Dual Quantization Continuous Time ΣΔ Modulators with Spectrally Shaped Feedback

Hossein Pakniat and Mohammad Yavari
Integrated Circuits Design Laboratory, Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran
E-mails: hoseinpakniat@aut.ac.ir, myavari@aut.ac.ir

***Abstract*— A new technique is proposed to eliminate the nonlinear error of a multibit digital-to-analog converter (DAC) used in a sigma-delta modulator where the solidity of multibit DAC against the clock jitter is reserved. By using dual quantization, this technique reduces the bit length of DAC to one bit and consequently eliminates its nonlinear error. A low-pass finite impulse response (FIR) filter with pseudo digital implementation is merged with this 1-bit DAC to improve its solidity against the clock jitter up to the conventional multibit DACs.**

I. INTRODUCTION

Daily increasing demand for bandwidth of analog to digital converters (ADCs) poses the continuous time ΣΔ modulators (CT-ΣΔMs) as a suitable candidate for high resolution and high speed applications [1]. Comparing to discrete time ΣΔ modulators (DT-ΣΔMs), the CT-ΣΔMs achieve wider bandwidth (with same power consumption) and have an inherent anti-alias filtering property. By the way, CT-ΣΔMs are more sensitive to the clock jitter compared to the DT-ΣΔMs. By using multibit quantization in CT-ΣΔMs, the sensitivity to clock jitter is significantly reduced, both the stability and accuracy are enhanced and the design specifications of opamps is relaxed [2]. Nonetheless, the elements of multibit DAC used in the feedback path have inherent mismatch which causes a nonlinear error. This error at outermost feedback path, where there is no shaping, limits the total accuracy of the modulator.

Shaping techniques are the most common solution for reduction of nonlinearity errors of a multi-bit DAC [3], but the latency and complexity of these techniques exponentially increase with the number of quantizer bits (N) and their efficiency decreases with the reduction of oversampling ratio (OSR). Calibration techniques are the other candidates to overcome the nonlinearity errors of multi-bit DACs [4]. These techniques usually impose significant circuit overhead to the modulator.

Dual quantization is another technique which reduces the number of bits of the outermost feedback DAC [5]. A digital quantizer truncates the lower significant bits of outermost feedback path, the most sensitive path, and lets the most significant bit to pass. Therefore, the 1-bit DAC of this path ensures the linearity. The truncation error of digital quantizer is injected to the second integrator, which benefits from the shaping of the first integrator, via a digital filter to compensate the truncation error of digital quantizer [5]. Hence, the dual quantization technique causes that the number of bits of the outermost feedback path to be reduced to one bit. Nonetheless, a 1-bit DAC is more sensitive to the clock jitter than a multibit DAC. This seems a restriction for using the dual quantization in a CT-ΣΔM. On the other hand, the spectrally shaping feedback scheme increases the effective number of bits of a 1-bit DAC without destroying its linearity, by using a low-pass FIR-filter in the feedback path of the modulator [6].

In this paper the idea of dual quantization is extended to CT-ΣΔMs to overcome the nonlinear errors of multi-bit DAC. Also, the idea of spectrally shaping feedback is combined with dual quantization to improve the clock jitter insensitivity of its 1-bit DAC up to the ideal multibit DACs. A straight forward design flow of modulator is also presented. The proposed technique eliminates the nonlinear errors of multibit DAC where its solidity against the clock jitter is simultaneously reserved.

II. DUAL QUANTIZATION WITH DΣΔM

In a dual quantization technique an additional quantizer is used in feedback path of the ΣΔM. This digital quantizer passes the most significant bit (MSB) and truncates the other bits of digital signal of feedback. A delay-free digital sigma-delta modulator (DΣΔM) depicted in Fig. 1, is used to shape the quantization error. This DΣΔM has no delay in forward path and its signal transfer function (STF) is equal to one, therefore it does not add any delay to the signal unless the delay added by the sum operation. The linear model of this modulator shows that the output signal of the modulator is equal to:

