JASTT

Effects of Rounding and Truncating Methods of Quantization Error and SQNR for Sine Signal

Alaaldin Hasso[1], Karwan Jacksi[1,*]

[1]*Department of Computer Science, University of Zakho, Kurdistan Region, Iraq, (alaaldin.hasso, karwan.jacksi)@uoz.edu.krd)*

Abstract

Within the Analog to Digital Conversion (ADC), quantization noise is a duplicate of a Quantization Error (QE) which is introduced by quantization. In signal processing and telecommunication systems, the noise is non-linear and depends on the signal type. During the analog, Sine signal converts to the digital (ADC) process, the two methods are used Rounding and Truncating in-order to eliminate the error produced in the digitization process. The rounding method quantize assigns each sample of sine signal to the nearest quantization level. However, making the Truncating would have assigned each sample of sine signal to the quantization level below it.

This paper compares the rounding and truncating methods of QE for sine signal, signal to quantization noise ratio, correlation coefficient, and regression equation of a line for both methods. Then, it calculates the residual sum of squares and compares it to the regression equations of the lines.

Keywords: ADC, SQNR, RSS, Quantization Error, Correlation Coefficient, Regression Equation

I. INTRODUCTION

The wide range of ceaseless sine signal has to be transformed to digital data with limited accuracy. This is known as the quantization [1]. The Quantization Error (QE) is the error generated from applying the incessant signal value by a limited set of discrete value levels [2].

During the Analog sinusoidal signal to Digital signal Converter (ADC), the adequacy of the analog sine signal to be transformed has endless accuracy [3]. Implementing a measurable methodology that adopting the QE of the sine signal is irregular sequence and augments to the original signal [4]. To decrease the additional digits, the truncating method can be used that allocates each sine signal models to a lower level of QE [5].

During the ADC process, as the number of bits rises in the digital world, the experimental and theoretical Signal to Quantization Noise Ratio (SQNR) of the sine signal using the rounding method are similar to theoretical values [6]. However, by means of the truncating method, the experimental SQNR value is permanently underneath the theoretical value, which means not as much of exact quantization of the sine signal that consequences in a lower SQNR. This was expected because instead of rounding the data, the truncating method occurred [3].

This paper reviews the QE, SQNR, correlation coefficient (r), regression equation of a line in section II. Calculating the Residual Sum of Square (RSS) for the rounding and truncating methods is also detailed in section II. The comparison between the regression equations of a line is presented in section III. Then, a comparison between the two methods is exposed in section IV. Finally, the discussions and conclusions are presented in section V and VI.

II. BACKGROUND

A. Rounding Method

This section reviews the QE, SQNR, correlation coefficient, regression equation of a line, and calculates RSS for the rounding method

From Fig. 1, it can be seen that the rang of QE of the sine analog signal is symmetric about zero. Furthermore, it is not resulting from a reduction of several significant bits as well as falls in the rang of all positive and negative numbers. The same experiment is achieved with the same values when the number of bits is 4, 8, and 16 [6]. Fig. 1. shows the original analog sine signal (**X2**), Quantization (**Xq**) and QE (**Xe**) using the rounding

method for some bits equal to 2 (n=2).

$$\left(-0.25 = -\frac{0.5}{2} = -\frac{\Delta}{2} \le QE\ Rounding \le \frac{\Delta}{2} = \frac{0.5}{2} = 0.25\right)$$
(1)

The investigation of the regression equation and the correlation coefficient (r) of the SQNR on the quantity of bits by means of rounding methods.
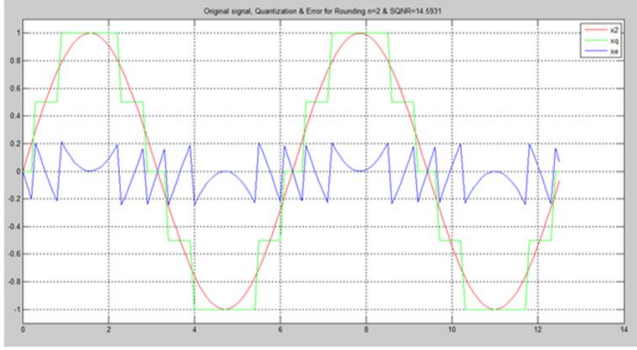

Fig. 1. *X2, Xq* and *Xe* using rounding method for (n=2)

The (r) states that how the data is close in a scatter-plot reduction sideways a straight line. The value of (r) which is nearer to 1 means the nearer the data is described by a linear equation. When the (r) is equal to 1, this means that the data set is faultlessly related, however once the (r) is close to zero, this means that the data set shows no straight-line relationship [6].

$$r_R = \frac{S_{ne}}{\sqrt{S_{nn}S_{ee}}}$$
(2)

$$r_R = \frac{1004.9884}{\sqrt{168 \times 6013.091}} = \frac{1004.9884}{1005.0866} = 0.99999022$$

The value of the r is (0.99999022) which means that it has a very strong relationship.

