RADIO SCIENCE, VOL. 42, RS3022, doi:10.1029/2006RS003585, 2007

Convenient formulas for quantization efficiency

A. R. Thompson,[1] D. T. Emerson,[2] and F. R. Schwab[1]

Received 6 November 2006; revised 18 January 2007; accepted 5 February 2007; published 19 June 2007.

[1] Digital quantization of signals prior to processing results in the insertion of a component of noise resulting from the finite number of quantization levels. In radio astronomy, for example, this is important because the number of levels tends to be limited by increasing sample rates, required by the use of increasingly wide bandwidths. We are here concerned with signals with Gaussian amplitude distribution that are processed by cross correlation. Quantization efficiency is the relative loss in signal-to-noise ratio resulting from the quantization process. We provide a method of calculating the quantization efficiency for any number of uniformly spaced levels, as a function of the level spacing, using formulas that are easily evaluated with commonly used mathematical programs. This enables a choice of level spacing to maximize sensitivity or to provide a compromise between the sensitivity and the voltage range of the input waveform.

Citation: Thompson, A. R., D. T. Emerson, and F. R. Schwab (2007), Convenient formulas for quantization efficiency, *Radio Sci.*, *42*, RS3022, doi:10.1029/2006RS003585.

1. Introduction

[2] The process of digital quantization of an analog waveform involves addition of an error component resulting from the finite number of bits in the digital representation. There is thus an increase in the uncertainty of any measurement using the digitized waveform, and hence a degradation of the signal-to-noise ratio (sensitivity) of the measurement. An early general analysis of quantization effects is given by *Bennett* [1948]. In radio astronomy, for example, cross correlations are formed of the received waveforms from spaced antennas or from a single antenna with different time delays. The probability distribution of the analog waveforms is close to Gaussian and the signal-to-noise ratio of the digitally correlated data, as a fraction of that for an equivalent ideal analog system, is referred to as the quantization efficiency η_Q. Expressions for η_Q for three- and four-level quantization in radio astronomy were first given by *Cooper* [1970]: for later derivations, see, e.g., the discussion by *Thompson et al.* [2001] and associated references. With advances in digital electronics, use of larger numbers of levels has become practical, and some analysis of performance in such cases is given by *Jenet and Anderson* [1998]. In deriving general expressions for η_Q we have used an approach which is based on the consideration that the quantization efficiency is equal to the variance of the original analog noise voltage divided by the equivalent noise variance of the digitized signal at one input of a correlator. Note that this applies to the situation in which the cross correlation of the signals at the two correlator inputs tends toward zero, which is generally the important case in radio astronomy. For this condition, we derive exact expressions for any number of levels using formulas that can easily be evaluated using widely available mathematical programs. Approximate expressions for η_Q for eight and higher numbers of levels are given by *Thompson et al.* [2001, pp. 273–276]. However, the quantization inequality (x-y) is used as an approximation for the quantization noise; that is, in effect α (see section 2) is taken to be 1. This is a useful approximation if the number of quantization levels is not too small, but it is now possible to provide exact expressions by using α_1 to select the random component.

2. Derivation of the Formulas

[3] Let x represent the voltage of the signal at the quantizer input. In radio astronomy such waveforms generally have a Gaussian probability distribution with variance σ^2. Let y represent the quantized values of x. The difference x-y represents an inequality introduced by the quantization. The inequality contains a component that is correlated with x, and an uncorrelated component that behaves as random noise. To separate these, consider the correlation coefficient between x and

[1]National Radio Astronomy Observatory, Charlottesville, Virginia, USA.
[2]National Radio Astronomy Observatory, Tucson, Arizona, USA.

