Chapter 20

Spectrum of Quantization Noise and Conditions of Whiteness

When the input to a quantizer is a sampled time series represented by $x_1, x_2, x_3, \ldots,$ the quantization noise is a time series represented by $\nu_1, \nu_2, \nu_3, \ldots$ Suppose that the input time series is stationary and that its statistics satisfy the conditions for multivariable QT II (it would be sufficient that two-variable QT II conditions were satisfied for x_1 and x_2, x_1 and x_3, x_1 and x_4, and so forth, because of stationarity). As such, the quantization noise will be uncorrelated with the quantizer input, and the quantization noise will be white, i.e. uncorrelated over time. The PQN model applies. The autocorrelation function of the quantizer output will be equal to the autocorrelation function of the input plus the autocorrelation function of the quantization noise.

Fig. 20.1(a) is a sketch of an autocorrelation function of a quantizer input signal. Fig. 20.1(b) shows the autocorrelation function of the quantization noise when the PQN model applies. Fig. 20.1(c) shows the corresponding autocorrelation function of the quantizer output.

Corresponding to the autocorrelation functions of Fig. 20.1, the power spectrum of the quantizer output is equal to the power spectrum of the input plus the power spectrum of the quantization noise. This spectrum is flat, with a total power of $q^2/12$.

When it is known that the PQN model applies perfectly or otherwise applies to a very close approximation, one can infer the autocorrelation function and power spectrum of the quantizer input from knowledge of the autocorrelation function and power spectrum of the quantizer output, since the autocorrelation function and power spectrum of the quantization noise are known, these only need to be subtracted.

Spectral analysis of quantization is very simple when the quantization noise is white and uncorrelated with the quantizer input signal. We will present methods for determining the whiteness condition based on the multivariable characteristic function of the quantizer input. Other methods for doing this exist in the literature, and it is the purpose of this chapter to to explore these methods, and in some cases, to enhance them.

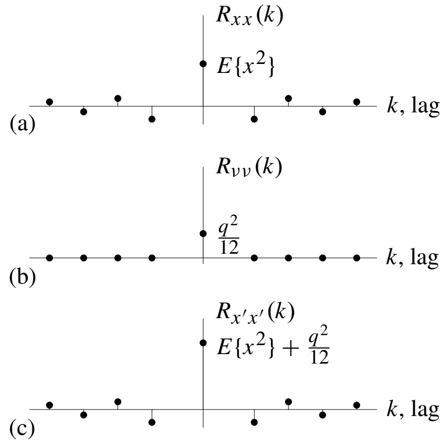

Figure 20.1 Autocorrelation functions when the PQN model applies: (a) input autocorrelation function, (b) quantization noise autocorrelation function, (c) output autocorrelation function.

20.1 QUANTIZATION OF GAUSSIAN AND SINE-WAVE SIGNALS

In this chapter the spectra of the quantized signal and of the quantization noise will be examined. For this, the correlation functions and the power spectral density (PSD) functions will be examined.

In order to obtain an overall impression about the spectral behavior of a quantized signal and quantization noise, let us consider two examples: quantization of a bandlimited Gaussian signal and of a sine wave.

Example 20.1 Quantization of a Bandlimited Gaussian Signal
A bandlimited Gaussian signal may be quantized by a rather rough quantizer, as illustrated in Fig. 20.2. The quantum size q as shown is approximately equal to the standard deviation σ of the input signal.

The most important difference between the shapes of $x(t)$ and that of $x'(t)$ and $\nu(t)$ is that the latter two contain many discontinuities. It is known from signal theory that discontinuities in the time function result in a relatively broad and slowly decaying spectrum.

The sparsely sampled version of the quantization noise $\nu(t_{i1})$ in Fig. 20.2(c) consists of seemingly independent (that is, uncorrelated) samples. Uncorrelated samples correspond to a white spectrum. Thus, if sampling is not too dense, whiteness may be a reasonable assumption.

Uncorrelatedness of the samples depends on the density of sampling. When sampling is denser, the samples are obviously correlated, as in Fig. 20.2(d).

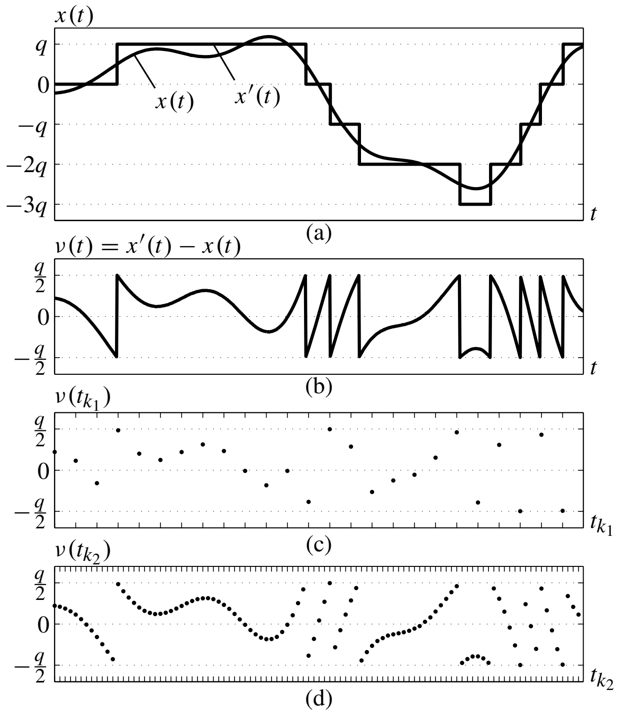

Figure 20.2 Quantization of a bandlimited Gaussian signal: (a) the input signal $x(t)$ and the quantized signal $x'(t)$; (b) the quantization noise $\nu(t)$; (c) sampled quantization noise, sparse sampling; (d) sampled quantization noise, dense sampling.

Example 20.2 Uniform Quantization of a Sine Wave
In Fig. 20.3 a sine wave of amplitude A is quantized using a quantum size $q = A/3.5$. It is clear that both $x'(t)$ and $\nu(t)$ are periodic, and, consequently, their spectra are discrete. Spectral broadening in the quantization noise means in this case that an infinite number of harmonics are produced, with a total power of about $q^2/12$.

Furthermore, though the samples of $\nu(t)$ do not exhibit a clear interdependence, periodicity yields a discrete spectrum, thus whiteness in the strict sense certainly does not hold. We will see instead that the power of the harmonics in a not too narrow frequency band, selected between 0 and the sampling frequency $f_s/2$, is more or less proportional to the bandwidth. This is also a kind of “whiteness”.

In this chapter the above observations are investigated in detail. First, the possibilities for an exact mathematical treatment will be discussed, then some results from the literature will be summarized. Either the power spectral density (PSD) function or the autocorrelation function will be examined. They contain the same information since they are Fourier transforms of each other:

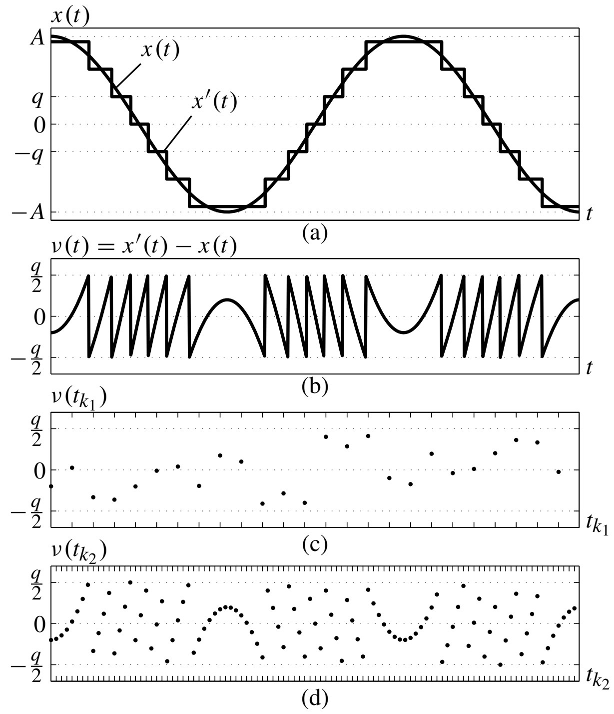

Figure 20.3 Quantization of a sine wave: (a) the input signal $x(t)$ and the quantized signal $x'(t)$; (b) the quantization noise $\nu(t)$; (c) sampled quantization noise, sparse sampling; (d) sampled quantization noise, dense sampling.

$$S(f) = \mathcal{F}\{R(\tau)\}\,. \tag{20.1}$$

After this, a more practical approach is used to develop simple approximate formulas for design purposes. The aim is to derive expressions that enable the estimation of the impact of quantization on signal spectra, in order to promote the proper design of measurement, control, and signal-processing systems.

20.2 CALCULATION OF CONTINUOUS-TIME CORRELATION FUNCTIONS AND SPECTRA

20.2.1 General Considerations

When dealing with spectra, in some derivations we deal with the PSD of $x'(t)$, in others with the PSD of $\nu(t)$. In most cases this is not of great consequence, since

$$R_{x'x'}(\tau) = \mathrm{E}\{x'(t)x'(t+\tau)\} = E\left\{\Big(x(t)+\nu(t)\Big)\Big(x(t+\tau)+\nu(t+\tau)\Big)\right\}$$
$$= R_{xx}(\tau) + R_{\nu\nu}(\tau) + \Big(R_{x\nu}(\tau) + R_{\nu x}(\tau)\Big) \tag{20.2}$$

and thus

$$S_{x'x'}(f) = S_{xx}(f) + S_{\nu\nu}(f) + \Big(S_{x\nu}(f) + S_{\nu x}(f)\Big)\,. \tag{20.3}$$

When x and ν are uncorrelated, and the mean value of ν is zero, that is, the PQN noise model is valid, the last term vanishes both in Eq. (20.2) and in Eq. (20.3), thus $S_{x'x'}(f)$ and $S_{\nu\nu}(f)$ depend only on each other, since $S_{xx}(f)$ is given. In the case of correlated x and ν, $S_{x'x'}(f)$ cannot be obtained by simple addition of $S_{xx}(f)$ and $S_{\nu\nu}(f)$.

