Ronney B. Panerai
University of Leicester

A. Roberto Ambrosini
Institute of Radioastronomy

C. Bortolotti
Institute of Radioastronomy

N. D'Amico
Institute of Radioastronomy

G. Grueff
Institute of Radioastronomy

S. Mariotti
Institute of Radioastronomy

S. Montebugnoli
Institute of Radioastronomy

A. Orfei
Institute of Radioastronomy

G. Tomassetti
University of L'Aquila

23
Spectrum Analysis and Correlation

23.1   FFT Spectrum Analysis and Correlation

Ronney B. Panerai

Most sensors and instruments described in previous sections of this handbook can produce continuous measurements in time or sequential measurements at fixed or variable time intervals, as represented in Figure 23.1. The temporal patterns resulting from such measurements are usually referred to as *signals.* Signals can either be *continuous* or *discrete in time* (Figure 23.1). The main objective of *spectral analysis* is to provide an estimate of the distribution of signal power at different frequencies. Spectral analysis and correlation techniques are an aid to the interpretation of signals and to the systems that generate them. These methods are now widely used for the analysis and interpretation of measurements performed in medicine, geophysics, vibration analysis, communications, and several other areas.

Although the original concept of a **signal** involves measurements as a function of time (Figure 23.1), this term has been generalized to include measurements along other dimensions, e.g., distance. In addition, signals can have multiple dimensions — the instantaneous velocity of an airplane can be regarded as a four-dimensional signal since it depends on time and three spatial coordinates.

With the growing availability of signal-processing computer packages and dedicated instruments, most readers will perform spectral analysis and correlation at the "touch of a button," visualizing results on a screen or as a computer plot. These "black-box" systems are useful for saving time and money, but users should be aware of the limitations of the fundamental techniques and circumstances in which inappropriate use can lead to misleading results. This chapter presents the basic concepts of spectral analysis and correlation based on the **fast Fourier transform** (**FFT**) approach. FFT algorithms allow the most efficient computer implementation of methods to perform spectral analysis and correlation and have become the most popular option. Nevertheless, other approaches, such as parametric techniques, wavelet transforms, and time-frequency analysis are also available. These will be briefly discussed and the interested reader will be directed to the pertinent literature for applications that might benefit from alternative approaches.
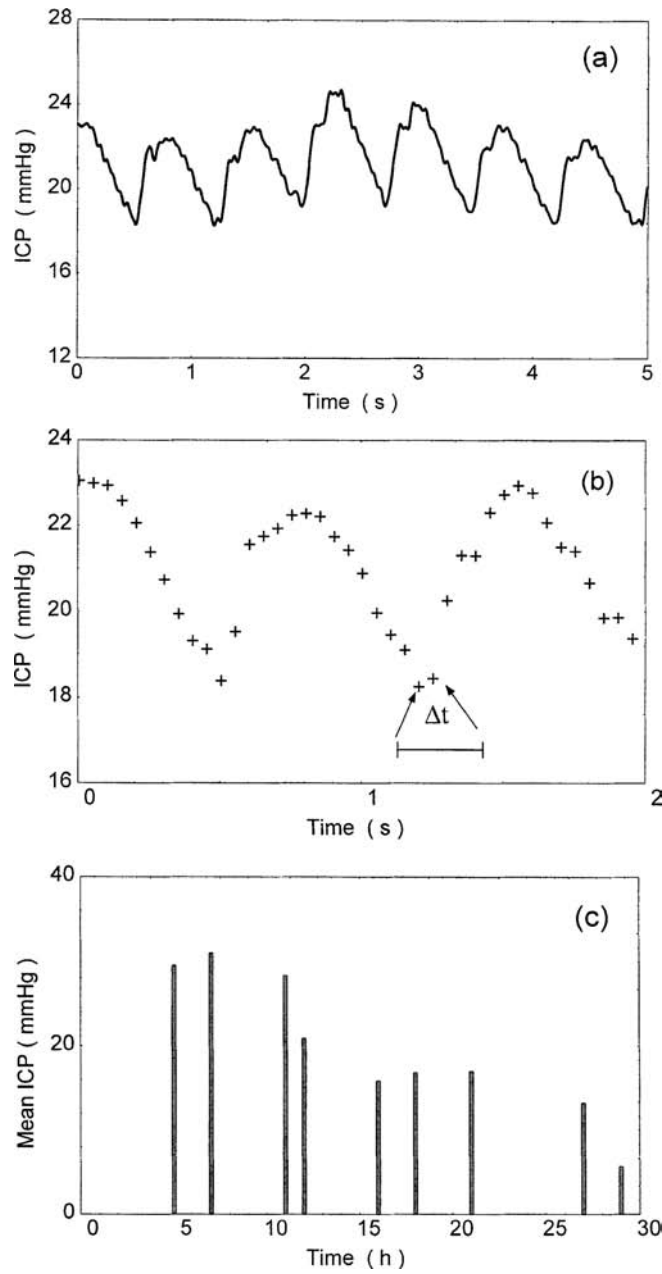
FIGURE 23.1 Examples of continuous and discrete-time signals. (a) Continuous recording of intracranial pressure in a head-injured patient. (b) Intracranial pressure measurements obtained at regular intervals of 50 ms. (c) Non-uniformly spaced measurements of mean intracranial pressure over a period of 30 h following surgery.

Fundamental Concepts

Spectral Analysis

Practical applications of spectral and correlation analysis are performed on discrete-time signals (Figure 23.1). These are obtained either from a sequence of discrete measurements or from the transformation of a continuous signal (Figure 23.1) to digital format using an **analog-to-digital converter (ADC).** When the latter is adopted to allow computer analysis of an originally continuous signal, two main

characteristics of the ADC need to be considered. The first is the number of bits available to represent each sample, as this will determine the resolution and accuracy of the sampled signal. The second important consideration is the *sampling interval* Δt (Figure 23.1). From the *Nyquist theorem*,[1] the maximum value of Δt must be such that the *sampling frequency* $f_s = 1/\Delta t$ is at least twice the highest frequency of interest in the original signal. If this rule is not followed, spectral and correlation estimations might be considerably distorted by a phenomenon called *aliasing*.[2] Low-pass filtering before ADC is always recommended to limit the bandwidth of the continuous signal to allow the correct choice of f_s or Δt. In practice, the sampling frequency is usually much higher than the minimum required by the Nyquist theorem to provide a better visual representation of the sampled data.

Let x_n represent a discrete-time signal with samples at $n = 0, 1, 2, \ldots, N-1$. The Fourier theorem[1,2] states that it is possible to decompose x_n as a sum of cosine and sine waveforms of different frequencies using an appropriate combination of amplitude coefficients. Therefore,

$$x_n = a_0 + \sum_{k=1}^{N-1} a_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^{N-1} b_k \sin\left(\frac{2\pi kn}{N}\right) \tag{23.1}$$

where $k = 1, 2, \ldots, N-1$ determines the frequency of each cosine and sine waveforms as $f_k = k/N\Delta t$. The corresponding coefficients are calculated from

$$a_0 = \frac{1}{N} \sum_{n=0}^{N-1} x_n \tag{23.2a}$$

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n \cos\left(\frac{2\pi kn}{N}\right) \tag{23.2b}$$

$$b_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n \sin\left(\frac{2\pi kn}{N}\right) \tag{23.2c}$$

Note that Equation 23.2a represents the mean value of x_n and that the argument $2\pi kn/N$ is the same for the *direct* (Equation 23.2) and *inverse* (Equation 23.1) **discrete Fourier transforms (DFT).**

From Euler's formula,[3] it is possible to combine the cosine and sine terms to express the DFT in exponential form:

$$e^{j\theta} = \cos\theta + j\sin\theta \tag{23.3}$$

leading to

$$x_n = \sum_{k=0}^{N-1} c_k e^{j(2\pi kn/N)} \tag{23.4}$$

with

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-j(2\pi kn/N)} \tag{23.5}$$

where c_k is now a complex value related to the original cosine and sine coefficients by

$$c_0 = a_0 \tag{23.6a}$$
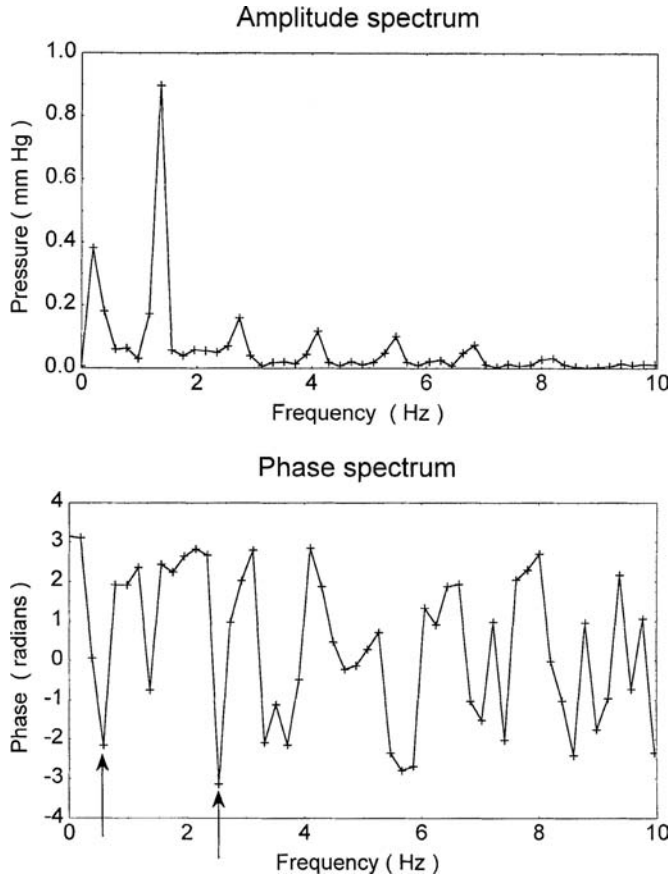
Amplitude spectrum

