

## Chapter 8

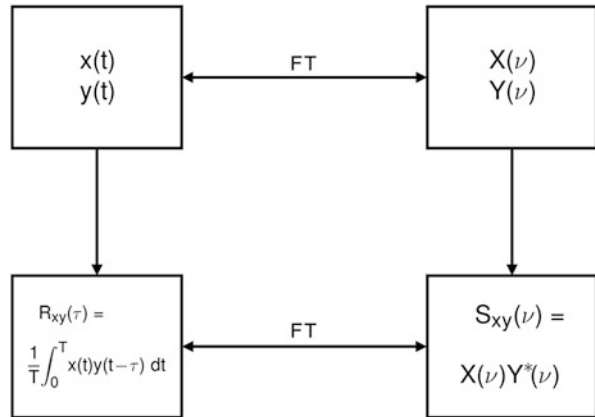
# Digital Signal Processing

The use of digital rather than analog instrumentation offers important practical advantages in the transmission of signals over long baselines, the implementation of compensating time delays, and the measurement of cross-correlation of signals. In digital delay circuits, the accuracy of the delay depends on the accuracy of the timing pulses in the system, and long delays accurate to tens of picoseconds are more easily achieved digitally than by using analog delay lines. Furthermore, there is no distortion of the signal by the digital units other than the calculable effects of quantization. In contrast, with an analog system, it is difficult to keep the shape of the frequency response within tolerances while delay elements are switched into and out of the signal channels. Correlators with wide dynamic range are readily implemented digitally, including those with multichannel output, as needed for spectral line observations. Analog multichannel correlators employ filter banks to divide the signal passband into many narrow channels. Such filters, when subject to temperature variations, can be a source of phase instability. Finally, except at the highest bit rates (frequencies), digital circuits need less adjustment than analog ones and are better suited to replication in large numbers for large arrays.

Digitization of the signal waveforms requires sampling of the voltages at periodic intervals and quantizing the sampled values so that each can be represented by a finite number of bits. The number of bits per sample is usually not large, especially in cases in which the signal bandwidth is large, requiring high sampling rates. However, coarse quantization results in a loss in sensitivity, since modification of the signal levels to the quantized values effectively results in the addition of a component of “quantization noise.” In most cases, this loss is small and is outweighed by the other advantages. In designing digital correlators, there are compromises to be made between sensitivity and complexity, and the number of quantization levels to use is an important consideration.

There are two ways to determine the spectrum of a random noise signal, as shown in Fig. 8.1. The autocorrelation function of the signal can be measured and

**Fig. 8.1** The relationship between two random processes,  $x(t)$  and  $y(t)$ , of duration  $T$ , and their cross-correlation function,  $R_{xy}(\tau)$  and the cross spectrum,  $S_{xy}(\nu)$ . If  $x(t)$  and  $y(t)$  are the same,  $R_{xx}(\tau)$  is the autocorrelation function, and  $S_{xx}(\nu)$  is the power spectrum. The spatial counterpart to this diagram is shown in Fig. 5.5.



then Fourier transformed into a power spectrum after a specified integration period. Alternately, the signal can be Fourier transformed first and the square modulus taken. In the first case, the resolution of the spectral estimate is approximately the reciprocal of the number of lags of the autocorrelation function calculated. In the direct Fourier transform route, the data stream must be segmented to control the spectral resolution, i.e., the resolution is approximately the reciprocal of the data segment length. The power spectra from all of the segments are summed over the integration period. To compare results between these methods, the number of lags in the correlator is set equal to the number of segment samples. For interferometry, the same two methods can be applied. The cross-correlation function can be calculated and Fourier transformed into a cross spectrum (called the XF technique), or the direct Fourier transform of one can be multiplied by the conjugate of the other to form the cross spectrum (the FX technique). These two methods are explored in detail in this chapter.

Digital signal processing in radio astronomy began in the early 1960s when Weinreb (1963) built a digital 64-channel autocorrelator that operated on the signal sampled at the Nyquist rate and quantized with one bit per sample.<sup>1</sup> At that time, the modern fast Fourier transform (FFT) algorithm (Cooley and Tukey 1965) was not known, although there are historical precedents in the mathematical literature going back to Gauss in the early nineteenth century. For the next two decades, virtually all spectrometers for single-dish and interferometric applications were based on the auto- or cross-correlation approach. By the 1990s and the advent of very large spectral processing systems (in terms of frequency channels and baselines), the advantages of the FX approach became apparent. All modern interferometers have spectral analysis capabilities, not only for observations of spectral lines but also for mitigation of the effects of radio frequency interference (RFI) and of instrumental bandwidth smearing.

<sup>1</sup>A similar device was used by Goldstein (1962) to detect radar echoes from Venus.

## 8.1 Bivariate Gaussian Probability Distribution

The bivariate normal probability function is central to all signal analysis. If  $x$  and  $y$  are joint Gaussian random variables with zero mean and variance  $\sigma^2$ , the probability that one variable is between  $x$  and  $x + dx$  and, simultaneously, the other is between  $y$  and  $y + dy$  is  $p(x, y)dx dy$ , where

$$p(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[\frac{-(x^2 + y^2 - 2\rho xy)}{2\sigma^2(1-\rho^2)}\right], \quad (8.1)$$

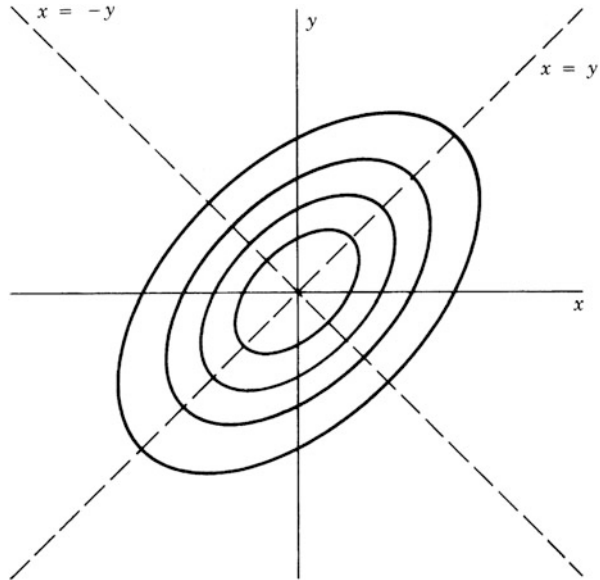
and  $\rho$  is the correlation coefficient equal to  $\langle xy \rangle / \sqrt{\langle x^2 \rangle \langle y^2 \rangle}$ , where  $\langle \rangle$  denotes the expectation, which, with the usual assumption of ergodicity, is approximated by the average over many samples. The form of this function is shown in Fig. 8.2. Note that  $-1 \leq |\rho| \leq 1$ . For  $|\rho| \ll 1$ , the exponential can be expanded, giving

$$p(x, y) \simeq \left[ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \right] \left[ \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-y^2}{2\sigma^2}\right) \right] \left(1 + \frac{\rho xy}{\sigma^2}\right), \quad (8.2)$$

which for  $\rho = 0$  is simply the product of two Gaussian functions. Equation (8.1) can also be written as

$$p(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right) \frac{1}{\sigma\sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-(y - \rho x)^2}{2\sigma^2(1-\rho^2)}\right]. \quad (8.3)$$

**Fig. 8.2** Contours of equal probability density from the bivariate Gaussian distribution in Eq. (8.1). The contours are given by  $x^2 + y^2 - 2\rho xy = \text{const.}$  For  $\rho = 0$ , they become circles; for  $\rho = 1$ , they merge into the line  $x = y$ ; and for  $\rho = -1$ , they merge into  $x = -y$ .



If this expression is integrated with respect to  $y$  from  $-\infty$  to  $+\infty$ , it reduces to a Gaussian function in  $x$ . As  $\rho$  approaches unity, Eq. (8.3) becomes the product of a Gaussian in  $x$  and a Gaussian in  $y - x$ ; the latter has a standard deviation  $\sigma \sqrt{1 - \rho^2}$ , which tends to zero as  $\rho$  approaches 1. Equations (8.1) and (8.2) will be used in examining the response of various types of samplers and correlators. For autocorrelators used with single antennas, the quantity to be measured is the autocorrelation function  $R(\tau) = \langle v(t)v(t - \tau) \rangle$ , where  $v$  is the received signal. This case can be treated with  $x = v(t)$  and  $y = v(t - \tau)$ .

## 8.2 Periodic Sampling

### 8.2.1 Nyquist Rate

If the signal is bandlimited, that is, its power spectrum is nonzero only within a finite band of frequencies, no information is lost in the sampling process as long as the sampling rate is high enough. This follows from the sampling theorem discussed in Sect. 5.2.1. Here, we sample a function of time and must avoid aliasing in the frequency domain. For a baseband (lowpass) rectangular spectrum with an upper cutoff frequency  $\Delta\nu$ , the width of the frequency spectrum, including negative frequencies, is  $2\Delta\nu$ . The function is fully specified by samples spaced in time with an interval no greater than  $1/(2\Delta\nu)$ , that is, a sampling frequency of  $2\Delta\nu$  or greater. This critical sampling frequency,  $2\Delta\nu$ , is called the *Nyquist rate*<sup>2</sup> for the waveform. For further discussion, see, for example, Bracewell (2000) or Oppenheim and Schaffer (2009). In some digital systems in radio astronomy, the waveform that is digitized has a baseband spectrum and is sampled at the Nyquist rate. For a rectangular passband of this type, the autocorrelation function, which by the Wiener–Khinchin relation is the Fourier transform of the power spectrum, is

$$R_{\infty}(\tau) = \frac{\sin(2\pi\Delta\nu\tau)}{2\pi\Delta\nu\tau}, \quad (8.4)$$

where the subscript  $\infty$  indicates unquantized sampling (that is, the accuracy is not limited by a finite number of quantization levels). Nyquist sampling can also be applied to bandpass spectra, and if the spectrum is nonzero only within a range of  $n\Delta\nu$  to  $(n + 1)\Delta\nu$ , where  $n$  is an integer, the Nyquist rate is again  $2\Delta\nu$ . Thus, for sampling at the Nyquist rate, the lower and upper bounds of the spectral band must be integral multiples of the bandwidth. The autocorrelation function of a signal that has a flat spectrum over such a band is

$$R_{\infty}(\tau) = \frac{\sin(\pi\Delta\nu\tau)}{\pi\Delta\nu\tau} \cos\left[2\pi\left(n + \frac{1}{2}\right)\Delta\nu\tau\right]. \quad (8.5)$$

---

<sup>2</sup>Shannon (1949) cites several references relevant to the development of this result, of which the earliest is Nyquist (1928).

Zeros in this function occur at time intervals  $\tau$  that are integral multiples of  $1/(2\Delta\nu)$ . Therefore, for a rectangular passband, successive samples at the Nyquist rate are uncorrelated. Sampling at frequencies greater or less than the Nyquist rate is referred to as oversampling or undersampling, respectively. For any signal, adjusting the center frequency so that the spectrum conforms to the bandpass sampling requirement described above minimizes the sampling rate required to avoid aliasing.

### 8.2.2 Correlation of Sampled but Unquantized Waveforms

We now investigate the response of a hypothetical correlator for which the input signals are sampled at the Nyquist rate but are not quantized. It is necessary to consider only single-multiplier correlators since complex correlators can be implemented as combinations of them, as indicated in Fig. 6.3. The system under discussion can be visualized as one in which the samples either remain as analog voltages or are encoded with a sufficiently large number of bits that quantization errors are negligible. Since no information is lost in sampling, the signal-to-noise ratio of the correlation measurement may be expected to be the same as would be obtained by applying the waveforms without sampling to an analog correlator. There is probably no reason, in practice, to build a correlator for inputs with unquantized sampling. However, by comparing the results with those for quantized sampling, which we discuss later, the effects of quantization are more easily understood.

Two bandlimited waveforms,  $x(t)$  and  $y(t)$ , are sampled at the Nyquist rate, and for each pair of samples, the multiplier within the correlator produces an output proportional to the product of the input amplitudes. The integrator allows the output to be averaged for any required time interval. Now the (normalized) cross-correlation coefficient of  $x(t)$  and  $y(t)$  for zero time delay between the two waveforms is

$$\rho = \frac{\langle x(t)y(t) \rangle}{\sqrt{\langle [x(t)]^2 \rangle \langle [y(t)]^2 \rangle}} . \quad (8.6)$$

(The cross-correlation coefficient  $\rho$  should not be confused with the autocorrelation function of  $x$  or  $y$ ,  $R_{\infty}$ .) Since  $x$  and  $y$  have equal variance  $\sigma^2$ ,

$$\langle x(t)y(t) \rangle = \rho\sigma^2 . \quad (8.7)$$

The left side is the averaged product of the two waveforms and thus represents the correlator output. The output of the digital correlator after  $N_N$  samples is

$$r_{\infty} = N_N^{-1} \sum_{i=1}^{N_N} x_i y_i , \quad (8.8)$$

where the subscript  $N$  denotes the Nyquist rate. Since the samples  $x_i$  and  $y_i$  obey the same Gaussian statistics as the continuous waveforms  $x(t)$  and  $y(t)$ , we can clearly write

$$\langle r_\infty \rangle = \rho \sigma^2 . \quad (8.9)$$

Thus, the output of the correlator is a linear measure of the correlation  $\rho$ . The variance of the correlator output is

$$\sigma_\infty^2 = \langle r_\infty^2 \rangle - \langle r_\infty \rangle^2 , \quad (8.10)$$

and

$$\begin{aligned} \langle r_\infty^2 \rangle &= N_N^{-2} \sum_{i=1}^{N_N} \sum_{k=1}^{N_N} \langle x_i y_i x_k y_k \rangle \\ &= N_N^{-2} \sum_{i=1}^{N_N} \langle x_i y_i \rangle^2 + N_N^{-2} \sum_{i=1}^{N_N} \sum_{k \neq i}^{N_N} \langle x_i y_i x_k y_k \rangle , \end{aligned} \quad (8.11)$$

where we have separated the terms for which  $i = k$  and  $i \neq k$ . The first summation on the right side of Eq. (8.11) has a value of  $\sigma^4(1 + 2\rho^2)N_N^{-1}$ : from Eq. (8.3), it can be shown that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^2 y^2 p(x, y) dx dy = \sigma^4(1 + 2\rho^2) . \quad (8.12)$$

The second summation term in Eq. (8.11) is readily evaluated by using the fourth-order moment relation in Eq. (6.36). Because successive samples of each signal are uncorrelated (a rectangular passband is assumed),  $\langle x_i y_i x_k y_k \rangle = \langle x_i y_i \rangle \langle x_k y_k \rangle$ , and the second summation term has a value of  $(1 - N_N^{-1})\rho^2 \sigma^4$ . Returning to Eq. (8.10), we can write

$$\begin{aligned} \sigma_\infty^2 &= (1 + 2\rho^2)\sigma^4 N_N^{-1} + (1 - N_N^{-1})\rho^2 \sigma^4 - \rho^2 \sigma^4 \\ &= \sigma^4 N_N^{-1} (1 + \rho^2) . \end{aligned} \quad (8.13)$$

The signal-to-noise ratio with unquantized sampling is

$$\mathcal{R}_{\text{sn}\infty} = \frac{\langle r_\infty \rangle}{\sigma_\infty} = \frac{\rho \sqrt{N_N}}{\sqrt{(1 + \rho^2)}} \simeq \rho \sqrt{N_N} , \quad (8.14)$$

where the approximation applies for  $\rho \ll 1$ . Note that the condition  $\rho \ll 1$  is satisfactory in many practical circumstances. For the case in which  $\rho \gtrsim 0.2$ , see Sect. 8.3.6. (The signal-to-noise ratio at the correlator output, which we are

calculating here, is of interest mainly for weak signals.) For a measurement period  $\tau$ ,  $N_N = 2\Delta\nu\tau$ , which is commonly  $10^6$ – $10^{12}$ . From Eq. (8.14), the threshold of detectability of a signal is given by  $\rho\sqrt{N_N} \simeq 1$ , that is,  $\rho \simeq 10^{-3}$ – $10^{-6}$ . In terms of the signal bandwidth and measurement duration,  $\mathcal{R}_{\text{sn}\infty} = \rho\sqrt{2\Delta\nu\tau}$ . Now for observations of a point source with identical antennas and receivers,  $\rho$  is equal to the ratio of the resulting antenna temperature to the system temperature,  $T_A/T_S$ . Thus, the present result is equal to that given by Eq. (6.45) for an analog correlator with continuous unsampled inputs and  $T_A \ll T_S$ .

Before leaving the subject of unquantized sampling, we should consider the effect of sampling at rates other than the Nyquist rate. Successive sample values from any one signal are then no longer independent. We consider a sampling frequency that is  $\beta$  times the Nyquist rate<sup>3</sup> and a number of samples  $N = \beta N_N$ . The sample interval is  $\tau_s = (2\beta\Delta\nu)^{-1}$ . Samples spaced by  $q\tau_s$ , where  $q$  is an integer, have a correlation coefficient that, from Eq. (8.4), is equal to

$$R_\infty(q\tau_s) = \frac{\sin(\pi q/\beta)}{\pi q/\beta} \quad (8.15)$$

for a rectangular baseband response. Since the samples are not independent, we must reconsider the evaluation of the second summation term on the right side of Eq. (8.11). For those terms for which  $q = |i - k|$  is small enough that  $R_\infty(q\tau_s)$  is significant, there will be an additional contribution given by

$$[\sigma^2 R_\infty(q\tau_s)]^2. \quad (8.16)$$

Now  $R_\infty^2$  is very small for all but a very small fraction of the  $N(N - 1)$  terms in the second summation in Eq. (8.11). From Eq. (8.15),  $R_\infty^2$ , at its maxima, is equal to  $(\beta/\pi q)^2$  and for  $q = 10^3$  is of order  $10^{-6}$ . However, as shown above,  $N$  is likely to be as high as  $10^6$ – $10^{12}$ . Thus, in the second summation in Eq. (8.11), the contribution made by the terms for which the  $i$  and  $k$  samples are effectively independent remains essentially unchanged. The products for which  $R_\infty^2$  is significant make an additional contribution equal to

$$2\sigma^4 N^{-2} \sum_{q=1}^{N-1} (N - q) R_\infty^2(q\tau_s) \simeq 2\sigma^4 N^{-1} \sum_{q=1}^{\infty} R_\infty^2(q\tau_s). \quad (8.17)$$

The variance of the correlator output now becomes

$$\sigma_\infty^2 = \sigma^4 N^{-1} \left[ 1 + 2 \sum_{q=1}^{\infty} R_\infty^2(q\tau_s) \right], \quad (8.18)$$

---

<sup>3</sup> $\beta$  is referred to as the oversampling factor.

and the signal-to-noise ratio of the correlation measurement is (see Appendix 8.1)

$$\mathcal{R}_{\text{sn}\infty} = \frac{\rho \sqrt{\beta N_N}}{\sqrt{1 + 2 \sum_{q=1}^{\infty} R_{\infty}^2(q\tau_s)}}. \quad (8.19)$$

Compare this result with Eq. (8.14) for Nyquist sampling. For values of  $\beta$  of  $\frac{1}{2}$ ,  $\frac{1}{3}$ ,  $\frac{1}{4}$ , and so on, which correspond to undersampling,  $R_{\infty} = 0$  and the denominator in Eq. (8.19) is unity. The sensitivity thus drops as one would expect from the decrease in the data. For oversampling,  $\beta > 1$ , and the summation of  $R_{\infty}^2(q\tau_s)$  in Eq. (8.19) is shown in Appendix 8.1 to be equal to  $(\beta - 1)/2$ . The denominator in Eq. (8.19) is then equal to  $\sqrt{\beta}$ , so the sensitivity is the same as that for sampling at the Nyquist rate. This is as expected, since in Nyquist sampling, no information is lost, and thus there is none to be gained by increased sampling. The result is different for quantized sampling, as will appear in the following sections.

### 8.3 Sampling with Quantization

In some sampling schemes, the signal is first quantized and then sampled, and in others, it is sampled and then quantized. Ideally, the end result is the same in either case, and in analyzing the process, we can choose the order that is most convenient. Suppose that a bandlimited signal is first quantized and then sampled. Quantization generates new frequency components in the signal waveform, so it is no longer bandlimited. If it is sampled at the Nyquist rate corresponding to the unquantized waveform, as is the usual practice, some information will be lost, and the sensitivity will be less than for unquantized sampling. Also, because quantization is a nonlinear operation, we cannot assume that the measured correlation of the quantized waveforms will be a linear function of  $\rho$ , which is what we want to measure. Thus, to utilize digital signal processing, there are three main points that should be investigated: (1) the relation between  $\rho$  and the measured correlation, (2) the loss in sensitivity, and (3) the extent to which oversampling can restore the lost sensitivity. Investigations of these points can be found in the work of Weinreb (1963), Cole (1968), Burns and Yao (1969), Cooper (1970), Hagen and Farley (1973), Bowers and Klingler (1974), Jenet and Anderson (1998), and Gwinn (2004).

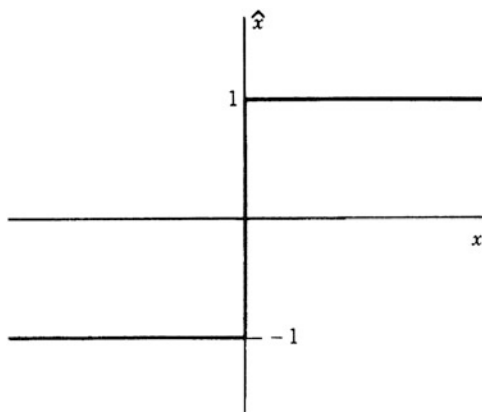
Note that in discussing sampling with quantization, it is common practice to refer to Nyquist sampling when what is meant is sampling at the Nyquist rate for the *unquantized* waveform. We also follow this usage.

#### 8.3.1 Two-Level Quantization

Sampling with two-level (one bit) quantization provided the earliest digital form of radio astronomy signals (Weinreb 1963). Although larger numbers of levels are



**Fig. 8.3** Characteristic curve for two-level quantization. The abscissa is the input voltage  $x$  and the ordinate is the quantized output  $\hat{x}$ .



now routinely used, this subsection is included as an introduction to the subject. The quantization characteristic for two-level sampling is shown in Fig. 8.3. The quantizing action senses only the sign of the instantaneous signal voltage. In many samplers, the signal voltage is first amplified and strongly clipped. The zero crossings are more sharply defined in the resulting waveform, and errors that might occur if the sampling time coincides with a sign reversal are thereby minimized.

The correlator for two-level signals consists of a multiplying circuit followed by a counter that sums the products of the input samples. The input signals are assigned values of +1 or -1 to indicate positive or negative signal voltages, and the products at the multiplier output thus take values of +1 or -1 for identical or different input values, respectively. We consider sampling both at the Nyquist rate and at multiples of it and represent by  $N$  the number of sample pairs fed to the correlator. The two-level correlation coefficient is

$$\rho_2 = \frac{(N_{11} + N_{\bar{1}\bar{1}}) - (N_{1\bar{1}} + N_{\bar{1}1})}{N}, \quad (8.20)$$

where  $N_{11}$  is the number of products for which both samples have the value +1,  $N_{\bar{1}\bar{1}}$  is the number of products in which the  $x$  sample has the value +1 and the  $y$  sample -1, and so on. The denominator in Eq. (8.20) is equal to the output that would occur if, for each sample pair, the signs of the signals were identical.  $\rho_2$  can be related to the correlation coefficient  $\rho$  of the unquantized signals through the bivariate probability distribution Eq. (8.1), from which

$$P_{11} = \frac{N_{11}}{N} = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty \exp\left[\frac{-(x^2 + y^2 - 2\rho xy)}{2\sigma^2(1-\rho^2)}\right] dx dy, \quad (8.21)$$

where  $P_{11}$  is the probability of the two unquantized signals being simultaneously greater than zero. The other required probabilities are obtained by changing the limits of the integrals in Eq. (8.21) as follows:  $\int_{-\infty}^0 \int_{-\infty}^0$  for  $P_{\bar{1}\bar{1}}$ ;  $\int_{-\infty}^0 \int_0^\infty$  for  $P_{\bar{1}1}$ ;

and  $\int_0^\infty \int_{-\infty}^0$  for  $P_{1\bar{1}}$ . Note that  $P_{11} = P_{\bar{1}\bar{1}}$  and  $P_{1\bar{1}} = P_{\bar{1}1}$ . Thus,

$$\rho_2 = 2(P_{11} - P_{1\bar{1}}) . \quad (8.22)$$

The integral in Eq. (8.21) is evaluated in Appendix 8.2, from which we obtain

$$P_{11} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho . \quad (8.23)$$

Similarly,

$$P_{1\bar{1}} = \frac{1}{4} - \frac{1}{2\pi} \sin^{-1} \rho , \quad (8.24)$$

so

$$\rho_2 = \frac{2}{\pi} \sin^{-1} \rho . \quad (8.25)$$

Equation (8.25), known as the Van Vleck relationship,<sup>4</sup> allows  $\rho$  to be obtained from the measured correlation  $\rho_2$ . For small values,  $\rho$  is proportional to  $\rho_2$ .

To determine the signal-to-noise ratio of the correlation measurement, we now calculate  $\sigma_2^2$ , the variance of the correlator output  $r_2$ :

$$\sigma_2^2 = \langle r_2^2 \rangle - \langle r_2 \rangle^2 , \quad (8.26)$$

where

$$r_2 = N^{-1} \sum_{i=1}^N \hat{x}_i \hat{y}_i . \quad (8.27)$$

In this chapter, the circumflex ( $\hat{\phantom{x}}$ ) is used to denote quantized signal waveforms. Since  $\rho_2 = \langle \hat{x} \hat{y} \rangle$ , then from Eq. (8.27),  $\langle r_2 \rangle = \rho_2$ . Thus,  $r_2$  is an unbiased estimator of  $\rho_2$ . The expression for  $\langle r_2^2 \rangle$  is equivalent to Eq. (8.11) for unquantized waveforms:

$$\langle r_2^2 \rangle = N^{-2} \sum_{i=1}^N \langle \hat{x}_i^2 \hat{y}_i^2 \rangle + N^{-2} \sum_{i=1}^N \sum_{k \neq i} \langle \hat{x}_i \hat{y}_i \hat{x}_k \hat{y}_k \rangle . \quad (8.28)$$

---

<sup>4</sup>This result was first derived by J. H. Van Vleck during World War II in a classified report, when studying the power spectrum of strongly clipped noise, which was used for electromagnetic jamming (Van Vleck 1943). The work was later declassified, and a brief summary of it appeared in Vol. 24 of MIT's Radiation Laboratory Series (Lawson and Uhlenbeck 1950). A fuller account was given by Van Vleck and Middleton (1966).

The first summation term on the right side of Eq. (8.28) is equal to  $N^{-1}$  since the products  $\hat{x}_i \hat{y}_i$  take values of  $\pm 1$  for two-level sampling. In evaluating the second summation term, the situation is similar to that for unquantized sampling. The factor  $\sigma^4$  in Eq. (8.17) is here replaced by the square of the variance of the quantized waveform, which is unity for two-level quantization. For all except a small fraction of the terms,  $q = |i-k|$  is large enough that samples  $i$  and  $k$  from the same waveform are uncorrelated. These terms make a total contribution closely equal to  $\rho_2^2$ . Those terms for which samples  $i$  and  $k$  are correlated make an additional contribution closely equal to

$$2N^{-1} \sum_{q=1}^{\infty} R_2^2(q\tau_s) , \quad (8.29)$$

where  $R_2(\tau)$  is the autocorrelation coefficient for a signal after two-level quantization. Thus,

$$\sigma_2^2 = N^{-1} + (1 - N^{-1})\rho_2^2 + 2N^{-1} \sum_{q=1}^{\infty} R_2^2(q\tau_s) - \rho_2^2 \quad (8.30a)$$

$$\simeq N^{-1} \left[ 1 + 2 \sum_{q=1}^{\infty} R_2^2(q\tau_s) \right] , \quad (8.30b)$$

where we have assumed that  $\rho_2 \ll 1$  and also that the term  $-N^{-1}\rho_2^2$  can be neglected, since here we are mostly interested in signals near the threshold of detectability. Then the signal-to-noise ratio is

$$\mathcal{R}_{\text{sn2}} = \frac{\langle r_2 \rangle}{\sigma_2} = \frac{2\rho\sqrt{N}}{\pi \sqrt{1 + 2 \sum_{q=1}^{\infty} R_2^2(q\tau_s)}} . \quad (8.31)$$

This ratio, relative to that for unquantized sampling at the Nyquist rate given by Eq. (8.14), defines an efficiency factor for the quantized correlation process:

$$\eta_2 = \frac{\mathcal{R}_{\text{sn2}}}{\mathcal{R}_{\text{sn}\infty}} = \frac{2\sqrt{\beta}}{\pi \sqrt{1 + 2 \sum_{q=1}^{\infty} R_2^2(q\tau_s)}} . \quad (8.32)$$

Here, we have used  $N = \beta N_N$ , so we are considering the same observing time as in the Nyquist-sampled case but sampling  $\beta$  times as rapidly. Note that  $\tau_s$  is correspondingly reduced.  $\eta_2$  is one case of the general quantization efficiency factor,  $\eta_Q$  (introduced in Sect. 6.2), where  $Q$  is the number of quantization levels.

