

Power Spectra Estimation

National Semiconductor
Application Note 255
November 1980



1.0 INTRODUCTION

Perhaps one of the more important application areas of digital signal processing (DSP) is the power spectral estimation of periodic and random signals. Speech recognition problems use spectrum analysis as a preliminary measurement to perform speech bandwidth reduction and further acoustic processing. Sonar systems use sophisticated spectrum analysis to locate submarines and surface vessels. Spectral measurements in radar are used to obtain target location and velocity information. The vast variety of measurements spectrum analysis encompasses is perhaps limitless and it will thus be the intent of this article to provide a brief and fundamental introduction to the concepts of power spectral estimation.

Since the estimation of power spectra is statistically based and covers a variety of digital signal processing concepts, this article attempts to provide sufficient background through its contents and appendices to allow the discussion to flow void of discontinuities. For those familiar with the preliminary background and seeking a quick introduction into spectral estimation, skipping to Sections 6.0 through 11.0 should suffice to fill their need. Finally, engineers seeking a more rigorous development and newer techniques of measuring power spectra should consult the excellent references listed in Appendix D and current technical society publications.

As a brief summary and quick lookup, refer to the Table of Contents of this article.

TABLE OF CONTENTS

Section	Description	Appendix A	Description
1.0	Introduction	A.0	Concepts of Probability, Random Variables and Stochastic Processes
2.0	What is a Spectrum?	A.1	Definitions of Probability
3.0	Energy and Power	A.2	Joint Probability
4.0	Random Signals	A.3	Conditional Probability
5.0	Fundamental Principles of Estimation Theory	A.4	Probability Density Function (pdf)
6.0	The Periodogram	A.5	Cumulative Distribution Function (cdf)
7.0	Spectral Estimation by Averaging Periodograms	A.6	Mean Values, Variances and Standard Deviation
8.0	Windows	A.7	Functions of Two Jointly Distributed Random Variables
9.0	Spectral Estimation by Using Windows to Smooth a Single Periodogram	A.8	Joint Cumulative Distribution Function
10.0	Spectral Estimation by Averaging Modified Periodograms	A.9	Joint Probability Density Function
11.0	Procedures for Power Spectral Density Estimates	A.10	Statistical Independence
12.0	Resolution	A.11	Marginal Distribution and Marginal Density Functions
13.0	Chi-Square Distributions	A.12	Terminology: Deterministic, Stationary, Ergodic
14.0	Conclusion	A.13	Joint Moments
15.0	Acknowledgements	A.14	Correlation Functions
		Appendices	
		B.	Interchanging Time Integration and Expectations
		C.	Convolution
		D.	References

2.0 WHAT IS A SPECTRUM?

A spectrum is a relationship typically represented by a plot of the magnitude or relative value of some parameter against frequency. Every physical phenomenon, whether it be an electromagnetic, thermal, mechanical, hydraulic or any other system, has a unique spectrum associated with it. In electronics, the phenomena are dealt with in terms of signals, represented as fixed or varying electrical quantities of voltage, current and power. These quantities are typically described in the time domain and for every function of time, $f(t)$, an equivalent frequency domain function $F(\omega)$ can be found that specifically describes the frequency-component content (frequency spectrum) required to generate $f(t)$. A study of relationships between the time domain and its corresponding frequency domain representation is the subject of Fourier analysis and Fourier transforms.

The *forward Fourier transform*, time to frequency domain, of the function $x(t)$ is defined

$$F[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt = X(\omega) \quad (1)$$

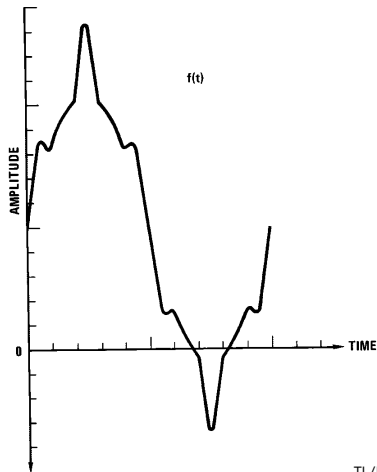


FIGURE 1. An Electrical Signal $f(t)$

TL/H/8712-1

and the *inverse Fourier transform*, frequency to time domain, of $X(\omega)$ is

$$F^{-1}[X(\omega)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{j\omega t} d\omega = x(t) \quad (2)$$

(For an in-depth study of the Fourier integral see reference 19.) Though these expressions are in themselves self-explanatory, a short illustrative example will be presented to aid in relating the two domains.

If an arbitrary time function representation of a periodic electrical signal, $f(t)$, were plotted versus time as shown in Figure 1, its Fourier transform would indicate a spectral content consisting of a DC component, a fundamental frequency component ω_0 , a fifth harmonic component $5\omega_0$ and a ninth harmonic component $9\omega_0$ (see Figure 2). It is illustratively seen in Figure 3 that the superposition of these frequency components, in fact, yields the original time function $f(t)$.

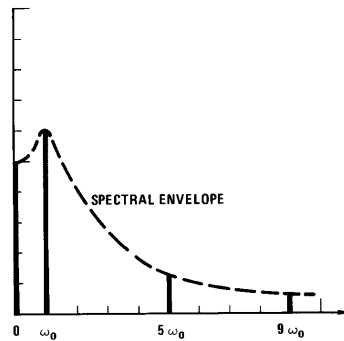


FIGURE 2. Spectral Composition or Spectrum $F(\omega)$ or $f(t)$

TL/H/8712-2

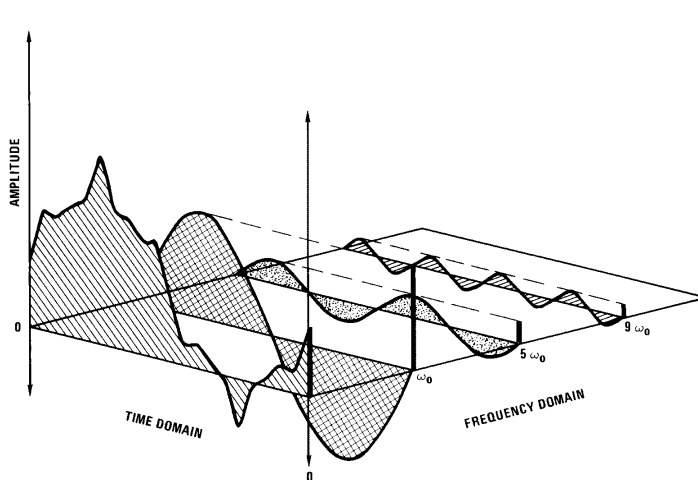


FIGURE 3. Combined Time Domain and Frequency Domain Plots

TL/H/8712-3

3.0 ENERGY AND POWER

In the previous section, time and frequency domain signal functions were related through the use of Fourier transforms. Again, the same relationship will be made in this section but the emphasis will pertain to signal power and energy.

Parseval's theorem relates the representation of energy, $\omega(t)$, in the time domain to the frequency domain by the statement

$$\omega(t) = \int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \int_{-\infty}^{\infty} F_1(f)F_2(f) df \quad (3)$$

where $f(t)$ is an arbitrary signal varying as a function of time and $F(f)$ its equivalent Fourier transform representation in the frequency domain.

The proof of this is simply

$$\int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \int_{-\infty}^{\infty} f_1(t)f_2(t) dt \quad (4a)$$

Letting $F_1(f)$ be the Fourier transform of $f_1(t)$

$$\int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} F_1(f)e^{j2\pi ft} df \right] f_2(t) dt \quad (4b)$$

$$= \int_{-\infty}^{\infty} \left[F_1(f) \int_{-\infty}^{\infty} e^{j2\pi ft} f_2(t) dt \right] df \quad (4c)$$

Rearranging the integrand gives

$$\int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \int_{-\infty}^{\infty} F_1(f) \left[\int_{-\infty}^{\infty} f_2(t)e^{j2\pi ft} dt \right] df \quad (4d)$$

and the factor in the brackets is seen to be $F_2(-f)$ (where $F_2(-f) = F_2^*(f)$ the conjugate of $F_2(f)$) so that

$$\int_{-\infty}^{\infty} f_1(t)f_2(t) dt = \int_{-\infty}^{\infty} F_1(f)F_2(-f) df \quad (4e)$$

A corollary to this theorem is the condition $f_1(t) = f_2(t)$ then $F(-f) = F^*(f)$, the complex conjugate of $F(f)$, and

$$\omega(t) = \int_{-\infty}^{\infty} f^2(t) dt = \int_{-\infty}^{\infty} F(f)F^*(f) df \quad (5a)$$

$$= \int_{-\infty}^{\infty} |F(f)|^2 df \quad (5b)$$

This simply says that the total energy[†] in a signal $f(t)$ is equal to the area under the square of the magnitude of its Fourier transform. $|F(f)|^2$ is typically called the *energy density*, *spectral density*, or *power spectral density* function and $|F(f)|^2 df$ describes the density of signal energy contained in the differential frequency band from f to $f + df$.

In many electrical engineering applications, the instantaneous signal power is desired and is generally assumed to be equal to the square of the signal amplitudes i.e., $f^2(t)$.

[†]Recall the energy storage elements

Inductor $v = L \frac{di}{dt}$

$$\omega(t) = \int_0^T v i dt = \int_0^T L \frac{di}{dt} i dt = \int_0^I L i di = \frac{1}{2} L i^2$$

Capacitor $i = C \frac{dv}{dt}$

$$\omega(t) = \int_0^T v i dt = \int_0^T v C \frac{dv}{dt} dt = \int_0^V C v dv = \frac{1}{2} C v^2$$

This is only true, however, assuming that the signal in the system is impressed across a 1Ω resistor. It is realized, for example, that if $f(t)$ is a voltage (current) signal applied across a system resistance R , its true instantaneous power would be expressed as $[f(t)]^2/R$ (or for current $[f(t)]^2 R$) thus, $[f(t)]^2$ is the true power only if $R = 1\Omega$.

So for the general case, power is always proportional to the square of the signal amplitude varied by a proportionality constant R , the impedance level in a circuit. In practice, however, it is undesirable to carry extra constants in the computation and customarily for signal analysis, one assumes signal measurement across a 1Ω resistor. This allows units of power to be expressed as the square of the signal amplitudes and the units of energy measured as volts²-second (amperes²-second).

For periodic signals, equation (5) can be used to define the average power, P_{avg} , over a time interval t_2 to t_1 by integrating $[f(t)]^2$ from t_1 to t_2 and then obtaining the average after dividing the result by $t_2 - t_1$ or

$$P_{avg} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} f^2(t) dt \quad (6a)$$

$$= \frac{1}{T} \int_0^T f^2(t) dt \quad (6b)$$

where T is the period of the signal.

Having established the definitions of this section, energy can now be expressed in terms of power, $P(t)$,

$$\omega(t) = \int_{-\infty}^{\infty} [f(t)]^2 dt \quad (7a)$$

$$= \int_{-\infty}^{\infty} P(t) dt \quad (7b)$$

with power being the time rate of change of energy

$$P(t) = \frac{d\omega(t)}{dt} \quad (8)$$

As a final clarifying note, again, $|F(f)|^2$ and $P(t)$, as used in equations (7b) and (8), are commonly called throughout the technical literature, *energy density*, *spectral density*, or *power spectral density* functions, PSD. Further, PSD may be interpreted as the average power associated with a bandwidth of one hertz centered at f hertz.

4.0 RANDOM SIGNALS

It was made apparent in previous sections that the use of Fourier transforms for analysis of linear systems is widespread and frequently leads to a saving in labor.

In view of using frequency domain methods for system analysis, it is natural to ask if the same methods are still applicable when considering a random signal system input. As will be seen shortly, with some modification, they *will* still be useful and the modified methods offer essentially the same advantages in dealing with random signals as with nonrandom signals.

It is appropriate to ask if the Fourier transform may be used for the analysis of any random sample function. Without proof, two reasons can be discussed which make the transform equations (1) and (2) invalid.

Firstly, $X(\omega)$ is a random variable since, for any fixed ω , each sample would be represented by a different value of the ensemble of possible sample functions. Hence, it is not a frequency representation of the process but only of one member of the process. It might still be possible, however, to use this function by finding its mean or expected value over the ensemble except that the second reason negates this approach. The second reason for not using the $X(\omega)$ of equations (1) and (2) is that, for stationary processes, it almost never exists. As a matter of fact, one of the conditions for a time function to be Fourier transformable is that it be integrable so that,

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \quad (9)$$

A sample from a stationary random process can never satisfy this condition (with the exception of generalized functions inclusive of impulses and so forth) by the argument that if a signal has nonzero power, then it has infinite energy and if it has finite energy then it has zero power (average power). Shortly, it will be seen that the class of functions having no Fourier integral, due to equation (9), but whose average power is finite can be described by statistical means.

Assuming $x(t)$ to be a sample function from a stochastic process, a truncated version of the function $x_T(t)$ is defined as

$$x_T(t) = \begin{cases} x(t) & |t| \leq T \\ 0 & |t| > T \end{cases} \quad (10)$$

and

$$x(t) = \lim_{T \rightarrow \infty} x_T(t) \quad (11)$$

This truncated function is defined so that the Fourier transform of $x_T(t)$ can be taken. If $x(t)$ is a power signal, then recall that the transform of such a signal is not defined

$$\int_{-\infty}^{\infty} |x(t)| dt \text{ not less than } \infty \quad (12)$$

but that

$$\int_{-\infty}^{\infty} |x_T(t)| dt < \infty \quad (13)$$

The Fourier transform pair of the truncated function $x_T(t)$ can thus be taken using equations (1) and (2). Since $x(t)$ is a power signal, there must be a power spectral density function associated with it and the total area under this density must be the average power despite the fact that $x(t)$ is non-Fourier transformable.

Restating equation (5) using the truncated function $x_T(t)$

$$\int_{-\infty}^{\infty} x_T^2(t) dt = \int_{-\infty}^{\infty} |X_T(f)|^2 df \quad (14)$$

and dividing both sides by $2T$

$$\frac{1}{2T} \int_{-\infty}^{\infty} x_T^2(t) dt = \frac{1}{2T} \int_{-\infty}^{\infty} |X_T(f)|^2 df \quad (15)$$

gives the left side of equation (15) the physical significance of being proportional to the average power of the sample function in the time interval $-T$ to T . This assumes $x_T(t)$ is a voltage (current) associated with a resistance. More precisely, it is the square of the *effective value* of $x_T(t)$

and for an ergodic process approaches the *mean-square* value of the process as T approaches infinity.

