Chapter 8 Energy and Power, Random Signals and Noise

8.1 Energy and Power in Deterministic Signals

Consider a deterministic signal $f(t)$. If we think of $f(t)$ as a voltage, the instantaneous power that it develops across a 1 Ω resistor is

$$P_f(t) = |f(t)|^2 \tag{8.1}$$

We shall adopt this as our definition of the *instantaneous power* in a signal.

8.1.1 Signals of Finite Energy

The *total energy* in $f(t)$ is then defined as the integral of $P_f(t)$ over all time, i.e.,

$$E_f = \int_{-\infty}^{\infty} |f(t)|^2 \,\mathrm{d}t = \int_{-\infty}^{\infty} |F(\nu)|^2 \,\mathrm{d}\nu \tag{8.2}$$

where $F(\nu)$ is the Fourier transform of $f(t)$ and we have used Rayleigh's theorem in the last equality.

Signals for which E is finite are said to be $\mathcal{L}^2$ functions or *finite energy signals*. All finite energy signals vanish as $t \to \pm\infty$.

We now wish to define the energy contained within a band of frequencies within a signal. For definiteness, consider the band of frequencies ν_1 to ν_2. If we consider an ideal band-pass filter with system function

$$H(\nu) = \begin{cases} 1 & \text{if } \nu_1 \le \nu \le \nu_2 \\ 0 & \text{otherwise} \end{cases} \tag{8.3}$$

we may think of passing $f(t)$ through such a bandpass filter which only lets through frequencies within this frequency range and measuring the total energy in the output $g(t)$ of such a filter. We find

$$\begin{aligned} E_g &= \int_{-\infty}^{\infty} |G(\nu)|^2 \,\mathrm{d}\nu = \int_{-\infty}^{\infty} |F(\nu)H(\nu)|^2 \,\mathrm{d}\nu \\ &= \int_{\nu_1}^{\nu_2} |F(\nu)|^2 \,\mathrm{d}\nu \end{aligned} \tag{8.4}$$

From this, we see that it makes sense to call $|F(\nu)|^2$ the *energy spectral density* of $f(t)$ since its integral over the range ν_1 to ν_2 gives the energy contained in this frequency range.

The inverse Fourier transform of the energy spectral density is

$$\begin{aligned} \int_{-\infty}^{\infty} |F(\nu)|^2 \exp(\mathrm{j}2\pi\nu\tau) \,\mathrm{d}\nu &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t) \exp(-\mathrm{j}2\pi\nu t) \,\mathrm{d}t \right)^* F(\nu) \exp(\mathrm{j}2\pi\nu\tau) \,\mathrm{d}\nu \\ &= \int_{-\infty}^{\infty} f^*(t) \left(\int_{-\infty}^{\infty} F(\nu) \exp(\mathrm{j}2\pi\nu(t+\tau)) \,\mathrm{d}\nu \right) \mathrm{d}t \\ &= \int_{-\infty}^{\infty} f^*(t) f(t+\tau) \,\mathrm{d}t \end{aligned} \tag{8.5}$$

This is called the *energy auto-correlation function* of $f(t)$ and is denoted $\phi^e_{ff}(\tau)$. In general, we define the *energy cross-correlation function* of two deterministic finite energy signals $f(t)$ and $g(t)$ to be

$$\phi^e_{fg}(\tau) = \int_{-\infty}^{\infty} f^*(t)g(t+\tau)\,\mathrm{d}t \tag{8.6}$$

when the two functions are the same, this reduces to the *energy auto-correlation function*

$$\phi^e_{ff}(\tau) = \int_{-\infty}^{\infty} f^*(t)f(t+\tau)\,\mathrm{d}t \tag{8.7}$$

We have thus shown that the energy spectral density is the Fourier transform of the energy auto-correlation function, i.e.,

$$\phi^e_{ff}(\tau) \leftrightarrow |F(\nu)|^2 = \Phi^e_{ff}(\nu) \leftrightarrow \phi^e_{ff}(\tau) \tag{8.8}$$

Exercise:

Show that the total energy is given by $\phi^e_{ff}(0)$ and that for all τ, $|\phi^e_{ff}(\tau)| \le \phi^e_{ff}(0)$. (*Hint:* Use the Cauchy-Schwarz inequality).

8.1.2 Signals of Finite Power

Many signals of practical importance have infinite total energy. Periodic signals are simple examples of such signals. The Fourier transform of signals with infinite energy are not absolutely square integrable and often contain generalized functions. Such signals do not have energy spectral density functions. We define the *average power* in a signal to be

$$P_f = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} |f(t)|^2\,\mathrm{d}t \tag{8.9}$$

Signals for which P_f is finite are said to have *finite power.*

For signals of finite power, we define power auto-correlation and power cross-correlation functions analogously to the energy correlation functions defined above. The *power cross-correlation function* of two deterministic finite power signals $f(t)$ and $g(t)$ is defined to be

$$\phi^p_{fg}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)g(t+\tau)\,\mathrm{d}t \tag{8.10}$$

similarly the *power auto-correlation function* of $f(t)$ is

$$\phi^p_{ff}(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)f(t+\tau)\,\mathrm{d}t \tag{8.11}$$

From this definition, it is clear that the average power in f is given by $\phi^p_{ff}(0)$.