Equation (8.25) gives the relationship between the correlation coefficients for a pair of signals before and after two-level quantization. This result includes the case of autocorrelation in which the two signals differ only because of a delay. Thus, we

may write

$$R_2(q\tau_s) = \frac{2}{\pi} \sin^{-1} [R_\infty(q\tau_s)] . \quad (8.33)$$

Equation (8.15) gives  $R_\infty(q\tau_s)$  for a rectangular baseband signal spectrum sampled at  $\beta$  times the Nyquist rate, and Eq. (8.33) becomes

$$R_2(q\tau_s) = \frac{2}{\pi} \sin^{-1} \left[ \frac{\beta \sin(\pi q/\beta)}{\pi q} \right] . \quad (8.34)$$

$R_2(q\tau_s)$  thus has zeros at the same values of  $q\tau_s$  that  $R_\infty(q\tau_s)$  does (the principal value is taken for the inverse sine function), and for  $\beta = 1, \frac{1}{2}, \frac{1}{3}$ , and so on, we obtain

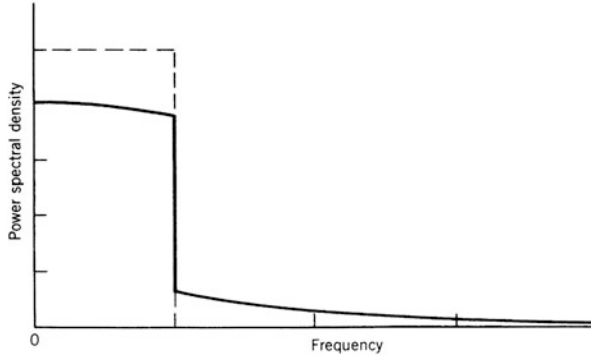
$$\sum_{q=1}^{\infty} R_2^2(q\tau_s) = 0 . \quad (8.35)$$

In these cases, the signal-to-noise ratio is a factor of  $2/\pi$  ( $= 0.637$ ) times that for unquantized sampling at the same rate given in Eq. (8.15). For oversampling with  $\beta = 2$  and  $\beta = 3$ , the corresponding signal-to-noise factors from Eqs. (8.32) and (8.34) are 0.744 and 0.773, respectively. Note, however, that the increased bit rate used in oversampling could produce a bigger increase in the signal-to-noise ratio if used to increase the number of quantization levels. Doubling the bit rate could be used to increase the number of levels to four, for which the signal-to-noise factor is 0.881 (as derived in Sect. 8.3.3). For a bit rate increase of three, the number of levels could be increased to eight, for which the signal-to-noise factor is 0.963. Note also that in the calculations given above, there is an implicit dependence on the bandpass shape of the signal through the assumption that  $\rho_2 \ll 1$  for samples for which  $i$  is not equal to  $k$  in Eq. (8.28). For  $\beta \geq 2$ , a further dependence on the bandpass shape enters through the autocorrelation function  $R_2(q\tau_s)$ .

It has been mentioned that quantization generates additional spectral components. We can compare the power spectra of a signal before and after quantization since these spectra are the Fourier transforms of autocorrelation functions that are related by Eq. (8.25). Figure 8.4 shows the spectrum, after two-level quantization, of noise with an originally rectangular spectrum. A fraction of the original bandlimited spectrum is converted into a broad, low-level skirt that dies away very slowly with frequency.

### 8.3.2 Four-Level Quantization

The use of two digital bits to represent the amplitude of each sample results in less degradation of the signal-to-noise ratio than is obtained with one-bit quantization.



**Fig. 8.4** Spectra of rectangular bandpass noise before and after two-level quantization. The unquantized spectrum is of lowpass form, as shown by the broken line. The spectrum after quantization is shown by the solid curve. The power levels of the two waveforms (represented by the areas under the curves) are equal, and the Fourier transforms of their spectra are related by Eq. (8.25).

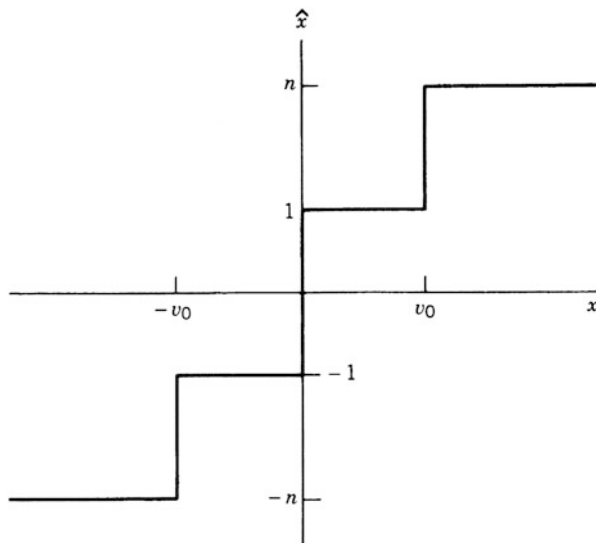
Consideration of two-bit sampling leads naturally to four-level quantization, the performance of which has been investigated by several authors, notably Cooper (1970) and Hagen and Farley (1973). The quantization characteristic is shown in Fig. 8.5, where the quantization thresholds are  $-v_0$ ,  $0$ , and  $v_0$ . The four quantization states have designated values  $-n$ ,  $-1$ ,  $+1$ , and  $+n$ , where  $n$ , which is not necessarily an integer, can be chosen to optimize the performance. Products of two samples can take the value  $\pm 1$ ,  $\pm n$ , or  $\pm n^2$ . The four-level correlation coefficient  $\rho_4$  can be specified by an expression similar to Eq. (8.20) for the two-level case, that is,

$$\rho_4 = \frac{2n^2N_{nn} - 2n^2N_{n\bar{n}} + 4nN_{1n} - 4nN_{1\bar{n}} + 2N_{11} - 2N_{1\bar{1}}}{(2n^2N_{nn} + 2N_{11})_{\rho=1}}, \quad (8.36)$$

where a bar on the subscript indicates a negative sign. The numerator is proportional to the correlator output and reduces to the form in the denominator for  $\rho = 1$ , that is, when the two input waveforms are identical. The numbers of the various level combinations can be derived from the corresponding joint probabilities. Thus, for example,

$$\begin{aligned} N_{nn} &= NP_{nn} \\ &= \frac{N}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_{v_0}^{\infty} \int_{v_0}^{\infty} \exp\left[\frac{-(x^2 + y^2 - 2\rho xy)}{2\sigma^2(1-\rho^2)}\right] dx dy, \end{aligned} \quad (8.37)$$

and, as in the two-level case, the other probabilities are obtained by using the appropriate limits for the integrals. For the case of  $\rho \ll 1$ , the approximate form of the probability distribution in Eq. (8.2) simplifies the calculation.



**Fig. 8.5** Characteristic curve for four-level quantization, with weighting factor  $n$  for outer levels. The abscissa is the unquantized voltage  $x$ , and the ordinate is the quantized output  $\hat{x}$ .  $v_0$  is the threshold voltage.

Although  $\rho_4$  can be evaluated from Eq. (8.36) in the above manner, an alternative derivation that provides a more rapid approach to the desired result is used here. This approach follows the treatment of Hagen and Farley (1973) and is based on a theorem by Price (1958). The form of the theorem that we require is

$$\frac{d\langle r_4 \rangle}{d\rho} = \sigma^2 \left\langle \frac{\partial \hat{x}}{\partial x} \frac{\partial \hat{y}}{\partial y} \right\rangle, \quad (8.38)$$

where  $r_4$  is the unnormalized correlator output, and  $\hat{x}$  and  $\hat{y}$  are again the quantized versions of the input signals. For four-level sampling,

$$\frac{\partial \hat{x}}{\partial x} = (n-1)\delta(x+v_0) + 2\delta(x) + (n-1)\delta(x-v_0), \quad (8.39)$$

where  $\delta$  is the delta function, and a similar expression can be written for  $\partial \hat{y} / \partial y$ . Equation (8.39) is the derivative of the function in Fig. 8.5. To determine the expectation of the product of the two derivatives on the right side of Eq. (8.38), the magnitudes of each of the nine terms in the product of the derivatives must be multiplied by the probability of occurrence. Thus, for example, the term  $(n-1)^2 \delta(x+v_0) \delta(y+v_0)$  has a magnitude of  $(n-1)^2$  and probability

$$\frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[\frac{-2v_0^2}{2\sigma^2(1+\rho)}\right]. \quad (8.40)$$

By consolidating terms with equal probabilities, we obtain

$$\frac{d\langle r_4 \rangle}{d\rho} = \frac{1}{\pi \sqrt{1-\rho^2}} \left\{ (n-1)^2 \left[ \exp\left(\frac{-v_0^2}{\sigma^2(1+\rho)}\right) + \exp\left(\frac{-v_0^2}{\sigma^2(1-\rho)}\right) \right] + 4(n-1) \exp\left(\frac{-v_0^2}{2\sigma^2(1-\rho^2)}\right) + 2 \right\}, \quad (8.41)$$

and

$$\langle r_4 \rangle = \frac{1}{\pi} \int_0^\rho \frac{1}{\sqrt{1-\xi^2}} \left\{ (n-1)^2 \left[ \exp\left(\frac{-v_0^2}{\sigma^2(1+\xi)}\right) + \exp\left(\frac{-v_0^2}{\sigma^2(1-\xi)}\right) \right] + 4(n-1) \exp\left(\frac{-v_0^2}{2\sigma^2(1-\xi^2)}\right) + 2 \right\} d\xi, \quad (8.42)$$

where  $\xi$  is a dummy variable of integration. To obtain the correlation coefficient  $\rho_4$ ,  $\langle r_4 \rangle$  must be divided by the expectation of the correlator output when the inputs are identical four-level waveforms, as in Eq. (8.36):

$$\rho_4 = \frac{\langle r_4 \rangle}{\Phi + n^2(1-\Phi)}, \quad (8.43)$$

where  $\Phi$  is the probability that the unquantized level lies between  $\pm v_0$ , that is,

$$\Phi = \frac{1}{\sigma \sqrt{2\pi}} \int_{-v_0}^{v_0} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx = \text{erf}\left(\frac{v_0}{\sigma \sqrt{2}}\right). \quad (8.44)$$

Equations (8.42)–(8.44) provide a relationship between  $\rho_4$  and  $\rho$  that is equivalent to the Van Vleck relationship for two-level quantization.

The choice of values for  $n$  and  $v_0$  is usually made to maximize the signal-to-noise ratio for weak signals, which we now derive. For  $\rho \ll 1$ , Eqs. (8.42) and (8.43) reduce to

$$(\rho_4)_{\rho \ll 1} = \rho \frac{2[(n-1)E + 1]^2}{\pi[\Phi + n^2(1-\Phi)]}, \quad (8.45)$$

where  $E = \exp(-v_0^2/2\sigma^2)$ . The variance in the measurement of  $r_4$  is

$$\sigma_4^2 = \langle r_4^2 \rangle - \langle r_4 \rangle^2 = \langle r_4^2 \rangle - \rho_4^2 [\Phi + n^2(1-\Phi)]^2. \quad (8.46)$$

The factor  $[\Phi + n^2(1-\Phi)]$  is the variance of the quantized waveform and here takes the place of  $\sigma^2$  in the corresponding equations for unquantized sampling. Again, we

follow the procedure explained for the unquantized case and write

$$\langle r_4^2 \rangle = N^{-2} \sum_{i=1}^N \langle \hat{x}_i^2 \hat{y}_i^2 \rangle + N^{-2} \sum_{i=1}^N \sum_{i \neq k} \langle \hat{x}_i \hat{y}_i \hat{x}_k \hat{y}_k \rangle. \quad (8.47)$$

To evaluate the first summation, note that  $(\hat{x}_i \hat{y}_i)^2$  can take values of 1,  $n^2$ , or  $n^4$ , and the sum of these values multiplied by their probabilities is equal to  $[\Phi + n^2(1 - \Phi)]^2$ . The contribution of the second summation is

$$(1 - N^{-1})\rho_4^2 [\Phi + n^2(1 - \Phi)]^2 + 2N^{-1}[\Phi + n^2(1 - \Phi)]^2 \sum_{q=1}^{\infty} R_4^2(q\tau_s), \quad (8.48)$$

where the second term represents the effect of oversampling and is similar to Eq. (8.17), and  $R_4$  is the autocorrelation function after four-level quantization. Thus, from Eq. (8.46), we have

$$\sigma_4^2 = N^{-1} [\Phi + n^2(1 - \Phi)]^2 \left[ 1 + 2 \sum_{q=1}^{\infty} R_4^2(q\tau_s) - \rho_4^2 \right]. \quad (8.49)$$

Since we have assumed  $\rho \ll 1$ , the  $\rho_4^2$  term can be neglected, and the signal-to-noise ratio for the four-level correlation measurement is

$$\mathcal{R}_{\text{sn}4} = \frac{\langle r_4 \rangle}{\sigma_4} = \frac{2\rho[(n-1)E + 1]^2 \sqrt{N}}{\pi [\Phi + n^2(1 - \Phi)] \sqrt{1 + 2 \sum_{q=1}^{\infty} R_4^2(q\tau_s)}}. \quad (8.50)$$

The signal-to-noise ratio relative to that for unquantized Nyquist sampling is obtained from Eq. (8.14) for  $N = \beta N_N$  and is

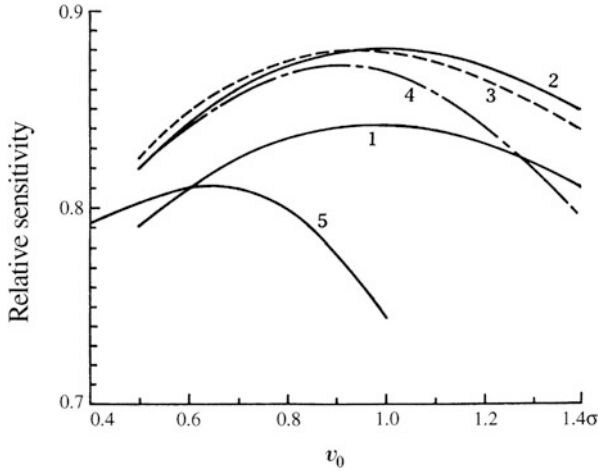
$$\eta_4 = \frac{\mathcal{R}_{\text{sn}4}}{\mathcal{R}_{\text{sn}\infty}} = \frac{2[(n-1)E + 1]^2 \sqrt{\beta}}{\pi [\Phi + n^2(1 - \Phi)] \sqrt{1 + 2 \sum_{q=1}^{\infty} R_4^2(q\tau_s)}}. \quad (8.51)$$

For sampling at the Nyquist rate,  $\beta = 1$  and

$$\eta_4 = \frac{\mathcal{R}_{\text{sn}4}}{\mathcal{R}_{\text{sn}\infty}} = \frac{2[(n-1)E + 1]^2}{\pi [\Phi + n^2(1 - \Phi)]}. \quad (8.52)$$

Values of  $\eta_4$  very close to optimum sensitivity are obtained for  $n = 3$  with  $v_0 = 0.996\sigma$ , and for  $n = 4$ , with  $v_0 = 0.942\sigma$ : see Table A8.1 in Appendix 8.3. Note that the choice of an integer for the value of  $n$  simplifies the correlator. For these two cases,  $\eta_4$ , the signal-to-noise ratio relative to that for unquantized sampling, is equal to 0.881 and 0.880, respectively. Curves of the relative sensitivity as a function of





**Fig. 8.6** Signal-to-noise ratio relative to that for unquantized correlation for the four-level system and several modifications of it. The abscissa is the quantization threshold  $v_0$  in units of the rms level of the waveforms at the quantizer input. The ordinate is sensitivity (signal-to-noise ratio) relative to an unquantized system. The curves are for: (1) full four-level system with  $n = 2$ ; (2) full four-level system with  $n = 3$ ; (3) full four-level system with  $n = 4$ ; (4) four-level system with  $n = 3$  and low-level products omitted; (5) three-level system. From Cooper (1970). © CSIRO 1970. Published by CSIRO Publishing, Melbourne, Victoria, Australia. Reproduced with permission.

$v_0/\sigma$  for  $n = 2, 3$ , and  $4$  are shown in Fig. 8.6. Similar conclusions are derived by Hagen and Farley (1973) and Bowers and Klingler (1974).

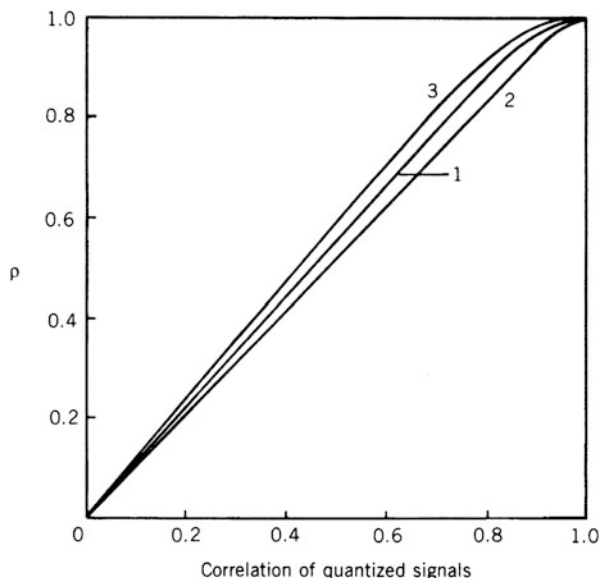
Having chosen values for  $n$  and  $v_0$ , we can now return to Eqs. (8.42) and (8.43) to examine the relationship of  $\rho$  and  $\rho_4$ . Curve 1 of Fig. 8.7 shows a plot of  $\rho$  and  $\rho_4$ . Extrapolation of a linear relationship with slope chosen to fit low values of  $\rho$  results in errors of 1% at  $\rho = 0.5$ , 2% at  $\rho = 0.7$ , and 2.8% at  $\rho = 0.8$ , where the error is a percentage of the true value of  $\rho$ . Thus, for many purposes, a linear approximation is satisfactory for values of  $\rho$  up to  $\sim 0.6$ . This linearity assumption simplifies the final step that we require in discussing four-level sampling, namely, calculation of the improvement in sensitivity resulting from oversampling.

The relationship between the autocorrelation function for unquantized noise  $R_\infty$  and that for the same waveform after four-level quantization is the same as for the corresponding cross-correlation functions in Eq. (8.45), so we can write

$$R_4 = \frac{2[(n-1)E + 1]^2 R_\infty}{\pi[\Phi + n^2(1 - \Phi)]}, \quad (8.53)$$

provided that  $R_\infty \lesssim 0.6$ . Now  $R_\infty$  as given by Eq. (8.15) fulfills this condition for  $q = 1$  with an oversampling factor  $\beta = 2$ . For  $n = 3$  and the corresponding optimum value of  $v_0$ ,  $E = 0.6091$ ,  $\Phi = 0.6806$ , and  $R_4 = 0.881R_\infty$ . For  $\beta = 2$ , we use Eqs. (8.15) and (8.53) and Eq. (A8.5) of Appendix 8.1 to evaluate the summation

**Fig. 8.7** Correlation coefficient  $\rho$  for unquantized signals plotted as a function of the correlation that would be measured after quantization. The curves are for: (1) full four-level system with  $n = 3$  and  $v_0 = \sigma$ , or  $n = 4$  and  $v_0 = 0.95\sigma$ ; (2) four-level system with low-level products omitted,  $n = 4$  and  $v_0 = 0.9\sigma$ ; (3) three-level system with  $v_0 = 0.6\sigma$ . From Cooper (1970). © CSIRO 1970. Published by CSIRO Publishing, Melbourne, Victoria, Australia. Reproduced with permission.



in the denominator of Eq. (8.51), and obtain  $\eta_4 = 0.935$ , which is a factor of 1.06 greater than for  $\beta = 1$ . Bowers and Klingler (1974) have pointed out that the optimum value of the quantization level  $v_0$  changes slightly with the oversampling factor. However, the optimum values are rather broad (see Fig. 8.6), and the effect on the sensitivity is very small.

In a discussion of two-bit quantization, Cooper (1970) considered the effect of omitting certain products in the multiplication process. For example, if all products of the two low-level bits are counted as zero instead of  $\pm 1$ , the loss in signal-to-noise ratio is approximately 1%, as shown in curve 4 of Fig. 8.6. The products to be accumulated are then only those counted as  $\pm n$  and  $\pm n^2$  in the full four-level system described above, and in the modified system, they can be assigned values of  $\pm 1$  and  $\pm n$ , respectively, thereby simplifying the counter circuitry of the integrator. An even greater simplification can be accomplished by omitting the intermediate-level products also and assigning values  $\pm 1$  to the high-level products. This last type of modification yields 92% of the sensitivity of a full four-level correlator. We shall not analyze the case where only the low-level products are omitted, but we note that to derive the correlation coefficient as a function of  $\rho$ , one can express the action of the correlator in terms of two different quantization characteristics (Hagen and Farley 1973) or else return to Eq. (8.36) and omit the appropriate terms. If both the low- and intermediate-level products are omitted, however, the action can be described more simply in terms of a new quantization characteristic, known as three-level quantization, without arbitrary omission of product terms.

### 8.3.3 Three-Level Quantization

Three-level quantization has proved to be an important practical technique, and the quantization characteristic is shown in Fig. 8.8. In this case, the approach using Price's theorem will again be followed.

The expressions for the operating characteristics of a three-level correlator can be obtained from those in the preceding section by omitting the terms that refer to low- and intermediate-level products and adjusting the weighting factors as appropriate. Thus, the equivalent derivative needed in Price's theorem is,

$$\frac{\partial \hat{x}}{\partial x} = \delta(x - v_0) + \delta(x + v_0) , \quad (8.54)$$

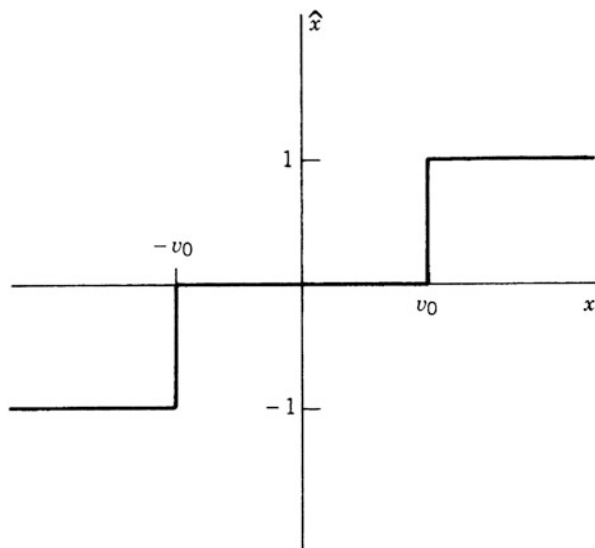
and the expectation of the correlator output  $\langle r_3 \rangle$  is, from Price's theorem,

$$\langle r_3 \rangle = \frac{1}{\pi} \int_0^\rho \frac{1}{\sqrt{1 - \xi^2}} \left[ \exp\left(\frac{-v_0^2}{\sigma^2(1 + \xi)}\right) + \exp\left(\frac{-v_0^2}{\sigma^2(1 - \xi)}\right) \right] d\xi , \quad (8.55)$$

where  $\xi$  is a dummy variable of integration. The normalized correlation coefficient is

$$\rho_3 = \frac{\langle r_3 \rangle}{1 - \Phi} , \quad (8.56)$$

**Fig. 8.8** Characteristic curve for three-level quantization. The abscissa is the unquantized voltage  $x$ , and the ordinate is the quantized output  $\hat{x}$ .  $v_0$  is the threshold voltage. Since the magnitude of  $\hat{x}$  takes only one nonzero value, it is perfectly general to set this value to unity.



where  $\Phi$  is given by Eq. (8.44). For  $\rho \ll 1$ , Eqs. (8.55) and (8.56) yield

$$(\rho_3)_{\rho \ll 1} = \rho \frac{2E^2}{\pi(1 - \Phi)}, \quad (8.57)$$

where  $E$  is defined following Eq. (8.45). The variance of  $r_3$  is

$$\sigma_3^2 = \langle r_3^2 \rangle - \langle r_3 \rangle^2 = N^{-1}(1 - \Phi)^2 \left[ 1 + 2 \sum_{q=1}^{\infty} R_3^2(q\tau_s) - \rho_3^2 \right], \quad (8.58)$$

where  $R_3$  is the autocorrelation coefficient after three-level quantization. If  $\rho_3^2$  in Eq. (8.58) can be neglected, the signal-to-noise ratio relative to a nonquantizing correlator is

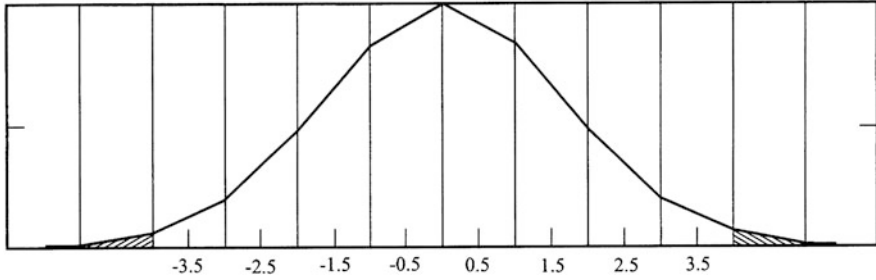
$$\eta_3 = \frac{\mathcal{R}_{\text{sn}3}}{\mathcal{R}_{\text{sn}\infty}} = \frac{\langle r_3 \rangle}{\sigma_3 \mathcal{R}_{\text{sn}\infty}} = \frac{2\sqrt{\beta}E^2}{\pi(1 - \Phi)\sqrt{1 + 2 \sum_{q=1}^{\infty} R_3^2(q\tau_s)}}. \quad (8.59)$$

For Nyquist sampling, the maximum sensitivity relative to the nonquantizing case is obtained with  $v_0 = 0.6120\sigma$ , for which  $\eta_3$  is equal to 0.810 (see curve 5 of Fig. 8.6). With this optimized threshold value,  $\Phi = 0.4595$ ,  $E = 0.8292$ , and we can write  $R_3(q\tau_s) = 0.810R_{\infty}(q\tau_s)$ , assuming that  $\rho$  is an approximately linear function of  $r_3$ . Then from Eqs. (8.15), (8.59), and Eq. (A8.5), we find that for a rectangular baseband spectrum with the oversampling factor  $\beta = 2$ ,  $\eta_3$  becomes 0.890, which is a factor of 1.10 greater than for  $\beta = 1$ .

### 8.3.4 Quantization Efficiency: Simplified Analysis for Four or More Levels

For quantization into two, three, or four levels, the quantization efficiency,  $\eta_Q$ , is 0.636, 0.810, and 0.881. For more quantization levels, the loss in efficiency resulting from the quantization decreases further, and an approximate method of calculating the loss (Thompson 1998) can be used, as follows. This is simpler than the more accurate method given in Sect. 8.3.3. In either case, the principle is to calculate the fractional increase in the variance of a signal that results from the quantization. The signal-to-noise ratio at the correlator output is inversely proportional to this variance.

