Behavioural Modelling Of Instrumentation Delta-Sigma ADC

V.I. Didenko, A.L. Movchan, J.S. Solodov

Moscow Power Engineering Institute (Technical University),
111250, Krasnokazarmennya st., 14, Moscow, Russia,
phone: +7 095 362 73 68, fax: +7 095 362 74 68, email: didenko@iit.mpei.ac.ru

Abstract - This paper examines the quantization error of a modulator for delta-sigma analog-to-digital converter (DSADC). In opposite to many papers, the analytical equations are found in time domain. This approach allows the authors to present the modulator as analog-to-digital converter (ADC) of input signal to code for given length of digital sequences at the modulator output. Such presentation is useful for investigation of quantization error for ordinary DSADC and can also be considered as the base for a new ADC type. Both analytical equations and results of simulations are used to find the sources of quantization errors: effect of the scale end, limited number of code transition levels, unequal length of code bin widths, non-ideal digital filters. Some suggestions are made to decrease influence of mentioned sources on quantization error including the new ADC type with conversion delay.

I. Introduction

Analog-to-digital converter is a key element for many applications. The most popular types of ADC are now devices based on delta-sigma modulation. The main problem of DSADC is a quantization error. For random input signal it can be considered as a random quantity or quantization noise. In investigations known, a modulator is analyzed in frequency domain without a digital filter. The digital filter is supposed to sample the modulator output at a frequency f_S/R, where f_S – is a comparator sampling frequency at the modulator output, R – oversampling ratio being a fundamental parameter at analysis and synthesis of different modulator types, mainly when investigating noise. The digital filter bandwidth is supposed to be the same as for an ideal low-pass filter, i.e. having unity gain from 0 Hz to $0.5f_S/R$ and having zero gain for all higher frequencies. The modulator properties are usually described in terms of *SNR* – the ratio of input sine wave RMS to noise RMS value. In [1] a following basic formula has been derived

$$SNR \approx 10\log\frac{3a^2(2n+1)R^{2n+1}}{2\alpha^2\pi^{2n}}, \tag{1}$$

where a – input signal amplitude referred to a reference voltage, n – modulator order, α – gain in the modulator channel, R – modulator oversampling ratio (in examples of our paper $a = 1$, $n = 2$, $\alpha = 0.5$). Graphs, corresponding to (1), have been demonstrated in [2]. In the paper [3] a particular case of (1) for $n = 3$ is given, while it is noted that (1) is quite approximate. In the paper [1] it was pointed at possible instability of the modulator and zeroes introduction has been proposed to improve the stability. These and other improvements, when done, changed the slope and proportionality factor of *SNR* vs. R function, while the linearity of the mentioned characteristic was not changing. Authors offered [4] to find parameters of (1) from experiments (amplitude coefficient instead of a/α and an exponent instead of $(2n+1)$ for two values of R). Then (1) became accurate enough for any other values of R. In metrology the measurement uncertainty is usually evaluated by multiplication of *RMS* into K coefficient that depends on chosen confident probability. At that the noise is usually thought to be normally distributed. Then for typical probability P=0.95 we have $K = 2$, while for maximum error $K = 3$ is taken, corresponding to $P = 0.9973$. In [5] the value $K = 3.3$ is recommended; this could result in to an idea to choose $P = 0.999$. However in [6] the authors have shown that the maximum values of noise *RMS* could differ from the typical ones by a factor of 2 approximately. Therefore, to calculate maximum error with probability $P = 0.9973$ for any value of input signal about zero at the value of *RMS* found from (1), it is necessary to take $K \approx 6$. Another important question is whether the *RMS* is constant as a function of the input signal. Relying on experimental data analysis as well as simulation results authors has shown [6] that the *RMS* significantly increases at the edges of input signal range in comparison with *RMS* for the signal about zero. Authors are not aware of theoretical investigations explaining this effect. This paper is an attempt to eliminate lack of comprehension for analysis of DSADC.

II. Analysis of delta-sigma ADC in time domain

The DSADC is known to consist of two main blocks: a delta-sigma modulator and a digital filter. Typical example of DSADC simulated in the paper is shown in Fig. 1.