Before going into details, two very general results will be presented. First, if the quantization noise is uniformly distributed, that is, either QT I or QT II or QT III/A is satisfied, the variance of the quantization noise, that is, the integral of its power spectral density is known, and depends on q only:

$$\int_{-\infty}^{\infty} S_{\nu\nu}(f)\,\mathrm{d}f = R_{\nu\nu}(0) = \mathrm{var}\{n\} = \frac{q^2}{12}\,. \tag{20.4}$$

Second, since $x'(t)$ contains a series of finite discontinuities ("jumps"), the envelope of its Fourier transform (if this exists) vanishes for $f \to \infty$ as $\mathcal{O}(1/f)$ (Bracewell, 1986, pp. 143–146). Similarly, the envelopes of $S_{x'x'}(f)$ and $S_{\nu\nu}(f)$ vanish as $\mathcal{O}(1/f^2)$.

Example 20.3 Power Spectral Density of a Random Bit Sequence
A random bit sequence is a stochastic process which may change its sign at equidistant time instants T (see Fig. 20.4). One can show that the autocorrelation

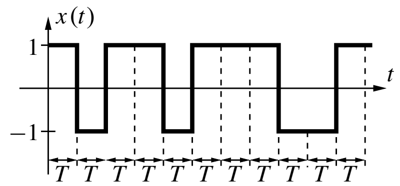

Figure 20.4 Random bit sequence.

function of the random bit sequence has the following form:

$$R_{xx}(\tau) = \begin{cases} \left(1 - \dfrac{|\tau|}{T}\right), & \text{if } |\tau| < T \\ 0 & \text{elsewhere.} \end{cases} \tag{20.5}$$

The power spectral density is the Fourier transform of the autocorrelation function:

$$S_{xx}(f) = \mathcal{F}\{R_{xx}(\tau)\} = T\left(\frac{\sin(\pi f T)}{\pi f T}\right)^2 . \tag{20.6}$$

$S_{xx}(f)$ vanishes like $\mathcal{O}(1/f^2)$, as it is expected.

Example 20.4 Power Spectral Density of a Random Telegraph Signal
The random telegraph signal is a binary (two-valued) stochastic process where the time instants of changes of the sign form a Poisson point process (that is, in any time interval the number of changes of the sign is a random variable with Poisson distribution). A time record is shown in Fig. 20.5. Let us determine the power spectral density function.

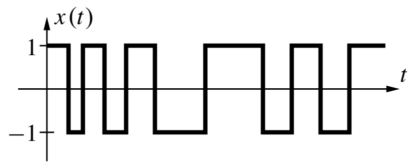

Figure 20.5 Random telegraph signal.

First the autocorrelation function will be calculated. The result of the multiplication of two samples can be ± 1:

- $x(t)x(t+\tau) = 1$ if in the interval τ the number of sign changes was even. The probability is $p_1 = \sum_{k=0}^{\infty} P_{2k}(\tau)$, where $P_{2k}(\tau)$ is the probability of having exactly $2k$ changes of sign in the interval τ.
- $x(t)x(t+\tau) = -1$ if in the interval τ the number of sign changes was odd. The probability is $p_{-1} = \sum_{k=0}^{\infty} P_{2k+1}(\tau)$.

From the above expressions,

$$\begin{aligned} R_{xx}(|\tau|) &= (1)\cdot p_1 + (-1)\cdot p_{-1} \\ &= \sum_{k=0}^{\infty}\left(\frac{(\lambda|\tau|)^{2k}}{(2k)!} - \frac{(\lambda|\tau|)^{2k+1}}{(2k+1)!}\right) e^{-\lambda|\tau|} \\ &= \sum_{n=0}^{\infty}(-1)^n \frac{(\lambda|\tau|)^n}{n!}\, e^{-\lambda|\tau|} \\ &= e^{-2\lambda|\tau|} . \end{aligned} \tag{20.7}$$

The PSD is the Fourier transform of the autocorrelation function:

$$S_{xx}(f) = \mathcal{F}\{R_{xx}(\tau)\} = \frac{4\lambda}{4\lambda^2 + (2\pi f)^2} \,. \tag{20.8}$$

The power spectral density function vanishes like $\mathcal{O}(1/f^2)$, as was expected from the jumps in the time record.

20.2.2 Direct Numerical Evaluation of the Expectations

It is possible to give mathematically correct general formulas for the correlation functions and the spectra. Let us introduce the following notation (see also Fig. 20.6):

- $Q(x)$ is the quantization characteristic: $x' = Q(x)$
- $Q_\nu(x)$ is the quantization noise characteristic: $\nu = Q_\nu(x)$
- $f(x_1, x_2, \tau)$ is the joint probability density function of $x(t)$ and $x(t+\tau)$, where $x(t)$ is a stationary random process.

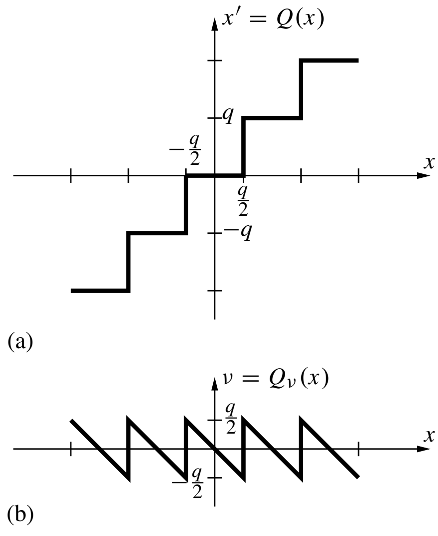

Figure 20.6 Quantizer characteristics of a rounding quantizer: (a) characteristic of the input–output quantizer; (b) characteristic which generates the quantization noise from the input.

Using this notation, the correlation functions are by definition

$$R_{x'x'}(\tau) = \mathrm{E}\left\{Q\Big(x(t)\Big)Q\Big(x(t+\tau)\Big)\right\} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} Q(x_1)Q(x_2)f(x_1, x_2, \tau)\,\mathrm{d}x_1\,\mathrm{d}x_2\,, \tag{20.9}$$

$$R_{\nu\nu}(\tau) = \mathrm{E}\left\{Q_\nu\Big(x(t)\Big)Q_\nu\Big(x(t+\tau)\Big)\right\} = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} Q_\nu(x_1)Q_\nu(x_2)f(x_1, x_2, \tau)\,\mathrm{d}x_1\,\mathrm{d}x_2\,. \tag{20.10}$$

The spectra can be obtained by Fourier transformation of the above expressions.

These expressions are theoretically correct. However, closed-form results usually do not exist because of the nonlinearity of Q and Q_ν. Therefore, their analytic evaluation can be troublesome.

Computer evaluation seems to be more hopeful because Q is piecewise constant and Q_ν is piecewise linear. The power of PQN (Eq. (20.4)) can be used to check the quality of approximations and numerical calculations. However, special care should be taken when evaluating Eq. (20.9) or Eq. (20.10) for $\tau = 0$: $f(x_1, x_2, \tau = 0)$ is a Dirac impulse sheet. Direct numerical computation may only be performed on the basis of Eq. (20.9) and Eq. (20.10) for $\tau \neq 0$, otherwise these equations must be modified as:

$$R_{x'x'}(0) = \int\limits_{-\infty}^{\infty} Q^2(x)f(x)\,\mathrm{d}x\,, \tag{20.11}$$

and similarly,

$$R_{\nu\nu}(0) = \int\limits_{-\infty}^{\infty} Q_\nu^2(x)f(x)\,\mathrm{d}x\,. \tag{20.12}$$

Nevertheless, since the computation is three-dimensional (two independent variables and the parameter τ), it may need much computing effort. Moreover, numerical results are generally appropriate for the purpose of analysis. To obtain usable results, restrictions (special distribution, special form of spectra, etc.) and approximations have to be introduced for well-defined cases.

20.2.3 Approximation Methods

Approximations are usually based on one of two key ideas.

Approximation Based on the Characteristic Function

The first one makes use of the fact that correlation, as a second-order joint moment, can be obtained from the joint characteristic function:

$$R(\tau) = \frac{1}{j^2} \frac{\partial^2 \Phi(u_1, u_2, \tau)}{\partial u_1 \partial u_2}\Bigg|_{u_1=u_2=0}, \tag{20.13}$$

where

$$\Phi(u_1, u_2, \tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, \tau)\, e^{j(u_1 x_1 + u_2 x_2)}\, \mathrm{d}x_1\, \mathrm{d}x_2 . \tag{20.14}$$

Equation (20.13) can be applied to the quantized signal or to the quantization noise (see Eqs. (8.23) and (9.4), respectively). From these expressions, the desired moments can be obtained for a rounding quantizer (Sripad and Snyder, 1977):

$$\begin{aligned}
R_{x'x'}(\tau) = R_{xx}(\tau)
&+ \frac{q}{2\pi j} \sum_{\substack{l_1=-\infty \\ l_1\neq 0}}^{\infty} \frac{\partial\, \Phi_{xx}\left(\frac{2\pi l_1}{q}, u_2, \tau\right)}{\partial u_2}\Bigg|_{u_2=0} \frac{(-1)^{l_1}}{l_1} \\
&+ \frac{q}{2\pi j} \sum_{\substack{l_2=-\infty \\ l_2\neq 0}}^{\infty} \frac{\partial\, \Phi_{xx}\left(u_1, \frac{2\pi l_2}{q}, \tau\right)}{\partial u_1}\Bigg|_{u_1=0} \frac{(-1)^{l_2}}{l_2} \\
&+ \frac{q^2}{4\pi^2} \sum_{\substack{l_1=-\infty \\ l_1\neq 0}}^{\infty} \sum_{\substack{l_2=-\infty \\ l_2\neq 0}}^{\infty} \Phi_{xx}\left(\frac{2\pi l_1}{q}, \frac{2\pi l_2}{q}, \tau\right) \frac{(-1)^{l_1+l_2+1}}{l_1 l_2},
\end{aligned} \tag{20.15}$$

and

$$R_{\nu\nu}(\tau) = \frac{q^2}{4\pi^2} \sum_{\substack{l_1=-\infty \\ l_1\neq 0}}^{\infty} \sum_{\substack{l_2=-\infty \\ l_2\neq 0}}^{\infty} \Phi_{xx}\left(\frac{2\pi l_1}{q}, \frac{2\pi l_2}{q}, \tau\right) \frac{(-1)^{l_1+l_2+1}}{l_1 l_2} . \tag{20.16}$$

Eqs. (20.15) and (20.16) usually cannot be given in closed form, but the higher-order terms in the infinite sums are often negligible (depending on the form of the CF), and an approximate closed form can be obtained.