$$Y_Q(z)=Y_q(z)+E_q(z)(1-S(z)) \qquad (1)$$

where $E_q(z)$ is the quantization noise of the digital quantizer. Choosing $S(z) = z^{-1}$ or $S(z) = 2z^{-1} - z^{-2}$, a first or a second order shaping can be achieved, respectively. Figure 2(a) shows a 3rd order DT-ΣΔM with chain of integrators with weighted feed-forward summation (CIFF) structure. This modulator is used as an exemplary modulator, but the techniques explained in this paper are general and can be used in any ΣΔ Modulator. Corresponding to Fig. 2(a), the DΣΔM is used in modulator to quantize the feedback signal and shape the quantized error. The

difference of the output and input signal of DΣΔM is injected to the second integrator via a digital filter, $I'(z)$. This digital filter is nominally equivalent to the analog integrator, $I(z)$. According to Fig. 2(a), the signal $U(z)$ is given by [5]:

$$\begin{aligned} U(z) &= I(z)\big(X(z)-Y_Q(z)\big)+I'(z)\big(Y_Q(z)-Y_q(z)\big) \\ &= I(z)\big\{X(z)-Y_q(z)-(1-\alpha)\big(1-S(z)\big)E_q(z)\big\} \end{aligned} \quad (2)$$

In this relation, for the sake of simplicity, $I'(z)$ is assumed to be $\alpha I(z)$ where α is ideally equal to one [5]. In case of perfect matching, $U(z) = I(z)$ ($X(z) - Y_q(z)$) and this modulator has the same expression as the conventional multibit one. However, in real cases, the mismatch error is shaped by $(1 - S(z))$ and therefore its effect is considerably mitigated.

Figure 2(b) illustrates the continuous time equivalent of Fig. 2(a). In this figure, unlike DT-ΣΔM, the feed forward paths in CT-ΣΔM are extracted before the summation nodes to prevent any unrealizable delay-free loop in modulator (in cases that $P(z)$ or $J(z)$ contain a delay-free terms). Two inner feedback paths can have the gain mismatch with the first feedback path.

The coefficients design of modulator shown in Fig. 2(b) may be started with the design of its discrete time counterpart. At first, the coefficient of multibit DT-ΣΔM, a_1, a_2 and a_3, are designed corresponding to its order, OSR and etc [2], and then these coefficients are used in modulator of Fig. 2(a). The modified z-transform can be used to transfer the discrete time modulator of Fig. 2(a) to the continuous time modulator of Fig. 2(b). It should be noted that there are two paths that should be matched in these two modulators. The first path is the outermost feedback path that transfers the $Y_q(z)$ to the quantizer output node. The second path is the inner feedback path that transfer the $(Y_q(z) - Y_Q(z))$. Assuming a rectangular shaped non-return-to-zero (NRZ) DAC pulse, the results of the first path matching are:

$$k_1 = a_1 - \frac{a_2}{2} + \frac{a_3}{3},\ k_2 = a_2 - a_3,\ k_3 = a_3 \quad (3)$$

In Fig. 2(b), $P(z)$ and $J(z)$ are used for the second path matching. Again using modified z-transform we have:

$$\begin{aligned} &P(z)\Big(Z_{m1}\big\{LF_{CT2}(s)/s\big\}-Z_{m2}\big\{LF_{CT2}(s)/s\big\}\Big)+ \\ &J(z)\Big(Z_{m1}\big\{LF_{CT3}(s)/s\big\}-Z_{m2}\big\{LF_{CT3}(s)/s\big\}\Big) \\ &= P(z)\left(\frac{k_2+k_3/2}{z-1}+\frac{k_3}{(z-1)^2}\right)+J(z)\left(\frac{k_3}{z-1}\right) \\ &= \frac{a_1}{(z-1)}+\frac{a_2}{(z-1)^2}+\frac{a_3}{(z-1)^3}=LF_{DT}(z) \end{aligned} \quad (4)$$

where $LF_2(s)=\left(k_2\dfrac{f_S}{s}+k_3\dfrac{(f_S)^2}{s^2}\right), LF_3(s)=\left(k_3\dfrac{f_S}{s}\right)$ were used.

The suitable candidates for $P(z)$ and $J(z)$ in (4) are as follows:

$$P(z) = p_0 + \frac{p_1}{z-1},\ J(z) = j_0 \quad (5)$$

where p_0, p_1 and j_0 are the constant coefficients. These selections give enough free parameters to solve the equation given in (4).