Figure 8.9 shows a piecewise linear approximation of the Gaussian probability distribution of a signal from one antenna. This approximation simplifies the analysis. The intersections with the vertical lines indicate exact values of the Gaussian. For eight-level sampling, the quantization thresholds are indicated by the positions of the vertical lines between the numbers  $\pm 3.5$  on the abscissa. The horizontal



**Fig. 8.9** Piecewise linear representation of the Gaussian probability distribution of the amplitude of a signal within the receiver. The intersections of the curve with the vertical lines denote exact values of the Gaussian. The abscissa is the signal amplitude (voltage) in units of  $\epsilon\sigma$ , and the numbers indicate the values assigned to the levels after quantization. For eight-level sampling the quantization thresholds are indicated by the seven vertical lines that lie between  $-3.5\epsilon\sigma$  and  $3.5\epsilon\sigma$  on the abscissa. For signal levels outside the range  $\pm 4\epsilon\sigma$ , indicated by the shaded areas, the assigned values are  $\pm 3.5\epsilon\sigma$ .

spacing between adjacent levels is represented by  $\epsilon$ , in units of the (unquantized) rms voltage,  $\sigma$ , i.e.,  $\epsilon\sigma$  is the spacing between the levels in volts. We consider first the case in which the number of levels is even, as in Fig. 8.9. Any one sample that falls between the two consecutive thresholds at  $m\epsilon\sigma$  and  $(m+1)\epsilon\sigma$  will be assigned a value  $(m + \frac{1}{2})\epsilon\sigma$ . The normalized trapezoidal probability distribution for the voltage in this segment of the overall probability distribution in Fig. 8.9 can be written as

$$p(v) = \frac{1}{\epsilon\sigma} \left[ v - \left( m + \frac{1}{2} \right) \epsilon\sigma \right] \Delta_m \quad m\epsilon\sigma < v < (m+1)\epsilon\sigma, \quad (8.60)$$

where  $\Delta_m$  is the change in probability, over the voltage range  $m\epsilon\sigma$  to  $(m+1)\epsilon\sigma$ . The extra variance that is incurred by quantizing the voltage is

$$\left\langle \left[ v - \left( m + \frac{1}{2} \right) \epsilon\sigma \right]^2 \right\rangle = \int_{m\epsilon\sigma}^{(m+1)\epsilon\sigma} \left[ v - \left( m + \frac{1}{2} \right) \epsilon\sigma \right]^2 p(v) dv. \quad (8.61)$$

If we make the substitution  $x = v - (m + \frac{1}{2})\epsilon\sigma$ , the excess variance becomes

$$\int_{-\epsilon\sigma/2}^{\epsilon\sigma/2} x^2 \left[ \frac{1}{\epsilon\sigma} + x\Delta_m \right] dx, \quad (8.62)$$

or

$$\frac{2}{\epsilon\sigma} \int_0^{\epsilon\sigma/2} x^2 dx = \frac{1}{3} \left( \frac{\epsilon\sigma}{2} \right)^2. \quad (8.63)$$

Note that the  $\Delta_m$  factor does not appear in Eq. (8.63). Hence, the excess variance is the same for all voltage bins from  $-4\epsilon\sigma$  to  $4\epsilon\sigma$ . The fraction of the area under the Gaussian probability curve that lies between these levels is

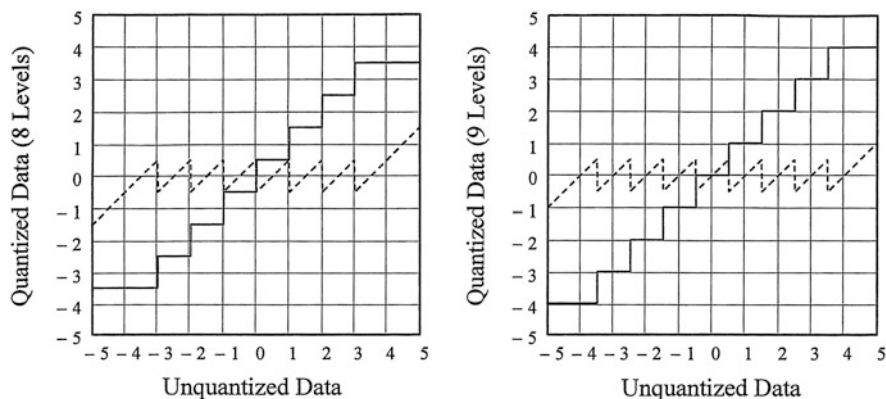
$$\frac{1}{\sqrt{2\pi}\sigma} \int_{-4\epsilon\sigma}^{4\epsilon\sigma} e^{-x^2/2\sigma^2} dx = \operatorname{erf}\left(\frac{4\epsilon}{\sqrt{2}}\right). \quad (8.64)$$

Thus, the variance resulting from quantization of the signal samples with amplitudes in the range  $\pm 4\epsilon\sigma$  is

$$\frac{1}{3} \left(\frac{\epsilon\sigma}{2}\right)^2 \operatorname{erf}\left(\frac{4\epsilon}{\sqrt{2}}\right). \quad (8.65)$$

We shall assume that the quantization error is essentially uncorrelated with the unquantized signal. In the extreme case of two-level sampling, the quantization error is highly correlated with the unquantized signal, so the treatment used here would not apply. Consider, however, the case of multilevel quantization, as in Fig. 8.10. If the signal voltage is increased steadily, the quantization error decreases from a maximum at each quantization threshold to zero when the voltage is equal to the midpoint of two thresholds. At each threshold, the quantization error changes sign, and the cycle repeats. This behavior greatly reduces any correlation between the quantization error and the signal waveform.

It is also necessary to take account of the effect of counting all signals below  $-4\epsilon\sigma$  as level  $-3.5\epsilon\sigma$ , and those above  $+4\epsilon\sigma$  as  $+3.5\epsilon\sigma$ . To make an approximate



**Fig. 8.10** Examples of quantization characteristics for (left diagram) an even number of levels (eight), and (right diagram) an odd number of levels (nine). Units on both axes are equal to  $\epsilon$ . The abscissa is the analog (unquantized) voltage, and the ordinate is the quantized output. The dotted curves show the analog level minus the quantized level. Note that for even numbers of levels, the thresholds occur at integral values on the abscissa, whereas for odd numbers of levels, they occur at values that are an integer plus one-half.

estimate of this effect, we divide the range of signal level outside of  $\pm 4\epsilon\sigma$  into intervals of width  $\epsilon\sigma$ . Consider, for example, the interval centered on  $6.5\epsilon\sigma$ . The probability of the signal falling within this level is equal to the corresponding area under the curve, which for the piecewise linear approximation is

$$\frac{1}{2} \frac{\epsilon}{\sqrt{2\pi}} \left[ e^{-(6\epsilon)^2/2} + e^{-(7\epsilon)^2/2} \right]. \quad (8.66)$$

The variance resulting from quantization of the signal within this range is closely approximated by  $[(6.5 - 3.5)\epsilon\sigma]^2$ , so the total variance of the quantization error for signals outside the range  $\pm 4\epsilon\sigma$  is

$$\frac{\epsilon^3 \sigma^2}{\sqrt{2\pi}} \sum_{m=4}^{\infty} (m-3)^2 \left[ e^{-m^2 \epsilon^2/2} + e^{-(m+1)^2 \epsilon^2/2} \right]. \quad (8.67)$$

In practice, the summation in (8.67) converges rapidly, and only a few terms are needed (i.e., those for  $m\epsilon \lesssim 3$ ). The quantization error resulting from the truncation of the signal values outside the range  $\pm 4\epsilon\sigma$  clearly has some degree of correlation with the unquantized signal level. However, this is a small effect because the fraction of samples for which the signal lies outside  $\pm 4\epsilon\sigma$  is less than 1.6% for eight-level quantization, with  $\epsilon$  optimized for sensitivity. The percentage decreases as the number of quantization levels increases. We shall therefore treat the quantization error resulting from the truncation of the signal peaks as uncorrelated with the signal, but bear in mind that this assumption may introduce a small uncertainty into the calculation.

The variance of the quantized signal is equal to the variance of the unquantized signal ( $\sigma^2$ ) plus the variance of the quantization errors in (8.65) and (8.67), that is,

$$\sigma^2 + \frac{1}{3} \left( \frac{\epsilon\sigma}{2} \right)^2 \operatorname{erf} \left( \frac{4\epsilon}{\sqrt{2}} \right) + \frac{\epsilon^3 \sigma^2}{\sqrt{2\pi}} \sum_{m=4}^{\infty} (m-3)^2 \left[ e^{-m^2 \epsilon^2/2} + e^{-(m+1)^2 \epsilon^2/2} \right]. \quad (8.68)$$

If the variance is the same for both signals at the correlator input, and if the correlation of the signals is small (i.e.,  $\rho \ll 1$ ), then the signal-to-noise ratio at the correlator output is inversely proportional to the variance. Thus, the quantization efficiency is

$$\eta_{(2N)} = \left\{ 1 + \frac{1}{3} \left( \frac{\epsilon}{2} \right)^2 \operatorname{erf} \left( \frac{N\epsilon}{\sqrt{2}} \right) + \frac{\epsilon^3}{\sqrt{2\pi}} \sum_{m=N}^{\infty} (m-N+1)^2 \left[ e^{-m^2 \epsilon^2/2} + e^{-(m+1)^2 \epsilon^2/2} \right] \right\}^{-1}. \quad (8.69)$$

**Table 8.1** Quantization efficiency and other factors for four or more levels

| Number of levels ( $Q$ ) | $N$ | $\epsilon$ | $P$    | $\eta_Q$ |
|--------------------------|-----|------------|--------|----------|
| 4                        | 2   | 1.08       | 0.03   | 0.86     |
| 8                        | 4   | 0.60       | 0.016  | 0.960    |
| 9                        | 4   | 0.55       | 0.013  | 0.968    |
| 16                       | 8   | 0.34       | 0.006  | 0.988    |
| 32                       | 16  | 0.19       | 0.002  | 0.996    |
| 256                      | 16  | 0.03       | <0.001 | 1.000    |

Here, the equation has been generalized for  $2N$  levels. For an odd number of levels,  $2N + 1$ , one of which is centered on zero signal level, the equivalent equation for the quantization efficiency is

$$\eta_{(2N+1)} = \left\{ 1 + \frac{1}{3} \left( \frac{\epsilon}{2} \right)^2 \operatorname{erf} \left( \frac{(N + \frac{1}{2})\epsilon}{\sqrt{2}} \right) + \frac{\epsilon^3}{\sqrt{2\pi}} \sum_{m=N+1}^{\infty} (m - N)^2 \left[ e^{-(m-\frac{1}{2})^2 \epsilon^2/2} + e^{-(m+\frac{1}{2})^2 \epsilon^2/2} \right] \right\}^{-1}. \quad (8.70)$$

Results from Eqs. (8.69) and (8.70) are given in Table 8.1. The values of  $\epsilon$  are those that maximize  $\eta_Q$ . The fourth column of the table gives  $P$ , which is the fraction of samples for which the signal amplitude is greater than  $\pm N\epsilon\sigma$  for an even number of levels or greater than  $\pm (N + \frac{1}{2})\epsilon\sigma$  for an odd number of levels. For eight levels,  $P$  is the fraction of signal samples that contribute to the variance in (8.67). The values of  $\eta_Q$  calculated here are accurate to about 2% for  $Q = 4$  and to 0.1% for  $Q = 8$  and higher.

### 8.3.5 Quantization Efficiency: Full Analysis, Three or More Levels

This section presents a general analysis of quantized systems for three or more levels, [e.g., Thompson et al. (2007)]. Let  $x$  represent the voltage of the unquantized signal samples, which have a Gaussian probability distribution with variance  $\sigma^2$ . Let  $\hat{x}$  represent the quantized values of  $x$ . The difference  $x - \hat{x}$  represents an inequality introduced by the quantization. The inequality contains a component that is correlated with  $x$ , and an uncorrelated component that behaves much like random noise. Consider the correlation coefficient between  $x$  and  $x' = x - \alpha\hat{x}$ , where  $\alpha$  is a scaling factor. The correlation coefficient is

$$\frac{\langle xx' \rangle}{x_{\text{rms}} x'_{\text{rms}}} = \frac{\langle x^2 \rangle - \alpha \langle x\hat{x} \rangle}{x_{\text{rms}} x'_{\text{rms}}}. \quad (8.71)$$



Here, the angle brackets  $\langle \rangle$  indicate the mean value. If  $\alpha = \langle x^2 \rangle / \langle x\hat{x} \rangle$ , then the correlation coefficient is zero, and  $x'$  represents purely random noise. We refer to this random component as the quantization noise, equal to  $x - \alpha_1 \hat{x}$ , where  $\alpha_1 = \langle x^2 \rangle / \langle x\hat{x} \rangle$ . Without loss of generality, we take  $\sigma^2 = \langle x^2 \rangle = 1$  in this analysis and use  $\alpha_1 = 1 / \langle x\hat{x} \rangle$ . The variance of the quantization noise is

$$\langle q^2 \rangle = \langle (x - \alpha_1 \hat{x})^2 \rangle = \langle x^2 \rangle - 2\alpha_1 \langle x\hat{x} \rangle + \alpha_1^2 \langle \hat{x}^2 \rangle = \alpha_1^2 \langle \hat{x}^2 \rangle - 1. \quad (8.72)$$

The total variance of the digitized signal is  $1 + \langle q^2 \rangle$ , and the quantization efficiency  $\eta_Q$  is equal to the variance of the unquantized signal expressed as a fraction of the total variance. Thus,

$$\eta_Q = \frac{1}{(1 + \langle q^2 \rangle)} = \frac{1}{\alpha_1^2 \langle \hat{x}^2 \rangle} = \frac{\langle x\hat{x} \rangle^2}{\langle \hat{x}^2 \rangle}. \quad (8.73)$$

Consider the case for an even number of equally spaced levels, as in the eight-level case in Fig. 8.10. When the number of levels is even, it is convenient to define  $N$  as half the number of levels. We first determine  $\langle x\hat{x} \rangle$ . Note that for each sample value,  $x$  and  $\hat{x}$  have the same sign, so  $x\hat{x}$  is always positive. Let  $\epsilon$  represent the spacing between adjacent quantization levels. The values of  $x$  that fall within the quantization level between  $m\epsilon$  and  $(m+1)\epsilon$  are assigned values  $\hat{x} = (m + \frac{1}{2})\epsilon$ , and their contribution to  $\langle x\hat{x} \rangle$  is

$$\frac{1}{\sqrt{2\pi}} \int_{m\epsilon}^{(m+1)\epsilon} \left(m + \frac{1}{2}\right) \epsilon x e^{-x^2/2} dx. \quad (8.74)$$

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

$$\begin{aligned} \langle x\hat{x} \rangle = & \sqrt{\frac{2}{\pi}} \left[ \left( \sum_{m=0}^{N-2} \int_{m\epsilon}^{(m+1)\epsilon} \left(m + \frac{1}{2}\right) \epsilon x e^{-x^2/2} dx \right) \right. \\ & \left. + \int_{(N-1)\epsilon}^{\infty} \left(N - \frac{1}{2}\right) \epsilon x e^{-x^2/2} dx \right]. \end{aligned} \quad (8.75)$$

The summation term contains one integral for each positive quantization level except the highest one. The integral on the lower line covers the highest level and the range of  $x$  above it, for both of which the assigned value is  $\hat{x} = (N - \frac{1}{2})\epsilon$ . Then, performing the integration, Eq. (8.75) reduces to

$$\langle x\hat{x} \rangle = \sqrt{\frac{2}{\pi}} \epsilon \left( \frac{1}{2} + \sum_{m=1}^{N-1} e^{-m^2 \epsilon^2 / 2} \right). \quad (8.76)$$

To evaluate the variance of  $\hat{x}$ , again consider first the contribution from values of  $x$  that fall between  $m\epsilon$  and  $(m+1)\epsilon$ . For this level, the quantized data  $\hat{x}$  all have the value  $(m + \frac{1}{2})\epsilon$ . The variance of  $\hat{x}$  for all values of  $x$  within this level is

$$\left(m + \frac{1}{2}\right)^2 \epsilon^2 \frac{1}{\sqrt{2\pi}} \int_{m\epsilon}^{(m+1)\epsilon} e^{-x^2/2} dx. \quad (8.77)$$

For negative  $x$ , we again include a factor of 2, sum over all positive quantization levels except the highest, and add a term for the highest level and the range of  $x$  above it. Thus, the total variance of  $\hat{x}$  is:

$$\begin{aligned} \langle \hat{x}^2 \rangle = & \sqrt{\frac{2}{\pi}} \left[ \left( \sum_{m=0}^{N-2} \left(m + \frac{1}{2}\right)^2 \epsilon^2 \int_{m\epsilon}^{(m+1)\epsilon} e^{-x^2/2} dx \right) \right. \\ & \left. + \left(N - \frac{1}{2}\right)^2 \epsilon^2 \int_{(N-1)\epsilon}^{\infty} e^{-x^2/2} dx \right]. \end{aligned} \quad (8.78)$$

The integrals in Eq.(8.78) can be represented by error functions. Then, using Eqs. (8.73), (8.76), and (8.78), we obtain

$$\eta_{(2N)} = \frac{\frac{2}{\pi} \left( \frac{1}{2} + \sum_{m=1}^{N-1} e^{-m^2 \epsilon^2/2} \right)^2}{\left(N - \frac{1}{2}\right)^2 - 2 \sum_{m=1}^{N-1} m \operatorname{erf}\left(\frac{m\epsilon}{\sqrt{2}}\right)}. \quad (8.79)$$

For the cases in which the number of levels is odd, the thresholds of the levels occur at values that are an integer plus  $\frac{1}{2}$ , as in the nine-level case in Fig. 8.10. We represent the odd level number by  $2N + 1$ . Consider the values of  $x$  that fall within the quantization level between  $(m - \frac{1}{2})\epsilon$  and  $(m + \frac{1}{2})\epsilon$ . These are assigned the value  $m\epsilon$ , i.e., zero for the level centered on  $x = 0$ . For this level, the contribution to  $\langle x\hat{x} \rangle$  is

$$\frac{1}{\sqrt{2\pi}} \int_{(m-\frac{1}{2})\epsilon}^{(m+\frac{1}{2})\epsilon} m\epsilon x e^{-x^2/2} dx. \quad (8.80)$$

Summing over all levels, as in Eq. (8.75), we obtain

$$\langle x\hat{x} \rangle = \sqrt{\frac{2}{\pi}} \left[ \left( \sum_{m=1}^{N-1} \int_{(m-\frac{1}{2})\epsilon}^{(m+\frac{1}{2})\epsilon} m\epsilon x e^{-x^2/2} dx \right) + \int_{(N-\frac{1}{2})\epsilon}^{\infty} N\epsilon x e^{-x^2/2} dx \right]. \quad (8.81)$$

Then, as in Eq. (8.78), we determine  $\langle \hat{x}^2 \rangle$ :

$$\langle \hat{x}^2 \rangle = \sqrt{\frac{2}{\pi}} \left[ \sum_{m=1}^{(N-1)} \left( \int_{(m-\frac{1}{2})\epsilon}^{(m+\frac{1}{2})\epsilon} (m\epsilon)^2 e^{-x^2/2} dx \right) + \int_{(N-\frac{1}{2})\epsilon}^{\infty} (N\epsilon)^2 e^{-x^2/2} dx \right]. \quad (8.82)$$

Performing the integration in Eqs. (8.81) and (8.82), from (8.73), we obtain

$$\eta_{(2N+1)} = \frac{\frac{2}{\pi} \left( \sum_{m=1}^N e^{-(m-\frac{1}{2})^2 \epsilon^2/2} \right)^2}{N^2 - 2 \sum_{m=1}^N \left( m - \frac{1}{2} \right) \operatorname{erf} \left( \frac{(m-\frac{1}{2})\epsilon}{\sqrt{2}} \right)}. \quad (8.83)$$

Equations (8.79) and (8.83) can easily be evaluated numerically and provide values of quantization efficiency for any number of equally spaced levels.

Since no significant approximations were made, the same method can be used for cases in which the number of quantization levels is small and consequently the quantization noise is relatively large. Values of  $\eta_Q$  for two, three, and four levels can be obtained by considering the effect of the quantization noise at a correlator input, following the method used above. In cases such as that in Appendix 8.3, for which the assigned values for the levels are chosen to optimize  $\eta_Q$ , or for which the spacing between the level thresholds is not uniform, the formulas derived here cannot be applied directly. However, the same general approach of considering the spacings between levels can be used. For three-level quantization, the levels for maximum quantization efficiency are  $\pm 0.612\sigma$  ( $\epsilon = 1.224$ ). Then we have

$$\langle x\hat{x} \rangle = \sqrt{\frac{2}{\pi}} \epsilon \int_{\epsilon/2}^{\infty} x e^{-x^2/2} dx, \quad (8.84)$$

$$\langle \hat{x}^2 \rangle = \sqrt{\frac{2}{\pi}} \epsilon^2 \int_{\epsilon/2}^{\infty} e^{-x^2/2} dx, \quad (8.85)$$

and

$$\eta_3 = \frac{\langle x\hat{x} \rangle^2}{\langle \hat{x}^2 \rangle} = \frac{\sqrt{\frac{2}{\pi}} e^{-0.612^2/2}}{1 - \operatorname{erf} \left( \frac{0.612}{\sqrt{2}} \right)}. \quad (8.86)$$

Examples of results derived using Eqs. (8.79), (8.83), and (8.86) are shown in Table 8.2. In each case, the value of  $\epsilon$  is chosen to maximize  $\eta_Q$ . The values of  $\eta_Q$  are given to five decimal places to show how they approach 1.0 as the number

**Table 8.2** Examples of quantization efficiency,  $\eta_Q$ , for sampling at the Nyquist rate

| Number of levels ( $Q$ ) | $N$ | $\epsilon$ | $\eta_Q$ |
|--------------------------|-----|------------|----------|
| 3                        | 1   | 1.224      | 0.80983  |
| 4                        | 2   | 0.995      | 0.88115  |
| 8                        | 4   | 0.586      | 0.96256  |
| 9                        | 4   | 0.534      | 0.96930  |
| 16                       | 8   | 0.335      | 0.98846  |
| 32                       | 16  | 0.188      | 0.99651  |
| 256                      | 128 | 0.0312     | 0.99991  |

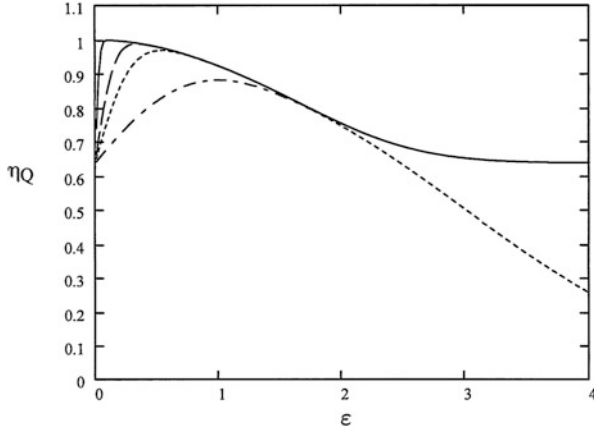
of levels increases. However, this is for the case of an ideal rectangular passband, which in a practical receiving system may be closely approximated. Figure 8.11 shows the quantization efficiency  $\eta_Q$  as a function of the threshold spacing  $\epsilon$ .

If the constant voltage spacing between adjacent thresholds for both input and output values is not maintained, the individual levels can sometimes be adjusted to obtain an improvement in  $\eta_Q$  of a few tenths of a percent, decreasing with increasing number of levels. The values of  $\eta_Q$  in Table 8.2 are in agreement with results by Jenet and Anderson (1998), who give detailed calculations of performance for two- to eight-bit quantization, for both uniform and nonuniform threshold spacing. See also Appendix 8.3 for optimization in the case of four-level quantization.

In recent designs of radio telescopes, the level increment  $\epsilon$  is frequently chosen so that signals at levels much higher than the rms system noise can be accommodated within the range of levels of the quantizer. This preserves an essentially linear response to interfering signals so that they can be eliminated or mitigated by further processing. For example, with 256 levels (8-bit representation) and  $\epsilon = 0.5$ , we find that  $\eta_Q = 0.9796$ . The range of  $\pm 128$  levels then corresponds to  $\pm 64\sigma$ , i.e.,  $\pm 36$  dB above the system noise, for a  $\sim 2\%$  sacrifice in signal-to-noise ratio.

### 8.3.6 Correlation Estimates for Strong Sources

The efficiency calculations of the previous sections are based on estimates of the correlation from the averaged signal products before or after quantization,  $\langle x_i y_i \rangle$  or  $\langle \hat{x}_i \hat{y}_i \rangle$ , in the limit of small correlation,  $|\rho| \ll 1$ . Johnson et al. (2013) show that when the correlation is small ( $|\rho| \ll 1$ ) and the signal variances are known (as is assumed when setting sampler thresholds), averaged products  $\langle \hat{x}_i \hat{y}_i \rangle$  do provide optimal estimates of the correlation. That is, when the correlation is small, no combination of the quantized signals will produce an unbiased estimate of correlation that has smaller variance than that of the correlator output, if suitable weights are chosen. This result arises from the form of the bivariate Gaussian distribution in this limit, which can be written such that the factor including the correlation coefficient  $\rho$  includes only terms of the form  $xy$  [see, e.g., Eq. (8.2)].



**Fig. 8.11** Quantization efficiency as a function of the threshold spacing,  $\epsilon$ , in units equal to the rms amplitude,  $\sigma$ . The curves are for 64-level (solid line), 16-level (long-dashed line), 9-level (short-dashed line), and 4-level (long-and-short-dashed line). As  $\epsilon$  becomes very small, the output of the quantizer depends mainly on the sign of the input, so the curves meet the ordinate axis at the two-level value of  $\eta_Q = 2/\pi$ . As  $\epsilon$  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. The working point in each case is chosen to be near the maximum of the curve.

However, when the correlation is large, alternative estimates of the correlation will have lower noise. Thus, in the high correlation regime, it is necessary to revise our expression for quantization efficiency. For instance, when the signals are not quantized, the optimal estimate of correlation for two zero-mean signals is Pearson's correlation coefficient [e.g., Wall and Jenkins (2012)],

$$r_p = \frac{\sum_{i=1}^{N_N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N_N} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{N_N} (y_i - \bar{y})^2}}, \quad (8.87)$$

where,  $\bar{x} = \frac{1}{N_N} \sum_{i=1}^{N_N} x_i$  is the sample mean, and the sums in the denominator are proportional to the sample variances. The standard error in the estimate of  $r_p$ , i.e.,  $\sigma_p$ , is

$$\sigma_p = N_N^{-1/2} (1 - \rho^2). \quad (8.88)$$

As  $\rho$  approaches unity,  $\sigma_p$  goes to zero. In this limit, the probability function  $p(x, y)$  given in Eq. (8.1) collapses to a one-dimensional Gaussian distribution along the line  $x = y$  (see Fig. 8.2). When  $\rho = 1$ , that line is perfectly defined by a

set of measurements of  $x_i$  and  $y_i$ , i.e., there are no deviations from the line, and the uncertainty in the estimate of  $\rho$  is zero. Perhaps, surprisingly, the estimate of correlation made without the sample means and variances, as in Eq. (8.8), i.e.,

$$r_\infty = \frac{1}{N_N} \sum_{i=1}^{N_N} x_i y_i, \quad (8.89)$$

has an error when normalized by  $\sigma^2$  of

$$\sigma_\infty = N_N^{-1/2} (1 + \rho^2)^{1/2} \quad (8.90)$$

[see Eq. (8.13)], which equals  $\sigma_p$  only for  $\rho = 0$ . For two-level quantization, for a rectangular passband and  $\beta = 1$ , the error on the correlation estimate is [see Eq. (8.30a) and Eq. (8.35)]

$$\sigma_2 = N_N^{-1/2} (1 - \rho^2)^{1/2}. \quad (8.91)$$

In the case of large correlation, the Van Vleck relation [Eq. (8.25)]

$$\rho = \sin \frac{\pi}{2} \rho_2, \quad (8.92)$$

will require a nonlinear scaling of the error in  $\rho_2$ , denoted  $\sigma_{2V}$ , which can be written

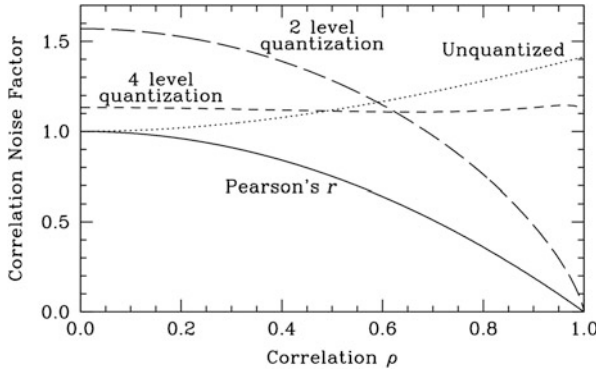
$$\sigma_{2V} = N_N^{-1/2} \left[ \left( \frac{\pi}{2} \right)^2 - (\sin^{-1} \rho)^2 \right] (1 - \rho^2)^{1/2}. \quad (8.93)$$

These errors in correlation for the various cases described above [ $\sigma_p$ ,  $\sigma_\infty$ , and  $\sigma_{2V}$  as well as the error for the case of four-level sampling, derived from formulas by Gwinn (2004)] are shown in Fig. 8.12. The interesting result is that the performance of the two-level correlator is better than that of the unquantized correlator for  $\rho > 0.6$  and approaches that of the Pearson estimator as  $\rho$  approaches unity. The peculiarity in the two-level scenario was noted by Cole (1968) and is related to the fact that the sample variance is irrelevant in the two-level quantization estimate.