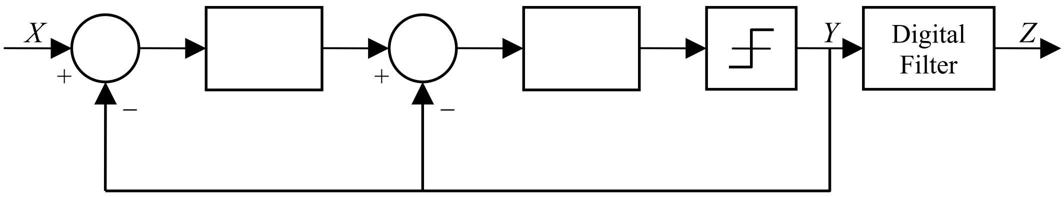

Figure 1. Typical structure of DSADC.

Since authors investigate instrumentation DSADC, the metrological approach to the modulator analysis appears to be worthwhile. The modulator input signal is a DC voltage. As used in literature, we shall utilize normalized value of input signal X that is calculated by dividing true value of the input signal into reference voltage. Quantization error is an even function of X. To reduce data volume we shall mainly give research results for the range $0 \leq X \leq 1$. At the output of the modulator a continuous stream of binary signal is present (zeros and ones). In measurement practice a conversion time is usually rated. We shall use the approach for the modulator. The modulator output signal Y is presented by a sequence of ones and zeros. Let us assign symbol L for the number of ones and zeros taking place during conversion. The L equals to the ratio of the conversion time to a period of the comparator sampling frequency $T_S = 1/f_S$. A number of such sequences are shown in Table 1 for the modulator shown in Fig. 1. It is supposed that L = 28 and $\alpha = 0.5$.

When $X = 0$ we have an initial sequence. We shall take for this sequence that $Y = Y_0 = 0$. When $X = 1$ the modulator output sequence contains only ones. We shall take for the sequence that $Y = Y_{MAX} = 1$. Then the nominal (ideal) modulator transfer function will be $Y = X$. Let us assign the total number of sequences having length L for X from 0 to 1 as N_Y. If initial values of voltages on the modulator capacitors are taken to be constant (e.g. zero) then to each value of X some certain sequence of ones and zeros should correspond. Under these conditions we shall assign value $Y_1 = 1/N_Y$ to the first sequence after initial one, the value $Y_2 = 2/N_Y$ – to the second sequence and so on. As a result the analog input signal X is expressed as digital one by means of staircase function, i.e. we can represent the modulator as ADC. The graphs shown in Fig. 2 correspond to the modulator shown in Fig. 1 for L=28 and α=0.5.

Like any ADC the converter $Y = Y(X)$ is characterized by a number of parameters, among of them the code bin width Q_i being of specific importance. This parameter determines errors of DSADC that cannot be reduced by means of any further digital filter for given conversion time. In other words the knowledge of Q_i enables us to determine potential accuracy of DSADC as a whole.

Table 1.

Input voltage X	Modulator output sequence Y
0.00000	1001100110011001100110011001
0.00265	1001100110011001100110011010
0.00363	1001100110011001100110100110
…	…
0.50143	1011110110111101101111011101
0.50293	1011110110111101110110111101
0.50530	1011110110111101110110111110
…	…
0.99716	1111111111111111111111111101
0.99736	1111111111111111111111111110
0.99754	1111111111111111111111111111