At this particular point, however, the limit as T approaches infinity cannot be taken since $X_T(f)$ is non-existent in the limit. Recall, though, that $X_T(f)$ is a random variable with respect to the ensemble of sample functions from which $x(t)$ was taken. The limit of the *expected value* of

$$\frac{1}{2T} |X_T(f)|^2$$

can be reasonably assumed to exist since its integral, equation (15), is always positive and certainly does exist. If the expectations $E\{\}$, of both sides of equation (15) are taken

$$E\left\{\frac{1}{2T} \int_{-\infty}^{\infty} x_T^2(t) dt\right\} = E\left\{\frac{1}{2T} \int_{-\infty}^{\infty} |X_T(f)|^2 df\right\} \quad (16)$$

then interchanging the integration and expectation and at the same time taking the limit as $T \rightarrow \infty$

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} \overline{x^2(t)} dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} E\{|X_T(f)|^2\} df \quad (17)$$

results in

$$\overline{x^2(t)} = \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E\{|X_T(f)|^2\}}{2T} df \quad (18)$$

where $\overline{x^2(t)}$ is defined as the mean-square value ($\overline{}$ denotes ensemble averaging and $\langle \rangle$ denotes time averaging).

For stationary processes, the time average of the mean-square value is equal to the mean-square value so that equation (18) can be restated as

$$\overline{x^2(t)} = \int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{E\{|X_T(f)|^2\}}{2T} df \quad (19)$$

The integrand of the right side of equation (19), similar to equation (5b), is called the *energy density* spectrum or *power spectral density* function of a random process and will be designated by $S(f)$ where

$$S(f) = \lim_{T \rightarrow \infty} \frac{E\{|X_T(f)|^2\}}{2T} \quad (20)$$

Recall that letting $T \rightarrow \infty$ is not possible before taking the expectation.

Similar to the argument in Section III, the physical interpretation of power spectral density can be thought of in terms of average power. If $x(t)$ is a voltage (current) associated with a 1Ω resistance, $x^2(t)$ is just the average power dissipated in that resistor and $S(f)$ can be interpreted as the average power associated with a bandwidth of one hertz centered at f hertz.

$S(f)$ has the units volts²-second and its integral, equation (19), leads to the mean square value hence,

$$\overline{x^2(t)} = \int_{-\infty}^{\infty} S(f) df \quad (21)$$

Having made the tie between the power spectral density function $S(f)$ and statistics, an important theorem in power spectral estimation can now be developed.

Using equation (20) and recalling that $X_T(f)$ is the Fourier transform of $x_T(t)$, assuming a nonstationary process,

$$S(f) = \lim_{T \rightarrow \infty} \frac{E[|X_T(f)|^2]}{2T} \quad (22)$$

$$S(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} E \left\{ \int_{-\infty}^{\infty} x_T(t_1) e^{j\omega t_1} dt_1 \int_{-\infty}^{\infty} x_T(t_2) e^{-j\omega t_2} dt_2 \right\} \quad (23)$$

Note that $|X_T(f)|^2 = X_T(f)X_T^*(f)$ and that in order to distinguish the variables of integration when equation (23) is re-manipulated the subscripts of t_1 and t_2 have been introduced. So, (see Appendix B)

$$S(f) = \lim_{T \rightarrow \infty} \left\{ \frac{1}{2T} E \left[\int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_1 \epsilon^{-j\omega(t_2-t_1)} x_T(t_1) x_T(t_2) \right] \right\} \quad (24)$$

$$= \lim_{T \rightarrow \infty} \left\{ \frac{1}{2T} \int_{-\infty}^{\infty} dt_2 \int_{-\infty}^{\infty} dt_1 \epsilon^{-j\omega(t_2-t_1)} E[x_T(t_1) x_T(t_2)] \right\} \quad (25)$$

Finally, the expectation $E[x_T(t_1) x_T(t_2)]$ is recognized as the autocorrelation, $R_{xx}(t_1, t_2)$ (Appendix A.14) function of the truncated process where

$$E[x_T(t_1) x_T(t_2)] = R_{xx}(t_1, t_2) \quad |t_1|, |t_2| \leq T$$

$$= 0 \quad \text{everywhere else.}$$

Substituting

$$t_2 - t_1 = \tau \quad (26)$$

$$dt_2 = d\tau \quad (27)$$

equation (25) next becomes

$$S(f) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\infty}^{\infty} d\tau \int_{-T}^T R_{xx}(t_1, t_1 + \tau) \epsilon^{-j\omega \tau} dt_1 \quad (28)$$

or

$$S(f) = \left\{ \left[\int_{-\infty}^{\infty} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T R_{xx}(t_1, t_1 + \tau) dt_1 \right] \epsilon^{-j\omega \tau} d\tau \right\} \quad (29)$$

We see then that the spectral density is the Fourier transform of the time average of the autocorrelation function. The relationship of equation (29) is valid for a nonstationary process.

For the stationary process, the autocorrelation function is independent of time and therefore

$$\langle R_{xx}(t_1, t_1 + \tau) \rangle = R_{xx}(\tau) \quad (30)$$

From this it follows that the spectral density of a stationary random process is just the Fourier transform of the autocorrelation function where

$$S(f) = \int_{-\infty}^{\infty} R_{xx}(\tau) \epsilon^{-j\omega \tau} d\tau \quad (31)$$

and

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} S(f) \epsilon^{j\omega \tau} df \quad (32)$$

are described as the Wiener-Khinchine theorem.

The Wiener-Khinchine relation is of fundamental importance in analyzing random signals since it provides a link between the time domain [correlation function, $R_{xx}(\tau)$] and the frequency domain [spectral density, $S(f)$]. Note that the uniqueness is in fact the Fourier transformability. It follows, then, for a stationary random process that the autocorrelation function is the inverse transform of the spectral density function. For the nonstationary process, however, the autocorrelation function cannot be recovered from the spectral density. Only the time average of the correlation is recoverable, equation (29).

Having completed this section, the path has now been paved for a discussion on the techniques used in power spectral estimation.

5.0 FUNDAMENTAL PRINCIPLES OF ESTIMATION THEORY

When characterizing or modeling a random variable, estimates of its statistical parameters must be made since the data taken is a finite sample sequence of the random process.

Assume a random sequence $x(n)$ from the set of random variables $\{x_n\}$ to be a realization of an ergodic random process (random for which ensemble or probability averages, $E[\cdot]$, are equivalent to time averages, $\langle \cdot \rangle$ where for all n ,

$$m = E[x_n] = \int_{-\infty}^{\infty} x f_x(x) dx \quad (33)$$

Assuming further that the estimate of the desired averages of the random variables $\{x_n\}$ from a finite segment of the sequence, $x(n)$ for $0 \leq n \leq N-1$, to be

$$m = \langle x_n \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x_n \quad (34)$$

then for each sample sequence

$$m = \langle x(n) \rangle = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N x(n) \quad (35)$$

Since equation (35) is a precise representation of the mean value in the limit as N approaches infinity then

$$\hat{m} = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \quad (36)$$

may be regarded as a fairly accurate estimate of m for sufficiently large N . The caret ($\wedge$) placed above a parameter is representative of an estimate. The area of statistics that pertains to the above situations is called *estimation theory*. Having discussed an elementary statistical estimate, a few common properties used to characterize the estimator will next be defined. These properties are *bias*, *variance* and *consistency*.

Bias

The difference between the mean or expected value $E[\hat{\eta}]$ of an estimate $\hat{\eta}$ and its true value η is called the *bias*.

$$\text{bias} = B_{\hat{\eta}} = \eta - E[\hat{\eta}] \quad (37)$$

Thus, if the mean of an estimate is equal to the true value, it is considered to be *unbiased* and having a *bias* value equal to zero.

Variance

The variance of an estimator effectively measures the width of the probability density (Appendix A.4) and is defined as

$$\sigma_{\hat{\eta}}^2 = E[(\hat{\eta} - E[\hat{\eta}])^2] \quad (38)$$

A good estimator should have a small variance in addition to having a small bias suggesting that the probability density function is concentrated about its mean value. The mean square error associated with this estimator is defined as

$$\text{mean square error} = E[(\hat{\eta} - \eta)^2] = \sigma_{\hat{\eta}}^2 + B_{\hat{\eta}}^2 \quad (39)$$

Consistency

If the bias and variance both tend to zero as the limit tends to infinity or the number of observations become large, the estimator is said to be consistent. Thus,

$$\lim_{N \rightarrow \infty} \sigma_{\hat{\eta}}^2 = 0 \quad (40)$$

and

$$\lim_{N \rightarrow \infty} B_{\hat{\eta}} = 0 \quad (41)$$

This implies that the estimator converges in probability to the true value of the quantity being estimated as N becomes infinite.

In using estimates the mean value estimate of m_x , for a Gaussian random process is the *sample mean* defined as

$$\hat{m}_x = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad (42)$$

the mean square value is computed as

$$E[\hat{m}_x^2] = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E[x_i x_j] \quad (43)$$

$$= \frac{1}{N^2} \left[\sum_{i=0}^{N-1} E[x_i^2] + \sum_{i=0}^{N-1} \sum_{\substack{j=0 \\ i \neq j}}^{N-1} E[x_i] \cdot E[x_j] \right] \quad (44)$$

$$= \frac{1}{N^2} \left[N(E[x_n^2]) + \left(\sum_{i=0}^{N-1} E[x_i] \right) \left(\sum_{\substack{j=0 \\ i \neq j}}^{N-1} E[x_j] \right) \right] \quad (45)$$

$$= \frac{1}{N} E[x_n^2] + (m_x)^2 \frac{N-1}{N} \quad (46)$$

thus allowing the variance to be stated as

$$\sigma_{\hat{m}_x}^2 = E[(\hat{m}_x)^2] - \{E[\hat{m}_x]\}^2 \quad (47)$$

$$= \frac{1}{N} E[x_n^2] + (m_x)^2 \frac{N-1}{N} - \{E[\hat{m}_x]\}^2 \quad (48)$$

$$= \frac{1}{N} E[x_n^2] + (m_x)^2 \frac{N-1}{N} - m_x^2 \quad (49)$$

$$= \frac{1}{N} (E[x_n^2] - m_x^2) \quad (50)$$

$$= \frac{\sigma_x^2}{N} \quad (51)$$

This says that as the number of observations N increase, the variance of the sample mean decreases, and since the bias is zero, the sample mean is a consistent estimator.

If the variance is to be estimated and the mean value is a known then

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - m_x)^2 \quad (52)$$

this estimator is consistent.

If, further, the mean and the variance are to be estimated then the sample variance is defined as

$$\hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \hat{m}_x)^2 \quad (53)$$

again $\hat{m}_x$ is the sample mean.

The only difference between the two cases is that equation (52) uses the true value of the mean, whereas equation (53) uses the estimate of the mean. Since equation (53) uses an estimator the bias can be examined by computing the expected value of $\hat{\sigma}_x^2$ therefore,

$$E[\hat{\sigma}_x^2] = \frac{1}{N} \sum_{i=0}^{N-1} (E[x_i] - E[\hat{m}_x])^2 \quad (54)$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} \left\{ E[x_i^2] - 2E[x_i \hat{m}_x] + E[\hat{m}_x^2] \right\} \quad (55)$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} E[x_i^2] - \frac{2}{N^2} \sum_{i=0}^{N-1} \left(\sum_{j=0}^{N-1} E[x_i x_j] \right) \quad (56)$$

$$+ \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E[x_i x_j]$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} E[x_i^2] - \frac{2}{N^2} \left(\sum_{i=0}^{N-1} E[x_i^2] + \sum_{i=0}^{N-1} \sum_{j=0, j \neq i}^{N-1} E[x_i] \cdot E[x_j] \right) + \frac{1}{N^2} \left(\sum_{i=0}^{N-1} \right) \quad (57)$$

$$\left(E[x_i^2] + \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} E[x_i] \cdot E[x_j] \right) = \frac{1}{N} \left(N \cdot E[x_i^2] \right) - \frac{2}{N^2} \left[N \cdot E[x_i^2] + N(N-1)m_x^2 \right] + \frac{1}{N^2} \left[N \cdot E[x_i^2] + N(N-1)m_x^2 \right] \quad (58)$$

$$= \frac{1}{N} \left(N \cdot E[x_i^2] \right) - \frac{2N}{N^2} \left(E[x_i^2] \right) - \frac{2N(N-1)}{N^2} m_x^2 + \frac{N}{N^2} E[x_i^2] + \frac{N(N-1)}{N^2} m_x^2 \quad (59)$$

$$= \frac{1}{N} \left(N \cdot E[x_i^2] \right) - \frac{2}{N} \left(E[x_i^2] \right) - \frac{2(N-1)}{N} m_x^2 + \frac{1}{N} E[x_i^2] + \frac{(N-1)}{N} m_x^2 \quad (60)$$

$$= \frac{1}{N} (N - E[x_i^2]) - \frac{1}{N} (E[x_i^2]) - \frac{(N-1)}{N} m_x^2 \quad (61)$$

$$= \frac{(N-1)}{N} (E[x_i^2]) - \frac{(N-1)}{N} m_x^2 \quad (62)$$

$$= \frac{(N-1)}{N} \sigma_x^2 \quad (63)$$

It is apparent from equation (63) that the mean value of the sample variance does not equal the variance and is thus biased. Note, however, for large N the mean of the sample variance asymptotically approaches the variance and the estimate virtually becomes unbiased. Next, to check for consistency, we will proceed to determine the variance of the estimate sample variance. For ease of understanding, assume that the process is zero mean, then letting

$$\psi = \hat{\sigma}_x^2 = \frac{1}{N} \sum_{i=0}^{N-1} x_i^2 \quad (64)$$

so that,

$$E[\psi^2] = \frac{1}{N^2} \sum_{i=1}^N \sum_{k=1}^N E[x_i^2 x_k^2] \quad (65)$$

$$= \frac{1}{N^2} \left[N E[x_n^4] + N(N-1) (E[x_n^2])^2 \right] \quad (66)$$

$$= \frac{1}{N} \left[E[x_n^4] + (N-1)(E[x_n^2])^2 \right] \quad (67)$$

the expected value

$$E[\psi] = E[x_n^2] \quad (68)$$

so finally

$$\text{var}[\hat{\sigma}_x^2] = E[\psi^2] - (E[\psi])^2 \quad (69)$$

$$= \frac{1}{N} \left[E[x_n^4] - (E[x_n^2])^2 \right] \quad (70)$$

Re-examining equations (63) and (70) as N becomes large clearly indicates that the sample variance is a consistent estimate. Qualitatively speaking, the accuracy of this estimate depends upon the number of samples considered in the estimate. Note also that the procedures employed above typify the style of analysis used to characterize estimators.

6.0 THE PERIODOGRAM

The first method defines the sampled version of the Wiener-Khinchine relation, equations (31) and (32), where the power spectral density estimate $S_{N_{xx}}(f)$ is the discrete Fourier transform of the autocorrelation estimate $R_{N_{xx}}(k)$ or

$$S_{N_{xx}}(f) = \sum_{k=-\infty}^{\infty} R_{N_{xx}}(k) e^{-j\omega kT} \quad (71)$$

This assumes that $x(n)$ is a discrete time random process with an autocorrelation function $R_{N_{xx}}(k)$.