Approximation Based on the Modulation Principle

The noise spectrum can directly be obtained as follows (Claasen and Jongepier, 1981). The quantization noise as a function of the signal amplitude (see Fig. 20.6(b)) can be developed into a Fourier series. For a rounding quantizer,

$$Q_\nu(x) = \frac{q}{\pi}\sum_{n=1}^{\infty}(-1)^n \frac{\sin\left(\frac{2\pi n x}{q}\right)}{n} \,. \tag{20.17}$$

Since $x = x(t)$ is a time function, Eq. (20.17) is a sum of phase-modulated sine waves. This phase modulation is usually wide-band, since the signal amplitude is generally larger than q. The spectrum can be approximated as follows. Each term can be expressed with the PDF of the derivative of the signal $x(t)$ (Claasen and Jongepier, 1981),

$$S_{\nu\nu}(f) \approx \frac{q^2}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{2n^2}\frac{f_{\dot{x}}\left(\frac{fq}{n}\right)\mathrm{d}\left(f\frac{q}{n}\right)}{\mathrm{d}f} \approx \frac{q^3}{2\pi^2}\sum_{n=1}^{\infty}\frac{f_{\dot{x}}\left(\frac{fq}{n}\right)}{n^3} \,. \tag{20.18}$$

The quality of approximation depends on the signal shape (Peebles, 1976, pp. 238–246).

20.2.4 Correlation Function and Spectrum of Quantized Gaussian Signals

For Gaussian signals several methods can be used. We will discuss five different approaches. We will need the PDF and CF of two jointly normal random variables, given in Appendix F.3.

Bennett's Direct Analysis

Bennett (1948) dealt with zero mean Gaussian signals. He accomplished a rather lengthy derivation for rounding quantizers, by directly evaluating Eq. (20.10). He obtained the following expression:

$$\begin{aligned} R_{\nu\nu}(\tau) = {} & \sigma_x^2\frac{\gamma}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}\, e^{-\frac{4n^2\pi^2}{\gamma}}\sinh\left(\frac{4n^2\pi^2\rho(\tau)}{\gamma}\right) \\ & + \sigma_x^2\frac{\gamma}{\pi^2}\sum_{\substack{n=1\\ n\neq m}}^{\infty}\sum_{m=1}^{\infty}\frac{1}{m^2-n^2}\, e^{-\frac{4(m^2+n^2)\pi^2}{\gamma}}\sinh\left(-\frac{4(m^2-n^2)\pi^2\rho(\tau)}{\gamma}\right) \end{aligned}$$

$$- \sigma_x^2 \frac{\gamma}{\pi^2} \sum_{\substack{n=1 \\ n \neq m}}^{\infty} \sum_{m=1}^{\infty} \frac{1}{(m-0.5)^2 - (n-0.5)^2} \, e^{-\frac{4\left((m-0.5)^2 + (n-0.5)^2\right)\pi^2}{\gamma}}$$
$$\times \sinh\left(- \frac{4\left((m-0.5)^2 - (n-0.5)^2\right)\pi^2 \rho(\tau)}{\gamma} \right), \qquad (20.19)$$

where $\rho(\tau) = R_{xx}(\tau)/R_{xx}(0)$ and $\gamma = q^2/\sigma_x^2$.

The complicated formula in Eq. (20.19) can be well approximated, noticing that $|\rho| \leq 1$, and usually $\gamma \ll 1$. The value of $R_{\nu\nu}(\tau)$ significantly differs from zero only if $|\rho| \approx 1$. Let us determine an approximation for $\rho \approx 1$:

$$R_{\nu\nu}(\tau) \approx \sigma_x^2 \frac{\gamma}{2\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} e^{-\frac{4n^2\pi^2(1-\rho)}{\gamma}} . \qquad (20.20)$$

Note that Eq. (20.20) gives the values of $R_{\nu\nu}(\tau)$ as a function of ρ (or of $R_{xx}(\tau)$). Its behavior is illustrated in Fig. 20.7, for $\sqrt{\gamma} = q/\sigma_x = 1/16$, exhibiting a strikingly rapid decrease as ρ moves away from 1. As a consequence of this behavior, the autocorrelation function of $\nu(t)$ will be much sharper at $\tau \approx 0$ than that of $x(t)$, and this corresponds to a much broader spectrum. We will come back to this observation later.

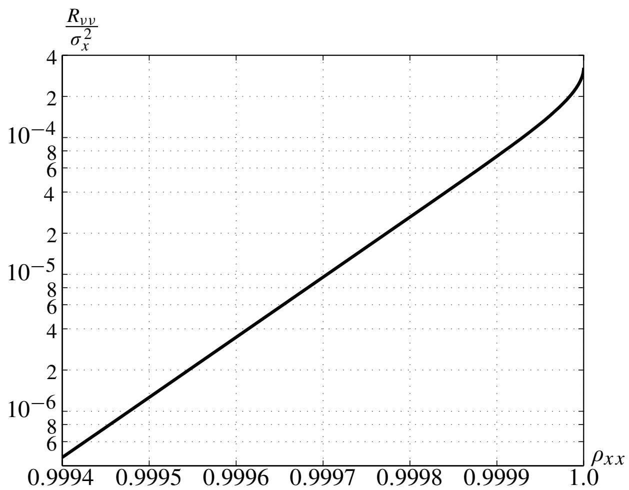

Figure 20.7 Correlation of quantization noises ("errors") as a function of the correlation coefficient of the signal samples. $2^7 q \approx 2 \cdot 4\sigma_x \rightarrow q \approx \sigma_x/16$. After Bennett (1948).

Example 20.5 Spectrum of the Quantization Noise of a Bandlimited Gaussian Signal
Equation (20.20) can be used to obtain the spectrum of the bandlimited white noise, substituting the first two terms of the power series expansion ($\tau \approx 0$ since $\rho \approx 1$),

$$\rho(\tau) = \frac{\sin(2\pi B\tau)}{2\pi B\tau} = 1 - \frac{(2\pi B\tau)^2}{3!} + \cdots , \tag{20.21}$$

where B is the bandlimit of the input signal.

The result is as follows:

$$S_{\nu\nu}(f) \approx \frac{q^2}{4\pi^3 B}\sqrt{\frac{3\gamma}{2\pi}}\,\eta\!\left(\frac{3\gamma f^2}{8\pi^2 B^2}\right), \tag{20.22}$$

where

$$\eta(y) = \sum_{n=1}^{\infty} \frac{e^{-y/n^2}}{n^3} . \tag{20.23}$$

Figure 20.8 illustrates the spectra obtained from Eq. (20.22) for some values of γ. The curves are parametrized by the bit number of the A/D converter: by setting the input range equal to $(-4\sigma_x, 4\sigma_x)$, the bit number can be expressed as $b = \log_2(8\sigma_x/q)$.

Application of the Characteristic Function Method

An alternative form of the correlation of the quantization noise was obtained for zero mean Gaussian signals by Sripad and Snyder (1977) by using Eq. (20.16):

$$\mathrm{E}\{\nu_1\nu_2\} = \frac{q^2}{\pi^2}\sum_{l_1=1}^{\infty}\sum_{l_2=1}^{\infty}\frac{(-1)^{l_1+l_2}}{l_1 l_2} e^{-2\pi^2\frac{\sigma_x^2}{q^2}(l_1^2+l_2^2)} \sinh\left(\frac{4\pi^2 l_1 l_2 \rho \sigma_x^2}{q^2}\right), \tag{20.24}$$

To prove the equivalence of formulas (20.19) and (20.24), consider the single sum in Eq. (20.19) as a special case of the first double sum with $m = 0$, and write both sums into a form where the indices run from $-\infty$ to ∞. By the following substitutions: in the first double sum $l_1 = m + n$, $l_2 = m - n$; in the second one $l_1 = m + n - 1$, $l_2 = m - n$, thus Eq. (20.24) is obtained. The equivalence of the different terms is illustrated in Fig. 20.9. The figure highlights the method by which Bennett (1948) summed the terms in his three sums, and shows why the sum of Sripad and Snyder (1977) is much simpler.

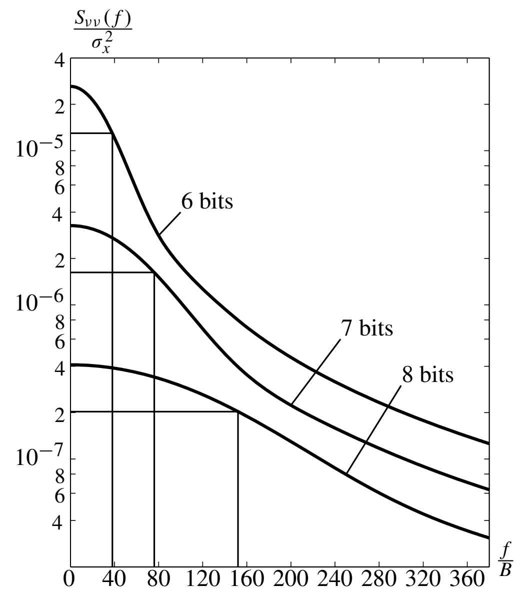

Figure 20.8 Quantization noise spectra of a bandlimited Gaussian signal (after (Bennett, 1948)). Frequency unit: bandwidth B of the original signal. Power unit: mean signal power (σ^2). The interval $(-4\sigma_x, 4\sigma_x)$ is equal to the input range of the A/D converter.

Analysis of the Autocorrelation Function of the Noise

Katzenelson (1962) obtained the same results by investigating the autocorrelation function. He made use of Eqs. (20.13) and (8.23), with the assumption that the conditions of QT II are approximately fulfilled. Accordingly,

$$R_{\nu\nu}(\tau) \approx R_{x'x'}(\tau) - R_{xx}(\tau) \tag{20.25}$$

may be used.[1]

By using the joint CF of the two-dimensional normal distribution (see Eq. (F.17)), he obtained Eq. (20.20) for $\rho \approx 1$. He suggested the following approximation for this expression:

$$R_{\nu\nu}(\tau) \approx \frac{q^2}{2\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} e^{-4n^2\pi^2(1-\rho(\tau))\frac{\sigma_x^2}{q^2}} \approx \frac{q^2}{12} e^{-4\pi^2(1-\rho(\tau))\frac{\sigma_x^2}{q^2}} . \tag{20.26}$$

[1] It is possible to make the first steps of his derivation exact, using Eq. (20.13) and (8.23). However, the infinite sum will need to be approximated in any event, and the final result will be the same.