We next want to show that the Fourier transform of the power auto-correlation function can be interpreted as a power spectral density. In order to establish this result, we again consider passing $f(t)$ through an ideal bandpass filter and look at the average power in the filter output. This can be done once we know how a linear filtering operation affects the correlation functions.

8.1.3 Linear Filters and Correlation Functions

Theorem: If a finite power signal $f(t)$ is passed through a filter with impulse response $h(t)$ yielding an output $g(t) = (f * h)(t)$, then

1. The cross-correlation of f and g and its Fourier transform are given by

$$\phi^p_{fg}(\tau) = (\phi^p_{ff} * h)(\tau) \tag{8.12}$$
$$\Phi^p_{fg}(\nu) = \Phi^p_{ff}(\nu)H(\nu) \tag{8.13}$$

2. The auto-correlation of g and its Fourier transform are given by

$$\phi^p_{gg}(\tau) = (\phi^p_{ff} * h * \tilde{h})(\tau) \tag{8.14}$$
$$\Phi^p_{gg}(\nu) = \Phi^p_{ff}(\nu)|H(\nu)|^2 \tag{8.15}$$

   where $\tilde{h}$ is defined by $\tilde{h}(t) = h^*(-t)$.

Proof:

1. From the definition,

$$\begin{aligned}
\phi^p_{fg}(\tau) &= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)g(t+\tau)\,\mathrm{d}t \\
&= \lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)\left(\int_{-\infty}^{\infty} f(t+\tau-u)h(u)\,\mathrm{d}u\right)\mathrm{d}t \\
&= \int_{-\infty}^{\infty}\left(\lim_{T\to\infty} \frac{1}{T}\int_{-T/2}^{T/2} f^*(t)f(t+\tau-u)\,\mathrm{d}t\right)h(u)\,\mathrm{d}u \\
&= \int_{-\infty}^{\infty} \phi^p_{ff}(\tau-u)h(u)\,du = (\phi^p_{ff} * h)(\tau)
\end{aligned} \tag{8.16}$$

   The Fourier transform of this relationship yields $\Phi^p_{fg}(\nu) = \Phi^p_{ff}(\nu)H(\nu)$.

2. Again from the definition, it is easy to show that

$$\phi^p_{gg}(\tau) = (\phi^p_{fg} * \tilde{h})(\tau) \tag{8.17}$$

   Substituting the previous result gives $\phi^p_{gg}(\tau) = (\phi^p_{ff} * h * \tilde{h})(\tau)$ and taking its Fourier transform gives $\Phi^p_{gg}(\nu) = \Phi^p_{ff}(\nu)|H(\nu)|^2$ since $\tilde{h}(\tau) \leftrightarrow H^*(\nu)$.

Note: This theorem also holds for signals of finite energy if all the power correlations ϕ^p are replaced by energy correlations ϕ^e.

8.1.4 The Power Spectral Density

Let us now return to the situation of passing a finite power signal $f(t)$ through an ideal bandpass filter with system function $H(\nu)$ as defined in (8.3). As usual we denote the filter impulse response by $h(t)$ and the filter output by $g(t) = (f * h)(t)$. The average power in $g(t)$ is given by $\phi^p_{gg}(0)$. Evaluating this in terms of its Fourier transform,

$$\begin{aligned}
\phi^p_{gg}(0) &= \int_{-\infty}^{\infty} \Phi^p_{gg}(\nu)\,\mathrm{d}\nu \\
&= \int_{-\infty}^{\infty} \Phi^p_{ff}(\nu)|H(\nu)|^2\,\mathrm{d}\nu \qquad \text{by the above theorem} \\
&= \int_{\nu_1}^{\nu_2} \Phi^p_{ff}(\nu)\,\mathrm{d}\nu \qquad (8.18)
\end{aligned}$$

Thus the average power in the output is just the integral of $\Phi^p_{ff}(\nu)$ in the range ν_1 to ν_2. This justifies calling $\Phi^p_{ff}(\nu)$ the *power spectral density* of $f(t)$.

This leads to the following important general rule

The energy (power) spectral density is the Fourier transform of the energy (power) auto-correlation function.

This result is known as the *Wiener-Khinchin* theorem. It also applies to signals with a random component as will be discussed later.

