REDUCING QUANTIZATION NOISE WITH RECURSIVE ΣΔ MODULATORS

Daniël Schinkel, Ed van Tuijl and Anne-Johan Annema

University of Twente, IC-Design Group, P.O.Box 217, 7500 AE Enschede, The Netherlands

ABSTRACT

This paper introduces a recursive multibit ΣΔ architecture that enables a high effective quantizer resolution while needing only a limited number of DAC elements. The recursive architecture consists of a set of ΣΔ modulators, whereby each stage cancels the quantization noise of the preceding stage. Conventional DEM algorithms can be used in each stage to reduce the sensitivity to mismatch. The architecture enables a significant reduction of both the signal-band and out-of-band quantization noise power, compared to conventional multibit ΣΔ converters.

1. INTRODUCTION

Digital to analog converters (DAC's) for high-resolution, low frequency applications such as high-quality audio have seen a significant evolution over the past few decades. Traditionally, near-Nyquist converters were used with fine quantization steps. These DACs usually consist of binary weighted DAC elements driven by a PCM code, and require a complex power consuming post-DAC anti-aliasing filter. This type of DAC suffers from linearity problems due to element mismatch. Oversampled ΣΔ converters, using an inherently linear 1-bit DAC, were subsequently introduced. However the high amount of quantization noise generated by a 1-bit ΣΔ DAC poses other problems, such as increased clock-jitter sensitivity and possible noise folding. Also here, a complex post-DAC filter is required, now to attenuate the out-of-band quantization noise to acceptable levels.

During the last decade multibit ΣΔ converters have become popular [1], [2]. Current multibit ΣΔ converters use a number of equally weighted DAC elements. Compared to 1-bit converters, the advantages of the multibit converters include an increase in modulator stability and performance [3], a reduction of the required oversampling ratio and a reduction of the out-of-band quantization noise, thereby relaxing the requirements of the post-DAC filter. The disadvantage is the DAC non-linearity due to mismatch. However, the redundancy that is present in a multibit ΣΔ DAC (many different code combinations lead to the same quantization level) can be exploited by dynamic element matching (DEM) algorithms to efficiently reduce the influence of mismatch on the linearity of the DAC [4]-[6].

Conventional multibit ΣΔ modulators with equally weighted DAC elements typically use a very moderate quantizer resolution because the number of DAC elements scales exponentially with the bit-resolution.

This paper introduces a multibit ΣΔ architecture that enables a high quantizer resolution while requiring only a moderate number of DAC elements. Compared to conventional multibit ΣΔ converters, the proposed DAC yields both a higher SNR and lower out-of-band noise, thereby relaxing the requirement for the post-DAC filter.

The proposed architecture exploits the fact that it is more efficient, compared to a conventional single-stage ΣΔ modulator, to use e.g two modulator stages, and use the second modulator to cancel the quantization noise of the first modulator. This second modulator can use smaller element weights and a smaller range. The stage-addition process can be repeated, yielding a recursive ΣΔ modulator. The advantages of less quantization noise, such as an increased dynamic range and relaxed post-filter requirements can easily outweigh the increased hardware requirements, especially because digital hardware can become increasingly compact as CMOS technologies advances.

The proposed architecture has some resemblance to the well-known MASH- or cascaded-architectures [7]. It is however a different architecture. The outputs of the different stages in a MASH structure need to be filtered prior to their addition and the main objective is to increase the noise-shaping order. With the proposed architecture, all modulator outputs are summed directly, with a more efficient reduction of the total quantization noise power.

The following section discusses the architecture of the proposed multi-stage converter. A concrete example with simulation results is given in section 3 and the conclusions are presented in section 4.

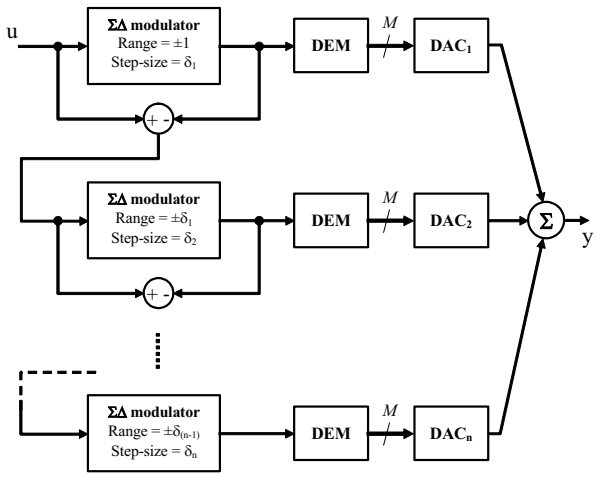

Fig. 1: recursive multi-stage ΣΔ converter architecture.