For a finite sequence of data

$$x(n) = \begin{cases} x_n & \text{for } n = 0, 1, \dots, N-1 \\ 0 & \text{elsewhere} \end{cases} \quad (72)$$

called a rectangular data window, the sample autocorrelation function (sampled form of equation A.14-9)

$$R_{N_{xx}}(k) = \frac{1}{N} \sum_{n=-\infty}^{\infty} x(n)x(n+k) \quad (73)$$

can be substituted into equation (71) to give the spectral density estimate

$$S_{N_{xx}}(f) = \frac{1}{N} |X_N(f)|^2 \quad (74)$$

called the periodogram.

$$\left(\text{Note: } \frac{|X_N(f)|^2}{N} = \frac{X_N(f)X_N^*(f)}{N} = \frac{X_N^2(f)_{\text{real}} + X_N^2(f)_{\text{imag}}}{N} \right. \\ \left. = F \left[R_{N_{xx}}(k) \right] = F \left[E[x(n)x(n+k)] \right] \right).$$

Hence,

$$S_{N_{xx}}(f) = \sum_{k=-\infty}^{\infty} R_{N_{xx}}(k) e^{-j\omega kT} \quad (75)$$

$$= \sum_{k=-\infty}^{\infty} \left[\frac{1}{N} \sum_{n=-\infty}^{\infty} x(n)x(n+k) \right] e^{-j\omega kT} \quad (76)$$

so letting $1 = e^{j\omega nT} e^{-j\omega nT}$

$$S_{N_{xx}}(f) = \frac{1}{N} \left(\sum_{n=-\infty}^{\infty} x(n) e^{j\omega nT} \sum_{k=-\infty}^{\infty} x(n+k) e^{-j\omega(n+k)T} \right) \quad (77)$$

and allowing the variable $m = n + k$

$$S_{N_{xx}}(f) = \frac{1}{N} X_N(f) X_N^*(f) = \frac{1}{N} |X_N(f)|^2 \quad (78)$$

in which the Fourier transform of the signal is

$$X_N(f) = \sum_{n=-\infty}^{\infty} x(n) e^{-j\omega nT} \quad (79)$$

The current discussion leads one to believe that the periodogram is an excellent estimator of the true power spectral density $S(f)$ as N becomes large. This conclusion is false and shortly it will be verified that the periodogram is, in fact, a poor estimate of $S(f)$. To do this, both the expected value and variance of $S_{N_{xx}}(f)$ will be checked for consistency as N becomes large. As a side comment it is generally faster to determine the power spectral density, $S_{N_{xx}}(f)$, of the random process using equation (74) and then inverse Fourier transforming to find $R_{N_{xx}}(k)$ than to obtain $R_{N_{xx}}(k)$ directly. Further, since the periodogram is seen to be the magnitude squared of the Fourier transformed data divided by N , the power spectral density of the random process is unrelated to the angle of the complex Fourier transform $X_N(f)$ of a typical realization.

Prior to checking for the consistency of $S_{N_{xx}}(f)$, the sample autocorrelation must initially be found consistent. Proceeding, since the sample autocorrelation estimate

$$R_{N_{xx}}(k) = \frac{x(0)x(k) + x(1)x(k+1) + \dots + x(N-1-k)x(N-1-k)}{N} \quad (80)$$

$$= \frac{1}{N} \sum_{n=0}^{N-1-|k|} x(n)x(n+|k|) \quad (81)$$

$$k = 0, \pm 1, \pm 2, \dots, \pm N-1$$

which averages together all possible products of samples separated by a lag of k , then, the mean value of the estimate is related to the true autocorrelation function by

$$E[R_{N_{xx}}(k)] = \left(\frac{1}{N} \sum_{n=0}^{N-1-|k|} E[x(n)x(n+|k|)] \right) \quad (82)$$

$$= \frac{N-|k|}{N} R(k)$$

where the true autocorrelation function $R(k)$ is defined as (the sample equivalent of equation A.14-8)

$$R(k) = E[x(n)x(n+k)] \quad (83)$$

From equation (82) it is observed that $R_{N_{xx}}(k)$ is a biased estimator. It is also considered to be asymptotically unbiased since the term

$$\frac{N-|k|}{N}$$

approaches 1 as N becomes large. From these observations $R_{N_{xx}}(k)$ can be classified as a good estimator of $R(k)$. In addition to having a small bias, a good estimator should also have a small variance. The variance of the sample autocorrelation function can thus be computed as

$$\text{var}[R_{N_{xx}}(k)] = E[R_{N_{xx}}^2(k)] - E^2[R_{N_{xx}}(k)] \quad (84)$$

Examining the $E[R_{N_{xx}}(k)]$ term of equation (84), substituting the estimate of equation (81) and replacing n with m , it follows that

$$E[R_{N_{xx}}^2(k)] = E \left\{ \left[\frac{1}{N} \sum_{n=0}^{N-1-|k|} x(n)x(n+|k|) \right] \right\} \quad (85)$$

$$= \frac{1}{N^2} \left(\sum_{n=0}^{N-1-|k|} \sum_{m=0}^{N-1-|k|} E[x(n)x(n+|k|)x(m)x(m+|k|)] \right) \quad (86)$$

$$E[x(n)x(n+|k|)x(m)x(m+|k|)]$$

If the statistical assumption that $x(n)$ is a zero-mean Gaussian process, then the zero-mean, jointly Gaussian, random variables symbolized as X_1, X_2, X_3 and X_4 of equation (86) can be described as [Ref. (30)].

$$E[X_1 X_2 X_3 X_4] = E[X_1 X_2] E[X_3 X_4] + E[X_1 X_3] E[X_2 X_4] + E[X_1 X_4] E[X_2 X_3] \quad (87)$$

$$= \left[E[x(n)x(n+|k|)] E[x(m)x(m+|k|)] + E[x(n)x(m)] E[x(n+|k|)x(m+|k|)] + E[x(n)x(m+|k|)] E[x(n+|k|)x(m)] \right] \quad (88)$$

Using equation (88), equation (84) becomes

$$\text{Var}[R_{N_{xx}}(k)] = \left\{ \frac{1}{N^2} \sum_{n=0}^{N-1-|k|} \sum_{m=0}^{N-1-|k|} \right. \quad (89)$$

$$\left. \begin{aligned} & R_{N_{xx}}(k) R_{N_{xx}}(k) + R_{N_{xx}}(n-m) R_{N_{xx}}(n-m) \\ & + R_{N_{xx}}(n-m-|k|) R_{N_{xx}}(n-m+|k|) \end{aligned} \right\} - \left[\frac{1}{N} \sum_{n=0}^{N-1-|k|} R_{N_{xx}}(k) \right]^2$$

$$\text{Var}[R_{N_{xx}}(k)] = \frac{1}{N^2} \sum_{n=0}^{N-1-|k|} \sum_{m=0}^{N-1-|k|} \left\{ R_{N_{xx}}^2(n-m) + R_{N_{xx}}(n-m-k) R_{N_{xx}}(n-m+|k|) \right\} \quad (90)$$

where the lag term $n - m$ was obtained from the lag difference between $\tau = n - m = (n + k) - (m + k)$ in the second term of equation (88). The lag term $n - k + m$ and $n - k - m$ was obtained by referencing the third term in equation (88) to n , therefore for

$$E[x(n) x(m + |k|)] \quad (91)$$

the lag term $\tau = n - (m + |k|)$ so

$$E[x(n) x(m + |k|)] = R_{N_{xx}}(n - m + |k|) \quad (92)$$

and for

$$E[x(n + |k|) x(m)] \quad (93)$$

first let $n - m$ then add $|k|$ so $\tau = n - m + |k|$ and

$$E[x(n + |k|) x(m)] = R_{N_{xx}}(n - m + |k|) \quad (94)$$

Recall that a sufficient condition for an estimator to be consistent is that its bias and variance both converge to zero as N becomes infinite. This essentially says that an estimator is consistent if it converges in probability to the true value of the quantity being estimated as N approaches infinity.

Re-examining equation (90), the variance of $R_{N_{xx}}(k)$, and equation (82), the expected value of $R_{N_{xx}}(k)$, it is found that both equations tend toward zero for large N and therefore $R_{N_{xx}}(k)$ is considered as a consistent estimator of $R(k)$ for fixed finite k in practical applications.

Having established that the autocorrelation estimate is consistent, we return to the question of the periodogram consistency.

At first glance, it might seem obvious that $S_{N_{xx}}(f)$ should inherit the asymptotically unbiased and consistent properties of $R_{N_{xx}}(k)$, of which it is a Fourier transform. Unfortunately, however, it will be shown that $S_{N_{xx}}(f)$ does not possess these nice statistical properties.

Going back to the power spectral density of equation (71).

$$S_{N_{xx}}(f) = \sum_{k=-\infty}^{\infty} R_{N_{xx}}(k) e^{-j\omega kT}$$

and determining its expected value

$$E[S_{N_{xx}}(f)] = \sum_{k=-\infty}^{\infty} E[R_{N_{xx}}(k)] e^{-j\omega kT} \quad (95)$$

the substitution of equation (82) into equation (95) yields the mean value estimate

$$E[S_{N_{xx}}(f)] = \sum_{k=-N}^N R(k) \left(1 - \frac{|k|}{N}\right) e^{-j\omega kT} \quad (96)$$

the $\left(1 - \frac{|k|}{N}\right)$

term of equation (96) can be interpreted as $a(k)$, the triangular window resulting from the autocorrelation of finite-sequence rectangular-data window $\omega(k)$ of equation (72). Thus,

$$a(k) = \begin{cases} 1 - \frac{|k|}{N} & |k| < N - 1 \\ 0 & \text{elsewhere} \end{cases} \quad (97a) \quad (97b)$$

and the expected value of the periodogram can be written as the finite sum

$$E[S_{N_{xx}}(f)] = \sum_{k=-\infty}^{\infty} R_{N_{xx}}(k) a(k) e^{-j\omega kT} \quad (98)$$

Note from equation (98) that the periodogram mean is the discrete Fourier transform of a product of the true autocorrelation function and a triangular window function. This frequency function can be expressed entirely in the frequency domain by the convolution integral. From equation (98), then, the convolution expression for the mean power spectral density is thus,

$$E[S_{N_{xx}}(f)] = \int_{-1/2}^{1/2} S(\eta) A(f - \eta) d\eta \quad (99)$$

where the general frequency expression for the transformed triangular window function $A(f)$ is

$$A(f) = \frac{1}{N} \left[\frac{\sin(2\pi f \frac{N}{2})}{\sin \frac{(2\pi f)}{2}} \right]^2 \quad (100)$$

Re-examining equation (98) or (96) it can be said that equation (71) or (74) gives an asymptotically unbiased estimate of $S(f)$ with the distorting effects of $a(k)$ vanishing as N tends toward infinity. At this point equation (98) still appears as a good estimate of the power spectral density function. For the variance $\text{var}[S_{N_{xx}}(f)]$ however, it can be shown [Ref. (10)] that if the data sequence $x(n)$ comes from a Gaussian process then the variance of $S_{N_{xx}}(f)$ approaches the square of the true spectrum, $S^2(f)$, at each frequency f . Hence, the variance is not small for increasing N ,

$$\lim_{N \rightarrow \infty} \text{var}[S_{N_{xx}}(f)] = S^2(f) \quad (101)$$

More clearly, if the ratio of mean to standard deviation is used as a kind of signal-to-noise ratio, i.e.

$$\frac{E[S_{N_{xx}}(f)]}{\{\text{var}[S_{N_{xx}}(f)]\}^{1/2}} \cong \frac{S(f)}{S(f)} = 1 \quad (102)$$

it can be seen that the true signal spectrum is only as large as the noise or uncertainty in $S_{N_{xx}}(f)$ for increasing N . In addition, the variance of equation (101), which also is approximately applicable for non-Gaussian sequences, indicates that calculations using different sets of N samples from the same $x(n)$ process will yield vastly different values of $S_{N_{xx}}(f)$ even when N becomes large. Unfortunately, since the variance of $S_{N_{xx}}(f)$ does not decrease to zero as N approaches infinity, the periodogram is thus an inconsistent estimate of the power spectral density and cannot be used for spectrum analysis in its present form.

7.0 SPECTRAL ESTIMATION BY AVERAGING PERIODOGRAMS

It was shown in the last section that the periodogram was not a consistent estimate of the power spectral density

function. A technique introduced by Bartlett, however, allows the use of the periodogram and, in fact, produces a consistent spectral estimation by averaging periodograms. In short, Bartlett's approach reduces the variance of the estimates by averaging together several independent periodograms. If, for example $X_1, X_2, X_3, \dots, X_L$ are uncorrelated random variables having an expected value $E[x]$ and a variance σ^2 , then the arithmetic mean

$$\frac{X_1 + X_2 + X_3 + \dots + X_L}{L} \quad (103)$$

has the expected value $E[x]$ and a variance of σ^2/L . This fact suggests that a spectral estimator can have its variance reduced by a factor of L over the periodogram. The procedure requires the observed process, an N point data sequence, to be split up into L nonoverlapping M point sections and then averaging the periodograms of each individual section.

To be more specific, dividing an N point data sequence $x(n)$, $0 \leq n \leq N-1$, into L segments of M samples each the segments $x_M^\ell(n)$ are formed. Thus,

$$x_M^\ell(n) = x(n + \ell M - M) \quad \begin{cases} 0 \leq n \leq M-1 \\ 1 \leq \ell \leq L \end{cases} \quad (104)$$

where the superscript ℓ specifies the segment or interval of data being observed, the subscript M represents the number of data points or samples per segment and depending upon the choice of L and M , we have $N \geq LM$. For the computation of L periodograms

$$S_M^\ell(f) = \left| \frac{1}{M} \sum_{n=0}^{M-1} x_M^\ell(n) e^{-j\omega n T} \right|^2 \quad 1 \leq \ell \leq L \quad (105)$$

If the autocorrelation function $R_{N_{xx}}(m)$ becomes negligible for m large relative to M , $m > M$, then it can be said that the periodograms of the separate sections are virtually independent of one another. The corresponding *averaged periodogram estimator* $\hat{S}_M^\ell(f)$ computed from L individual periodograms of length M is thus defined

$$\hat{S}_M^\ell(f) = \frac{1}{L} \sum_{\ell=1}^L S_M^\ell(f) \quad (106)$$

Since the L subsidiary estimates are identically distributed periodograms, the averaged spectral estimate will have the same mean or expected value as any of the subsidiary estimates so

$$E[\hat{S}_M^\ell(f)] = \frac{1}{L} \sum_{\ell=1}^L E[S_M^\ell(f)] \quad (107)$$

$$= E[S_M^\ell(f)] \quad (108)$$

From this, the expected value of the Bartlett estimate can be said to be the convolution of the true spectral density with the Fourier transform of the triangular window function corresponding to the M sample periodogram where $M \leq N/L$ equations (98) or (99) we see that

$$E[\hat{S}_M^\ell(f)] = E[S_M^\ell(f)] = \frac{1}{M} \int_{-1/2}^{1/2} S(\eta) A(f - \eta) d\eta \quad (109)$$

where $A(f)$ is the Fourier transformed triangular window function of equation (100). Though the averaged estimate is no different than before, its variance, however, is smaller. Recall that the variance of the average of L identical independent random variables is $1/L$ of the individual variances, equation (51). Thus, for L statistically independent periodograms, the variance of the averaged estimate is

$$\text{var}[\hat{S}_M^\ell(f)] = \frac{1}{L} \text{var}[S_{N_{xx}}(f)] \approx \frac{1}{L} [S(f)]^2 \quad (110)$$

So, again, the averaging of L periodograms results in approximately a factor of L reduction in power spectral density estimation variance. Since the variance of equation (110) tends to zero as L approaches infinity and through equation (98) and (99) $\hat{S}_M^\ell(f)$ is asymptotically unbiased, $\hat{S}_M^\ell(f)$ can be said to be a consistent estimate of the true spectrum.