Johnson et al. (2013) derive maximum likelihood estimators (MLEs) for the unsampled case in which the signal variance is known. Its standard deviation,  $\sigma_q$ , falls slightly below  $\sigma_p$ . The authors also derive MLEs for various quantization levels and show that their performance  $\sigma_q(Q)$  approaches  $\sigma_q$  for large values of  $Q$ .

## 8.4 Further Effects of Quantization

Various forms of analysis in radio astronomy involve cross-correlation of signals from different antennas or autocorrelation of a signal as a function of time. The values of the correlation of quantized signals deviate from the true correlation of



**Fig. 8.12** The correlation noise factor,  $N_N^{1/2}$  times the standard deviation, vs. correlation  $\rho$ . The correlation factors for  $\rho = 0$  are equal to  $n_Q^{-1}$ . The curves labeled unquantized and two-level quantization give correlation factors based on the standard signal-product estimate of  $\rho$  [see Eqs. (8.90) and (8.93) for unquantized and two-level quantization, respectively]. The factor for the Pearson's  $r$  curve, given in Eq. (8.88), is based on the estimator  $r_p$ , which involves the sample mean and variance. Adapted from Johnson et al. (2013).

the unquantized signals to an extent that is most serious for two-level sampling, and the deviation decreases as the number of levels is increased. Correction for this effect requires determination of how the cross-correlation of the quantized data, here designated  $R$ , is related to the true cross-correlation,  $\rho$ . To examine the effect of quantization, we consider the effect of a time offset  $\tau$  on two Gaussian waveforms that are otherwise identical. In the case of two-level sampling, the required relationship is given by the Van Vleck equation [Eq. (8.25)] and is

$$R_2(\tau) = \frac{2}{\pi} \sin^{-1} \rho(\tau) . \quad (8.94)$$

For more than two quantization levels, the relationship is more complicated, and although the nonlinearity of the quantized correlation becomes less serious with an increasing number of levels, correction may still be necessary. As very large instruments come into operation, it becomes increasingly important to remove the responses to strong radio sources in order to study the fainter emission from the most distant regions of the Universe. This requires very accurate calibration of the received signal strengths.

### 8.4.1 Correlation Coefficient for Quantized Data

Let  $x$  and  $y$  represent two Gaussianly distributed data streams that differ only by a time offset,  $\tau$ . The correlation coefficient,  $\rho(\tau)$ , is equal to  $\langle xy \rangle / \langle x^2 \rangle$ . The quantized values of  $x$  and  $y$  are identified by circumflex accents, i.e.,  $\hat{x}$  and  $\hat{y}$ . The correlation

coefficient of the quantized variables is

$$R(\tau) = \langle \hat{x}\hat{y} \rangle / \langle \hat{x}^2 \rangle. \quad (8.95)$$

To determine  $\rho(\tau)$  as a function of the correlation coefficient  $\rho$ , we need to consider the probabilities of occurrence of the unquantized variables  $x$  and  $y$  within each quantization interval. First, consider the case in which the number of quantization intervals is even and equal to  $2N$ . Thus, there are  $N$  positive intervals plus  $N$  negative ones. The mean value of the products of pairs of the quantized values,  $\langle \hat{x}\hat{y} \rangle$ , is obtained by considering each of the  $2N \times 2N = 4N^2$  possible pairings of the levels of  $\hat{x}$  and  $\hat{y}$ . Only half of these need be calculated, since if the  $x$  and  $y$  values are interchanged, the probability remains the same. The probability of the unquantized variables  $x$  and  $y$  falling within any pair of intervals is given by integration of the Gaussian bivariate probability distribution, Eq. (8.1), over the corresponding range of  $x$  and  $y$ . In Eq. (8.1),  $x$  and  $y$  have variance  $\sigma$  and cross-correlation coefficient  $\rho$ . Here, we are concerned with samples of  $x$  and  $y$  taken at the Nyquist interval  $\tau_s$ , and  $n$  is the number of Nyquist intervals between the pairs of samples considered. For a rectangular passband of width  $\Delta\nu$ , the correlation coefficient is given by

$$\rho(n\tau_s) = \frac{\sin(\pi n\tau_s)}{\pi n\tau_s}. \quad (8.96)$$

To calculate  $\langle \hat{x}\hat{y} \rangle$  for each combination of two quantization intervals, the joint probability of the required unquantized variables falling within these intervals is multiplied by the product of the corresponding values assigned in the quantization process. These results are then summed for all the pairs of intervals. Since the probability distributions of  $\hat{x}$  and  $\hat{y}$  are both symmetrical about zero, first consider the case in which both of these variables are positive and run from zero to  $N$ . As noted above, we take the step size to be unity. Let  $L(i)$  be the series of  $N+1$  values that define the positive quantization steps, i.e.,  $0, 1, 2, \dots, (N-1), \dots, \infty$ . Thus, for  $i = 1$  to  $N$ ,  $L(i) = i-1$ , and  $L(N+1) = \infty$ . For  $y$ , there is an identical series of levels represented as  $L(j)$ . Then the component of  $\langle \hat{x}\hat{y} \rangle$  that results from the positive ranges of  $x$  and  $y$  is

$$\sum_{i=1}^N (i-1/2) \left[ \sum_{j=1}^N (j-1/2) \int_{L(j)}^{L(j+1)} \int_{L(i)}^{L(i+1)} p(x, y) dx dy \right], \quad (8.97)$$

where  $(i-1/2)$  and  $(j-1/2)$  are the values of the digital data assigned to the corresponding quantization intervals, and  $p(x, y)$  is the Gaussian bivariate probability distribution, Eq. (8.1). The case in which both  $x$  and  $y$  are negative provides an equal component of  $\langle \hat{x}\hat{y} \rangle$ . Thus, the component of  $\langle \hat{x}\hat{y} \rangle$  for cases in



which  $x$  and  $y$  have the same sign is

$$\langle \hat{x}\hat{y} \rangle = \frac{1}{\pi\sigma^2\sqrt{1-\rho^2}} \sum_{i=1}^N (i-1/2) \left\{ \sum_{j=1}^N (j-1/2) \int_{L(j)}^{L(j+1)} \int_{L(i)}^{L(i+1)} \left[ \exp\left(\frac{-(x^2+y^2-2\rho xy)}{2\sigma^2(1-\rho^2)}\right) \right] dx dy \right\}. \quad (8.98)$$

For the cases in which  $x$  and  $y$  have opposite signs, one of either  $(i-1/2)$  or  $(j-1/2)$  is negative, and the sign of either  $x$  or  $y$  within the exponential function in Eq. (8.98) is negative. When the corresponding expression is included (with negative sign since the component of  $\langle \hat{x}\hat{y} \rangle$  is negative), we obtain

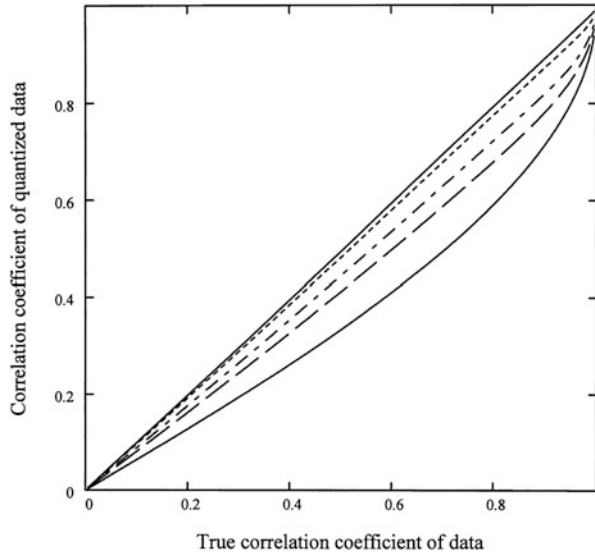
$$\langle \hat{x}\hat{y} \rangle = \frac{1}{\pi\sigma^2\sqrt{1-\rho^2}} \sum_{i=1}^N (i-1/2) \left\{ \sum_{j=1}^N (j-1/2) \int_{L(j)}^{L(j+1)} \int_{L(i)}^{L(i+1)} \left[ \exp\left(\frac{-(x^2+y^2-2\rho xy)}{2\sigma^2(1-\rho^2)}\right) - \exp\left(\frac{-(x^2+y^2+2\rho xy)}{2\sigma^2(1-\rho^2)}\right) \right] dx dy \right\}. \quad (8.99)$$

Equation (8.99) shows how  $\langle \hat{x}\hat{y} \rangle$  is derived using the usual form of the bivariate distribution in Eq. (8.1). An equivalent form of the probability distribution of  $x$  and  $y$  in Eq. (8.1) is given by Abramowitz and Stegun (1968, see Eqs. 26.2.1 and 26.3.2), which avoids the explicit use of the double integrals. Equation (8.99) can then be written as follows:

$$\begin{aligned} \langle \hat{x}\hat{y} \rangle = & \frac{1}{\sqrt{2\pi}\sigma} \left\{ \sum_{i=1}^N \left[ (i-1/2) \sum_{j=1}^N \left[ (j-1/2) \int_{L(i)}^{L(i+1)} \operatorname{erfc}\left(\frac{L(j)-\rho x}{\sigma\sqrt{2(1-\rho^2)}}\right) \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \right] \right. \right. \\ & - \sum_{i=1}^N \left[ (i-1/2) \sum_{j=1}^N \left[ (j-1/2) \int_{L(i)}^{L(i+1)} \operatorname{erfc}\left(\frac{(L(j+1)-\rho x)}{\sigma\sqrt{2(1-\rho^2)}}\right) \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \right] \right] \\ & - \sum_{i=1}^N \left[ (i-1/2) \sum_{j=1}^N \left[ (j-1/2) \int_{L(i)}^{L(i+1)} \operatorname{erfc}\left(\frac{L(j)+\rho x}{\sigma\sqrt{2(1-\rho^2)}}\right) \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \right] \right] \\ & \left. \left. + \sum_{i=1}^N \left[ (i-1/2) \sum_{j=1}^N \left[ (j-1/2) \int_{L(i)}^{L(i+1)} \operatorname{erfc}\left(\frac{(L(j+1)+\rho x)}{\sigma\sqrt{2(1-\rho^2)}}\right) \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \right] \right] \right\}, \end{aligned} \quad (8.100)$$

where  $\operatorname{erfc}$  is the complementary error function  $(1 - \operatorname{erf})$ .

**Fig. 8.13** Curves of the correlation coefficient of quantized data as a function of the true correlation (i.e., the correlation of the unquantized data). The lowest (solid) curve is for 2-level quantization, and moving upward, the curves are for 3 levels (long dashes), 4 levels (long and short dashes), 8 levels (small dashes) and 16 levels (solid line). Similar curves for three and four quantization levels are given in Fig. 8.7.

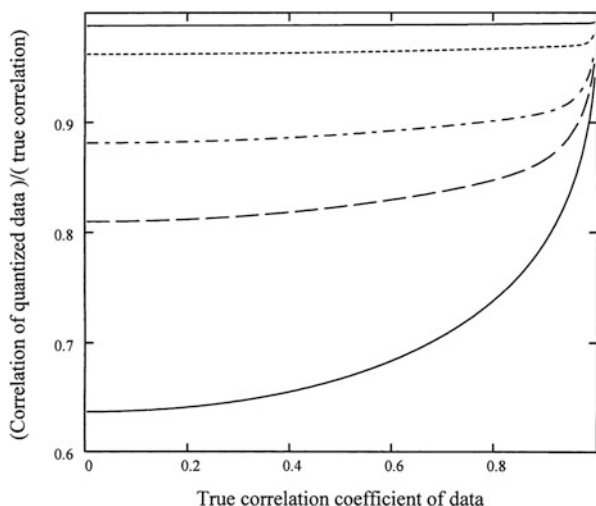


To calculate  $\hat{x}^2$ , the Gaussian probability function for a single variable is used, taking double the expression for the positive range of  $x$ :

$$\begin{aligned}\hat{x}^2 &= \frac{\sqrt{2}}{\sqrt{\pi}\sigma} \sum_{i=1}^N (i - 1/2)^2 \int_{L(i)}^{L(i+1)} \exp\left(\frac{-x^2}{2\sigma^2}\right) dx \\ &= \sum_{i=1}^N (i - 1/2)^2 \left[ \operatorname{erfc}\left(\frac{L(i)}{\sqrt{2}\sigma}\right) - \operatorname{erfc}\left(\frac{L(i+1)}{\sqrt{2}\sigma}\right) \right].\end{aligned}\quad (8.101)$$

Thus, for a given value of the time interval between samples, the correlation coefficient for the quantized data is as given in Eqs. (8.95), (8.99) or (8.100), and (8.101). Note that the ratio  $\langle xy \rangle / \langle x^2 \rangle$  is independent of the frequency response of the system considered and is based on a Gaussian distribution of the amplitude.

Figures 8.13 and 8.14 show examples of the relationship between the correlation of the quantized signals and the true signal correlation. Both of the figures result from the same analysis, but the presentation in Fig. 8.14, in which the correlation of the quantized data is shown as a fraction of the true correlation, helps to emphasize the nonlinearity in the response. A linear response would appear as a horizontal line in Fig. 8.14, and the curves approach this condition as the number of quantization



**Fig. 8.14** Curves of the correlation coefficient of quantized data, expressed as a fraction of the true correlation. These are from the same data as used in Fig. 8.13, but here they are plotted as fractions of the true (unquantized) correlation, in which the nonlinearity appears as a deviation from a horizontal line. The lowest curve is for 2-level quantization, and moving upward, the curves are for 3, 4, 8, and 16 levels. The points at which the curves meet the left vertical axis indicate the reduction in correlation resulting from quantization when the signal-to-noise ratio is low, as given in Table 8.2. The signal-to-noise ratio increases as the curves move from left to right, and the correlation coefficients of both the quantized and unquantized data move toward 1.0 for the theoretical case of complete correlation between the two signals.

levels increases. Except for observations of the strongest sources, the signal-to-noise ratio from an individual element of a synthesis array is small. Thus, the working point on the curves in Figs. 8.13 and 8.14 is generally near the left side, where the linearity for signals from cross-correlated pairs is best. As the number of quantization levels increases, the accuracy of the correlation increases. The curves provide an indication of the extent to which the quantization affects the measurement of cross-correlation of signals with Gaussian amplitude distribution. A detailed discussion of the effects of quantization of the signal amplitude is given by Benkevitch et al. (2016). This includes the case in which the cross-correlated signals have different amplitudes, and the effects of quantization as the cross-correlation of the analog waveforms approach unity.

For ease of computation, the correlation can be expressed as a rational function, or similar approximation, of the correlator output: See Appendix 8.3 for four-level quantization. For three-level quantization, procedures for determination of the cross-

correlation,  $\rho$ , from the correlator output are given by Kulkarni and Heiles (1980) and D'Addario et al. (1984). However, with the continuing increase in computer power, larger numbers of levels are generally used.

### 8.4.2 Oversampling

Sampling of signals at the Nyquist rate results in no loss of information, but quantization causes a reduction in sensitivity as represented by the quantization efficiency. Some of the loss due to quantization can be recovered by oversampling, that is, sampling faster than the Nyquist rate. For sampling of random noise with an ideal rectangular spectrum of width  $\Delta\nu$ , the time interval between adjacent Nyquist samples is  $1/(2\Delta\nu)$ . With Nyquist sampling, the noise within each sample is uncorrelated with respect to the noise in any other sample, and when such data are combined, the noise combines additively in power. Consider the case of oversampling in which the number of samples per second is  $\beta$  times the Nyquist rate. When the sample rate exceeds the Nyquist rate, the samples are no longer independent, and for any particular sample, there are components of the noise within other samples that are correlated with the noise in the sample considered. [Note, however, that for any two samples spaced by  $\beta$  times the sample interval (i.e., spaced at the Nyquist interval), or by an integral multiple of the Nyquist interval, the noise is uncorrelated.<sup>5</sup>] The correlated components of the noise in different samples combine additively in voltage, rather than additively in power, as is the case for uncorrelated noise.

To illustrate how the components of noise combine, consider one pair of antennas and, for example, just four consecutive samples at the correlator output. Let  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  be these voltages, which are proportional to the product of the voltages at the correlator inputs. Then we have for the squared sum of these correlated noise voltages, i.e., the total noise power,

$$[a_1 + a_2 + a_3 + a_4]^2 = a_1^2 + a_2^2 + a_3^2 + a_4^2 + 2(a_1a_2 + a_1a_3 + a_1a_4 + a_2a_3 + a_2a_4 + a_3a_4) . \quad (8.102)$$

The autocorrelation coefficient of the quantized signals at the correlator input is  $R(n\tau_s)$ , where  $n$  is an integer and  $\tau_s$  is the spacing in time between adjacent samples. The output of the correlator consists of values that are the product of two input samples, so the autocorrelation coefficient of the samples at the correlator output is

---

<sup>5</sup>It can be assumed that the noise components of the signals from any two antennas are uncorrelated, because noise from the sky background that is received in separate antennas is resolved, and generally the antennas are sufficiently far apart that cross talk of instrumental noise can be ignored.

$R^2(n\tau_s)$ . The mean noise power is given by the mean of the terms in the right side of Eq. (8.102), in which each of the  $a_n^2$  terms can be replaced by the mean squared noise amplitude  $\langle a^2 \rangle$ , and each of the  $a_m a_n$  terms by  $\langle a^2 \rangle R^2(|n - m|\tau_s)$ . Thus, the squared sum of the four noise voltages becomes

$$4\langle a^2 \rangle + 2\langle a^2 \rangle [3R^2(\tau_s) + 2R^2(2\tau_s) + R^2(3\tau_s)] . \quad (8.103)$$

If the four noise terms were uncorrelated, i.e., if the  $R^2$  terms were zero, the noise power would be the sum of the individual noise powers,  $4\langle a^2 \rangle$ . The effect of the correlation of the noise is to increase the averaged noise power by a factor equal to (8.103) divided by  $\langle 4a^2 \rangle$ :

$$1 + 2[(3/4)R^2(\tau_s) + (1/2)R^2(2\tau_s) + (1/4)R^2(3\tau_s)] . \quad (8.104)$$

In the general case, averaging a total of  $N$  samples at the correlator output, this factor becomes

$$1 + 2 \left[ \left( \frac{N-1}{N} \right) R^2(\tau_s) + \left( \frac{N-2}{N} \right) R^2(2\tau_s) + \left( \frac{N-3}{N} \right) R^2(3\tau_s) + \dots \right. \\ \left. + \left( \frac{1}{N} \right) R^2[(N-1)\tau_s] \right] . \quad (8.105)$$

In practice, in radio astronomy, the rate at which the data are sampled is in the range of MHz to GHz. The averaging times are in the range milliseconds to seconds, so  $N$  is likely to be within the range  $10^3$  to  $10^9$ . The autocorrelation coefficient decreases as the time interval between samples increases, and in practice,  $R^2(n\tau_s)$  becomes very small for  $n\tau_s \gtrsim 200$  times the Nyquist sample interval. Thus, for the terms within the square brackets in Eq. (8.105), those after about the first  $\sim 200\beta$  can be neglected. Since, in most cases,  $N \gg 200\beta$ , the squared sum of the noise voltages simplifies to

$$1 + 2[R^2(\tau_s) + R^2(2\tau_s) + R^2(3\tau_s) + \dots] = 1 + 2 \sum_{n=1}^{\infty} R^2(n\tau_s) . \quad (8.106)$$

Equation (8.106) is the fractional increase in the squared noise voltage (i.e., the noise power) that results from the fact that the noise in the samples is no longer independent when the data are oversampled. The quantization efficiency  $\eta_Q$  is equal to the quantization efficiency for Nyquist sampling,  $\eta_{QN}$ , multiplied by  $\sqrt{\beta}$  to take account of the increase in the number of samples, but divided by the square root of Eq. (8.106) because the noise in different samples is no longer independent. Thus,

**Table 8.3** Variation of quantization efficiency,  $\eta_Q$ , with oversampling factor  $\beta$ 

| No. of levels | $\epsilon$ | $\beta = 1$ | $\beta = 2$ | $\beta = 4$ | $\beta = 8$ | $\beta = 16$ | $\beta = 32$ |
|---------------|------------|-------------|-------------|-------------|-------------|--------------|--------------|
| 2             |            | 0.6366      | 0.744       | 0.784       | 0.795       | 0.798        | 0.799        |
| 3             | 1.224      | 0.8098      | 0.882       | 0.912       | 0.920       | 0.922        | 0.923        |
| 4             | 0.995      | 0.8812      | 0.930       | 0.951       | 0.958       | 0.960        | 0.960        |
| 8             | 0.586      | 0.9626      | 0.980       | 0.987       | 0.991       | 0.991        | 0.992        |
| 16            | 0.335      | 0.9885      | 0.994       | 0.996       | 0.998       | 0.998        | 0.998        |

noting that  $\tau_s = 1/(2\beta\Delta\nu)$ , we obtain

$$\eta_Q = \frac{\eta_{QN}\sqrt{\beta}}{\sqrt{1 + 2 \sum_{n=1}^{\infty} R^2 \left( \frac{n}{2\beta\Delta\nu} \right)}}. \quad (8.107)$$

To illustrate the effect of oversampling, examples of the quantization efficiency  $\eta_Q$ , derived using Eqs. (8.95), (8.100), (8.101), and (8.107), are shown in Table 8.3. These are for 2-, 3-, 4-, 8-, and 16-level sampling and values of  $\beta$  equal to 1, 2, 4, 8, 16, and 32. In each case, the value of  $\epsilon$  used is the one that maximizes  $\eta_Q$  for Nyquist sampling, as given in Thompson et al. (2007).<sup>6</sup> Note that as  $\beta$  is increased, the improvement gained by each further increase declines, because the correlation between adjacent samples increases, and thus, the new information provided by finer sampling becomes progressively smaller.

### 8.4.3 Quantization Levels and Data Processing

At this point, it is useful to put into perspective the characteristics of quantization schemes, which are summarized in Tables 8.2 and 8.3. It should be remembered that the assumption  $\rho \ll 1$  was used in determining these values. In considering the relative advantages of different quantization schemes, we note first that both the quantization efficiency  $\eta_Q$  and the receiving bandwidth  $\Delta\nu$  may be limited by the size and speed of the correlator system. The overall sensitivity is proportional to  $\eta_Q\sqrt{\Delta\nu}$ . Consider two conditions. In the first, the observing bandwidth is limited by factors other than the capacity of the digital system. This can occur in spectral line observing or when the interference-free band is of limited width. The sensitivity limitation imposed by the correlator system then involves only the quantization efficiency  $\eta_Q$  in Table 8.2, and the choice of quantization scheme is one between

<sup>6</sup>In this reference,  $\sigma$  is taken to be unity and  $\epsilon$ , the size of the quantization steps, is chosen to maximize the quantization efficiency.

**Table 8.4** Sensitivity factor  $\frac{\eta_Q}{\sqrt{\beta N_b}}$  for a correlator-limited system

| Number of<br>quantization levels ( $Q$ ) | $\beta = 1$ | $\beta = 2$ |
|--|-------------|-------------|
| 2 (1 bit)                                | 0.64        | 0.53        |
| 3 (2 bits)                               | 0.57        | 0.44        |
| 4 (2 bits)                               | 0.62        | 0.47        |
| 8 (3 bits)                               | 0.56        | 0.40        |
| 16 (4 bits)                              | 0.49        | 0.35        |

simplicity and sensitivity. In the second case, the observing bandwidth is set by the maximum bit rate that the digital system can handle, as may occur in continuum observation in the higher-frequency bands. For a fixed bit rate  $\nu_b$ , the sample rate is  $\nu_b/N_b$ , where  $N_b$  is the number of bits per sample, and the maximum signal bandwidth  $\Delta\nu$  is  $\nu_b/(2\beta N_b)$ , where  $\beta$  is the oversampling factor. Thus, the sensitivity is proportional to  $\eta_Q/\sqrt{\beta N_b}$ , and this factor is listed for various systems in Table 8.4, in which  $N_b = 1$  for  $Q = 2$  and  $N_b = 2$  for  $Q = 3$  or 4. Note that oversampling always reduces the performance under these conditions. For those situations in which the capacity of the correlator is limited by the maximum bit rate, the value of 0.64 for Nyquist sampling with two-level quantization results in the highest overall performance. Four-level sampling is almost as good, and four or more levels would be preferred if the bandwidth is limited, as in spectral line observations.

A three-level  $\times$  five-level correlator, for which the quantization efficiency  $\eta_Q$  is 0.86, was constructed by Bowers et al. (1973) for spectral line imaging with a two-element interferometer.

A further point to be noted is that with an analog correlator, the  $\sin \times \sin$  and  $\cos \times \cos$  products for signals from two antennas provide, in principle, exactly the same information. However, with a digital correlator, the quantization noise is largely uncorrelated between the sine and cosine components of the signal, so the quantization loss can be reduced by generating both products and averaging them.

## 8.5 Accuracy in Digital Sampling

Deviations from ideal performance in practical samplers result in errors that, if not corrected for, can limit the accuracy of images synthesized from the data. Once the signal is in digital form, however, the rate at which errors are introduced is usually negligibly small.

Two-level samplers, which sense only the sign of the signal voltages, are the simplest to construct. The most serious error that is likely to occur is in the definition of the zero level, in which a small voltage offset may occur. The effect of offsets in the samplers is to produce small offsets of positive or negative polarity

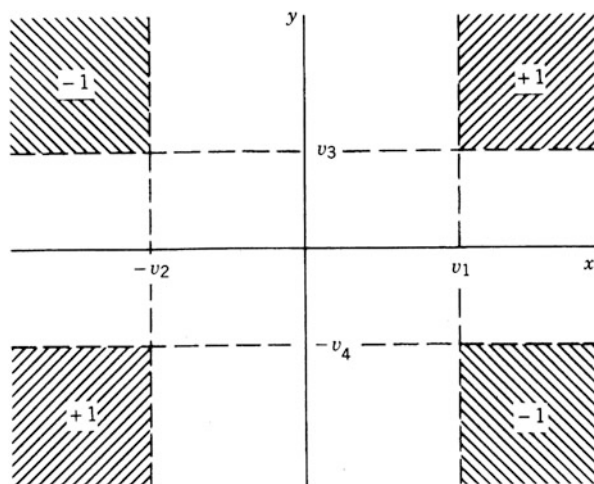
in the correlator outputs, which can be largely eliminated by phase switching, as described in Sect. 7.5. Alternately, the offsets in the samplers can be measured by incorporating counters to compare the numbers of positive and negative samples produced. Correction for the offsets can then be applied to the correlator output data [see, e.g., Davis (1974)].

In samplers with three or more quantization levels, the performance depends on the specification of the levels with respect to the rms signal level,  $\sigma$ . An automatic level control (ALC) circuit is therefore sometimes used at the sampler input. Errors resulting from incorrect signal amplitude become less important as the number of quantization levels is increased; with many levels, the signal amplitude becomes simply a linear factor in the correlator output. In systems using complex correlators, two samplers are usually required for each signal, one at each output of a quadrature network. The accuracy of the quadrature network and the relative timing of the two sample pulses are also important considerations.

### 8.5.1 Tolerances in Digital Sampling Levels

This section provides an example of the accuracy required in sampling. It is based on a study of errors in three-level sampling thresholds by D'Addario et al. (1984). We start by considering the diagram in Fig. 8.15, which shows the sampling thresholds for a pair of signals to be correlated. Thresholds  $v_1$  and  $-v_2$  apply to the signal waveform  $x(t)$  and  $v_3$  and  $-v_4$  to  $y(t)$ . The Gaussian probability distribution of  $x$  and  $y$  is given by Eq. (8.1), and the correlator output is proportional to this probability integrated over the  $(x, y)$  plane with the weighting factors  $\pm 1$  and zero indicated in the figure. This approach enables one to investigate the effect of deviations of the sampler thresholds from the optimum,  $v_0 = 0.612\sigma$ . For three-level sampling, the

**Fig. 8.15** Threshold diagram for a correlator, the inputs of which are three-level quantized signals.  $x$  and  $y$  represent the unquantized signals, and the shaded areas show the combinations of input levels for which the output is nonzero.





correlator output can be written

$$\langle r_3(\boldsymbol{\alpha}, \rho) \rangle = [L(\alpha_1, \alpha_3, \rho) + L(\alpha_2, \alpha_4, \rho) - L(\alpha_1, \alpha_4, -\rho) - L(\alpha_2, \alpha_3, -\rho)] , \quad (8.108)$$

where  $\alpha_i = v_i/\sigma$ , and

$$L(\alpha_i, \alpha_k, \rho) = \int_{\alpha_i}^{\infty} \int_{\alpha_k}^{\infty} \frac{1}{2\pi \sqrt{1-\rho^2}} \exp \left[ \frac{-(X^2 + Y^2 - 2\rho XY)}{2(1-\rho^2)} \right] dX dY . \quad (8.109)$$

Here,  $X = x/\sigma$ ,  $Y = y/\sigma$ , and the integrand in Eq. (8.109) is equivalent to the expression in Eq. (8.1) but with the variables measured in units of  $\sigma$ .