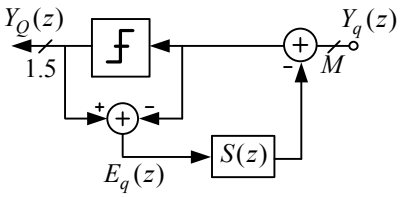

Figure 1. Block diagram of a digital sigma delta modulator (DΣΔM).

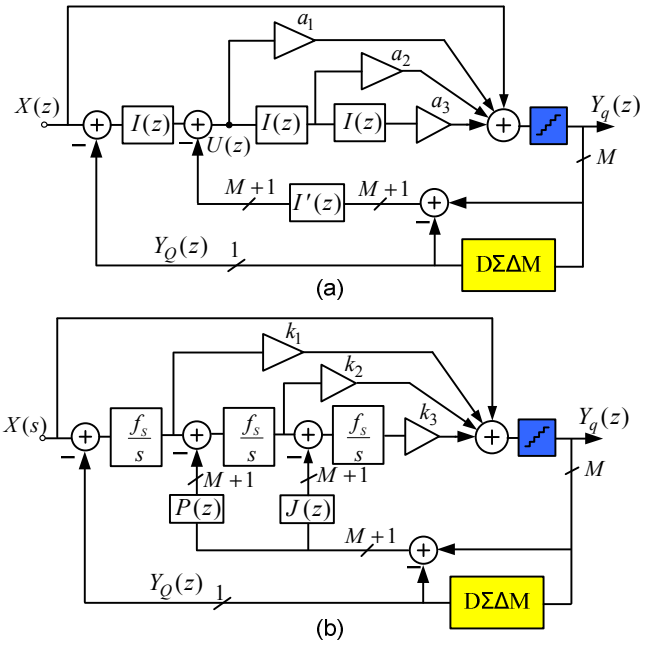

Figure 2. Dual quantization technique used in a 5-bit 3rd order examplary modulator with CIFF structure, a) discrete time and b) continuous time.

By substituting the relation (5) in (4), we have:

$$\begin{aligned} &p_0 = (\,a_2 - a_3\times(k_2/k_3+0.5)\,)/k_3,\quad p_1 = a_3/k_3 \\ &j_0 = \big(a_1 - p_0\times(k_2+k_3/2)\big)/k_3 \end{aligned} \quad (6)$$

The number of bits of output signal of I'(z) in Fig. 2(a) and also the number of bits of output signals of P(z) and J(z) in Fig. 2(b), were acquired by simulation where a -4.4 dBFS sine-wave was used. Although the dual quantization of a CT-ΣΔM needs two new (M+1)-bit DACs, but, these DACs benefit from the shaping effect of the previous integrators and so their design are not critical. Besides, the number of bits of these two DACs can be reduced (by truncating their LSBs) without significant degradation of modulator performance [5]. This truncation error is added to E_q(z) introduced in (2) which itself is shaped by (1–S(z)).

III. PROPOSED TECHNIQUE

Usually, there is a least significant bit (LSB) variation between two adjacent output samples of a multibit DAC. Therefore, the influence of clock jitter on a multibit DAC is an order of magnitude less than a single bit DAC. By increasing the bit length, a NRZ DAC benefits more clock jitter insensitivity than a return-to-zero (RZ) DAC [7]. The dual quantization technique, introduced in section II, decreases the number of bits of the

outermost feedback path. This bit length reduction leads to increment of clock jitter sensitivity. Spectrally shaped feedback is one of techniques to reduce the clock jitter sensitivity of a 1-bit feedback path in CT-ΣΔMs [6].

Combination of dual quantization and spectrally shaped feedback is proposed for a CT-ΣΔM to eliminate the nonlinearity problem of its multibit DAC where the clock jitter insensitivity of multibit DAC is approximately reserved. Figure 3 illustrates a low-pass FIR-filter used in feedback path. This filter smoothes the variation of 1-bit feedback signal and reduces the effect of clock jitter like a multibit DAC. This filter is merged in the structure of DAC as a pseudo digital implementation [8]. For example in modulator of Fig. 3, with 1-bit feedback signal, $y_Q[.]$, a chain of (N-1) flip-flops makes different delay tabs of FIR-filter and there is a vector of N samples of $y_Q[.]$ on this tabs as Y=[$y_Q[n]$, $y_Q[n-1]$,…., $y_Q[n-N+1]$]. The coefficients of FIR-filter are implemented by elements of DAC with different values. The mismatch of elements of DAC is equal to mismatch of coefficients of FIR filter and just makes a linear error. At a specific time, each sample of vector Y excites one of elements of DAC. Often there is a small variation between two adjacent output samples of this DAC, like a conventional multibit DAC, which leads to a low clock jitter sensitivity.