8.2 Stochastic Processes

We now consider signals which are non-deterministic. These are useful in many applications, for example in describing noise in electrical systems, the random motion of a particle undergoing Brownian motion, the concentration of a certain chemical during the course of a reaction, the motion of particles as they diffuse, the intensity of light from a partially coherent source etc. These are generally called *stochastic processes* and are scalar or vector valued quantities which evolve in time. Unlike a deterministic process for which there is a definite value for the quantity at each instant of time, a stochastic process can only be described statistically. This is because we do not want a stochastic process just to describe a particular situation which occurs when an experiment is performed once but rather it should give information about the possible range of outcomes in many realizations of the same experiment.

8.2.1 The Concept of an Ensemble

Let us consider how we might describe the Johnson noise voltage generated in a resistor which is maintained at some nonzero temperature T. At a given instant of time, say at $t = 0$, a given resistor will have a certain voltage across it but this value is not predictable. What we really want to give is a probabilistic description of the possible voltages across the resistor at this time. To do this, it is useful to imagine a whole collection of resistors, all maintained under identical conditions and to look at the probability density formed by considering all their voltages at the same time $t = 0$. This collection is called an *ensemble*. The probability density functions that we talk about in stochastic processes always refer to probabilities defined over such an ensemble. It may be useful to imagine a whole collection of parallel universes in which a physical situation is set up under the same conditions (i.e., identical to within the constraints specified by the problem) and to regard the ensemble as being composed of these “copies” of the system of interest. When we focus on a system in a particular universe, the process that it undergoes is said to be a single *realization* of the stochastic process.

Once we think about an ensemble of "potential realities", it makes sense to regard the voltage at a given time, say $x(0)$, as being a random variable. (From now on, we shall discontinue the convention of using capital letters to distinguish random variables from the values that they can take.) We can then consider the probability density function for $x(0)$ and statistics such as the mean and variance of this random variable. The notation $p(x, 0)$ is often used for the probability density of x at time $t = 0$. The mean and variance of $x(0)$ are then given by

$$\mathrm{E}\left[x\left(0\right)\right] = \int_{-\infty}^{\infty} xp(x, 0)\,\mathrm{d}x \tag{8.19}$$

$$\mathrm{E}\left[x\left(0\right)^2\right] = \int_{-\infty}^{\infty} (x - \mathrm{E}\left[x\left(0\right)\right])^2 p(x, 0)\,\mathrm{d}x \tag{8.20}$$

$$= \int_{-\infty}^{\infty} x^2 p(x, 0)\,\mathrm{d}x - \left(\int_{-\infty}^{\infty} xp(x, 0)\,\mathrm{d}x\right)^2 \tag{8.21}$$

The expectation value symbol $\mathrm{E}[\cdot]$ is usually used to refer to an ensemble average. In a similar way we can define higher order moments of the random variable.

8.2.2 Characterizing a Stochastic Process

In general, characterizing a stochastic process is quite complicated. We have seen that at each instant of time, the value of the stochastic process is a random variable. We thus need an infinite number of probability densities $p(x, t)$ (one for each t) just to describe the process at each instant time. Further thought shows that even this infinite collection is not enough to fully describe the process. For example, we may be interested in the joint probability density for the voltage at two times t_1 and t_2. This will be specified by a probability density of the form $p(x_1, t_1; x_2, t_2)$. These two-time probability densities tell us whether knowledge about the process at time t_1 gives us any knowledge about the process at time t_2. These two-time probability densities form a doubly infinite set of joint probability densities. Similarly, we may consider the joint probability function for the values of the process at n distinct times. All of these functions for all values of n are usually assumed to be sufficient to characterize the process. Processes for which this is possible are said to be *separable.*

The need to specify such a large number of probability density functions to characterize a general stochastic process means that such processes are seldom studied. Instead, special cases are considered in which all the joint probability functions can be calculated in terms of simpler quantities. As an example, if the values of the process at different times are known to be independent, all joint probability densities factorize into a product of single-time probability densities.

8.2.3 Mean, auto-correlation and auto-covariance functions of a stochastic process

Corresponding to the moments of a random variable, we define correlation functions of a stochastic process. These functions are all defined in terms of ensemble averages over all possible realizations of the process. The *mean* of the process is

$$\mu_x(t) = \mathrm{E}\left[x\left(t\right)\right] = \int_{-\infty}^{\infty} x\,p(x, t)\,\mathrm{d}x \tag{8.22}$$

Note that there is a mean for each t since we are not averaging over time but across the ensemble for each time.

The two-time auto-correlation function is

$$\phi_{xx}(t_1, t_2) = \mathrm{E}\left[x\left(t_1\right) x\left(t_2\right)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2\, p(x_1, t_1; x_2, t_2)\, \mathrm{d}x_1 \mathrm{d}x_2 \tag{8.23}$$

and the two-time auto-covariance function is

$$\gamma_{xx}(t_1, t_2) = \mathrm{E}\left[\left(x\left(t_1\right) - \mu_x\left(t_1\right)\right)\left(x\left(t_2\right) - \mu_x\left(t_2\right)\right)\right] \tag{8.24}$$

Note: For complex-valued stochastic processes, we take the complex conjugate of the value at t_1 in the above definitions.