A few notes are next in order. First, the L fold variance reduction or $(L)^{1/2}$ signal-to-noise ratio improvement of equation (102) is not precisely accurate since there is some dependence between the subsidiary periodograms. The adjacent samples will be correlated unless the process being analyzed is white.

However, as indicated in equation (110), such a dependence will be small when there are many sample intervals per periodogram so that the reduced variance is still a good approximation. Secondly, the bias of $\hat{S}_M^\ell(f)$, equation (106), is greater than $\hat{S}_M^\ell(f)$, equation (105), since the main lobe of the spectral window is larger for the former. For this situation, then, the bias can be thought of as effecting spectral resolution. It is seen that increasing the number of periodograms for a fixed record length N decreases not only the variance but, the samples per periodograms M decrease also. This decreases the spectral resolution. Thus when using the Bartlett procedure the actual choice of M and N will typically be selected from prior knowledge of a signal or data sequence under consideration. The tradeoff, however, will be between the spectral resolution of bias and the variance of the estimate.

8.0 WINDOWS

Prior to looking at other techniques of spectrum estimation, we find that to this point the subject of spectral windows has been brought up several times during the course of our discussion. No elaboration, however, has been spent explaining their spectral effects and meaning. It is thus appropriate at this juncture to diverge for a short while to develop a fundamental understanding of windows and their spectral implications prior to the discussion of Sections 9 and 10 (for an in depth study of windows see *Windows, Harmonic Analysis and the Discrete Fourier Transform*; Frederic J. Harris; submitted to IEEE Proceedings, August 1976).

In most applications it is desirable to taper a finite length data sequence at each end to enhance certain characteristics of the spectral estimate. The process of terminating a sequence after a finite number of terms can be thought of as multiplying an infinite length, i.e., impulse response sequence by a finite width window function. In other words, the window function determines how much of the original impulse sequence can be observed through this window,

see Figures 4a, 4b, and 4c. This tapering by multiplying the sequence by a data window is thus analogous to multiplying the correlation function by a lag window. In addition, since multiplication in the time domain is equivalent to convolution in the frequency domain then it is also analogous to convolving the Fourier transform of a finite-length-sequence with the Fourier transform of the window function, Figures 4d, 4e, and 4f. Note also that the Fourier transform of the rectangular window function exhibits significant oscillations and poor high frequency convergence, Figure 4e. Thus, when convolving this spectrum with a desired amplitude response, poor convergence of the resulting amplitude response may occur. This calls for investigating the use of other possible window functions that minimize some of the difficulties encountered with rectangular function.

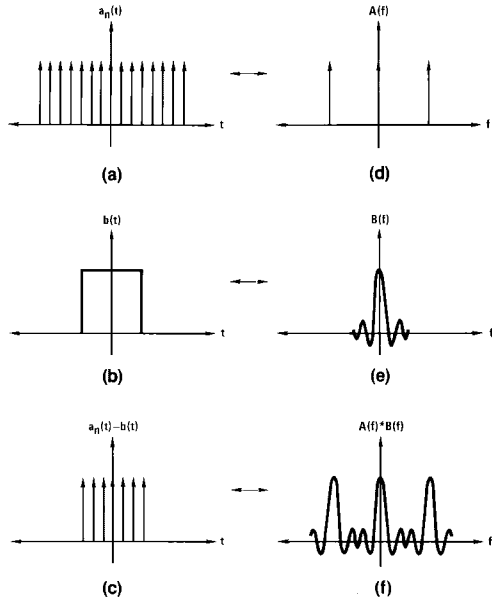


FIGURE 4

In order for the spectrum of a window function to have minimal effects upon the desired amplitude response, resulting from convolving two functions, it is necessary that the window spectrum approximate an impulse function. This implies that as much of its energy as possible should be concentrated at the center of the spectrum. Clearly, an ideal impulse spectrum is not feasible since this requires an infinitely long window.

In general terms, the spectrum of a window function typically consists of a main lobe, representing the center of the spectrum, and various side lobes, located on either side of the main lobe (see Figures 6 thru 9). It is desired that the window function satisfy two criteria; (1) that the main lobe should be as narrow as possible and (2) relative to the

main lobe, the maximum side lobe level should be as small as possible. Unfortunately, however, both conditions cannot be simultaneously optimized so that, in practice, usable window functions represent a suitable compromise between the two criteria. A window function in which minimization of the main lobe width is the primary objective, fields a finer frequency resolution but suffers from some oscillations, i.e., the spectrum passband and substantial ripple in the spectrum stopband. Conversely, a window function in which minimization of the side lobe level is of primary concern tends to have a smoother amplitude response and very low ripple in the stopband but, yields a much poorer frequency resolution. Examining Figure 5 assume a hypothetical impulse response, Figure 5a, whose spectrum is Figure 5b. Multiplying the impulse response by the rectangular window, Figure 4b, yields the windowed impulse response, Figure 5c, implying the convolution of the window spectrum, Figure 4e, with the impulse response spectrum, Figure 5b. The result of this convolution is seen in Figure 5d and is a distorted version of the ideal spectrum, Figure 5b, having passband oscillations and stopband ripple. Selecting another window, i.e., Figure 9 with more desirable spectral characteristics, we see the appropriately modified windowed data, Figure 5e, results in a very good approximation of Figure 5b.

This characteristically provides a smoother passband and lower stopband ripple level but sacrifices the sharpness of the roll-off rate inherent in the use of a rectangular window (compare Figures 5d and 5f). Concluding this brief discussion, a few common window functions will next be considered.

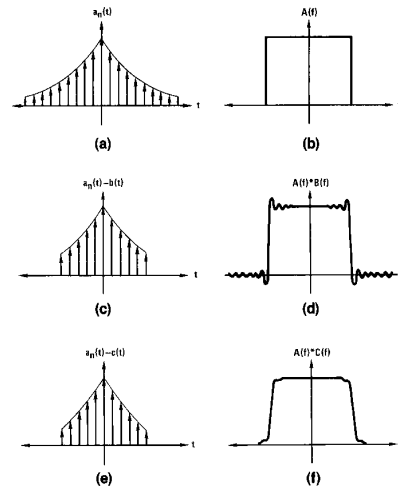


FIGURE 5. (a)(b) Unmodified Data Sequence
(c)(d) Rectangular Windowed Data Sequence
(e)(f) Hamming Windowed Data Sequence

Rectangular window: *Figure 6*

$$w(n) = 1 \quad |n| < \frac{N-1}{2} \quad (111)$$

$$= 0 \quad \text{otherwise}$$

Bartlett or triangular window: *Figure 7*

$$w(n) = 1 - \frac{2|n|}{N} \quad |n| < \frac{N-1}{2} \quad (112)$$

$$= 0 \quad \text{otherwise}$$

Hann window: *Figure 8*

$$w(n) = 0.5 + 0.5 \cos\left(\frac{2\pi n}{N}\right) \quad |n| < \frac{N-1}{2} \quad (113)$$

$$= 0 \quad \text{otherwise}$$

Hamming window: *Figure 9*

$$w(n) = 0.54 + 0.46 \cos\left(\frac{2\pi n}{N}\right) \quad |n| < \frac{N-1}{2} \quad (114)$$

$$= 0 \quad \text{otherwise}$$

Again the reference previously cited should provide a more detailed window selection. Nevertheless, the final window choice will be a compromise between spectral resolution and passband (stopband) distortion.

9.0 SPECTRAL ESTIMATION BY USING WINDOWS TO SMOOTH A SINGLE PERIODOGRAM

It was seen in a previous section that the variance of a power spectral density estimate based on an N point data sequence could be reduced by chopping the data into shorter segments and then averaging the periodograms of the individual segments. This reduced variance was acquired at the expense of increased bias and decreased spectral resolution. We will cover in this section an alternate way of computing a reduced variance estimate using a smoothing operation on the single periodogram obtained from the entire N point data sequence. In effect, the periodogram is smoothed by convolving it with an appropriate spectral window. Hence if $S_{XX}(f)$ denotes the smooth periodogram then,

$$S_{WXX}(f) = \int_{-1/2}^{1/2} S_{NXX}(\eta) W(f-\eta) d\eta = S_{NXX}(\eta) * W(\eta) \quad (115)$$

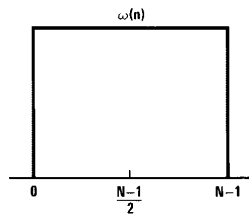


FIGURE 6. Rectangular Window

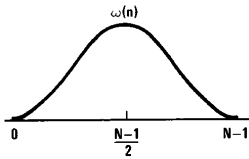


FIGURE 7. Bartlett or Triangular Window

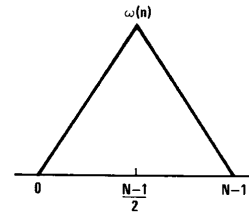
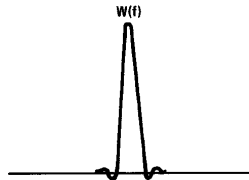


FIGURE 8. Hann Window

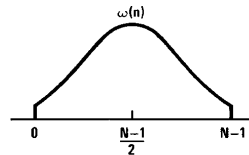


FIGURE 9. Hamming Window

where $W(f - \eta)$ is the spectral window and $*$ denotes convolution. Since the periodogram is equivalent to the Fourier transform of the autocorrelation function $R_{N_{xx}}(k)$ then, using the frequency convolution theorem

$$F\{x(t)y(t)\} = X(f) * Y(\eta - f) \quad (116)$$

where $F\{\cdot\}$ denotes a Fourier transform, $S_{xx}(f)$ is the Fourier transform of the product of $R_{N_{xx}}(k)$ and the inverse Fourier transform of $W(f)$. Therefore for a finite duration window sequence of length $2K - 1$,

$$S_{W_{xx}}(f) = \sum_{k=-(K-1)}^{K-1} R_{N_{xx}}(k) w(k) e^{-j\omega kT} \quad (117)$$

where

$$w(k) = \sum_{-1/2}^{1/2} W(f) e^{j\omega kT} df \quad (118)$$

References (10)(16)(21) proceed further with a development to show that the smoothed single windowed periodogram is a consistent estimate of the power spectral density function. The highlights of the development, however, show that a smoothed or windowed spectral estimate, $S_{W_{xx}}(f)$, can be made to have a smaller variance than that of the straight periodogram, $S_{N_{xx}}(f)$, by β the variance ratio relationship

$$\beta = \frac{\text{var}[S_{W_{xx}}(f)]}{\text{var}[S_{N_{xx}}(f)]} = \frac{1}{N} \sum_{m=-(M-1)}^{M-1} w^2(m) \quad (119)$$

$$= \frac{1}{N} \int_{-1/2}^{1/2} W^2(f) df$$

where N is the record length and $2M - 1$ is the total window width. Note that the energy of the window function can be found in equation (119) as

$$E_w = \sum_{m=-(M-1)}^{M-1} w^2(m) = \int_{-1/2}^{1/2} W^2(f) df \quad (120)$$

Continuing from equation (119), it is seen that a satisfactory estimate of the periodogram requires the variance of $S_{W_{xx}}(f)$ to be small compared to $S_{N_{xx}}^2$ so that

$$\beta \ll 1 \quad (121)$$

Therefore, it is the adjusting of the length and shape of the window that allows the variance of $S_{W_{xx}}(f)$ to be reduced over the periodogram.

Smoothing is like a low pass filtering effect, and so, causes a reduction in frequency resolution. Further, the width of the window main lobe, defined as the symmetric interval between the first positive and negative frequencies at which $W(f) = 0$, in effect determines the bandwidth of the smoothed spectrum. Examining Table I for the following defined windows;

Rectangular window

$$w(m) = 1 \quad |m| \leq M - 1 \quad (122)$$

$$= 0 \quad \text{otherwise}$$

Bartlett or triangular window

$$w(m) = 1 - \frac{|m|}{M} \quad |m| \leq M - 1 \quad (123)$$

$$= 0 \quad \text{otherwise}$$

Hann window

$$w(m) = 0.5 + 0.5 \cos\left(\frac{\pi m}{M - 1}\right) \quad |m| \leq M - 1 \quad (124)$$

$$= 0 \quad \text{otherwise}$$

Hamming window

$$w(m) = 0.54 + 0.46 \cos\left(\frac{\pi m}{M - 1}\right) \quad |m| \leq M - 1 \quad (125)$$

$$= 0 \quad \text{otherwise}$$

We see once again, as in the averaging technique of spectral estimation (Section 7), the smoothed periodogram technique of this discussion also makes a trade-off between variance and resolution for a fixed N . A small variance requires a small M while high resolution requires a large M .