Phase spectrum

FIGURE 23.2 Amplitude and phase spectra of the intracranial pressure signal represented in Figure 23.1a after analog-to-digital conversion with a sampling interval of 20 ms. The main peak in the amplitude spectrum corresponds to the frequency of the cardiac cycle in Figure 23.1a. Wraparound of the phase spectrum is apparent in the third and 13th harmonics (arrows). Both spectra have been plotted to 10 Hz only.

$$c_k = a_k - jb_k \quad k = 1, 2, \ldots, N-1 \tag{23.6b}$$

A graphic representation of the a_k, b_k, or c_k coefficients for each value of k (or f_k) constitutes the frequency spectrum of x_n, expressing the relative contribution of different sinusoidal frequencies to the composition of x_n (Equation 23.4). Since c_k is complex (Equation 23.6b), a more meaningful physical interpretation of the spectrum is obtained with the *amplitude* and *phase* spectra, defined as

$$A_k = \left(a_k^2 + b_k^2\right)^{1/2} = |c_k| \tag{23.7a}$$

$$\theta_k = \tan^{-1}\left(-\frac{b_k}{a_k}\right) \tag{23.7b}$$

Figure 23.2 shows the amplitude (or magnitude) and phase spectra for the signal in Figure 23.1a, sampled at intervals $\Delta t = 20$ ms. The signal was low-pass-filtered at 20 Hz before ADC. The total duration is given by $T = N\Delta t = 5$ s, corresponding to $N = 250$ samples. Before calculating the spectral coefficients, the mean value of the complete record was removed (dc term) and any linear trends were removed by fitting a straight line to the data (detrending). As will be discussed below, it is also important to apply a window

to the data, to minimize the phenomenon of leakage. For $k > N/2$ both spectra present symmetrical values. This can be easily demonstrated from the fact that cosine (and sine[2]) functions have even symmetry while sine has odd symmetry. From Equation 23.7 it follows that A_k and θ_k have *even* and *odd* symmetry, respectively.[4] Consequently, only half the spectral components ($k \leq N/2$) are required to give a complete description of x_n in the frequency domain.

The amplitude spectra indicates the combined amplitude of the cosine and sine terms to reconstruct x_n; the phase spectra reflects the relative phase differences (or time delays) between the sinusoidal waveforms to generate the temporal pattern of x_n. The amplitude spectra also reflects the signal power at different frequencies. For simplicity, the power spectrum can be defined as

$$P_k = A_k^2 = |c_k|^2 \tag{23.8}$$

Direct implementation of Equation 23.8, however, leads to spectral power estimates which are biased and inconsistent. More appropriate procedures for estimating the **power spectrum** (or *power density spectrum*) will be discussed later.

Parseval's theorem[5] demonstrates that the total signal energy can be computed either in time or frequency domain:

$$\frac{1}{N} \sum_{n=0}^{N-1} x_n^2 = \sum_{k=0}^{N-1} P_k \tag{23.9}$$

If x_n has zero mean, the left-hand side of Equation 23.9 is the biased estimator of signal variance.[6] Although most applications of spectral analysis concentrate on the characteristics of the amplitude or power spectra, it is important to bear in mind that the phase spectrum is also responsible for the temporal pattern of x_n. As an example, while both the Dirac impulse function and white noise have a flat, constant amplitude (or power) spectra,[6] it is the difference in the phase spectra which accounts for the different morphologies in the time domain.

Interpretation of the amplitude and phase spectra of both theoretical functions and sampled data is facilitated by taking into account several properties of the DFT (Equations 23.4 and 23.5), namely, *symmetry, linearity, shifting, duality,* and *convolution.*[7] To these, a very important property of Equations 23.1 and 23.4 must be added. Since cosine and sine functions are periodic, and exist for $-\infty < t < \infty$, Equations 23.1 and 23.4 will reconstruct x_n not only in the interval of interest ($0 \leq t \leq T$) but also at all other multiple intervals $pT \leq t \leq (p+1)T$ ($p = 0, \pm 1, \pm 2, \ldots$). As a consequence, spectral estimations obtained with the DFT inherently assume that x_n is *periodic* with period $T = N/\Delta t$. As discussed in the following sections, this property needs to be taken into account when performing spectral analysis with the DFT and FFT.

Correlation Analysis

The basic concept of the correlation coefficient, as a measure of the strength of linear relationship between two variables[6] can be extended to signal analysis with the definition of the **cross-correlation function (CCF)** as:[5]

$$r_{xy}(p) = \frac{1}{N} \sum_{n=0}^{N-1} x_n y_{n-p} \qquad p = 0, \pm 1, \pm 2, \ldots \tag{23.10}$$

where x_n and y_n are zero-mean, discrete-time signals defined in the interval $n = 0, 1, 2, \ldots, N-1$. For each value of p, the cross correlation is computed by shifting y_n by $p\Delta t$ and calculating the average product in Equation 23.10. If x_n and y_n are unrelated, the sum of positive and negative products will tend to zero. Conversely, if y_n tends to follow x_n, but with a time delay D, $r_{xy}(p)$ will show a peak at $p = D/\Delta t$. This property of the CCF is illustration in Figure 23.3. As noted by Bergland,[8] cross correlation can be viewed as "one signal searching to find itself in another signal."
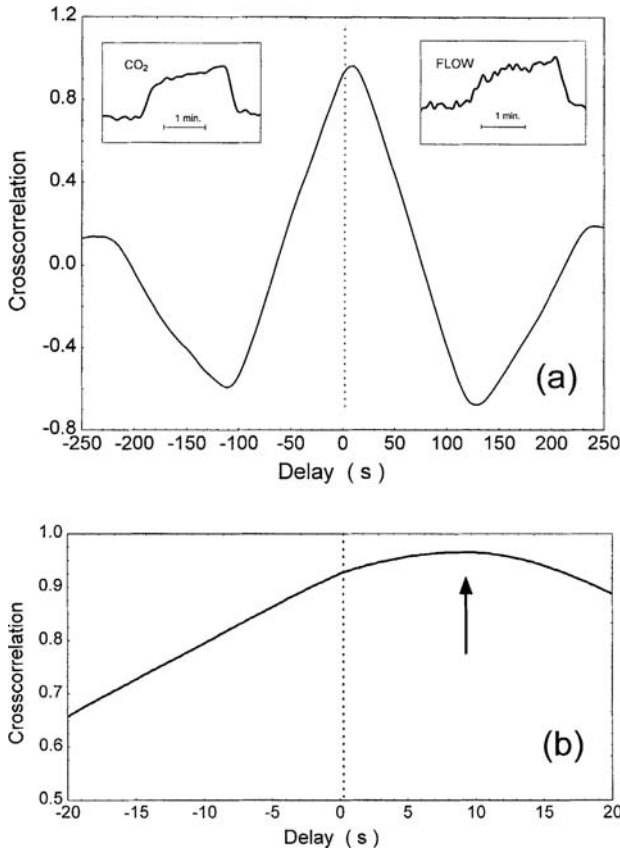
FIGURE 23.3　CCF between changes in arterial CO_2 and blood flow to the brain. Arterial CO_2 was estimated from end-tidal measurements and cerebral blood flow with Doppler ultrasound in the middle cerebral artery. (a) CCF and original signals (inserts). The cross-correlation value of approximately 1.0, observed at time delays near zero, reflects the similar temporal patterns between the two measurements. The negative cross correlations are obtained when either signal is shifted by approximately the duration of the plateau phase, which lasts 2 min. (b) Enlarging the scale around delay = 0 shows that the peak cross correlation occurs at 10 s, reflecting the time it takes for the flow to respond to the CO_2 change. (Data kindly provided by Dr. Joanne Dumville, Mr. A. Ross Naylor, and Prof. David H. Evans, University of Leicester, U.K.)

For $y_n = x_n$, $r_{xy}(p)$ becomes the **autocorrelation function (ACF):**

$$r_{xx}(p) = \frac{1}{N} \sum_{n=0}^{N-1} x_n x_{n-p}$$ (23.11)

and it is intuitive that the maximum value of $r_{xx}(p)$ occurs for $p = 0$ with

$$r_{xx}(0) = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2$$ (23.12)

which represents the signal variance or total energy. Therefore, for signals with unit standard deviation, the autocorrelation peak is equal to 1.

The Wiener–Khintchine theorem[9] demonstrates that the autocorrelation function and the power spectrum constitute a Fourier transform pair, that is,
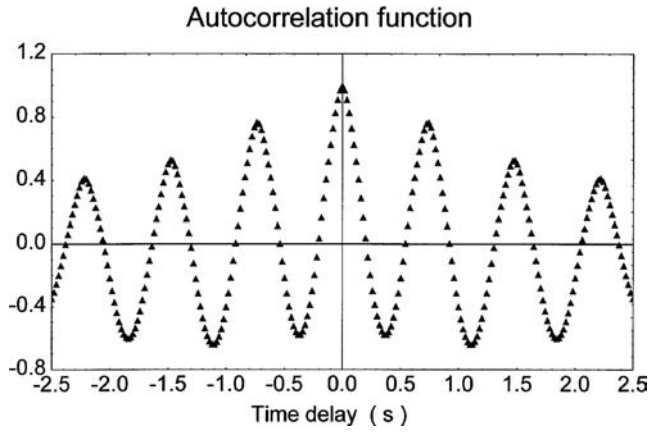
FIGURE 23.4 ACF of the discrete-time version of the signal in Figure 23.1a. The periodicity of the ACF reflects the quasi-periodic pattern of the intracranial pressure signal (Figure 23.1a).

$$S_k = \sum_{p=0}^{N-1} r_{xx}(p) e^{-j(2\pi kp/N)}$$

(23.13)

where S_k is usually called the autospectra of x_n.[6] Equation 23.13 indicates that it is possible to estimate the power spectra from a previous estimate of the autocorrelation function. As a transform pair, the autocorrelation function can also be derived from the autospectra by substituting S_k for c_k in Equation 23.4.