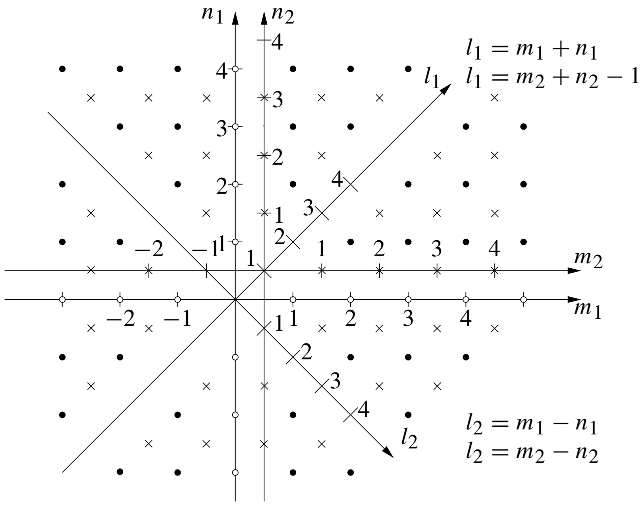

Figure 20.9 Illustration of the equivalence of the infinite sums (20.19) and (20.24): ∘ = terms of the simple sum in (20.19); • = terms of the first double sum in (20.19); × = terms of the second double sum in (20.19).

The last expression can be obtained by realizing that for $\rho \approx 1$ the exponentials in the first few terms of the sum in Eq. (20.26) are all approximately equal to one, and that the following sum converges rapidly:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} \,. \tag{20.27}$$

Series Representation of the Joint Normal PDF

For Gaussian signals there is still another possibility to compute spectra, based on (Amiantov and Tikhonov, 1956). Based on their work, Velichkin (1962) used the following series representation of the two-dimensional Gaussian probability density with zero mean:

$$f_{x_1,x_2}(x_1, x_2) = \sum_{k=1}^{\infty} \frac{\mathrm{d}^k f_{x_1}(x_1)}{\mathrm{d}x_1^k} \frac{\mathrm{d}^k f_{x_2}(x_2)}{\mathrm{d}x_2^k} \frac{\rho^k}{k!} \sigma_x^{2k} \,. \tag{20.28}$$

This formula can be proved e.g. by using Price's theorem (Papoulis, 1991, p. 161). From Eqs. (20.28) and (20.9), by using $\rho(\tau)^k \sigma_x^{2k} = R_{xx}^k(\tau)$, and assuming a stationary input signal ($f_{x_1}(x) = f_{x_2}(x) = f_x(x)$),

$$R_{x'x'}(\tau) = \sum_{k=1}^{\infty} \left(\int_{-\infty}^{\infty} Q(x) \frac{\mathrm{d}^k f_x(x)}{\mathrm{d}x^k} \,\mathrm{d}x \right)^2 \frac{R_{xx}^k(\tau)}{k!} = \sum_{k=1}^{\infty} \frac{A_k}{\sigma_x^{2k}} R_{xx}^k(\tau)\,, \qquad (20.29)$$

and

$$S_{x'x'}(f) = \sum_{k=1}^{\infty} \frac{A_k}{\sigma_x^{2k}} \mathcal{F}\left\{R_{xx}^k(\tau)\right\} = \sum_{k=1}^{\infty} \frac{A_k}{\sigma_x^{2k}} \mathcal{C}_{k-1}\left\{S_{xx}(f)\right\}\,, \qquad (20.30)$$

where $\mathcal{C}_k\{\,\}$ denotes the $(k-1)$th self-convolution.

This is a very interesting expression of the spectrum. The coefficients A_k do not depend on the spectrum (or the autocorrelation) of the input signal. This means that the spectral behavior of $S_{x'x'}(f)$ directly depends on the $\mathcal{F}\{R_{xx}^k(\tau)\}$ terms. When $k = 1$, the corresponding component is the original input spectrum multiplied by A_1/σ_x^2. It can be shown that

$$\lim_{q \to 0} \frac{A_1}{\sigma_x^2} = 1\,. \qquad (20.31)$$

For other values of n, the Fourier transform of $R_{xx}^n(\tau)$ corresponds to an $(n-1)$th self-convolution of $S_{xx}(f)$. That is, with the increase of n the spectrum is increasingly "smeared" along the frequency axis. This is related to the large bandwidth of the quantization noise.

Equation (20.30) can be used for the calculation of spectra, though its convergence is rather slow (Robertson, 1969). This convergence can be supervised on-line, realizing that on the one hand

$$R_{x'x'}(0) = \sum_{k=1}^{\infty} A_k \left(\frac{\mu_x^2 + \sigma_x^2}{\sigma_x^2} \right)^k\,, \qquad (20.32)$$

and on the other hand

$$R_{x'x'}(0) = \int_{-\infty}^{\infty} Q^2(x) f_x(x)\,\mathrm{d}x\,. \qquad (20.33)$$

Equation (20.33) can be quickly evaluated, and the result can be used for checking the convergence of the sum in Eq. (20.32). Another possibility for checking the convergence is to use

$$R_{x'x'}(0) \approx \mu_x^2 + \sigma_x^2 + \frac{q^2}{12}\,, \qquad (20.34)$$

which is a good approximation of the mean square value for $q < \sigma_x$.

Use of the Modulation Principle

Let us use Eq. (20.18) to obtain the quantization noise spectrum. The approximation is very good, since for bandlimited Gaussian signals the so-called RMS modulation index (Peebles, 1976, pp. 245–246), which is approximately equal in our case to $\beta_{\text{rms}} \approx 22\sigma_x/q$, fulfills

$$\beta_{\text{rms}} > 5\,. \tag{20.35}$$

This condition is amply fulfilled in practical cases.

The derivative of a Gaussian process is Gaussian too. The standard deviation can be obtained for the bandlimited white noise from the integral of the spectrum of the derivative:

$$\sigma_{\dot{x}}^2 = \int_{-\infty}^{\infty} S_{\dot{x}\dot{x}}(f)\,\mathrm{d}f = \int_{-\infty}^{\infty} (2\pi f)^2\, S_{xx}(f)\,\mathrm{d}f = \frac{4\pi^2}{3} B^2 \sigma_x^2\,. \tag{20.36}$$

By substituting the Gaussian probability density with parameter $\sigma_{\dot{x}}$ into Eq. (20.18) we obtain

$$S_{\nu\nu}(f) \approx \frac{q^3}{2\pi^2} \frac{1}{\sqrt{2\pi}\,\frac{2\pi}{\sqrt{3}} B\sigma_x} \sum_{k=1}^{\infty} \frac{1}{k^3}\, e^{-\frac{1}{2}\frac{\left(\frac{fq}{k}\right)^2}{\left(\frac{2\pi}{\sqrt{3}}\sigma_x B\right)^2}} \tag{20.37}$$

which is the same as Eq. (20.22). This is by no means surprising, since Peebles used the same second-order approximation of the autocorrelation function as we did when deriving Eq. (20.22).

20.2.5 Spectrum of the Quantization Noise of a Quantized Sine Wave

Quantization of sine waves presents a much more difficult problem to analyze than the quantization of Gaussian signals. The CF of a sinusoidal signal decreases slowly (see Appendix I.6, Fig. I.6), therefore the autocorrelation function and the spectrum cannot be approximated by just a few terms of the series representations. Moreover, the quantization noise is periodic having the same period length as the signal, so its spectrum is discrete, with a lot of harmonics.

Example 20.6 Autocorrelation Function of the Quantization Noise of a Sine Wave
The quantization noise of a sine wave can be analyzed numerically. In Fig. 20.10

the autocorrelation function is evaluated for $A = 4q$. Let us observe the irregular form of the autocorrelation function. This illustrates why is it difficult to handle analytically.

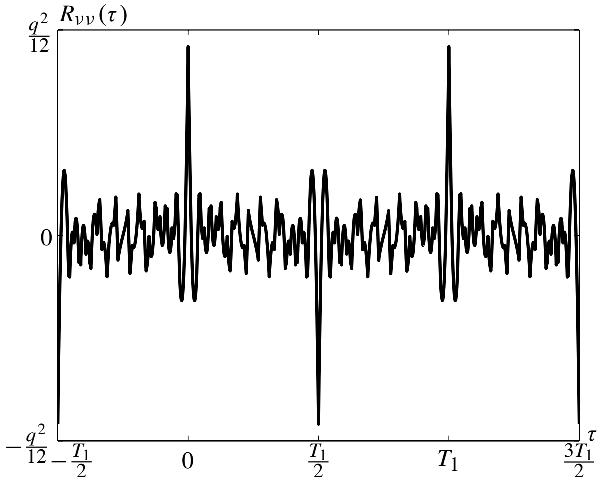

Figure 20.10 Calculated autocorrelation function of the quantization noise of a sine wave. $A = 4q$, $T_1 = 1/f_1$.

Example 20.7 Power Spectral Density of the Quantization Noise of a Sine Wave

The power spectral density function of the quantization noise of a sine wave can be analyzed using a computer, performing numerical Fourier series expansion.

The spectrum is full of harmonics, placed in a rather irregular manner. However, some strange "periodic" form shows up in the envelope, and this phenomenon can be described analytically, although providing a rough approximation only.

In the following considerations a reasonable approximation of the quantization noise spectrum will be obtained. Since the characteristic function vanishes very slowly, the approximation of Eq. (20.15) or (20.16) by the first terms will not work, and so the method of phase-modulated sine waves will be used instead.