2. RECURSIVE MULTIBIT ΣΔ MODULATORS

Fig. 1 shows the architecture of the proposed multi-stage ΣΔ converter. The input signal of each stage is equal to the (shaped) quantization-noise of the preceding stage, except for the first stage that operates on the input signal. If the input to output transfer (signal transfer) of each modulator is unity for all frequencies, then the sum of all outputs only contains the quantization noise of the bottom modulator. Most conventional ΣΔ modulators only have about unity signal transfer in the frequency region where the loop-gain of the modulator is still well above unity. However, two ΣΔ structures are known that have a signal transfer that is exactly unity over the entire band. These two structures are shown in Fig. 2.

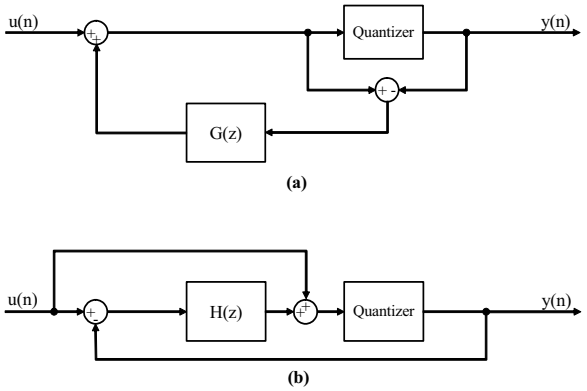

Fig. 2: Two ΣΔ modulators with unity signal transfer.

The topology shown in Fig. 2a is a well-known noise-shaping structure. The topology shown in Fig. 2b is also a standard topology, except for the feed-forward of the input. Control theory can be used to show that both topologies have a unity signal transfer-function (STF), under the assumption that the quantizer can be modeled as a linear, unity gain element with additive uncorrelated quantization noise injection:

$$a)\ STF(z)=\frac{Y(z)}{U(z)}=1$$
$$b)\ STF(z)=\frac{Y(z)}{U(z)}=\big(1+H(z)\big)\frac{1}{1+H(z)}=1 \qquad (1)$$

The noise-transfer function (NTF, the transfer of the quantization noise (E) to the output) for the topologies in Fig. 2a and 2b are respectively:

$$a)\ NTF(z)=\frac{E(z)}{U(z)}=1-G(z)$$
$$b)\ \ NTF(z)=\frac{1}{1+H(z)} \qquad (2)$$

Standard ΣΔ design procedures (e.g. [8]) can be used for the design of the filters G(z) or H(z). First, the desired NTF is chosen. The corresponding loop-filter can then be calculated by taking the inverse of Eq. 2a or 2b. Note that this inverse only leads to implementable (strictly causal) filters if the order of the numerator and denominator polynomial of the NTF are equal, and if the highest order polynomial terms are equal.

In base-band ΣΔ converters, the NTF has a high-pass characteristic. As an implication, the loop-filter H(z) of Fig. 2b is a low-pass filter. The filter H(z) can be implemented with a cascade of single- and second-order filter sections, allowing increasingly smaller word-lengths for sections closer to the filter output (because internal quantization errors are also noise-shaped). The filter G(z) in Fig. 2a has an all-pass nature and does not lend itself for implementations with tapered word-lengths. Therefore, although the two topologies in Fig. 2 are functionally equivalent for equal NTF, the second topology (b) is more efficient and hence preferred.

It is possible to use the same loop-filter for every modulator in the recursive multi-stage architecture that is shown in Fig. 1. Exploiting this property can considerably reduce hardware costs: in a dedicated-hardware implementation, the filter logic can be implemented once and can be used for all stages with data-multiplexing.

Noise-shaping in general comes at a price of increased total (in- and out-of-band) quantization noise power. The quantization noise amplitude at the output of a ΣΔ modulator will at least be as high as the step-size (δ) of its quantizer (while the quantization noise amplitude of a stand-alone quantizer is only equal to half the step-size) [1]. ΣΔ modulators with aggressive noise-shaping properties can produce quantization noise amplitudes higher than their step-size.