Similarly, the n-time auto-correlation function is

$$\begin{aligned} \phi_{xx\ldots x}(t_1, t_2, ..., t_n) &= \mathrm{E}\left[x\left(t_1\right) x\left(t_2\right) \ldots x\left(t_n\right)\right] \qquad (8.25)\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} x_1 x_2 ... x_n\, p(x_1, t_1; x_2, t_2; ...; x_n, t_n)\, \mathrm{d}x_1 \mathrm{d}x_2 \ldots \mathrm{d}x_n \end{aligned}$$

and the n-time auto-covariance function is

$$\gamma_{xx\ldots x}(t_1, t_2, ..., t_n) = \mathrm{E}\left[\left(x\left(t_1\right) - \mu_x\left(t_1\right)\right)\left(x\left(t_2\right) - \mu_x\left(t_2\right)\right) \ldots \left(x\left(t_n\right) - \mu_x\left(t_n\right)\right)\right] \tag{8.26}$$

The collection of the mean and all the auto-correlation functions are usually assumed to completely characterize the stochastic process. Note that all these functions are deterministic even though they are calculated from the stochastic process.

8.2.4 Markov Processes

In physics, an important class of stochastic processes are the Markov processes. In a Markov process, the conditional probability for any event in the future given a set of values at times $t_1 > t_2 > ... > t_n$ is the same as the conditional probability of that event given the value at the most recent time t_1. For example,

$$p(x, t | x_1, t_1; x_2, t_2; ...; x_n, t_n) = p(x, t | x_1, t_1) \tag{8.27}$$

where $t > t_1 > ... > t_n$. In other words, knowledge of the process at the set of times $t_1, t_2, ..., t_n$ gives us no more useful information for predicting the future evolution of the process than the knowledge of the process at the most recent time t_1.

As a simple example of a Markov process, consider counting the number of particles that a radioactive material has produced from some time origin up to a given time t. Since the probability of decay per unit time is constant, our ability to predict the number of decays up to time t given that there have been n decays up to time $t_1 < t$ is completely independent of the history of decays up to t_1. On the other hand, if the decays occur with a constant delay between successive events, this is not a Markov process. Given that there have been n decays up to time t_1, we would predict a different number of decays by time t depending on exactly when the last decay occurred.

For a Markov process, arbitrary time-ordered joint probabilities can be evaluated as a product of *transition probabilities*. For example, if $t_1 > t_2 > t_3$,

$$\begin{aligned} p(x_1, t_1; x_2, t_2; x_3, t_3) &= p(x_1, t_1 | x_2, t_2; x_3, t_3)\, p(x_2, t_2 | x_3, t_3)\, p(x_3, t_3) \\ &= p(x_1, t_1 | x_2, t_2)\, p(x_2, t_2 | x_3, t_3)\, p(x_3, t_3) \qquad (8.28) \end{aligned}$$

Thus a complete description of a Markov process involves the specification of an initial condition $p(x_0, t_0)$ together with its transition probability density function $p(x, t | x', t')$.

The transition probability density function for a Markov process cannot be specified completely arbitrarily but must satisfy a consistency requirement called the *Chapman-Kolmogorov equation*

$$p(x_1, t_1 | x_3, t_3) = \int_{-\infty}^{\infty} p(x_1, t_1 | x_2, t_2)\, p(x_2, t_2 | x_3, t_3)\, \mathrm{d}x_2 \tag{8.29}$$

for any t_2 satisfying $t_1 > t_2 > t_3$. The Chapman-Kolmogorov equation can be written as a differential equation for $p(x, t|x', t')$. Special cases of this differential equation are known as the *Master equation*, the *Liouville equation* and the *Fokker-Planck equation* which are of great importance in the theory of stochastic processes (see for example *Handbook of Stochastic Methods* by C.W. Gardiner).

8.2.5 Stationary, Wide-sense Stationary and Ergodic Processes

Another way in which a general stochastic process may be simplified is if its statistics are independent of the choice of the time origin. Such a process is said to be *stationary.* Instead of an infinity of one-time probability densities $p(x, t)$, one for each t, a stationary process has a single probability density $p(x)$. The mean of a stationary random process is a single value rather than a function of time. Similarly the n-time joint probability density satisfies

$$p(x_1, t_1; x_2, t_2; ...; x_n, t_n) = p(x_1, t_1 + u; x_2, t_2 + u; ...; x_n, t_n + u) \tag{8.30}$$

for any u. The two-time auto-correlation and auto-covariance functions now depend on the time-difference only and we may write, for example

$$\begin{aligned}\phi_{xx}(\tau) &= \mathrm{E}\left[x^*(t)\, x(t+\tau)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^* x_2 p(x_1, t; x_2, t+\tau)\, \mathrm{d}x_1 \mathrm{d}x_2\end{aligned} \tag{8.31}$$

where the result is independent of t due to the stationarity of the process.