The Fourier spectrum of a sine wave which is phase-modulated by another sine wave, is given in the literature (Peebles, 1976, p. 237). It consists of an infinite sum of harmonics, weighted with values of Bessel functions. For $x(t) = A\cos\left(2\pi f_0 t + \beta\cos(2\pi f_m t)\right)$,

$$S_{xx}(f) = \frac{A^2}{4}\sum_{k=-\infty}^{\infty} J_k^2(\beta)\left(\delta(f - f_0 - kf_m) + \delta(f + f_0 + kf_m)\right) . \qquad (20.38)$$

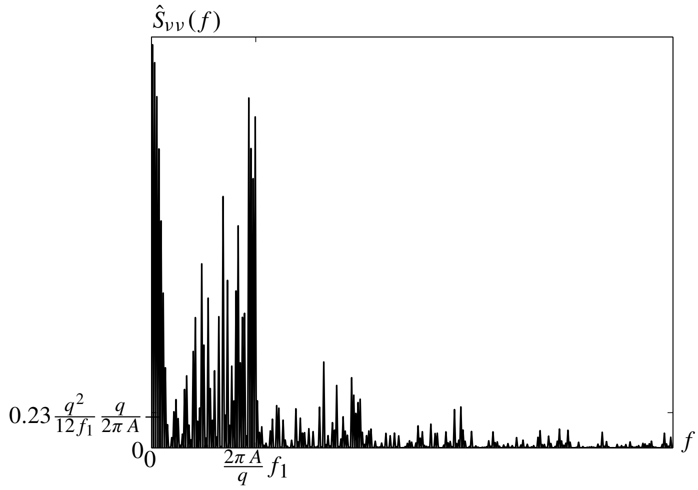

Figure 20.11 Numerical calculation of the power spectral density of the quantization noise of a sine wave. $A = 15.5q$. The mark on the vertical axis shows the value of the analytical approximation (Eq. (20.41), Fig. 20.13).

By using Eq. (20.17), a second summation is to be accomplished, and an absolute squaring of the obtained coefficients at every frequency line is necessary, in order to obtain power values (Fujii and Azegami, 1967). It is obvious that the result is not fruitful, thus another approach is to be chosen.

The spectrum of a quantized sine wave was investigated by Claasen and Jongepier (1981), using Eq. (20.18).

If $x(t)$ is a sine wave,

$$x(t) = A\sin(2\pi f_1 t + \varphi)\,, \tag{20.39}$$

its derivative is also sinusoidal, and the PDF of the derivative is

$$f_{\dot{x}}(x) = \frac{1}{2\pi^2 f_1 A}\frac{1}{\sqrt{1-\left(\dfrac{x}{2\pi f_1 A}\right)^2}}\,, \quad \text{for } |x| < 2\pi f_1 A\,. \tag{20.40}$$

By substituting Eq. (20.40) into Eq. (20.18), the desired formula is obtained:

$$S_{\nu\nu}(f) \approx \frac{q^3}{4\pi^4 f_1 A}\sum_{n=1}^{\infty}\frac{\dfrac{1}{\sqrt{1-\left(\dfrac{fq}{2\pi f_1 A n}\right)^2}}}{n^3}\,, \quad \text{for } |fq| < 2\pi f_1 A n\,, \tag{20.41}$$

where the condition means that for each value of f only those terms are summed for which the condition is satisfied.

Equation (20.18) is clearly a rather rough approximation in this case, since this is continuous, while the quantization noise is in reality periodic with the fundamental frequency f_1 (see Fig. 20.10), and has a discrete spectrum (see Fig. 20.11). However, Eq. (20.18) was successfully verified by Claasen and Jongepier, using a spectrum analyzer. The character of a measured spectrum is shown in Fig. 20.12.

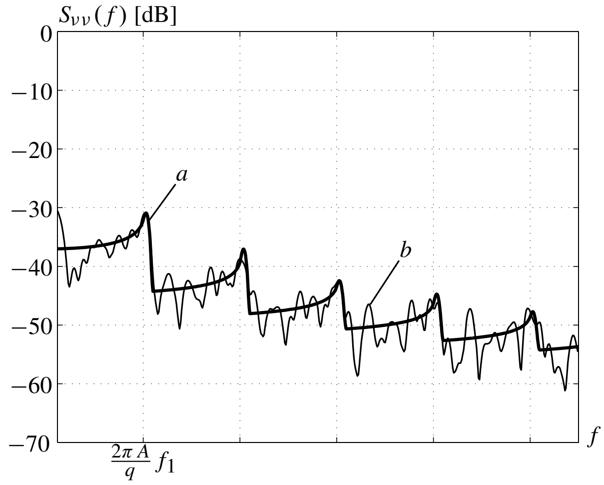

Figure 20.12 Verification of the model of Eq. (20.18) for the case of a sine wave (after Claasen and Jongepier (1981), ©1981 IEEE). Signal amplitude: $A = 15.5q$. a – spectrum obtained from Eq. (20.18), after smoothing; b – spectrum measured by a spectrum analyzer, with $\Delta f = 4.9 f_1$.

The analyzer had a resolution of

$$\Delta f \approx \frac{1}{20} \frac{2\pi A}{q} f_1 \approx 4.9 f_1 \,. \tag{20.42}$$

The infinite peaks in both the true spectrum and in Eq. (20.41) (see Fig. 20.13) were smoothed by the finite resolution. Verification was successful, because the *smoothed versions* of the two spectra are very similar. The approximate spectrum obtained from Eq. (20.41) and Eq. (20.40) contains about the same power as the true spectrum in a given bandwidth, if $\Delta f > f_1$ and $A \gg q$. In this sense this is a successful attempt to provide a more or less accurate, usable expression of the spectrum of a quantized sine wave.

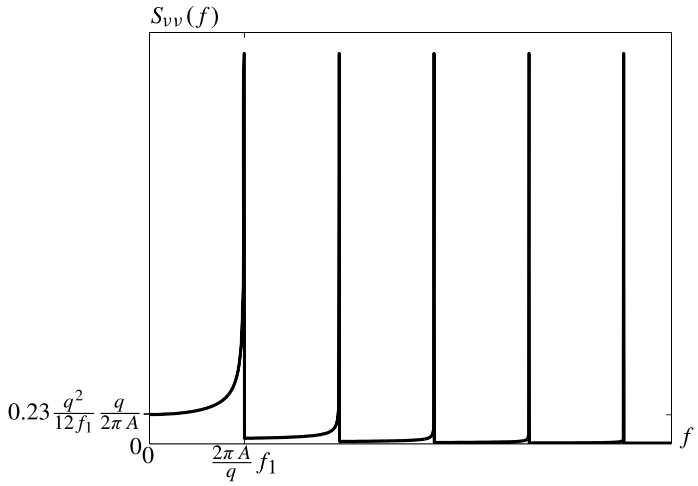

Figure 20.13 Approximate spectrum of the quantization noise of a sine wave, based on Eq. (20.41). $A = 15.5q$.

20.3 CONDITIONS OF WHITENESS FOR THE SAMPLED QUANTIZATION NOISE

Before going into details, let us briefly survey why a white noise spectrum is favorable:

(a) The white spectrum does not depend on the form of the input spectrum, thus the spectral behavior of the quantization noise does not depend on that of the input signal.

(b) The white spectrum means that the noise samples are uncorrelated, thus well-known design formulas elaborated for measurements with uncorrelated samples can be used.

(c) The power of the quantization noise is uniformly distributed between 0 and $f_s/2$, and this is advantageous in the measurement of spectra.

Example 20.8 Quantization Noise in Audio
In digital audio techniques it is of outstanding importance that the quantization noise does not result in an annoying audible noise superimposed on the music. With present technology the noise level is not reduced below the hearing threshold, thus at soft parts of music the noise is audible. There are two conditions that should to be fulfilled:

- The noise spectrum should be white, so that the noise is heard and it is not a "tone" due to a more or less concentrated frequency spectrum.
- The noise should be uncorrelated with the input signal, otherwise it would act as a kind of distortion component.

The first condition will be studied in the next part of this chapter. The second one was treated in Chapter 6.

It should be mentioned here that by appropriate feedback the noise spectrum can be modified in such a way that, at the price of remarkably higher variance, a large part of the quantization noise power is "transferred" to higher frequencies. More details can be found in (Spang and Schultheiss, 1962).

We have already seen in the previous sections that the spectrum of the quantization noise is usually broad and flat in a relatively wide band. If only this band is of interest, the spectrum can be considered as being white. However, the quantization noise, considered as a continuous-time stochastic process, cannot have a white spectrum, since its variance is approximately $q^2/12$ (see Eq. (20.4)), while a continuous-time white spectrum would in theory have an infinite variance.

On the other hand, quantization is usually performed in connection with sampling, or on already sampled data. The effect of sampling a signal is the repetition of its original spectrum. From Fig. 20.14 it is obvious that in the case of smooth spectra and significant overlapping (sufficiently small f_s values) the resulting spectrum of the samples is more or less white. Thus, to provide a white spectrum, an upper bound for f_s must be determined.

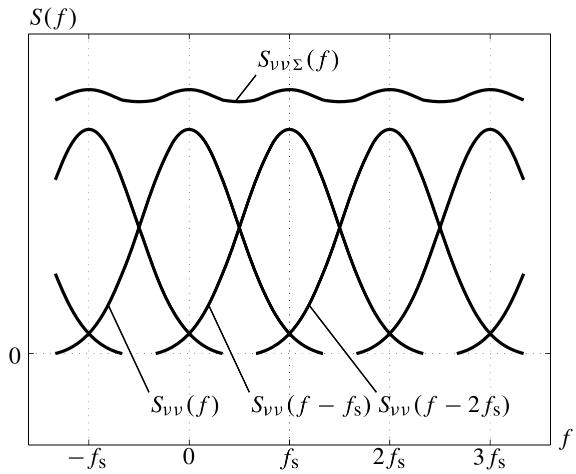

Figure 20.14 The effect of sampling in frequency domain.

In the time domain, white spectrum means uncorrelated samples. This feature will be useful in understanding the following discussions; on the other hand, this explains why whiteness of the spectrum is advantageous when the effect of quantization is investigated in many systems.

Let us formulate now the conditions for a white quantization noise spectrum. The general condition of whiteness was given by Sripad and Snyder (1977). Using the characteristic function method they showed that the necessary and sufficient condition for the uncorrelatedness of noise samples is that the characteristic function of the sample pair of $x(t)$ fulfills the equation

$$\Phi_{x_1 x_2}\left(\frac{2\pi l_1}{q}, \frac{2\pi l_2}{q}\right) = 0 \tag{20.43}$$

for every integer value of l_1 and l_2, except $(l_1, l_2) = (0, 0)$.