From Equation 23.11 it is clear that $r_{xx}(p)$ has even symmetry, that is, $r_{xx}(+p) = r_{xx}(-p)$. This property is apparent in Figure 23.4, which shows the estimated autocorrelation function for the signal in Figure 23.1a. Another characteristic of ACF, which can be visualized in Figure 23.4, is the occurrence of secondary peaks reflecting the presence of an oscillatory component in x_n (Figure 23.1a).

Fast Fourier Transform

The FFT is not a single algorithm but rather a large family of algorithms which can increase the computational efficiency of the DFT. The main ideas behind the formulation of FFT algorithms are discussed below. A detailed description of the different algorithms that have been proposed is beyond the scope of this introduction; this can be found in References 5 through 7 and 10 through 14.

For both software and hardware implementations of Equations 23.4 and 23.5, the computational efficiency is usually expressed by the number of complex multiplications and additions required or, simply, by the *number of operations*.[10] Straight implementation of either Equation 23.4 or 23.5 leads to N^2 operations. Typically, FFT algorithms can reduce this number to $N \log_2 N$. For $N = 1024$ the FFT algorithm is 100 times faster than the direct implementation of Equation 23.4 or 23.5.

The essence of all FFT algorithms is the periodicity and symmetry of the exponential term in Equations 23.4 and 23.5, and the possibility of breaking down a transform into a sum of smaller transforms for subsets of data. Since n and k are both integers, the exponential term is periodic with period N. This is commonly represented by

$$W_N = e^{-j(2\pi/N)}$$

(23.14)

and Equation 23.5 can be written as

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n W_N^{kn} \quad k = 0, 1, 2, \ldots, N-1$$

(23.15)

In many applications the terms W_N^{kn} are called **twiddle factors.** Assuming $N = 8$, calculation of the DFT with Equation 23.15 will require 64 values of W_8^{kn}. Apart from the minus sign, a simple calculation can show that there are only four different values of this coefficient, respectively: 1, j, $(1 + j)/\sqrt{2}$, and $(1 - j)/\sqrt{2}$.[4] Consequently, only these four complex factors need to be computed, representing a significant savings in number of operations.

Most FFT algorithms are based on the principle of **decimation-in-time,** involving the decomposition of the original time (or frequency) sequence into smaller subsequences. To understand how this decomposition can reduce the number of operations, assume that N is even. In this case it is possible to show that Equation 23.15 can be written as:[4,5,7,11]

$$c_k = \frac{1}{N} \sum_{r=0}^{(N/2)-1} x_r^e \cdot W_{N/2}^{kr} + \frac{1}{N} \sum_{r=0}^{(N/2)-1} x_r^o \cdot W_{N/2}^{kr} \qquad (23.16)$$

where x_r^e and x_r^o represent the even- and odd-order samples of x_n, respectively. Comparing Equations 23.15 and 23.16, it is clear that the latter represents two DFTs with dimension $N/2$, involving $2(N/2)^2$ operations rather than the N^2 operations required by Equation 23.15. This process of decimation-in-time can be carried out further to improve computational performance. In the general case, N can be decomposed into q factors:

$$N = \prod_{i=1}^{q} r_i = r_1 r_2 \ldots r_q \qquad (23.17)$$

The number of operations required is then:[6]

$$\text{number of operations} = N \sum_{i=1}^{q} r_i \qquad (23.18)$$

In the original algorithm of Cooley and Tukey,[10] $r_i = 2$ and $N = 2^q$. In this case the theoretical number of operations required would be $2Nq = 2N \log_2 N$. As pointed out in Reference 6, further improvements in efficiency are possible because of the symmetry of the twiddle factors. The efficiency gain of most FFT algorithms using radix-2, i.e., $N = 2^q$ is

$$\text{efficiency gain} = \frac{N^2}{N \log_2 N} = \frac{N}{\log_2 N} \qquad (23.19)$$

For $N = 1024$, $q = 10$ and the efficiency gain is approximately 100. Specific applications might benefit from other decompositions of the original sequence. Cases of particular interest are radix-4 and radix-8 FFTs.[14] However, as shown by Rabiner and Gold,[11] (p. 585), it is not possible to generalize the superiority of radix-8 over radix-4 algorithms.

In general, most FFT algorithms accept complex x_n sequences in Equation 23.5. By limiting x_n to the most common situation of real-valued signals, it is possible to obtain more efficient algorithms as demonstrated by Sorensen et al.[15] Uniyal[16] performed a comparison of different algorithms for real-valued sequences showing that performance is architecture dependent. For machines with a powerful floating point processor, the best results were obtained with Brunn's algorithm.[17]

The application of FFT algorithms for spectral and correlation analysis is discussed in the following sections.

FFT Spectral Analysis

For some deterministic signals, x_n can be expressed by a mathematical function and the amplitude and phase spectra can be calculated as an exact solution of Equations 23.5 and 23.7. The same is true for the power spectra (Equations 23.8 and 23.13). Examples of this exercise can be found in many textbooks.[1,2,4,5,7]

In most practical applications, there is a need to perform spectral analysis of experimental measurements, corresponding to signals which, in general, cannot be described by simple mathematical functions. In this case the spectra has to be estimated by a numerical solution of Equations 23.5 through 23.8, which can be efficiently implemented on a digital computer with an FFT algorithm. For estimation of the power spectrum, this approach is often classified as *nonparametric,* as opposed to other alternatives which are based on parametric modeling of the data such as *autoregressive* methods.[18] Considerable distortions can result from applications of the FFT unless attention is paid to the following characteristics and properties of the measured signal and the DFT/FFT.

Limited Observation of Signal in Time

Limited observation of a signal x_n in time can be seen as the multiplication of the original signal x_n^∞ by a rectangular window of duration $T = N\Delta t$ as exemplified for a single sinusoid in Figure 23.5. The DFT assumes that x_n is periodic, with period T, as mentioned previously. Instead of a single harmonic at the frequency of the original sinusoid, the power spectrum estimated with the FFT will have power at other harmonics as indicated by the spectrum in Figure 23.5c. The spectral power, which should have been concentrated on a single harmonic (Figure 23.5c, dashed line), has "leaked" to neighboring harmonics and for this reason this phenomenon is usually called *leakage.* The morphology of the distorted spectrum of Figure 23.5c can be explained by the fact that the Fourier transform of a rectangular window function (Figure 23.5b) is given by a *sinc* function (sin x/x) which presents decreasing side lobes.[1,2] Multiplication in time corresponds to the convolution operation in the frequency domain.[1,2] In the general case of signals comprising several harmonics, the *sinc* functions will superimpose and the resulting spectrum is then a distorted version of the "true" spectrum. As the individual *sinc* functions superimpose to produce the complete spectrum, a *picket-fence* effect is also generated.[8] This means that spectral leakage not only adds spurious power to neighboring harmonics but also restricts the frequency resolution of the main spectral peaks. The effects of spectral leakage can be reduced by (1) increasing the period of observation and (2) multiplying the original signal x_n by a window function with a smooth transition as represented by the dashed line window in Figure 23.5b. The Fourier transform of a window function with tapered ends has smaller side lobes, thus reducing the undesirable effects leakage. A large number of tapering windows have been proposed, as reviewed by Harris.[19] As an example, the four-term Blackman–Harris window, defined as

$$w_n = a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) + a_2 \cos\left(\frac{4\pi n}{N}\right) - a_3 \cos\left(\frac{6\pi n}{N}\right) \quad n = 0, 1, 2, \ldots N-1 \qquad (23.20)$$

produces side lobe levels of –92 dB if the a_i coefficients are chosen as $a_0 = 0.35875$, $a_1 = 0.48829$, $a_2 = 0.14128$, $a_3 = 0.01168$.[19] Windows also play an important role in the sampling properties of power spectral estimates, as will be discussed later. Windowing attenuates the contribution of signal samples at the beginning and end of the signal and, therefore, reduces its effective signal duration. This effect is reflected by the equivalent noise bandwidth (ENBW) defined as[19]

$$\text{ENBW} = \frac{\sum\limits_{n=0}^{N-1} w_n^2}{\left[\sum\limits_{n=0}^{N-1} w_n\right]^2} \qquad (23.21)$$

For a rectangular window ENBW = 1.0 and for the Blackman–Harris window (Equation 23.20) the corresponding value is 2.0. The majority of other window shapes have intermediate values of ENBW.[19]

Effects of "Zero-Padding"