TABLE I

Window Function	Width of Main Lobe* (approximate)	Variance Ratio* (approximate)
Rectangular	$\frac{2\pi}{M}$	$\frac{2M}{N}$
Bartlett	$\frac{4\pi}{M}$	$\frac{2M}{3N}$
Hann	$\frac{3\pi}{M}$	$2M \left[\frac{(0.5)^2 + (0.5)^2}{2N} \right]$
Hamming	$\frac{3\pi}{M}$	$2M \left[\frac{(0.54)^2 + (0.46)^2}{2N} \right]$

* Assumes $M \gg 1$

10.0 SPECTRAL ESTIMATION BY AVERAGING MODIFIED PERIODOGRAMS

Welch [Ref. (36)(37)] suggests a method for measuring power spectra which is effectively a modified form of Bartlett's method covered in Section 7. This method, however, applies the window $w(n)$ directly to the data segments before computing their individual periodograms. If the data sequence is sectioned into

$$L \leq \frac{N}{M}$$

segments of M samples each as defined in equation (104), the L modified or windowed periodogram can be defined as

$$I_M^\ell(f) = \frac{1}{UM} \left| \sum_{n=0}^{M-1} x_M^\ell(n) w(n) e^{-j\omega nT} \right|^2 \quad 1 \leq \ell \leq L \quad (126)$$

where U , the energy in the window is

$$U = \frac{1}{M} \sum_{n=0}^{M-1} w^2(n) \quad (127)$$

Note the similarity between equations (126) and (105) and that equation (126) is equal to equation (105) modified by functions of the data window $w(n)$.

The spectral estimate $\hat{I}_M^\ell$ is defined as

$$\hat{I}_M^\ell(f) = \frac{1}{L} \sum_{\ell=1}^L I_M^\ell(f) \quad (128)$$

and its expected value is given by

$$E[\hat{I}_M^\ell(f)] = \int_{-1/2}^{1/2} S_{N_{xx}}(\eta) W(f-\eta) d\eta = S_{N_{xx}}(\eta) * W(\eta) \quad (129)$$

where

$$W(f) = \frac{1}{UM} \left| \sum_{n=0}^{M-1} w(n) e^{-j\omega nT} \right|^2 \quad (130)$$

The normalizing factor U is required so that the spectral estimate $\hat{I}_M^\ell(f)$, of the modified periodogram, $I_M^\ell(f)$, will be asymptotically unbiased [Ref. (34)]. If the intervals of $x(n)$ were to be nonoverlapping, then Welch [Ref. (37)] indicates that

$$\text{var}[\hat{I}_M^\ell(f)] \cong \frac{1}{L} \text{var}[S_{N_{xx}}(f)] \cong \frac{1}{L} [S(f)]^2 \quad (131)$$

which is similar to that of equation (110). Also considered is a case where the data segments overlap. As the overlap increases the correlation between the individual periodograms also increase. We see further that the number of M point data segments that can be formed increases. These two effects counteract each other when considering the variance $I_M^\ell(f)$. With a good data window, however, the increased number of segments has the stronger effect until the overlap becomes too large. Welch has suggested that a 50 percent overlap is a reasonable choice for reducing the variance when N is fixed and cannot be made arbitrarily large. Thus, along with windowing the data segments prior to computing their periodograms, achieves a variance reduction over the Bartlett technique and at the same time smooths the spectrum at the cost of reducing its resolution. This trade-off between variance and spectral resolution or bias is an inherent property of spectrum estimators.

11.0 PROCEDURES FOR POWER SPECTRAL DENSITY ESTIMATES

Smoothed single periodograms [Ref. (21)(27)(32)]

1. Determine the sample autocorrelation $R_{N_{xx}}(k)$ of the data sequence $x(n)$
2. Multiply $R_{N_{xx}}(k)$ by an appropriate window function $w(n)$
3. Compute the Fourier transform of the product

$$R_{N_{xx}}(k) w(n) \longleftrightarrow S_{W_{xx}}(f) \quad (71)$$

1. Compute the Fourier transform of the data sequence $x(n)$

$$x(n) \longleftrightarrow X(f)$$

2. Multiply $X(f)$ by its conjugate to obtain the power spectral density $S_{N_{xx}}(f)$

$$S_{N_{xx}}(f) = \frac{1}{N} |X(f)|^2 \quad (74)$$

3. Convolve $S_{N_{xx}}(f)$ with an appropriate window function $W(f)$

$$S_{W_{xx}}(f) = S_{N_{xx}}(f) * W(f) \quad (115)$$

1. Compute the Fourier transform of the data sequence $x(n)$

$$x(n) \longleftrightarrow X(f)$$

2. Multiply $X(f)$ by its conjugate to obtain the power spectral density $S_{N_{xx}}(f)$

$$S_{N_{xx}}(f) = \frac{1}{N} |X(f)|^2 \quad (74)$$

3. Inverse Fourier transform $S_{N_{xx}}(f)$ to get $R_{N_{xx}}(k)$
4. Multiply $R_{N_{xx}}(k)$ by an appropriate window function $w(n)$
5. Compute the Fourier transform of the product to obtain $S_{W_{xx}}(f)$

$$S_{W_{xx}}(f) \longleftrightarrow R_{N_{xx}}(k) \bullet w(n) \quad (117)$$

Averaging periodograms [Ref. (32)(37)(38)]

1. Divide the data sequence $x(n)$ into $L \leq N/M$ segments, $x_M^\ell(n)$
2. Multiply a segment by an appropriate window
3. Take the Fourier transform of the product
4. Multiply procedure 3 by its conjugate to obtain the spectral density of the segment
5. Repeat procedures 2 through 4 for each segment so that the average of these periodogram estimates produce the power spectral density estimate, equation (128)

12.0 RESOLUTION

In analog bandpass filters, resolution is defined by the filter bandwidth, Δf_{BW} , measured at the passband half power points. Similarly, in the measurement of power spectral density functions it is important to be aware of the resolution capabilities of the sampled data system. For such a system resolution is defined as

$$\Delta f_{BW} = \frac{1}{NT} \quad (132)$$

and for;

$$\text{Correlation resolution} \quad \tau_{\max} = mT, \text{ where } m \quad (133)$$

is the maximum value allowed to produce $\tau_{\max}$, the maximum lag term in the correlation computation

Fourier transform (FFT) resolution

$$\Delta f_{BW} = \frac{1}{P_L} = \frac{m}{NT} = \frac{1}{LT} \quad \text{where } P \text{ is the record length, } N, \text{ the number of data points and } m, \text{ the samples within each } L \text{ segment,}$$

$$L = \frac{N}{M}. \quad (134)$$

Note that the above Δf_{BW} 's can be substantially poorer depending upon the choice of the data window. A loss in degrees of freedom (Section 13) and statistical accuracy occurs when a data sequence is windowed. That is, data at each end of a record length are given less weight than the data at the middle for anything other than the rectangular window. This loss in degrees of freedom shows up as a loss in resolution because the main lobe of the spectral window is widened in the frequency domain.

Finally, the limits of a sampled data system can be described in terms of the maximum frequency range and minimum frequency resolution. The maximum frequency range is the Nyquist or folding frequency, f_c ,

$$f_c = \frac{f_s}{2} = \frac{1}{2T_s} \quad (135)$$

where f_s is the sampling frequency and T_s the sampling period. And, secondly, the resolution limit can be described in terms of a $(\Delta f_{BW}) NT$ product where

$$\Delta f_{BW} \geq \frac{1}{NT} \quad (136)$$

or

$$(\Delta f_{BW}) NT \geq 1 \quad (137)$$

13.0 CHI-SQUARE DISTRIBUTIONS

Statistical error is the uncertainty in power spectral density measurements due to the amount of data gathered, the probabilistic nature of the data and the method used in deriving the desired parameter. Under reasonable conditions the spectral density estimates approximately follow a chi-square, χ_n^2 , distribution. χ_n^2 is defined as the sum of the squares of χ_n , $1 \leq n \leq N$, independent Gaussian variables each with a zero mean and unit variance such that

$$\chi_N^2 = \sum_{n=1}^N \chi_n^2 \quad (138)$$

The number n is called the *degrees of freedom* and the χ_n^2 probability density function is

$$f(\chi_n^2) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} \left[\left(\frac{\chi_n^2}{2} \right)^{\frac{n-2}{2}} \right] e^{-\frac{\chi_n^2}{2}} \quad (139)$$

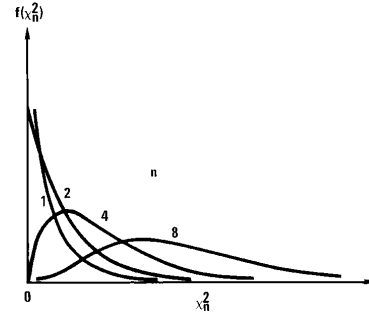
where $\Gamma(\frac{n}{2})$ is the statistical gamma function (Ref. (14)).

Figure 10 shows the probability density function for several n values and it is important to note that as n becomes large the chi-square distribution approximates a Gaussian distribution. We use this χ_n^2 distribution in our analysis to discuss the variability of power spectral densities. If χ_n has a zero mean and N samples of it are used to compute the power spectral density estimate $S(f)$ then, the probability that the true spectral density, $S(f)$, lies between the limits

$$A \leq S(f) \leq B \quad (140)$$

is

$$P = (1 - \alpha) = \text{probability} \quad (141)$$



TL/H/8712-10

FIGURE 10

The lower A and upper B limits are defined as

$$A = \frac{n\hat{S}(f)}{\chi_{n; \frac{\alpha}{2}}^2} \quad (142)$$

and

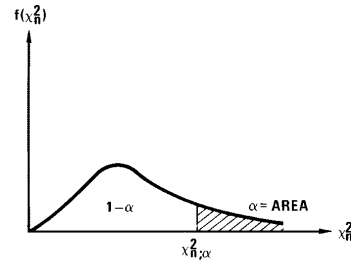
$$B = \frac{n\hat{S}(f)}{\chi_{n; 1 - \frac{\alpha}{2}}^2} \quad (143)$$

respectively. $\chi_{n; \alpha}^2$ is defined by

$$\chi_{n; \alpha}^2 = [\nu \text{ so that } \int_{\nu}^{\infty} f(\chi_n^2) d\chi_n^2 = \alpha] \quad (144)$$

see Figure 11 and the interval A to B is referred to as a confidence interval. From Otnes and Enrochson [Ref. (35) pg. 217] the degrees of freedom can be described as

$$n = 2(\Delta f_{BW}) NT = 2(\Delta f_{BW}) P_L \quad (145)$$



TL/H/8712-11

FIGURE 11

and that for large n i.e., $n \geq 30$ the χ_n^2 distribution approaches a Gaussian distribution so that the standard deviation or standard error, ϵ_o , can be given by

$$\epsilon_o = \frac{1}{\sqrt{\Delta f_{BW} NT}} \quad (146)$$

The degrees of freedom and associated standard error for the correlation and Fourier transform are as follows:

$$\text{correlation: } n = \frac{2N}{m} \quad \epsilon_o = \sqrt{\frac{m}{N}} \quad (147)$$

$$\text{FFT: } n = 2M \quad \epsilon_o = \sqrt{\frac{1}{M}} \quad (148)$$

where M is the number of $|X(f)|^2$ values

$$M = NT (\Delta f_{BW})_{\text{desired}} \quad (149)$$

and m is the maximum lag value.

An example will perhaps clarify the usage of this information.

Choosing $T = 100$ ms, $N = 8000$ samples and $n = 20$ degrees of freedom then

$$f_c = \frac{1}{2T} = 5 \text{ Hz}$$

$$n = 2(NT) (\Delta f_{BW})$$

so

$$\Delta f_{BW} = \frac{n}{2NT} = 0.0125 \text{ Hz}$$

If it is so desired to have a 95% confidence level of the spectral density estimate then

$$P = (1 - \alpha)$$

$$0.95 = 1 - \alpha$$

$$\alpha = 1 - 0.95 = 0.05$$

the limits

$$B = \frac{n\hat{S}(f)}{\chi_{n; 1 - \alpha/2}^2} = \frac{20\hat{S}(f)}{\chi_{20; 0.975}^2}$$

$$A = \frac{n\hat{S}(f)}{\chi_{n; \alpha/2}^2} = \frac{20\hat{S}(f)}{\chi_{20; 0.025}^2}$$

yield from Table II

$$\chi_{20; 0.975}^2 = 9.59$$

$$\chi_{20; 0.025}^2 = 34.17$$

so that

$$\frac{20\hat{S}(f)}{34.17} \leq S(f) \leq \frac{20\hat{S}(f)}{9.59}$$

$$0.5853 \hat{S}(f) \leq S(f) \leq 2.08 \hat{S}(f)$$

There is thus a 95% confidence level that the true spectral density function $S(f)$ lies within the interval $0.5853 \hat{S}(f) \leq S(f) \leq 2.08 \hat{S}(f)$.

As a second example using equation (148) let $T = 1$ ms, $N = 4000$ and it is desired to have $(\Delta f_{BW})_{\text{desired}} = 10$ Hz.

Then,

$$NT = 4$$

$$f_c = \frac{1}{2T} = 500 \text{ Hz}$$

$$\epsilon_o = \sqrt{\frac{1}{M}} = \sqrt{\frac{1}{NT (\Delta f_{BW})_{\text{desired}}}} = 0.158$$

$$N = 2M = 2NT (\Delta f_{BW})_{\text{desired}} = 80$$

If it is again desired to have a 95% confidence level of the spectral density estimate then

$$\alpha = 1 - p = 0.05$$

$$\chi_{80; 0.975}^2 = 5.75$$

$$\chi_{80; 0.025}^2 = 106.63$$

and we thus have a 95% confidence level that the true spectral density $S(f)$ lies within the limits

$$0.75 \hat{S}(f) \leq S(f) \leq 1.39 \hat{S}(f)$$

It is important to note that the above examples assume Gaussian and white data. In practical situations the data is typically colored or correlated and effectively results in reducing number of degrees of freedom. It is best, then, to use the white noise confidence levels as guidelines when planning power spectral density estimates.

14.0 CONCLUSION

This article attempted to introduce to the reader a conceptual overview of power spectral estimation. In doing so a wide variety of subjects were covered and it is hoped that this approach left the reader with a sufficient base of "tools" to indulge in the mounds of technical literature available on the subject.

15.0 ACKNOWLEDGEMENTS

The author wishes to thank James Moyer and Barry Siegel for their support and encouragement in the writing of this article.

TABLE II. Percentage Points of the Chi-Square Distribution

n	α							
	0.995	0.990	0.975	0.950	0.050	0.025	0.010	0.005
1	0.000039	0.00016	0.00098	0.0039	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.1030	5.99	7.38	9.21	10.60
3	0.0717	0.115	0.216	0.352	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.150	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	79.08	83.80	88.38	91.95
70	43.28	45.44	48.76	51.74	90.53	95.02	100.43	104.22
80	51.17	53.54	57.15	60.39	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	124.34	129.56	135.81	140.17

APPENDIX A

A.0 CONCEPTS OF PROBABILITY, RANDOM VARIABLES AND STOCHASTIC PROCESSES

In many physical phenomena the outcome of an experiment may result in fluctuations that are random and cannot be precisely predicted. It is impossible, for example, to determine whether a coin tossed into the air will land with its head side or tail side up. Performing the same experiment over a long period of time would yield data sufficient to indicate that on the average it is equally likely that a head or tail will turn up. Studying this average behavior of events allows one to determine the frequency of occurrence of the outcome (i.e., heads or tails) and is defined as the notion of *probability*.