To cancel the quantization noise, a subsequent modulator requires a peak-to-peak range (R_i) that is at least twice the quantization step-size of the previous modulator. With aggressive loop-filters, the range will have to be higher, but this reduces the efficiency of the architecture and requires additional scaling logic. Simulations verified that a range R_i equal to $\pm\delta_{i-1}$ is adequate to ensure stable operation with most loop-filters, even if the input occasionally exceeds the modulator range.

With the specified range dependency of $R_i=\pm\delta_{i-1}$ it is possible to relate the total quantization noise of the multi-stage architecture to the number of stages and to the number of DAC-elements per stage. The total quantization noise can be derived from the step-size of the bottom modulator (δ_n), which determines the effective overall quantizer resolution. The intrinsic quantization error of the bottom quantizer (with a power of $\delta_n^2/12$), translates to the output noise by a multiplication with the NTF of the modulator.

The number of 1-bit DAC elements (*M*) in a single-loop multi-level DAC is usually equal to its peak-to-peak output range (*R*), divided by the step size (δ). Assuming that each DAC in the multi-stage architecture uses an equal number of elements, the following relations can be formulated (given a normalized peak-to-peak range of 2 for the first modulator):

$$\delta_i=\frac{R_i}{M}=\frac{2\delta_{(i-1)}}{M}\ \Rightarrow\ \delta_n=\delta_1\left(\frac{2}{M}\right)^{(n-1)}=\left(\frac{2}{M}\right)^n \qquad (3)$$

[1] Consider for example a ΣΔ modulator with a DC input signal that has a value slightly below a quantizer level. The output of the modulator will occasionally alternate between the mentioned quantizer level and one level lower in order to equal the average output of the modulator to its input.

Eq. 3 shows that the step-size can only decrease for subsequent stages if more than two elements are used per stage. Compared to binary-weighted PCM converters, more elements are needed to reach a certain quantizer resolution. These additional elements provide the code-redundancy, necessary for the DEM algorithms.

An indication for the (hardware) efficiency of the recursive architecture is the total number of DAC elements that is required for a certain resolution. Below, it is shown that this efficiency depends on the number of elements per stage and that an optimum can be found.

The bit-resolution of a quantizer equals the binary logarithm of the number of quantization levels. The number of quantization levels equals the range divided by the step-size plus one. For simplicity reasons, this last level can be ignored if the resolution is more than a few bits:

$$bitres = \log_2\left(\frac{2}{\delta_n}+1\right) \approx \log_2\left(\frac{2}{\delta_n}\right) = 1 + n\cdot\left(\log_2(M)-1\right) \qquad (4)$$

Clearly, each additional stage adds $(log_2(M) - 1)$ bits to the effective resolution. The increase of the resolution per DAC-element is thus:

$$\Delta\, bits\ \ per\ element = \frac{\log_2(M)-1}{M} \qquad (5)$$

Fig. 3 gives a graphical representation of Eq. 5. Equation 4 should be used to calculate the bit-resolution, Eq. 5 only predicts the resolution increase for subsequent stages. It is clear from the figure that the optimum number of DAC elements per stage is five or six.

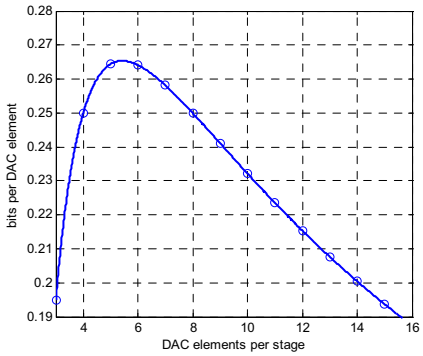

Fig. 3: Resolution per DAC element as a function of the number of elements per stage.

From an implementation point of view, it is more practical to use either four or eight elements per stage. For these cases, the element weights decrease by an integer factor (of respectively a factor of two per stage with four elements per stage and a factor of four with eight elements per stage). Note that in both cases, each DAC element (of the second and subsequent stages) adds exactly one quarter of a bit of effective quantizer resolution.