Most FFT algorithms operate with $N = 2^q$ samples; the choice of q is many times critical. Since frequency resolution is inversely proportional to N, in many circumstances a value of q leading to $2^q > N$ is preferable
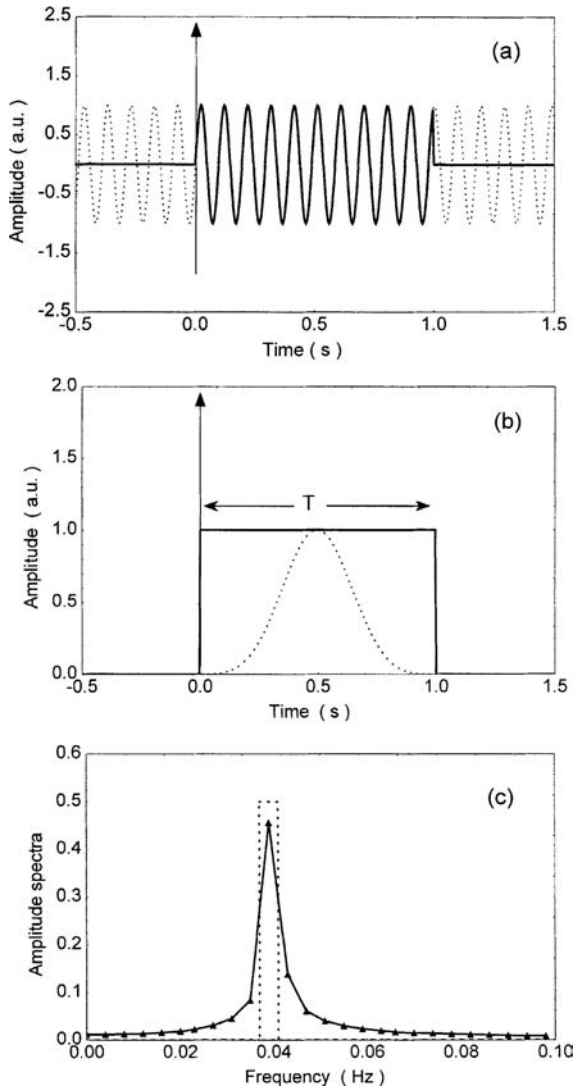
FIGURE 23.5 Effect of limited observation time T on the amplitude spectra of a sinusoidal component. (a) Observation of a single harmonic (dashed line) for a limited period of time T is equivalent to the multiplication for the rectangular function represented in (b). The Blackman–Harris window is also represented in (b) (dashed line). (c) Truncating a single harmonic produces spectral estimates smeared by *leakage* (solid line) as compared with the theoretical result (dashed line) with width equal to the frequency resolution ($f_r \approx 0.004$ Hz).

to the option of limiting the signal to $N' = 2^{q-1}$ samples with $N' < N$. The most common and simple way of extending a signal to comply with the 2^q condition is by **zero-padding.** For signals with zero mean and with first and last values around zero, this can be accomplished by complementing the signal with Q zeros to achieve the condition $N + Q = 2^q$. For signals with end points different from zero, these values can be used for padding. If initial and final values differ significantly, a linear interpolation from the last to the first point is also a practical option. However, with the application of windowing, most signals will have similar initial and final points and these can be used for zero-padding. As discussed in the next section, zero-padding has important applications for the estimation of correlation functions via FFT. For spectral analysis, it is relatively simple to demonstrate that adding Q zeros corresponds to oversampling the N point original spectrum with a new frequency resolution which is $(N + Q)/N$ times greater than

the original resolution. Consequently, although zero-padding does not introduce major distortions, it produces the false illusion of higher resolution than warranted by the available *N* measured signal samples.

Phase Spectrum Estimation

The use of Equation 23.7b to estimate the phase spectrum is fraught with a different kind of problem, resulting from the indetermination of the *tan*$^{-1}$ function to discriminate between phase angles with absolute values greater than π. This problem is illustrated in Figure 23.2b showing that phase angles decrease continuously until reaching –π and then "jump" to continue decreasing from the +π value. This feature of the phase spectrum is called **wraparound.** Methods to "unwrap" the phase spectrum have been proposed,[20] but a general satisfactory solution to this problem is not available. In some cases the shifting property of the DFT[5,7] can be used to "rotate" the original signal in time, thus minimizing the slope of the phase spectrum and, consequently, the occurrence of wraparound.

Sampling Properties of Spectral Estimators

The most straightforward approach to computing the power spectrum is to use Equation 23.8. This method is known as the **periodogram**.[5] Application is limited to signals which are **stationary,** meaning stable statistical properties (such as the mean and the variance) along time. For measurements performed on *nonstationary* systems, such as speech or systems with time-varying parameters, other methods of spectral estimation are available and will be mentioned later. It is possible to demonstrate that when the period of observation *T* tends to infinity, Equation 23.8 gives an unbiased estimate of the power spectrum. In practice, due to finite values of *T,* the phenomenon of spectral leakage described above will lead to power spectral estimates which are *biased.*

   The second inherent problem with the periodogram is the *variance* of the resulting spectral estimates. Assuming x_n to follow a Gaussian distribution, it follows that a_k and b_k will also be Gaussian because Equation 23.2 represents a linear transformation. Since Equations 23.7a and 23.8 involve the sum of two squared Gaussian variates, P_k will follow a χ^2 distribution with two degrees of freedom.[6] In this case the mean and the standard deviation of the power spectral estimate will be the same, *independently of the frequency considered.* As a consequence, power spectral estimates obtained from Equation 23.8, using a simple sample x_n, should be regarded as highly unreliable. In addition, the variance or standard deviation of this χ^2 distribution does not decrease with increases in sample duration *N*. This indicates that the periodogram (Equation 23.8) is an *inconsistent* estimator of the power spectrum.

   For a χ^2 distribution with *m* degrees of freedom, the coefficient of variation is given by

$$CV\left[\chi_m^2\right] = \frac{\sqrt{2m}}{m} = \sqrt{\frac{2}{m}} \tag{23.22}$$

showing that it is possible to improve the reliability of power spectral estimates by increasing *m*. This can be achieved by replacing Equation 23.8 by[21]

$$\hat{P}_k = \frac{1}{L}\sum_{l=1}^{L} c_{k,l}^2 \quad k = 0, 1, 2, \ldots N-1 \tag{23.23}$$

with *L* representing a number of separate samples x_n each with length $T = N\Delta t$. If only one record of x_n can be obtained under stationary conditions, it is possible to break down this record into *L* segments to obtain an improved estimate of the power spectrum with variance reduced by a factor of *L*. However, the spectral resolution, given by $f_r = 1/T$, will be reduced by the same factor *L,* thus indicating an inescapable compromise between resolution and variance.

   A *modified periodogram* was introduced by Welch[21] consisting of the multiplication of x_n by a triangular, or other window shape, before computing the individual spectral samples with Equation 23.5. The application of a window justifies overlapping adjacent segments of data by as much as 50%. For a signal

with a total duration of N samples, the combination of overlapping with segmentation (Equation 23.23) can lead to a further reduction of the spectral variance by a factor of 11/18.

Averaging L spectral samples as indicated by Equation 23.23 represents one approach to improve spectral estimation by means of *smoothing*. A similar effect can be obtained with the **correlogram**. Equation 23.13 indicates that it is possible to estimate the power spectrum from the autocorrelation function. Limiting the number of shifts of the autocorrelation function to $p \ll N$ is equivalent to smoothing the original spectrum by convolution with the Fourier transform of a Bartlett (triangular) window.[22] As discussed in the next section, the autocorrelation function can also be computed more efficiently with the FFT and it can be shown that in this case it involves a smaller number of numerical operations than the Welch method based on the periodogram.[5]

FFT Correlation Analysis

Before considering the application of FFT algorithms to compute auto- and cross-correlation functions, it is important to discuss their sampling properties using Equations 23.10 and 23.11 as estimators. Assuming that variables x_n and y_n are not defined outside the interval $0 \leq n \leq N - 1$, it follows from Equation 23.10 that as p increases and the two functions "slide" past each other, the effective number of summed products is $N - |p|$ rather than N as implied by Equations 23.10 and 23.11. For this reason these equations are often rewritten as

$$r_{xy}(p) = \frac{1}{N - |p|} \sum_{n=0}^{N-|p|-1} x_n y_{n-1} \quad p = 0, \pm 1, \pm 2, \ldots \tag{23.24}$$

The main justification for this modification, however, is that Equations 23.10 and 23.11 lead to biased estimations of correlation functions while Equation 23.24 is *unbiased*.

Equation 23.24 normally assumes that x_n and y_n are standardized variables with zero mean and unit variance. If the mean values are different from zero, Equation 23.10 and 23.11 will produce distorted estimates with a "pyramid effect" due to the presence of the dc term. However, this effect is compensated for in Equation 23.24 and in this case the effect of the mean value is to add a constant term:

$$r_{xy}(p) = r'_{xy}(p) - m_x m_y \tag{23.25}$$

where $r'_{xy}(p)$ is the cross correlation of variables with mean values m_x and m_y, respectively.

Similarly to the DFT, Equations 23.10 and 23.11 involve N^2 operations and Equation 23.24 slightly less. Since the autocorrelation and the power spectra constitute a Fourier transform pair (Equation 23.13), the computation of correlation functions can also be sped up by means of an FFT algorithm. For the sake of generality, the **cross spectrum** of x_n and y_n can be defined as[6]

$$C_{xy}(f_k) = X(f_k)Y^*(f_k) \tag{23.26}$$

with $X(f_k)$ and $Y(f_k)$ representing the Fourier transforms of x_n and y_n, respectively. The generalized Wiener–Khintchine theorem then gives the cross-correlation function as

$$r_{xy}(p) = \sum_{k=0}^{N-1} C_{xy}(f_k) e^{-j(2\pi kp/N)} \tag{23.27}$$

Therefore, "fast" correlation functions can be computed[23] using the forward FFT to calculate $X(f_k)$ and $Y(f_k)$ and then the inverse FFT to obtain $r_{xy}(p)$ with Equation 23.27. Obviously, when autocorrelation functions are being computed with this method, only one transform is necessary to obtain the autospectra (Equation 23.13) instead of the cross spectra (Equation 23.26).

When correlation functions are computed with the FFT, it is critical to pay attention again to the periodicity of the transformed variables as an intrinsic property of the DFT. When the two functions in either Equation 23.10 or 23.11 are displaced by p samples, for periodic functions there will be nonzero products outside the range $0 \leq n \leq N - 1$, thus leading to significant errors in the estimated auto- or cross-correlation functions. In this case the resulting estimates are called *circular* correlations.[6] This error can be avoided by zero-padding the original signals from $n = N$ to $n = 2N - 1$ and computing the FFTs with $2N$ points. The resulting correlation functions will be noncircular and, in the range $0 \leq p \leq N - 1$, will agree with correlations computed with the original Equations 23.10 or 23.11. Finally, to remove bias the results of Equation 23.27 should also be multiplied by $N/(N - |p|)$ to agree with Equation 23.24.