The equation of the regression line in general is:

$$e = a + bn$$
(3)

Where (bR) is the slope of the line, and (aR) is the intercept

$$b_R = \frac{S_{ne}}{S_{nn}} = \frac{1004.9884}{168} = 5.9820738$$

$$a_R = \bar{e} - b\bar{n} = 56.109362 - 5.9820738\ x\ 9$$
$$= 56.109362 - 53.838664 = 2.270698$$

Experimental rounding

$$SQNRR = aR + bR\ *n$$
(4)

$$SQNR_R = 2.270698 + 5.9820738 * n$$

Theoretical for sinusoidal signal

$$SQNR_{Theo} = 1.76 + 6.02 * n$$
(5)

Calculating RSS for the rounding method

To calculate RSS for the rounding method, as shown in Table I,**Error! Reference source not found.** according to the following formula (formula 6) which is taken from [7].

$$RSS = [SQNR_R - SQNR_{THeo}]^2$$
(6)

$$RSS = 1.64713551$$

Where

SQNR_R is the experimental rounding observed value

SQNR_THeo. is the corresponding theoretical value

TABLE I. CALCULATING RSS FOR ROUNDING METHOD

RSS = [SQNR Exp.(R) - SQNR (Theo)]2	SQNR Exp. Round	SQNR Theo.	No. of bits
0.628849	14.593	13.8	2
0.130321	26.201	25.84	4
3.61E-06	37.8781	37.88	6
0.02053489	49.7767	49.92	8
0.04687225	61.7435	61.96	10
0.74200996	74.8614	74	12
0.07817616	85.7604	86.04	14
0.00036864	98.0608	98.08	16
RSS = 1.64713551			

B. Truncating method

This section reviews the QE, SQNR, correlation coefficient, regression equation of a line, and calculates the RSS for truncating method. Fig. 2. illustrates the original sine analog signal (*X2*), Quantization (*Xq*) and QE (*Xe*) using Truncating for a number of bits (n=2) which is taken from [3].
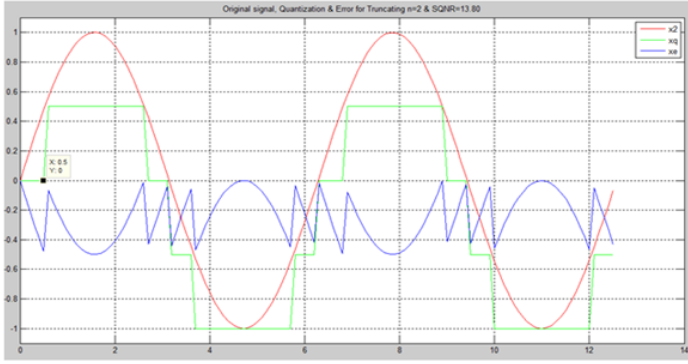
Fig. 2. *X2, Xq*, and *Xe* using Truncating for (n=2)

From Fig. 2., the rang of QE using the truncating method of the sine analog signal is not symmetric about zero, and it is a result of a reduction of a number of important bits as well as falls in the rang of all negative numbers. The same experiment for the truncating method is accomplished in [3] with the same values when the number of bits is 4, 8, and 16.

$$(-\Delta = -0.5 \leq QE\ (\text{Truncating}) \leq 0) \qquad (7)$$

The correlation coefficient for truncating (rT)

$$rT = \frac{S_{ne}}{\sqrt{S_{nn}S_{ee}}} \qquad (8)$$

$$rT = \frac{1015.41}{\sqrt{168 \times 6137.507}} = \frac{1015.41}{1015.4314} = 0.9999789$$

The value of the correlation coefficient is (0.9999789) which means it has a strong relationship.