Independence implies uncorrelatedness, thus Eq. (20.43) is a sufficient condition. The exact expression of $\mathrm{E}\{\nu_1 \nu_2\} = R_{\nu_1 \nu_2}(\tau)$ was given in Eq. (20.16), and this has to be equal to zero for $|\tau| > 0$ in order to provide uncorrelatedness.

However, to formulate a usable *general* condition on the basis of Eq. (20.16) seems to be very difficult. Thus, in the following considerations Eq. (20.43) will be used, or Eq. (20.16) will be evaluated for special cases.

In the following sections we will deal first with the conditions for a white quantization noise spectrum in the case of the two most often investigated signal types, the Gaussian noise and the sine wave, based on the previous considerations. Then a uniform condition will be formulated, which does not rely on the signal shape.

20.3.1 Bandlimited Gaussian Noise

A Condition Based on Approximate Correlation

First let us consider expression Eq. (20.16) of the noise autocorrelation as applied in the Gaussian case (see Appendix F, Eq. (F.17)). The highest quantization noise correlation occurs for $\rho \approx 1$, where

$$\Phi_{x_1 x_2}(u_1, u_2) \approx e^{-4\pi^2(1-\rho)\frac{\sigma_x^2}{q^2}} . \tag{20.44}$$

Under this condition, the dominating terms in expression (20.16) belong to $l_1 = -l_2 = \pm 1$, that is, $u_1 = -u_2 = \pm 2\pi/q$.

To ensure that the correlation coefficient of the quantization noise samples is less than, say, 0.1,

$$\frac{\mathrm{E}\{\nu_1 \nu_2\}}{q^2/12} \approx \frac{1}{q^2/12} 2 \frac{q^2}{4\pi^2} e^{-4\pi^2(1-\rho)\frac{\sigma_x^2}{q^2}} < 0.1 \, , \tag{20.45}$$

the inequality

$$\rho < 1 - \frac{1.8}{4\pi^2} \frac{q^2}{\sigma_x^2} = 1 - 0.046 \frac{q^2}{\sigma_x^2} \tag{20.46}$$

has to be fulfilled, and, using the first two terms of the power series expansion of the correlation function of a bandlimited white noise again (see Eq. (20.21)):

$$\rho(\tau) = \frac{\sin(2\pi B\tau)}{2\pi B\tau} = 1 - \frac{(2\pi B\tau)^2}{3!} + \cdots , \qquad (20.21)$$

the desired condition can be obtained, using $f_s = 1/\tau$:

$$f_s < \frac{2\pi}{\sqrt{3! \cdot 0.046}} \frac{\sigma_x}{q} B = 12 \frac{\sigma_x}{q} B . \qquad (20.47)$$

An exact formula for the correlation of the quantization noise can be obtained using Eq. (9.7). This was obtained by Sripad and Snyder (1977) as Eq. (20.24).

However, this is a rather complicated formula. Widrow (1956b) and Korn (1965) presented a first-order approximation of $\mathrm{E}\{\nu_1\nu_2\}$ for $\rho \approx 1$:

$$\mathrm{E}\{\nu_1\nu_2\} \approx \frac{q^2}{12} e^{-(1-\rho)4\pi^2 \frac{\sigma_x^2}{q^2}} , \qquad (20.48)$$

which is the same as Eq. (20.26).

Using Eq. (20.48), the condition of approximate uncorrelatedness can be formulated as follows:

$$\frac{R_{\nu\nu}(\tau)}{R_{\nu\nu}(0)} < 0.1 \quad \text{if} \quad \rho < 1 - 0.058 \frac{q^2}{\sigma_x^2} . \qquad (20.49)$$

In the case of quantization of a bandlimited Gaussian white noise, the first two terms of Eq. (20.21) can be used again, and applying $f_s = 1/\tau$, the following condition is obtained:

$$f_s < \sqrt{\frac{(2\pi)^2}{3! \cdot 0.058}} \frac{\sigma_x}{q} B \approx 10.7 \frac{\sigma_x}{q} B , \qquad (20.50)$$

which is in very good agreement with Eq. (20.47).

Equation (20.50) can be re-written as

$$\frac{q}{\sigma_x} < 10.7 \frac{B}{f_s} . \qquad (20.51)$$

The Nyquist sampling frequency in this case is $f_N = 2B$. Accordingly, for whiteness of the quantization noise,

$$\frac{q}{\sigma_x} < 5.3 \frac{f_N}{f_s} . \qquad (20.52)$$

This is a useful result.

When sampling with $f_s = f_N$, Eq. (20.52) is weaker than its application condition $q < \sigma_x$, therefore, as an example, we will discuss the use of $f_s = 10 f_N$. When sampling at this frequency, the quantization noise will be white if the quantization grain size q is less than 0.5 standard deviations of the quantizer input signal x. One should note that meeting the condition for whiteness (20.52) does not guarantee that the quantization noise will be uncorrelated with signal x. In fact, the grain size q should be less than about $0.25\sigma_x$, in order to have very low crosscorrelation between the quantization noise and signal x.

A Condition Based on the Bandwidth of the Spectrum

By considering Fig. 20.8 it can be observed that the quantization noise spectra are more or less smooth, and, by decreasing the quantum step size, they broaden and flatten. Sampling repeats the spectrum at integer multiples of the sampling frequency. To have a white spectrum for the sampled quantization noise, the spectral replicas should sufficiently overlap. It seems to be reasonable to choose the sampling frequency not greater than, e.g., double the frequency spacing of the 3 dB (half-power) points of the quantization noise spectrum, since in this case, the summed replicas everywhere will yield a large enough spectrum.

In Eq. (20.23), $\eta(y)$ has its 3 dB-point at $y \approx 0.8$, and thus from

$$y = \frac{3\frac{q^2}{\sigma_x^2}\left(\frac{f_\mathrm{s}}{2}\right)^2}{8\pi^2 B^2} < 0.8\,, \tag{20.53}$$

the following condition can be given:

$$f_\mathrm{s} < 9.2\frac{\sigma_x}{q}B\,, \tag{20.54}$$

which is practically equivalent to Eq. (20.47) or (20.50). The condition for q/σ_x is

$$\frac{q}{\sigma_x} < 4.6\frac{f_\mathrm{N}}{f_\mathrm{s}}\,, \tag{20.55}$$

very close to (20.52).

The equivalent quantization noise bandwidth (Bendat and Piersol, 1986) can be calculated as follows:

$$B_\mathrm{e} = \frac{\int\limits_{-\infty}^{\infty} S_{\nu\nu}(f)\,\mathrm{d}f}{2\,S_{\nu\nu}(0)} \approx \frac{\frac{q^2}{12}}{2\frac{q^2}{4\pi^3 B}\sqrt{\frac{3\gamma}{2\pi}}\sum\limits_{n=1}^{\infty}\frac{1}{n^3}} \approx 6.3\frac{\sigma_x}{q}B\,, \tag{20.56}$$

from which once again practically the same condition as above,

$$f_\mathrm{s} < 12.6\frac{\sigma_x}{q}B \tag{20.57}$$

is obtained.

The condition for the quantization grain size, expressed similarly to (20.52), is

$$\frac{q}{\sigma_x} < 6.3 \frac{f_{\mathrm{N}}}{f_{\mathrm{s}}} . \tag{20.58}$$

A Condition Based on Computer Evaluation of Spectra

On the basis of computer evaluation of the series expansion of the spectrum of the quantized variable, Eq. (20.30), Robertson (1969) gave a simple rule of thumb: provided that $\sigma_x > q$, the noise spectra will be white even in the case of colored signal spectra, if the sampling frequency is not much greater (e.g. $f_{\mathrm{s}} < 6B$) than the Nyquist rate for the signal being quantized. For the quantization of bandlimited white noise, in the case of $\sigma_x \gg q$, the limit on sampling frequency is to be increased by the factor σ_x/q (the density of the jumps in the quantization noise increases in proportion with σ_x/q, which means that its spectrum broadens similarly), the resulting condition,

$$f_{\mathrm{s}} < 6 \frac{\sigma_x}{q} B \tag{20.59}$$

is again in good agreement with Eq. (20.54) etc. Condition (20.59) is stricter, since it is derived from a condition valid for colored spectra as well.

The condition for the quantization grain size is, similarly to (20.52),

$$\frac{q}{\sigma_x} < 3 \frac{f_{\mathrm{N}}}{f_{\mathrm{s}}} . \tag{20.60}$$

Robertson's result can be used in the case of narrow-band Gaussian noise, too. Refer to Fig. 20.15. The center frequency of the Gaussian noise spectrum is f_0. The limit on the sampling frequency is

$$f_{\mathrm{s}} < 2 \cdot 2 f_0 \frac{\sigma_x}{q} = 4 \frac{\sigma_x}{q} f_0 . \tag{20.61}$$

However, this should be taken with due precaution, since this is based on a rather qualitative statement of Robertson.

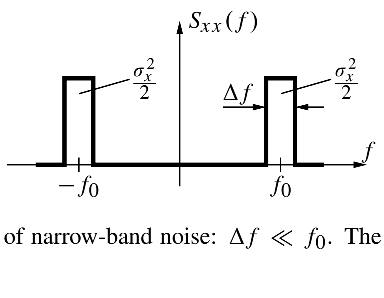

Figure 20.15 Spectrum of narrow-band noise: $\Delta f \ll f_0$. The area of each rectangle is $\sigma_x^2/2$.

The center frequency of the input signal is f_0, and the Nyquist sampling frequency for this case is approximately

$$f_{\mathrm{N}} = 2 f_0 \,. \tag{20.62}$$

Accordingly, for whiteness of quantization noise,

$$\frac{q}{\sigma_x} < 2 \frac{f_{\mathrm{N}}}{f_{\mathrm{s}}} \,. \tag{20.63}$$

This is a useful result. It is a tighter condition than (20.52), and it reflects the idea that the lowpass Gaussian signal is, at its Nyquist rate, more uncorrelated over time than the bandpass Gaussian signal is at its Nyquist rate.

20.3.2 Sine Wave

Using the results of Claasen and Jongepier (1981), a condition for the sampling rate can be formulated when a sine wave is quantized (see also Section 20.2.5). From Fig. 20.12 one can see that $S_{\nu\nu}(f)$ is flat for

$$|f| < \frac{\pi A}{q} f_1 \,. \tag{20.64}$$

Claasen and Jongepier (1981) suggested this or an even smaller upper bound for the sampling frequency. Considering that the first peak at $f = 2\pi A f_1/q$ contains about 10% of the total signal power, the condition

$$f_{\mathrm{s}} < 0.5 \frac{\pi A}{q} f_1 \tag{20.65}$$

may be suggested.[2] The resulting spectrum will contain some ripples for any value of f_{s}. However, the power associated with them will be small in comparison to the total power of the quantization noise.