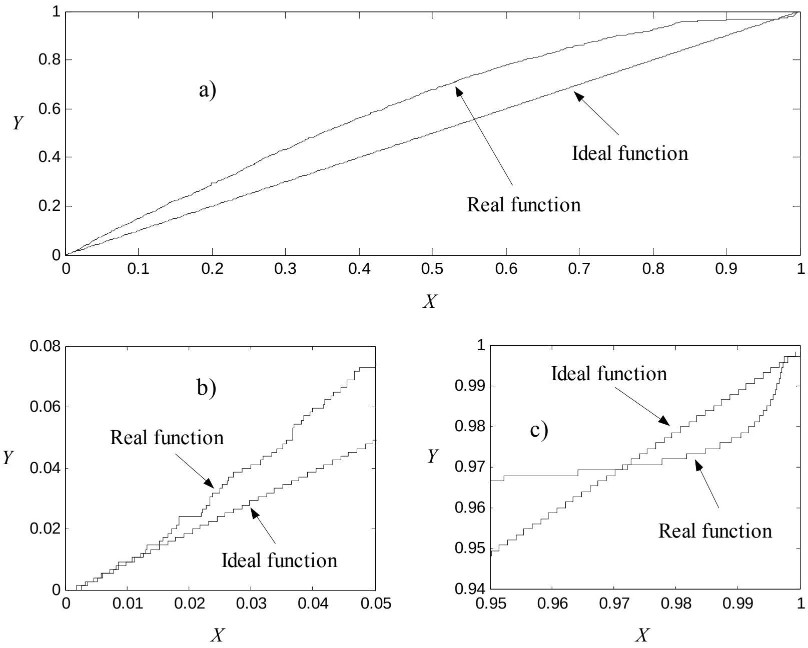

Fig. 2. Modulator transfer characteristics.

Distinction of real errors of DSADC from the errors determined by Q_i characterizes the extent of digital filter ideality if the modulator and its operation mode are given. Let us consider what errors can be upper estimated using known values of Q_i. Maximum absolute error of DSADC cannot be less than $\Delta_{\max} = \pm 0.5 Q_{i.max}$. This error can be achieved with a device at the modulator output which is difficult to build but completely eliminates linearity error of conversion $Y = Y(X)$. If all the code bins were of the same width then the maximum error would have been $\Delta_{\max} = \pm 1/2N_Y$. If the input quantity is uniformly distributed within the operation range then the modulator error becomes random quantity, its standard deviation being calculated by means of the following formula

$$\sigma = \frac{1}{2\sqrt{3}} \sqrt{\sum_{i=0}^{N_Y} Q_i^3} \ . \tag{2}$$

If all the code bins were of the same width then it would have been

$$\sigma = \frac{1}{N_Y 2\sqrt{3}} \ . \tag{3}$$

Now let us consider possible values of number of code bins in transfer characteristic $Y = Y(X)$. When input signal is not negative the number of combinations in the sequence of length L is determined by (4)

$$N_Y = 2^{L-1}. \tag{4}$$

Equation (4) corresponds to a number of states of successive approximation ADC having L bits. In other words, under the same resolution DSADC can potentially have the same conversion time as that of the successive approximation ADC without accurate multi-bit DAC. However, there are at least four hindrances to implement this condition. Let us consider these hindrances and the ways to overcome them.

Firstly, the value of N_Y is non-uniformly distributed within the range of X. Possible amount of code bins having given number M of ones (for positive X a number of ones is more than the number of zeros) is determined by formula

$$N_{Y.M} = \frac{L!}{M!(L-M)!}. \quad (5)$$

For example, for $L = 4$ (in practice this value is usually much bigger) there are total $2^4 = 16$ code bins. According to (5) the number of combinations having two ones ($M = 2$) is six (three for $X<0$ and three for $X>0$); for $M = 3$ there are four combinations and for $M=4$ there is only one. It can easily be noted that when the sequence length increases the number of combinations corresponding to the input signal about zero increases strongly. Meanwhile, the number of combinations having (L–1) ones (one "0") increases proportionally. Let us find resolution (average code bin width) for the sequence having only one "0". Extreme relationship between input voltage V_{IN} (in terms of reference voltage) and the number of ones M for any DSADC having quite big L is expressed by (6)

$$V_{IN} = \frac{2M-L}{L}. \quad (6)$$

It follows from (6), that the input signal increment for 2/L-th of full scale must correspond to the increment of M by 1. The sequence, having (M-1) ones, must correspond to 2/L-th of full scale, in the first approximation. On this part average code bin width turns out to be $2/L^2$. According to (4) the average code bin width must be equal to 2^{1-L}. For big enough L used in practice, the average code bin width for (M-1) ones turns out to be much greater than it could be expected according to (4). For example, for L=100 the calculation yields $2/L^2=2\cdot10^{-2}$ and $2^{1-L} = 2\cdot10^{-99}$. To reduce the average code bin width one can use only part of the range, as done in some sigma-delta ADC models. Unfortunately, here the full-scale error rises that is bound with any additive error constituents.