The regression equation of a line for truncating

$$b_T = \frac{S_{ne}}{S_{nn}} = \frac{1015.41}{168} = 6.0441071$$

$$a_T = \bar{e} - b\bar{n} = 49.827237 - 6{,}0441071\ x\ 9$$
$$= 49.827237 - 54.396963 = -4.569726$$

Experimental Truncating

$$SQNR_T = a_T + b_T * n \qquad (9)$$
$$SQNR_T = -4.569726 + 6.0441071 * n$$

TABLE II. CALCULATING RSS FOR THE TRUNCATING METHOD

RSS = [SQNR Ex. Tr- SQNR Theo]²	SQNR Ex. Tr	SQNR Theo.	No. of Bits
38.49582025	7.5955	13.8	2
41.78458881	19.3759	25.84	4
39.08625361	31.6281	37.88	6
34.37594161	44.0569	49.92	8
38.85776896	55.7264	61.96	10
34.07173641	68.1629	74	12
34.56616849	80.1607	86.04	14
38.05039225	91.9115	98.08	16
RSS = 299.2886704			

From Table II RSS for the truncating (RSS)$_T$ is

$$RSS_T = 299.2886704$$

III.   REGRESSION EQUATION COMPARISONS

A comparison between regression equations of lies for equations (4, 5, and 9) which are mentioned and detailed in the above sections are presented. The comparison is accomplished using a multi-paradigm numerical computing environment MATLAB.

```
%n=Number of Bits
<<n=2:2:16 ;
%S_th=Signal Quantization to Noise Ratio Theory
<<S_th=1.76 + 6.02*n            ;
<<plot(n,S_th,'r');
<<grid on; >> hold on;
%S_R=Signal Quantization to Noise Ratio   Experimental using
Rounding Method
<<S_R=2.2702 + 5.98207*n    ;
<<plot(n,S_R,'b'); >> hold on;
%S_T=Signal Quantization to Noise Ratio   Experimental using
Truncating Method
<<S_T=-4.56972 + 6.04411*n;
<<plot(n,S_T,'g');
<<xlabel('n'); >> ylabel('SQNR');
<<legend('S_th,r','S_R,b','S_T,g');
<<title('SQNR for Theory, Experimental Rounding, Experimental
Truncating');
```

The lines in Fig. 3 are drawn from the above code which presents a comparison among the three lines from equations 4, 5 and 9.
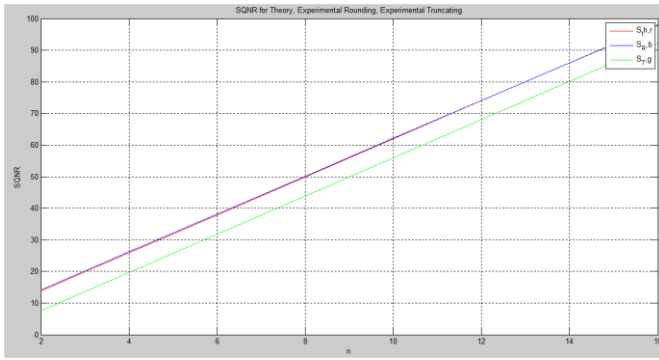
Fig. 3. Lines drawn from equations (4, 5, and 9)

Similarly, by using MATLAB program as shown in Fig. 3., as the quantity of bits amplified in the digital world, the experimental and theoretical SQNR of the sine signal using rounding method is very similar, but as using the truncating method, the experimental SQNR value was always lower than the theoretical value, which leads to less accurate quantization of the signal and a lower SQNR.

IV. COMPARISON OF ROUNDING AND TRUNCATING METHODS

In this section, a detailed comparison for rounding and truncating methods for sine signal is presented. The precision of the model decreases when the standard deviation and mean absolute deviation increase, it means using rounding will be more accurate as shown in Table III. The R-Squared statistic indicates that the model as fitted explains that (99.71793519%) of the variability in SQNR using a rounding method while (99.70530110%) using the truncating method. The value of the correlation coefficient for rounding is (0.99999022) this means that it has a very strong relationship, while for the truncating method (0.99997890) has only a strong relationship.

TABLE III. COMPARISON OF ROUNDING AND TRUNCATING METHODS

		Rounding	Truncating
1	Rang of QE	$(-\frac{\Delta}{2} \leq QE \leq \frac{\Delta}{2})$	$(-\Delta = -0.5 \leq QE \leq 0)$
2	Standard Diversion Error (SD)	0.03720387388	0.08552677275
3	Mean Absolute Diversion Error	0.02485336575	0.057049666
4	R-square (R²) Error	99.71793519 %	99.7053011 %
5	SEE	0.00213459133	0.005014927716
6	Coefficient of Variation (CV) % Error	11.66849539 %	12.06942713 %
7	Correlation Coefficient (r)	0.99999022	0.9999789

8	The Equation of Regression a ine	SQNR = 2.270698 + 5.9820738 * n	SQNR = −4.569726 + 6.0441071 * n
9	RSS	1.64713551	299.2886704

The Standard Error of Estimate (SEE) using the truncating method is (0.005014927716). This can be confirmed, as the inverse relationships between R-squared and SEE. In other words, the precision of the model increases as the SEE decreases.