The Nyquist frequency of the sine wave of frequency f_1 is

$$f_{\mathrm{N}} = 2 f_1 \,. \tag{20.66}$$

Substituting, the whiteness condition for quantization noise with a sinusoidal input becomes

$$\frac{q}{A} < 0.25\pi \frac{f_{\mathrm{N}}}{f_{\mathrm{s}}} \,. \tag{20.67}$$

This is a useful result.

It is interesting to compare (20.67) with the corresponding condition for the narrow-band bandpass Gaussian case given by (20.63), since both the Gaussian signal and the sinusoidal signal are narrow-band. The sine wave amplitude is bounded between $\pm A$. The Gaussian signal is not bounded, but almost all of its probability

[2] This coincides well with (G.6), obtained using the characteristic function

lies between $\pm 2.5\sigma_x$. So let the narrow-band Gaussian signal have an "amplitude" of

$$A_\mathrm{G} = 2.5\sigma_x \,. \tag{20.68}$$

Substituting this into (20.63), we have

$$\frac{q}{A_\mathrm{G}} < \frac{2}{2.5} \cdot \frac{f_\mathrm{N}}{f_\mathrm{s}} \,. \tag{20.69}$$

Whiteness condition (20.69) for the narrow-band Gaussian signal is almost the same as (20.67) for the sinusoidal signal since

$$0.25\pi = 0.7854, \quad \text{and}$$
$$2/2.5 = 0.8 \,.$$

This is a very nice result.

Example 20.9 Quantization Noise in a Spectrum Analyzer
In a spectrum analyzer with the bandwidth of 25 kHz (Pendergrass and Farnbach, 1978), a 27 kHz sine wave of amplitude $A_d \approx 11.5q$ is used at the input as an additive dither[3] to linearize and thus improve the performance of the analyzer's 12-bit A/D converter. The dither creates approximate conditions for the satisfaction of QT II at the quantizer input and allows the PQN model of quantization to apply to a very close approximation.

Let us calculate the bias of the spectral estimate due to quantization noise, if the sampling frequency is $f_\mathrm{s} \approx 100$ kHz. Let us compute the signal-to-quantization-noise ratio in the analyzer bandwidth for the case of the maximal amplitude input sine wave, if the number of processed samples is $N = 512$. Note that the dither itself does not distort the spectrum since its frequency of 27 kHz is outside the frequency range of the analyzer.

The worst case occurs when a useful signal of very low level is analyzed. From the point of view of quantization, in this case, practically the dither itself is quantized. According to (20.65), the quantization noise spectrum is white if

$$f_\mathrm{s} = 100\,\mathrm{kHz} < 1.6\frac{A_d}{q} f_1 = 1.6 \cdot 11.5 \cdot 27 \approx 497\,\mathrm{kHz}\,, \tag{20.70}$$

which is amply fulfilled. Thus, the whole power, $q^2/12$, will be approximately uniformly distributed between zero and f_s, thus the increase in the PSD due to quantization is

$$\Delta S = \frac{q^2}{12 f_\mathrm{s}}\,, \tag{20.71}$$

[3]Dither is the subject of Chapter 19. It is briefly mentioned here to illustrate an application of the theory of sine wave quantization.

where ΔS is the bias in the spectrum due to quantization noise. This is the "noise floor" of the spectrum analyzer.

Let us compare this to the spectral peak belonging to the maximal input sine wave. Suppose that the sampling frequency is approximately an integer multiple of the sine frequency so that the "picket fence effect" does not show up. Using $A_{\mathrm{max}} \approx 2^{11}q$, the dynamic range of the spectrum analyzer is

$$d = 10 \log \left(\frac{\frac{A_{\mathrm{max}}^2}{4} N}{\frac{q^2}{12}} \right) \approx 98\,\mathrm{dB}\,. \tag{20.72}$$

20.3.3 A Uniform Condition for White Noise Spectrum

In the previous sections different techniques were used to derive upper bounds of the sampling frequency for the quantization of certain signal types. In this section we formulate an approximate uniform condition which is independent of concrete signal parameters.

In Fig. 20.2 the quantization of a Gaussian noise is illustrated. We observe that the waveform of the quantization noise is in most time intervals very similar to a "saw-tooth" signal, with varying slope.

It seems to be quite natural to assume that the samples of the quantization noise are uncorrelated, if not more than 1–2 samples are taken from a "period" (see Fig. 20.2(c)). The average length of these periods depends on the average slope of the signal, and can be expressed as follows:

$$T_{\mathrm{p}} = \frac{q}{\mathrm{E}\{|\dot{x}(t)|\}}\,, \tag{20.73}$$

where $\dot{x}(t)$ is the first derivative of the signal, and $x(t)$ is assumed to be stationary. From Eq. (20.73) an upper bound can be given for the sampling frequency:

$$f_{\mathrm{s}} < K \frac{\mathrm{E}\{|\dot{x}(t)|\}}{q}\,, \tag{20.74}$$

where the constant K may be somewhere in the range $(1, 2)$.

Let us check this condition for the cases treated in the previous sections. For zero mean Gaussian variables,

$$\mathrm{E}\{|x|\} = \sqrt{\frac{2}{\pi}}\sigma_x\,. \tag{20.75}$$

Differentiation of the time domain signal causes a factor $(j2\pi f)$ to appear in the Fourier domain. Therefore, the expression of the variance of the derivative is

$$\sigma_{\dot{x}}^2 = \int_{-\infty}^{\infty} (2\pi f)^2 S_{xx}(f)\,\mathrm{d}f \,, \tag{20.76}$$

and from this, we obtain for the bandlimited (lowpass) input signal

$$\mathrm{E}\{|\dot{x}(t)|\} = \sqrt{\frac{2}{\pi}}\sigma_{\dot{x}} \approx \sqrt{\frac{8\pi}{3}}\sigma_x B \,, \tag{20.77}$$

and for the narrow-band (bandpass) input signal

$$\mathrm{E}\{|\dot{x}(t)|\} = \sqrt{\frac{2}{\pi}}\sigma_{\dot{x}} \approx \sqrt{8\pi}\,\sigma_x f_0 \,. \tag{20.78}$$

For the sine wave,

$$\mathrm{E}\{|\dot{x}(t)|\} = 4 f_1 A \,. \tag{20.79}$$

By comparing Eqs. (20.77), (20.78), and (20.79) with Eqs. (20.54), (20.61), and (20.65), it is clear that in each case Eq. (20.74) can be used, with $K = 3.2, 0.8, 0.4$, respectively. It can be observed that K is of the same order of magnitude for each type of signal (though its value varies more than it was supposed above; it is not surprising that the "worse" the signal behaves ("bad" means in our interpretation to have very spiked spectra, as in the case with the latter two signals), the smaller is the value of K that must be chosen within the approximate range from 3.2 to 0.4.

On the basis of the above results, Eq. (20.74) turns out to be very useful, especially because $\mathrm{E}\{|\dot{x}(t)|\}$ can usually be measured directly. The appropriate value of K depends slightly on the waveform, nevertheless since Eq. (20.74) is an inequality, K can be chosen to be sufficiently small in order to have a safe upper bound on the sampling frequency.

It has been shown in this section that Eq. (20.74) can be used for two types of Gaussian signals and also for sinusoidal signals, to provide white quantization noise spectra. There are further examples (see e.g. Example 20.10 below), in which quantization noises behave similarly. However, it is an open question, what is the *maximum* value of K for a class of signals? Intuitively it is clear that if the input signal does not have broad nearly constant sections or nearly constant-slope sections (as the square wave and the triangle wave), then Eq. (20.74) can be used e.g. with $K = 0.4$. Greater values of K can be used if the signal is of stochastic nature. Moreover, dithering (see Chapter 19) can always be effective in assuring the necessary randomness of the signal to be quantized.

Example 20.10 Signal Processing in an Industrial Weightchecker
This is an illustrative example to show the use of the general condition of whiteness in a real-life problem.

In many measurements like the measurement of temperature or weight, the system has a limited bandwidth which does not allow quick readout. The classical approach is to wait long enough to allow the transients to die out. However, if the system dynamics is known, the steady-state value can be predicted by using LS estimation.

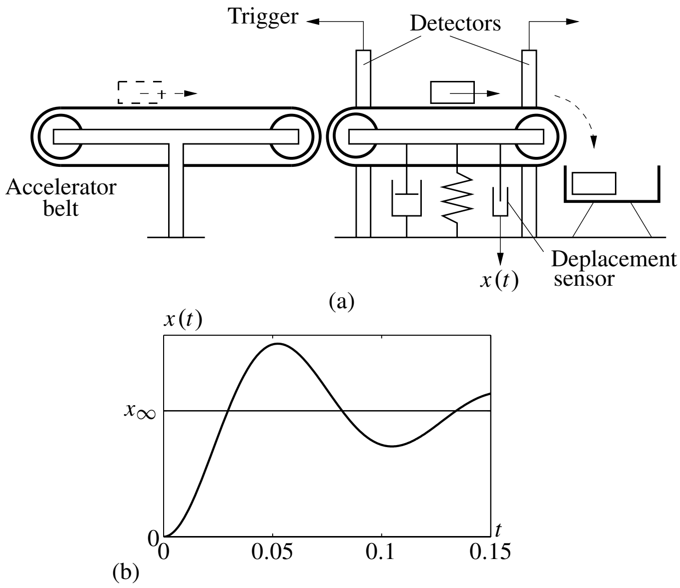

Figure 20.16 Transient response of a weightchecker: (a) schematic diagram; (b) measured response.