D'Addario et al. (1984) point out that since less than 5% loss in signal-to-noise ratio occurs for threshold departures of  $\pm 40\%$  from optimum, the required accuracy of the threshold settings, in practice, depends mainly on the algorithm used to correct the result. Suppose that the thresholds are kept close to, but not exactly equal to, the optimum value. For the  $x$  sampler in Fig. 8.15, the deviations from the ideal threshold value  $\alpha_0$  can be expressed in terms of an even part

$$\Delta_{gx} = \frac{1}{2}(\alpha_1 + \alpha_2) - \alpha_0 , \quad (8.110)$$

and an odd part

$$\Delta_{ox} = \frac{1}{2}(\alpha_1 - \alpha_2) . \quad (8.111)$$

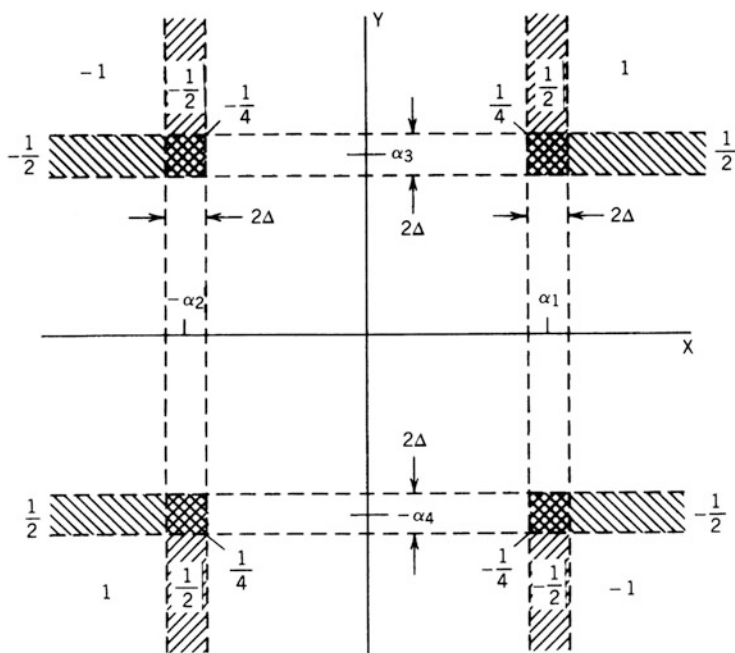
For the  $y$  sampler,  $\Delta_{gy}$  and  $\Delta_{oy}$  are similarly defined. The  $\Delta_g$  terms produce gain errors. They are equivalent to an error in the level of the signal at the sampler, and they have the effect of introducing a multiplicative error in the measured cross-correlation. The  $\Delta_o$  terms produce offset errors in the correlator output and are potentially more damaging since such errors can be large compared with the low levels of cross-correlation resulting from weak sources. The offset errors, however, can be removed with high precision by phase switching. The cancellation of the offset results from the sign reversal of the digital samples, or of the correlator output, as described in Sect. 7.5. The correlator output of a phase-switched system is of the form

$$r_{3s}(\boldsymbol{\alpha}, \rho) = \frac{1}{2} [r_3(\boldsymbol{\alpha}, \rho) - r_3(\boldsymbol{\alpha}, -\rho)] . \quad (8.112)$$

If all  $\alpha$  values are within  $\pm 10\%$  of  $\alpha_0$ , the output is always within  $10^{-3}$  (relative error) of the output of a correlator with the same gain errors, but no offset errors, in

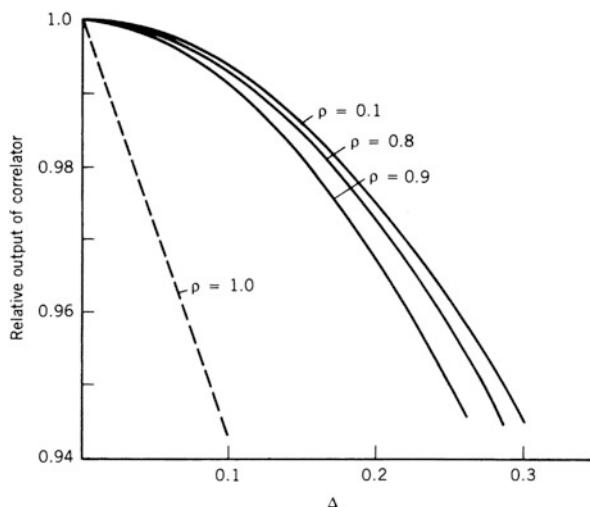
the samplers. Thus, with phase switching, errors of up to  $\sim 10\%$  in the thresholds may be tolerable. Also, corrections can be made for gain errors if the actual threshold levels are known. Since the probability density distribution of the signal amplitudes can be assumed to be Gaussian, the threshold levels can be determined by counting the relative numbers of  $+1$ ,  $0$ , and  $-1$  outputs from each sampler. When  $\rho$  is small (a few percent), a simple correction for the gain error can be obtained by dividing the correlator output by the arithmetic mean of the numbers of high-level ( $\pm 1$ ) samples for the two signals. Then  $10\%$  errors in the threshold settings result in errors of less than  $1\%$  in  $\rho$ .

Another nonideal aspect of the behavior of the sampler and quantizer is that the threshold level may not be precisely defined but may be influenced by effects such as the direction and rate of change of the signal voltage, the previous sample value (hysteresis), and noise in the sampling circuitry. The result can be modeled by including an indecision region in the sampler response extending from  $\alpha_k - \Delta$  to  $\alpha_k + \Delta$ . It is assumed that a signal that falls within this region results in an output that takes either of the two values associated with the threshold randomly and with equal probability. The three-level threshold diagram with indecision regions included is shown in Fig. 8.16.



**Fig. 8.16** Threshold diagram for a three-level correlator showing indecision regions and the shaded areas within them for which the response is nonzero. The figures  $\pm 1$ ,  $\pm \frac{1}{2}$ , and  $\pm \frac{1}{4}$  indicate the correlator response. The diagram shows the  $(X, Y)$  plane in which the signals are normalized to the rms value  $\sigma$ .

**Fig. 8.17** Effect of indecision regions on the output of a three-level correlator. The thresholds are assumed to be set to the optimum value  $0.612\sigma$ , and the widths of the indecision regions are  $2\sigma\Delta$ . The output is given as a fraction of the output for  $\Delta = 0$ .



The weighting in the indecision regions depends on the probability of the random sample values and is  $1/4$  when both signals fall within indecision regions, and  $1/2$  when one signal is within an indecision region and the other produces a nonzero output. As before, the correlator output can be obtained by integrating the weighted probability of the signal values over the  $(X, Y)$  plane. Figure 8.17 shows the decrease in the correlator output as a function of  $\Delta$  for several values of  $\rho$ , computed by expressing the output decrease as a Maclaurin series in  $\Delta$  (D'Addario et al. 1984). For all cases except those in which  $\rho$  approaches unity, the relatively small decrease in output results from the fact that when one input waveform falls within an indecision region, the other generally does not. For the particular case of  $\rho = 1$ , the input waveforms are identical and fall within these regions simultaneously. The output decrease is then proportional to  $\Delta$ , as shown by the broken line in Fig. 8.17: However, this case is only of limited practical importance. For a 1% maximum error,  $\Delta$  must not exceed  $0.11\sigma$ , so the indecision region can be as large as  $\pm 18\%$  of the threshold value. For a maximum error of 0.1%, the above limits must be divided by  $\sqrt{10}$ . Thus, the indecision regions have large enough tolerances that their effect may be negligible.

## 8.6 Digital Delay Circuits

Time delays that are multiples of the sample interval can be applied to streams of digital bits by passing them through shift registers that are clocked at the sampling frequency. Shift registers with different numbers of stages thus provide different fixed delays. A method of using two shift registers to obtain a delay that is variable in increments of the clock pulse interval is described by Napier et al. (1983). However,

integrated circuits for random access memory (RAM), developed for computer applications, provide an economical solution for large digital delays.

Another useful technique is serial-to-parallel conversion, that is, the division of a bit stream at frequency  $\nu$  into  $n$  parallel streams at frequency  $\nu/n$ , where  $n$  is a power-of-two integer. This allows the use of slower and more economical types of digital circuits for delay, correlation, and other processes.

The precision required in setting a delay has been discussed in Sect. 7.3.5 and is usually some fraction of the reciprocal of the signal bandwidth. In any form of delay that operates at the frequency of the sampler clock, the basic delay increment is the reciprocal of the sampling frequency. A finer delay step can be obtained digitally by varying the timing of the sample pulse in a number of steps, for example, 16, between the basic timing pulses. Thus, if an extra delay of, say,  $5/16$  of a clock interval is required, the sampler is activated  $11/16$  of a clock interval after the previous clock pulse, and the data are held for  $5/16$  of an interval to bring them into phase with the clock-pulse timing. Correction for delay steps equal to the sampling interval can also be made after the signals have been cross-correlated, by applying a phase correction to the cross power spectrum.

## 8.7 Quadrature Phase Shift of a Digital Signal

We have mentioned that complex correlators for digital signals can be implemented by introducing the quadrature phase shift in the analog signal, as in Fig. 6.3, and then using separate samplers for the signal and its phase-shifted version. The Hilbert transformation that the phase shift represents can also be performed on the digital signal, thus eliminating the quadrature network and saving samplers and delay lines, but the accuracy is limited. Hilbert transformation is mathematically equivalent to convolution with the function  $(-\pi\tau)^{-1}$ , which extends to infinity in both directions [see, e.g., Bracewell (2000), p. 364]. A truncated sequence of the same form, for example,  $\frac{1}{3}, 0, 1, 0, -1, 0, -\frac{1}{3}$ , provides a convolving function for the digital data that introduces the required phase shift. However, the truncation results in convolution of the resulting signal spectrum with the Fourier transform of the truncation function, that is, a sinc function. This introduces ripples and degrades the signal-to-noise ratio by a few percent. Also, the summation process in the digital convolution increases the number of bits in the data samples, but the low-order bits can be discarded to avoid a major increase in the complexity of the correlator. This results in a further quantization loss. The overall result is that the imaginary output of the correlator suffers spectral distortion and some loss in signal-to-noise ratio relative to the real output. These effects are most serious in broad-bandwidth systems, in which the high data rate permits only simple processing. Lo et al. (1984) have described a system in which the real part of the correlation is measured as a function of time offset, as described below for the spectral correlator, and the imaginary part is then computed by Hilbert transformation.

## 8.8 Digital Correlators

### 8.8.1 *Correlators for Continuum Observations*

In continuum observations, the average correlation over the signal bandwidth is measured, and data on a finer frequency scale may not be required. In such cases, the correlation of the signals is measured usually for zero time-delay offset. Digital correlators can be designed to run at the sampling frequency of the signals or at a submultiple resulting from dividing the bit stream from the sampler into a number of parallel streams. In the latter case, the number of correlator units must be proportionally increased, and their outputs can subsequently be additively combined. Two-level and three-level correlators, for which the products are represented by values of  $-1$ ,  $0$ , and  $+1$ , are the simplest. Correlators in which one of the inputs is a two-level or three-level signal and the other input is more highly quantized also have a degree of simplicity. In this case, the correlator is essentially an accumulating register into which the higher-quantization value is entered. The two-level or three-level value is used to specify whether the other number is to be added, subtracted, or ignored. In correlators in which both inputs have more than three levels of quantization, the multiplier output for any single product can be one of a range of numbers. One method of implementing such a multiplier is to use a read-only memory unit as a lookup table in which the possible product values are stored. The input bits to be multiplied are used to specify the address of the required product in the memory.

The output of a multiplier can take both positive and negative values, and, ideally, an up-down counter is required as an integrator. Since such counters are usually slower than simple adding counters, two of the latter are sometimes used to accumulate the positive and negative counts independently. Another technique is to count, for example,  $-1$ ,  $0$ , and  $+1$  as  $0$ ,  $1$ , and  $2$ , and then subtract the excess values, in this case equal to the number of products, in the subsequent processing.

Spectral line (multichannel) correlators are used with most large general-purpose arrays. For continuum observations, they offer advantages such as the ability to reject narrowband interfering signals or to divide a band into narrower sub-bands to reduce the smearing of spectral details.

### 8.8.2 *Digital Spectral Line Measurements*

In spectral line observations, measurements at different frequencies across the signal band are required. These measurements can be obtained by digital techniques using a spectral correlator system, which is commonly implemented by measuring the correlation of the signals as a function of time offset. The Fourier transform of this quantity is the cross power spectrum, which can be regarded as the complex visibility as a function of frequency. (This Fourier transform relationship is a form of

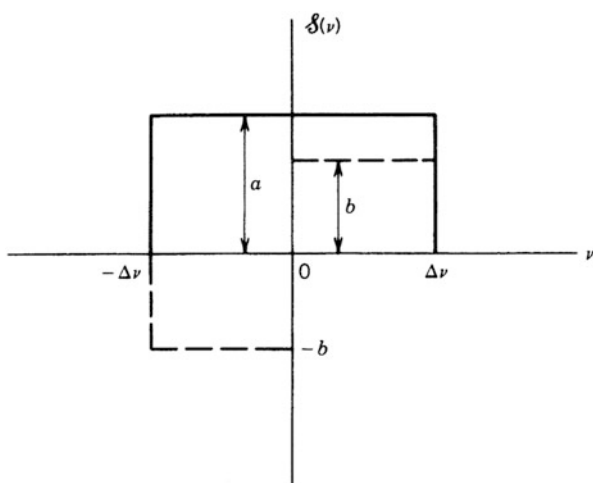
the Wiener–Khinchin relation discussed in Sect. 3.2.) In the case of an autocorrelator (for use with a single antenna), the two input signals are the same waveform with a time offset. Thus, the autocorrelation function is symmetric, and the power spectrum is entirely real and even. However, the cross power spectrum of the signals from two different antennas is complex, and the cross-correlation function has odd as well as even parts.

The output of a spectral correlator system provides values of the visibility at  $N$  frequency intervals across the signal band. These intervals are sometimes spoken of as frequency channels and their spacing as the channel bandwidth. To explain the action of a digital spectral correlator, we consider the cross power spectrum  $S(\nu)$  of the signals from two antennas, as shown in idealized form in Fig. 8.18. Here it is assumed that the source under observation has a flat spectrum with no line features, and the final IF amplifier before the sampler has a rectangular baseband response. In Fig. 8.18, we have included the negative frequencies since they are necessary in the Fourier transform relationships. For  $-\Delta\nu \leq \nu \leq \Delta\nu$ , the real and imaginary parts of  $S(\nu)$  have magnitudes  $a$  and  $b$ , respectively, and the corresponding visibility phase is  $\tan^{-1}(b/a)$ . The cross-correlation function  $\rho(\tau)$  is the Fourier transform of  $S(\nu)$ , where  $\tau$  is the time offset:

$$\begin{aligned}\rho(\tau) &= (a - jb) \int_{-\Delta\nu}^0 e^{j2\pi\nu\tau} d\nu + (a + jb) \int_0^{\Delta\nu} e^{j2\pi\nu\tau} d\nu \\ &= 2\Delta\nu \left[ a \frac{\sin(2\pi\Delta\nu\tau)}{2\pi\Delta\nu\tau} - b \frac{1 - \cos(2\pi\Delta\nu\tau)}{2\pi\Delta\nu\tau} \right].\end{aligned}\quad (8.113)$$

Thus,  $\rho(\tau)$  has an even component of the form  $(\sin x)/x$ , which is related to the real part of  $S(\nu)$ , and an odd component of the form  $(1 - \cos x)/x$ , which is related to the imaginary part. The spectral correlator measures  $\rho(\tau)$  for integral values

**Fig. 8.18** Cross power spectrum  $S(\nu)$  of two signals for which the power spectra are rectangular bands extending in frequency from zero to  $\Delta\nu$ . Negative frequencies are included. The solid line represents the real part of  $S(\nu)$  and the dashed line the imaginary part. The corresponding correlation function is derived in Eq. (8.113).



of the sampling interval  $\tau_s$ . We consider the case of Nyquist sampling, for which  $\tau_s = 1/(2\Delta\nu)$ . The measured cross-correlation refers to the quantized waveforms, and the analysis in Sect. 8.4.1 shows how this is related to the cross-correlation of the unquantized waveforms. For correlation levels that are not too large, the two quantities are closely proportional, so for simplicity, we assume that Eq. (8.113) represents the behavior of the measured cross-correlation. The measurements are made with  $2N$  time offsets from  $-N\tau_s$  to  $(N-1)\tau_s$  between the signals, and Fourier transformation of these discrete values yields the cross power spectrum at frequency intervals of  $(2N\tau_s)^{-1} = \Delta\nu/N$  for Nyquist sampling. The  $N$  complex values of the positive frequency spectrum are the data required. Of these, the imaginary part comes from the odd component of the correlator output  $r(\tau)$ . Thus, in the correlation measurement, it suffices to use single-multiplier correlators to measure  $2N$  real values of  $r(\tau)$  over both positive and negative values of  $\tau$  for one antenna with respect to the other. As an alternative to measuring only the real part of the correlation, complex correlators could be used to measure both the real and imaginary parts for a range of time offsets from zero to  $(N-1)\tau_s$ . However, complex correlators require broadband quadrature networks.

Measurement of the cross-correlation over the limited time offset range is equivalent to measuring  $r(\tau)$  multiplied by a rectangular function of width  $2N\tau_s$ . The cross power spectrum derived from the limited measurements is therefore equal to the true cross power spectrum convolved with the Fourier transform of the rectangular function, that is, with the sinc function

$$\frac{\sin(\pi\nu N/\Delta\nu)}{\pi\nu}, \quad (8.114)$$

which is normalized to unit area with respect to  $\nu$ . Any line feature within the spectrum is broadened by the sinc function (8.114) and, depending on its frequency profile, may show the characteristic oscillating skirts. The width of the sinc function at the half-maximum level is  $1.2\Delta\nu/N$ , that is, 1.2 times the channel separation, and this width defines the effective frequency resolution.

The oscillations of the sinc function introduce structure in the frequency spectrum similar to the sidelobe responses of an antenna beam. They result from the sharp edges of the rectangular function that multiplies the correlation function. Such sidelobes are undesirable and can be reduced by choosing weighting functions, other than rectangular truncation, that are constrained to be zero outside the measurement range. It is desirable that weighting functions should taper smoothly to zero at  $|\tau| = N\tau_s$ , thereby reducing unwanted ripples in the smoothing (convolving) function, but also to be as wide as possible in order to keep the width of the smoothing function as narrow as possible. These requirements are not generally compatible, so weighting functions that produce smoothing functions with very low sidelobes have poor frequency resolution. Some commonly used weighting functions are listed in Table 8.5.

Hann weighting, also known as raised cosine weighting, reduces the first sidelobe by a factor of 9 but degrades the resolution by 1.67, compared with uniform

**Table 8.5** Commonly used weighting functions

| Weighting<br>function $w(\tau)$ | $[w(\tau) = 0,  \tau  > \tau_1 = N\tau_s]$   | Half-amplitude<br>width<br>(Unit= $\Delta\nu/N$ ) | Peak<br>sidelobe |
|---------------------------------|--|---|------------------|
| Uniform                         | $w(\tau) = 1$  | 1.21  | 0.22             |
| Bartlett                        | $w(\tau) = 1 - ( \tau /\tau_1)$  | 1.77  | 0.047            |
| Hann <sup>a</sup>               | $w(\tau) = 0.5 + 0.5 \cos(\pi\tau/\tau_1)$   | 2.00  | 0.027            |
| Hamming                         | $w(\tau) = 0.54 + 0.46 \cos(\pi\tau/\tau_1)$   | 1.82  | 0.0073           |
| Blackman                        | $w(\tau) = 0.42 + 0.50 \cos(\pi\tau/\tau_1)$<br>$+ 0.08 \cos(2\pi\tau/\tau_1)$   | 2.30  | 0.0011           |
| Blackman-Harris                 | $w(\tau) = 0.35875 + 0.48829 \cos(\pi\tau/\tau_1)$<br>$+ 0.14128 \cos(2\pi\tau/\tau_1)$<br>$+ 0.0106411 \cos(3\pi\tau/\tau_1)$ | 2.67  | 0.000025         |

<sup>a</sup>Hann weighting is named after the nineteenth-century meteorologist Julius von Hann and is sometimes colloquially referred to as “Hanning weighting.” Hamming weighting is named after R. W. Hamming, an engineer at Bell Telephone Laboratories.

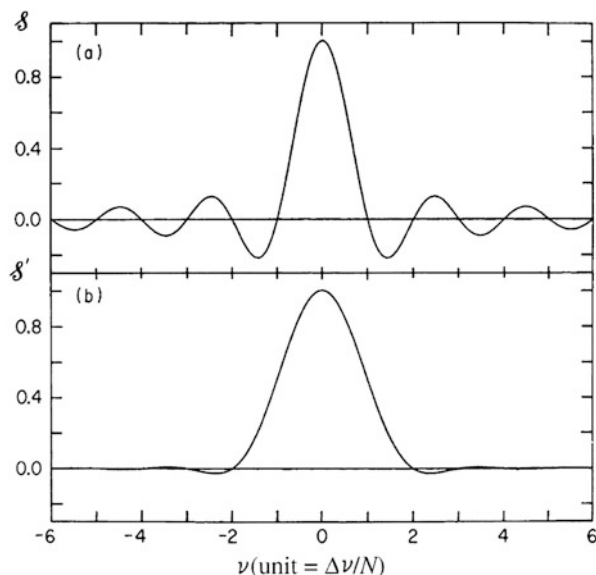
weighting. The Fourier transform of the Hann weighting function is the sum of three sinc functions of relative amplitudes 0.25, 0.5, and 0.25. This is the smoothing function in the spectral domain shown in Fig. 8.19b, which corresponds to Hann weighting. For the usual case in which the number of points in the discretely sampled spectrum equals the number of points in the correlation function (i.e., no zero padding, as in the FX correlator, Sect. 8.8.4), the smoothing or convolution can be implemented as a three-point running mean with relative weights of 0.25, 0.5, and 0.25. Thus, the smoothed value of the cross power spectrum at frequency channel  $n$  is given by

$$S' \left( \frac{n\Delta\nu}{N} \right) = \frac{1}{4} S \left[ \frac{(n-1)\Delta\nu}{N} \right] + \frac{1}{2} S \left( \frac{n\Delta\nu}{N} \right) + \frac{1}{4} S \left[ \frac{(n+1)\Delta\nu}{N} \right]. \tag{8.115}$$

The Hamming weighting function is very similar to the Hann function and would appear to be superior because it produces a better resolution and a lower peak sidelobe level. However, the sidelobes of the Hamming smoothing function do not decrease in amplitude as rapidly as those of the Hann smoothing function. Weighting functions are discussed in detail by Blackman and Tukey (1959) and Harris (1978).

A further effect of the finite time-offset range complicates the calibration of the instrumental frequency response in the following way (Willis and Bregman 1981). The frequency responses of the amplifiers associated with the different antennas may not be exactly identical, as discussed in Sect. 7.3. To calibrate the response of each antenna pair over the spectral channels, it is usual to measure the cross power spectrum of an unresolved source for which the actual radiated spectrum is known to be flat across the receiving passband. We can consider the result in terms of the idealized power spectra in Fig. 8.18. If no special weighting function is used, the real and imaginary parts are both convolved with the sinc function (8.114).

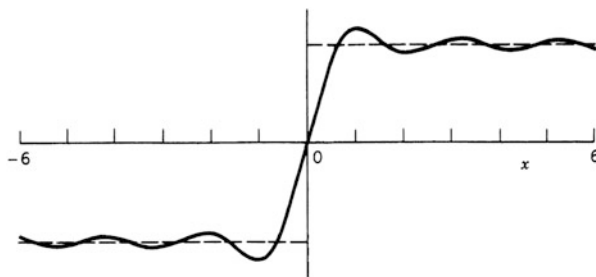




**Fig. 8.19** (a) The ordinate is the sinc function  $\sin(\pi \nu N / \Delta \nu) / (\pi \nu N / \Delta \nu)$ , which represents the frequency response of a spectral correlator with channels of width  $\Delta \nu / N$  to a narrow line at  $\nu = 0$ . The abscissa is frequency  $\nu$  measured with respect to the center of the received signal band. (b) The same curve after the application of Hann smoothing, as in Eq. (8.115).

When a function with a sharp edge is convolved with a sinc function, the result is the appearance of oscillations (the Gibbs phenomenon) near the edge, as shown in Fig. 8.20. The point here is that the real component of  $S(\nu)$  in Fig. 8.18 is continuous through zero frequency, but the imaginary part shows a sharp sign reversal. Thus, near zero frequency, the observed imaginary part of  $S(\nu)$  will show oscillations that may be as high as 18% in peak amplitude, whereas the real component will show relatively small oscillations at that point (see also Fig. 10.14b and associated text). As a result, the magnitude and phase measured for  $S(\nu)$  will show oscillations or ripples, the amplitude of which will depend on the relative amplitudes of the real and imaginary parts, that is, on the phase of the uncalibrated visibility. The uncalibrated phase measured for any source depends on instrumental factors such as the lengths of cables as well as the source position, which may not be known. In general, the phase will not be the same for the source under investigation and the calibrator. Hence, near zero frequency, some precautions must be taken in applying the calibration. Possible solutions to the problem include (1) calibrating the real and imaginary parts separately, (2) observing over a wide enough band that the end channels in which the ripples are strongest can be discarded, or (3) applying smoothing in frequency to reduce the ripples.

Another problem encountered when observing a spectral line in the presence of a continuum background is caused by reflections in the antenna structure. These reflections cause a sinusoidal gain variation across the passband, the period of which



**Fig. 8.20** Convolution of a step function at the origin (broken line) with the sinc function  $(\sin \pi x)/\pi x$ . Here,  $x = vN/\Delta v$ , and the half-cycle period of the ripple is approximately equal to the width of a spectral channel.

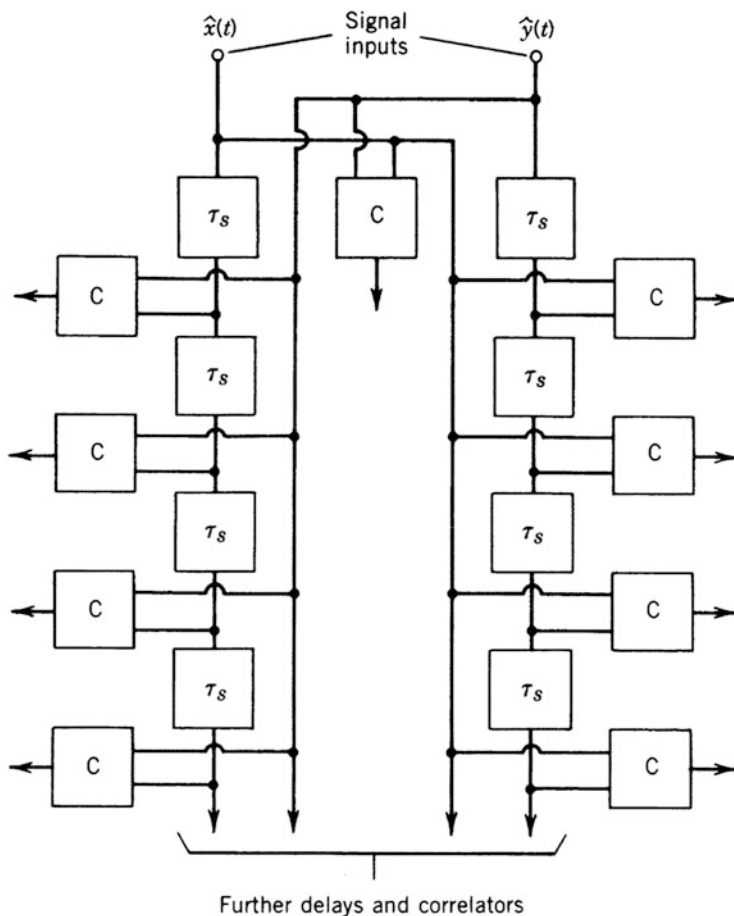
is equal to the reciprocal of the delay of the signal caused by the reflection. In a correlation interferometer, the magnitude of the ripple is a nearly constant fraction of the correlated continuum flux density, and the ripple is removed when the spectrum of the source under investigation is divided by the spectrum of the calibration source.

### 8.8.3 Lag (XF) Correlator

Correlators can be classified as two general types. In a lag (or XF) correlator, cross-correlation is followed by Fourier transformation, and in an FX correlator, Fourier transformation is followed by cross-correlation. A simplified schematic diagram of a lag correlator is shown in Fig. 8.21. Practical systems are often more complicated and are designed to take full advantage of the flexibility of digital processing techniques. The bandwidths of channels required for spectral line studies vary greatly, from a few tens of hertz to hundreds of megahertz. This versatility is necessary because the widths of spectral features are influenced by Doppler shifts, which are proportional to the rest frequencies of the lines and the velocities of the emitting atoms and molecules. The correlator of the upgraded VLA system (Perley et al. 2009) is fundamentally an XF design, as is the ALMA system, following its digital filter (Escoffier et al. 2000).