The second hindrance on the way to implement limiting resolution of DSADC is a capability of the concrete modulator scheme to fulfill quantization of the input signal with limited number of quanta. Let us find this value for the scheme shown in Fig. 1. Let us consider the change of standardized input signal Q_{av}, corresponding to the average code bin width on the modulator transfer characteristic. For switched-capacitors modulator, the voltage at the output of the first integrator through l interval after input signal change will get the increment (7):

$$V_{OUT.1} = \alpha \cdot l \cdot Q_{av}, \quad (7)$$

where α – gain in the channel (in our investigations equals to 0.5), l – sequence length at the modulator output from input signal step to the end of conversion time L.

In the case of second-order modulator, the voltage at the output of the second integrator through l interval after input signal change will get the increment (8), which can be calculated as a sum of first l members of an arithmetical progression (7):

$$V_{OUT.2} = 0.5 \cdot \alpha \cdot l \cdot (l+1) \cdot Q_{av}. \quad (8)$$

For n-th order modulator an increment of the voltage at the output of the n-th integrator can be calculated similarly.

The voltage calculated from (8) for 2nd order modulator either similar expression for higher order modulators are balanced with voltage proportional to the reference voltage. In [1] the range of possible comparator input voltages is taken to be equal to quadruple reference voltage, though possible significant distinctions from the value are noted as well. First we shall take the same assumption. Additionally we shall assume that the average value of a balancing voltage equals to one half of the range, i.e. double reference voltage. The last assumption results from an admission made in [1] about uniform distribution of the comparator input voltage. Equating the left part of (8) to double reference voltage one gets

$$Q_{av} = \frac{2}{0.5 \cdot \alpha \cdot l \cdot (l-1)}. \quad (9)$$

The amount of code bins for non-negative input signals ($0 \le X \le 1$) produced by changing the last term of the sequence having length l on the modulator output is calculated from (9) as

$$N_{Y\,l} = \frac{0.5 \cdot \alpha \cdot l \cdot (l+1)}{2}. \tag{10}$$

Total amount of code bins for non-negative X is found as a sum of all the code bins produced on conversion interval L:

$$N_Y = \sum_{i=l-L+1}^{l} \frac{0.5 \cdot \alpha \cdot i \cdot (i+1)}{2}. \tag{11}$$

If one introduces into consideration a coefficient η proposed in [1], then (11) is transformed to the following form

$$N_Y = \sum_{i=l-L+1}^{l} \frac{0.5 \cdot \alpha \cdot \eta \cdot i \cdot (i+1)}{2}. \tag{12}$$

The expression (12) is not less fundamental for the modulator analysis in time domain than (1) is for frequency domain. The values of N_Y found from (12) are several orders less than those found from (4). In other words, the second-order modulator doesn't use all possible combinations that could be implemented in sequence with given l. To increase the number of combinations one can increase the modulator order. However, along with hardware complication, the danger of self-excitations increases as well [1]. Analysis of equation (12) enables us to propose a new method to increase the number of code bins for a given conversion time. It is necessary to choose $l>L$ for this purpose. In other words, a conversion should start some time interval after the input voltage has been applied to the modulator input. At the limit one can get the number of code bins determined by (4). However, at unnecessarily big excess of l above L the repetition of sequences is possible, when the input signal rises. In this case the modulator stops being a monotonic ADC. The approach mentioned is, essentially, a proposition to design a new ADC class. The class completely uses the DSADC hardware while the software is essentially different. It assumes writing sequences to the ADC memory for different signals for chosen values of l and L. A logic device should compare a received sequence with a sequence written to the memory and yield a corresponding input signal. In comparison with ordinary DSADC, the new one has significantly smaller quantization error for the same conversion time either smaller conversion time for the same quantization error. The drawback is worse reduction of disturbances.

The third hindrance on the way to implement limiting resolution of DSADC is unequal code bin widths (differential nonlinearity of the modulator represented as an ADC). This effect can be seen in Fig. 2.

