28, 2002, pp. 35–38.

[11] E. B. Hogenauer, "An economical class of digital filters for decimation and interpolation," *IEEE Trans. Acoust., Speech Signal Processing*, vol. ASSP-29, pp. 155–162, Apr. 1981.

[12] J. C. Candy, "Decimation for sigma delta modulation," *IEEE Trans. Commun.*, vol. 34, pp. 72–76, Jan. 1986.

[13] J. Robert and P. Deval, "A second-order high-resolution incremental A/D converter with offset and charge injection compensation," *IEEE J. Solid-State Circuits*, vol. 23, pp. 736–741, June 1988.

[14] J. Silva, U. Moon, J. Steensgaard, and G. C. Temes, "A wideband low-distortion delta-sigma ADC topology," *Electron. Lett.*, vol. 37, no. 12, pp. 737–738, June 2001.

[15] W. L. Lee and C. G. Sodini, "A topology for higher order interpolative coders," in *Proc. IEEE Int. Symp. Circuits and Systems*, Philadelphia, PA, May 4–7, 1987, pp. 459–462.

[16] R. Schreier. (2003) The Delta-Sigma Toolbox v6.0 (Delsig). [Online]. Available: http://www.mathworks.com/matlabcentral/fileexchange/

[17] G. C. Temes, "Finite amplifier gain and bandwidth effects in switched-capacitor filters," *IEEE J. Solid-State Circuits*, vol. 15, pp. 358–61, June 1980.

[18] G. Cauwenberghs and G. C. Temes, "Adaptive digital correction of analog errors in MASH ADC's-part I. Off-line and blind on-line calibration," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 621–28, July 2000.

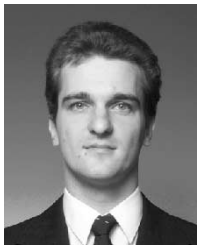
[19] C. C. Enz and G. C. Temes, "Circuit techniques for reducing the effects of op-amp imperfections: Autozeroing, correlated double sampling, and chopper stabilization," *Proc. IEEE*, vol. 84, pp. 1584–1614, Nov. 1996.

[20] R. T. Baird and T. S. Fiez, "Linearity enhancement of multibit delta-sigma A/D and D/A converters using data weighted averaging," *IEEE Trans. Circuits Syst. II*, vol. 42, pp. 753–762, Dec. 1995.

[21] U. Moon, J. Silva, J. Steensgaard, and G. C. Temes, "A switched-capacitor DAC with analog mismatch correction," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2000)*, Geneva, Switzerland, May 28–31, 2000, pp. 421–424.

[22] P. Kiss, U. Moon, J. Steensgaard, J. T. Stonick, and G. C. Temes, "Multibit delta-sigma ADC with mixed-mode DAC error correction," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS'2001)*, Sydney, Australia, May 6–9, 2001, pp. 280–283.

[23] L. Lo Presti and A. Akhdar, "Efficient antialiasing decimation filter for $\Delta\Sigma$ converters," in *Proc. 5th IEEE Int. Conf. Electronics, Circuits, and Systems (ICESC '98)*, vol. 1, Lisboa, Portugal, Sept. 7-10, 1998, pp. 367–370.



János Márkus (S'02) received the M.S. degree in electrical engineering from the Budapest University of Technology and Economics (BUTE), Budapest, Hungary, in 1999. He is currently working toward the Ph.D. degree at the same university.

In 2000, he worked also as a Design Engineer at Panda Audio Ltd. (a subsidiary of Akai Professional). In 2001 and 2002, he was an exchange student at Oregon State University, Corvallis, in the Department of Electrical and Computer Engineering. His current research interests include quantization errors in digital signal processing systems, analog-to-digital and digital-to-analog converter testing and Delta-Sigma converter structures.

Gabor C. Temes (SM'66–F'73–LF'98) received his undergraduate education at the Technical University and Eötvös University in Budapest, Hungary, from 1948 to 1956, and the Ph.D. degree in electrical engineering from the University of Ottawa, Ottawa, ON, Canada, in 1961. He received an honorary doctorate from the Technical University of Budapest, Budapest, Hungary, in 1991.

He held academic positions at the Technical University of Budapest, Stanford University, Stanford, CA, and the University of California at Los Angeles (UCLA), and worked in industry at Northern Electric R&D Laboratories (now Bell-Northern Research), as well as at Ampex Corporation. He is now a Professor in the Department of Electrical and Computer Engineering at Oregon State University (OSU), Corvallis. He served as Department Head at both UCLA and OSU. He is coeditor and coauthor of *Modern Filter Theory and Design* (New York: Wiley, 1973), coauthor of *Introduction to Circuit Synthesis and Design* (New York: McGraw-Hill, 1977), coauthor of *Analog MOS Integrated Circuits for Signal Processing* (New York: Wiley, 1986), and coeditor and coauthor of *Oversampling Delta-Sigma Data Converters* (New York: IEEE Press, 1992) and *Delta-Sigma Data Converters* (Piscataway, NJ: IEEE Press, 1997), as well as a contributor to several other edited volumes. He has published approximately 300 papers in engineering journals and conference proceedings. His recent research has dealt with CMOS analog integrated circuits, as well as data converters and integrated sensor interfaces.

Dr. Temes is a Life Fellow of the IEEE. He was Associate Editor of the J*ournal of the Franklin Institute*, Editor of the IEEE TRANSACTIONS ON CIRCUIT THEORY and Vice President of the IEEE Circuits and Systems (CAS) Society. In 1968 and in 1981, he was co-winner of the CAS Darlington Award, and in 1984 winner of the Centennial Medal of the IEEE. He received the Andrew Chi Prize Award of the IEEE Instrumentation and Measurement Society in 1985, the Education Award of the IEEE Circuits and Systems Society in 1987, and the Technical Achievement Award of the same Society in 1989. He received the IEEE Graduate Teaching Award in 1998, and the IEEE Millennium Medal as well as the IEEE CAS Golden Jubilee Medal in 2000.