A recirculating correlator is one that can store blocks of data and process them multiple times through the correlator. This can be done only when the correlator is capable of running faster than the incoming data rate. These multiple passes allow the number of correlator channels to be increased. For example, if data samples are processed by the correlator twice, the range of delays can be doubled, so the spectral resolution is improved by a factor of two.

To implement the above scheme, recirculator units are required, which are basically memories that store blocks of input samples and allow them to be read out



**Fig. 8.21** Simplified schematic diagram of a lag (XF) spectral correlator for two sampled signals.  $\tau_s$  indicates a time delay equal to the sampling interval and  $C$  indicates a correlator. The correlation is measured for zero delay, for the  $\hat{x}$  input delayed with respect to the  $\hat{y}$  input (left correlator bank), and for  $\hat{y}$  delayed with respect to  $\hat{x}$  (right correlator bank). The delays are integral multiples of  $\tau_s$ .

at the correlator input rate. These memory units are required in pairs, so that one is filled with data at the Nyquist rate appropriate to the chosen signal bandwidth, while the other is being read at the maximum data rate. One memory becomes filled in the time that the other is read for the required number of times, and the two are then interchanged. Examples of recirculating lag correlators are described by Ball (1973) and Okumura et al. (2000). The WIDAR correlator on the VLA uses recirculation (Perley et al. 2009).

### 8.8.4 *FX Correlator*

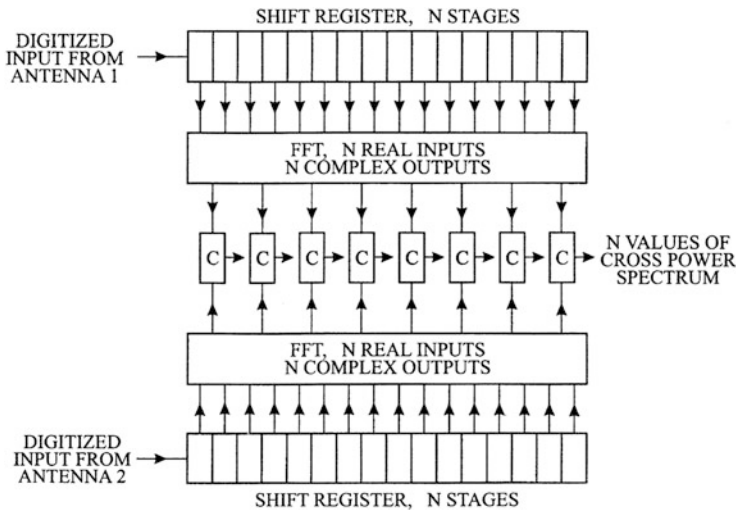
The designation FX indicates a correlator in which Fourier transformation to the frequency domain is performed before cross multiplication of data from different antennas. In such a correlator, the input bit stream from each antenna is converted to a frequency spectrum by a real-time FFT, and then for each antenna pair, the complex amplitudes for each frequency are multiplied to produce the cross power spectrum. A major part of the computation occurs in the Fourier transformation, for which the total number of operations is proportional to the number of antennas. In comparison, in a lag correlator, the total computation is largely proportional to the number of antenna pairs. Thus, the FX scheme offers economy in hardware, especially if the number of antennas is large (see Sect. 8.8.5). The principle of the FX correlator, based on the use of the FFT algorithm, was discussed by Yen (1974) and first used in a large practical system by Chikada et al. (1984, 1987). Description of system designed for the VLBA are given by Benson (1995) and Romney (1999).

Two slightly different implementations of the FX correlator have been used. In one, both in-phase and quadrature components of the signal are sampled to provide a sequence of  $N$  complex samples, which is then Fourier-transformed to provide  $N$  values of complex amplitude, distributed in frequency over positive and negative frequencies. In the other,  $N$  real samples are transformed to provide  $N$  values of complex amplitude. However, the negative frequencies are redundant, and only  $N/2$  spectral points need be retained. We follow the second scheme in the discussion below.

Figure 8.22 is a schematic diagram of the basic operations of an FX correlator. The input sample stream from an antenna is Fourier transformed in contiguous sequences of length- $N$  samples, where  $N$  is usually a power-of-two integer for efficiency in the FFT algorithm. The output of each transformation is a series of  $N$  complex signal amplitudes as a function of frequency. The frequency spacing of the data after transformation is  $1/(Nt_s)$ , where  $t_s$  is the time interval between samples of the signals. In the cross-multiplication process that follows the FFT stage, the complex amplitude from one antenna of each pair is multiplied by the complex conjugate of the amplitude of the other. These multiplications occur in the correlator elements in Fig. 8.22. Note that the data in any one input sequence are combined only with data from other antennas for the same time sequence. This leads to some differences in the effective weighting of the data in the FX and XF designs.

### 8.8.5 *Comparison of XF and FX Correlators*

*Spectral Response.* In the FX configuration, the F engine (DFT processor) operates on short segmented blocks of data in order to control the spectral resolution. The equivalent correlation function constructed from a block of data has  $N$  ways, or  $N$  possible multiplications for the zero lag component. There are progressively fewer

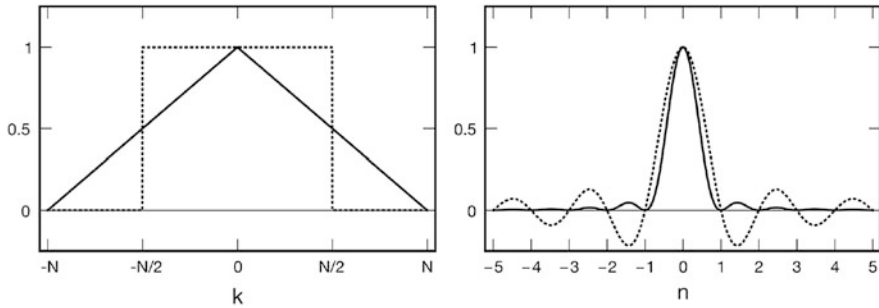


**Fig. 8.22** Simplified schematic diagram of an FX correlator for two antennas. The digitized signals are read into the shift registers and an FFT performed at intervals of  $N$  sample periods. The correlator elements, indicated by C, form products of one signal with the complex conjugate of the other. In an array with  $n_a$  antennas, the outputs of each FFT are split  $(n_a - 1)$  ways for combination with the complex amplitudes from all other antennas.

multiplications available for increasing lags because of the data block boundaries. The correlator function at the maximum lags of  $\pm(N - 1)t_s$  can be obtained in only one way. Hence, the density of lag multiplications has a triangular shape as a function of lag over the range  $\pm Nt_s$ , as shown in Fig. 8.23 (see also Moran 1976). Hence, the spectral response, the Fourier transform of this triangular function, is  $\text{sinc}^2(Nt_s\nu) = \text{sinc}^2(n)$ , where  $\nu = n/Nt_s$  and  $n$  is the spectral channel number. An alternate derivation of this result is given in Sect. A8.4.1, where it is shown that the spectral response to a sine wave is a  $\text{sinc}^2$  function.

For the XF configuration, the spectral resolution depends on the length of the correlation function, calculated as described in Sect. 8.8.2. Since the correlation function is calculated on a segment of data that is much longer than the block length, the density of lag multiplications is essentially uniform, except for a very small end effect. Hence, the spectral response is  $\text{sinc}(Nt_s\nu)$  or  $\text{sinc}(n)$ . The spectral responses for the FX and XF correlators are shown in Fig. 8.23.

Note that the integral over frequency of both of these spectral responses is unity. Therefore, the flux or the area under a spectral line profile, the convolution of the source spectrum with the spectral response function, is conserved. The peak amplitude of a spectral feature narrower than the resolution will depend on where it falls with respect to the spectral channels. A line that falls midway between two channels will have its peak amplitude reduced by  $\text{sinc}^2(1/2) = 0.41$  for the FX processor compared with  $\text{sinc}(1/2) = 0.81$  for the XF processor. This is the well-known effect called scalloping. It can be mitigated by the technique of padding with zeros to obtain an interpolated spectrum (see Sect. A8.4.2).



**Fig. 8.23** (left) The density of lag calculations, or intrinsic weighting function, for an FX correlator (solid line) and an XF correlator (dotted line).  $N$  is the segment size for the FX correlator. For comparison purposes, the width of the function for the XF correlator is chosen to make the number of spectral channels the same. (right) The spectral response for the FX correlator (solid line), a sinc<sup>2</sup> function (see Appendix 8.4 for additional explanation) and the XF correlator (dotted line), a sinc function given by Eq. (8.114). Adapted from Romney (1999) and Deller et al. (2016).

In some cases, it may be desirable to actually calculate the correlation functions from the output of the F engine, such as for application of a full nonlinear quantization correction. It is well known (Press et al. 1992) that it is necessary to pad the spectrum with  $N$  zeros in order to obtain the correct result [see discussion after Eq. (A8.40)]. The implementation of this calculation is discussed by O’Sullivan (1982) and Granlund (1986).

**Signal-to-Noise Ratio.** The fundamental difference between the FX and XF processors is the density weighting in the lag domain. Both systems have the same number of equivalent multiplications, as can be seen in Fig. 8.23. The FX covers twice the number of lags as the XF processor but has lower density as the lag number increases. In particular, the FX provides half the lag density for  $k = N/2$ . For a continuum source, the signal-to-noise ratio of the FX and XF systems is the same. This can be appreciated by the fact that only the zero lag multiplications are important, and they are equal in both systems. Similarly, the signal-to-noise ratio for a very-narrow-bandwidth source, less than the resolution, is also the same because the total number of equivalent multiplications is the same.

There is a small difference in response for signals that have line widths about equal to the resolution. In particular, for this case, the amplitude of a spectral line is reduced by a factor of about 0.82 (Okumura et al. 2001; Bunton 2005). This is a problem only for slightly resolved spectral features. In any event, most spectrometers are designed to produce several channels per resolution element in order to properly analyze the lines. This perceived deficiency in the FX correlator is due to the distribution of lags. The FX correlator has a larger range but fewer multiplications at lag( $k$ )  $\pm N/2$  (see Fig. 8.23). There are several approaches to recovering this loss of information. The classic method (Welch 1967; Percival and Walden 1993) is to overlap the segments in the block processing in the F engine.

A 50% overlap recovers most of the lost signal-to-noise ratio but at a cost of doubling the processing time in the F engine. This overlap feature was available in the original FX VLBA processors but rarely, if ever, used (Romney 1995). Another approach is to simply channel-average the spectrum, but this wastes the resolution capability of the F engine. Note that with polyphase filter banks, scalloping for narrow spectral lines and signal-to-noise ratio loss are very small.

*Number of Operations.* We can make an approximate comparison of the workload requirements of XF and FX signal processors by comparing the number of multiplications needed in each system. For this rather simplistic analysis, we assume that the data are streams of real numbers at the Nyquist interval appropriate for bandwidth  $\Delta\nu$ , i.e.,  $t_s = 1/(2\Delta\nu)$ . To make this comparison, we further assume that the number of lags computed in the X engine (lag correlator),  $N$ , is equal to the data segment length into the F engine. This makes the spectral resolution of both systems approximately equal (see Fig. 8.23 for exact responses).

Consider the analysis of one second of data, i.e.,  $2\Delta\nu$  samples. For the XF system, a lag correlator is required for each baseline. Thus,  $2N\Delta\nu$  multiplications are required for each baseline. Since  $Nt_s \ll 2\Delta\nu$ , the edge effects in calculating the correlation function are negligible (i.e., all lags have almost the same number of multiplications approaching  $N$ ), and the workload of the single Fourier transform at the end of the integration period is negligible compared with the workload of calculating the correlation function. Thus, the rate of multiplications (multiplies per second),  $r_{XF}$ , is

$$r_{XF} = 2\Delta\nu N n_b, \quad (8.116)$$

where  $n_b$  is the number of baselines. For the FX processor, one DFT engine is required for each antenna. We assume that the number of multiplications for the FFT implementation of the  $N$ -point DFT is  $N \log_2 N$ . (Some variation exists, depending on the FFT implementation, e.g., an FFT with  $N$  being a power of four would run somewhat faster.) The cross power spectrum calculation requires the pairwise cross multiplication of the outputs of DFT engines for all baselines. These multiplications are complex, requiring four real multiplications each. In addition, only the  $N/2$  spectral points at positive frequencies need to be calculated and retained. The number of multiplications is therefore  $[n_a N \log_2 N + 4N n_b / 2]M$ , where  $M$  is the number of segments processed,  $2\Delta\nu/N$ . Since  $MN = 2\Delta\nu$ , the aggregate multiplication rate is

$$r_{FX} = 2\Delta\nu [n_a \log_2 N + 2n_b]. \quad (8.117)$$

The workload ratio,  $\mathcal{R} = r_{XF}/r_{FX}$  is therefore

$$\mathcal{R} = \frac{n_b N}{n_a \log_2 N + n_b}. \quad (8.118)$$

The  $n_a \log_2 N$  factor reflects the  $\log_2 N$  advantage and the antenna-based processing of the DFT engine. The  $n_b N$  factor reflects the baseline processing of the X engine. Since  $n_b = n_a(n_a - 1)/2$ , we can rewrite Eq. (8.118) as

$$\mathcal{R} = \frac{N}{\frac{2 \log_2 N}{n_a - 1} + 2} . \quad (8.119)$$

Note that this relation holds for  $n_a \geq 2$  because no single antenna spectra are calculated. [Analysis for a spectrometer on a single antenna would yield  $\mathcal{R} = N/(\log_2 N + 1)$ .] The limiting forms of Eq. (8.119) are

$$\begin{aligned} \mathcal{R} &= N/2 , & n_a &\gg 1 + \log_2 N , \\ \mathcal{R} &= \frac{N(n_a - 1)}{2 \log_2 N} & n_a &\ll 1 + \log_2 N . \end{aligned} \quad (8.120)$$

In general, the larger the values of  $N$  or  $n_a$ , the more the FX design is favored. For example, with  $n_a = 10$  and  $N = 1024$ ,  $\mathcal{R} \sim 240$ . Perhaps the most important limitation in Eq. (8.119) is that the X engine operates on one or a few-bit representation of the signal and multiplication can be achieved by simple table lookup, whereas the F engine needs more bits per sample and there is additional bit growth in its internal operations. Furthermore, the detailed architecture of chips has a major influence on calculation speed. Hence, Eq. (8.119) is a useful guide for the general dependence of  $\mathcal{R}$  on  $N$  and  $n_a$  but does not accurately specify a crossover point favoring one design over the other. The advantage clearly shifts to the FX design for very large  $n_a$  or  $N$ .

*Digital Fringe Rotation.* In early systems, fringe rotation was often applied to the signal as an analog process, but generally it is advantageous to implement it after digitization. For example, in VLBI observations in which the data are recorded as digital samples, it is useful to be able to repeat the analysis with different fringe rates if the position of the source on the sky is not known with sufficient accuracy before the observation. Digital fringe rotation is usually applied to the digitized IF waveform before it goes to the correlator and involves multiplication with a digitized fringe rotation waveform. It is desirable to use a multibit representation for the rotated data to maintain the required accuracy, and thus, the number of bits in the input data to the correlator may be increased. Increasing the number of bits per sample in a lag correlator results in a proportional increase in complexity. Thus, it may be necessary to truncate the data before input to the correlator, which effectively introduces the quantization loss a second time. In contrast, in the FX design, multibit data representation is required in the FFT processing, so the bit increase that fringe rotation presents is more easily accommodated. See Sect. 9.7.1 for more details.

*Fractional Sample Delay Correction.* In digital implementation of the compensating delays, one way of adjusting the delay in steps smaller than the sampling interval is to adjust the timing of the sampler pulses, as described in Sect. 8.6. Another way



of introducing a fractional sample period delay is done after transformation to the frequency domain by incrementing the phase values by an amount that varies in proportion to the frequency across the IF band. In the FX correlator, this is easily done because the signals appear as an amplitude spectrum every FFT cycle, and the correction can be applied as required for each antenna before the data are combined in antenna pairs. With a lag correlator, there are two problems in this process. First, the transformation to a spectrum occurs after the data are combined for antenna pairs, so many more values require correction. Second, for long baselines, the corrections required may occur more rapidly than the rate at which cross-correlation values are transformed to cross power spectra. Thus, it may be possible to apply only a statistical correction rather than an exact one. See Sect. 9.7.3 for a description of the statistical corrections.

*Quantization Correction.* The nonlinearity of the amplitude of the cross-correlation measured using coarsely quantized samples is seen in the Van Vleck relationship [Eq. (8.25)]. Application of a correction for the nonlinearity in quantization in the lag (XF) correlator is a relatively straightforward process because the cross-correlation values are directly calculated. To obtain the cross-correlation values in the FX correlator, the cross power spectrum at the correlator output must be Fourier transformed from the frequency domain to the lag domain. After applying the correction, the data must then be transformed back to a frequency spectrum. The correction is necessary only if the correlation of the total waveform (signal plus noise) is large for any pair of antennas. This condition implies observation of a source that is largely unresolved and sufficiently strong that the signal power in the receiver is comparable to the noise or greater. In the case of a spectral line observation, it is the power averaged over the receiver bandwidth that is important.

*Adaptability.* The FX design is somewhat more easily expanded or adapted to special requirements because more of the system is modularized per antenna rather than per baseline, as in the lag correlator. Addition of an extra antenna to an FX correlator requires less modification of the reduction procedure than is necessary for a lag correlator. Thus, the FX design is convenient for projects in which the number of antennas is planned to increase over time and is more efficient for larger arrays (Parsons et al. 2008).

*Pulsar Observations.* For pulsar observations, a gating system at the correlator output is required to separate data received during the pulsar-on period, so that the sensitivity is not degraded by noise received when the pulsar is off. For many pulsars, which have periods  $\geq 0.1$  s, time resolution of order 1 ms is adequate in the gating.<sup>7</sup> With an FX correlator, it is necessary to collect data in complete sequences of  $N$  samples, so the gating process has to accommodate data that arrive

---

<sup>7</sup>Many arrays can also be used in a phased-array mode (e.g., for VLBI, see Sect. 9.9), which provides one signal output per polarization. A specially designed pulsar processor can then provide measurements with high time resolution for study of the pulse profile and timing. In such cases, the array is used only to provide a large collecting area for high sensitivity. See Sect. 9.9.

at time intervals of  $\sim N$  times the sample interval  $t_s$ . For example, with  $N = 1,024$  and a total bandwidth of 10 MHz,  $Nt_s \simeq 500 \mu\text{s}$ . Again, this might restrict flexibility for the fastest pulsars. However, a nice feature of the FX correlator is that complete spectra are obtained during each  $Nt_s$  interval in time. In the subsequent time averaging, it is possible to process the frequency channels individually and to vary the time of the gating pulse for each one so as to match the variation in pulse timing that results from dispersion in the interstellar medium.

*Choice of Correlator Design.* Because the relative advantages of the lag and FX schemes discussed above involve a number of different features, the best choice of architecture for any particular application may not be immediately obvious. Detailed design studies for different approaches, taking account of the precise requirements and the implementation of the very-large-scale integrated (VLSI) circuits, are required. For discussions of lag and FX correlators, see D’Addario (1989), Romney (1995, 1999), and Bunton (2003). The widespread use of polyphase filter banks for precise channel definitions and radio frequency interference (RFI) excision favors the FX approach (see Sect. 8.8.9).

### 8.8.6 Hybrid Correlator

In designing a broadband correlator, it may be advantageous or necessary to divide the analog signal of total bandwidth,  $\Delta\nu$ , from each antenna into  $n_f$  contiguous narrow sub-bands. A separate digital sampler is used for each such sub-band, and the correlator is designed as  $n_f$  sections operating in parallel to cover the full signal band. A system of this type that incorporates both analog filtering and digital frequency analysis is referred to as a *hybrid correlator*. If the digital part uses a lag design, then the rate of digital operations is reduced by a factor  $n_f$  relative to the rate for a lag correlator that processes the whole bandwidth without subdivision. This can be seen from Eq. (8.116), where for one sub-band, the bandwidth is  $\Delta\nu_s = \Delta\nu/n_f$ , the number of channels required is  $N/n_f$ , but  $n_f$  such sections of digital processing are required. We can write a cost equation for a hybrid correlator (Weinreb 1984), as

$$C = A_1 \frac{\Delta\nu n_a (n_a - 1) N}{n_f} + A_2 n_f n_a + A_3, \quad (8.121)$$

where  $A_1$  and  $A_2$  are coefficients for the digital and analog hardware, respectively, and  $A_3$  is another constant.

In this equation, the cost can be minimized with respect to  $n_f$ , with the result that

$$n_f = \left[ \frac{A_1}{A_2} \Delta\nu (n_a - 1) N \right]^{1/2}. \quad (8.122)$$

**Table 8.6** Hybrid channelization

| Instrument      | Year<br>Commissioned | $\Delta\nu$<br>(GHz) | $\Delta\nu_s$<br>(MHz) | $n_f$           | Reference                |
|-----------------|----------------------|----------------------|------------------------|-----------------|--------------------------|
| SMA             | 2003                 | 2.0 <sup>b</sup>     | 104                    | 24 <sup>a</sup> | Ho et al. (2004)         |
| SMA             | 2015                 | 8.0                  | 2000                   | 4               | —                        |
| Plateau de Bure | 1992                 | 0.5                  | 50                     | 10              | Guilloteau et al. (1992) |
| Plateau de Bure | 2013                 | 8.0                  | 2000                   | 4               | Jan (2008)               |
| ALMA            | 2013                 | 4.0 <sup>b</sup>     | 2000 <sup>c</sup>      | 2               | Escoffier et al. (2007)  |

<sup>a</sup> $n_f$  is only approximately  $\Delta\nu/\Delta\nu_s$  because of sub-band overlap.

<sup>b</sup>Two polarizations or two bands.

<sup>c</sup>These sub-bands are subsequently reduced to 128 channels of 62.5 MHz each by digital filtering. Each 2000-MHz band can be positioned independently.

Equation (8.122) is useful only if the digital electronics are fast enough to handle a bandwidth of  $\Delta\nu/n_f$ . Over the last decades, the sampling rates have steadily risen and the costs have dropped for digital hardware, while the cost of analog electronics has remained relatively flat. The evolution of design in hybrid correlators can be seen in Table 8.6. A general disadvantage of the hybrid correlator is that very careful calibration of the frequency responses of the sub-bands is required to avoid discontinuities in gain at the sub-band edges. In general, it is advantageous to use the fastest samplers to minimize the analog filtering required. However, at millimeter wavelengths, where very wide bandwidths are needed and can be accommodated by receivers, the restriction on digital sampling speed requires some channelization. If an FX implementation is used for the digital section, a similar cost equation can be written, but there is less reduction in the number of operations since in Eq. (8.117),  $N$  enters logarithmically.

### 8.8.7 Demultiplexing in Broadband Correlators

The bit rate for the VLSI circuits used in large correlator systems is generally slower than that of the digital samplers that are used with broadband correlators. Serial-to-parallel conversion at the sampler output, that is, demultiplexing in the time domain, allows use of optimum bit rates for the correlator. Consider a system in which each sampler output is demultiplexed into  $n$  streams, and assume for simplicity that there is one bit per sample; parallel architecture accommodates multiple bits. Any  $n$  contiguous samples all go to different streams. To obtain all the products required in a lag correlator for a pair of IF signals with this configuration of the data, it would be necessary to include cross-correlations between each stream of one signal with every stream of the other signal. To simplify the system, Escoffier (1997) developed a scheme in which the  $n$  demultiplexed bit streams from each signal are fed into a large random access memory (RAM) and read out in reordered form. Each demultiplexed stream then contains a series of discontinuous blocks of

$\sim 10^5$  samples. Each block contains data contiguous in time, as sampled. Cross-correlations are performed between data in corresponding blocks only. Thus, for any pair of input signals,  $n$  cross-correlators running at the demultiplexed rate are required for each value of lag. Also, each signal requires two RAM units so that one is filled as the other is read out. In Escoffier's system, the sample rate is  $4 \text{ Gbit s}^{-1}$ ,  $n = 32$ , and the length of a block of the demultiplexed data is approximately 1 ms. Since cross-correlations do not extend across the boundaries of any given block, there is a very small loss of efficiency, which in this case is about 0.2%. Another possible approach is based on demultiplexing in the frequency domain, as in the case of the hybrid correlator. It is then necessary only to cross-correlate corresponding frequency channels between each antenna, so the number of cross-correlators per signal pair is again equal to  $n$  for each lag. Carlson and Dewdney (2000) have described an all-digital development of the frequency demultiplexing principle used in the hybrid correlator. This is used with the expanded VLA (Perley et al. 2009), and the system is described as a WIDAR correlator. Broadband signals are digitized at full bandwidth, divided into frequency channels using digital filters, and resampled at the appropriate lower rate before cross-correlation between all antenna pairs. (The use of digital filters avoids the small differences in the responses of analog filters, which in some systems provide the initial channelization.) As a final step, the cross-correlated data are Fourier transformed to the frequency domain. This scheme is sometimes referred to as an FFX system. Both Escoffier's reordering scheme and the WIDAR system of demultiplexing provide approaches to the design of large broadband correlators. The latter requires fewer lags because the digital filters provide part of the spectral resolution.

For filtering sampled signals, digital filters of the FIR (finite impulse response) type can be used, in which the incoming sample stream is convolved with a series of numbers, referred to as tap weights, the Fourier transform of which represents the filter response (Escoffier et al. 2000). The tap weights can be stored in a RAM and readily changed as required. An important advantage of digital filters is the freedom from individual variations of the characteristics. However, it may be necessary to truncate the output data samples to match the number of bits per sample that can be handled by the correlator, and thus a further quantization loss may be incurred.

### 8.8.8 Examples of Bandwidths and Bit Data Quantization

The initial observing bandwidth of the 27-antenna VLA, when it came into operation in the early 1980s, was 100 MHz per polarization with three-level (2-bit) sampling. The expanded system that came into operation around 2010, covering a frequency range of 1–50 GHz, has a maximum observing bandwidth of 8 GHz per polarization with 3-bit sampling, or 8-bit sampling with a reduced bandwidth (Perley et al. 2009). This large increase in data capacity is possible as a result of the increase in computing speed and in signal transmission capacity using optical fiber.

The Atacama Large Millimeter/submillimeter Array, which came into operation in 2012, covers bandwidths of 8 GHz per polarization with 3-bit (8-level) quantization (Wootten and Thompson 2009). The number of antennas is 64 and the correlator is FX with initial digital filtering, sometimes referred to as an FFX system.

In the meter-wavelength range, observing bandwidths are generally narrower than at shorter wavelengths, but the spectrum is often more heavily used by transmitting services, so the requirement for avoiding or removing interfering signals is important. Larger numbers of bits allow for greater dynamic range in the system response, which helps to reduce the probability that interfering signals will cause overloading. The LWA (Long Wavelength Array) covers 20–80 MHz using 8-bit (256-level) sampling with the option of 12-bit (4096-level) sampling. The sample frequency is 196 mega-samples/sec (Ellingson et al. 2009). The LOFAR system covers 15–80 MHz and 110–240 MHz using 12-bit digitization (de Vos et al. 2009).

### 8.8.9 Polyphase Filter Banks

Polyphase filtering is a digital signal-processing technique that was developed for applications such as the separation of signals in multichannel communication systems with high interchannel rejection (Bellanger et al. 1976). The disadvantages of the nonoverlapping-segment discrete Fourier transform (DFT) processing, which we will call the single-block Fourier transform (SBFT) method, have been noted in earlier sections of this chapter and are also described in Appendix 8.4. Namely, this approach has high spectral leakage since the spectral response is a sinc-squared function that has sidelobe levels as high as  $-13.5$  db. In addition, the amplitude of a monochromatic signal, or unresolved cosmic line, depends on its relative location with respect to the channel boundaries, going from 1 at channel center to  $(2/\pi)^2 = 0.41$  at the edge. This effect is called scalloping. There is a slight loss in sensitivity for signals whose line widths are close to the spectral resolution, which is related to the effective lag distribution of the DFT (see Sect. 8.8.5).

Polyphase filtering and polyphase filter banks (PFBs) correct these deficiencies at a modest computational overhead. PFBs have become an important tool in radio astronomy as a way of excising radio frequency interference since they make possible the elimination of only the specific channels in which the interference occurs. It is also helpful in spectroscopic observations of some cosmic sources such as masers, where a very strong and narrow line in the passband makes it difficult to study other nearby lines because of the effect of spectral leakage. For detailed treatments of PFBs, see Crochiere and Rabiner (1981), Vaidyanathan (1990), and Harris et al. (2003). Useful tutorials are available by Harris (1999) and Chennamangalam (2014). For applications to radio interferometry, see Buntun (2000, 2003).