Further Information

Software for FFT special analysis is available from multiple sources. Off-the-shelf software ranges from specialized packages for digital signal processing, such as DADiSP, to statistical packages which include FFT analysis of time series. Mathematical and engineering packages such as MATLAB also include routines for FFT spectral and correlation analysis. For a review of available options see Reference 24. For readers who want to implement their own software, FFT routines can be found in Reference 4, 13 through 15, and 18. Additional references are 25 through 27.

Hardware implementations of FFT algorithms are common in areas requiring real-time spectral analysis as in the case of blood flow velocity measurement with Doppler ultrasound. For a review of hardware implementations see References 11 and 28. Developments in this area follow the pace of change in VLSI technology.[29]

One of the limitations of the FFT is the fact that frequency resolution is the inverse of the signal observation time. Improved resolution can be obtained with *parametric* methods of spectral analysis and their application is particularly relevant when only short segments of data are available or when it is necessary to discriminate between frequency harmonics which are closely spaced in the spectrum. Broadly speaking, parametric methods assume that the data follow spectral densities with a known pole-zero structure of variable complexity, characterized by a given *model order*. All-zero models correspond to the *moving average* structure while the all-pole version represents the *autoregressive* model. The general case is the autoregressive-moving average model (ARMA). For a comprehensive review of these methods see Reference 30; further information and software implementations can be found in References 18 and 27.

Nonstationary signals present a particular problem. In cases where the signal statistical properties change relatively slowly with time, it is possible to select short segments of quasi-stationary data and to use the DFT or parametric methods to estimate the spectra as mentioned previously. However, when these changes in systems parameters or statistical moments are fast in relation to the phenomenon under observation (e.g., speech or seismic data), this approach is not feasible because of the poor frequency resolution resulting from short observation times. Methods proposed to cope with signal nonstationarity often depend on the underlying cause of nonstationary behavior.[9,31] More general methods, known as *time-frequency distributions,* are now favored by most investigators.[32] The Wigner–Ville and Choi–Williams transforms are some of the more widely used of these time-frequency distributions. In each case the signal is described by a simultaneous function of time *and* frequency and hence is graphically represented by a three-dimensional plot having time and frequency as dependent variables.

A different approach to the analysis of nonstationary data is the application of *wavelets*.[33] This alternative also has advantages in the representation of fast transients and in applications requiring data compression and pattern classification. Similarly to the sine and cosine functions, which are the basis of Fourier analysis, wavelets are orthogonal functions which can be used to decompose and reconstruct signals using a finite set of coefficients obtained by a *wavelet transform* (WT). The main difference between wavelets and sinusoids, however, is that the former are limited in time. In addition, the complete orthogonal set of wavelets can be obtained simply by expansion (or compression) and scaling of a single function, known as the *mother wavelet.* Because of their limited time duration wavelets can provide a much more synthetic decomposition of fast transients, or sharp edges in image analysis, than it is possible

to obtain with the DFT. Their property of expansion/contraction of a single mother wavelet can also overcome a major limitation of the DFT, that is, to allow good frequency resolution at both low and high frequencies. For applications of the WT and commercially available software see References 34 and 35.

Defining Terms

Analog-to-digital conversion: The process of converting a continuous signal to a discrete time sequence of values usually sampled at uniform time intervals.

Autocorrelation function (ACF): A measure of longitudinal variability of a signal which can express the statistical dependence of serial samples.

Correlogram: Numerical calculation and graphical representation of the ACF or CCF.

Cross-correlation function (CCF): A measure of similarity between signals in the time domain which also allows the identification of time delays between transients.

Cross-spectrum: The complex product of the power spectra of two different signals.

Decimation-in-time: The process of breaking down a time series into subsequences to allow more efficient implementations of the FFT.

Discrete Fourier transform (DFT): The usual method to obtain the Fourier series of a discrete time signal.

Fast Fourier transform (FFT): Algorithm for the efficient computation of the DFT.

Periodogram: A family of methods to estimate the power spectrum using the DFT.

Power spectrum: The distribution of signal power as a function of frequency.

Signal: Continuous or discrete representation of a variable or measurement as a function of time or other dimension.

Stationarity: Property of signals which have statistical moments invariant with time.

Twiddle factors: Exponential term in the DFT whose periodicity allows repeated use and hence considerable savings of computation time in the FFT.

Wraparound: Overflow of phase spectral estimations above $|\pi|$ due to the uncertainty of the $\tan^{-1}$ function.

Zero-padding: Extension of a signal with zeros, constant values, or other extrapolating functions.

References

1. B.P. Lathi, *Communication Systems,* New York: John Wiley & Sons, 1968.
2. R.N. Bracewell, *The Fourier Transform and Its Applications,* Englewood Cliffs, NJ: Prentice-Hall, 1988.
3. E. Kreyszig, *Advanced Engineering Mathematics,* New York: John Wiley & Sons, 1962.
4. P.A. Lynn and W. Fuerst, *Introductory Digital Signal Processing with Computer Applications,* Chichester: John Wiley & Sons, 1989.
5. J.G. Proakis and D.G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications,* 2nd ed., New York: Macmillan, 1992.
6. J.S. Bendat and A.G. Piersol, *Random Data: Analysis and Measurement Procedures,* 2nd ed., New York: John Wiley & Sons, 1986.
7. A.V. Oppenheim and R.W. Schafer, *Discrete-Time Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1989.
8. G.D. Bergland, A guided tour of the fast Fourier transform, *IEEE Spectrum,* 6: 41–52, 1969.
9. M.B. Priestley, *Spectral Analysis and Time Series,* London: Academic Press, 1981.
10. J.W. Cooley and J.W. Tukey, An algorithm for the machine computation of complex Fourier series, *Math. Computation,* 19: 297–301, 1965.
11. L.R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing,* Englewood Cliffs, NJ: Prentice-Hall, 1975.
12. E.O. Brigham, *The Fast Fourier Transform and Its Applications,* 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 1988.
13. Digital Signal Processing Committee, *Programs for Digital Signal Processing,* New York: IEEE Press, 1979.

14. R.C. Singleton, An algorithm for computing the mixed radix fast Fourier transform, *IEEE Trans. Audio Electroacoust.,* 17: 93–103, 1969.

15. H.V. Sorensen, D.L. Jones, M.T. Heideman, and C.S. Burrus, Real-valued fast Fourier transform algorithms, *IEEE Trans. Acoust. Speech Signal Proc.,* 35: 849–863, 1987.

16. P.R. Uniyal, Transforming real-valued sequences: fast Fourier versus fast Hartley transform algorithms, *IEEE Trans. Signal Proc.,* 41: 3249–3254, 1994.

17. G. Brunn, z-Transform DFT filters and FFT, *IEEE Trans. Acoust. Speech Signal Proc.,* 26: 56–63, 1978.

18. S.M. Kay, *Modern Spectral Estimation,* Englewood Cliffs, NJ: Prentice-Hall, 1988.

19. F.J. Harris, On the use of windows for harmonic analysis with the discrete Fourier transform, *Proc. IEEE,* 66: 51–83, 1978.

20. J.M. Tribolet, A new phase unwrapping algorithm, *IEEE Trans. Acoust. Speech Signal Proc.,* 25: 170–177, 1977.

21. P.D. Welch, 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.,* 15: 70–73, 1967.

22. R.B. Blackman and J.W. Tukey, *The Measurement of Power Spectra,* New York: Dover Publications, 1958.

23. T.G. Stockham, Jr., High-speed convolution and correlation, *1966 Spring Joint Computer Conference, AFIPS Proc.,* 28: 229–233, 1966.

24. R. Braham, Math & visualization: new tools, new frontiers, *IEEE Spectrum,* 32(11): 19–36, 1995.

25. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes: The Art of Scientific Computing,* Cambridge, UK: Cambridge University Press, 1986.

26. D.M. Monro, Complex discrete fast Fourier transform, *Appl. Stat.,* 24: 153–160, 1975.

27. S.L. Marple, Jr., *Digital Spectral Analysis with Applications,* Englewood Cliffs, NJ: Prentice-Hall, 1987.

28. G.D. Bergland, Fast Fourier transform hardware implementations — an overview, *IEEE Trans. Audio Electroacoust.,* 17: 104–108, 1969.

29. E. Bidet, D. Castelain, C. Joanblanq, and P. Senn, A fast single-chip implementation of 8192-complex point FFT, *IEEE J. Solid-State Circuits,* 30: 300–305, 1995.

30. S.M. Kay and S.L. Marple, Jr., Spectrum analysis — a modern perspective, *Proc. IEEE,* 69: 1380–1419, 1981.

31. J. Leuridan, H.V. Van der Auweraer, and H. Vold, The analysis of nonstationary dynamic signals, *Sound Vibration,* 28: 14–26, 1994.

32. L. Cohen, Time-frequency distributions — a review, *Proc. IEEE,* 77: 941–981, 1989.

33. I. Daubechies, Orthonormal bases of compactly supported wavelets, *Commun. Pure Appl. Math.,* 41: 909–996, 1988.

34. A. Aldroubi and M. Unser, Eds., *Wavelets in Medicine and Biology,* Boca Raton, FL: CRC Press, 1996.

35. A. Bruce, D. Donoho, and H.Y. Gao, Wavelet analysis, *IEEE Spectrum,* 33: 26–35, 1996.

23.2   RF/Microwave Spectrum Analysis[1]

A. Ambrosini, C. Bortolotti, N. D'Amico, G. Grueff, S. Mariotti, S. Montebugnoli, A. Orfei, and G. Tomassetti

A *signal* is usually defined by a time-varying function carrying some sort of information. Such a function most often represents a time-changing electric or magnetic field, whose propagation can be in free space or in dielectric materials constrained by conductors (waveguides, coaxial cables, etc.). A signal is said to be periodic if it repeats itself exactly after a given time T called the period. The inverse of the period T, measured in seconds, is the frequency f measured in hertz (Hz).