An architecture with four elements per stage requires twice as many stages as an architecture with eight elements per stage. However, the complexity of DEM algorithms tends to increase more than proportional with the number of DEM elements per stage. Therefore, the optimum number of elements per stage (either 4 or 8) is a trade-off between the DEM complexity and the modulator complexity.

3. SYSTEM EXAMPLES

Simulations with both a multi-stage architecture and a conventional single-stage multibit ΣΔ converter were done to estimate the gain in performance of the proposed architecture.

As a concrete example, 32 DAC-elements are used in both systems. The multi-stage architecture divides the 32 elements over 4 stages, using 8 elements per stage. The effective quantizer resolution of the multi-stage architecture is 9 bits (see Eq. 4), compared to 5 bits of resolution for the conventional multibit converter. For the measurements of the in-band SNR and out-of-band noise power, an oversampling ratio of 64 is chosen. The well-known data-weighted averaging (DWA) algorithm is used for the dynamic element matching [5], resulting in a first-order shaping of the mismatch-induced noise. Simple second-order loop-filters are used for all the ΣΔ modulators:

$$H(z) = \frac{(z-0.5)}{(z-1)^2} \qquad (6)$$

More aggressive noise-shaping with a second-order filter is possible, but that will result in overloading and incomplete noise-cancellation in the stages of the multi-stage architecture.

Higher-order loop-filters could be used to further decrease the in-band quantization noise, provided that they are implemented in combination with a higher-order mismatch-shaping DEM algorithm. For the current system, the mismatch-induced noise in the signal band is already dominant over the quantization noise and a further decrease of the quantization noise has no effect on the performance.

The impact of mismatch is simulated with Monte-Carlo simulations, applying random deviations on the DAC elements before each trial. The relative mismatch between the DAC elements is assumed to be dependent on the square root of the active area occupied by each element, to model real-world random mismatch effects. The active area occupied by each element is assumed to be linearly dependent to the element weights. Both converters are given roughly the same total area for the active parts of the DAC-elements, with a distribution over the various elements such as shown in Fig. 4.

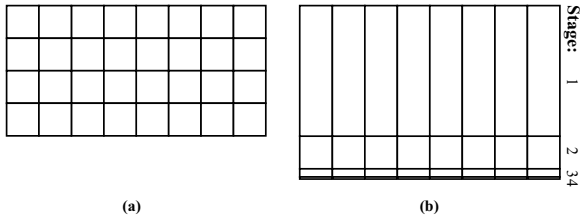

Fig. 4: Area distribution of the DAC-elements for the single-stage (a) and multi-stage (b) converter.

For the single-stage converter (a), a relative element mismatch of 0.2% is assumed (1σ). Consequently, the elements of the four stages of the multi-stage converter (b) will get a relative

mismatch of respectively 0.1%, 0.2%, 0.4% and 0.8% from the first to the last stage.

Fig. 5 and Fig. 6 show the spectra of the single-stage and multi-stage converter for a typical Monte Carlo trial. In both cases an input signal is applied with a normalized frequency of $f/f_s=8.2\cdot10^{-4}$ and an amplitude of 0.4 (-8 dB FS). The mismatch-induced first-order high-pass error signal is clearly visible at low frequencies in the case of the multi-stage converter in Fig. 6. The quantization noise of the single-stage converter in Fig. 5 is higher and is still dominant over the mismatch noise. The spikes that are visible in the spectrum of the multi-stage converter are typical for DWA algorithms, because many tonal limit-cycles are present in the first-order noise-shaped individual elements.