Associated with the concept of probability are probability density functions and cumulative distribution functions which find their use in determining the outcome of a large number of events. A result of analyzing and studying these functions may indicate regularities enabling certain laws to be determined relating to the experiment and its outcomes; this is essentially known as *statistics*.

A.1 DEFINITIONS OF PROBABILITY

If n_A is the number of times that an event A occurs in N performances of an experiment, the frequency of occurrence of event A is thus the ratio n_A/N . Formally, the probability, $P(A)$, of event A occurring is defined as

$$P(A) = \lim_{N \rightarrow \infty} \left[\frac{n_A}{N} \right] \quad (\text{A.1-1})$$

Where it is seen that the ratio n_A/N (or fraction of times that an event occurs) asymptotically approaches some mean value (or will show little deviation from the exact probability) as the number of experiments performed, N, increases (more data).

Assigning a number,

$$\frac{n_A}{N},$$

to an event is a measure of how likely or probable the event. Since n_A and N are both positive and real numbers and $0 \leq n_A \leq N$; it follows that the probability of a given event cannot be less than zero or greater than unity. Furthermore, if the occurrence of any one event excludes the occurrence of any others (i.e., a head excludes the occurrence of a tail in a coin toss experiment), the possible events are said to be *mutually exclusive*. If a complete set of possible events A_1 to A_n are included then

$$\frac{n_{A_1}}{N} + \frac{n_{A_2}}{N} + \frac{n_{A_3}}{N} + \dots + \frac{n_{A_n}}{N} = 1 \quad (\text{A.1-2})$$

or

$$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1 \quad (\text{A.1-3})$$

Similarly, an event that is absolutely certain to occur has a probability of one and an impossible event has a probability of zero.

In summary:

1. $0 \leq P(A) \leq 1$
2. $P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) = 1$, for an entire set of events that are mutually exclusive
3. $P(A) = 0$ represents an impossible event
4. $P(A) = 1$ represents an absolutely certain event

A.2 JOINT PROBABILITY

If more than one event at a time occurs (i.e., events A and B are not mutually exclusive) the frequency of occurrence of the two or more events at the same time is called the *joint probability*, $P(AB)$. If n_{AB} is the number of times that event A and B occur together in N performances of an experiment, then

$$P(A,B) = \lim_{N \rightarrow \infty} \left[\frac{n_{AB}}{N} \right] \quad (\text{A.2-1})$$

A.3 CONDITIONAL PROBABILITY

The probability of event B occurring given that another event A has already occurred is called *conditional probability*. The dependence of the second, B, of the two events on the first, A, will be designated by the symbol $P(B|A)$ or

$$P(B|A) = \frac{n_{AB}}{n_A} \quad (\text{A.3-1})$$

where n_{AB} is the number of joint occurrences of A and B and N_A represents the number of occurrences of A with or without B. By dividing both the numerator and denominator of equation (A.3-1) by N, conditional probability $P(B|A)$ can be related to joint probability, equation (A.2-1), and the probability of a single event, equation (A.1-1)

$$P(B|A) = \left(\frac{n_{AB}}{n_A} \right) \left(\frac{\frac{1}{N}}{\frac{1}{N}} \right) = \frac{P(A,B)}{P(A)} \quad (\text{A.3-2})$$

Analogously

$$P(A|B) = \frac{P(A,B)}{P(B)} \quad (\text{A.3-3})$$

and combining equations (A.6) and (A.7)

$$P(A|B) P(B) = P(A, B) = P(B|A) P(A) \quad (\text{A.3-4})$$

results in Bayes' theorem

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (\text{A.3-5})$$

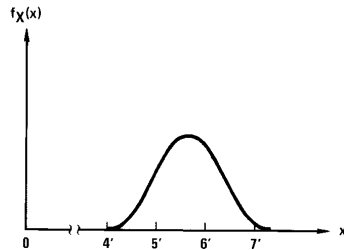
Using Bayes' theorem, it is realized that if $P(A)$ and $P(B)$ are *statistically independent* events, implying that the probability of event A does not depend upon whether or not event B has occurred, then $P(A|B) = P(A)$, $P(B|A) = P(B)$ and hence the joint probability of events A and B is the product of their individual probabilities or

$$P(A,B) = P(A) P(B) \quad (A.3-6)$$

More precisely, two random events are statistically independent only if equation (A.3-6) is true.

A.4 PROBABILITY DENSITY FUNCTIONS

A formula, table, histogram, or graphical representation of the probability or possible frequency of occurrence of an event associated with variable X, is defined as $f_X(x)$, the *probability density function* (pdf) or *probability distribution function*. As an example, a function corresponding to height histograms of men might have the probability distribution function depicted in Figure A.4.1.



TL/H/8712-12

FIGURE A.4.1

The *probability element*, $f_X(x) dx$, describes the probability of the event that the random variable X lies within a range of possible values between

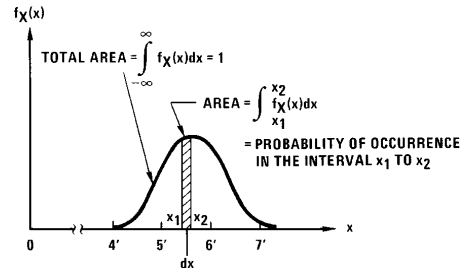
$$\left(x - \frac{\Delta x}{2}\right) \text{ and } \left(x + \frac{\Delta x}{2}\right)$$

i.e., the area between the two points 5'5" and 5'7" shown in Figure A.4.2 represents the probability that a man's height will be found in that range. More clearly,

$$\text{Prob} \left[\left(x - \frac{\Delta x}{2}\right) \leq X \leq \left(x + \frac{\Delta x}{2}\right) \right] = \int_{x - \frac{\Delta x}{2}}^{x + \frac{\Delta x}{2}} f_X(x) dx \quad (A.4-1)$$

or

$$\text{Prob} [5'5" \leq X \leq 5'7"] = \int_{5'5"}^{5'7"} f_X(x) dx$$



TL/H/8712-13

FIGURE A.4.2

Continuing, since the total of all probabilities of the random variable X must equal unity and $f_X(x) dx$ is the probability that X lies within a specified interval

$$\left(x - \frac{\Delta x}{2}\right) \text{ and } \left(x + \frac{\Delta x}{2}\right),$$

then,

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (A.4-2)$$

It is important to point out that the density function $f_X(x)$ is in fact a mathematical description of a curve and is not a probability; it is therefore not restricted to values less than unity but can have any non-negative value. Note however, that in practical application, the integral is normalized such that the entire area under the probability density curve equates to unity.

To summarize, a few properties of $f_X(x)$ are listed below.

1. $f_X(x) \geq 0$ for all values of x or $-\infty < x < \infty$

2. $\int_{-\infty}^{\infty} f_X(x) dx = 1$

3. $\text{Prob} \left[\left(x - \frac{\Delta x}{2}\right) \leq X \leq \left(x + \frac{\Delta x}{2}\right) \right]$

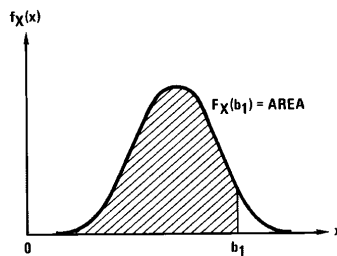
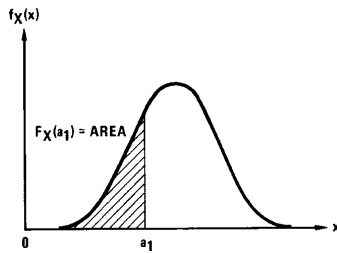
$$= \int_{x - \frac{\Delta x}{2}}^{x + \frac{\Delta x}{2}} f_X(x) dx$$

A.5 CUMULATIVE DISTRIBUTION FUNCTION

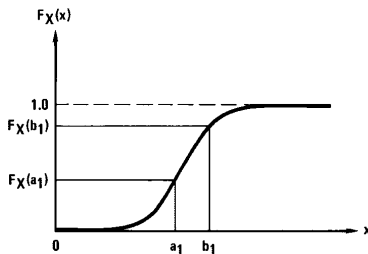
If the entire set of probabilities for a random variable event X are known, then since the probability element, $f_X(x) dx$, describes the probability that event X will occur, the accumulation of these probabilities from $x = -\infty$ to $x = \infty$ is unity or an absolutely certain event. Hence,

$$F_X(x) \int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (A.5-1)$$

where $F_X(x)$ is defined as the *cumulative distribution function* (cdf) or *distribution function* and $f_X(x)$ is the pdf of random variable X . Illustratively, Figures A.5.1a and A.5.1b show the probability density function and cumulative distribution function respectively.



(a) Probability Density Function



(b) Cumulative Distribution Function

FIGURE A.5.1

In many texts the notation used in describing the cdf is

$$F_X(x) = \text{Prob}[X \leq x] \quad (\text{A.5-2})$$

and is defined to be the probability of the event that the observed random variable X is less than or equal to the allowed or conditional value x . This implies

$$F_X(x) = \text{Prob}[X \leq x] = \int_{-\infty}^x f_X(x) dx \quad (\text{A.5-3})$$

It can be further noted that

$$\text{Prob}[x_1 \leq x \leq x_2] = \int_{x_1}^{x_2} f_X(x) dx = F_X(x_2) - F_X(x_1) \quad (\text{A.5-4})$$

and that from equation (A.5-1) the pdf can be related to the cdf by the derivative

$$f_X(x) = \frac{d[F_X(x)]}{dx} \quad (\text{A.5-5})$$

Re-examining Figure A.5.1 does indeed show that the pdf, $f_X(x)$, is a plot of the differential change in slope of the cdf, $F_X(x)$.

$F_X(x)$ and a summary of its properties.

1. $0 \leq F_X(x) \leq 1$ $-\infty < x < \infty$ (Since $F_X = \text{Prob}[X < x]$ is a probability)
2. $F_X(-\infty) = 0$ $F_X(+\infty) = 1$
3. $F_X(x)$ the probability of occurrence increases as x increases
4. $F_X(x) = \int f_X(x) dx$
5. $\text{Prob}(x_1 \leq x \leq x_2) = F_X(x_2) - F_X(x_1)$

A.6 MEAN VALUES, VARIANCES AND STANDARD DEVIATION

The procedure of determining the average weight of a group of objects by summing their individual weights and dividing by the total number of objects gives the average value of x . Mathematically the discrete *sample mean* can be described

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{A.6-1})$$

for the continuous case that *mean value* of the random variable X is defined as

$$\bar{x} = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (\text{A.6-2})$$

where $E[X]$ is read "the expected value of X ".

Other names for the same *mean value* $\bar{x}$ or the *expected value* $E[X]$ are average value and statistical average.

It is seen from equation (A.6-2) that $E[X]$ essentially represents the sum of all possible values of x with each value being weighted by a corresponding value of the probability density function of $f_X(x)$.

Extending this definition to any function of X for example $h(x)$, equation (A.6-2) becomes

$$E[h(x)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx \quad (\text{A.6-3})$$

An example at this point may help to clarify matters. Assume a uniformly dense random variable of density $1/4$ between the values 2 and 6, see Figure A.6.1. The use of equation (A.6-2) yields the expected value

$$\bar{x} = E[X] = \int_2^6 x \frac{1}{4} dx = \frac{x^2}{8} \Big|_2^6 = 4$$

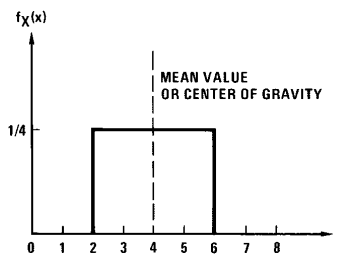


FIGURE A.6.1

which can also be interpreted as the first moment or center of gravity of the density function $f_X(x)$. The above operation is analogous to the techniques in electrical engineering where the DC component of a time function is found by first integrating and then dividing the resultant area by the interval over which the integration was performed.

Generally speaking, the time averages of random variable functions of time are extremely important since essentially no statistical meaning can be drawn from a single random variable (defined as the value of a time function at a given single instant of time). Thus, the operation of finding the *mean value* by integrating over a specified range of possible values that a random variable may assume is referred to as *ensemble averaging*.

In the above example, $\bar{x}$ was described as the first moment m_1 or DC value. The mean-square value $\bar{x}^2$ or $E[X^2]$ is the second moment m_2 or the total average power, AC plus DC and in general the n th moment can be written

$$m_n = E[X^n] = \int_{-\infty}^{\infty} [h(x)]^n f_X(x) dx \quad (\text{A.6-4})$$

Note that the first moment squared m_1^2 , $\bar{x}^2$ or $E[X]^2$ is equivalent to the DC power through a 1Ω resistor and is *not* the same as the second moment m_2 , $\bar{x}^2$ or $E[X^2]$ which, again, implies the total average power.

A discussion of central moments is next in store and is simply defined as the moments of the difference (deviation) between a random variable and its mean value. Letting $[h(x)]^n = (X - \bar{x})^n$, mathematically

$$\overline{(X - \bar{x})^n} = E[(X - \bar{x})^n] = \int_{-\infty}^{\infty} (X - \bar{x})^n f_X(x) dx \quad (\text{A.6-5})$$

For $n = 1$, the first central moment equals zero i.e., the AC voltage (current) minus the mean, average or DC voltage (current) equals zero. This essentially yields little information. The second central moment, however, is so important that it has been named the *variance* and is symbolized by σ^2 . Hence,

$$\sigma^2 = E[(X - \bar{x})^2] = \int_{-\infty}^{\infty} (X - \bar{x})^2 f_X(x) dx \quad (\text{A.6-6})$$

Note that because of the squared term, values of X to either side of the mean $\bar{x}$ are equally significant in measuring variations or deviations away from $\bar{x}$ i.e., if $\bar{x} = 10$, $X_1 = 12$ and $X_2 = 8$ then $(12 - 10)^2 = 4$ and $(8 - 10)^2 = 4$ respectively. The variance therefore is the measure of the variability of $[h(x)]^2$ about its mean value $\bar{x}$ or the expected square deviation of X from its mean value. Since,

$$\sigma^2 = E[(X - \bar{x})^2] = E[X^2 - 2\bar{x}X + \bar{x}^2] \quad (\text{A.6-7a})$$

$$= E[X^2] - 2\bar{x}E[X] + \bar{x}^2 \quad (\text{A.6-7b})$$

$$= \bar{x}^2 - 2\bar{x}\bar{x} + \bar{x}^2 \quad (\text{A.6-7c})$$

$$= \bar{x}^2 - \bar{x}_1^2 \quad (\text{A.6-7d})$$

or

$$m_2 - m_1^2 \quad (\text{A.6-7e})$$

The analogy can be made that variance is essentially the average AC power of a function $h(x)$, hence, the total average power, second moment m_2 , minus the DC power, first moment squared m_1^2 . The positive square root of the variance.

$$\sqrt{\sigma^2} = \sigma$$

is defined as the *standard deviation*. The mean indicates where a density is centered but gives no information about how spread out it is. This spread is measured by the standard deviation σ and is a measure of the spread of the density function about $\bar{x}$, i.e., the smaller σ the closer the values of X to the mean. In relation to electrical engineering the standard deviation is equal to the root-mean-square (rms) value of an AC voltage (current) in circuit.