0048-6604/07/2006RS003585$11.00

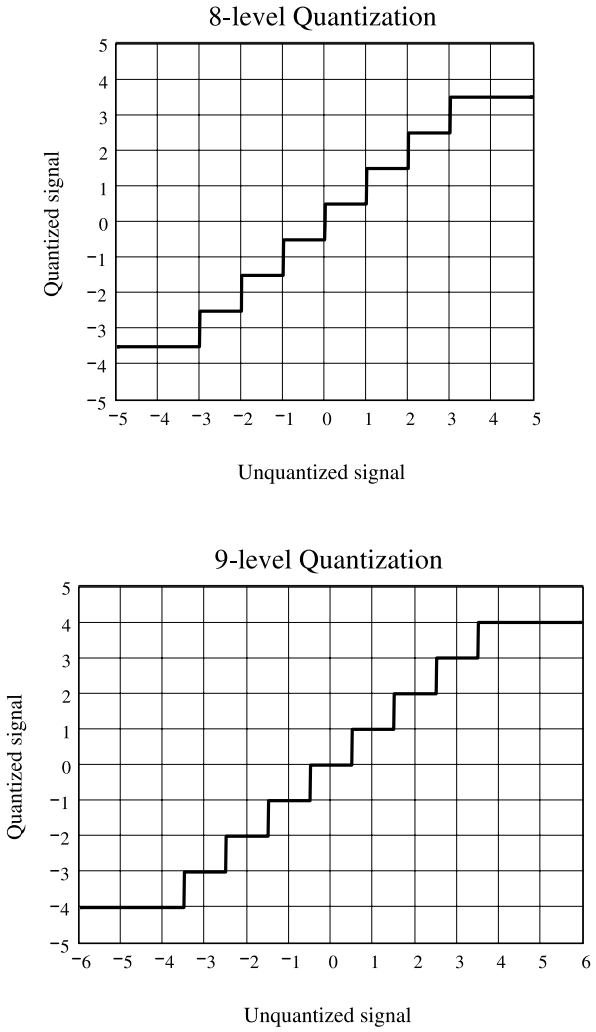

Figure 1. Examples of quantization characteristics with (top) an even number of levels (eight) and (bottom) an odd number of levels (nine). In each case, $\mathcal{N} = 4$. Units on both axes are equal to ϵ. The vertical sections of the staircase functions represent the thresholds between the levels. Note that for even numbers of levels, the thresholds occur at integral values on the abscissa, whereas for odd numbers of levels, the thresholds occur at values that are an integer $\pm\frac{1}{2}$.

$\Delta = x - \alpha y$, where α is a scaling factor. The correlation coefficient is

$$\frac{\langle x\Delta\rangle}{x_{rms}\Delta_{rms}} = \frac{\langle x^2\rangle - \alpha\langle xy\rangle}{x_{rms}\Delta_{rms}}. \tag{1}$$

[4] Here the angle brackets $\langle\ \rangle$ indicate the mean value. If $\alpha = \langle x^2\rangle/\langle xy\rangle$, then the correlation coefficient is zero, and Δ represents purely random noise. We refer to this random component as the quantization noise, q, equal to $x - \alpha_1 y$ where $\alpha_1 = \langle x^2\rangle/\langle xy\rangle$. Note that for each sample, x and y have the same sign so xy is always positive. Without loss of generality, we take $\sigma^2 = \langle x^2\rangle = 1$ and use $\alpha_1 = 1/\langle xy\rangle$. Thus the variance of the quantization noise is

$$\langle q^2\rangle = \langle (x - \alpha_1 y)^2\rangle = \alpha_1^2\langle y^2\rangle - 1, \tag{2}$$

and since the total variance of the digitized signal is $1 + \langle q^2\rangle$, we obtain

$$\eta_Q = \frac{1}{(1 + \langle q^2\rangle)} = \frac{1}{\alpha_1^2\langle y^2\rangle} = \frac{\langle xy\rangle^2}{\langle y^2\rangle}. \tag{3}$$

[5] As a verification of this result, consider the case of two-level quantization, which was particularly important in early radio astronomy correlators. Here y is assigned the value of 1 when $x > 0$, and -1 when $x < 0$. Thus $\langle y^2\rangle = 1$ and $\langle xy\rangle = \langle |x|\rangle$. Then we have

$$\langle |x|\rangle = \frac{2}{\sqrt{2\pi}}\int_0^\infty x\, e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}}, \tag{4}$$

and from equation (3) $\eta_Q = 2/\pi$, which is a well-known result that follows from a study by *Van Vleck and Middleton* [1966]. These authors also give the correction for linearity that can be applied to the autocorrelation or cross-correlation terms produced after quantization with two levels, and similar corrections for larger numbers of levels have been derived elsewhere. These corrections will scale the RMS level by the same factor as the signal within a given correlation term, so the resultant signal-to-noise ratio is unaffected by the linearity correction. An interesting historical detail concerning this reference is that the work was done during World War II and described in Radio Research Laboratory Report 51 of Harvard University, dated 1943, at which time it was classified.