In Fig. 3, there are two paths from $Y_Q(z)$ to the input of the second and third integrators, via two filters $L(z)$ and $N(z)$, respectively. These two paths compensate the effect of $F(z)$ on both the modulator signal and noise transfer functions. $P(z)$, $J(z)$ are the same as what explained in Fig. 2(b). Like modulator of Fig. 2(b), the design of modulator shown in Fig. 3 may be started from the DT-ΣΔM of Fig. 2(a). There are two paths that should be matched between the modulator shown in Fig. 2(a) and the modulator of Fig .3. The first path transfers $Y_Q(z)$ to input of quantizer through $F(z)$, $L(z)$ and $N(z)$. For matching of this path, the modified z-transform is used by assuming a rectangular shaped NRZ DAC pulse:

$$\begin{aligned} &F(z)\left(\frac{k_1+k_2/2+k_3/6}{z-1}+\frac{k_2+k_3}{(z-1)^2}+\frac{k_3}{(z-1)^3}\right)\\ &+L(z)\left(\frac{k_2+k_3/2}{z-1}+\frac{k_3}{(z-1)^2}\right)+N(z)\left(\frac{k_3}{z-1}\right)\\ &=\frac{a_1}{(z-1)}+\frac{a_2}{(z-1)^2}+\frac{a_3}{(z-1)^3} \end{aligned} \tag{7}$$

Since $F(z) = \Sigma f_k z^{-k}$ is a low-pass FIR filter which should not perform any attenuation in low frequencies ($z \to 1$), therefore:

$$\sum_{k=1}^{N} f_k = 1 \tag{8}$$

$F(z)$ has N terms of power of z^{-1} and there are three terms of power of $1/(z-1)$ in (7), consequently there are 3×N equations in (7). Although $L(z)$ is enough for compensating the effect of $F(z)$ [6], by using both $L(z)$ and $N(z)$, the solution to (7) will be more straight forward.

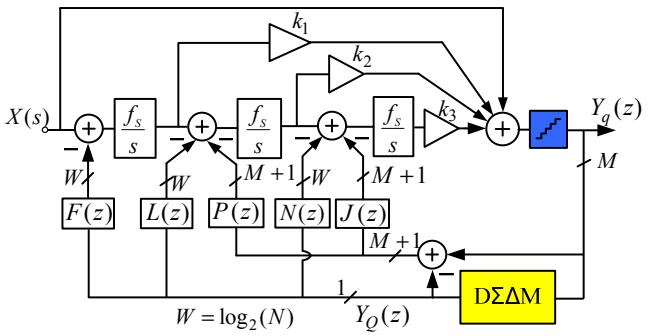

Figure 3. Proposed combinational mudulator (dual quantization with spectrally shaped feedback used in a 5-bit, 3rd order CT-ΣΔM).

Regarding (8), a golden solution for (7) independent of $F(z)$, can be obtained by choosing:

$$L(z)=\big(1-F(z)\big)\big(l_0+l_1/(z-1)\big),\ \ N(z)=n_0\big(1-F(z)\big) \tag{9}$$

Equations (8) and (9) let coefficients k_1, k_2 and k_3 in (7) to be calculated from (3), identical to the conventional case $F(z)$ =1. After calculation of k_1, k_2 and k_3, the coefficients l_0, l_1 and n_0 are calculated independently of $F(z)$ as:

$$\begin{aligned} &l_1=\frac{a_3}{k_3},\quad l_0=\big(a_2-l_1\big(k_2+k_3/2\big)\big)/k_3\\ &n_0=\big(a_1-l_0\big(k_2+k_3/2\big)\big)/k_3 \end{aligned} \tag{10}$$

According to (9), since $F(z)$ is a FIR filter, $N(z)$ is also a FIR filter and in cases that $(1-F(z))$ has a zero at $z = 1$, $L(z)$ will also be a FIR filter. These FIR filters can be implemented by pseudo digital implementation and just add some flip-flops as overhead.