In order for a stochastic process to be stationary the auto-correlation functions of all orders must be independent of the time origin. If only the mean and the two-time auto-correlation function have this property, the process is said to be *wide-sense stationary.* In practice, it is usually only feasible to verify wide-sense stationarity.

Exercise: Show that a wide-sense stationary Markov process is in fact stationary.

For a stationary process, it is sometimes possible that ensemble averages can be replaced by time averages over a single realization of the process. For example, calculating the mean of a single realization by averaging over time may give the same answer as an ensemble average. Similarly we may try to use the definition of the power auto-correlation (8.11) for a single realization as a way of calculating the stationary auto-correlation (8.31). Processes for which this is possible are said to be *ergodic.* All ergodic processes are stationary but the converse is not always true. For many physical processes, we make the assumption that the process is ergodic despite the fact that it may not be possible to prove that this holds. As a simple example of a non-ergodic process, consider an ensemble consisting of a collection of batteries and the process being the battery voltage. Measurements on a single battery as a function of time gives no information about the statistical properties of the ensemble.

8.3 Power Spectrum of a Stationary Stochastic Process

For a stationary stochastic process $x(t)$, we define the *average power* by

$$P_x = \mathrm{E}\left[x(t)^2\right] \tag{8.32}$$

Despite the fact that the right-hand side contains the variable t, it is in fact independent of t by stationarity. We similarly define the (stationary two-time) auto-correlation and cross-correlation functions by

$$\phi_{xx}(\tau) = \mathrm{E}\left[x^*(t)\, x(t+\tau)\right] \tag{8.33}$$

$$\phi_{xy}(\tau) = \mathrm{E}\left[x^*(t)\, y(t+\tau)\right] \tag{8.34}$$

Both of which are independent of t. It should be emphasized again that these correlation functions are deterministic and are a property of the whole ensemble - they *do not* vary from one realization to another.

It is easy to see that if we pass a stationary stochastic process $x(t)$ through a filter with impulse response $h(t)$ yielding $y(t) = (x * h)(t)$, the two-time cross-correlation of x and y is

$$\begin{aligned}
\phi_{xy}(t_1, t_2) &= \mathrm{E}\left[x^*(t_1)\, y(t_2)\right] \\
&= \mathrm{E}\left[x^*(t_1) \int_{-\infty}^{\infty} h(\tau) x(t_2 - \tau)\, \mathrm{d}\tau\right] \\
&= \int_{-\infty}^{\infty} h(\tau) \mathrm{E}\left[x^*(t_1)\, x(t_2 - \tau)\right] \mathrm{d}\tau \\
&= \int_{-\infty}^{\infty} h(\tau) \phi_{xx}(t_2 - t_1 - \tau)\, \mathrm{d}\tau = (\phi_{xx} * h)(t_2 - t_1)
\end{aligned} \tag{8.35}$$

where we adopt the convention that a two-time correlation function with only one argument indicates an explicitly stationary correlation. Since $\phi_{xy}(t_1, t_2)$ is a function of $t_2 - t_1$ alone, it is in fact a stationary correlation and we can write

$$\boxed{\phi_{xy}(\tau) = (\phi_{xx} * h)(\tau)} \tag{8.36}$$

Similarly

$$\boxed{\phi_{yy}(\tau) = \left(\phi_{xy} * \tilde{h}\right)(\tau) = \left(\phi_{xx} * h * \tilde{h}\right)(\tau)} \tag{8.37}$$

where $\tilde{h}(t) = h^*(-t)$. Following the same line of argument as was presented for power correlation functions of deterministic signals, it should be clear that the Fourier transform $\Phi_{xx}(\nu)$ of the stationary auto-correlation function $\phi_{xx}(\tau)$ is the *power spectral density* of the stochastic process $x(t)$.

It is important to notice that the power spectrum of a stochastic process depends on the *correlation* between the noise values at different times and not on the probability density function of the noise at a given time. In particular, it is not necessary that noise have a Gaussian probability density at each time, although Gaussian noise is probably the most common. If Gaussian noise is passed through any linear filter, the output is still Gaussian noise but in general the spectrum of the output will be different from that of the input.

8.3.1 White Noise

A stationary noise process is said to be *white* if its power spectral density is constant. If the power per unit bandwidth is $N_0\,\mathrm{W\,m^{-2}}$, this corresponds to having

$$\Phi_{xx}(\nu) = \frac{N_0}{2} \tag{8.38}$$

The factor of two is present because a range of frequencies f_a to f_b over which the noise power is measured corresponds to the union of the intervals $[f_a, f_b]$ and $[-f_b, -f_a]$ in the two-sided power spectral density.