The fourth factor that results in increase of the quantization error is difference of real characteristics of a digital filter from ideal ones. Let us consider this question on the example of a second-order modulator. The *RMS* of discreteness error for $a=1$, $\alpha=0.5$ and $n=2$ is found from (1) as

$$\sigma = \frac{\pi^2}{R^{2.5}\sqrt{15}}. \tag{13}$$

From (3) and (13) under the same assumptions for $l = L$ one can find that

$$\sigma = \frac{1}{\sqrt{3}\sum_{i=1}^{L} 0.5i(i+1)}. \tag{14}$$

Equating (12) and (13) we shall get a relationship between the conversion time L and the oversampling ratio R:

$$R = \left(\frac{\pi^2}{\sqrt{5}}\sum_{i=1}^{L} 0.5i(i+1)\right)^{0.4}. \tag{15}$$

Reminding that L equals to the number of binary pulses taken from the comparator input for chosen sequence length, i.e. L equals to a ratio of the ADC conversion time to a period of pulses repetition. For the 3-stage moving average filter [5] the settling time equals to $3R$ if synchronization is used. The expression in brackets changes vs. l with an exponent 2.95. It means that R rises faster than l and conversion time of DSADC with a digital filter increases continuously in comparison with L.

Theoretical conclusions are illustrated in Table 2.

Table 2.

l		28	30	40	50	60	70
$L = l$	$N_{Y.SIM}$	754	927	2121	4115	7071	11089
	$N_{Y.CALC}$	746	910	2110	4236	7071	11243
	R	50	54	76	99	123	147
	$3R/L$	5.4	5.4	5.7	5.9	6.1	6.3
$L = 28$	$N_{Y.SIM}$	754	927	2112	4011	6636	9897
	$N_{Y.CALC}$	746	905	2043	3863	6014	8874
	$\sigma_{id}\times10^{-4}$	3.8	3.1	1.4	0.72	0.44	0.29
	$\sigma_{tr}\times10^{-4}$	10.4	9.2	4.9	3.2	2.2	1.6
	$\sigma_{id} / \sigma_{tr}$	0.36	0.34	0.29	0.22	0.20	0.18

The number of code bins for $L = l$ found from simulation ($N_{Y.SIM}$) and calculation ($N_{Y.CALC}$) with (12) for $\eta = 0.731$ distinguish less than 2%. The value of R was found from (15). True conversion time for 3-stage moving average filter is more than ideal conversion time from 5.4 to 6.3 times when L changes from 28 to 70. The number of code bins for $l \geq L$=28 found from simulation ($N_{Y.SIM}$) and calculation ($N_{Y.CALC}$) with (12) for $\eta = 0.731$ distinguish less than 15%. The standard deviation for equal code bin width (σ_{id}) found with (3) and for true code bin width (σ_{tr}) found with (2) distinguish from 0.36 to 0.18 when l changes from 28 to 70.

III. Conclusions

1. A modulator of DSADC can be considered as an independent ADC if one fixes sequence length at its output and initial conditions on the modulator integrators.
2. Maximum achievable ADC resolution (discreteness error) along the full scale has been found. The quantization error was shown to rise at the scale edges. A possibility to reduce the effect using only a part of the scale has been shown.
3. A general approach to the analysis of ADC resolution has been proposed by finding analytical expressions in time domain. Reliability of the received analytical expressions is confirmed by simulation.
4. A method to reduce the ADC quantization error has been proposed by delaying the conversion start with respect to applying signal to the modulator having preset initial voltages on capacitors of the modulator integrator.

References

[1] P. Benabes, P. Aldebert, R. Rielbasa, "Analog-to-digital sigma-delta converters modelling for simulation and synthesis," *Proceedings of International Workshop on ADC Modelling and testing*, Bordeaux, France, pp. 3-14, September 9-10, 1999.

[2] *System application guide*, Analog Devices technical reference books, U.S.A., 1993.

[3] A. J. Davis, G. Fisher, "Behavioural modelling of delta-sigma modulators", *Computer Standards & Interfaces,* 19 pp. 189-203, 1998.

[4] V. Didenko, A. Movchan, "Minimization of number of metrological parameters for data acquisition systems", *IEEE transactions on Instrumentation and Measurement*, vol. 51, No. 1, pp. 88-91, Feb. 2002.

[5] *Signal Conditioning ADC AD7714*, Analog Devices, U.S.A., Data Sheet, 1997.

[6] V.I. Didenko, A.L. Movchan, J.S. Solodov, "Modelling of instrumentation sigma-delta analog-to-digital converters", *Proceeding of the 8th International Workshop on ADC Modelling and Testing*, Perugia, Italy, pp.75-80, 2003.