[6] To apply equation (3) to cases with larger numbers of levels we need general expressions for $\langle xy\rangle$ and $\langle y^2\rangle$. Values of xy and y^2 are determined by the sample values of x, so the mean values over many samples can be expressed in terms of the Gaussian probability function of x. We consider only cases in which the spacing between adjacent quantization thresholds is constant, and begin with even numbers of levels as in the 8-level case in Figure 1. We Define ϵ (measured in units of σ) as the spacing in the x coordinate between adjacent level thresholds, and $\mathcal{N}$ as half the number of levels. We first determine $\langle xy\rangle$. The values of x that fall within the quantization level between $m\epsilon$ and $(m + 1)\epsilon$ are assigned

Table 1. Values of ϵ and η_Q for Several Numbers of Levels

Number of Levels	$\mathcal{N}$	ϵ	η_Q
2			0.636620
3	1	1.224	0.809826
4	2	0.995	0.881154
8	4	0.586	0.962560
9	4	0.534	0.969304
16	8	0.335	0.988457
32	16	0.188	0.996505
64	32	0.104	0.998960
128	64	0.0573	0.999696
256	128	0.0312	0.999912

the value $y = (m + 1/2)\epsilon$. (Since the digitized values are specified in units of ϵ, choice of ϵ introduces a gain factor, but this does not affect the signal-to-noise ratios with which we are concerned.) The contribution to $\langle xy \rangle$ from this level is

$$\left(m+\frac{1}{2}\right)\epsilon\frac{1}{\sqrt{2\pi}}\int_{m\epsilon}^{(m+1)\epsilon} x\,e^{-x^2/2}\,dx. \tag{5}$$

[7] The contribution from the level between $-m\epsilon$ and $-(m + 1)\epsilon$ is the same as the expression above, so to obtain $\langle xy \rangle$ we sum the integrals for the positive levels and include a factor of two:

$$\langle xy\rangle = \sqrt{\frac{2}{\pi}}\left[\left(\sum_{m=0}^{\mathcal{N}-2}\left(m+\frac{1}{2}\right)\epsilon\int_{m\epsilon}^{(m+1)\epsilon} xe^{-x^2/2}\,dx\right) + \left(\mathcal{N}-\frac{1}{2}\right)\epsilon\int_{(\mathcal{N}-1)\epsilon}^{\infty} xe^{-x^2/2}\,dx\right]. \tag{6}$$

[8] The summation term contains one integral for each positive quantization level except the highest one. The integral on the lower line covers the range of x above the highest threshold, for which the assigned value is $y = (\mathcal{N} - 1/2)\epsilon$. Then since $\int xe^{-x^2/2}dx = -e^{-x^2/2}$, equation (6) reduces to

$$\langle xy\rangle = \sqrt{\frac{2}{\pi}}\epsilon\left(\frac{1}{2}+\sum_{m=1}^{\mathcal{N}-1} e^{-m^2\epsilon^2/2}\right). \tag{7}$$

[9] To evaluate the variance of y, again consider first the contribution from values of x that fall between $m\epsilon$ and $(m + 1)\epsilon$. The variance of y for all values of x within this level is

$$\left(m+\frac{1}{2}\right)^2\epsilon^2\frac{1}{\sqrt{2\pi}}\int_{m\epsilon}^{(m+1)\epsilon} e^{-x^2/2}\,dx. \tag{8}$$

[10] For negative x we again include a factor of 2, sum over all positive quantization levels below the highest threshold, and add a term for the range of x above the highest threshold. Thus the total variance of y is

$$\langle y^2\rangle = \sqrt{\frac{2}{\pi}}\left[\left(\sum_{m=0}^{(\mathcal{N}-2)}\left(m+\frac{1}{2}\right)^2\epsilon^2\int_{m\epsilon}^{(m+1)\epsilon} e^{-x^2/2}\,dx\right) + \left(\mathcal{N}-\frac{1}{2}\right)^2\epsilon^2\int_{(\mathcal{N}-1)\,\epsilon}^{\infty} e^{-x^2/2}\,dx\right]. \tag{9}$$