The second feedback path of modulator of Fig. 3, transfers $(Y_q(z) - Y_Q(z))$ to the input of quantizer through $P(z)$ and $J(z)$. The equations of matching of this path are completely similar to (4) and consequently the results of (5) and (6) are valid here.

Design of $F(z)$ involves the design of both its order and its coefficients. Increasing the order of $F(z)$ has at least two drawbacks. Firstly, it may leads to very small (unrealizable) DAC elements. Secondly, using a linear model of modulator in Fig. 3 and writing the transfer function of $X(z)$ to the output signal of first integrator, it is evident that $F(z)$ may increase the output swing of the first integrator. Therefore a moderate order should be selected for $F(z)$. Since the jitter sensitivity of a NRZ DAC is proportional to the variation of feedback signal [9], $F(z)$ should minimize the variation of its input signal. The optimal design of coefficients of $F(z)$ is beyond the scope of this paper and it is not discussed here. An optimal design of coefficients of $F(z)$ for a RZ DAC is presented in [6].

The proposed technique replaces a relaxed W-bit DAC, $\{W=\log_2(N)\}$, of an M-bit DAC in a conventional multi-bit ΣΔM. But, it adds W flip-flops and some additional relaxed DACs (two $(M+1)$-bit and two W-bit DACs). As discussed before the number of bits of two $(M+1)$-bit DACs can be reduced by truncation. Also, this technique works independently of OSR. In contrast, the shaping techniques like DWA [3] need

a logarithmic shifter with 2^M paths and M stages which may has considerable latency. Also the performance of this shaping technique depends on the OSR.

IV. SIMULATION RESULTS

Figure 4 shows the histogram of signal-to-noise and distortion ratio (SNDR) in four different cases (for each case 500 simulations were performed): (a) and (b) conventional 3rd order 5-bit quantizer where 0.2% and 0.5% intentional mismatch was considered in DAC elements, respectively, (c) and (d) proposed modulator of Fig. 3 where respectively 0.2% and 0.5% intentional mismatch for gains of its feedback paths and also 0.2% and 0.5% relative mismatch between the coefficients of its FIR filter were assumed. A 2nd order DΣΔM ($S(z) = 2z^{-1} - z^{-2}$) was used in modulator of Fig. 3. In all simulations the OSR was 16. In perfect matching, both modulators have an SNDR of 97 dB. According to Fig. 4, it is evident that mismatch causes a serious performance degradation in multibit modulator where the performance degradation of proposed modulator is negligible.

Table 2 compares the solidity against the clock jitter of tree modulators: conventional 5-bit modulator, modulator of Fig. 2(b) (dual quantization technique) and the proposed modulator of Fig. 3 with $F(z) = (1+z^{-1}+z^{-2})/3$ and $N = 3$. A perfect element matching was assumed in tree modulators and the clock jitter was modeled according to [9]. The input signal was a sinusoidal wave at f_s /512 with -4.4 dBFS amplitude. The results confirm that the proposed technique has approximately a clock jitter insensitivity of a multibit modulator where the dual quantization technique without spectrally shaped feedback suffers from the significant performance degradation.

Figure 5 illustrates the simulated SNDR versus different input values in tree modulators: ideal modulator, proposed modulator with 0.5% mismatch, conventional multibit modulator with 0.5% mismatch between the DAC elements. In cases 2 and 3, the clock jitter of $\sigma_t = 0.01\ \% \times T_s$ ($T_s = 1/f_s$) was also exerted. As is seen, in spite of mismatch and clock jitter, the proposed technique improves the performance of a conventional multibit modulator up to the ideal case.