We see that the autocorrelation function of white noise is

$$\phi_{xx}(\tau) = \frac{N_0}{2}\delta(\tau) \tag{8.39}$$

Assuming that the process has zero mean, this shows that the values of the process at any two distinct times are uncorrelated. This represents an idealization since it implies that the instantaneous power $\phi_{xx}(0)$ is infinite so that the value of the process at each time can be arbitrarily large. This should not be too surprising since a flat power spectral density corresponds to infinite power when integrated over all frequencies.

In practice the power spectral density always tends to zero for sufficiently high frequencies. The model of white noise is useful if we are only interested in a frequency range over which the power spectral density is approximately constant.

Sampling a white noise process directly is not meaningful since the variance of each sample is infinite. For theoretical purposes, if we sample with a time interval of T, it is convenient to think of prefiltering the process using an ideal lowpass filter which cuts off at the Nyquist frequency $1/(2T)$. It can be shown that if zero-mean white noise with $\Phi_{xx}(\nu) = N_0/2$ is fed into such a filter, the samples taken at these times are *uncorrelated* random variables with zero mean and variance $N_0/(2T)$. In particular, if the input noise is Gaussian, the sample values are independent identically distributed Gaussian random variables.

Exercise: By considering the impulse response of the ideal low-pass filter and the autocorrelation of the output of the filter at lags kT for integer values of k, confirm that the samples are uncorrelated and have the variance stated above.

8.4 Signals in Noise

We are often interested in observing known signal pulses embedded within a stationary noise process. If $x(t)$ is the deterministic signal and $n(t)$ represents a realization of a noise process, the received signal $y(t)$ satisfies

$$y(t) = x(t) + n(t) \tag{8.40}$$

There are many ways of defining the signal to noise ratio (SNR) depending on the application, but for the purposes of signal detectability (i.e., seeing the signal within the noise), a convenient definition is

$$\mathrm{SNR} = \frac{\text{Peak signal power}}{\text{Total noise power}} \tag{8.41}$$

For example, if $x(t)$ is a toneburst with amplitude A, the peak signal power is A^2 and if the noise power spectral density is $\Phi_{nn}(\nu)$ the total noise power is its integral over all frequencies of interest.

8.4.1 Matched filtering

We wish to investigate what is the "best" filter to pass $y(t)$ through if we wish to optimize the detectability of the (known) signal $x(t)$ in the midst of noise $n(t)$ of known power spectral density $\Phi_{nn}(\nu)$. We shall assume that we are not interested in preserving the shape of $x(t)$ but wish to maximize the signal to noise ratio as defined above.

Let us suppose that the filter through which $y(t)$ is passed has impulse response $h(t)$. Using the usual convention that the function denoted by an upper-case letter refers to the Fourier transform of the function denoted by the lower-case letter, the filter output $g(t)$ is given by

$$g(t) = \int_{-\infty}^{\infty} H(\nu)Y(\nu)\exp(\mathrm{j}2\pi\nu t)\,\mathrm{d}\nu \tag{8.42}$$

We shall design the filter to give a maximum output at some time $t = t_0$. Since y is composed of signal x and noise n, we consider the effects on each of these separately. The peak signal power in the output is given by the peak of the filtered signal, i.e.,

$$\text{Peak signal power} = \left| \int_{-\infty}^{\infty} H(\nu)X(\nu)\exp(\mathrm{j}2\pi\nu t_0)\,\mathrm{d}\nu \right|^2 \tag{8.43}$$

The filtered noise has power spectral density $\Phi_{nn}(\nu)|H(\nu)|^2$. The total noise power in the output is

$$\text{Total noise power} = \int_{-\infty}^{\infty} \Phi_{nn}(\nu)|H(\nu)|^2\,\mathrm{d}\nu \tag{8.44}$$

Thus the signal to noise ratio which we wish to maximize is

$$\mathrm{SNR} = \frac{\left|\int_{-\infty}^{\infty} H(\nu)\,X(\nu)\exp(\mathrm{j}2\pi\nu t_0)\,\mathrm{d}\nu\right|}{\int_{-\infty}^{\infty} \Phi_{nn}(\nu)\,|H(\nu)|^2\,\mathrm{d}\nu} \tag{8.45}$$