The SEE and coefficient of variation of the mean deviation of the quantization error using the rounding method are lower than using the truncating method for the same signal. This means that the rounding method gives more accurate results.

The most appropriate statistics for making a comparison between these two methods is to calculate the RSS for each method separately. From the calculations, the RSS using the rounding method is much less than the RSS in Table II, using the truncating method. This means that as the value of RSS decreases the accuracy of the method increases, and vice versa.

V. DISCUSSION

In an ideal ADC, where the quantization error is uniformly distributed and the signal has a uniform distribution covering all quantization levels, the SQNR can be calculated as follow:

$$SQNR = 6.02*n \tag{10}$$

The most common test signals that complete this is the full amplitude triangle signals [8]. for example, a 16-bit ADC has a maximum SQNR of $6.02 \times 16 = 96.3$ dB.

But when the input signal is a full-amplitude, the sine wave distribution of the signal is no longer uniform distribution and does not cover all quantization level, and the corresponding equation (5) is given

$$SQNR = 1.76 + 6.02*n$$

The (1.76) difference in SQNR only occurs due to the sine wave signal being a full-scale, and again when the input signal has high amplitude and a wide frequency spectrum, the SQNR is assumed to be uniformly distributed. In this case, a 16-bit ADC has a maximum SQNR of 98.09 dB.

For the rounding method, the corresponding equation is (4)

$$SQNR_R = 2.270698 + 5.9820738 * n$$

The 2.270698 difference in SQNR occurs due to a full-scale sine wave signal. A 16-bit ADC has a maximum SQNR of 97.9838788 dB

For the truncating method, the corresponding equation is (9)

$$SQNR_T = -4.569726 + 6.0441071 * n$$

The (-4.569726) difference in SQNR occurs also due to a full-scale sine signal, and a 16-bit ADC has a maximum SQNR of 92.1359876 dB

The differences in these equations have not occurred, if the input sine wave signal has a uniform distribution and covering all quantization levels, as in equation (10)

VI. CONCLUSION

By using the STATGRAPHICS PROGRAM, and MATLAB programs, the value of SQNR using the truncating method is always lower than the theoretical and rounding values, which leads to less accurate quantization of the sine signal.

The designed filter for truncating method is running to remove the 16-bits least significant bit (LSB) and to keep the most significant 16-bits (MSB) of a 32- bit in the digital world, this means that ignoring the lower 16-bits of the digital word of the output signals and take only the upper 16 bits.

The problem with the truncating method is that it makes no use of the information contained in the LSB, which gives us a smarter approach to round the upper 16-bits either up or down 1-bit based on the lower 16-bits. Selecting of the 16-bits gives a more accurate sign of the original word and also eliminates bias that is a DC component in our signal spectrum.

The advantage of the truncating method over the rounding method is that designing a microprocessor's filter for the truncating will be smaller in size, speedily and less latency, because fewer electric circuits needed to build.

REFERENCES

[1] A. V. Oppenheim, "Applications of digital signal processing," *Englewood Cliffs NJ Prentice-Hall Inc 1978 510 P*, 1978.

[2] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 139–149, 1982.

[3] A. Hasso, K. Jacksi, and K. Smith, "Effect of Quantization Error and SQNR on the ADC Using Truncating Method to the Nearest Integer Bit," presented at the 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 112–117.

[4] S. Pavan, R. Schreier, and G. C. Temes, *Understanding delta-sigma data converters*. John Wiley & Sons, 2017.

[5] B. Widrow and I. Kollár, "Quantization noise," *Camb. Univ. Press*, vol. 2, p. 5, 2008.

[6] A. Hasso and S. Mohammed, "The Effect of ADC Quantization Error by rounding to nearest integer bit," in *1st international conference on Engineering and innovative Technology, SU-ICETT*, Salahaddin University, 2016.

[7] W. M. Mendenhall and T. L. Sincich, *Statistics for Engineering and the Sciences*. CRC Press, 2016.

[8] K. C. Pohlmann, *Principles of digital audio*. McGraw-Hill, Inc., 1995.