Before describing the PFB, consider an elementary design of a digital filter bank based on a conventional analog filter bank with  $M$  equally spaced filters spanning the

frequency range 0 to  $\Delta\nu$ . Suppose the input voltage is  $x(t)$ , which is a bandlimited Gaussian process in the frequency range 0 to  $\Delta\nu$ .  $x(t)$  can be represented by a digital sequence  $x(n)$ , sampled at the Nyquist interval  $1/2\Delta\nu$ . A crude lowpass filter can be constructed by taking a running mean of  $M$  samples in the time domain. The spectral response to this “boxcar” averaging is a sinc function with its first null at  $2\Delta\nu/M$ . Obtaining a perfect lowpass filter response in the frequency domain with a cutoff at  $\nu_c = \Delta\nu/M$  would require the convolution in the time domain with a sinc function [ $\text{sinc}(x) = \sin(\pi x)/(\pi x)$ ],

$$h(t) = \text{sinc}(2\nu_c t) = \text{sinc}(n/M) . \quad (8.123)$$

Note that for  $M = 1$ ,  $h(t) = 1$  for  $n = 0$  and otherwise is zero, so  $x(n)$  remains unchanged. However, for  $M > 1$ , perfect lowpass filtering action requires a convolution over infinite time. As an approximation, we can use  $N$ -point smoothing. The filter shape will be

$$H(\nu) = \sum_{n=0}^{N-1} \text{sinc}(n/M) e^{j(\pi \nu n / \Delta\nu)} , \quad (8.124)$$

which has a fairly sharp cutoff at  $\nu = \Delta\nu/M$ .  $y(n)$ , the smoothed version of  $x(n)$ , will be oversampled by a factor of about  $M$ . The normal process at this point is to resample  $y(n)$ , taking every  $M$ th sample. This process is called decimation,<sup>8</sup> or downsampling. To make the rest of the filter bank, multiply  $x(n)$  by  $e^{j2\pi\nu t}$ , where  $\nu = m/\Delta\nu$  and  $m = 1$  to  $M - 1$ , filter each stream by  $h(n)$ , and downsample. This process is inefficient since the downsampling discards most of the arithmetic computations. The PFB provides a more efficient processing structure to obtain a filter shape with a sharp cutoff.

We now describe the PFB, following the analysis of Bunton (2000, 2003). Consider a sample sequence of data,  $x(n)$ , of length  $N$ , which is multiplied by a window function  $h(n)$ . Its DFT is

$$X(k) = \sum_{n=0}^{N-1} h(n)x(n) e^{-j(2\pi/N)nk} , \quad (8.125)$$

where  $k$  ranges from 0 to  $N - 1$ . The frequency steps are  $2\Delta\nu/N$ , i.e., covering both positive and negative frequencies. If  $H(k)$ , the DFT of  $h(n)$ , has a width of approximately  $2\Delta\nu/M$ , then  $X(k)$  will be oversampled, and only  $r = N/M$  samples need be retained.  $N$  and  $M$  are chosen so that  $r$  is an integer. If  $H(k)$  is desired to be

---

<sup>8</sup>“Decimation” formally means a reduction to 1/10; however, the broader definition is to “reduce drastically, especially in number.”

the discrete idealization of a perfect lowpass filter, i.e.,

$$\begin{aligned} H(k) &= 1, & N - M < k \leq M, \\ &= 0, & \text{otherwise,} \end{aligned} \quad (8.126)$$

then

$$h(n) \simeq \text{sinc} \left[ \left( \frac{n - \frac{N}{2}}{N} \right) r \right]. \quad (8.127)$$

The decimated spectrum, i.e., taking every  $r$ th point of  $X(k)$  in Eq. (8.125), is

$$X(k') = \sum_{n=0}^{N-1} h(n)x(n) e^{-j(2\pi/N)nrk'}, \quad (8.128)$$

where  $k'$  goes from 0 to  $M - 1$ . We can rewrite Eq. (8.128) as a double summation over  $r$  subsegments, each of length  $M$ , as

$$X(k') = \sum_{m=0}^{r-1} \sum_{n=0}^{M-1} h(n + mM) x(n + mM) e^{-j(2\pi/N)(n+mM)rk'}. \quad (8.129)$$

Notice that

$$e^{-j(2\pi/N)(n+mM)rk'} = e^{-j(2\pi nk'/M)} e^{-j2\pi mk'}. \quad (8.130)$$

The rightmost exponential factor is unity. Hence,

$$X(k') = \sum_{m=0}^{r-1} \sum_{n=0}^{M-1} h(n + mM) x(n + mM) e^{-j(2\pi/M)nk'}. \quad (8.131)$$

In Eq. (8.131), there are  $r$  DFTs of length  $M$ , and in Eq. (8.128), there is one DFT of length  $N = rM$ , so there is only a slight reduction in the workload, approximated by the number of multiplications required. Note that the FFT algorithm, which has a workload proportional to  $M \log_2 M$ , is used for the DFT calculation.

The kernel of the exponential in Eq. (8.131) does not contain  $r$ , so we can interchange the order of summation and rewrite it as

$$X(k') = \sum_{n=0}^{M-1} \left[ \sum_{m=0}^{r-1} h(n + mM) x(n + mM) \right] e^{-j(2\pi/M)nk'}. \quad (8.132)$$

This step reduces the calculation from  $r$  DFTs of length  $M$  to one DFT of length  $M$ . The workload for applying the window function  $h(n)$  remains proportional

to  $N$ . Hence, the workload for Eq. (8.128) is  $N + N \log_2 N$ , while the workload for Eq. (8.131) is  $N + M \log_2 M$ . The workload is thus reduced by a factor of  $\mathcal{R}$ , given by

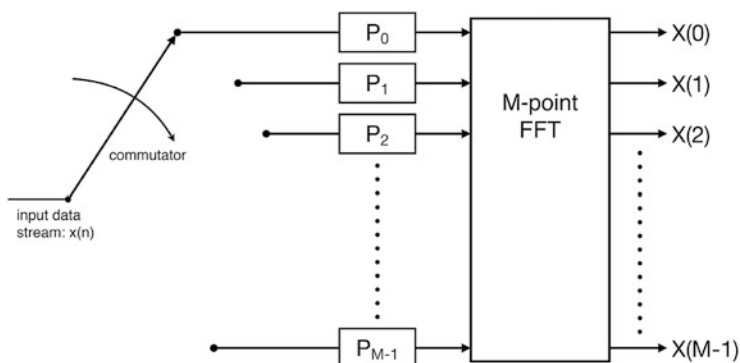
$$\mathcal{R} = \frac{N + N \log_2 N}{N + M \log_2 M} = \frac{1 + \log_2 N}{1 + \frac{1}{r} (\log_2 N - \log_2 r)} \simeq r, \quad (8.133)$$

where the approximation holds for  $N \gg 1$ .

After the calculation in Eq. (8.132) is performed, the  $N$ -point window is moved by  $M$  steps, and the process is repeated. Each segment of  $M$  points is thus processed  $r$  times. Therefore, the input and output data rates are the same, except when spectral values at negative frequencies are discarded.

The calculation in Eq. (8.132) is expressed diagrammatically in Fig. 8.24. This process may seem counterintuitive in the following sense. The data stream is severely decimated by the action of the commutator, which distributes the time samples among the branches, or “partitions,” with a cycling period  $M$ . That is, the data samples into each of the  $M$  partitions are

$$\begin{aligned} & x(0), \quad x(M), \quad x(2M), \quad \dots, \quad x(rM - M) \\ & x(1), \quad x(M + 1), \quad x(M + 2), \quad \dots, \quad x(rM - M + 1) \\ & x(2), \quad x(M + 2), \quad x(M + 3), \quad \dots, \quad x(rM - M + 2) \\ & \vdots \\ & x(M - 1), \quad x(M), \quad x(M + 1), \quad \dots, \quad x(rM - 1). \end{aligned} \quad (8.134)$$



**Fig. 8.24** A diagram of a polyphase filter bank, which converts a set of  $N$  data samples into an  $M$ -point spectrum. The input data stream is distributed among the  $M$  filter partitions by a commutator. Each partition receives a data stream that has been downsampled by a factor  $M$ . In each partition,  $P_M$  represents the action of the decimated version of  $h(n)$ , as described by the term in brackets in Eq. (8.132). The nonaliased  $M$ -point spectrum is assembled by the action of the FFT. Note that if the data samples are real numbers, then only  $M/2$  values of the spectrum, corresponding to the positive frequencies, need be retained.



Each of these decimated data streams is undersampled by a factor of  $M$ , and its corresponding spectrum is heavily aliased. The action of the PFB undoes this aliasing.

Consider an example where  $N = 1024$ ,  $M = 256$ , and  $r = 4$  (a four-tap polyphase filter), as shown in Fig. 8.25. The first polyphase partition,  $P_0$ , calculates only the four-term sum  $x(0)h(0) + x(256)h(256) + x(512)h(512) + x(768)h(768)$ , and  $P_1$  calculates  $x(1)h(1) + x(257)h(257) + x(513)h(513) + x(769)h(769)$ .

We can now compare the performance and requirements of the SBFT and the PFB. The SBFT produces an  $M$ -point spectrum for each  $M$  data samples. It moves successively from block to block, so the data rate remains the same. The PFB takes in  $N$  data samples and produces an  $M$ -point spectrum and then steps by  $M$  samples for the next spectral calculation. Hence, its data rate also remains the same. The overhead in the PFB is due to the windowing. Hence, the workload ratio  $\mathcal{R}$  needed for the PFB with respect to the SBFT is

$$\mathcal{R} = \frac{N + M \log_2 M}{M \log_2 M} = 1 + \frac{r}{\log_2 M} . \quad (8.135)$$

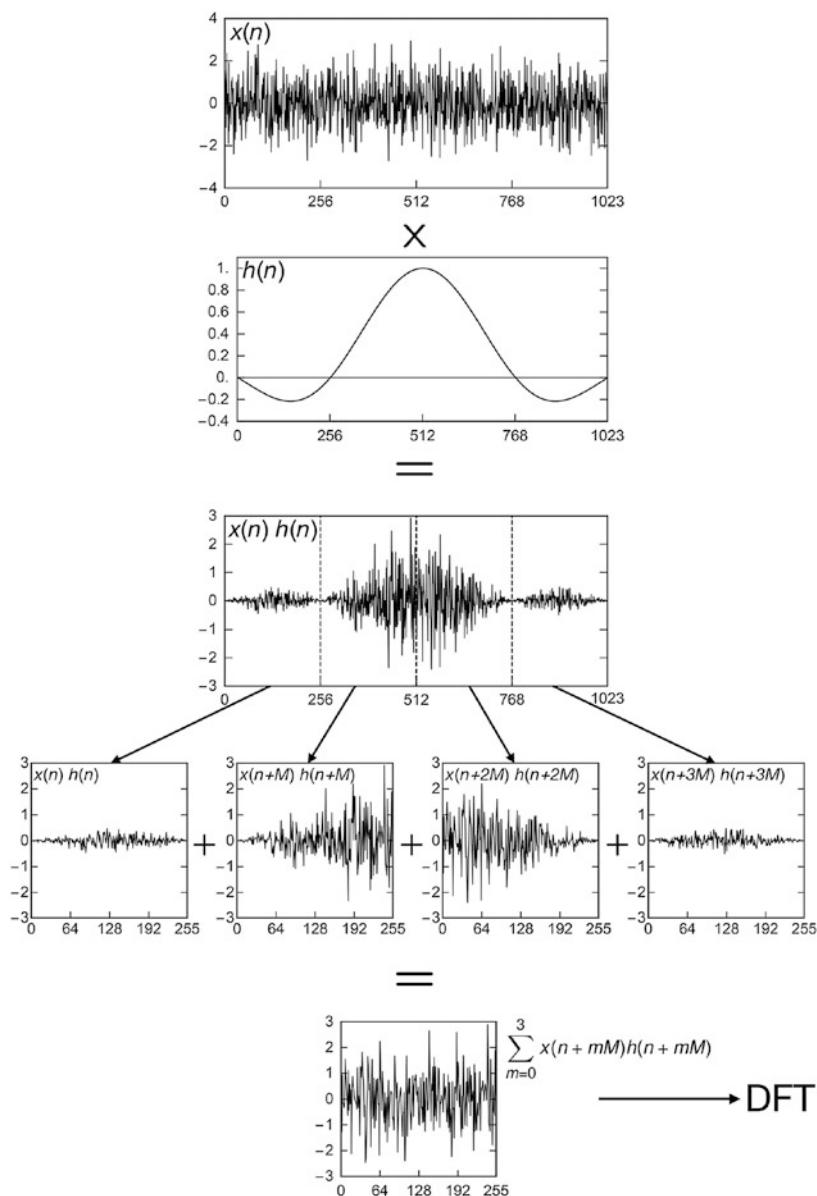
For  $M = 1024$  and  $r = 4$ , there is a 40% overhead incurred with the PFB structure. The flat response on low leakage from the PFB is made possible because there are  $N$  samples available to provide the filter action rather than  $M$ . Note that in hardware implementation, the buffering requirement increases with  $r$ .

It is advantageous to apply additional weighting to  $h(n)$  such as Hann, Hamming, or Blackman weighting to further reduce spectral leakage. This does not reduce the resolution significantly as long as the weighting function remains at a level of  $\sim 1$  over  $M$  samples. Examples of PFB and SBFT filter shapes are shown in Fig. 8.26. If the weighting is applied in the SBFT mode over  $M$  samples, the leakage is reduced, but the resolution is also reduced.

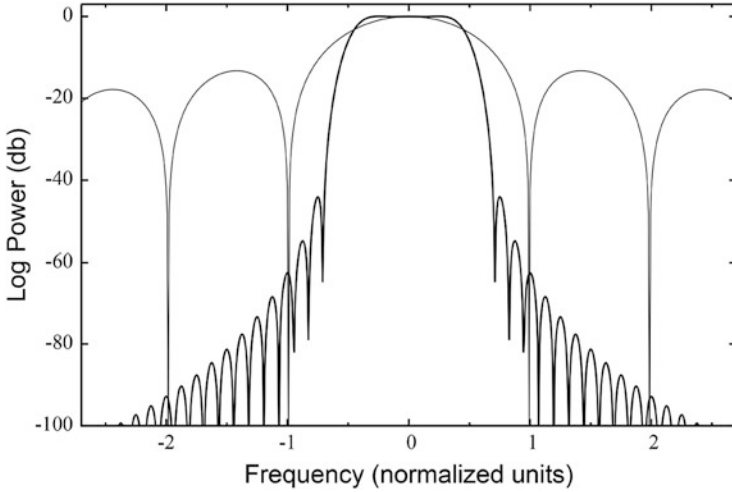
Note that PFBs can be concatenated. The output of any subset or all of the channels of a PFB can be fed into an additional PFB to obtain finer resolution. The Murchison Widefield Array uses such a scheme. Another application is to use a PFB only for coarse channelization. Its output can then be fed to an XF or FX correlator.

### 8.8.10 Software Correlators

Since, in practice, the signals for which cross-correlations are formed are in digital form, having also been subjected to a digital delay system, the cross multiplication and averaging processes can be carried out in a computer system. This is useful in



**Fig. 8.25** A graphical representation of the action of a polyphase filter bank with  $r = 4$ , or four taps. A random noise data stream represented by a set of  $N$  independent Gaussianly distributed random noise (i.e., white noise) is shown in the top panel. It is multiplied by a window function  $h(n)$ , the envelope of which is shown in the next panel. Here,  $h(n)$  is chosen to be a sinc functions with exactly four zero crossings, equal to the number of taps. The result is separated into four segments, which are coadded to form the  $M$ -term time series shown in the lowest panel, which is then Fourier transformed into an  $M$ -point spectrum,  $-\Delta\nu$  to  $\Delta\nu$ , as formulated by Eq. (8.132). After this calculation, the window is moved by  $M$  samples, and the process is repeated. Adapted from Gary (2014).



**Fig. 8.26** The thick line shows the response of a filter element in a PFB having  $r = 8$  and Hann weighting applied. The thin line shows the response for a SBFT, a  $\text{sinc}^2$  function. Both filters have a response of about  $(2/\pi)^2$  at the filter edge of  $\pm 0.5$  in normalized frequency units of  $2\Delta\nu/M$ .

small systems for which the development of special correlator hardware is avoided. Also, in the case of large systems in which antennas are brought into operation over a period of years, changes in the correlation requirements are more easily accommodated. An example of a software correlator and the advantages of the design are described by Deller et al. (2007). Most VLBI processing is done in software correlators.

## Appendix 8.1 Evaluation of $\Sigma_{q=1}^{\infty} R_{\infty}^2(q\tau_s)$

The periodic function  $f(t)$  can be expressed as a Fourier series as follows:

$$f(t) = \frac{a_0}{2} + \sum_{q=1}^{\infty} \left[ a_q \cos\left(\frac{2\pi qt}{\beta}\right) + b_q \sin\left(\frac{2\pi qt}{\beta}\right) \right], \quad (\text{A8.1})$$

where  $\beta$  is the period and

$$a_q = \frac{2}{\beta} \int_0^{\beta} f(t) \cos\left(\frac{2\pi qt}{\beta}\right) dt, \quad (\text{A8.2a})$$

$$b_q = \frac{2}{\beta} \int_0^{\beta} f(t) \sin\left(\frac{2\pi qt}{\beta}\right) dt. \quad (\text{A8.2b})$$

Parseval's theorem for Eq. (A8.1) takes the form

$$\frac{2}{\beta} \int_0^\beta f^2(t) dt = \frac{a_0^2}{2} + \sum_{q=1}^{\infty} (a_q^2 + b_q^2). \quad (\text{A8.3})$$

Now let  $f(t)$  be a series of rectangular functions of unit height and width, one centered on  $t = 0$  and the others centered on integral multiples of  $\pm\beta$ . Then, one obtains

$$a_0 = \frac{2}{\beta}, \quad a_q = \frac{2}{\beta} \frac{\sin(\pi q/\beta)}{\pi q/\beta}, \quad b_q = 0, \quad \int_0^\beta f^2(t) dt = 1. \quad (\text{A8.4})$$

From Eqs. (A8.3) and (A8.4),

$$\sum_{q=1}^{\infty} \left[ \frac{\sin(\pi q/\beta)}{\pi q/\beta} \right]^2 = \frac{\beta - 1}{2}, \quad (\text{A8.5})$$

which, from Eq. (8.15), is the summation needed to evaluate Eq. (8.19).

## Appendix 8.2 Probability Integral for Two-Level Quantization

The probability integration required in Eq. (8.21) can be performed as follows. The integral is

$$P_{11} = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty \exp\left[\frac{-(x^2 + y^2 - 2\rho xy)}{2\sigma^2(1-\rho^2)}\right] dx dy. \quad (\text{A8.6})$$

Restore circular symmetry in the integral by the substitutions

$$z = \frac{y - \rho x}{\sqrt{1-\rho^2}}, \quad dy = \sqrt{1-\rho^2} dz. \quad (\text{A8.7})$$

Then

$$P_{11} = \frac{1}{2\pi\sigma^2} \int_0^\infty dx \int_{\frac{-\rho x}{\sqrt{1-\rho^2}}}^\infty \exp\left[\frac{-(x^2 + z^2)}{2\sigma^2}\right] dz. \quad (\text{A8.8})$$

Next, substitute  $x = r \cos \theta$  and  $z = r \sin \theta$ . The lower limit of the  $z$  integral in Eq. (A8.8) represents the line  $z = -\rho x/\sqrt{1-\rho^2}$ , which makes an angle  $\theta$  with the  $x$  axis given by  $\theta = -\sin^{-1} \rho$ . The integral covers an area of the  $(x, z)$  plane

between this line and the  $z$  axis ( $\theta = \pi/2$ ). Thus,

$$P_{11} = \frac{1}{2\pi\sigma^2} \int_0^\infty dr \int_{-\sin^{-1}\rho}^{\pi/2} r \exp\left(\frac{-r^2}{2\sigma^2}\right) d\theta . \quad (\text{A8.9})$$

Finally, substitute  $u = r^2/2\sigma^2$ :

$$P_{11} = \frac{1}{2\pi} \int_0^\infty du \int_{-\sin^{-1}\rho}^{\pi/2} e^{-u} d\theta . \quad (\text{A8.10})$$

Equation (A8.10) can be integrated directly to give

$$P_{11} = \frac{1}{4} + \frac{1}{2\pi} \sin^{-1} \rho . \quad (\text{A8.11})$$

### Appendix 8.3 Optimal Performance for Four-Level Quantization

Schwab (1986) has investigated various aspects of the performance of correlators with four-level quantization. These include precise values for optimal thresholds and quantization efficiencies, and expressions for computation of the cross-correlation as a function of the correlator output. The threshold values and the efficiencies are given Table A8.1.

The values of quantization efficiency  $\eta_4$  for  $n = 3$  and 4 are within 0.3% of the highest value and are useful because nonintegral values of the weighting factor  $n$  would require more complicated implementation in a lag-type correlator. Rational approximations for the cross-correlation  $\tilde{\rho}$  are minimax solutions; that is, they minimize the maximum relative error. The variable  $r_N$  is the normalized correlator output, that is, the measured output divided by the corresponding output for  $\rho = 1$ . The first three approximations given below are valid for all  $|r_N| \leq 1$ .

For  $n = 3$  and the corresponding value of  $v_0/\sigma$  in Table A8.1, the following approximation yields a maximum relative error of  $1.51 \times 10^{-4}$ :

$$\tilde{\rho}(r_N) = \frac{1.1347043 - 3.0971312r_N^2 + 2.9163894r_N^4 - 0.89047693r_N^6}{1 - 2.6892104r_N^2 + 2.4736683r_N^4 - 0.72098190r_N^6} r_N . \quad (\text{A8.12})$$

For  $n \approx 3.3359$  and the corresponding value of  $v_0/\sigma$  in Table A8.1, the following approximation yields a maximum relative error of  $1.46 \times 10^{-4}$ :

$$\tilde{\rho}(r_N) = \frac{1.1329552 - 3.1056902r_N^2 + 2.9296994r_N^4 - 0.90122460r_N^6}{1 - 2.7056559r_N^2 + 2.5012473r_N^4 - 0.73985978r_N^6} r_N . \quad (\text{A8.13})$$

**Table A8.1** Optimal thresholds and efficiencies for four-level quantization

| $n$       | $v_0/\sigma$ | $\eta_4$     |
|-----------|--------------|--------------|
| 3         | 0.99568668   | 0.8811539496 |
| 3.3358750 | 0.98159883   | 0.8825181522 |
| 4         | 0.94232840   | 0.8795104597 |

For  $n = 4$  and the corresponding value of  $v_0/\sigma$  in Table A8.1, the following approximation yields a maximum relative error of  $1.50 \times 10^{-4}$ :

$$\tilde{\rho}(r_N) = \frac{1.1368256 - 3.0533973r_N^2 + 2.8171512r_N^4 - 0.85148929r_N^6}{1 - 2.6529114r_N^2 + 2.4027335r_N^4 - 0.70073934r_N^6} r_N. \quad (\text{A8.14})$$

The following approximation also applies for  $n = 4$  and the corresponding value of  $v_0/\sigma$  in Table A8.1 but is valid for only  $|r_N| \leq 0.95$ . It yields a maximum relative error of  $2.77 \times 10^{-5}$ :

$$\tilde{\rho}(r_N) = \frac{1.1369813 - 1.2487891r_N^2 + 4.5380174 \times 10^{-2}r_N^4 - 9.1448344 \times 10^{-3}r_N^6}{1 - 1.0617975r_N^2} r_N. \quad (\text{A8.15})$$

## Appendix 8.4 Introduction to the Discrete Fourier Transform

This appendix provides a brief introduction to the discrete Fourier transform (DFT), with specific emphasis on applications important to topics covered in this book. For more comprehensive discussion, see Bracewell (2000) or Oppenheim and Schaffer (2009).

Consider the Fourier transform integral of a function  $x(t)$ , a bandlimited signal (0 to  $\Delta\nu$ ), which has finite duration  $T$ .

$$X(\nu) = \int_0^T x(t) e^{-j2\pi\nu t} dt. \quad (\text{A8.16})$$

We can approximate this integral as

$$X(\nu) \simeq \Delta \sum_{n=0}^{N-1} x(t_n) e^{-j2\pi\nu t_n}, \quad (\text{A8.17})$$

where  $x(t_n)$  is a sampled version of  $x(t)$  at the Nyquist interval  $\Delta = 1/2\Delta\nu$  so that  $t_n = n\Delta$ . For simplicity, we assume  $x(t)$  to be a real function. We calculate  $X(\nu)$  at a set of  $N$  frequencies,  $\nu_k = 2k\Delta\nu/N$ , where  $k = 0$  to  $N - 1$ , as

$$X(\nu_k) \simeq \Delta \sum_{n=0}^{N-1} x(t_n) e^{-j2\pi kn/N} . \quad (\text{A8.18})$$

The important next step is to use Eq. (A8.18) as the basis for a definition of the DFT by writing

$$X_k \equiv \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} , \quad (\text{A8.19})$$

where  $x_n$  are the samples  $x(t_n)$ , and  $X_k$  are the corresponding spectral components,  $X(\nu_k)$ . For  $k = 0$ ,

$$X_0 = \sum_{n=1}^{N-1} x_n , \quad (\text{A8.20})$$

which corresponds to the component of  $X$  at  $\nu = 0$ . For  $k = N/2$ ,

$$X_{N/2} = \sum_{n=1}^{N-1} x_n e^{-j\pi n} = \sum_{n=1}^{N-1} x_n (-1)^n , \quad (\text{A8.21})$$

which corresponds to  $X$  at  $\nu = \Delta\nu$ . The negative frequency components lie between  $k = N/2$  and  $N - 1$ , and  $X_N = X_0$ . The inverse DFT is

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{j2\pi kn/N} . \quad (\text{A8.22})$$

We can show that Eq. (A8.22) is indeed the inverse discrete transform of Eq. (A8.19) by substituting Eq. (A8.19) into Eq. (A8.22), that is,

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} \left( \sum_{\ell=0}^{N-1} x_\ell e^{-j2\pi k\ell/N} \right) e^{j2\pi kn/N} . \quad (\text{A8.23})$$

We introduced time index  $\ell$  to distinguish it from  $n$ . Interchanging the order of summation gives

$$x_n = \frac{1}{N} \sum_{\ell=0}^{N-1} x_\ell \left( \sum_{k=0}^{N-1} e^{j2\pi k(n-\ell)/N} \right) . \quad (\text{A8.24})$$

In the summation in parentheses in Eq. (A8.24), the phasor steps uniformly around the complex plane exactly  $n - \ell$  times. Hence,

$$\frac{1}{N} \sum_{k=0}^{N-1} e^{j2\pi k(n-\ell)/N} = \delta_{n\ell} , \quad (\text{A8.25})$$

where  $\delta_{n\ell}$  is called the Kronecker delta function, which has the properties

$$\begin{aligned} \delta_{n\ell} &= 0 , & n &\neq \ell , \\ &= 1 , & n &= \ell . \end{aligned} \quad (\text{A8.26})$$

The Kronecker delta function is nonzero only for  $\ell = n$ , so Eq. (A8.24) yields  $x_n = x_n$  and demonstrates that  $x_n$  is recovered from the original data after a DFT followed by an inverse DFT. Note that  $x_n$  and  $X_k$  are both periodic with period  $N$ . Thus,

$$X_{k+mN} = X_k \quad (\text{A8.27})$$

and

$$x_{n+mN} = x_n , \quad (\text{A8.28})$$

where  $m$  is the period number. Thus, for example,  $X_N = X_0 = \sum_{n=0}^{N-1} x_n$ .