To carry out this maximization, recall the *Cauchy-Schwarz inequality* which states that for any two functions $U(\nu)$ and $V(\nu)$,

$$\left|\int_{-\infty}^{\infty} U(\nu)V(\nu)\,\mathrm{d}\nu\right|^2 \le \int_{-\infty}^{\infty} |U(\nu)|^2\,\mathrm{d}\nu \int_{-\infty}^{\infty} |V(\nu)|^2\,\mathrm{d}\nu \tag{8.46}$$

where equality holds if and only if $U(\nu) = kV^*(\nu)$ where k is a constant. If we identify

$$U(\nu) = \sqrt{\Phi_{nn}(\nu)}\,H(\nu) \tag{8.47}$$

$$V(\nu) = \frac{X(\nu)\exp(\mathrm{j}2\pi\nu t_0)}{\sqrt{\Phi_{nn}(\nu)}} \tag{8.48}$$

and substitute into the Cauchy-Schwarz inequality, we find after minor rearranging that

$$\mathrm{SNR} = \frac{\left|\int_{-\infty}^{\infty} H(\nu)X(\nu)\exp(\mathrm{j}2\pi\nu t_0)\,\mathrm{d}\nu\right|^2}{\int_{-\infty}^{\infty} \Phi_{nn}(\nu)|H(\nu)|^2\,\mathrm{d}\nu} \le \int_{-\infty}^{\infty} \frac{|X(\nu)|^2}{\Phi_{nn}(\nu)}\,\mathrm{d}\nu \tag{8.49}$$

We notice that the right hand side is independent of $H(\nu)$ and gives an upper bound for the attainable signal to noise ratio. The signal to noise ratio is maximized when equality holds. This happens if

$$\sqrt{\Phi_{nn}(\nu)}\,H(\nu) = k\left[\frac{X(\nu)\exp(\mathrm{j}2\pi\nu t_0)}{\sqrt{\Phi_{nn}(\nu)}}\right]^* \tag{8.50}$$

or

$$H(\nu) = \frac{kX^*(\nu)\exp(-\mathrm{j}2\pi\nu t_0)}{\Phi_{nn}(\nu)} \tag{8.51}$$

The maximum signal to noise ratio is then given by

$$(\mathrm{SNR})_{\max} = \int_{-\infty}^{\infty} \frac{|X(\nu)|^2}{\Phi_{nn}(\nu)}\,\mathrm{d}\nu \tag{8.52}$$

Equations (8.51) and (8.52) define the *Middleton-North matched filter* which is of importance in the design of signal detection systems (e.g., radar and sonar). Notice how the amplitude of the frequency response of the filter $|H(\nu)|$ is large where $|X(\nu)|$ is large and $P_{nn}(\nu)$ is small. Intuitively, the filter lets through the signal and rejects the noise as well as it can.

Now consider the special case of white stationary noise for which $\Phi_{nn}(\nu) = N_0/2$. In this case we find that

$$H(\nu) = \frac{2k}{N_0}X^*(\nu)\exp(-\mathrm{j}2\pi\nu t_0) \tag{8.53}$$

and

$$(\mathrm{SNR})_{\max} = \frac{2}{N_0}\int_{-\infty}^{\infty} |X(\nu)|^2\,\mathrm{d}\nu \tag{8.54}$$

1. We see that the optimum signal to noise ratio is proportional to the signal energy and inversely proportional to the noise power spectral density.
2. The filter impulse response is the inverse Fourier transform of $H(\nu)$. We see that

$$h(t) = \frac{2k}{N_0}x^*(t_0 - t) \tag{8.55}$$

   This is just the signal $x(t)$ conjugated, reversed in time and delayed by t_0. The value of k is arbitrary since it only affects the overall amplitude and not the signal to noise ratio.
3. As a result of passing the signal component through the matched filter, its shape is changed. In fact

$$(x * h)(t) = \frac{2k}{N_0}(x(t) * x^*(t_0 - t)) = \frac{2k}{N_0}\phi^e_{xx}(t - t_0) \tag{8.56}$$

   Thus in the absence of noise, the output of the matched filter is a scaled and delayed version of the energy autocorrelation function of the signal. The peak of this is at t_0.
4. The matched filter can be implemented digitally using a shift register as a delay line to carry out the convolution.

When choosing a signal for matched filtering, it is usually best to find one with a highly peaked autocorrelation. If the time of arrival of the signal is not known (e.g. an echo from a target), the peak of the matched filter output may be used to give an estimate. A signal which is often used is a chirp which has a linearly swept carrier frequency.

$$\nu = \nu_0 + \mu t \tag{8.57}$$

The phase is the integral of the frequency with respect to time and so the form of the chirp could be

$$x(t) = \sin\left[2\pi\left(\nu_0 t + \frac{1}{2}\mu t^2\right)\right] \tag{8.58}$$

The following MatLab code illustrates matched filtering with a chirp signal. The function `filter` carries out the discrete convolution using a digital filter.

```
%
% MATCHED.M illustrates matched filtering with the Middleton-North filter
%
t = linspace(0,1,256);
s = sin(2*pi*(2*t+5*t.^2));   % Set up chirp
%
tbase = [0:1023]*(t(2)-t(1));
x = zeros(size(tbase));
x(1:length(s)) = s;           % Put chirp at start of frame
h = conj(s(length(s):-1:1));  % Impulse response of matched filter
%
y = x + randn(size(x));       % Corrupt signal with uncorrelated noise
%
% Show result with no noise
%
g = filter(h,1,x);             % Matched filter
subplot(4,1,1); plot(tbase,x); % Clean signal
title('Noise-free signal');
subplot(4,1,2); plot(tbase,g); % Matched filter output
title('Matched filtered output');
%
% Matched filtering of noisy signal
%
g = filter(h,1,y);             % Matched filter
subplot(4,1,3); plot(tbase,y); % Noisy signal
title('Noisy signal');
subplot(4,1,4); plot(tbase,g); % Matched filter output
title('Matched filtered output');
```

The results of this program are shown in Figure 8.1. The signal is clearly seen in the midst of the noise after applying the matched filter.