The signal model for the output of an industrial weightchecker illustrated in Fig. 20.16 is as follows:

$$x(t) = x_\infty + A\, e^{-\frac{t}{T}} \cos(\omega_0 t) + B\, e^{-\frac{t}{T}} \sin(\omega_0 t) + n(t)\,. \tag{20.80}$$

where x_∞ is the desired steady-state value, $\omega_0 = 60\,\text{Hz}$, the natural frequency of the system, $T = 0.083\,\text{s}$, $n(t)$ is the measurement noise. A and B are unknown parameters corresponding to the slightly random initial conditions of each measurement. For the initial conditions the following equations hold:

$$\text{E}\{x(0)\} = 0\,, \quad \text{E}\{\dot{x}(0)\} = 0 \quad \Rightarrow \quad \text{E}\{A\} = -x_\infty\,, \quad \text{E}\{B\} = -\frac{x_\infty}{\omega_0 T}\,. \tag{20.81}$$

For samples of the measured signal this model takes the following form,

$$x(t_k) = x_\infty + A\,e^{-\frac{t_k}{T}}\cos(\omega_0 t_k) + B\,e^{-\frac{t_k}{T}}\sin(\omega_0 t_k) + k(t_k)\,, \quad k = 0, 1, 2, \ldots, N-1\,, \tag{20.82}$$

or in matrix form

$$\mathbf{x} = \mathbf{U}\begin{bmatrix} x_\infty \\ A \\ B \end{bmatrix} + \mathbf{n}\,, \tag{20.83}$$

where N is the number of samples, and

$$\mathbf{U} = \begin{bmatrix} 1 & e^{-\frac{0\Delta t}{T}}\cos(\omega_0 0\Delta t) & e^{-\frac{0\Delta t}{T}}\sin(\omega_0 0\Delta t) \\ 1 & e^{-\frac{1\Delta t}{T}}\cos(\omega_0 1\Delta t) & e^{-\frac{1\Delta t}{T}}\sin(\omega_0 1\Delta t) \\ \vdots & \vdots & \vdots \\ 1 & e^{-\frac{(N-1)\Delta t}{T}}\cos(\omega_0 (N-1)\Delta t) & e^{-\frac{(N-1)\Delta t}{T}}\sin(\omega_0 (N-1)\Delta t) \end{bmatrix} \tag{20.84}$$

The LS estimate can be obtained from the discrete model in the form

$$\hat{x}_\infty = \mathbf{e}_1^{\mathrm{T}}(\mathbf{U}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{U}^{\mathrm{T}}\mathbf{x} = \sum_{k=1}^{N} a_k x(t_k)\,, \tag{20.85}$$

with $\mathbf{e}_1^{\mathrm{T}} = [1\ 0\ 0]$.

Suppose that Eq. (20.85) will be evaluated, with data obtained by using an 8-bit A/D converter with input range $(0, 2x_\infty)$. Equation (20.85) is a weighted average of the samples, therefore it effectively reduces the variance due to quantization noise, at least when the noise samples are uncorrelated. The roundoff noise of the representation of the coefficients a_k and that of the calculations will be neglected. Therefore, it is reasonable to increase the sampling frequency in order to reduce the uncertainty of the estimate. What is the maximum number of samples to be taken in a given measurement time $T_{\mathrm{m}} = 0.15$ s that already provides optimal suppression of the quantization noise? We need to determine the minimum standard deviation of Eq. (20.85) with respect to q if only the quantization noise is taken into account as a source of variance.

Let us use the general formula Eq. (20.74). Since the spectral behavior of $x(t)$ is lowpass, $K \approx 3$ seems to be a reasonable choice.

First the average of $|\dot{x}(t)|$ has to be calculated:

$$\overline{|\dot{x}(t)|} = \frac{\text{Total variation}}{\text{Measurement time}} \approx \frac{2.3x_\infty}{T_{\mathrm{m}}}\,. \tag{20.86}$$

The total variation was calculated from the data derived from simulation of the system pictured in Fig. 20.16.

From Eq. (20.71) we obtain for the number of samples:

$$N = f_\mathrm{s} T_\mathrm{m} < K \frac{\overline{|\dot{x}(t)|}}{q} T_\mathrm{m} \approx 3 \frac{\dfrac{2.3 x_\infty}{0.15}}{\dfrac{2 x_\infty}{2^8}} 0.15 \approx 883\,. \tag{20.87}$$

Simulation results give for the limit of effective averaging $N \approx 1300$, which corresponds to a value of $K \approx 5$. Our calculated upper bound is consequently somewhat low, however, it is a safer limit. Equation (20.87) provides in fact independent noise samples. On the other hand, the value of K given for bandlimited Gaussian noise ($K = 3.2$) lies in the near neighborhood of $K \approx 5$. Accordingly, from the point of view of quantization, the function given in Eq. (20.80) behaves similarly to a bandlimited Gaussian noise.

The variance of the quantization noise is equal to approx. $q^2/12$. Let us use the result numerically obtained (Kollár, 1983) for the case illustrated in Fig. 20.16:

$$\sum_{k=1}^{N-1} a_k^2 \approx \frac{1.13}{N}\,. \tag{20.88}$$

Assuming that the noise samples are uncorrelated, for $N = 1300$

$$\sqrt{\mathrm{var}\{\hat{x}_\infty\}} = \sqrt{\frac{q^2}{12} \sum_{k=1}^{N} a_k^2} \approx \sqrt{\frac{q^2}{12} \frac{1.13}{N}} \approx 0.008q\,. \tag{20.89}$$

Because of the nonlinear error of the ADC, this result is somewhat smaller than experienced in practice, but well illustrates the validity of the theory.

20.4 SUMMARY

If the sampled input to a quantizer is $x_1, x_2, x_3, \ldots$, and if two-variable QT II conditions are satisfied for x_1 and x_2, x_1 and x_3, x_1 and x_4, and so forth, the quantization noise will be white with zero mean and a variance of $q^2/12$, and it will be uncorrelated with the signal being quantized. This situation occurs in practice in many cases, and is closely approximated in many more cases. Spectral analysis of quantization noise is especially simple under these circumstances because the autocorrelation of the quantizer output is equal to the autocorrelation of the quantizer input plus the autocorrelation of the quantization noise, and the spectrum of the output of the quantizer is equal to the spectrum of the input plus the spectrum of the quantization noise. The spectrum of the noise is flat, with total power of $q^2/12$.

Conditions for whiteness of the quantization noise based on the characteristic function were derived by Widrow (1956b) and Sripad and Snyder (1977). Based on a number of different approaches, other conditions for whiteness were derived by Bennett (1948), Claasen and Jongepier (1981), Peebles (1976), Katzenelson (1962), Bendat and Piersol (1986), Robertson (1969), and Kollár (1986). The basic issue is, when performing analog-to-digital conversion on an input signal, how fine must the quantization be in order that the quantization noise be white?

In general, the higher the sampling rate, the finer the quantization must be to achieve whiteness of the quantization noise. When a bandlimited lowpass Gaussian signal is quantized, relation (20.47) gives a limit on the sampling frequency to insure whiteness, with the application condition $q < \sigma_x$. Other approximations lead to similar limits such as (20.50), (20.54), and (20.57). The most conservative is (20.54),

$$f_\text{s} < 9.2\frac{\sigma_x}{q}B\,. \tag{20.54}$$

This leads to

$$\frac{q}{\sigma_x} < 4.6\frac{f_\text{N}}{f_\text{s}}\,, \tag{20.55}$$

with f_N being the Nyquist sampling frequency.

When sampling with $f_\text{s} = 10 f_\text{N}$, the quantization noise will be white if the quantization grain size q is less than 0.5 standard deviations of the quantizer input signal x. However, the grain size q should be less than about $0.25\sigma_x$, in order to have very low crosscorrelation between the quantization noise and signal x.

If the Gaussian signal is more oversampled, for example if the sampling rate is 100 times the Nyquist rate, expression (20.52) tells us that the quantization grain size q should be smaller than $0.05\sigma_x$ for the quantization noise to be white. Under these circumstances, the quantization noise would be highly uncorrelated with the input signal x.

Quantization of a narrow-band bandpass Gaussian signal has the whiteness condition

$$f_\text{s} < 4\frac{\sigma_x}{q}f_0\,, \tag{20.61}$$

with the application condition $q < \sigma$. From this,

$$\frac{q}{\sigma_x} < 2\frac{f_\text{N}}{f_\text{s}}\,. \tag{20.63}$$

Quantization of a sine wave has the whiteness condition

$$f_\text{s} < 0.5\pi\frac{A}{q}f_1\,, \tag{20.65}$$

with the application condition $q \ll A$. This condition can be rearranged as

$$\frac{q}{A} < 0.25\pi \frac{f_{\mathrm{N}}}{f_{\mathrm{s}}} . \tag{20.67}$$

This is usually fulfilled as a consequence of $q \ll A$, if oversampling is not very high.

20.5 EXERCISES

20.1 A sine wave of amplitude $A = 3.5$ V is sampled and quantized with a 10-bit A/D converter working over the range [± 5 V]. The sampling frequency is f_{s}, and the sine wave frequency is f_1.

- **(a)** Determine numerically the amplitude spectrum of the quantization noise for $f_{\mathrm{s}} = 3000 f_1$, using a DFT with $N = 4096$. Plot the spectrum with logarithmic amplitude scale (dB).
- **(b)** Determine the highest sampling frequency f_{s} such that the approximate condition for quantization noise whiteness is satisfied.
- **(c)** Plot the amplitude spectrum of the quantization noise, sampling with the frequency determined in (b), with $N = 4096$.
- **(d)** Repeat the above calculations applying white dither with uniform distribution between ± 0.0049 V.

20.2 The DFT of a random white Gaussian sequence with variance σ^2 is calculated.

- **(a)** Determine the probability distribution at frequency k of the result for $0 < k < N/2$.
- **(b)** Calculate the 95% upper bound of the magnitude parametrically, and evaluate it for $N = 1024$.

20.3 The N-point digital Fourier transform of an input signal is calculated with double precision floating-point arithmetic. This calculation may be considered to be perfectly accurate. The input signal is uniformly quantized with a quantum step size of q.

- **(a)** At frequency index k, $0 < k < N/2$, calculate the theoretical bounds on the real and imaginary parts of the quantization noise.
- **(b)** Calculate the theoretical bound on the magnitude of the quantization noise.
- **(c)** At frequency index k, calculate the variance of the real and imaginary parts of the quantization noise, and calculate the standard deviation of the quantization noise.
- **(d)** For $N = 1024$, calculate the theoretical bound of the magnitude of the quantization noise, and the standard deviation of it at frequency index $k = 128$. Compare these values.
- **(e)** Give the distribution of the magnitude square of the quantization noise ($|X_{\mathrm{q}k}|^2$), and of the magnitude.