A very useful concept is to think of  $x_n$  and  $X_k$  as lying on a circle instead of on a line. Most of the well-known theorems in Fourier transform theory have counterparts in DFT theory where the data lie on a circle. The shift theorem of Fourier transforms illustrates the circular nature of the DFT. The DFT of a circularly shifted, by one step, sequence of  $x_n$ , i.e.,  $y_n = x_{n-1}$ , or

$$y = \{x_{N-1}, x_0, x_1, x_2, \dots, x_{N-2}\} , \quad (\text{A8.29})$$

is

$$Y_k = \sum_{n=0}^{N-1} y_n e^{-j2\pi kn/N} , \quad (\text{A8.30})$$

$$= \sum_{n=1}^{N-1} x_{n-1} e^{-j2\pi kn/N} + x_{N-1} , \quad (\text{A8.31})$$

$$= \sum_{n=0}^{N-2} x_n e^{-j2\pi k(n+1)/N} + x_{N-1} , \quad (\text{A8.32})$$



$$= e^{-j2\pi k/N} \sum_{n=0}^{N-2} x_n e^{-j2\pi kn/N} + x_{N-1} , \quad (\text{A8.33})$$

$$= e^{-j2\pi k/N} \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} . \quad (\text{A8.34})$$

The last step of absorbing  $x_{N-1}$  into the summation was accomplished by recognizing that the  $x_{N-1}$  term of the summation in Eq. (A8.34) is

$$e^{-j2\pi k/N} x_{N-1} e^{-j2\pi k(N-1)/N} = x_{N-1} . \quad (\text{A8.35})$$

Thus,

$$Y_k = e^{-j2\pi k/N} X_k . \quad (\text{A8.36})$$

In general, for a shift of  $\ell$  steps,

$$y_n = x_{n-\ell} , \quad (\text{A8.37})$$

the DFT becomes

$$Y_k = e^{-j2\pi \ell k/N} X_k . \quad (\text{A8.38})$$

Hence, the shift theorem for DFT is clearly based on a circular shift of  $x_n$ . It is straightforward to prove the cyclic convolution and correlation theorems:

$$x_n * y_n \xleftrightarrow{\text{DFT}} \frac{1}{N} X_k Y_k , \quad (\text{A8.39})$$

$$x_n \star y_n \xleftrightarrow{\text{DFT}} \frac{1}{N} X_k Y_k . \quad (\text{A8.40})$$

It is important to understand that to use expressions (A8.39) and (A8.40) to calculate either convolution or correlation function, it is necessary to pad the spectral array with  $N$  zeros to avoid unwanted products in the circular correlation (see Sect. A8.4.2).

Parseval's theorem for the DFT can be easily proved, by using Eq. (A8.19) and writing

$$\sum_{k=0}^{N-1} X_k X_k^* = \sum_{k=0}^{N-1} \left[ \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N} \right] \left[ \sum_{\ell=0}^{N-1} x_\ell e^{j2\pi k\ell/N} \right] . \quad (\text{A8.41})$$

Interchanging the order of summations gives

$$\sum_{k=0}^{N-1} X_k X_k^* = \sum_{n=0}^{N-1} \sum_{\ell=0}^{N-1} x_n x_\ell^* \sum_{k=0}^{N-1} e^{-j2\pi k(n-\ell)/N} . \quad (\text{A8.42})$$

The rightmost sum is proportional to the Kronecker delta function [Eq. (A8.25)], so

$$\sum_{k=0}^{N-1} X_k X_k^* = N \sum_{n=0}^{N-1} x_n^2 . \quad (\text{A8.43})$$

If  $x_n$  is complex, then the general form of Eq. (A8.43) becomes<sup>9</sup>

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2 . \quad (\text{A8.44})$$

#### A8.4.1 Response to a Complex Sine Wave

We now calculate the DFT response to a complex sine wave,

$$x_n = e^{j2\pi \nu t_n} , \quad t_n = \frac{n}{2\Delta\nu} . \quad (\text{A8.45})$$

We introduce a normalizing frequency  $\nu' = \frac{\nu}{\Delta\nu} \frac{N}{2}$ . For the positive frequency range of 0 to  $\Delta\nu$ ,  $\nu'$  ranges from 0 to  $N/2$ . Note that  $\nu'$  does not have to be an integer. The DFT of  $x_n$  is

$$X_k = \sum_{n=0}^{N-1} e^{-j2\pi n(k-\nu')/N} . \quad (\text{A8.46})$$

We use the formula for the sum of a geometric series,

$$\sum_{n=0}^{N-1} y^n = \frac{1 - y^N}{1 - y} , \quad (\text{A8.47})$$

---

<sup>9</sup>In some DFT formulations, the  $N$  factor of Eq. (A8.22) appears in Eq. (A8.19). In that case, the  $N$  factor in Eq. (A8.44) moves to the numerator on the left side.

where  $y = e^{-j2\pi(k-v')/N}$ , to write Eq. (A8.46) as

$$X_k = \frac{1 - e^{j2\pi(k-v')}}{1 - e^{j2\pi(k-v')/N}}. \quad (\text{A8.48})$$

By factoring out  $e^{j\pi(k-v')}$  from the numerator and  $e^{j\pi(k-v')/N}$  from the denominator of Eq. (A8.48), we can write

$$X_k = \left[ \frac{e^{j\pi(k-v')}}{e^{j\pi(k-v')/N}} \right] \left[ \frac{\sin \pi(k-v')}{\sin(\pi(k-v')/N)} \right]. \quad (\text{A8.49})$$

We are interested in the power response

$$S_k = |X_k|^2 = \left[ \frac{\sin \pi(k-v')}{\sin(\pi(k-v')/N)} \right]^2. \quad (\text{A8.50})$$

This is the circular form of the sinc function, which repeats on the interval  $N$ , that is,  $X_{k+N} = X_k$ , or  $S_{k+N} = S_k$ .

We approximate the denominator of Eq. (A8.50) as  $\pi(k-v')/N$ , so that

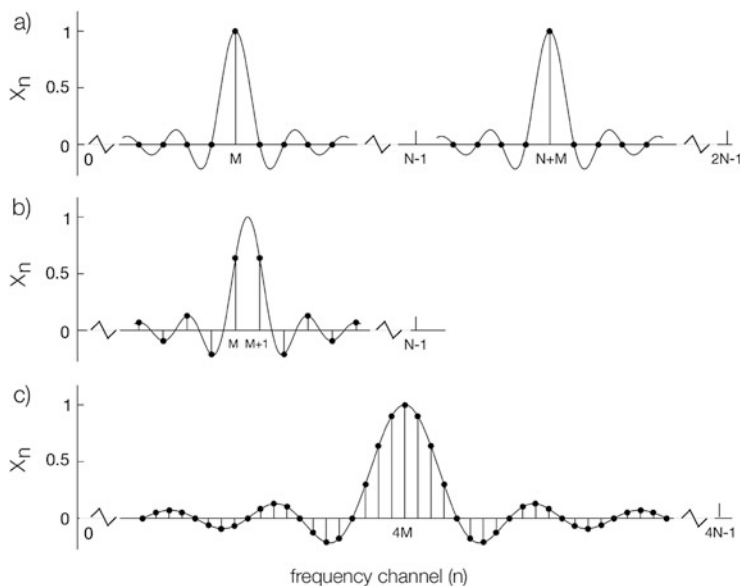
$$S_k = |X|^2 \simeq N \text{sinc}^2(k-v'). \quad (\text{A8.51})$$

If  $v' = m$ , an integer, then

$$\begin{aligned} S_k &= N, & k &= m, \\ &= 0, & k &\neq m. \end{aligned} \quad (\text{A8.52})$$

In Fig. A8.1, we show the response to the complex sine wave for  $v' = m$  and  $v' = m + 1/2$ . Unless  $v'$  corresponds exactly to a DFT channel,  $S_k$  will be nonzero in every channel, demonstrating the problem of spectral leakage.

The DFT of Eq. (A8.50) gives the corresponding response for the correlation function, which is a triangle function. This reflects the fact that the number of ways the correlation function, given by Eq. (A8.39), can be computed from a segment of data decreases linearly from  $N$  ways for lag zero to one way for lag  $N-1$ , as shown in Fig. 8.23. If it is desired to calculate the correlation function from the power spectrum, i.e., Eq. (A8.39), it is important to note that the spectrum must be padded with zeros to length  $2N$ .



**Fig. A8.1** (a) The response of the DFT of length  $N$  to a complex sine wave of frequency  $\nu' = m$ . This is a plot of Eq. (A8.49) without the phase factor. The continuous envelope [e.g., as calculated from Eq. (A8.54)] is shown along with the function values at the sample points. One repetition of this periodic function is shown. (b) Response where  $\nu' = m + 1/2$  (the frequency falls midway between two DFT channels). (c) The data array size has been increased from  $N$  to  $4N$  by padding with zeros, which results in a more finely defined spectrum.

### A8.4.2 Padding with Zeros

Padding with zeros is a very important concept in DFT theory. Padding with zeros means adding a block of zeros to the data sequence, usually at the end, to increase its length from  $N$  to  $N'$ . The three main reasons to pad with zeros are:

1. It provides a way to interpolate  $X_k$  and define it at more finely spaced frequency intervals.
2. If  $N$  is not a power of two,  $x_n$  can be padded with zeros to  $N'$ , a power of two. This makes the FFT used to calculate the DFT more computationally efficient. The  $N'$  spectrum is a properly interpolated spectrum in the Nyquist sense.
3. If a linear correlation function of two functions is to be properly calculated from  $X_n Y_n$  via the circular convolution theorem [Eq. (A8.39)], then  $X_n$  and  $Y_n$  must be padded with zeros to  $2N$  to avoid unwanted multiplications.

To understand how interpolation is achieved, consider a data sequence of length  $N$  to which  $M$  zeros are added, giving a sequence of length  $N' = N + M$ . The DFT of the new data set is

$$\begin{aligned} X_k &= \sum_{n=0}^{N-1} x_n e^{-j2\pi kn/N'} + \sum_{n=N}^{N'-1} 0 \cdot e^{-j2\pi kn/N'} \\ &= \sum_{n=0}^{N-1} x_n e^{j2\pi kn/Nr}, \quad 0 \leq k \leq N' - 1, \end{aligned} \quad (\text{A8.53})$$

where  $r = N'/N$ . The frequency spacing interval is now  $2\Delta\nu/Nr$ . Hence, if  $r = 2$ , the spectrum  $X_k$  gives a halfway interpolation of the unpadded version of  $X_k$ .

Padding with zeros can provide arbitrarily fine definition of  $X_k$ . However, it is often helpful to define a continuous spectrum associated with the discrete series  $x_n$  as

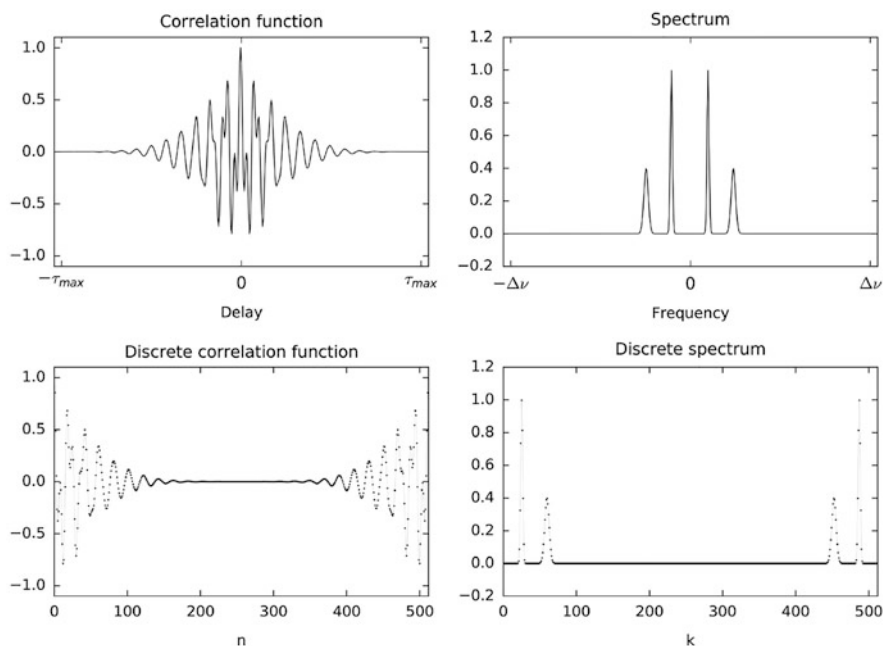
$$X(\nu) = \sum_{n=0}^{N-1} x_n e^{-j\pi n \nu'}, \quad (\text{A8.54})$$

where  $\nu' = \nu/\Delta\nu$ . This is sometimes called the discrete time Fourier transform (DTFT). It can be calculated for arbitrary values of  $\nu'$ .

It is often useful to load the FFT in such a way as to avoid unwanted phase factors from appearing in the transform. For example, suppose we want to calculate the spectrum from an autocorrelation function  $R(\tau) = R_n$ , where  $n = 0, N/2$  ranging over positive delays only. Load  $R_n$  into the array

$$\begin{aligned} R'_n &= R_n, & n &= 0, N/2, \\ R'_{N-1-n} &= R_n. \end{aligned} \quad (\text{A8.55})$$

The DFT of  $R'_n$  will have a real valued spectrum  $S_k$  in the positive frequencies  $k = 0$  to  $k = N/2$ , with the negative frequencies being in  $k = N$  to  $N - N/2 - 1$ , as shown in Fig. 8.2.



**Fig. 8.2** An example of the loading of an autocorrelation function into a DFT array. **(top)** A continuously defined autocorrelation function (left) and its power spectrum (right) via Fourier transform. **(bottom)** Positive lags loaded into 0 to  $N/2$  indices, and negative lags loaded into  $N/2$  to  $N - 1$  indices and its spectrum (right) via DFT. Loading the data in this manner gives a real power spectrum. Zero padding should be done in the middle of the delay values to keep the spectrum real valued.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



## Further Reading

- Deller, A.T., Romney, J.D., Gunaratne, T., and Carlson, B., Digital Signal Processing and Cross Correlators, in *Synthesis Imaging in Radio Astronomy III*, Mioduszewski, A., Ed., *Publ. Astron. Soc. Pacific* (2016), in press
- Harris, F.J., *Multirate Signal Processing for Communication Systems*, Prentice Hall Professional Technical Reference, Upper Saddle River, NJ (2008)

- Lyons, R.G., *Understanding Digital Signal Processing*, Addison-Wesley Longman Inc., Reading, MA (1997)
- Oppenheim, A.V., and Schaffer, R.W., *Discrete-Time Signal Processing*, Pearson Prentice Hall, Upper Saddle River, NJ (2009)

## References

- Abramowitz, M., and Stegun, I.A., *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC (1968)
- Ball, J.A., The Harvard Minicorrelator, *IEEE Trans. Instrum. Meas.*, **IM-22**, 193 (1973)
- Bellanger, M.G., Bonnerot, G., and Coudreuse, M., Digital Filtering by Polyphase Network: Application to Sample-Rate Alteration and Filter Banks, *IEEE Trans. Acoust., Speech, and Signal Proc.*, **ASSP-24**, 109–114 (1976)
- Benkevitch, L.V., Rogers, A.E.E., Lonsdale, C.J., Cappallo, R.J., Oberoi, D., Erickson, P.J., and Baker, K.D., Van Vleck Correction Generalization for Correlators with Multilevel Quantization, arXiv: 1608.04367v1
- Benson, J.M., The VLBA Correlator, in *Very Long Baseline Interferometry and the VLA*, Zensus, J.A., Diamond, P.J., and Napier, P.J., Eds., Astron. Soc. Pacific Conf. Ser., **82**, 117–131 (1995)
- Blackman, R.B., and Tukey, J.W., *The Measurement of Power Spectra*, Dover, New York (1959)
- Bowers, F.K., and Klingler, R.J., Quantization Noise of Correlation Spectrometers, *Astron. Astrophys. Suppl.*, **15**, 373–380 (1974)
- Bowers, F.K., Whyte, D.A., Landecker, T.L., and Klingler, R.J., A Digital Correlation Spectrometer Employing Multiple-Level Quantization, *Proc. IEEE*, **61**, 1339–1343 (1973)
- Bracewell, R.N., *The Fourier Transform and Its Applications*, McGraw-Hill, New York (2000) (earlier eds. 1965, 1978)
- Bunton, J.D., An Improved FX Correlator, ALMA Memo 342 (2000)
- Bunton, J.D., Multi-Resolution FX Correlator, ALMA Memo 447 (2003)
- Bunton, J.D., New Generation Correlators, in Proceedings of the 28th URSI General Assembly in New Delhi (2005). [http://www.ursi.org/Proceedings/Proc/GA05/pdf/J06.6\(0311\).pdf](http://www.ursi.org/Proceedings/Proc/GA05/pdf/J06.6(0311).pdf)
- Burns, W.R., and Yao, S.S., Clipping Noise Loss in the One-Bit Autocorrelation Spectral Line Receiver, *Radio Sci.*, **4**, 431–436 (1969)
- Carlson, B.R., and Dewdney, P.E., Efficient Wideband Digital Correlation, *Electronics Lett.*, **36**, 987–988 (2000)
- Chennamangalam, J., The Polyphase Filter Bank Technique, CASPER Memo 41 (2014). <https://casper.berkeley.edu/wiki/Memos> (no. 42 in list but labeled no. 41 in the memo)
- Chikada, Y., Ishiguro, M., Hirabayashi, H., Morimoto, M., Morita, K., Miyazawa, K., Nagane, K., Murata, K., Tojo, A., Inoue, S., Kanzawa, T., and Iwashita, H., A Digital FFT Spectro-Correlator for Radio Astronomy, in *Indirect Imaging*, Roberts, J.A., Ed., Cambridge Univ. Press, Cambridge, UK (1984), pp. 387–404
- Chikada, Y., Ishiguro, M., Hirabayashi, H., Morimoto, M., Morita, K.I., Kanzawa, T., Iwashita, H., Nakazimi, K., Ishiwaka, S.I., Takashi, T., and seven coauthors, A  $6 \times 320$ -MHz 1024-Channel FFT Cross Spectrum Analyzer for Radio Astronomy, *Proc. IEEE*, **75**, 1203–1210 (1987)
- Cole, T., Finite Sample Correlations of Quantized Gaussians, *Aust. J. Phys.*, **21**, 273–282 (1968)
- Cooley, J.W., and Tukey, J.W., An Algorithm for the Machine Calculation of Complex Fourier Series, *Math. Comp.*, **19**, 297–301 (1965)
- Cooper, B.F.C., Correlators with Two-Bit Quantization, *Aust. J. Phys.*, **23**, 521–527 (1970)
- Crochiere, R.E., and Rabiner, L.R., Interpolation and Decimation of Digital Signals—A Tutorial Review, *Proc. IEEE*, **69**, 300–331 (1981)
- D’Addario, L.R., Cross Correlators, in *Synthesis Imaging in Radio Astronomy*, Perley, R.A., Schwab, F.R., and Bridle, A.H., Eds., Astron. Soc. Pacific Conf. Ser., **6**, 59–82 (1989)

- D'Addario, L.R., Thompson, A.R., Schwab, F.R., and Granlund, J., Complex Cross Correlators with Three-Level Quantization: Design Tolerances, *Radio Sci.*, **19**, 931–945 (1984)
- Davis, W.F., Real-Time Compensation for Autocorrelation Clipper Bias, *Astron. Astrophys. Suppl.*, **15**, 381–382 (1974)
- Deller, A.T., Tingay, S.J., Bailes, M., and West, C., DiFX: A Software Correlator for Very Long Baseline Interferometry Using Multiprocessor Computing Environments, *Publ. Astron. Soc. Pacific*, **119**, 318–336 (2007)
- Deller, A.T., Romney, J.D., Gunaratne, T., and Carlson, B., Digital Signal Processing and Cross Correlators, in *Synthesis Imaging in Radio Astronomy III*, Mioduszewski, A., Ed., *Publ. Astron. Soc. Pacific* (2016), in press
- de Vos, M., Gunst, A.W., and Nijboer, R., The LOFAR Telescope: System Architecture and Signal Processing, *Proc. IEEE*, **97**, 1431–1437 (2009)
- Ellingson, S.W., Clarke, T.E., Cohen, A., Craig, J., Kassim, N.E., Pihlström, Y., Rickard, L.J., and Taylor, G.B., The Long Wavelength Array, *Proc. IEEE*, **97**, 1421–1430 (2009)
- Escoffier, R.P., The MMA Correlator, MMA Memo 166, National Radio Astronomy Observatory (1997)
- Escoffier, R.P., Webber, J.C., D'Addario, L.R., and Broadwell, C.M., A Wideband Digital Filter Using FPGAs, in *Radio Telescopes*, *Proc. SPIE*, **4015**, 106–113 (2000)
- Escoffier, R.P., Comoretto, G., Webber, J.C., Baudry, A., Broadwell, C.M., Greenberg, J.H., Treacy, R.R., Cais, P., Quertier, B., Camino, P., Bos, A., and Gunst, A.W., The ALMA Correlator, *Astron. Astrophys.*, **462**, 801–810 (2007)
- Gary, D.E., Digital Cross-Correlators, Radio Astronomy Lecture 8, New Jersey's Science and Technology University, fall (2014). <http://web.njit.edu/~gary/728/Lecture8.html>
- Goldstein, R.M., A Technique for the Measurement of the Power Spectra of Very Weak Signals, *IRE Trans. Space Electron. Telem.*, **8**, 170–173 (1962)
- Granlund, J., O'Sullivan's Zero-Padding, VLBA Correlator Memo 66, National Radio Astronomy Observatory (1986)
- Guilloteau, S., Delannoy, J., Downes, D., Greve, A., Guélin, M., Lucas, R., Morris, D., Radford, S.J.E., Wink, J., Cernicharo, J., and seven coauthors, The IRAM Interferometer on Plateau de Bure, *Astron. Astrophys.*, **262**, 624–633 (1992)
- Gwinn, C.R., Correlation Statistics of Quantized Noiselike Signals, *Publ. Astron. Soc. Pacific*, **116**, 84–96 (2004)
- Hagen, J.B., and Farley, D.T., Digital Correlation Techniques in Radio Science, *Radio Sci.*, **8**, 775–784 (1973)
- Harris, F., Tutorial on Polyphase Transforms (1999). <http://ubm.io/1XzFUbY>
- Harris, F., Dick, C., and Rice, M., Digital Receivers and Transmitters Using Polyphase Filter Banks for Wireless Communications, *IEEE Trans. Microwave Theory Tech.*, **51**, 1395–1412 (2003)
- Harris, F.J., The Use of Windows for Harmonic Analysis with the Discrete Fourier Transform, *Proc. IEEE*, **66**, 51–83 (1978)
- Ho, P.T.P., Moran, J.M., and Lo, K.-Y., The Submillimeter Array, *Astrophys. J. Lett.*, **616**, L1–L6 (2004)
- Jan, M.T., Main Technical Features of the WideX Correlator, 2008. <http://www.iram.fr/IRAMFR/TA/backend/WideX>
- Janet, F.A., and Anderson, S.B., The Effects of Digitization on Nonstationary Stochastic Signals with Applications to Pulsar Signal Baseband Recording, *Publ. Astron. Soc. Pacific*, **110**, 1467–1478 (1998)
- Johnson, M.D., Chou, H.H., and Gwinn, C.R., Optimal Correlation Estimators for Quantized Signals, *Astrophys. J.*, **765**:135 (7pp) (2013)
- Kulkarni, S.R., and Heiles, C., How to Obtain the True Correlation from a Three-Level Digital Correlator, *Astron. J.*, **85**, 1413–1420 (1980)
- Lawson, J.L., and Uhlenbeck, G.E., *Threshold Signals*, McGraw-Hill, New York (1950). Reprinted by Dover Publications (p. 58), New York, 1965



- Lo, W.F., Dewdney, P.E., Landecker, T.L., Routledge, D., and Vaneldik, J.F., A Cross-Correlation Receiver for Radio Astronomy Employing Quadrature Channel Generation by Computed Hilbert Transform, *Radio Sci.*, **19**, 1413–1421 (1984)
- Moran, J.M., Very Long Baseline Interferometer Systems, in *Methods of Experimental Physics*, Vol. 12, Part C (*Astrophysics: Radio Observations*), Meeks, M.L., Ed., Academic Press, New York (1976), pp. 174–197
- Napier, P.J., Thompson, A.R., and Ekers, R.D., The Very Large Array; Design and Performance of a Modern Synthesis Radio Telescope, *Proc. IEEE*, **71**, 1295–1320 (1983)
- Nyquist, H., Certain Topics in Telegraph Transmission Theory, *Trans. Am. Inst. Elect. Eng.*, **47**, 617–644 (1928)
- Okumura, S.K., Momose, M., Kawaguchi, N., Kanzawa, T., Tsutsumi, T., Tanaka, A., Ichikawa, T., Suzuki, T., Ozeki, K., Natori, K., and Hashimoto, T., 1-GHz Bandwidth Digital Spectro-Correlator System for the Nobeyama Millimeter Array, *Publ. Astron. Soc. Japan*, **52**, 393–400 (2000)
- Okumura, S.K., Chikada, Y., Momose, M., and Iguchi, S., Feasibility Study of the Enhanced Correlator for Three-Way ALMA, ALMA Memo 350 (2001)
- Oppenheim, A.V., and Schaffer, R.W., *Digital-Time Signal Processing*, 3rd ed., Pearson, Prentice Hall, Upper Saddle River, NJ (2009)
- O'Sullivan, J.D., *Efficient Digital Spectrometers—A Survey of Possibilities*, Note 375, Netherlands Foundation for Radio Astronomy, Dwingeloo (1982)
- Parsons, A., Backer, D., Siemion, A., Chen, H., Werthimer, D., Droz, P., Filiba, T., Manley, J., McMahon, P., Paarsa, A., MacMahon, D., and Wright, M., A Stable Correlator Architecture Based on Modular FPGA Hardware, Reusable Gateway, and Data Packetization, *Publ. Astron. Soc. Pacific*, **120**, 1207–1221 (2008)
- Percival, B.D., and Walden, A.T., *Spectral Analysis for Physical Applications*, Cambridge Univ. Press, Cambridge, UK (1993), p. 289
- Perley, R., Napier, P., Jackson, J., Butler, B., Carlson, B., Fort, D., Dewdney, P., Clark, B., Hayward, R., Durand, S., Revnell, M., and McKinnon, M., The Expanded Very Large Array, *Proc. IEEE*, **97**, 1448–1462 (2009)
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P., *Numerical Recipes*, 2nd ed., Cambridge Univ. Press, Cambridge, UK (1992)
- Price, R., A Useful Theorem for Nonlinear Devices Having Gaussian Inputs, *IRE Trans. Inf. Theory*, **IT-4**, 69–72 (1958)
- Romney, J.D., Theory of Correlation in VLBI, in *Very Long Baseline Interferometry and the VLA*, Zensus, J.A., Diamond, P.J., and Napier, P.J., Eds., Astron. Soc. Pacific Conf. Ser., **82**, 17–37 (1995)
- Romney, J.D., Cross Correlators, in *Synthesis Imaging in Radio Astronomy II*, Taylor, G.B., Carilli, C.L., and Perley, R.A., Eds., Astron. Soc. Pacific Conf. Ser., **180**, 57–78 (1999)
- Schwab, F.R., Two-Bit Correlators: Miscellaneous Results, VLBA Correlator Memo 75, National Radio Astronomy Observatory (1986)
- Shannon, C.E., Communication in the Presence of Noise, *Proc. IRE*, **37**, 10–21 (1949)
- Thompson, A.R., Quantization Efficiency for Eight or More Sampling Levels, MMA Memo 220, National Radio Astronomy Observatory (1998)
- Thompson, A.R., Emerson, D.T., and Schwab, F.R., Convenient Formulas for Quantization Efficiency, *Radio Sci.*, **42**, RS3022 (2007)
- Vaidyanathan, P.P., Multirate Digital Filters, Filter Banks, Polyphase Networks, and Applications: A Tutorial, *Proc. IEEE*, **78**, 56–93 (1990)
- Van Vleck, J.H., The Spectrum of Clipped Noise, Radio Research Lab., Harvard Univ., Report RRL-51 (July 21, 1943)
- Van Vleck, J.H., and Middleton, D., The Spectrum of Clipped Noise, *Proc. IEEE*, **54**, 2–19 (1966)
- Wall, J.V., and Jenkins, C.R., *Practical Statistics for Astronomers*, 2nd ed., Cambridge Univ. Press, Cambridge, UK (2012)
- Weinreb, S., *A Digital Spectral Analysis Technique and Its Application to Radio Astronomy*, Technical Report 412, Research Lab. for Electronics, MIT, Cambridge, MA (1963)

- Weinreb, S., Analog-Filter Digital-Correlator Hybrid Spectrometer, *IEEE Trans. Instrum. Meas.*, **IM-34**, 670–675 (1984)
- Welch, P.D., The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified Periodograms, *IEEE Trans. Audio Electroacoust.*, **AU-15**, 70–73 (1967)
- Willis, A.G., and Bregman, J.D., Effects in Fourier Transformed Spectra, in *User's Manual for Westerbork Synthesis Radio Telescope*, Netherlands Foundation for Radio Astronomy, Westerbork, the Netherlands (1981), ch. 2, app. 2
- Wootten, A., and Thompson, A.R., The Atacama Large Millimeter/Submillimeter Array, *Proc. IEEE*, **97**, 1463–1471 (2009)
- Yen, J.L., The Role of Fast Fourier Transform Computers in Astronomy, *Astron. Astrophys. Suppl.*, **15**, 483–484 (1974)