---

[1]All figures have been reproduced courtesy of Hewlett Packard, Rohde Schwarz, Hameg, Tektronix companies, and IEEE *Microwave Measurements.*
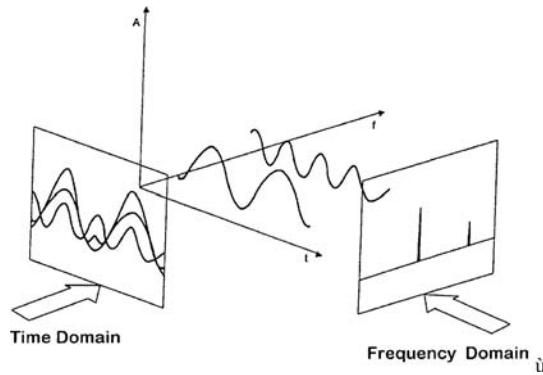
FIGURE 23.6 How the same signal can be displayed.

A periodic signal can always be represented in terms of a sum of several (possibly infinite) sinusoidal signals, with suitable amplitude and phase, and having frequencies that are integer multiples of the signal frequency. Assuming an electric signal, the square of the amplitudes of such sinusoidal signals represent the power in each sinusoid, and is said to be the power spectrum of the signal. These concepts can be generalized to a nonperiodic signal; in this case, its representation (spectrum) will include a continuous interval of frequencies, instead of a discrete distribution of integer multiples of the fundamental frequency. The representation of a signal in terms of its sinusoidal components is called Fourier analysis. The (complex) function describing the distribution of amplitudes and phases of the sinusoids composing a signal is called its Fourier transform (FT). The Fourier analysis can be readily generalized to functions of two or more variables; for instance, the FT of a function of two (spatial) variables is the starting point of many techniques of image processing. A time-dependent electrical signal can be analyzed directly as a function of time with an *oscilloscope* which is said to operate in the *time domain*. The time evolution of the signal is then displayed and evaluated on the vertical and horizontal scales of the screen.

The *spectrum analyzer* is said to operate in the *frequency domain* because it allows one to measure the harmonic content of an electric signal, that is, the power of each of its spectral components. In this case the vertical and horizontal scales read powers and frequencies. The two domains are mathematically well defined and, through the FT algorithm, it is not too difficult to switch from one response to the other. Their graphical, easily perceivable representation is shown in Figure 23.6 where the two responses are shown lying on orthogonal planes. It is trivial to say that the easiest way to make a Fourier analysis of a time-dependent signal is to have it displayed on a spectrum analyzer. Many physical processes produce (electric) signals whose nature is not deterministic, but rather stochastic, or random (noise). Such signals can also be analyzed in terms of FT, although in a statistical sense only.

A time signal is said to be band-limited if its FT is nonzero only in a finite interval of frequencies, say $(F_{max} - F_{min}) = B$. Usually, this is the case and an average frequency F_0 can be defined. Although the definition is somewhat arbitrary, a (band-limited) signal is referred to as RF (radio frequency) if F_0 is in the range 100 kHz to 1 GHz and as a microwave signal in the range 1 to 1000 GHz. The distinction is not fundamental theoretically, but it has very strong practical implications in instrumentation and spectral measuring techniques. A band-limited signal can be described further as narrowband, if $B/F_0 \ll 1$, or wideband otherwise.

The first step in performing a spectral analysis of a narrowband signal is generally the so-called heterodyne downconversion: it consists in the mixing ("beating") of the signal with a pure sinusoidal signal of frequency F_L, called local oscillator (LO). In principle, mixing two signals of frequency F_0 and F_L in any nonlinear device will result in a signal output containing the original frequencies as well as the difference $(F_0 - F_L)$ and the sum $(F_0 + F_L)$ frequencies, and all their harmonic (multiple) frequencies. In the practical case, a purely quadratic mixer is used, with an LO frequency $F_L < F_0$; the output will include the frequencies $(F_0 - F_L)$, $2F_L$, $2F_0$, and $(F_0 + F_L)$, and the first term (called the intermediate frequency

or IF) will be easily separated from the others, which have a much higher frequency. The bandwidth of the IF signal will be the same as the original bandwidth B; however, to preserve the original information fully in the IF signal, stringent limits must be imposed on the LO signal, because any deviation from a pure sinusoidal law will show up in the IF signal as added phase and amplitude noise, corrupting the original spectral content. The process of downconverting a (band-limited) signal is generally necessary to perform spectral analysis in the very high frequency (microwave) region, to convert the signal to a frequency range more easily handled technically.

When the heterodyne process is applied to a wideband signal (or whenever $F_L > F_{min}$) "negative" frequencies will appear in the IF signal. This process is called *double sideband* mixing, because a given IF bandwidth B (i.e., $(F_L + B/2)$ will include two separate bands of the original signal, centered at $F_L +$ IF ("upper" sideband) and $F_L -$ IF ("lower" sideband). This form of mixing is obviously undesirable in spectrum analysis, and input filters are generally necessary to split a wideband signal in several narrow-band signals before downconversion. Alternatively, special mixers can be used that can deliver the upper and lower sidebands to separate IF channels. A band-limited signal in the frequency interval $(F_{max} - F_{min}) = B$ is said to be converted to baseband when the LO is placed at $F_L = F_{min}$, so that the band is converted to the interval $(B–0)$. No further lowering of frequency is then possible, unless the signal is split into separate frequency bands by means of filters.

After downconversion, the techniques employed to perform power spectrum analysis vary considerably depending on the frequencies involved. At lower frequencies, it is possible to employ analog-to-digital converters (ADC) to get a discrete numerical representation of the analog signal, and the spectral analysis is then performed numerically, either by direct computation of the FT (generally via the fast Fourier transform, FFT, algorithm) or by computation of the signal autocorrelation function, which is directly related to the square modulus of the FT via the Wiener–Khinchin theorem. Considering that the ADC must sample the signal at least at the Nyquist rate (i.e., at twice the highest frequency present) and with adequate digital resolution, this process is feasible and practical only for frequencies (bandwidths) less than a few megahertz. Also, the possibility of a real-time analysis with high spectral resolution may be limited by the availability of very fast digital electronics and special-purpose computers. The digital approach is the only one that can provide extremely high spectral resolution, up to several hundred thousand channels. For high frequencies, several analog techniques are employed.

A Practical Approach to Spectrum Analysis [1]

Spectrum analysis is normally done in order to verify the harmonic content of oscillators, transmitters, frequency multipliers, etc. or the spurious components of amplifiers and mixer. Other specialized applications are possible, such as the monitoring of radio frequency interference (RFI), electromagnetic interference (EMI), and electromagnetic compatibility (EMC). These applications, as a rule, require an antenna connection and a low-noise, external amplifier. Which are then the specifications to look for in a good spectrum analyzer? We would suggest:

1. It should display selectable, very wide bands of the EM radio spectrum with power and frequency readable with good accuracy.
2. Its selectivity should range, in discrete steps, from few hertz to megahertz so that sidebands of a selected signal can be spotted and shown with the necessary details.
3. It should possess a very wide dynamic range, so that signals differing in amplitude six to eight orders of magnitude can be observed at the same time on the display.
4. Its sensitivity must be compatible with the measurements to be taken. As already mentioned, specialized applications may require external wideband, low-noise amplifiers and an antenna connection.
5. Stability and reliability are major requests but they are met most of the time.

Occasionally a battery-operated option for portable field applications may be necessary. A block diagram of a commercial spectrum analyzer is shown in Figure 23.7.
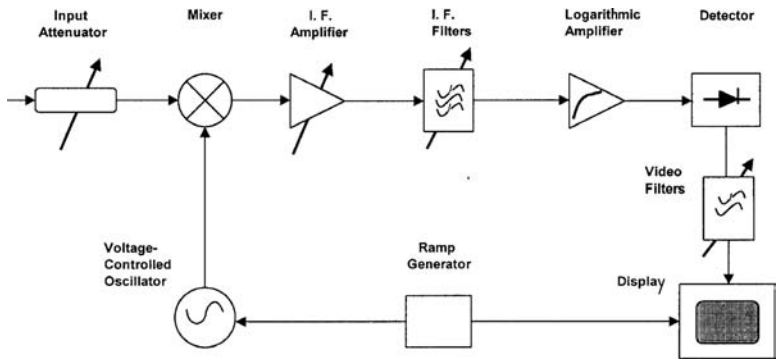
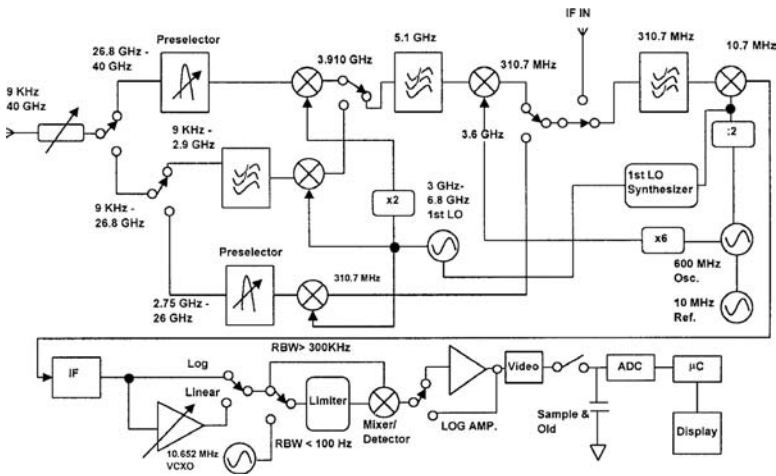**FIGURE 23.7** Block diagram of a commercial spectrum analyzer.

FIGURE 23.8 Standard block diagram of a modern spectrum analyzer.

Referring to Figure 23.7 we can say that we are confronted with a radio-receiver-like superhet with a wideband input circuit. The horizontal scale of the instrument is driven by a ramp generator which is also applied to the voltage-controlled LO [2].