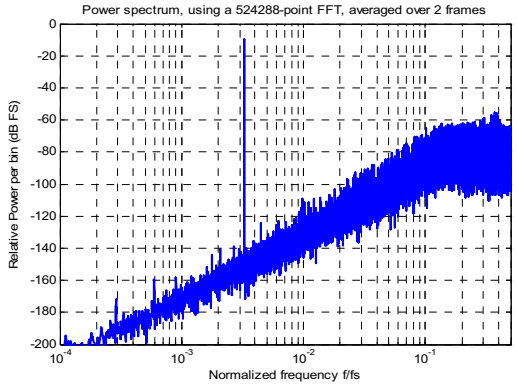

Fig. 5: Simulated spectrum of the single-stage modulator

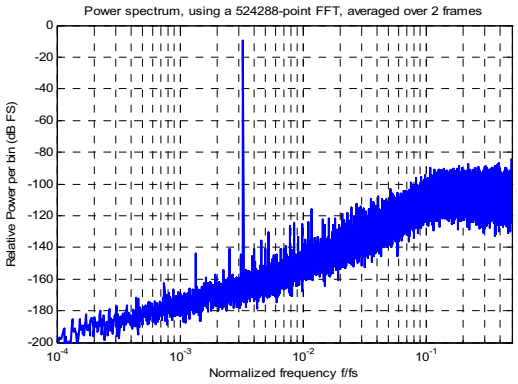

Fig. 6: simulated spectrum of the multi-stage modulator

Some quantitative simulation-results are given in Table 1, using average values from the Monte Carlo simulations. The multi-stage converter outperforms the single-stage variant by roughly 22 dB, except for the SNR that is simulated in the presence of mismatch (The 120 dB of dynamic range that is obtained in the presence of mismatch is limited by the mismatch-induced noise). These results comply with the predicted increase in effective resolution (3.7 bits actual increase, in stead of 4).

In the high-frequency region, the quantization noise is still dominant over the mismatch-induced noise. This means that the out-of-band noise of the multi-stage converter can be further reduced with additional stages, up to the point where the mismatch-induced noise starts to dominate in all frequency regions.

Table 1: Simulation results

	Single-stage ΣΔ converter	Multi-stage ΣΔ converter
usable input range	100%	100%
over-sampling ratio	64	64
SNR (dyn. range)		
ideal	104 dB FS	126 dB FS
with mismatch	104 dB FS	120 dB FS
Out-of-band noise		
ideal	-28 dB FS	-50 dB FS
with mismatch	-28 dB FS	-50 dB FS

4. CONCLUSIONS

This paper introduces a recursive ΣΔ modulator architecture. Each stage in the architecture cancels the quantization noise of the preceding stage, such that the overall quantization noise equals the low quantization noise of the last stage.

The increase in effective quantizer resolution leads to a decrease of the total quantization noise power (both in-band and out-of-band). Less quantization noise implies that a desired SNR can be reached with a lower modulator order and also relaxes the constraints for the post-DAC filter.

Optimal implementations of the proposed architecture need four DAC elements for every added bit of quantizer resolution. In contrast, a single-stage ΣΔ converter requires a doubling of the number of DAC elements for every additional bit of resolution.

The mismatch-induced noise is determined by the DEM algorithm and its power is comparable in the conventional and in the proposed recursive ΣΔ modulator. In the 32 element example, the quantization noise of the recursive architecture is 22 dB lower than that in the conventional architecture.

5. REFERENCES

[1] I. Fujimori, A. Nogi and T. Sugimoto, "A Multibit Delta-Sigma Audio DAC with 120-dB Dynamic Range," IEEE J. Solid-State Circuits, vol. 35, pp. 1066 - 1073, Aug. 2000.

[2] R. Adams, K. Nguyen and Karl Sweetland, "A 113dB SNR oversampling DAC with segmented noise-shaped scrambling," IEEE JSSC, vol. 33, pp. 1871-1878, Dec. 1998.

[3] R.T. Baird and T.S. Fiez, 'Stability analysis of high-order delta-sigma modulation for ADC's', IEEE Tr. Circuit Syst. II, vol. 41, pp. 59 - 62, Jan. 1994.

[4] R. Schreier and B. Zhang, "Noise-shaped multi-bit D/A converter employing unit elements," Electron. Lett., vol. 31, pp. 1712-1713. Sept. 1995.

[5] R.T. Baird and T.S. Fiez, "Linearity enhancement of multibit ΔΣ A/D and D/A converters using data weighted averaging," IEEE Tr. Circuit Syst. II, vol. 42, pp. 753 - 762, Dec. 1995.

[6] J. Welz et al., "Simplified logic for first-order and second-order mismatch-shaping digital-to-analog converters," IEEE Tr. Circuit Syst. II, vol. 48, pp. 1014 - 1027, Nov. 2001.

[7] Y. Matsuya et al., "A 17-bit oversampling D-to-A conversion technology using multistage noise-shaping," IEEE J. Solid-State Circuits, vol. 24, pp. 969-975, Aug. 1989.

[8] B. Adams et al., "Theory and Practical Implementation of a Fifth-Order Sigma-Delta A/D Converter," J. of the AES, Vol. 39, No. 7/8, page 515-528, 1991.