A summary of the concepts covered in this section are listed in Table A.6.1.

A.7 FUNCTIONS OF TWO JOINTLY DISTRIBUTED RANDOM VARIABLES

The study of jointly distributed random variables can perhaps be clarified by considering the response of linear systems to random inputs. Relating the output of a system to its input is an example of analyzing two random variables from different random processes. If on the other hand an attempt is made to relate present or future values to its past values, this, in effect, is the study of random variables coming from the same process at different instances of time. In either case, specifying the relationship between the two random variables is established by initially developing a probability model for the joint occurrence of two random events. The following sections will be concerned with the development of these models.

A.8 JOINT CUMULATIVE DISTRIBUTION FUNCTION

The joint cumulative distribution function (cdf) is similar to the cdf of Section A.5 except now two random variables are considered.

$$F_{XY}(x,y) = \text{Prob}[X \leq x, Y \leq y] \quad (\text{A.8-1})$$

defines the joint cdf, $F_{XY}(x,y)$, of random variables X and Y . Equation (A.8-1) states that $F_{XY}(x, y)$ is the probability associated with the joint occurrence of the event that X is less than or equal to an allowed or conditional value x and the event that Y is less than or equal to an allowed or conditional value y .

TABLE A.6-1		
Symbol	Name	Physical Interpretation
$\bar{X}, E[X], m_1: \int_{-\infty}^{\infty} x f_X(x) dx$	Expected Value, Mean Value, Statistical Average Value	- Finding the mean value of a random voltage (current) is equivalent to finding its DC component. - First moment; e.g., the first moment of a group of masses is just the average location of the masses or their center of gravity. - The range of the most probable values of x.
$E[X]^2, \bar{X}^2, m_1^2$		DC power
$\bar{X}^2, E[X^2], m_2: \int_{-\infty}^{\infty} x^2 f_X(x) dx$	Mean Square Value	- Interpreted as being equal to the time average of the square of a random voltage (current). In such cases the mean-square value is proportional to the total average power (AC plus DC) through a 1 Ω resistor and its square root is equal to the rms or effective value of the random voltage (current). - Second moment; e.g., the moment of inertia of a mass or the turning moment of torque about the origin. - The mean-square value represents the spread of the curve about $\bar{x} = 0$
$\text{var}[], \sigma^2, (X - \bar{X})^2, E[(X - \bar{X})^2], m_2: \int_{-\infty}^{\infty} (X - \bar{X})^2 f_X(x) dx$	Variance	- Related to the average power (in a 1 Ω resistor) of the AC components of a voltage (current) in power units. The square root of the variances is the rms voltage (current) again not reflecting the DC component. - Second movement; for example the moment of inertia of a mass or the turning moment of torque about the value $\bar{x}$. - Represents the spread of the curve about the value $\bar{x}$.
$\sqrt{\sigma^2}, \sigma$	Standard Deviation	- Effective rms AC voltage (current) in a circuit. - A measure of the spread of a distribution corresponding to the amount of uncertainty or error in a physical measurement or experiment. - Standard measure of deviation of X from its mean value $\bar{x}$.
$(\bar{X})^2$ is a result of smoothing the data and then squaring it and $\overline{(X^2)}$ results from squaring the data and then smoothing it.		

A few properties of the joint cumulative distribution function are listed below.

$$1. 0 \leq F_{XY}(x, y) \leq 1 \quad \begin{array}{l} -\infty < x < \infty \\ -\infty < y < \infty \end{array}$$

(since $F_{XY}(x, y) = \text{Prob}[X \leq x, Y \leq y]$ is a probability)

$$2. F_{XY}(-\infty, y) = 0$$

$$F_{XY}(x, -\infty) = 0$$

$$F_{XY}(-\infty, -\infty) = 0$$

$$3. F_{XY}(+\infty, +\infty) = 1$$

$$4. F_{XY}(x, y) \quad \text{The probability of occurrence increases as either } x \text{ or } y, \text{ or both increase}$$

A.9 JOINT PROBABILITY DENSITY FUNCTION

Similar to the single random variable probability density function (pdf) of sections A.4 and A.5, the joint probability density function $f_{XY}(x, y)$ is defined as the partial derivative of the joint cumulative distribution function $F_{XY}(x, y)$. More clearly,

$$f_{XY}(x, y) = \frac{d^2}{dx dy} F_{XY}(x, y) \quad (\text{A.9-1})$$

Recall that the pdf is a *density* function and must be integrated to find a probability. As an example, to find the probability that (X, Y) is within a rectangle of dimension $(x_1 \leq X \leq x_2)$ and $(y_1 \leq Y \leq y_2)$, the pdf of the joint or two-dimensional random variable must be integrated over both ranges as follows,

$$\text{Prob}[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \quad (\text{A.9-2})$$

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{XY}(x, y) dy dx = 1$$

It is noted that the double integral of the joint pdf is in fact the cumulative distribution function

$$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x, y) dx dy = 1 \quad (\text{A.9-3})$$

analogous to Section A.5. Again $f_{XY}(x, y) dx dy$ represents the probability that X and Y will jointly be found in the ranges

$$x \pm \frac{dx}{2} \text{ and } y \pm \frac{dy}{2},$$

respectively, where the joint density function $f_{XY}(x, y)$ has been normalized so that the volume under the curve is unity.

A few properties of the joint probability density functions are listed below.

$$1. f_{XY}(x, y) > 0 \quad \text{For all values of } x \text{ and } y \text{ or } -\infty < x < \infty \text{ and } -\infty < y < \infty, \text{ respectively}$$

$$2. \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1$$

$$3. F_{XY}(x, y) = \int_{-\infty}^y \int_{-\infty}^x f_{XY}(x, y) dx dy$$

$$4. \text{Prob}[x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2] = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f_{XY}(x, y) dx dy$$

A.10 STATISTICAL INDEPENDENCE

If the knowledge of one variable gives no information about the value of the other, the two random variables are said to be statistically independent. In terms of the joint pdf

$$f_{XY}(x, y) = f_X(x) f_Y(y) \quad (\text{A.10-1})$$

and

$$f_X(x) f_Y(y) = f_{XY}(x, y) \quad (\text{A.10-2})$$

imply statistical independence of the random variables X and Y . In the same respect the joint cdf

$$F_{XY}(x, y) = F_X(x) F_Y(y) \quad (\text{A.10-3})$$

and

$$F_X(x) F_Y(y) = F_{XY}(x, y) \quad (\text{A.10-4})$$

again implies this independence.

It is important to note that for the case of the expected value $E[XY]$, statistical independence of random variables X and Y implies

$$E[XY] = E[X] E[Y] \quad (\text{A.10-5})$$

but, the converse is *not* true since random variables can be uncorrelated but not necessarily independent.

In summary

1. $F_{XY}(x, y) = F_X(x) F_Y(y)$ reversible
2. $f_{XY}(x, y) = f_X(x) f_Y(y)$ reversible
3. $E[XY] = E[X] E[Y]$ non-reversible

A.11 MARGINAL DISTRIBUTION AND MARGINAL DENSITY FUNCTIONS

When dealing with two or more random variables that are jointly distributed, the distribution of each random variable is called the *marginal distribution*. It can be shown that the marginal distribution defined in terms of a joint distribution can be manipulated to yield the distribution of each random variable considered by itself. Hence, the marginal distribution functions $F_X(x)$ and $F_Y(y)$ in terms of $F_{XY}(x, y)$ are

$$F_X(x) = F_{XY}(x, \infty) \quad (\text{A.11-1})$$

and

$$F_Y(y) = F_{XY}(\infty, y) \quad (\text{A.11-2})$$

respectively.

The marginal density functions $f_X(x)$ and $f_Y(y)$ in relation to the joint density $f_{XY}(x, y)$ is represented as

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy \quad (\text{A.11-3})$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx \quad (\text{A.11-4})$$

respectively.

A.12 TERMINOLOGY

Before continuing into the following sections it is appropriate to provide a few definitions of the terminology used hereafter. Admittedly, the definitions presented are by no means complete but are adequate and essential to the continuity of the discussion.

Deterministic and Nondeterministic Random Processes: A random process for which its future values cannot be exactly predicted from the observed past values is said to be *nondeterministic*. A random process for which the future values of any sample function can be exactly predicted from a knowledge of all past values, however, is said to be a *deterministic* process.

Stationary and Nonstationary Random Processes: If the marginal and joint density functions of an event do not depend upon the choice of i.e., time origin, the process is said to be *stationary*. This implies that the mean values and moments of the process are constants and are not dependent upon the absolute value of time. If on the other hand the probability density functions do change with the choice of time origin, the process is defined *nonstationary*. For this case one or more of the mean values or moments are also time dependent. In the strictest sense, the stochastic process $x(t)$ is stationary if its statistics are not affected by the shift in time origin i.e., the process $x(t)$ and $x(t + \tau)$ have the same statistics for any τ .

Ergodic and Nonergodic Random Processes: If every member of the ensemble in a stationary random process exhibits the same statistical behavior that the entire ensemble has, it is possible to determine the process statistical behavior by examining only one typical sample function. This is defined as an *ergodic* process and its mean value and moments can be determined by time averages as well as by ensemble averages. Further ergodicity implies a stationary process and any process not possessing this property is *nonergodic*.

Mathematically speaking, any random process or, i.e., wave shape $x(t)$ for which

$$\lim_{T \rightarrow \infty} \overline{x(t)} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt = E[x(t)]$$

holds true is said to be an ergodic process. This simply says that as the averaging time, T , is increased to the limit $T \rightarrow \infty$, time averages equal ensemble averages (the *expected value* of the function).

A.13 JOINT MOMENTS

In this section, the concept of joint statistics of two continuous dependent variables and a particular measure of this dependency will be discussed.

The joint moments m_{ij} of the two random variables X and Y are defined as

$$m_{ij} = E[X^i Y^j] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^i y^j f_{XY}(x, y) dx dy \quad (\text{A.13-1})$$

where $i + j$ is the order of the moment.

The second moment represented as μ_{11} or σ_{XY} serves as a measure of the dependence of two random variables and is given a special name called the *covariance* of X and Y . Thus

$$\mu_{11} = \sigma_{XY} = E[(X - \bar{x})(Y - \bar{y})] = \quad (\text{A.13-2})$$

$$\begin{aligned} & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (X - \bar{x})(Y - \bar{y}) f_{XY}(x, y) dx dy \\ &= E[XY] - E[X] E[Y] \end{aligned} \quad (\text{A.13-3})$$

or

$$= m_{11} - \bar{x}\bar{y} \quad (\text{A.13-4})$$

If the two random variables are independent, their covariance function μ_{11} is equal to zero and m_{11} , the average of the product, becomes the product of the individual averages hence.

$$\mu_{11} = 0 \quad (\text{A.13-5})$$

implies

$$m_{11} = E[(X - \bar{x})(Y - \bar{y})] = E[X - \bar{x}] E[Y - \bar{y}] \quad (\text{A.13-6})$$

Note, however, the converse of this statement in general is not true but does have validity for two random variables possessing a joint (two dimensional) Gaussian distribution.

In some texts the name cross-covariance is used instead of covariance, regardless of the name used they both describe processes of two random variables each of which comes from a separate random source. If, however, the two random variables come from the same source it is instead called the *autovariance* or *auto-covariance*.

It is now appropriate to define a normalized quantity called the *correlation coefficient*, ρ , which serves as a numerical measure of the dependence between two random variables. This coefficient is the covariance normalized, namely

$$\rho = \frac{\text{covar}[X, Y]}{\sqrt{\text{var}[X] \text{var}[Y]}} = \frac{E\{[X - E[X]][Y - E[Y]]\}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \quad (\text{A.13-7})$$

$$= \frac{\mu_{11}}{\sigma_X \sigma_Y} \quad (\text{A.13-8})$$

where ρ is a dimensionless quantity

$$-1 \leq \rho \leq 1$$

Values close to 1 show high correlation of i.e., two random waveforms and those close to -1 show high correlation of the same waveform except with opposite sign. Values near zero indicate low correlation.

A.14 CORRELATION FUNCTIONS

If $x(t)$ is a sample function from a random process and the random variables

$$x_1 = x(t_1)$$

$$x_2 = x(t_2)$$

are from this process, then, the *autocorrelation* function $R(t_1, t_2)$ is the joint moment of the two random variables;

$$R_{xx}(t_1, t_2) = E[x(t_1)x(t_2)] \quad (\text{A.14-1})$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_{x_1 x_2}(x_1, x_2) dx_1 dx_2$$

where the autocorrelation is a function of t_1 and t_2 .

The auto-covariance of the process $x(t)$ is the covariance of the random variables $x(t_1)$ $x(t_2)$

$$c_{xx}(t_1, t_2) = E\{[x(t_1) - \overline{x(t_1)}][x(t_2) - \overline{x(t_2)}]\} \quad (\text{A.14-2})$$

or rearranging equation (A.14-1)

$$\begin{aligned} c(t_1, t_2) &= E\{[x(t_1) - \overline{x(t_1)}][x(t_2) - \overline{x(t_2)}]\} \\ &= E\{x(t_1)x(t_2) - x(t_1)\overline{x(t_2)} - \overline{x(t_1)}x(t_2) + \overline{x(t_1)}\overline{x(t_2)}\} \\ &= E\{x(t_1)x(t_2) - x(t_1)E[x(t_2)] - E[x(t_1)]x(t_2) + E[x(t_1)]E[x(t_2)]\} \\ &= E[x(t_1)x(t_2)] - E[x(t_1)]E[x(t_2)] \\ &\quad - E[x(t_1)]E[x(t_2)] + E[x(t_1)]E[x(t_2)] \\ &= E[x(t_1)x(t_2)] - E[x(t_1)]E[x(t_2)] \quad (\text{A.14-3}) \end{aligned}$$

or

$$= R(t_1, t_2) - E[x(t_1)]E[x(t_2)] \quad (\text{A.14-4})$$

The autocorrelation function as defined in equation (A.14-1) is valid for both stationary and nonstationary processes. If $x(t)$ is stationary then all its ensemble averages are independent of the time origin and accordingly

$$R_{xx}(t_1, t_2) = R_{xx}(t_1 + T, t_2 + T) \quad (\text{A.14-5a})$$

$$= E[x(t_1 + T)x(t_2 + T)] \quad (\text{A.14-5b})$$

Due to this time origin independence, T can be set equal to $-t_1$, $T = -t_1$, and substitution into equations (A.14-5a, b)

$$R_{xx}(t_1, t_2) = R_{xx}(0, t_2 - t_1) \quad (\text{A.14-6a})$$

$$= E[x(0)x(t_2 - t_1)] \quad (\text{A.14-6b})$$

imply that the expression is only dependent upon the time difference $t_2 - t_1$. Replacing the difference with $\tau = t_2 - t_1$ and suppressing the zero in the argument $R_{xx}(0, t_2 - t_1)$ yields

$$R_{xx}(\tau) = E[x(t_1)x(t_1 - \tau)] \quad (\text{A.14-7})$$

Again since this is a stationary process it depends only on τ . The lack of dependence on the particular time, t_1 , at which the ensemble was taken allows equation (A.14-7) to be written without the subscript, i.e.,

$$R_{xx}(\tau) = E[x(t)x(t + \tau)] \quad (\text{A.14-8})$$

as it is found in many texts. This is the expression for the autocorrelation function of a stationary random process.