8.4.2 Coherent Averaging

A very simple yet powerful technique for improving the signal-to-noise ratio when investigating the response of a time-invariant system to a signal is the process of *coherent averaging.* A repetitive signal is sent in time frames and the responses are added together coherently.

Suppose that the peak signal amplitude is A and the noise power is σ^2. The signal to noise ratio for one frame is then

$$(\mathrm{SNR})_{1\ \mathrm{frame}} = A^2/\sigma^2 \tag{8.59}$$

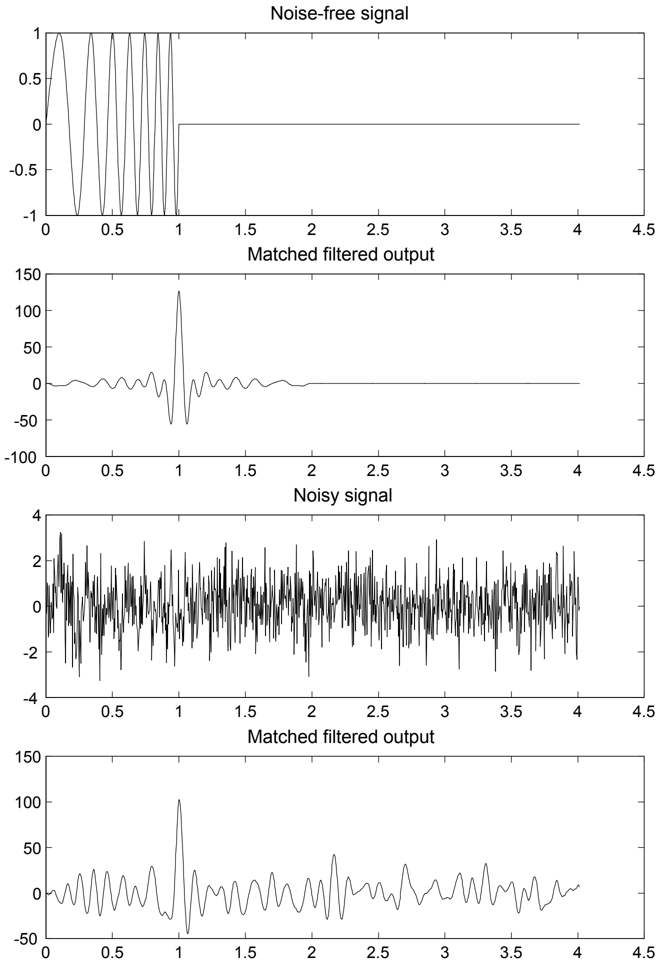

Figure 8.1 Middleton-North matched filter

As a result of adding together L frames, the signal *amplitudes* add because they are coherent whereas the noise *powers* add because they are incoherent. The signal to noise ratio after L frames is thus

$$(\mathrm{SNR})_{L \text{ frames}} = (LA)^2/(L\sigma^2) = L \times (\mathrm{SNR})_{1 \text{ frame}} \tag{8.60}$$

If L can be made large (this usually depends on the timescale on which the system may be regarded as being time-independent), we can get substantial enhancement in the signal to noise ratio.

The ratio of signal to noise *energies* in a time frame can be estimated by a process called double-period averaging. If two frames are of duration T and contain $y_1(t)$ and $y_2(t)$ we find the energies

$$E_+ = \int_0^T |y_1(t) + y_2(t)|^2 \,\mathrm{d}t \tag{8.61}$$

$$E_- = \int_0^T |y_1(t) - y_2(t)|^2 \,\mathrm{d}t \tag{8.62}$$

We assume that the contents of the frames can be written as a deterministic signal together with noise which is different between the frames, i.e.,

$$y_1(t) = x(t) + n_1(t) \quad \text{and} \quad y_2(t) = x(t) + n_2(t) \tag{8.63}$$

Then if E_s denotes the signal energy and E_n the expected noise energy, we see that the expected values of E_+ and E_- are

$$E_+ = 4E_s + 2E_n \quad \text{and} \quad E_- = 2E_n \tag{8.64}$$

Hence

$$\frac{E_s}{E_n} = \frac{1}{2}\left(\frac{E_+}{E_-} - 1\right) \tag{8.65}$$

This may then be used to estimate the signal to noise ratio.