A problem arises when dealing with a broadband mixing configuration like the one shown above, namely, avoiding receiving the image band.

The problem is successfully tackled here by upconverting the input band to a high-valued IF. An easily designed input low-pass filter, not shown in the block diagram for simplicity, will now provide the necessary rejection of the unwanted image band.

Nowadays, with the introduction of YIG bandpass filter preselectors, tunable over very wide input bands, upconversion is not always necessary. Traces of unwanted signals may, however, show up on the display although at very low level (less than –80 dBc) on good analyzers.

A block diagram of a commercial spectrum analyzer exploiting both the mentioned principles is shown in Figure 23.8. This instrument includes a very important feature which greatly improves its performance: the LO frequency is no longer coming from a free-running source but rather from a synthesized unit referenced to a very stable quartz oscillator. The improved quality of the LO both in terms of its own noise and frequency stability, optimizes several specifications of the instrument, such as frequency determining accuracy, finer resolution on display, and reduced noise in general.
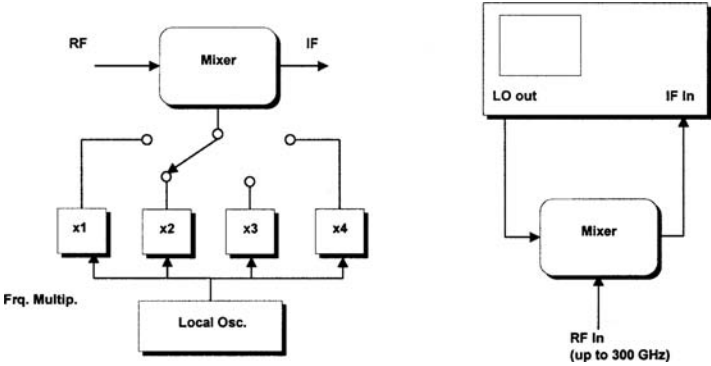
FIGURE 23.9   Encreasing the input bandwidth characteristics.

Further, a stable LO generates stable harmonics which can then be used to widen the input-selected bands up to the millimeter region. As already stated, this option requires external devices, e.g., a mixer-amplifier as shown in Figure 23.9a and b.

The power reference on the screen is the top horizontal line of the reticle. Due to the very wide dynamic range foreseen, the use of a log scale (e.g., 10 dB/square) seems appropriate. Conventionally, 1 mW is taken as the zero reference level: accordingly, dBm are used throughout.

The noise power level present on the display without an input signal connected (noise floor) is due to the input random noise multiplied by the IF amplifier gain. Such a noise is always present and varies with input frequency, IF selectivity, and analyzer sensitivity (in terms of noise figure).

The "on display dynamic range" of the analyzer is the difference between the maximum compression-free level of the input signal and the noise floor. As a guideline, the dynamic range of a good instrument could be of the order of 70 to 90 dB.

An input attenuator, always available on the front panel, allows one to apply more power to the analyzer while avoiding saturation and nonlinear readings. The only drawback is the obvious sensitivity loss. One should not expect a spectrum analyzer to give absolute power level readings to be better than a couple of dB.

For the accurate measurement of power levels, the suggestion is to use a power meter. An erratic signal pattern on display and a fancy level indication may be caused by the wrong setting of the "scan time" knob. It must be realized that high-resolution observation of a wide input band requires the proper scanning time. An incorrect parameter setting yields wrong readings but usually an optical alarm is automatically switched on to warn the operator.

The knowledge of the noise floor level allows a good valuation of the noise temperature, T_n (and therefore of the sensitivity), of the analyzer, a useful parameter on many occasions. The relations involved are as follows.

The Nyquist relation states that

$$P = kT_n B$$

where P = noise floor power level read on the display (W)
   k = Boltzmann constant = 1.38×10^{-23} (J/K)
   B = passband of the selected IF (Hz)

therefore,

$$T_n = P/(kB)$$

Usually engineers prefer to quote the noise figure of receivers. By definition we can write
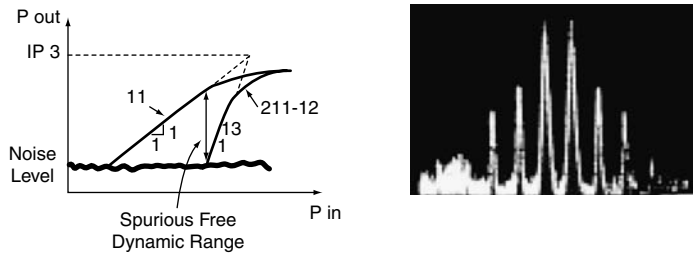
FIGURE 23.10 (a) Spurious free dynamic range. (b) Higher-order spurious.

$$N = \left(T_0/T_0\right) + 1$$

where $\qquad$ N = noise factor

$\qquad$ T_0 = 290 K

F (noise figure) = 10 log N

A typical F for a good spectrum analyzer is of the order of 30 dB.

It must be said, however, that the "ultimate sensitivity" of the spectrum analyzer will depend not only on its noise figure but also on the setting of other parameters like the video filter, the IF bandwidth, the insertion of averaging functions, the scan speed, the detector used, etc.

As a rough estimate a noise floor level of –130/–140 dBm is very frequently met by a good instrument.

Another criterion to select a spectrum analyzer is a good "IMD dynamic range," that is, the tendency to create spurious signals by intermodulation due to saturation.

This figure is generally quoted by the manufacturers, but it is also easily checked by the operator by injecting two equal amplitude sinusoidal signals at the input socket of the analyzer. The frequency separation between the two should be at least a couple of "resolution bandwidths," i.e., the selected IF bandwidth. As the input levels increase, spurious lines appear at the sum and difference frequencies and spacing of the input signals.

The range in decibels between the nonoverloaded input signals on display and the barely noticeable spurious lines is known as the "spurious free dynamic range," shown graphically in Figure 23.10a, where the third-order "intercept point" is also graphically determined. If input power is increased, higher-order spurious signals appear, as shown in Figure 23.10b. The input connector of most spectrum analyzers is of the 50 Ω coaxial type. Past instruments invariably used N-type connectors because of their good mechanical and electrical behavior up to quite a few gigahertz. Today SMA or K connectors are preferred.

External millimeter wave amplifiers and converters use waveguide input terminations. As is discussed in the next section, multipurpose analyzers are available where power meter, frequency counter, tracking generator, etc. can all be housed in the same cabinet. The economic and practical convenience of these units must be weighed on a case-by-case basis.

Finally, we mention that spectrum analyzers are available equipped with AM and FM detectors to facilitate their use in the RFI monitoring applications.

What Is the Right Spectrum Analyzer for My Purpose?

Several manufacturers offer a large number of spectrum analyzer models; the choice may be made on the basis of application field (i.e., CATV, mobile telephony, service, surveillance, R&D, etc.), performance (resolution bandwidth, frequency range, accuracy, battery operation etc.), or cost.

In addition, it is important to know that most spectrum analyzers need some accessories generally not furnished as a standard: for example, a connectorized, coaxial, microwave cable is always required; a directional coupler, or power divider, or handheld sampler antenna may be very useful to pick up the signals; and a personal computer is useful to collect, store, reduce, and analyze the data.

There are four main families of RF and microwave spectrum analyzers.

Family 1

The bench instruments are top performance, but also large, heavy, and the most expensive class, intended for metrology, certification, factory reference, and for radio surveillance done by government and military institutions.

The frequency ranges span from a few tens of hertz up to RF (i.e., 2.9 GHz), up to microwave region (i.e., 26.5 GHz), or up to near millimeter wavelength (i.e., 40 GHz). This class of instruments includes lower noise figures, approximately 20 dB, and may be decreased down to 10 to 15 dB with an integrated preamplifier. The synthesized local oscillator has a good phase noise (typically 10 dB better than other synthesized spectrum analyzers) for precise, accurate, and stable measurement. Also this class of instruments, by sharing the display unit, can be integrated with plug-in instruments like a power meter (for more accurate power measurements) or a tracking generator (for network analysis and mixer testing).

The interface to a computer (and a printer) such IEEE-488 or RS-232 is standard; it allows remote control and data readings; this class of spectrum analyzer often has a powerful microprocessor, RAM, and disks for storing data and performing statistical and mathematical analysis.

The best known families are the Hewlett-Packard series, 71xxxx [3] and the Rhode & Schwarz series FSxx. [4]. Indicative prices are between $50,000 and $90,000.

Family 2

Less expensive bench instruments, the workhorse class of spectrum analyzers, portable and lightweight, are associated with a synthesized local oscillator, that includes a frequency range from a few kilohertz up to RF region (i.e., 2.9 GHz), microwave region (i.e., 26.5 GHz), or near millimeter wavelengths (i.e., 40 to 50 GHz). A typical noise figure of 30 dB is good enough to ensure most measurements. A large number of filters down to few hertz of resolution are offered; digital filters are preferable to analog ones, because they give a faster refresh rate of the trace on the display. This kind of spectrum analyzer nearly always has the capability to extend the frequency range up to millimeter and submillimeter wavelengths with an external mixer. One of the most important features for a spectrum analyzer in this class is the quality of the local oscillator; it should be synthesized (PLL) to achieve stability, precision, accuracy, and low phase noise. Demodulation is also an important feature to listen to AM, FM on the loudspeaker and to display TV pictures or complex modulations onto the screen, which is often required by people working on surveillance, TV, and mobile telephone. The interface to a computer such as IEEE-488 or RS232 is standard in a large number of spectrum analyzers, and allows the remote control and data reading, storing, and manipulation.

This kind of instrument may integrate a tracking generator, a frequency counter, and other instruments that can transform the spectrum analyzer into a compact, full-featured RF and microwave laboratory.

The most popular families are the Hewlett-Packard series 856xx [3, 5], Rhode & Schwarz series FSExxx [4], Anritsu series MS26x3 [6], IFR mod. AN930 [7], and Marconi Instruments series 239x [9]. The Tektronix production should be taken in account. Prices typically span from $30,000 to $60,000.