[11] The right-hand side of equation (9) can be simplified by expressing the integrals in terms of the error function erf(): $\mathrm{erf}(\xi/\sqrt{2}) = \sqrt{2/\pi}\int_0^{\xi}\exp(-t^2/2)\,dt$. Then, using equations (3) and (7), we obtain

$$\eta_{Q(2\mathcal{N})} = \frac{\frac{2}{\pi}\left(\frac{1}{2}+\sum_{m=1}^{\mathcal{N}-1} e^{-m^2\epsilon^2/2}\right)^2}{\left(\mathcal{N}-\frac{1}{2}\right)^2-2\sum_{m=1}^{\mathcal{N}-1} m\,\mathrm{erf}\left(\frac{m\epsilon}{\sqrt{2}}\right)}. \tag{10}$$

[12] For the case where the number of levels is odd the thresholds occur at values that are an integer $\pm\frac{1}{2}$, as in the 9-level case in Figure 1. Values of x that fall within the quantization level between $(m - \frac{1}{2})\epsilon$ and $(m + \frac{1}{2})\epsilon$ are assigned the quantized value $m\epsilon$. We represent the odd level number by $2\mathcal{N} + 1$. Then following the steps as outlined for the even-number levels we obtain

$$\eta_{Q(2\mathcal{N}+1)} = \frac{\frac{2}{\pi}\left(\sum_{m=1}^{\mathcal{N}} e^{-\left(m-\frac{1}{2}\right)^2\epsilon^2/2}\right)^2}{\mathcal{N}^2-2\sum_{m=1}^{\mathcal{N}}\left(m-\frac{1}{2}\right)\mathrm{erf}\left(\frac{\left(m-\frac{1}{2}\epsilon\right)}{\sqrt{2}}\right)}. \tag{11}$$

3. Results

[13] For even and odd numbers of levels equations (10) and (11), respectively, provide values of η_Q from starting values of ϵ and $\mathcal{N}$. They can be evaluated rapidly in Mathcad, Mathmatica or similar programs. Examples of results derived are shown in Table 1. Values of ϵ are in units of σ and are chosen empirically to maximize η_Q. Curves showing η_Q as a function of ϵ are shown in Figure 2. As $\epsilon \rightarrow 0$ the output of the quantizer depends only on the sign of the input, so the curves meet the ordinate axis at the two-level value of η_Q, $2/\pi$. As ϵ increases, more of the higher (positive and negative) levels contain only values in the extended tails of the Gaussian distribution, so the number of levels that make

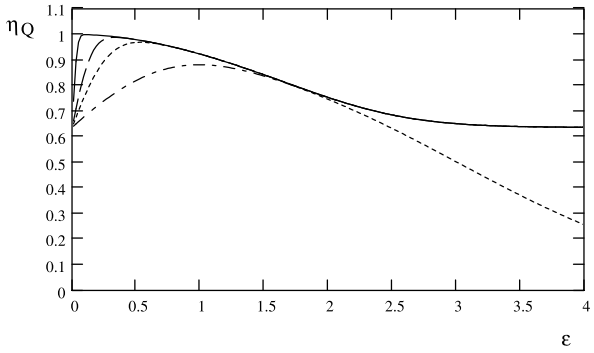

Figure 2. Quantization efficiency as a function of threshold spacing in units of σ. The curves are for 64-level (solid curve), 16-level (long-dashed curve), 9-level (short-dashed curve), and 4-level (long-and-short-dashed curve) quantization. As $\epsilon \rightarrow 0$, the output of the quantizer depends only on the sign of the input, so the curves meet the ordinate axis at the two-level value of η_Q, $2/\pi$. As ϵ increases, more of the higher (positive and negative) levels contain only values in the extended tails of the Gaussian distribution, so the number of levels that make a significant contribution to the output decreases, and the curves merge together. The curves for even level numbers move asymptotically to the two-level value, and curves for odd level numbers move toward zero.

a significant contribution to the output decreases, and the curves merge together. The curves for even level numbers move asymptotically to the two-level value, and curves for odd level numbers move toward zero. In situations where there are different numbers of levels used for the two inputs of a correlator, the output signal-to-noise ratio is proportional to the geometric mean of the input signal-to-noise ratios, i.e., to the geometric mean of the quantization efficiencies.