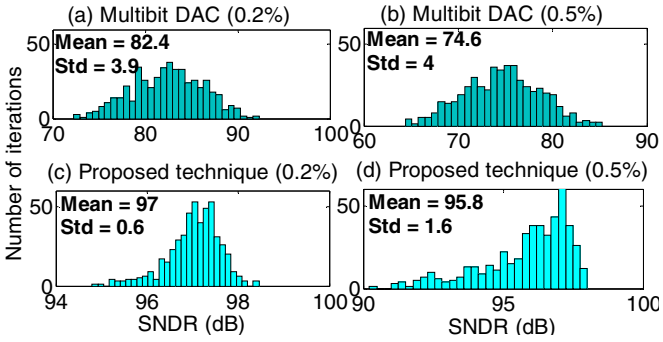

Figure 4. Comparison of SNDR degradation due to implementation mismatch, between conventional multibit modulator and proposed modulator.

TABLE I. CLOCK JITTER SOLIDITY COMPARISON FOR TREE DIFFERENT CASES WITH TWO DIFFERENT VALUES OF CLOCK JITTER.

Modulator type	SNDR (dB)	
	$\sigma_t = 0.01\ \%\ T_S$	$\sigma_t = 0.02\ \%\ T_S$
Dual quantization (1-bit DAC)	85	79
Proposed technique (FIR filter merged to 1-bit DAC)	93.5	88
Multibit modulator (ideal 5-bit DAC)	93.5	88
Ideal ΣΔM	97	

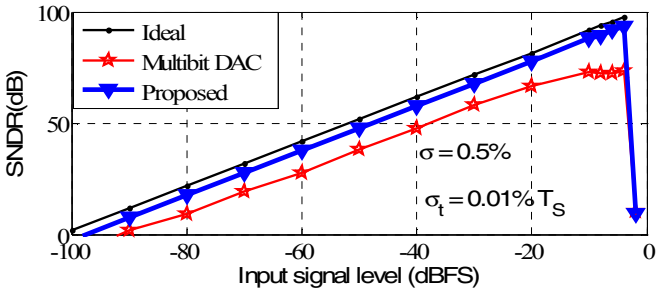

Figure 5. Simulated SNDR versus the input signal amplitude

V. CONCLUSIONS

A combination of dual quantization and spectral shaped feedback was proposed to overcome the mismatch errors in the feedback DAC of multibit CT-ΣΔMs. This combinational technique completely eliminates the mismatch error of multibit DAC where approximately the same jitter insensitivity of a multibit ΣΔM is achieved. The performance of the proposed modulator is approximately equal to a multibit modulator with an ideal DAC.

References

[1] J. M. de la Rosa, "Sigma-delta modulators: tutorial overview, design guide, and state-of-the-art survey," *IEEE Trans. Circuits Syst.-I, Reg. Papers*, vol. 58, no. 1, pp. 1-21, Jan. 2011.

[2] R. Schreier and G. C. Temes, *Understanding delta-sigma data converters*, IEEE Press, 2005.

[3] R. T. Barid and T. S. Fiez, "Linearity enhancement of multibit ΣΔ A/D and D/A converters using data weighted averaging," *IEEE Trans. Circuits Syst. II*, pp. 753 -762, vol. 42, no. 12, Dec. 1995.

[4] P. Witte and M. Ortmanns, "Digital background DAC error estimation using a pseudo random noise based correlation technique for sigma-delta analog-to-digital converters," *IEEE Trans. circuits syst. I*, pp. 1500-1512, vol. 57, no. 7, Jul. 2010.

[5] F. Colodro, et al., "digital noise-shaping of residues in dual-quantization sigma–delta modulators," *IEEE Trans. circuits syst. I*, pp. 225-232, vol. 51, no. 2, Feb. 2004.

[6] Oliaei, O. "Sigma-delta modulator with spectrally shaped feedback," *IEEE Trans. Circuits Syst.-II*, vol. 50, no. 9, pp. 518–530, Sept. 2003.

[7] M. Ortmanns and F. Gerfers, *Continuous-time sigma-delta A/D conversion*, Springer, 2005.

[8] D. K. Su and b. A. Wooley, "A CMOS oversampling D/A converter with a current-mode semi-digital reconstruction filter," *IEEE J. Solid-State Circuits*, vol. 28, pp. 1224–1233, Dec. 1993.

[9] L. Hernández et al., "Modeling and optimazation of low pass continuous-time sigma-delta modulators for clock jitter noise reduction," in *Proc. ISCAS*, pp. 1072-1074, 2004.