Family 3

The entry level, a more economical class of spectrum analyzer, is intended for field use or for one specific application. If your need is mainly EMI/EMC, CATV, mobile telephone, or surveillance, perhaps you do not need the extreme stability of a synthesized local oscillator, and a frequency range up to 2 GHz may be enough; however, if you need some special functions such as "quasi-peak detector" or "occupied bandwidth measurement," two functions that are a combination of a mathematical treatment with some legislative aspects, these are easily measured with a spectrum analyzer including those functions. As the normatives can change, the capability to easily upgrade the measurement software is important; some models come with a plug-in memory card, some others with 3.5″ disks.

A large number of spectrum analyzer models are tailored to meet the specific needs of a customer. This is the case with the HP series 859x [3], Tektronix series 271x [10], IFR series A-xxxx [8], Anritsu MS2651 [6], and Advantest series U4x4x [4]. Costs typically are around $10, 000 to $20,000.

Family 4

The most economical class of spectrum analyzer, with prices around \$2,000 to \$6000, includes instruments that perform only the basic functions with a limited frequency range and filter availability and without digital capability. They are intended for service, for general-purpose measurements (i.e., IP_3, harmonic distortion) or for precertification in EMI/EMC measurements. One of the most popular series is the Hameg series HM50xx [11].

In this class are some special spectrum analyzers that come on a personal computer (PC) board. Such spectrum analyzers, generally cheap (typically \$3,000 to \$5,000), with frequency range up to 2 GHz, may include PLL local oscillators, tracking generators, and other advanced characteristics. The input is through a coaxial connector on the board, the output and the control is done by a virtual instrument running on the PC. One model is made by DKD Instruments [12].

Other unusual RF spectrum analyzers working in conjunction with a PC and worth noting are the instruments for EMI/EMC measurements and reduction in power lines and power cords. For this type of instrument, the core is not the hardware but the software that performs the measurement according to international standards and may guide the engineer to meet the required compatibility. An example is given by Seaward Electronic Sceptre [13].

Advanced Applications

New technological approaches and the use of spectrum analysis concepts in radioastronomy constitute some advanced spectrum analysis applications. Autocorrelators, with a typical frequency resolution of ~5/25 kHz, have been extensively used in radioastronomy. Their performance is well documented; the autocorrelation function is computed online and recorded. Later the FFT of the function is computed off line in order to get the power spectrum. Recently, the Tektronix 3054 Fourier Analyzer, based on a bank of programmable filters, was introduced as an alternative approach. The state of the art in integrated digital signal processors (DSPs) allows an alternative approach to spectrum analysis. By paralleling several of these DSPs, one is able to compute online the FFT directly on a very wide input bandwidth (several tens of megahertz).

By using this technique, high time and frequency resolution can be achieved.

A system based on the Sharp LH9124-LH9320 chip set is described in Figure 23.11. It is based on VME boards: one or two 10-bit, 40-MS/s ADCs and two boards in charge to compute the FFT of the incoming streams of data, in real time [14]. A following block computes the power and averages on the board up to 64 K spectra before storing the result on disk or tape. The FFT boards are powered by one of the fastest state-of-the-art DSPs (Sharp LH9124). The overall system is controlled by an embedded FORCE 3 Sparcstation. The LH 9124 DSP works with 24+24 bits (with 6 exponent bit) in block floating point. The system architecture allows expansion of the input bandwidth and the number of channels by paralleling more DSP boards. All the computing core is housed in a VME crate and is able to produce single-sided spectra from 1024 frequency bins to 131072 bins at an input bandwidth of 12 MHz without losing data or with 56% of time efficiency at 20 MHz. Single- or double-channel operation mode is provided. In a single-channel mode the main features of the system are reported as

Input bandwidth	0.5–20 MHz
Time efficiency	100% at 12 MHz (56% at 20 MHz)
FFT size	1K, 2K, 256K (points)
Avgs out format	<256 averages → integer 24 bits
	>256 averages → float 32 bits
Windows	Hanning, Hamming, Kaiser Bessel

This spectrometer was developed (1993) as a cost-effective system for both the NASA-SETI (Search for Extraterrestrial Intelligence) program [15,16] and for radioastronomical spectroscopy [17] at the CNR Institute of Radio Astronomy of Bologna. The digital spectrometer was first used to investigate the effects of the Jupiter/SL9 comet impacts (July 1994) [18,19]. In this application, the high time resolution
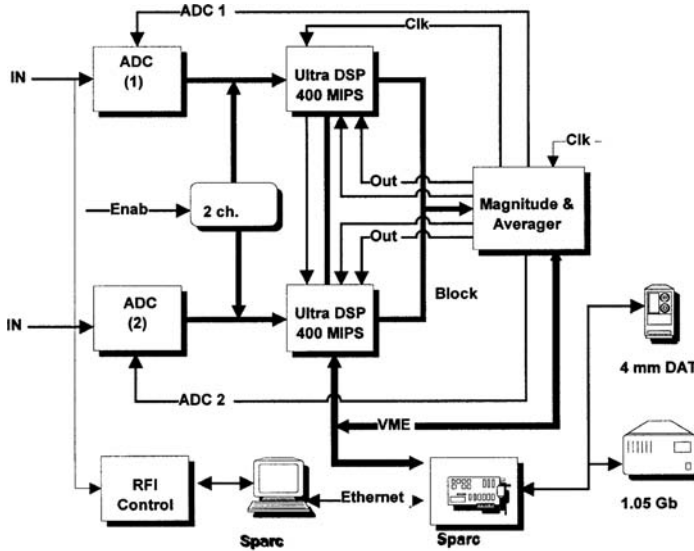
FIGURE 23.11   Block diagram of spectrum analyzer for radioastronomy.

of the spectrometer (a 16K points FFT every 1.3 ms) was exploited to compensate for the fast planet rotational velocity Doppler shift.

The system has been successfully used at the 32 m dish radiotelescope near Bologna in many line observations with unique results. Note that the use of such a high time and resolution system in radio astronomy may help to observe the molecular line in a very precise and unusual way. The whole pattern (a couple of megahertz wide) of a NH_3 molecule line coming from the sky was obtained in flash mode with a frequency resolution high enough to distinguish the different components. The same machine can be used for high-time-resolution observations of pulsar and millisecond pulsar. In those cases, the possibility of performing the FFT of the RF signal online allows coherent dedispersion of the pulses. This new technological approach in computing the FFT may be successfully addressed to many different fields, such as image processing, medical diagnostic systems, radio surveillance, etc.

References

1. A.E. Bailey, Ed., *Microwave Measurements,* 2nd ed., Peter Peregrins on behalf of IEEE, London, 1989.
2. Hewlett-Packard, *A.N 63* and *A.N 243,* Hewlett-Packard Company.
3. Hewlett-Packard, *Test & Measurement Catalog 1997,* pp. 224–257, Hewlett-Packard Company.
4. Rohde & Schwarz, *Catalog 96/97 Test & Measurement Products* pp. 109–145, Rohde & Schwarz Company.
5. Hewlett-Packard, *HP 8560E, HP8561E and HP8563E Portable Spectrum Analizers Technical Data,* Hewlett-Packard Company, pp. 5091–3274E.
6. Anritsu-Wiltron, *Electronic Measuring Instruments,* pp. 67–100, Anritsu-Wiltron Company.
7. IFR, *The AN 930 Microwave Spectrum Analyzer,* IFR System, Inc.
8. IFR, *The A-8000 & A-7500 Spectrum Analyzer,* IFR System, Inc.
9. Marconi Instruments, *Test and Measurement 96/97 Edition Instrument System*, pp. 4-1–4-5, Marconi Instruments Company.
10. Tektronix, *Measurement Products Catalog 1997/1998,* pp. 145–171, Tektronix Company.
11. Hameg Instruments, *Catalog 1994,* Hameg GmbH.
12. DKD Instruments, *Model 1800 Data Sheet*, DKD Instruments.
13. C. Galli, Analizzatori di spettro: rassegna commerciale, *Sel. Elettron.,* 15, 1996 [in Italian].

14. S. Montebugnoli et al., "A new 6 MHz 128000 channels spectrum analyzer," IAA-95-IAA.9.1.10. 46th International Astronautical Congress, October, Oslo, Norway, 1995.

15. S. Montebugnoli, C. Bortolotti, S. Buttaccio, A. Cattani, N. D'Amico, G. Grueff, A. Maccaferri, G. Maccaferri, A. Orfei, M. Roma, G. Tuccari, and M. Tugnoli, "Upgrade of the mini spectrum analyzer" in *5th Proceedings of the International Conference on Bioastronomy,* Capri, Vol. I, July 1996.

16. A. Orfei, S. Montebugnoli, C. Miani, J. Monari, and G. Tuccari, "The SETI facilities at the Medicina/ Noto sites," (preprint) IAA-96-IAA.9.1.11, 47th International Astronautical Congress, October, Beijing, China, 1996.

17. S. Montebugnoli, C. Bortolotti, S. Buttaccio, A. Cattani, N. D'Amico, G. Grueff, A. Maccaferri, G. Maccaferri, A. Orfei, M. Roma, G. Tuccari, and M. Tugnoli, "A new high-resolution digital spectrometer for radioastronomy applications," *Rev. Sci. Instrum.* 67, 365–370, 1996.

18. C.B. Cosmovici et al., "Detection of the 22 GHz line of water during and after the SL-9/Jupiter event," European SL-9/Jupiter Workshop–February, ESO Headquarters, Munchen, Germany, 1995.

19. C.B. Cosmovici, S. Montebugnoli, A. Orfei, S. Pogrebenko, and P. Colom, "First evidence of planetary water MASER emission induced by the comet/Jupiter catastrophic impact," *Planetary Space Sci.,* 44, 735–773, 1996.