[14] If the constant voltage spacing between adjacent thresholds for both input and output values is not maintained, the individual levels can be adjusted to obtain an improvement in η_Q of a few tenths of a percent, decreasing with increasing number of levels. Level values optimized in this way are given by *Jenet and Anderson* [1998] for several numbers of levels. A highly detailed analysis of quantization effects which also includes threshold optimization is in preparation (F. R. Schwab, Optimal quantization functions for multi-level digital correlators, manuscript in preparation, 2007).

[15] In the analysis by *Jenet and Anderson* [1998] the assigned value for a signal that falls between adjacent level thresholds is equal to the RMS value of the corresponding Gaussian distribution between these thresholds. In the present case we want to be able to maintain linearity of response over voltage ranges which include non-Gaussian interfering signals (see section 4), and have used assigned values that are the mean of the threshold values between which the input voltage falls. This is also generally applicable to commercially available digital quantizers. Jenet and Anderson adjust the quantization parameters to minimize the RMS difference between the unquantized and quantized values of the input waveforms, whereas we have adjusted the spacing between thresholds, ϵ, to maximize the quantization efficiency. Differences in the results, however, are small and comparison of η_Q values in Table 1 with corresponding values by Jenet and Anderson shows that our value for 4 levels is 1.9% higher, but in other cases differences are only in the fourth or higher decimal places. Jenet and Anderson list values of a parameter l which is equal to $1 - \eta_Q$.

4. Choice of Level-Threshold Spacing

[16] Often the requirement for calculation of the quantization efficiency is simply to find the value of ϵ that provides the maximum sensitivity for a particular number of levels. In recent systems, however, ϵ is sometimes chosen so that signal voltages much higher than the RMS system noise can be accommodated within the range of the quantizer. This preserves an essentially linear response to interfering signals so that they can subsequently be mitigated by filtering or other processes. For example, with 256 levels (8-bit representation) and $\epsilon = 0.0312$ to maximize sensitivity, ± 128 levels corresponds to $\pm 4\sigma$, i.e., 6 dB above the RMS system level. However with $\epsilon = 0.25$, equation (10) shows that $\eta_Q = 0.9948$ and ± 128 levels then corresponds to 30 dB above the RMS level. Thus with 256 levels, a sacrifice of 0.5% in signal-to-noise ratio can permit an increase of 24 dB in the headroom above the interference-free power level. Such an arrangement is particularly useful at the lower frequencies used in radio astronomy observations where interference is common and bandwidths used are narrower allowing larger numbers of levels without incurring undesirably high bit rates.

[17] **Acknowledgment.** The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

References

Bennett, W. R. (1948), Spectra of quantized signals, *Bell Syst. Tech. J.*, *27*, 446–472.

Cooper, B. F. C. (1970), Correlators with two-bit quantization, *Aust. J. Phys.*, *23*, 521–527.

Jenet, F. A., and S. B. Anderson (1998), The effects of digitization on nonstationary stochastic signals with applications to pulsar signal baseband recording, *Publ. Astron. Soc. Pac.*, *110*, 1467–1478.

Thompson, A. R., J. M. Moran, and G. W. Swenson (2001), *Interferometry and Synthesis in Radio Astronomy*, 2nd ed., John Wiley, New York.

Van Vleck, J. H., and D. Middleton (1966), The spectrum of clipped noise, *Proc. IEEE*, *54*, 2–19.

D. T. Emerson, National Radio Astronomy Observatory, 949 North Cherry Ave., Campus Building 65, Tucson, AZ 85721 USA. (demerson@nrao.edu)

F. R. Schwab and A. R. Thompson, National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, VA 22903 USA. (fschwab@nrao.edu; athompso@nrao.edu)