For the autocorrelation function of a nonstationary process where there is a dependence upon the particular time at which the ensemble average was taken as well as on the time difference between samples, the expression must be written with identifying subscripts, i.e., $R_{xx}(t_1, t_2)$ or $R_{xx}(t_1, \tau)$.

The time autocorrelation function can be defined next and has the form

$$R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t + \tau) dt \quad (\text{A.14-9})$$

For the special case of an ergodic process (Ref. Appendix A.12) the two functions, equations (A.14-8) and (A.14-9), are equal

$$R_{xx}(\tau) = R_{xx}(\tau) \quad (\text{A.14-10})$$

It is important to point out that if $\tau = 0$ in equation (A.14-7) the autocorrelation function

$$R_{xx}(0) = E[x(t_1)x(t_1)] \quad (\text{A.14-11})$$

would equal the mean square value or total power (AC plus DC) of the process. Further, for values other than $\tau = 0$, $R_x(\tau)$ represents a measure of the similarity between its waveforms $x(t)$ and $x(t + \tau)$.

In the same respect as the previous discussion, two random variables from two different jointly stationary random processes $x(t)$ and $y(t)$ have for the random variables

$$\begin{aligned}x_1 &= x(t_1) \\ y_2 &= y(t_1 + \tau)\end{aligned}$$

the *crosscorrelation* function

$$\begin{aligned}R_{xy}(\tau) &= E\{x(t_1) y(t_1 + \tau)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 y_2 f_{x_1 y_2}(x_1, y_2) dx_1 dy_2\end{aligned}\quad (\text{A.14-12})$$

The crosscorrelation function is simply a measure of how much these two variables depend upon one another.

Since it was assumed that both random processes were jointly stationary, the crosscorrelation is thus only dependent upon the time difference τ and, therefore

$$R_{xy}(\tau) = R_{yx}(\tau) \quad (\text{A.14-13})$$

where

$$\begin{aligned}y_1 &= y(t_1) \\ x_2 &= x(t_1 + \tau)\end{aligned}$$

and

$$\begin{aligned}R_{yx}(\tau) &= E\{y(t_1) x(t_1 + \tau)\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y_1 x_2 f_{y_1 x_2}(y_1, x_2) dy_1 dx_2\end{aligned}\quad (\text{A.14-14})$$

The time crosscorrelation functions are defined as before by

$$R_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) y(t + \tau) dt \quad (\text{A.14-15})$$

and

$$R_{yx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} y(t) x(t + \tau) dt \quad (\text{A.14-16})$$

and finally

$$R_{xy}(\tau) = R_{xy}(\tau) \quad (\text{A.14-17})$$

$$R_{yx}(\tau) = R_{yx}(\tau) \quad (\text{A.14-18})$$

for the case of jointly ergodic random processes.

APPENDIX B

B.0 INTERCHANGING TIME INTEGRATION AND EXPECTATIONS

If $f(t)$ is a nonrandom time function and $a(t)$ a sample function from a random process then,

$$E \left[\int_{t_1}^{t_2} a(t) f(t) dt \right] = \int_{t_1}^{t_2} E[a(t)] f(t) dt \quad (\text{B.0-1})$$

This is true under the condition

$$\text{a) } \int_{t_1}^{t_2} E[|a(t)|] |f(t)| dt < \infty \quad (\text{B.0-2})$$

b) $a(t)$ is bounded by the interval t_1 to t_2 . [t_1 and t_2 may be infinite and $a(t)$ may be either stationary or nonstationary]

APPENDIX C

C.0 CONVOLUTION

This appendix defines convolution and presents a short proof without elaborate explanation. For complete definition of convolution refer to National Semiconductor Application Note AN-237.

For the *time convolution* if

$$f(t) \longleftrightarrow F(\omega) \quad (\text{C.0-1})$$

$$x(t) \longleftrightarrow X(\omega) \quad (\text{C.0-2})$$

then

$$y(t) = \int_{-\infty}^{\infty} x(\tau) f(t - \tau) d\tau \longleftrightarrow Y(\omega) = X(\omega) * F(\omega) \quad (\text{C.0-3})$$

or

$$y(t) = x(t) * f(t) \longleftrightarrow Y(\omega) = X(\omega) * F(\omega) \quad (\text{C.0-4})$$

proof:

Taking the Fourier transform, $F[\]$, of $y(t)$

$$F[y(t)] = Y(\omega) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x(\tau) f(t - \tau) d\tau \right] e^{-j\omega t} dt \quad (\text{C.0-5})$$

$$Y(\omega) = \int_{-\infty}^{\infty} x(\tau) \left[\int_{-\infty}^{\infty} f(t - \tau) e^{-j\omega t} dt \right] d\tau \quad (\text{C.0-6})$$

and letting $k = t - \tau$, then, $dk = dt$ and $t = k + \tau$.

Thus,

$$Y(\omega) = \int_{-\infty}^{\infty} x(\tau) \left[\int_{-\infty}^{\infty} f(k) e^{-j\omega(k+\tau)} dk \right] d\tau \quad (C.0-7)$$

$$= \int_{-\infty}^{\infty} x(\tau) e^{-j\omega\tau} d\tau \int_{-\infty}^{\infty} f(k) e^{-j\omega k} dk \quad (C.0-8)$$

$$Y(\omega) = X(\omega) \bullet F(\omega) \quad (C.0-9)$$

For the frequency convolution of

$$f(t) \longleftrightarrow F(\omega) \quad (C.0-10)$$

$$x(t) \longleftrightarrow X(\omega) \quad (C.0-11)$$

then

$$H(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\nu) X(\omega - \nu) d\nu \longleftrightarrow h(t) = f(t) \bullet x(t) \quad (C.0-12)$$

or

$$H(\omega) = \frac{1}{2\pi} F(\omega) * X(\omega) \longleftrightarrow h(t) = f(t) \bullet x(t) \quad (C.0-13)$$

proof:

Taking the inverse Fourier transform $F^{-1}[\]$ of equation (C.0-13)

$$\begin{aligned} h(t) &= F^{-1} \left[\frac{X(\omega) * F(\omega)}{2\pi} \right] \quad (C.0-14) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} F(\nu) X(\omega - \nu) d\nu \right] e^{j\omega t} d\omega \\ &= \left(\frac{1}{2\pi} \right)^2 \int_{-\infty}^{\infty} F(\nu) \int_{-\infty}^{\infty} X(\omega - \nu) e^{j\omega t} d\omega d\nu \quad (C.0-15) \end{aligned}$$

and letting $g = \omega - \nu$, then $dg = d\omega$ and $\omega = g + \nu$.

Thus,

$$F^{-1} \frac{X(\omega) * F(\omega)}{2\pi} \quad (C.0-16)$$

$$h(t) = \left(\frac{1}{2\pi} \right)^2 \int_{-\infty}^{\infty} F(\nu) \int_{-\infty}^{\infty} X(g) e^{j(g+\nu)t} dg d\nu$$

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\nu) e^{j\nu t} d\nu \bullet \int_{-\infty}^{\infty} X(g) e^{jg t} dg \quad (C.0-17)$$

$$h(t) = f(t) \bullet x(t) \quad (C.0-18)$$

APPENDIX D

D.0 REFERENCES

1. Brigham, E. Oran, *The Fast Fourier Transform*, Prentice-Hall, 1974.
2. Chen, Carson, *An Introduction to the Sampling Theorem*, National Semiconductor Corporation Application Note AN-236, January 1980.
3. Chen, Carson, *Convolution: Digital Signal Processing*, National Semiconductor Corporation Application Note AN-237, January 1980.
4. Connors, F.R., *Noise*.
5. Cooper, George R.; McGillen, Clare D., *Methods of Signal and System Analysis*, Holt, Rinehart and Winston, Incorporated, 1967.
6. Enochson, L., *Digital Techniques in Data Analysis*, Noise Control Engineering, November-December 1977.
7. Gabel, Robert A.; Roberts, Richard A., *Signals and Linear Systems*.
8. Harris, F.J. *Windows, Harmonic Analysis and the Discrete Fourier Transform*, submitted to IEEE Proceedings, August 1976.
9. Hayt, William H., Jr.; Kemmerly, Jack E., *Engineering Circuit Analysis*, McGraw-Hill, 1962.
10. Jenkins, G.M.; Watts, D.G., *Spectral Analysis and Its Applications*, Holden-Day, 1968.
11. Kuo, Franklin F., *Network Analysis and Synthesis*, John Wiley and Sons, Incorporated, 1962.
12. Lathi, B.P., *Signals, Systems and Communications*, John Wiley and Sons, Incorporated, 1965.
13. Liu, C.L.; Liu, Jane W.S., *Linear Systems Analysis*.
14. Meyer, Paul L., *Introductory Probability and Statistical Applications*, Addison-Wesley Publishing Company, 1970.
15. Mix, Dwight F., *Random Signal Analysis*, Addison-Wesley Publishing Company, 1969.
16. Oppenheim, A.V.; Schafer, R.W., *Digital Signal Processing*, Prentice-Hall, 1975.
17. Otnes, Robert K.; Enochson, Loran, *Applied Time Series Analysis*, John Wiley and Sons, Incorporated, 1978.
18. Otnes, Robert K.; Enochson, Loran, *Digital Time Series Analysis*, John Wiley and Sons, Incorporated, 1972.
19. Papoulis, Athanasios, *The Fourier Integral and Its Applications*, McGraw-Hill, 1962.
20. Papoulis, Athanasios, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1965.
21. Papoulis, Athanasios, *Signal Analysis*, McGraw-Hill, 1977.
22. Rabiner, Lawrence R.; Gold, Bernard, *Theory and Application of Digital Signal Processing*, Prentice-Hall, 1975.

23. Rabiner, L.R.; Schafer, R.W.; Dlugos, D., *Periodogram Method for Power Spectrum Estimation*, Programs for Digital Signal Processing, IEEE Press, 1979.

24. Raemer, Harold R., *Statistical Communications Theory and Applications*, Prentice-Hall EE Series.

25. Roden, Martin S., *Analog and Digital Communications Systems*, Prentice-Hall, 1979.

26. Schwartz, Mischa, *Information Transmission Modulation, and Noise*, McGraw-Hill, 1959, 1970.

27. Schwartz, Mischa; Shaw, Leonard, *Signal Processing: Discrete Spectral Analysis, Detection, and Estimation*, McGraw-Hill, 1975.

28. Silvia, Manuel T.; Robinson, Enders A., *Digital Signal Processing and Time Series Analysis*, Holden-Day Inc., 1978.

29. Sloane, E.A., *Comparison of Linearly and Quadratically Modified Spectral Estimates of Gaussian Signals*, IEEE Transactions on Audio and Electroacoustics Vol. Au-17, No. 2, June 1969.

30. Smith, Ralph J., *Circuits, Devices, and Systems*, John Wiley and Sons, Incorporated, 1962.

31. Stanley, William D., *Digital Signal Processing*, Reston Publishing Company, 1975.

32. Stearns, Samuel D., *Digital Signal Analysis*, Hayden Book Company Incorporated, 1975.

33. Taub, Herbert; Schilling, Donald L., *Principles of Communication Systems*, McGraw-Hill, 1971.

34. Tretter, Steven A., *Discrete-Time Signal Processing*, John Wiley and Sons, Incorporated, 1976.

35. Turkey, J.W.; Blackman, R.B., *Discrete-Time Fourier Transform*, Dover Publications Incorporated, 1959.

36. Welch, P.D., *Fast Fourier Transform of the Sinusoidal Amplitude Spectrum*, IBM Watson Research Center, Yorktown Heights, N.Y. 10598.

37. Welch, P.D., *Power Spectral Density Estimation by the Modified Periodogram Method*, IEEE Transactions on Audio and Electroacoustics, June 1967.

38. Programs for *Power Spectra Estimation*, IEEE Press, 1979.

39. Bendat, J.S.; Piersol, A.G., *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, 1971.

LIFE SUPPORT POLICY

NATIONAL'S PRODUCTS ARE NOT AUTHORIZED FOR USE AS CRITICAL COMPONENTS IN LIFE SUPPORT DEVICES OR SYSTEMS WITHOUT THE EXPRESS WRITTEN APPROVAL OF THE PRESIDENT OF NATIONAL SEMICONDUCTOR CORPORATION. As used herein:

1. Life support devices or systems are devices or systems which, (a) are intended for surgical implant into the body, or (b) support or sustain life, and whose failure to perform, when properly used in accordance with instructions for use provided in the labeling, can be reasonably expected to result in a significant injury to the user.
2. A critical component is any component of a life support device or system whose failure to perform can be reasonably expected to cause the failure of the life support device or system, or to affect its safety or effectiveness.

	National Semiconductor Corporation 1111 West Bardin Road Arlington, TX 76017 Tel: 1(800) 272-9959 Fax: 1(800) 737-7018	National Semiconductor Europe Fax: (+49) 0-180-530 85 86 Email: cnjwge@tevm2.nsc.com Deutsch Tel: (+49) 0-180-530 85 85 English Tel: (+49) 0-180-532 78 32 Français Tel: (+49) 0-180-532 93 58 Italiano Tel: (+49) 0-180-534 16 80	National Semiconductor Hong Kong Ltd. 19th Floor, Straight Block, Ocean Centre, 5 Canton Rd. Tsimshatsui, Kowloon Hong Kong Tel: (852) 2737-1600 Fax: (852) 2736-9960	National Semiconductor Japan Ltd. Tel: 81-043-299-2309 Fax: 81-043-299-2408
---	---	---	--	--

National does not assume any responsibility for use of any circuitry described, no circuit patent licenses are implied and National reserves the right at any time without notice to change said circuitry and specifications.